

# Homogenous Base Rates for Malingering in Neuropsychological Examination of Litigants

Ronald M. Ruff<sup>1</sup> · Jasmin Klopfer<sup>1</sup> · Jeremy Blank<sup>1</sup>

Received: 30 December 2015 / Accepted: 17 May 2016 / Published online: 26 May 2016  
© Springer Science+Business Media New York 2016

**Abstract** Base rates for malingering are often obtained and averaged across multiple clinicians who apply heterogeneous methods for detection (Mittenberg et al., *J Clin Exp Neuropsychol* 24: 1094–1102, 2002; Young, *Psychol Inj Law* 8: 200–218, 2015). Our aims of obtaining homogenous base rates included the following: (a) evaluation of all our legal cases in accordance with the guidelines set forth in the position papers by both the National Academy of Neuropsychology and the Association for Psychological Advancement in Psychological Injury and Law, (b) minimal variation between our comprehensive neuropsychological examinations, and (c) determination of base rates of failed effort in 150 consecutively examined legal cases in one medical setting. To assess the various levels of volitional exaggeration, we introduced four gradations of poor effort definitions, which relied on performance validity tests (PVTs). A comparison between two consecutive samples of 75 litigants indicated less frequent poor effort with increasingly more conservative criteria. In our analysis of a subset of litigants who sustained traumatic brain injuries ( $N=115$ ), the four base rates for mild versus moderate-severe TBI groups were equivalent for the two more lenient malingering definitions but varied for the two more conservative definitions. Specifically, for the mild TBI cases investigated, the percentage of three PVT failures (or one PVT failure significantly below chance) arrived at 3.4 %. Our final aim was to compare the base rates of poor effort that were obtained with PVTs to the base rates of emotional and physical symptom endorsement, which were obtained with symptom validity tests (SVTs). No significant

correlations emerged in this analysis. The discussion emphasizes the relatively lower base rates of poor effort found in the convenience sample studied in neuropsychological evaluations relative to the higher estimates in the literature (40 +/- 10 %, Larrabee et al., *Clin Neuropsychol* 23: 841–849, 2009) but not others based on comprehensive review (Young, *Psychol Inj Law* 8: 200–218, 2015).

**Keywords** Malingering · Base rates · Neuropsychological assessment

## Introduction

Just a few decades ago, the assessment of malingering in neuropsychological examinations was both rare and largely idiosyncratic. Slick, Sherman and Iverson (1999) authored a seminal paper that set forth carefully chosen and balanced guidelines for the assessment of malingering and thus laid the groundwork for a more comprehensive approach to the question. Instead of an all-or-none approach, they proposed a diagnostic framework that allowed for gradations of “possible,” “probable,” and “definite” malingering of neurocognitive dysfunction. Incorporating these concepts, the National Academy of Neuropsychology published in 2005 the first official position paper on malingering, which stated the following:

Symptom exaggeration or fabrication occurs in a sizeable minority of neuropsychological examinees, with greater prevalence in forensic contexts. Adequate assessment of response validity is essential in order to maximize confidence in the results of neurocognitive and personality measures and in the diagnoses and recommendations that are based on the results. Symptom

✉ Ronald M. Ruff  
ronruff@mindspring.com

<sup>1</sup> University of California at San Francisco, San Francisco, CA, USA

validity assessment may include specific tests, indices, and observations... Assessment of response validity, as a component of a medically necessary evaluation, is medically necessary. When determined by the neuropsychologist to be necessary for the assessment of response validity, administration of specific symptom validity tests are also medically necessary. (Bush, Ruff, Tröster, Barth, Koffler, Pliskin & Silver, 2005, p. 419)

Over the years, other neuropsychological organizations have reinforced the medical necessity of including symptom validity testing and performance validity testing in neuropsychological examinations. For example, the American Academy of Clinical Neuropsychology published a “Consensus Conference Statement on the Neuropsychological Assessment of Effort, Response Bias, and Malingering” (Heilbronner, Sweet, Morgan, Larrabee, Millis & Participants, 2009). Similarly, the Association for Scientific Advancement in Psychological Injury and Law published their official position entitled “Psychological Assessment of Symptom and Performance Validity, Response Bias, and Malingering” (Bush, Heilbronner & Ruff, 2014). Thus, a consensus has emerged that mandates the assessment of test takers’ efforts as essential parts of both neuropsychological and psychodiagnostic evaluations. Note that performance validity tests (PVTs) are ones that typically are embedded in neuropsychological assessments, but they can be separate stand-alone ones as well. Also, symptom validity tests (SVTs) typically are separate scales in omnibus and personality tests that examine response bias, such as negative ones indicate possible exaggeration, feigning, and malingering.

Young (2015) published a comprehensive review of studies that have attempted to capture the base rate, or percentage, of individuals who “exaggerated or malingered” during neuropsychological examinations. These base rates vary significantly. For example, Trueblood and Schmidt (1993) found a base rate of 7.5 % in a sample of 106 patients who sought a neuropsychological evaluation. In this study, malingering was only diagnosed when performance on validity testing fell below chance. Much higher base rates have been reported when various gradations of malingering were examined. For example, Youngjohn, Burrows and Erdal (1995) classified 48 % of their litigants as having questionable motivation and 33 % as having insufficient effort. Base rates between 23 and 35 % were captured by Green, Rohling, Lees-Haley and Allen (2001). Particularly high rates of malingering have been documented in individuals with mild traumatic brain injuries (mTBI). For example, Larrabee (2003) reviewed 11 studies that published base rates and concluded that the overall prevalence rate was 40 % in 1363 compensation-seeking patients who presented as having sustained mild TBIs. In the largest study to date, Mittenberg, Patton, Canyock and Condit (2002) solicited the membership of the American Board of Clinical

Neuropsychology for their percentages of cases that, during the past year, were judged to be “malingering.” Base rates of probable malingering or symptom exaggeration were determined from survey responses of 131 neuropsychologists in active practice. A total sample of 6371 cases who claimed personal injuries was collated, and the base rates for malingering were determined. Two key issues stood out in this study. First, highly variable methods were used to determine poor effort. Second, within the subsample of individuals diagnosed with TBI, significantly different base rates were found for patients diagnosed with mild TBIs and the moderate-severe TBI group; that is, the reported base rate for symptom exaggeration was 38.50 % for mild TBI patients, whereas for the moderate and severe TBI patients it was only 8.82 %.

Larrabee, Millis and Meyers (2009) famously captured the base rate of malingering in the forensic neuropsychological context as 40 +/- 10 %. However, Young (2015) pointed out inconsistencies in the literature on the question and argued the rate could be more like 15 +/- 15 % in forensic disability assessments, although admittedly higher for mTBI/postconcussive cases.

Across, and often even within, the studies referenced above, heterogeneous definitions and procedures were used for determining malingering base rates. What also stood out was the great variation in the interpretations of PVTs. On the one hand, a score below the cutoff on one (or more) PVTs might be sufficient to determine poor effort whereas, on the other hand, malingering could be defined as a below-chance score on one (or more) PVTs. Heterogeneous instructions are also common; as an example, the above-mentioned position paper by the National Academy of Neuropsychology (NAN; Bush et al., 2005) posits that examiners should: “Inform the examinee at the outset of the evaluation and as needed during the evaluation that good effort and honesty will be required (the examiner may inform the examinee that such factors will be directly assessed).” We followed this guideline as well as all the other guidelines (see Appendix 1). However, we know many neuropsychologists who do not follow these recommendations, which can potentially affect base rates.

Thus, our first aim was to examine base rates of poor effort on PVTs that were obtained in a highly homogenous and clearly standardized manner. That is, in our examinations all the instructions were provided in accordance with both the NAN Position Paper’s Guidelines (Bush et al., 2005) and those of the Association for Psychological Advancement in Psychological Injury and Law (Bush et al., 2014). As recommended in these position papers, we placed our emphasis on the data that were derived from multiple PVTs. Furthermore, in each case, we relied (although, as recommended in the NAN Guidelines) to a lesser degree on our subjective impressions and clinical judgment. This included, for example, screening for blatant symptom exaggerations, striking inconsistencies within the neuropsychological test battery or

examinee behavior, contradictions between examinations of different professionals, and/or atypical suspicious behaviors during the examination. The second aim was to examine if base rates across our four grades of poor effort varied between two samples of 75 consecutively evaluated individuals, all of whom were similarly referred in the medical-legal context. Note that in our practice all of these individuals were evaluated in a fairly homogenous manner (e.g., same neuropsychological test battery, same individuals responsible for picking up on subjective indicators of malingering). Since the majority of our patients sustained TBIs, our third aim was to compare the base rates of litigants with diagnosed mild versus more moderate-severe TBIs. Our fourth aim was to examine, within two standardized psychodiagnostic tests, the extent to which our litigants' clinical complaints were exaggerated.

After our data of 150 cases were analyzed, we found that in all of the cases we had clinically, and thus subjectively, determined that malingering was likely, we also consistently found that these same individuals failed one or more of our PVTs. Given this complete overlap, we decided that there was no need for us to engage in a detailed compilation of these subjective indicators. Instead, we focus our study on determining the base rates according to performances on the data derived from the multiple PVTs that were administered to each litigant. Note that although in research, focusing on PVTs is admissible if there is a complete overlap with subjective indicators for each examinee, no neuropsychologist should utilize this approach clinically where it is essential to rely on both objective and subjective indicators.

## Method

### Subjects

Based on our exclusion criteria (i.e., below the age of 18, non-English speaking, or unable to complete a standard neuropsychological examination due to global aphasia, severe amnesia, or blindness), 43 individuals needed to be dropped from 193 consecutively examined cases; therefore, we obtained in our study a sample of 150 consecutively examined cases. All of these individuals claimed to be injured by a third party and thus were referred for forensic neuropsychological examinations to an outpatient clinic in San Francisco. Plaintiff attorneys referred a majority of the cases (86.7 %), and defense attorneys sent a minority of the individuals (13.3 %). Despite this discrepancy, the potential existed for all of these litigants to magnify or fake their symptoms, and moreover, we tested for discrepancies in the base rates of poor effort, as per below.

Our sample was comprised of 63 women and 87 men. Their average age was 46.7 years and ranged from 18 to 78 years ( $SD = 15.8$  years). The average education was 14.6 years and ranged between 8 and 22 years

( $SD = 14.6$  years). Within our sample of 150 litigants, 59.3 % of individuals sustained mild TBIs, 17.3 % moderate to severe TBIs, and 23.4 % were diagnosed with other types of brain damage (e.g., carbon monoxide poisoning, accidental overmedicating).

### Assessing Malingering or Intentional Exaggeration for Financial Gain

As pointed out by Rogers (1988), a diagnosis of “malingering,” per se, cannot be objectively diagnosed without having a competent test-taker, after having failed one or more PVTs and/or volitionally exaggerating symptoms, admit that he or she intentionally withheld effort and exaggerated for personal financial gain. However, such an admission is exceedingly rare and was not the case in any of our 150 litigants. Therefore, we decided not to use the terms “malingering” or even “exaggeration” in this article, and kept to the immediate influence of failing a PVT since these terms are most often used in the literature, to signify that such a failure indicates poor or suboptimal effort as more likely than not, even in the absence of an overt admission by the test-taker.

### Test Procedure

A comprehensive neuropsychological test battery was administered that consistently included three PVTs and one or more SVTs. Moreover, all patients were evaluated in the medical-legal context according to the guidelines published in a NAN Position Paper, which are contained in Appendices 1 and 2. In sum, all of our patients were informed that it was important that they try their very best and, if they needed any additional breaks, to let us know.

### Performance Validity Tests (PVTs)

As recommended by the NAN guidelines, we interspersed multiple PVTs throughout the test battery. Because some of our patients had been previously tested and PVTs had already been administered, we at times varied our test selection of administering at least three PVTs. However, most often, we administered the Test of Memory Malingering (Tombaugh, 1996), the Dot Counting Test (Boone, Lu & Herzberg, 2002), and the Rey Fifteen Item Test (Rey, 1964), and less frequently we gave the Medical Symptom Validity Test (Green, 2004), the Reliable Digit Span (Wechsler, 2009a), or the Word Choice Subtest of the WAIS-IV Advanced Clinical Solutions (Wechsler, 2009b). The cut scores used to determine PVT failure were taken from the test manuals. They were as follows, respectively: 44/50 for the TOMM; 21 for the DCT; 7/15 for the RFIT; 82.5 % for the MSVT; 6 digits for Reliable Digit Span; and 39/50 for the ACS.

## Symptom Validity Tests (SVTs)

Our test battery also included two psychodiagnostic tests, namely the Millon Clinical Multiaxial Inventory (MCMI-III; Millon, Millon, Davis & Grossman, 2009) and the Ruff Neurobehavioral Inventory (RNBI; Ruff & Hibbard, 2003), both of which contain respondent validity scales.

## Determining Base Rates

As stated in the NAN guidelines, data from SVTs/PVTs “should generally be given substantially greater weight than subjective indicators of suboptimal effort. Subjective indicators, such as examinee statements and examiner observations, should be afforded less weight due to the lack of scientific evidence supporting their validity (see Appendix 1). As noted above, we found that in our sample of 150 consecutive examined cases all individuals who exhibited subjective indicators of malingering also failed one or more SVT or PVT. Thus, our data presented below focuses only on the quantifiable and objective performances gathered from our validity testing.

Instead of relying on an all-or-none approach, we relied on the above-mentioned guidelines by Slick et al. (1999) that proposed a diagnostic framework, which allowed for gradations of “possible,” “probable,” and “definite” malingering of neurocognitive dysfunction. In this section of the paper, we keep their language referring to malingering, but remind readers that more than PVT analysis is needed to arrive at an attribution of malingering. In addition to these three gradations, we added a fourth, “highly sensitive,” category, which in our experience some clinicians use. These four gradations were defined as follows:

1. *Highly Sensitive*: This classification applied for failing any part of a single PVT. For example on the TOMM test, if one or more of the three trials fell below 45, then the criteria for suboptimal effort was met. Note that after a time delay the patient repeats the TOMM and, according to the author of the test, the following applies: if the delayed score— independent of the previous scores—is 45 or higher, the patient passes this test. Thus, in this category, the aim was to flag any score below 45 across the three trials to achieve an utmost sensitivity to possible poor effort, one that some clinicians may find useful.
2. *Possible*: Any individual who failed one but passed the other two PVTs was classified as possibly malingering.
3. *Probable*: Those that failed two PVTs and passed one PVT were identified as probably malingering (i.e., “more likely than not”).
4. *Definite*: Those that failed three PVTs or, alternatively, scored below chance on one or more PVTs were classified as definitively malingering (i.e., “highly likely”).

Given that all of our neuropsychological examinations included psychodiagnostic measures, we also examined what portion of our litigants scored above the scale cutoffs of these SVTs used, which can indicate exaggerations in their rating of clinical status.

## Results

### Comparison of Four Different Gradations/Definitions of Malingering

Out of our total sample of 150 individuals, the percentage of patients that was identified according to our four gradations of poor effort varied significantly ( $\chi^2(3)=122.690$ ,  $p < .001$ , Cochran  $Q$  Test; see Table 1). Whereas the frequencies between the *definite* and the *probable* definitions did not differ significantly,  $p = .125$ , the frequencies for the *probable* and the *possible* definitions did differ significantly,  $p < .001$ . The difference in frequency between the *possible* and the *highly sensitive* definition was significant as well,  $p < .001$ . Note that the  $\alpha$ -level used was corrected for multiple testing with a Bonferroni correction. The new resulting  $\alpha$ -level was  $\alpha = .0167$ . Table 1 indicates that the base rate for definite malingering according to the conservative definition provided was only 2 %. These results are a far cry from the one advocated by Larrabee et al. (2009), but are consistent with other studies as found in Young (2015).

Although plaintiff attorneys referred the majority of our cases (86.7 %) compared to the referrals by defense attorneys (13.3 %), the base rates of suboptimal effort between the two referral sources was insignificant (Fisher’s Exact Test,  $p = .594$ ). This indicates that the low base rate of definite malingering found in the present sample was not due to referral bias.

### How Reliable Were the Base Rates?

To examine the stability of our base rates, we first analyzed our initial 75 consecutively examined cases, and thereafter

**Table 1** Base rate comparison of poor effort on performance validity tests

Grades of poor effort (malingering)	Base rates of total sample (N=150)
Highly sensitive	38.0 %
Possible	21.3 %
Probable	4.7 %
Definite	2.0 %

Highly sensitive = failure on any one sub-part within any of the three PVTs; Possible = failure on one PVT according to the test’s recommended cutoffs; Probable = failure on two PVTs; Definite = failure on three PVTs or, alternatively, a below-chance performance on any PVT



compared these base rates to our subsequent sample of 75 consecutively examined individuals. As noted above, all of these individuals were examined in the same manner (i.e., encouraged to do their best, informed that trying their best was important). According to our four grades of effort, the percentage of patients with suboptimal effort as defined across four levels was comparable in both groups (see Table 2).

### Comparison of Base Rates in Patients with Mild Vs. Moderate-Severe TBIs

Within our total sample of 150 individuals, 115 sustained TBIs. The majority, or 89, individuals (77.4 %) sustained mild TBIs and 26 (22.6 %) were diagnosed with moderate to severe TBIs. The difference in the group sizes is fairly consistent with the incidence rates, since 80 % of all TBI patients have a mild TBI, with 10 % having a moderate TBI and another 10 % having a severe TBI (Bruns & Hauser, 2003). Table 3 compares the percentages according to the four gradations separately for the mild versus the moderate-to-severe TBI subgroups.

Our data did not support the common impression that mild TBI patients have consistently higher rates of poor effort across the board. Instead, when subdivided according to combining the *highly sensitive* and the *possible* grades, the two severity groups were more or less equivalent (with even a mild inverse trend of higher rates in the moderate to severe group). However, according to both of the more stringent grades (*probable* and *definite*), only mild TBI patients were identified as having malingered, which is aligned with the findings reported in the literature. We do acknowledge, however, that our moderate-to-severe sample was quite small. The percentage of mild TBI cases that were classified as definite malingerers according to the present definition is quite small (3.4 %) relative to the higher estimates in the literature (e.g., Larrabee et al. 2009). Granted procedures, tests, and definitions are not comparable across all studies, so some variation is expected, but in the present case, the lower estimate of the two mentioned seems more reliable because of the care taken in conducting the study.

**Table 2** Stability of base rates in two samples of 75 patients

Grades of poor effort (malingering)	First sample (N=75)	Second sample (N=75)
Highly sensitive	36.0 %	40.0 %
Possible	20.0 %	22.7 %
Probable	6.7 %	2.7 %
Definite	1.3 %	2.7 %

**Table 3** Comparisons between mild vs. moderate-severe TBI patients

Grades of poor effort (malingering)	Mild TBI (N=89)	Moderate-to-severe TBI (N=26)
Highly sensitive	36.0 %	42.3 %
Possible	20.2 %	26.9 %
Probable	6.7 %	None
Definite	3.4 %	None

### Poor Effort According to Psychodiagnostic Measures

As a rule, psychodiagnostic tests are administered in most forensic neuropsychological examinations. We follow this practice and typically administer the MCMI-III (Millon et al., 2009) as well as the RNBI (Ruff & Hibbard, 2003). Invalid response patterns can be captured in both the MCMI-III and RNBI when test takers endorse an unusually high number of severe symptoms that exceed a level endorsed by patients with genuine psychopathology and no financial incentives to exaggerate.

Unlike the MCMI-III, which was specifically developed to capture psychopathology, the RNBI is focused on biopsychosocial assessment, which is more pertinent for individuals who have sustained brain damage. Thus, in addition to examining the likelihood of emotional symptoms, the RNBI also assesses, in an equivalent manner, the patients' (a) physical, (b) cognitive, (c) emotional, and (d) quality of life statuses. The RNBI was standardized on a sample of over 1000 volunteers and then validated according to multiple patient groups. This included studies that examined the interactions across the four groups of scales among inpatients with spinal cord injuries (Murray, Asghari, Egorov, Rutkowski, Siddall, Soden & Ruff, 2007) and TBI patients with and without anger (Johansson, Jamora, Ruff & Pack, 2008), with and without pain (Jamora, Schroeder & Ruff, 2013), as well as with or without post-traumatic stress disorder (Schroeder, Ruff & Jäncke, 2015).

The RNBI is divided into two parts, the first assessing a respondent's current functioning levels and the second part soliciting the patient's status prior to his or her alleged illness or injury. Thus, the premorbid ratings can be compared with the patients' postmorbid status. This test also contains four validity scales to assess overly positive or negative ratings as well as infrequent and inconsistent response patterns for both the pre and post ratings.

In our sample of 150 litigants, 48 patients were assessed only with the Beck Depression and Anxiety Scales due to time constraints and neither of these screening tests included validity scales. All of the remaining 102 litigants completed both the MCMI-III and RNBI.

According to the RNBI Postmorbid Negativity Scale, 13.73 % of subjects exaggerated their postmorbid validity

scales but none exaggerated the premorbid validity scales. According to the Debasement scale included in the MCMI-III, 14.71 % exaggerated their symptoms. Note that SVT failures such as these do not automatically denote malingering. Once more, the full file must be consulted before any malingering attribution can be made in conclusion.

### Do the Concepts of PVTs and SVTs Overlap?

In our sample, 102 litigants completed both the MCMI-III and RNBI. On the RNBI Postmorbid Negativity Scale, 13.73 % exaggerated their problems and 5 % of these 102 litigants also concurrently exaggerated their responses on one of more PVTs. In the same sample, 14.71 % exaggerated their problems on the MCMI-III according to the Debasement scale and concurrently 4 % of these individuals also exaggerated their problems on one or more PVTs.

There are several recent findings that support the thesis that PVTs and SVTs are different concepts and thus do not overlap. For example, Dyke, Millis, Axelrod and Hanks (2013) found a three-factor model, based on confirmatory factor analyses, in which (a) cognitive performance, (b) SVT results and standard self-report, and (c) PVT results were separate factors. Based on this finding, they concluded that failure in symptom validity does not necessarily invalidate performance validity. Further, Ruocco, Swirsky-Sacchetti, Chute, Mandel, Platek and Zillmer (2008) found that the TOMM results are not correlated with the MCMI-III validity scales (Disclosure, Debasement, Desirability). A factor analysis of the TOMM Trial 1, the Reliable Digit Span (RDS; Greiffenstein, Baker & Gola 1994), and the MCMI-III validity scales led to a two-factor model that accounted for 67.4 % of the variance. The first factor loaded highly on the MCMI-III validity scales (accounting for 40.6 % of the variance); the second factor loaded on the TOMM Trial 1 and the Reliable Digit Span (accounting for 16.8 % of the variance). Thus, this study also did not find links between the performance and symptom validity measures. In contrast, Whiteside, Dunbar-Mayer and Waters (2009) found a significant correlation between TOMM values and the validity scales of the Personality Assessment Inventory (PAI; Morey, 2007). Their factor analysis led to two factors for cognitive and personality components, but the Negative Impression Management Scale had factor loadings on the cognitive as well as on the personality components; therefore, they concluded that, although significant, the relationship between the two concepts was modest. Finally, Haggerty, Frazier, Busch and Naugle (2007) also found a significant modest relationship between the Victoria Symptom Validity Test (VSVT; Slick, Hopp, Strauss & Thompson, 1997/2005) and the Negative Impression Management Scale of the PAI.

To conclude, though PVTs and SVTs generally measure separate concepts, one can find in the literature some

indications of a modest relationship. Nonetheless, the literature as well as our data suggests that it is inappropriate on the basis of an invalid SVT performance measure of clinical symptoms of personality to reach the conclusion that the neurocognitive test results are also unreliable, and vice versa.

It should be noted that in neuropsychological assessments, malingering or its absence cannot be confidently determined on the basis of PVTs alone. For example, it cannot be assumed that the neuropsychological findings are valid on the basis of a person passing 2/3 or even 3/3 PVTs. Or, when someone with previously normal cognitive functioning who has sustained an mTBI passes 2/3 or 3/3 PVTs, and in the same examination produces atypical, inconsistent, or substantially impaired performance on multiple neuropsychological tests, the issue of likely flawed effort must be examined. That being said, an examination of the whole clinical file and the pattern of the test data in the case needs to be considered in each case. This includes relying on relevant data in the scientific literature, which has provided what test performances are consistent versus inconsistent with an mTBI. For example, atypical cognitive scores can occur due to physical injuries, pre-existing emotional disorders, secondary gain, or any combination thereof, and thus not brain injury per se. In such complicated cases, PVTs have their limited role, and instead of using them solely, the decision making in such complex judgements need to rely on the full reliable data collected, and include the clinician's judgment.

## Discussion

### Summary

The overall aim of this study was to determine base rates for suboptimal and poor effort in a medicolegal sample of litigants that was assessed in a uniform manner. To achieve this uniformity, we took the following steps. First, all litigants were examined in adherence with the guidelines and methods that are delineated in the position papers by the National Academy of Neuropsychology (Bush et al., 2005) and the Association for Scientific Advancement in Psychological Injury and Law (Bush et al., 2014). These guidelines recommend that the examiners “establish trust and rapport with all parties about the nature of the evaluation.” This includes (a) instructing each litigant on the importance of maintaining his or her best effort during the neuropsychological examination and (b) allow for additional breaks, if needed.

These NAN guidelines also encourage the implementation of and reliance on performance and symptom validity testing as well as the examiners' clinical judgments in the interpretation of neuropsychological test data and the whole data set collected in the examination. As noted in the introduction, we found to our surprise that in our 150 consecutively

examined litigants, every individual whose neuropsychological test scores were clinically judged to lack sufficient effort also failed one of more of either the PVTs and/or SVTs. Thus, none of our cases' poor effort was captured exclusively by examiner subjective impression.

As recommended by Slick et al. (1999), we introduced a diagnostic framework that allowed for three gradations of “possible,” “probable,” and “definite” malingering of neurocognitive dysfunction, and we added a fourth “highly sensitive” definition that some neuropsychologists use (e.g., Mittenberg et al., 2002). As expected, in the 150 litigants, and in our subset of 105 TBI litigants, we captured higher base rates for the “highly sensitive” and “possible” malingering definitions (ranging between 20 and 42 %), when compared with the “probable” and “definite” definitions (ranging between 0 and 6.7 %). In sum, different diagnostic definitions of malingering impact base rates in a substantial way.

Our second aim was to compare the four base rates in two samples of 75 litigants to determine their reliability. As noted above, all of these individuals were examined in the same manner (i.e., encouraged to do their best, informed that trying their best was important; and similar effort tests were administered). The percentage of patients with suboptimal and poor effort was comparable across four levels in the two groups (see Table 2).

Our third aim was to determine the base rates of suboptimal/poor effort for those litigants who sustained a mild TBI versus those with moderate-to-severe TBIs. As discussed above, the literature has reported significantly higher base rates of malingering for mild TBI patients when compared to patients with moderate and severe TBIs. It is important to note that our sample size for the moderate-to-severe TBI group was much smaller than that for the mTBI group. Nonetheless, by applying the four malingering gradations, the two TBI severity groups were equivalent for both the *highly sensitive* and the *possible* malingering definitions. In contrast, for the *probable* and *definite* definitions, the expected trend was confirmed; that is, litigants with mild TBIs are more likely to malingering when compared to moderate and severely brain injured individuals. Yet, again, it should be noted that our sample sizes were uneven between the two severity groups and thus this finding needs to be validated in future studies.

Our final aim was to compare two different types of SVTs, one assessing personality/psychopathology (the MCMI-III) and the other capturing the biospsychosocial status of the litigants (RNBI). Across our sample, the RNBI's postmorbidity validity scales and the MCMI-III's validity scales captured exaggerated profiles in patients with or without exaggerated PVTs. Thus, our SVT results are in line with the recent literature, in that we did not find a significant relationship between the measures of SVTs and PVTs so that they tap different aspects of negative response bias and one cannot stand as representing the other.

Further, we did a post hoc analysis concerning the RNBI postmorbidity scales for the Emotional Domain, the Cognitive Domain, the Physical Domain, and the Quality of Life Domain. A comparison of individuals with poor effort and individuals with good effort captured the following results. Poor effort was classified by failing three out of three PVTs (or one failure significantly below chance) and good effort was classified by passing all three PVTs. We found significantly elevated values for the poor effort group for all four scales. Four *t* tests for independent samples, which were corrected for multiple testing with a Bonferroni correction for a significance level of 5 %, resulted in significant differences ( $\alpha$ -level at  $\alpha = .125$ ). The comparison between the poor effort and the good effort groups for the RNBI Cognitive Domain Scale was significant ( $t(49) = 3.3, p = .002$ ), as was the comparison between the poor effort and the good effort groups for the RNBI Emotional Domain Scale ( $t(49) = 2.9, p = .005$ ). Further, the comparison between the poor effort and the good effort groups for the RNBI Physical Domain Scale was significant ( $t(49) = 3.5, p = .001$ ). Finally, the comparison between the poor effort and the good effort groups for the RNBI Quality of Life Domain Scale was also significant ( $t(49) = 3.1, p = .003$ ). In sum, those individuals who had insufficient effort (failed three PVTs) endorsed a significantly greater number of problems across all the domains of the RNBI when compared to those individuals who provided good effort (passed three PVTs).

### What Can Explain Variability Across Base Rate Studies?

Most of the base rate studies combine samples across many clinicians. In some studies, the respondents included self-selected clinicians who voluntarily submitted their base rates across a specified time frame, such as during the last year. However, when the base rates are derived from clinicians who are asked to report the percentages of their cases that malingered, the sample size can vary significantly and thus this can result in inaccurate base rates. Let us illustrate the error source according to an extreme example. If a clinician tested only three forensic cases during 1 year and all three met the criteria of malingering, then 100 % percent of the legal cases malingered. If a second clinician who tested six litigants determined that one malingered, then the percentage is 16.7 %. Thus, by averaging these two base rates, 58.3 % malingered, when in fact the base rate should have been for the combined sample of nine cases, at 44.4 %. Thus, averaging base rates across smaller sample sizes can potentially inflate or deflate the percentages of the base rates.

### What Is a Reasonable Base Rate?

In the most comprehensive literature review of base rates to date, Young (2015) concluded that the base rate for

malingering proposed by Larrabee (2012) of 40 % or greater was unreasonably high. Instead, he recommended after carefully examining and recalculating the data that a more realistic base rate is 15 %. Young also acknowledged that base rates vary due to the different methods and definitions that are applied, and this variation was estimated to range between 0 to 30 %.

With the exception of our “highly sensitive” base rate, all of the three more typically used base rates (i.e., “possible”, “likely” and highly likely”) fall within the range of  $15 \pm 15$  % that Young proposed. Thus, we concur with the NAN recommendation that the “highly sensitive” base rate should not be relied upon for determining malingering (i.e., Appendix 2 (d) states (with reference to SVTs, which used to be the term used for both SVTs and PVTs): Performance slightly below cutoff on one SVT may not justify an interpretation of biased responding; converging evidence from additional indicators may be required).

Finally, we postulate that our relatively low base rate of 6.8 % (4.7 % for probable and 2.0 % for likely) is in large part due to (a) our strict adherence with the above-mentioned position papers; (b) our homogeneous patient instructions and test methods; and (c) our strict reliance on four malingering grades, which helps differentiate likely and less likely malingering cases. However, future research is needed to confirm this finding by comparing if base rates tend to be lower when they are derived in a homogenous manner versus relying on more heterogeneous methods. Also, the lack of difference in malingering rates between plaintiff and defense cases tests in any one setting needs to be confirmed widely.

## Weaknesses

Our study contains a number of weaknesses. One weakness is our uneven TBI sample, because it is comprised of a much larger group of mild versus moderate-severe TBI litigants. As previously noted, the difference in the group sizes is, however, fairly consistent with the incidence rates in the literature, in that 80 % of all TBI patients have a mild TBI, 10 % have a moderate TBI, and another 10 % have a severe TBI (Bruns & Hauser, 2003). Nonetheless, more balanced sample sizes would have strengthened (a) the accuracy of the base rates for the moderate-severe TBI group and (b) the comparison between the different severity grades. Nonetheless, clinically we have noticed over the years that the more severely injured patients are often less concerned about the sufficiency of their deficits, especially if their neurocognitive problems are significant and documented consistently throughout the medical records by multiple healthcare providers.

A second weakness is our uneven referral base, which raised the following question: Is it feasible that plaintiff’s experts have lower base rates as compared to defense experts? Our study was comprised of 86.7 % of referrals by plaintiff

attorneys versus 13.3 % from defense attorneys. Thus, this issue could not be directly addressed with our data and in this sense represents a weakness of our study. What does the literature tell us about such imbalances? For example, Kaufmann and Greiffenstein (2013) stated:

“A simple ratio of plaintiff to defense cases is not compelling evidence for objectivity versus partisanship. It is a reality that a [forensic neuropsychologist’s] career trajectory increasingly attracts retention by one side or the other. There are many reasons for this including word of mouth, aggressive versus conservative neurodiagnostic approaches, and scientist-practitioner ethos versus pure clinical orientation” (p. 57).

Larrabee (2012) posited that if a fully trained neuropsychologist relies on scientifically based methods, then such a forensic expert should be equally conversant and willing to accept both defense and plaintiff cases. Lees-Haley (1999) commented on the value of aiming to achieve a balanced referral stream as follows: “the desirable and ideal 50–50 forensic referral pattern as an...unfounded but widely circulated myth” (p. 14). Lees-Haley (1999) further pointed out that testifying 50 % for plaintiff and 50 % defense is in itself not sufficient evidence of absence of bias, and stated: “This myth is a problem in the context of debiasing because it is used to imply lack of bias when ‘50–50’ may actually be evidence of just the opposite. The 50–50 myth is a classic case of an unexamined proposition that survives by repetition without critical review” (p. 14).

We concur with our colleagues that neuropsychologists have no direct control over their referral stream and most often do not have a balanced ratio between their referrals from defense and plaintiff attorneys. Nonetheless, forensic neuropsychologists can become typecast as either a “liberal plaintiff neuropsychologist” or “conservative defense expert.” However, the source of the referral does not automatically indicate a bias in and of itself. Thus, if the focus is on both our scientific methods and balanced interpretations, a neuropsychologist should be open to accept both plaintiff and defense cases. A 50–50 referral pattern is rare and not in itself proof of a lack of bias.

Nonetheless, it is the expert’s responsibility to examine his or her biases on an ongoing basis. This is stated as follows by Bush et al. (2014): “Because interpretation of invalid test performances is made by examiners, it is essential that examiners are aware of their own biases. This awareness can be achieved only if examiners make an effort to identify their biases. With biases identified, examiners can strive to reduce the effects of the biases on their opinions. Just as it is important for a psychotherapist to examine countertransference, it is imperative that examiners explore their thoughts and feelings about examinees, especially in the context of litigation (Ruff, 2009).”

The third weakness of the present study is borne out of a variability that exists between base rates because: (a) Multiple



definitions for malingering are used. In fact, we know of neuropsychologists who use no gradation at all, but diagnose *definite* malingering even if only one PVT of multiple PVTs falls below the cutoff score. (b) The application of different malingering tests also represents a likely source of variance. Clearly, some effort tests are more rigorous than others. Using three soft versus three rigorous tests can play a role. (c) Even if two neuropsychologists rely on the same gradation of degree of effort, but one relies on three PVTs to determining malingering attribution and another clinician relies also on three stand-alone PVTs but additionally calculates three or more embedded malingering indices, then the likelihood of diagnosing malingering may increase, as does the likelihood for false positives.

Finally, a potential weakness of this study is that we did not include analyses of clinical judgment or subjective impression but instead chose to focus on analyses of the data derived from PVTs and SVTs. However, because our focus was on accurate base rates based on gradations of malingering suboptimal/poor effort, there would have been no advantage of adding our subjective malingering indicators because in each case, among our participants with problematic clinical judgment about their effort, these individuals were also identified with our PVTs as being either below chance or having two PVTs that fell below the cutoffs.

## Conclusions and Future Considerations

How can we reduce the divergence that currently exists in neuropsychologist's determination of malingering, which in turn results in highly different estimates of base rates? A first step in the field was achieved by multiple neuropsychological organizations publishing position papers; thus, in many ways, a consensus has emerged for the standard assessment of effort in neuropsychology. All examiners should adhere to these guidelines and, if they alter them, then, the examiner should be able to justify any modification. The NAN position paper also states that neuropsychological examinations should include multiple PVTs and at least one SVT. However, none of the position papers have recommended an upper limit as to the number of effort tests that is reasonable. We suggest that future guidelines address potential trade-offs of administering more than three PVTs plus adding multiple embedded effort indicators.

Similar to neuropsychologists selecting different overall test batteries, a variation in the type and number of PVTs and SVTs selected will remain. However, in our opinion, standards should emerge that caution neuropsychologists to avoid using overly sensitive malingering indicators as conclusive evidence of malingering. This suggestion is supported by our study, where the “highly sensitive” malingering gradation

flagged possible malingering but fell short of capturing conclusive evidence of malingering.

When three PVTs are given, adhering to the rule of two or more positive PVTs for attributing likely malingering is, in our opinion, a preferable approach, with the exception of (a) when, on a single PVT, the score falls below chance and (b) concurrent subjective indicators buttress this interpretation. Further, three PVT failures out of three PVTs administered could result in attributing definite malingering, depending on the full set of reliable data in the file at hand. Conversely, if two effort tests are passed but on one the performance falls short of the cutoff, then the issue of variable effort is indicated and this should be stated accordingly. This conservative approach avoids false positives. Moreover, this approach is also supported by the fact that outliers do occur when testing non-litigants or even volunteers who have no financial incentives whatsoever on various neuropsychological tests. For example, we have found significant outliers in our sample of volunteers that were tested during the standardization that was required for the development of new tests. Similarly, Robert Heaton, who standardized the Halstead-Reitan Battery, also found outliers in non-clinical volunteers during the standardization of the battery (personal communication). Thus, it is reasonable to postulate that if outliers occur in volunteers then they can also occur in litigants.

**Acknowledgments** We wish to extend our thanks to Dr. Maryann Bens for her thorough editing of and commentary on this manuscript.

### Compliance with Ethical Standards

**Conflict of Interest** The authors have no conflict of interest to declare.

**Informed Consent** All procedures for archival data were followed in accordance with the ethical standards of St. Francis Medical Center.

**Animal Rights** No animal studies were carried out by the authors for this article.

## Appendix 1

### Method Used for Determining Poor Effort—Procedures

- (a) Remain abreast of trends in the symptom validity assessment literature. Approach the assessment of symptom validity proactively.
- (b) Assess whether cognitive, psychiatric, and/or behavioral symptoms are embellished.
- (c) Use a *multi-method* approach. A distinction is made between a *multi-method* approach and a *multi-test* approach. Whereas the administration of multiple tests may or may not contribute incrementally to the validity of the clinical determination, the use of multiple methods that extend beyond testing is likely to contribute to such validity.
- (d) Inform the examinee at the outset of the evaluation and as needed during the evaluation that good effort and honesty will be required.

(the examiner may inform the examinee that such factors will be directly assessed).

- (e) Use SVTs with the most appropriate psychometric properties, given the characteristics of the examinee and setting.
- (f) Disperse SVTs or measures with symptom validity indicators throughout the evaluation, with administration of at least one SVT early in the evaluation process.
- (g) Report the results of symptom validity assessment

Source: Bush et al. (2005).

## Appendix 2

### Further Methods Used for Determining Poor Effort

- (a) Data from SVTs should generally be given substantially greater weight than subjective indicators of suboptimal effort. Subjective indicators, such as examinee statements and examiner observations, should be afforded less weight due to the lack of scientific evidence supporting their validity. Invalid performance on a measure of personality does not allow for an a priori conclusion that the neurocognitive test results are also unreliable, and vice versa.
- (b) The examiner must consider the nature of the performance on SVTs and other evaluation findings when generalizing from the results of SVTs to other test results.
- (c) Strong evidence of invalid performance on SVTs or other indicators of symptom validity raise doubt about the validity of all neurocognitive test results. In the presence of invalid performance on measures or indices of symptom validity, interpretation of performances on other tests as valid would need to be justified.
- (d) When evidence of invalid performance exists, scores on cognitive ability tests may be interpreted as representing the examinee's minimum level of ability.
- (e) Performance slightly below cutoff on one SVT may not justify an interpretation of biased responding; converging evidence from additional indicators may be required.
- (f) If an evaluation that has been discontinued due to insufficient effort or invalid responding is later continued, the confidence that could be placed in the validity of the results would remain limited.

Appropriate probabilistic language based on the nature and extent of convergent evidence should be employed when offering explanations for symptom exaggeration or fabrication (see, for example, Slick et al., 1999). Vague or misleading terminology to describe invalid performance should be avoided.

Source: Bush et al. (2005).

## References

- Boone, K., Lu, P., & Herzberg, D. (2002). *The Dot Counting Test*. Los Angeles: Western Psychology Service.
- Bruns, J., & Hauser, W. A. (2003). The epidemiology of traumatic brain injury: A review. *Epilepsia*, *44*, 2–10.
- Bush, S. S., Heilbronner, R. L., & Ruff, R. M. (2014). Psychological assessment of symptom and performance validity, response bias, and malingering: Official position of the Association for Scientific Advancement in Psychological Injury and Law. *Psychological Injury and Law*, *7*, 197–205. doi:10.1007/s12207-014-9198-7
- Bush, S. S., Ruff, R. M., Tröster, A. I., Barth, J. T., Koffler, S. P., Pliskin, N. H., & Silver, C. H. (2005). Symptom validity assessment: Practice issues and medical necessity: NAN Policy & Planning Committee. *Archives of Clinical Neuropsychology*, *20*, 419–426. doi:10.1016/j.acn.2005.02.002
- Dyke, S. A. V., Millis, S. R., Axelrod, B. N., & Hanks, R. A. (2013). Assessing effort: Differentiating performance and symptom validity. *The Clinical Neuropsychologist*, *27*, 1234–1246. doi:10.1080/13854046.2013.835447
- Green, P. (2004). *Medical Symptom Validity Test (MSVT)*. Edmonton: Green's Publishing Inc.
- Green, P., Rohling, M. L., Lees-Haley, P. R., & Allen, L. M. A. (2001). Effort has a greater effect on test scores than severe brain injury in compensation claimants. *Brain Injury*, *15*, 1045–1060. doi:10.1080/02699050110088254
- Greiffenstein, M. F., Baker, W. J., & Gola, T. (1994). Validation of malingering amnesia measures with a large clinical sample. *Psychological Assessment*, *6*, 218–224.
- Haggerty, K. A., Frazier, T. W., Busch, R. M., & Naugle, R. I. (2007). Relationships among Victoria Symptom Validity Test indices and Personality Assessment Inventory validity scales in a large clinical sample. *The Clinical Neuropsychologist*, *21*, 917–928. doi:10.1080/13854040600899724
- Heilbronner, R., Sweet, J., Morgan, J., Larrabee, G., Millis, S., & Participants, C. (2009). American Academy of Clinical Neuropsychology Consensus Conference Statement on the neuropsychological assessment of effort, response bias, and malingering. *The Clinical Neuropsychologist*, *23*, 1093–1129. doi:10.1080/13854040903155063
- Jamora, C. W., Schroeder, S. C., & Ruff, R. M. (2013). Pain and mild traumatic brain injury: The implications of pain severity on emotional and cognitive functioning. *Brain Injury*, *27*, 1134–1140. doi:10.3109/02699052.2013.804196
- Johansson, S. H., Jamora, C. W., Ruff, R. M., & Pack, N. M. (2008). A biopsychosocial perspective of aggression in the context of traumatic brain injury. *Brain Injury*, *22*, 999–1006. doi:10.1080/02699050802530573
- Kaufmann, P. M., & Greiffenstein, M. F. (2013). Forensic neuropsychology: Training, scope of practice, and quality control. *National Academy of Neuropsychology Bulletin*, *27*, 11–15.
- Larrabee, G. J. (2003). Detection of malingering using atypical performance patterns on standard neuropsychological tests. *The Clinical Neuropsychologist (Neuropsychology, Development and Cognition: Section D)*, *17*, 410–425. doi:10.1076/clin.17.3.410.18089
- Larrabee, G. J. (2012). Performance validity and symptom validity in neuropsychological assessment. *Journal of the International Neuropsychological Society*, *18*, 625–630. doi:10.1017/S1355617712000240
- Larrabee, G. J., Millis, S. R., & Meyers, J. E. (2009). 40 plus or minus 10, a new magical number: Reply to Russell. *The Clinical Neuropsychologist*, *23*, 841–849.
- Lees-Haley, P. R. (1999). Commentary on Sweet and Moulthrop's debiasing procedures. *Journal of Forensic Neuropsychology*, *1*(3), 43–47. doi:10.1300/J151v01n03\_04
- Millon, T., Millon, C., Davis, R., & Grossman, S. (2009). *MCMIII: Millon Clinical Multiaxial Inventory-III*. San Antonio: Pearson.
- Mittenberg, W., Patton, C., Canyock, E. M., & Condit, D. C. (2002). Base rates of malingering and symptom exaggeration. *Journal of Clinical and Experimental Neuropsychology (Neuropsychology, Development and Cognition: Section A)*, *24*, 1094–1102. doi:10.1076/jcen.24.8.1094.8379
- Morey, L. C. (2007). *Personality assessment inventory: Professional manual* (2nd ed.). Lutz: Psychological Assessment Resources.
- Murray, R. F., Asghari, A., Egorov, D. D., Rutkowski, S. B., Siddall, P. J., Soden, R. J., & Ruff, R. (2007). Impact of spinal cord injury on self-

- perceived pre- and postmorbid cognitive, emotional, and physical functioning. *Spinal Cord*, 45, 429–436. doi:10.1038/sj.sc.3102022
- Rey, A. (1964). *The clinical examination in psychology*. Paris: Presses Universitaires de France.
- Rogers, R. (1988). *Clinical assessment of malingering and deception*. New York: The Guilford Press.
- Ruff, R. (2009). Best practice guidelines for forensic neuropsychological examinations of patients with traumatic brain injury. *The Journal of Head Trauma Rehabilitation*, 24, 131–140.
- Ruff, R. M., & Hibbard, K. M. (2003). *RNBI: Ruff Neurobehavioral Inventory*. Lutz: PAR.
- Ruocco, A. C., Swirsky-Sacchetti, T., Chute, D. L., Mandel, S., Platek, S. M., & Zillmer, E. A. (2008). Distinguishing between neuropsychological malingering and exaggerated psychiatric symptoms in a neuropsychological setting. *The Clinical Neuropsychologist*, 22, 547–564. doi:10.1080/13854040701336444
- Schroeder, S. C., Ruff, R. M., & Jäncke, L. (2015). Posttraumatic stress disorder exacerbates emotional complaints but not cognitive impairments in individuals suffering from postconcussional disorder after mild traumatic brain injury. *Zeitschrift Für Neuropsychologie*, 26, 35–50. doi:10.1024/1016-264X/a000132
- Slick, D. J., Hopp, G., Strauss, E., Thompson, G. B. (1997/2005). *Victoria Symptom Validity Test: Professional manual*. Odessa, FL: Psychological Assessment Resources.
- Slick, D. J., Sherman, E. M., & Iverson, G. L. (1999). Diagnostic criteria for malingering neurocognitive dysfunction: Proposed standards for clinical practice and research. *The Clinical Neuropsychologist*, 13, 545–561.
- Tombaugh, T. N. (1996). *TOMM Test of Memory Malingering*. North Tonawanda: Multi-Health Systems.
- 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. doi:10.1080/01688639308402580
- Wechsler, D. (2009a). *Advanced clinical solutions for the WAIS-IV and WMS-IV*. San Antonio: Pearson.
- Wechsler, D. (2009b). *WAIS-IV—Administration and scoring manual*. San Antonio: Pearson.
- Whiteside, D. M., Dunbar-Mayer, P., & Waters, D. P. (2009). Relationship between TOMM performance and PAI validity scales in a mixed clinical sample. *The Clinical Neuropsychologist*, 23, 523–533. doi:10.1080/13854040802389169
- Young, G. (2015). Malingering in forensic disability-related assessments: Prevalence 15±15%. *Psychological Injury and Law*, 8, 200–218. doi:10.1007/s12207-015-9232-4
- Youngjohn, J. R., Burrows, L., & Erdal, K. (1995). Brain damage or compensation neurosis? The controversial post-concussion syndrome. *The Clinical Neuropsychologist*, 9, 112–123. doi:10.1080/13854049508401593