Historical, Conceptual and Empirical Factors in Performance and Symptom Validity Assessment

Glenn J. Larrabee, Ph.D., ABPP-Cn

---

Financial Disclosure

I have financial relationships to disclose:
I receive royalties from Oxford University Press from sales of Assessment of Malingered Neuropsychological Deficits (Larrabee, 2007), and Forensic Neuropsychology: A Scientific Approach (Larrabee, 2012).
I receive royalties from Psychological Assessment Resources for sales of the Continuous Visual Memory Test
Psychological and Neuropsychological Testing

• Requires accurate self report on tests such as the MMPI-2
• Requires a valid performance on tests of ability, such as the Wechsler Intelligence and Memory scales
• Since both self report and performance can be controlled by the examinee, it is critical that symptom validity and performance validity is evaluated in each assessment

Performance Validity Tests or PVTs

• PVTs tell you whether the examinee is providing an accurate measure of their actual abilities
• These can be stand-alone like the Test of Memory Malingering (TOMM) or Word Memory Test (WMT)
• These can be embedded/derived measures based on neurologically atypical patterns or levels of performance on standard tests like the Auditory Verbal Learning Test, Wisconsin Card Sorting Test, Finger Tapping, or Reliable Digit Span
Symptom Validity Tests or SVTs

- SVTs tell you whether the examinee is providing an accurate report of their actual symptom experience
- SVTs are included on personality tests like the MMPI-2-RF (F-r, Fp-r, Fs, FBS-r, RBS)
- SVTs have also been developed for pain scales such as the Modified Somatic Perception Questionnaire or Pain Disability Index

Manfred Greiffenstein
1952-2016

- Skeptics approach to forensic neuropsychology
- Myths of Forensic Neuropsychology
- Developer of the Reliable Digit Span
- Studies of MTBI litigants
  - MTBI, PTSD, Chronic Pain and Toxic Exposure
  - MMPI profiles
    - Pre-existing profiles
    - Examiner effects
    - FBS and medical conditions
    - Academic records
  - Interests in history & evolutionary psychology
PVT Failure Distorts Expected Relationships with Validity Criteria

- GPA does not show the expected relationship with IQ until those failing PVTs are excluded (Greiffenstein & Baker, 2003)
- Olfactory identification does not correlate with TBI severity until those failing PVTs are excluded (Green et al., 2003)
- Memory complaints only correlate with memory performance in those failing PVTs (Gervais et al., 2008)
- Memory test scores correlate .49 with hippocampal volume in MCIs who pass PVTs, but correlate -.11 for those who fail PVTs (Rienstra et al., 2013)

PVT Failure Distorts Expected Relationships with Validity Criteria

- Memory performance does not differ between examinees with and without CT scan abnormalities until those failing PVTs are excluded (Green, 2007)
- A TBI/neurologically impaired group did not differ on neuropsychological tests from psychiatric, pain or mTBI subjects until those failing PVTs were excluded (Green et al., 2001)
- Neuropsychological performance was associated with presence/absence of brain injury only in those Ss passing PVTs (Fox, 2011)
Frequency of PVT failure in Settings with External Incentive

- 40% in litigation/compensation-seeking mTBI (Larrabee, 2003)
- 54.3% in criminal defendants (Ardolf et al. 2007)
- 48.5% in Social Security claimants (Chafetz, 2008)
- 40% in persons claiming environmental or toxic exposure (Greve et al. 2006)

Costs for SSD (SSI and SSDI) in Billions
Chafetz & Underhill, 2013

<table>
<thead>
<tr>
<th>Frequency of PVT Failure</th>
<th>Mental Disorders</th>
<th>Musculoskeletal Disorder (Chronic Pain)</th>
<th>Combined Mental Disorders/Musculoskeletal Disorder</th>
</tr>
</thead>
<tbody>
<tr>
<td>40%</td>
<td>$20.022 Billion</td>
<td>$14.176 Billion</td>
<td>$34.198 Billion</td>
</tr>
<tr>
<td>50%</td>
<td>$25.028 Billion</td>
<td>$17.72 Billion</td>
<td>$42.748 Billion</td>
</tr>
</tbody>
</table>
Research on SVTs and PVTs In Psychology and Neuropsychology Has Lengthy and Extensive History

- F scale included in MMPI (1943)
- Rey (1941) Dot Counting Test
- Rey (1964) 15-item Test
- Research on feigned BVRT performance (Benton & Spreen, 1961)
- Two alternative forced choice testing (Pankratz, Fausti & Peed, 1975)
- Pattern analysis on standard neuropsychological tests (Heaton, Smith, Lehman & Vogt, 1978)
- Slick et al. (1999) diagnostic criteria for malingered neurocognitive dysfunction
- Bianchini et al. (2005) criteria for malingered pain related disability

Explosion in PVT and SVT Research 1990 to Present (Sweet & Guidotti Breting, 2013)

- 14 Meta-analytic reviews of PVTs and SVTs, 1991-2011
Position Papers from Major Organizations

• NAN Position Paper, 2005: Symptom exaggeration or fabrication occurs in a sizeable minority of exams, with greater prevalence in forensic contexts; use of PVTs/SVTs maximizes confidence in results and is medically necessary
• AACN Consensus Statement, 2009: 30 experts; PVT/SVT assessment is important and necessary, particularly in secondary gain contexts, but also in routine clinical practice

How are PVTs created?

Simulation Design

• Simulation design usually compares two groups of Ss on the PVT being developed:
• 1) a group of non-injured persons asked to believably feign deficits in an imagined personal injury suit
• 2) Persons with moderate/severe TBI who do not have external incentive
How are PVTs Created?
Known Groups or Criterion Groups Design

• Usually compares two groups on the PVT being studied:
  • 1) a litigating/compensation-seeking group of uncomplicated mTBI (normal CT scan, no or brief LOC, limited PTA), who also fail 2 or more PVTs independent of the PVT being studied
  • 2) a group of Ss with moderate/severe TBI who do not have any external incentive

Key Difference: PVTs vs. Traditional Test Scores

• PVTs represent performance that is atypical in pattern or degree compared to performances produced by neurologically, psychiatrically, or developmentally impaired subjects
• PVTs are criterion-referenced
• Traditional Test scores are norm-referenced (compared to the performance of “normal” [normative] subjects)
PVT Survey, Martin et al., 2015, TCN, v. 29, 741-776)

- Surveyed 316 practicing neuropsychologists, recruited from major professional listserves (e.g. npych, AACN)
- Extensive and comprehensive survey
- This was followed up by an additional survey, of experts alone (Schroeder et al., TCN, 2016, 515-535), n = 24, 12 also in prior study, defined as members of AACN Consensus Group or first author on at least 4 articles/book chapters in past 5 years

### General Clinician Survey Number of PVTs Used in Forensic Cases

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Stand Alone</th>
<th>Embedded</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean (SD)</td>
<td>2.4 (.97)</td>
<td>3.9 (2.10)</td>
<td>6.3 (2.56)</td>
</tr>
<tr>
<td>Median</td>
<td>2</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>Minimum</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Maximum</td>
<td>6</td>
<td>12</td>
<td>16</td>
</tr>
</tbody>
</table>
### Top 10 Free Standing and Embedded/Derived for Clinicians

<table>
<thead>
<tr>
<th>Free Standing</th>
<th>Percent Using</th>
<th>Embedded/Derived</th>
<th>Percent Using</th>
</tr>
</thead>
<tbody>
<tr>
<td>TOMM</td>
<td>77.8</td>
<td>Digit Span</td>
<td>96.2</td>
</tr>
<tr>
<td>WMT</td>
<td>58.2</td>
<td>CVLT-2</td>
<td>74.7</td>
</tr>
<tr>
<td>MSVT</td>
<td>28.2</td>
<td>Finger Tapping</td>
<td>24.4</td>
</tr>
<tr>
<td>Dot Counting</td>
<td>24.7</td>
<td>WCST</td>
<td>24.7</td>
</tr>
<tr>
<td>Rey 15</td>
<td>24.1</td>
<td>WMS-III/IV</td>
<td>23.1</td>
</tr>
<tr>
<td>VSVT</td>
<td>19.3</td>
<td>Rey Figure</td>
<td>17.4</td>
</tr>
<tr>
<td>NV-MSVT</td>
<td>12.0</td>
<td>RBANS</td>
<td>13.2</td>
</tr>
<tr>
<td>ACS WCT</td>
<td>11.1</td>
<td>RAVLT</td>
<td>9.8</td>
</tr>
<tr>
<td>b Test</td>
<td>9.5</td>
<td>Trail Making</td>
<td>5.7</td>
</tr>
<tr>
<td>VIP</td>
<td>4.7</td>
<td>CAT Test</td>
<td>5.4</td>
</tr>
</tbody>
</table>

### Top Ten Free Standing and Embedded/Derived-Experts

<table>
<thead>
<tr>
<th>Free Standing</th>
<th>Percent Using</th>
<th>Embedded/Derived</th>
<th>Percent Using</th>
</tr>
</thead>
<tbody>
<tr>
<td>TOMM</td>
<td>79.2</td>
<td>Digit Span</td>
<td>91.7</td>
</tr>
<tr>
<td>WMT</td>
<td>62.5</td>
<td>CVLT-2</td>
<td>54.2</td>
</tr>
<tr>
<td>MSVT</td>
<td>41.7</td>
<td>Finger Tapping</td>
<td>37.5</td>
</tr>
<tr>
<td>Dot Counting</td>
<td>33.3</td>
<td>WCST</td>
<td>33.3</td>
</tr>
<tr>
<td>Rey 15</td>
<td>33.3</td>
<td>WMS-III/IV</td>
<td>33.3</td>
</tr>
<tr>
<td>VSVT</td>
<td>25.0</td>
<td>RBANS</td>
<td>20.8</td>
</tr>
<tr>
<td>b Test</td>
<td>20.8</td>
<td>Trail Making</td>
<td>12.5</td>
</tr>
<tr>
<td>NV-MSVT</td>
<td>16.7</td>
<td>CPT-II</td>
<td>12.5</td>
</tr>
<tr>
<td>ACS WCT</td>
<td>16.7</td>
<td>WAIS (exclude D.Sp)</td>
<td>12.5</td>
</tr>
<tr>
<td>Rey Word Recog</td>
<td>12.5</td>
<td>Recall v. Recog</td>
<td>12.5</td>
</tr>
</tbody>
</table>
Diagnostic Statistics/Terminology

- **Condition of Interest (COI):** what is being detected, e.g. invalid performance
- **Sensitivity:** the % of Ss detected who actually have the COI
- **Specificity:** the % of Ss who are correctly identified who do *not* have the COI
- **False negative rate:** % of Ss with the COI who are *not* detected
- **False positive rate:** the % of Ss without the COI who are incorrectly identified as having the COI

Probability of a Diagnosis

- **Referred to as the Positive Predictive Power (PPP) or Positive Predictive Value (PPV)**
- **Is broadly defined as Specificity/Specificity + (1-Specificity)**
- **Alternatively defined as True Positives/True positives + False Positives**
- **PPP is base rate dependent**
Example of PPP

• Assume a base rate of 40% (e.g. 40% invalid, 60% valid), and Sensitivity (Sn) of .50 and Specificity (Sp) of .90
• With 100 total cases, Sn of .50 identifies 20 cases (.5 x 40), 1-Sp identifies 6 cases (.1 x 60), and PPP is 20/20 + 6 = .79

PVTs, SVTs and Effect of Minimizing the False Positive Rate

• Use of non-litigating patients with moderate and severe TBI, with history of coma and/or CT scan abnormalities, plus a PVT/SVT cutoff with ≤10% false positive (FP) rate
• Specifying the characteristics of Ss who are false positive
• Because of the attempt to minimize false positive error, sensitivity is notably lower
• PVT meta-analysis of Vickery et al. (2001) found mean sensitivity of .56 with mean specificity of .95
• Sollman and Berry (2011) found mean sensitivity of .69 with mean specificity of .90
PVTs are Easy to Pass for Examinees who have Bona Fide Problems

• Mean TOMM correct is 98.7% for aphasics

• Mean TOMM correct is 98.3% for severe TBIs (>1 day up to 3 months of coma); 1 S with GSW, 38 days coma, & right frontal lobectomy scored perfectly (Tombaugh, 1996)

• 3 patients with bilateral hippocampal damage due to hypoxia passed the PVT trials of the Word Memory Test (Goodrich-Hunsaker & Hopkins, 2009)

• 94% of Ss with temporal lobe damage (57% left/43% right), intractable epilepsy, & AVLT delay = T39, pass the Effort Index of the AVLT (Silverberg & Barrash, 2005)

• Performance on 3 PVTs and 1 SVT was not related to presence or absence of CT/MRI lesions or to frontal vs. non-frontal location (McBride et al., 2013)
PVTs are Easy to Pass in Patients with Bona Fide Problems

- Cold pressor does not impact performance on Reliable Digit Span or on the TOMM (Etherton, Bianchini, Ciota et al., 2005; Etherton, Bianchini, Greve et al., 2005)
- “Diagnosis Threat” does not affect WMT performance (Suhr & Gunstad, 2005)

- Depression (Rees et al., 2001), depression and anxiety (Ashendorf et al., 2004) and depression with chronic pain (Iverson et al., 2007) do not affect TOMM performance
- 100% of fibromyalgia and rheumatoid arthritis Ss not seeking compensation passed the CARB (Gervais, Russell et al., 2001)
TOMM Scores

TBI Sample (N = 45; Tombaugh, 1996)

- Trial 1
- Trial 2
- Retention

Aphasia Sample (N = 21; Tombaugh, 1996)

- Trial 1
- Trial 2
- Retention
The Positive Likelihood Ratio
Grimes & Schulz 2005

- LR+ is defined as sensitivity/1 – specificity
- LR+ gives likelihood the score came from the group with the condition of interest (COI) as opposed to the group without the COI
- LR+ multiplied by the base rate odds gives the post-test odds, which can be converted back to a diagnostic probability by the formula odds/odds + 1
LR+ for Vickery et al. PVT Meta Analyses

• Vickery et al. PVT meta analysis, mean Sn = .56, Sp = .95
• For the Vickery et al. PVT meta analysis, LR+ is .56/.05 or 11.2
• PPP = base rate odds x LR+ yielding post-test odds, which are converted to a probability by odds/odds + 1
• Post test odds = .40/.60 (for 40% base rate) or .67 x 11.2 = 7.5; 7.5/8.5 = .88

LR+ for Sollman & Berry PVT Meta Analyses

• Sollman & Berry PVT meta analysis determined a mean Sn of .69 and Sp of .90
• For the Sollman & Berry meta analysis, LR+ is .69/.10 or 6.9
• Post-test odds = .67 x 6.9 = 4.623
• PPP = 4.623/5.623 = .82
LR+ for Various Clinical Disorders: Heaton et al., 2004; Center For Evidence Based Medicine, Toronto

- LR+ for Vickery et al., 11.2 and Sollman & Berry, 6.9
- Brain damage vs. no brain damage using AIR from Halstead Reitan (Heaton et al. 2004), LR+ = .771/.146 = 5.28
- Alzheimer’s positive APOE, LR+ = 2.0
- Myocardial Infarction, Cardiac Specific Troponin T, LR+ at 0-2 hours = 6.3
- Chronic Obstructive Airway Disease by spirometry, LR+ = 7.3

Slick et al. Criteria for Malingered Neurocognitive Dysfunction

- Slick, Sherman and Iverson (1999): diagnostic criteria for determination of malingered neurocognitive dysfunction (MND)
- MND: a) presence of a substantial external incentive, b) evidence from neuropsychological testing and/or self report, and c) evidence from b cannot be primarily accounted for by bona fide neurologic, psychiatric or developmental disorder
Slick Criteria Are Probabilistic

- Significantly worse-than-chance performance on two-alternative forced choice testing = definite MND, in the presence of substantial external incentive, and absence of rule-out conditions
- In the absence of < chance performance, multiple types of neuropsychological test evidence, or one type of test evidence combined with one pattern of atypical self report, in the presence of external incentive and absence of rule-out conditions, can define probable MND

Impact of Slick Criteria on Research

- Subsequent to the appearance of the Slick Criteria in 1999, allowing an operational definition of malingering, there has been a veritable explosion in research on PVTs, SVTs, and malingering
- One of the early influences was to encourage research on the relationship of multiple independent PVTs to detection of invalid performance/malingering
Larrabee (2003, TCN, 17, 410-425)

- Used up to 31 cases with moderate/severe TBI, and up to 26 cases, primarily mild TBI, without objective radiologic abnormalities who were litigating/compensation-seeking, and who performed significantly worse-than-chance on the PDRT
- Cutting scores were determined for 4 embedded/derived PVTs (Visual Form Discrimination, Finger Tapping, Reliable Digit Span, WCST FMS) and 1 SVT (MMPI-2 FBS) with the goal of keeping the FP rate for moderate/severe TBI <16%.

<table>
<thead>
<tr>
<th>Test and Cutting Score</th>
<th>Definite MND identified</th>
<th>TBI identified</th>
</tr>
</thead>
<tbody>
<tr>
<td>Visual Form Discrimination &lt;26</td>
<td>12/25 (48%)</td>
<td>27/29 (93.1%)</td>
</tr>
<tr>
<td>Finger Tapping Combined &lt;63</td>
<td>10/25 (40%)</td>
<td>29/31 (93.5%)</td>
</tr>
<tr>
<td>Reliable Digit Span &lt;8</td>
<td>13/26 (50%)</td>
<td>29/31 (93.5%)</td>
</tr>
<tr>
<td>WCST Failure to Maintain Set &gt;1</td>
<td>12/25 (48.0%)</td>
<td>27/31 (87.1%)</td>
</tr>
<tr>
<td>MMPI-2 FBS raw score &gt;21</td>
<td>21/26 (80.8%)</td>
<td>25/29 (86.2%)</td>
</tr>
<tr>
<td>Average Correctly Identified</td>
<td>53.5%</td>
<td>90.7%</td>
</tr>
</tbody>
</table>
Characteristics of False Positive Cases

Case 1: Male, in 20s, severe TBI, 1 month coma, 2 months PTA, bilateral CT abnormalities, post-traumatic dementia (FIQ = 75, ID twin = 102). In litigation, Rey 15 = 12, TOMM T2 and Retention = 50, 24 on PDRT easy, 24 on PDRT Hard. RDS = 7, FMS = 2 (both scores just beyond cutoff).

Case 2: Non-litigating female in 40s, severe TBI as child, CT shows prior craniotomy and bilateral frontal lobe encephalomalacia. FMS = 3, FBS = 24.

Case 3: Litigating female, in 60s, moderate TBI, GCS = 14, CT showed lesion in left internal capsule/lateral thalamus, PTA 3 days, 14 on Rey 15, PDRT Easy = 32 9/9 on Hard, Warrington Words = 39. FMS = 2, FBS = 24.
Positive Predictive Power (PPP)

• Represents the probability of detecting a condition of interest (COI)
• PPP is defined by relationship of True Positives (TP) and False Positives (FP)
  – PPP = TP / (TP + FP)
• For the 2 or more out of 5 failure rate
  – PPP = 21/(21 + 3) = 21/24 = .875
• For the 3 or more out of 5 failure rate
  – PPP = 13/(13 + 0) = 1.00

Importance of FP rate

• The preceding slide shows that keeping the FP rate low keeps the PPP high
• So, low FP rate (high Specificity) is good for ruling in a COI
• When the false positive rate is zero, only persons with the COI fall in this range
Cross Validation Results

• 15/17 (88.2%) of probable MND, and 100% of neurologic and psychiatric subjects correctly identified by failure of 2 or more PVTs/SVT using the criteria for the definite MND/TBI per test cutoffs

• 8/17 (47%) of probable MND and 100% of neurologic and psychiatric cases correctly identified by failure of 3 or more PVTs/SVT

Comparison of Definite MND and Probable MND

• Definite MND and Probable MND did not differ on the 5 embedded/derived measures of PVT and SVT

• Definite MND and Probable MND did not differ on 4 sensitive measures of neuropsychological ability: COWA, Trails B, Verbal Selective Reminding, and CVMT

• These performance similarities support the definition of probable MND as comprised of multiple PVT and SVT failures (also see Chapter 13, in Larrabee, 2007, *Assessment of Malingered Neuropsychological Deficits*)
Victor et al. (2009, TCN, 23, 297-313)

- 66 credible patients, who had no external incentive, failed ≤1 PVT, and did not have dementia or mental retardation
- 37 non-credible patients who had external incentive (litigation, disability claim) and failed at least 2 of 5 PVTs
- PVTs included Reliable Digit Span, RO Complex Figure Test Effort Equation, RAVLT Effort Equation, and Dominant Hand Finger Tapping

<table>
<thead>
<tr>
<th>PVT</th>
<th>Hit Rate</th>
<th>Sensitivity</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>RDS</td>
<td>70.9</td>
<td>43.2</td>
<td>86.4</td>
</tr>
<tr>
<td>R-O</td>
<td>87.5</td>
<td>81.3</td>
<td>91.7</td>
</tr>
<tr>
<td>RAVLT</td>
<td>84.2</td>
<td>86.1</td>
<td>83.1</td>
</tr>
<tr>
<td>FTT</td>
<td>69.4</td>
<td>47.1</td>
<td>81.3</td>
</tr>
<tr>
<td>Any one test</td>
<td>68.0</td>
<td>94.6</td>
<td>53.0</td>
</tr>
<tr>
<td>Any two tests</td>
<td>90.3</td>
<td>83.8</td>
<td>93.9</td>
</tr>
<tr>
<td>Any three tests</td>
<td>81.6</td>
<td>51.4</td>
<td>98.5</td>
</tr>
</tbody>
</table>
Victor et al. Characteristics of False Positive Cases

- 30 year old Hispanic female, 8 years education, borderline IQ (77), spoke 10% English while growing up, history of learning disability, bipolar disorder, and substance abuse
- 17 year old African American male with 10 years education, borderline IQ (72), history of severe TBI with post-traumatic seizures, alcohol dependence and learning disability

- 53 year old Hispanic male with one year of education in Mexico and borderline IQ (79), English as second language, history of schizophrenia, panic disorder and major depressive disorder
- 50 year old Caucasian female 14 years education, diagnosed with bipolar disorder and a rule out for somatoform disorder
Victor et al False Positive Cases

- The first 3 Ss described were already on disability, had no other identifiable external incentives to feign, and all failed the same two PVTs: RDS and Finger Tapping
- The fourth false positive was not on disability, failed 3 embedded PVTs but passed all freestanding PVTs

<table>
<thead>
<tr>
<th>Value</th>
<th>Victor Fail ≥2/4 PVTs MND = 37 Clinical = 66</th>
<th>Larrabee Fail ≥2/5 PVT/SVT MND = 41 Clinical = 54</th>
<th>Victor Fail 3 ≥4 PVT MND = 37 Clinical = 66</th>
<th>Larrabee Fail ≥3/5 PVT/SVT MND = 41 Clinical = 54</th>
<th>Victor Logistic Regress</th>
<th>Larrabee Logistic Regress</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sensitivity</td>
<td>83.8</td>
<td>87.8</td>
<td>51.4</td>
<td>51.2</td>
<td>85.7</td>
<td>79.2</td>
</tr>
<tr>
<td>Specificity</td>
<td>93.9</td>
<td>94.4</td>
<td>98.5</td>
<td>100.0</td>
<td>95.6</td>
<td>85.2</td>
</tr>
<tr>
<td>Hit Rate</td>
<td>90.3</td>
<td>91.6</td>
<td>81.6</td>
<td>78.9</td>
<td>91.8</td>
<td>82.4</td>
</tr>
</tbody>
</table>
Likelihood Ratios

• Positive Likelihood Ratio (LR+)
  – LR+ = Sensitivity / (1 – Specificity)
  – LR+ = Sensitivity / false positive rate

• For Preceding Data
  – Victor fail ≥2 PVTs: LR+ = 83.8 / 6.1 = 13.74
  – Larrabee fail ≥2 P/SVTs: LR+ = 87.8 / 5.6 = 15.68

Larrabee (2008, TCN, 22, 666-679)

• Discussed linking of LR+ as a model of multiple PVT failure, suggesting 2 or more or 3 or more failures provide evidence for probable malingering
• You can link LR+ if individual PVTs are uncorrelated in the manner of (base rate odds) x LR+ (for 1st failed PVT) x LR+ for second PVT; post-test odds are then converted to PPP by odds/odds + 1
• Based on Larrabee (2003) actual PPP for each pair wise combination ranged from .800 to 1.00, whereas PPP based on linked LR+ computed for these combinations ranged from .939 to .977
Likelihood Ratios and Diagnostic Probability

- LR+ multiplied by the base rate odds of the COI gives the post-test odds of a positive test result.
- Victor fail ≥ 2/4 PVTs
  - Base rate = \(\frac{37}{37 + 66} = \frac{37}{103} = .359\)
  - Base rate odds = \(\frac{.359}{1 - .359} = .359/.641 = .56\)
  - Given LR+ = 13.74, post test odds = \(.56(13.74) = 7.69\)
  - Converting odds to probability (i.e., odds/odds+1), post test probability = \(7.69/8.69 = .88\)

Improving Diagnostic Accuracy by Using Multiple Indicators (Larrabee, 2008)

- LR+ can be chained, if the individual PVTs and SVTs are independent and either weakly correlated or uncorrelated.
- Consider 2 independent, uncorrelated PVTs, each with a sensitivity of .50 and specificity of .90, and a base rate probability of PVT failure of .40 (yielding base rate odds of .40/1-.40 = .67)
- Post test odds after failing the 1st PVT are \(.67 \times 5.0 = 3.35\), for a post-test probability of \(3.35/4.35 = .77\)
- The post-test odds of 3.35 can now be used to premultiply the LR+ for a 2nd failed PVT, which also has an LR+ of 5.0, yielding new post-test odds of 16.75, for a post-test probability of \(16.75/17.75 = .94\)
Further Support for Use of Multiple PVTs

- Boone (2009, TCN, 23, 729-741): you need continuous sampling of PVTs
- Proto et al. (2014, ACN, 29, 614-624) found reduced neuropsychological performance for mTBI in association with failure of 1 of 6 PVT v. 0 fails, effect sizes, $d = -.52$ ranging to -.74 on CVLT, PSI and TMB
- Proto et al. found linear increase in effect size as a function of number of PVT failures on CVLT 1-5:
  - $d = -0.72$ for fail 1 v. 0
  - $d = -1.11$ for fail 2 v. 0
  - $d = -1.42$ for fail $\geq 3$ v. 0

Statistically Based Criticisms of Use of Multiple PVTs

- Medici (2013, J of Forensic Psychology Practice, 13, 68-78): the false positive error rate increases with multiple PVTs as per the normal approximation to the binomial
- Berthelson et al. (2013, Brain Injury, 27, 909-916): Monte Carlo simulation is preferred to the binomial since Monte Carlo allows for PVT intercorrelation while the binomial does not
Binomial Equation

\[
\binom{n}{n_1} p^{n_1} q^{n-n_1}
\]

Example for Failure of \(\geq 2\) PVTs each with a FP rate of .10
Using the Binomial Equation

- \(n =\) total number of items, \(n_1 =\) subset selected, \(p = .10, q = .90;\) to compute probability of failing \(\geq 2\) of 5, this is \(p\) for 2/5, + \(p\) for 3/5...+ \(p\) for 5/5
- For 2 of 5 indicators: \([5!/(2!)(3!)] \times (.1^2)(.9^3)\) or .073
- For 3 of 5 indicators: \([5!/(3!)(2!)] \times (.1^3)(.9^2)\) or .008
- For 4 of 5 indicators: \([5!/(4!)(1!)] \times (.1^4)(.9^1)\) or .00045
- For 5 of 5: .1 to the 5th, or essentially zero
- Thus, for failure \(\geq 2\) of 5 PVTs, each with a .1 FP rate, the combined probability is .073 + .008 + .00045 or .08145
Monte Carlo Estimation
Crawford et al. (2007, Neuropsychology, 21, 419-430)

- Requires correlation matrix, R of variables
- Compute square root of the matrix (the Choleski decomposition)
- Generate random vector of k independent standard normal variates
- Post-multiply by Choleski decomposition matrix to yield an observation from the desired multivariate normal distribution with a mean vector of 0 and covariance matrix R; repeat preceding steps multiple times
- This process generates observations with a mean of 0 and SD of 1.0 (the standard normal distribution)

Monte Carlo Estimation Continued

- Berthelson et al. (2013, Brain Injury, 27, 909-916) cite 3 separate investigations showing accurate estimation of actual frequency of failure rates on standard neuropsychological tests
- Crawford et al. offered caveats to the use of their procedure
- An important caveat is the assumption of multivariate normality
Use of Multiple PVTs Does Not Negatively Impact the Per Test False Positive Rate

- Berthelson et al. (2013) used Monte Carlo simulation to estimate failure rates for subsets of PVTs failed at either a 10% or 15% per-test false positive rate
- Berthelson et al. found what they interpreted as excessive false positive rates associated with use of multiple PVTs

<table>
<thead>
<tr>
<th>Number of PVTs Used</th>
<th>≥1 PVTs Failed</th>
<th>≥2 PVTs Failed</th>
<th>≥3 PVTs Failed</th>
</tr>
</thead>
<tbody>
<tr>
<td>5 (ACS; Pearson, 2009)</td>
<td>22.0</td>
<td>7.0</td>
<td>2.0</td>
</tr>
<tr>
<td>5 (Binomial, Medici, 2013)</td>
<td>55.6</td>
<td>16.4</td>
<td>2.6</td>
</tr>
<tr>
<td>5 (Monte Carlo, Berthelson, 2013, r = .31)</td>
<td>45.7</td>
<td>19.4</td>
<td>7.3</td>
</tr>
<tr>
<td>4 (Victor et al., 2009)</td>
<td>47.0</td>
<td>6.0</td>
<td>1.0</td>
</tr>
<tr>
<td>4 (Binomial)</td>
<td>47.8</td>
<td>11.0</td>
<td>1.2</td>
</tr>
<tr>
<td>4 (Monte Carlo)</td>
<td>40.4</td>
<td>14.6</td>
<td>4.3</td>
</tr>
</tbody>
</table>
Actual PVT Failure Rates in Clinical Patients Contradicts Berthelson et al.

- Davis and Millis (TCN, 2014, 199-214) found that increasing the number of PVTs administered did not show a significant association with PVT failure.
- Davis and Millis attributed the difference in their findings compared to those of Berthelson et al. to violation of the Monte Carlo assumption of multivariate normality.
#PVTs administered and # failed is consistent in 3 Samples
Davis (from Larrabee & Davis AACN, 2015)

<table>
<thead>
<tr>
<th>Sample</th>
<th>N</th>
<th>$r_s$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Litigants/Disability Claimants</td>
<td>175</td>
<td>-.06</td>
</tr>
<tr>
<td>Clinical-VA</td>
<td>863</td>
<td>-.10*</td>
</tr>
<tr>
<td>Clinical-PM&amp;R</td>
<td>158</td>
<td>.13</td>
</tr>
<tr>
<td>Clinical-PM&amp;R NNI</td>
<td>87</td>
<td>.09</td>
</tr>
</tbody>
</table>

* $p = .003$

---

Logistic Regression Models, 1-9 PVTs Predicting MSVT pass/fail (Davis from Larrabee & Davis, 2015, AACN)

<table>
<thead>
<tr>
<th>#</th>
<th>PVT</th>
<th>Model</th>
<th>N</th>
<th>AUC</th>
<th>SEN</th>
<th>SPE</th>
<th>%Cor</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>TMA-A</td>
<td>P1</td>
<td>676</td>
<td>.72</td>
<td>.23</td>
<td>.96</td>
<td>75.4</td>
</tr>
<tr>
<td>P2</td>
<td>CVLT-FC</td>
<td>P1-2</td>
<td>659</td>
<td>.78</td>
<td>.40</td>
<td>.95</td>
<td>79.5</td>
</tr>
<tr>
<td>P3</td>
<td>RDS</td>
<td>P1-3</td>
<td>612</td>
<td>.79</td>
<td>.40</td>
<td>.94</td>
<td>79.6</td>
</tr>
<tr>
<td>P4</td>
<td>FTTD</td>
<td>P1-4</td>
<td>496</td>
<td>.81</td>
<td>.34</td>
<td>.95</td>
<td>79.8</td>
</tr>
<tr>
<td>P5</td>
<td>WC FMS</td>
<td>P1-5</td>
<td>410</td>
<td>.82</td>
<td>.38</td>
<td>.95</td>
<td>81.5</td>
</tr>
<tr>
<td>P6</td>
<td>RCF TPR</td>
<td>P1-6</td>
<td>194</td>
<td>.79</td>
<td>.30</td>
<td>.97</td>
<td>82.9</td>
</tr>
<tr>
<td>P7</td>
<td>LM REC</td>
<td>P1-7</td>
<td>110</td>
<td>.91</td>
<td>.62</td>
<td>.96</td>
<td>89.1</td>
</tr>
<tr>
<td>P8</td>
<td>VR REC</td>
<td>P1-8</td>
<td>109</td>
<td>.90</td>
<td>.60</td>
<td>.97</td>
<td>89.9</td>
</tr>
<tr>
<td>P9</td>
<td>VPA REC</td>
<td>P1-9</td>
<td>104</td>
<td>.93</td>
<td>.63</td>
<td>.98</td>
<td>91.4</td>
</tr>
</tbody>
</table>
Why Monte Carlo Simulated Data do not Match Actual PVT False Positive Rates

- Berthelson et al. Monte Carlo simulated data creates variables with a mean of 0 and SD of 1
- These create simulated data that follow the standard normal curve
- PVT and SVT data do not follow the normal curve, as they are skewed, with performance commonly at ceiling

Examples of Skewed Data From the TOMM Manual

- 21 aphasic subjects averaged 98.7% correct on Trial 2, with 16 achieving perfect scores
- 22 severe TBI patients (> 1 day to 3 months of coma) averaged 98.2% correct on Trial 2 with 14 achieving perfect scores
- These data demonstrate skewed distributions with performances occurring at ceiling, a common finding with PVTs
TOMM Scores

TBI Sample (N = 45; Tombaugh, 1996)
- Trial 1
- Trial 2
- Retention

Aphasia Sample (N = 21; Tombaugh, 1996)
- Trial 1
- Trial 2
- Retention
Larrabee (2008, TCN, 22, 666-679)

- Presents data on chaining of LR+
- If the PVTs are relatively independent, LR+ can be chained, that is, multiplied sequentially
- Thus, one can chain base rate odds multiplied by PVT 1, PVT 2, etc.
- In Larrabee (2008) I showed how this chained approach approximated the failure rates for failure of ≥2 of 5 PVTs/SVT, and ≥3 of 5

Criticisms of Larrabee (2008)

- In Larrabee (2008) I pointed out that once one fails 2 or more out of 5 PVTs, linking of new LR+, even for tests falling far below cutoff, would not reduce the post-test odds since the value of true positives/false positives would never fall below 1.0, since true positives would always be expected to be greater than false positives
- Frederick (npsych listserv, 04/15/15) has pointed out that the appropriate use of chaining requires linking of both LR+ and LR-
Linking of LR+ and LR-

- The procedure still requires pre-multiplying by the base rate odds
- The number of PVTs/SVTs exceeding cutoff identifies the PVT/SVT as contributing to LR+
- The number of PVTs/SVTs not exceeding cutoff contributes to LR-
- LR- = False Negatives/True Negatives (1 – sensitivity/specificity)

Frederick’s example (npsych, 04/15/15 at 3:35 PM)

- Example uses a base rate of .4, yielding base rate odds of .4/(1-.4) = .667
- TOMM TPR = .56, FPR = .02, WMT TPR = .85, FPR = .30, Rey 15-item TPR = .469, FPR = .028, VIP FPR Non-Verbal TPR = .662, FPR = .10
- For the example of failing the WMT but passing the TOMM, 15-item test, and VIP, the computation is (.667)(.85/.30 WMT)(.44/.98TOMM)(.531/.972 15-Item)(.338/.9VIP) = .174 for post-test odds
- PPP is .174/1.174 = .148
Actual Probability of Malingering vs. Probability Determined by Chaining of LR+ and LR-

- Victor et al. (2009), from Table 7, 6% of 66 cases, 4, were FP failing ≥2, and 84% of 37 cases, 31, were TP failing ≥2, so \( PPP = \frac{TP}{TP + FP} \) or \( \frac{31}{35} = .886 \); base rate odds = .56
- For RDS, LR+ = 3.18, LR- = .66, for the Rey-Osterrieth, LR+ = 9.80, LR- = .20, for the RAVLT, LR+ = 5.09, LR- = .17, for FTT, LR+ = 2.52, LR- = .65

Chaining of LR+ and LR- continued

- For failure of RDS and Rey-O, post test odds are (.56) \( (3.18)(9.80)(.17)(.65) = 1.93 \)
- This value must be added to 5 other computations of pairwise failure to get the post-test odds of failing 2 PVTs out of 4, which are 17.74
- PPP of failing 2 of 4 is thus 17.74/18.74 or .95, a value higher than the actual PPP of .886 for failure of ≥2 PVTs
Frederick, 2015, Too Much Information: Problems When Using Multiple Malingering Tests

• This was Frederick’s Address for the Distinguished Contributions To Forensic Psychology Award
• This address provides further examples of linking of likelihood ratios
• The following example from this address shows problems with the way Frederick chained the likelihood ratios, by not considering all possible combinations of passing vs failing PVTs

Frederick 2015 Address Continued

• Slides 31-33 from this talk provide an example of a case where 7 failed PVTs were chained, using LR+ alone: These included 1) Rey 15 with Recognition, 2) Rey Word, 3) ACS Word Choice Test, 4) Logical Memory II Recognition, 5) Verbal Paired Associates II Recognition, 6) AVLT Logistic Regression, and 7) WCST FMS
• Frederick demonstrated that by chaining the 7 LR+ associated with the failed PVTs, and the 17 LR- associated with the passed PVTs, the posterior probability (at a base rate of .40, base-rate odds of .667) was actually quite small, at .0003
Frederick’s Computations

- Chaining of the 7 LR+ and 17 LR- values yielded a chained LR of .000384456
- Multiplying this by .667 yielded a post-test odds of .000256432152
- This value, divided by 1 + this value, yields a post-test probability (rounded off) of .0003

Is the Post-Test Probability of .0003 Accurate?

- If it seems that a p of .0003 for failing 7 PVTs is too low, it is
- What Frederick computed was the chained LR for failing one particular set of 7 PVTs, and passing one particular set of 17 PVTs
- Although the computed probability for this one particular set of passed and failed PVTs is accurate, there are 346,104 ways of failing 7 out of 24 PVTs
- This is easily determined by n!/r!(n-r)! Or 24!/7!(24-7)!
Implications of Frederick’s Computation of the likelihood of Failing 7 and passing 17 PVTs

- Frederick only analyzed 1 out of 346,104 combinations of 7 out of 24 PVT failures
- In so doing, he grossly underestimates the probability of failing 7 out of 24 PVTs
- His example is equivalent to saying: “I am going to administer 24 PVTs, but I am only going to conclude invalid performance if the S fails 1) Rey 15 with Recognition, 2) Rey Word, 3) ACS Word Choice Test, 4) Logical Memory II Recognition, 5) Verbal Paired Associates II Recognition, 6) AVLT Logistic Regression, and 7) WCST FMS, but passes the 17 other PVTs
- The post-test odds for 346,103 other combinations of failure of 7 out of 24 PVTs are ignored

Putting It All Together

- Failure of multiple PVTs and SVTs indicates that there is a high probability of invalid data
- By itself, this indicates that low scores in an evaluation are more likely due to invalid performance, particularly with absence of risk factors for severe impairment such as history of coma
- Similarly, normal range scores themselves may be low estimates of actual level of ability
Putting it All Together

- PVT and SVT failure alone does not equate to malingering
- Multiple PVT and SVT failure, in the context of external incentive, with no clear evidence of neurologic, psychiatric or developmental contributions to test performance, meets general criteria for probable malingering (Slick et al., 1999; Bianchini et al. 2005)
- Below chance on 2-alternative forced choice testing, in the context of external incentive, meets criteria for definite malingering, Pankratz’ (1990) “smoking gun of intent”

---

Putting it All Together
Larrabee et al 2007 Chapter 13

- Definite malingerers (< chance) perform similarly to non-injured simulators, establishing < chance as intentional
- Definite malingerers (< chance) perform similarly to criterion group probable malingerers (multiple PVT/SVT failures), establishing validity of probable malingering criteria as supporting intentionally poor performance
- Dose effect (Bianchini et al., 2006) showing positive correlation of malingering with increasing external incentive shows that PVT/SVT failure is reinforced by pursuit of external incentives
Percentage of Clinical and Malingering Cases Failing PVTs and SVTs Larrabee 2014

<table>
<thead>
<tr>
<th># PVT &amp; SVT Failures</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clinical Cases n=54</td>
<td>51.9%</td>
<td>37.0%</td>
<td>7.4%</td>
<td>3.7%</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Malingering Cases n=41</td>
<td>0</td>
<td>2.4%</td>
<td>9.8%</td>
<td>24.4%</td>
<td>26.8%</td>
<td>19.5%</td>
<td>17.1%</td>
<td>0</td>
</tr>
</tbody>
</table>

Sensitivity and Specificity for Larrabee 2014 Multiple Indicator Paper

<table>
<thead>
<tr>
<th>PVT/SVT Failure</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Likelihood Ratio +</th>
</tr>
</thead>
<tbody>
<tr>
<td>≥2 of 7</td>
<td>97.6</td>
<td>88.9</td>
<td>8.79</td>
</tr>
<tr>
<td>≥3 of 7</td>
<td>87.8</td>
<td>96.3</td>
<td>23.73</td>
</tr>
<tr>
<td>≥4 of 7</td>
<td>63.4</td>
<td>100.0</td>
<td>Cannot compute LR+, however PPP = TP/TP + FP, = 100%</td>
</tr>
</tbody>
</table>
Improving Diagnostic Accuracy Results from Controlling per-test False Positive rate and use of Multiple PVTs

- PVT and SVT research focuses on keeping the False+ (FP) rate at 10% or less per test in subjects with significant neurologic, psychiatric and developmental problems
- Current PVT and SVT research specifies the characteristics of FP cases in derivation studies
- FP typically occur in Ss who have undeniably severe deficits, often requiring 24 hour supervised care
- Requirement of multiple PVTs and SVTs improves diagnostic accuracy by reducing False Positive and False Negative rates

Summary on PVTs and SVTs

- The science behind PVTs and SVTs dates back over 70 years, with extensive research in the past 20 years on both stand alone and embedded/derived measures
- Diagnostic accuracy has focused on keeping the per-test false positive rate (FP) at 10% or less which is further enhanced by requiring use of multiple independent PVTs/SVTs
- PVT/SVT failure is rare in non-demented patients; rare in well-motivated patients, but frequent (40-50% of the time) in patients having external incentives
- Costs for mental disorder and pain disability based on invalid data are in the billions of dollars per year
Strategies for Analysis of False Positive Cases

- Determination as to whether an examinee belongs to a patient group at risk for false positive identification (severe TBI, dementia, mental retardation/intellectual disability, schizophrenia with negative symptoms)
- If an examinee falls in one of these groups, the PVT cutoff can be adjusted (e.g. b Test), or # of failure criterion can be increased (Smith et al., 2014, TCN, 28, 1048-1070)
- If not in one of these at-risk groups, risk of false positive error is reduced

Reducing False Positive Error by Combining PVTs or Specifying Minimum Levels of Performance

- False positive error for single PVTs can be reduced by combining two or more PVTs
- False positive error for single PVTs can be reduced by specifying minimum levels of performance on actual measures of neuropsychological function
Loring et al. RDS and AVLT Recognition in AD, aMCI and Normal Elderly, ACN, 2016, 313-331

<table>
<thead>
<tr>
<th>PVT</th>
<th>AD, n = 178, age 75.7 False Positive Rate</th>
<th>aMCI, n = 365, age 74.9 False Positive rate</th>
<th>Controls, n = 206, age 76.0 False Positive Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>RDS ≤6</td>
<td>13%</td>
<td>4%</td>
<td>3%</td>
</tr>
<tr>
<td>AVLT Recognition ≤9</td>
<td>70%</td>
<td>42%</td>
<td>9%</td>
</tr>
<tr>
<td>AVLT Recognition ≤5</td>
<td>37%</td>
<td>13%</td>
<td>5%</td>
</tr>
<tr>
<td>RDS ≤6 &amp; AVLT Recognition ≤9</td>
<td>6%</td>
<td>1%</td>
<td>0%</td>
</tr>
<tr>
<td>RDS ≤6 &amp; AVLT ≤5</td>
<td>4%</td>
<td>&lt;1%</td>
<td>0%</td>
</tr>
</tbody>
</table>

False Positive Rates As a Function of Minimum Level of Performance on Trails B and AVLT Long Delay Free Recall

<table>
<thead>
<tr>
<th>PVT</th>
<th>Total Sample False Positive Rate</th>
<th>False Positive Rate for Trails B ≥ SS 10</th>
<th>False Positive Rate for AVLT Long Delay Free Recall ≥ SS 10</th>
</tr>
</thead>
<tbody>
<tr>
<td>RDS ≤6</td>
<td>6.1%</td>
<td>2.2%</td>
<td>3.4%</td>
</tr>
<tr>
<td>AVLT Recognition ≤9</td>
<td>39.8%</td>
<td>27%</td>
<td>5.8%</td>
</tr>
</tbody>
</table>
Normal Performance on Actual Tests Mimicked by PVTs

- PVTs mimic tests of actual ability
- RDS mimics working memory, hence performing normally on Arithmetic, Letter Number Sequencing, or AVLT List B shows sufficient native ability to pass RDS
- The same is true for failure of the TOMM and WMT in the context of normal performance on WMS-IV Visual Reproduction, or normal AVLT
- Performing normally on the Grooved Pegboard test shows adequate motor ability for passing Finger Tapping

Summary of Minimizing False Positive Rate

- Evaluate for clinical condition known to be associated with elevated fp rate (severe neurologic, psychiatric or developmental disorder requiring supervised care)
- If clinical condition shows fp risk, adjust individual PVT cutoff or # failures required
- Use multiple PVTs and SVTs that are independent
- Evaluate domains of actual ability that the PVT mimics; e.g. evidence of normal attention (Arithmetic, Stroop), normal fine motor skills (GPB), normal memory (AVLT) and normal problem solving skills (WAIS BD) shows sufficient native ability to pass RDS, FT, ACS Word Choice Test, and WCST FMS
References


References


References


References


References


