Assessment of Noncredible Test Performance

March 23, 2010

Charles J. Vella, Ph.D.

Thanks to:

- Kyle Boone, Ph.D.
- William Lynch, Ph.D.
- Paul Green, Ph.D.

History 1

- Truth verification has been around in some form for thousands of years. The first recorded criminal investigation was the homicide of Abel by his brother Cain. God asked Cain if he had killed Abel and Cain responded, "Am I my brothers keeper?". Answering a question with a question is one of many tell tale signs of deception.
- During the Dark Ages in Europe, many devices of torture were used to facilitate "confessions" from crime suspects or those who were suspected of witchcraft. If you have heard of "The Rack" or "The Iron Maiden", you may not have known that those were used to "extract" confessions from suspects.
- One could argue that the tests during the middle ages to prove if people were witches was a form of malingering detection. . .e.g., if, when thrown into a river, if you drowned and sunk to the bottom you were said to have been innocent, but if you could swim or floated on top of the water you were said to be a witch and then executed.

Hartman, 2008

History 2

- Dot Counting (A. Rey, 1941)
- The Rey 15-Item Memory Test (1964)
- 48-Pictures Test (Signoret, 1979).
- Pankratz (1979) introduced the term "symptom validity testing."
- Pankratz used this term (SVT) "for basing validity judgments on statistical probability, always testing the symptoms of which the patient complained."

Definition of Effort

- Definition: whether or not a person is performing to the best of their ability
- Effort is the investment in performing at capacity levels, the investment to do well.
- Lack of Effort is defined as the low motivation of depressed subjects and subjects who are indifferent to testing to apply their cognitive abilities to neuropsychological tests. This results in worse performance *and lower scores than they are capable*
- David Hartman: 'E = O - A' Where Effort equals optimal performance minus actual performance. Optimal being defined as neuropsychological performance (or variance accounted for by it) the level of actual neurological capacity or condition; Actual being defined as the (variance accounted for by) score on the test.

Determining Effort Level

- Malingering is rare
- Wide range of less than optimal motivation on testing
- Current consensus: Insufficient Effort Level, noncredible test performance

Malingering

- Malingerer puts forth good effort: to fail rather than to succeed
- Malingering tests: error totals that exceed chance

Malingering 2

- Malingering is as old as war, work, school, and bad dates, splitting headache, case of flu.
- We've all done it

Multiple Causes of Less than Full Effort Performance

- Malingering Syndrome
- Factitious Disorder
- Learned Illness Behavior
- Conversion, Pain Syndrome, or other somatoform disorders
- Depressive Disorders
- Test Anxiety
- Fatigue
- Medication of psychoactive drug effects
- Lowered self efficacy expectations
- Need to gain recognition of symptoms

Multiple Causes of Less than Full Effort Performance 2

- Distraction due to anxiety/stress
- Feeling overwhelmed
- Depression
- Loss of self-confidence
- Pain
- Personality issues
- Secondary gain issues (fear of failure, dependency, etc)
- Intentional malingering

Malingering Syndrome

- Intentional production of false or exaggerated physical or psychological symptoms, motivated by ***external incentives*** (financial, evade criminal prosecution, obtain drugs)
- Often is used to identify all types of less than full effort performances
- APA (DSM-IV) requires two of the following:
  - Medicolegal context
  - Marked discrepancy between claimed impairment and objective findings
  - Lack of cooperation during testing
  - Antisocial Personality Disorder
- Estimated incidence: 33-60% of patients seen in NP clinics

Volitional exaggeration

- Conversion disorder, somatoform disorder and other psychiatric conditions cannot be diagnosed if there is VOLITIONAL production or exaggeration of symptoms.
- See DSM-IV manual.
- Patients with high to very high levels of neuroticism (on the NEO-FFI) can pass the MSVT with no problem.

Factitious Disorder

- Intentional production of physical or psychological signs and symptoms to assume or maintain the ***sick role***
- Distinguished from malingering in that FD focuses on the sick role
  - If the incentive were absent, FD would persist, Malingering would not
- Distinguished from Learned Illness Behavior in that FD is conscious and intentional

Who produces Noncredible Test Performance

- Chronic pain cases claim more memory problems than people with brain tumors, strokes, ruptured aneurysm, or very severe TBI
- Disability seekers
- Somatoform patients
- mTBI patients
- Depressed patients

False Positives

- All SVTs based on easy tasks have the limitation that they will produce false positives in people with very severe impairment, notably those with MR or dementia.
- On any single easy test which is failed (e.g. TOMM) in which the score is not worse than chance, someone with dementia or MR will look like someone with poor effort
- While people with dementia often fail the easy WMT recognition memory subtests, they score very much worse on the harder subtests, creating a profile that makes sense for someone with truly major impairment. In such cases, the easy-hard difference will be above a certain value (30 for WMT, 20 for MSVT and NV-MSVT).

Noncredible Test Performance is critical

- The viability of psychological and neuropsychological assessment hinges on ability to verify that scores are true and accurate
  - Psychological assessments are worthless, if feigned performance cannot be detected

Most prominent issue in clinical neuropsychology today

- >300 articles in last 15 years

Strategies For Detection Of Feigned Cognitive Symptoms

- A. Noncredible pattern on "effort" tests (SVTs: symptom validation tests)
- B. Noncredible pattern on standard cognitive tests
- C. Inconsistency in scores within/across cognitive evaluation
- D. Inconsistency between test scores and ADLs
- E. Inconsistency between injury specifics and test scores (improbable outcome)
- F. MMPI-2 Profile (FBS (fake bad scale), Response Bias Scale)

Cognitive “Effort” Tests

- Rationale: based on faulty information held by the general public regarding the effects of brain injury, specifically, that the following skills are impaired:
  - overlearned information (alphabet, simple calculations, sight reading)
  - recognition memory versus free recall

Reliance on a single SVT (incorrectly) assumes that

- Response bias is constant across an exam
- Response bias presents in the same manner in all individuals (i.e., that all patients use the same strategies when feigning)

Instead:

- Response bias typically fluctuates across an exam
- Even if effort is constant, individuals differ in the strategies they use when feigning cognitive symptoms
- Therefore, need *continuous sampling* of effort using differing effort indicators

Noncredible patients are heterogeneous

- There is no one "noncredible" profile
  - Some "malingerers" will do well on some tests and this does not negate the fact that they are not credible

Current recommended practice

- "Disperse SVTs or measures with symptom validity indicators *throughout* the evaluation, with administration of at least one SVT early in the evaluation process"
  - NAN position paper:
    - Symptom validity assessment: Practice issues and medical necessity, NAN Policy and Planning Committee, 2005 (ACN, 20, 419-426)

Surveys of Clinical Practice

- Approximately 79% of experts reported giving at least 1 SVT in every litigant assessment
  - Rey 15-item and TOMM most frequently given SVTs
    - Slick et al. (2004)
- 56% of NAN member respondents reported often or always including a measure of effort
  - TOMM, MMPI-2 F-K, FBS, Rey 15-item, CVLT most frequently administered
    - Sharland et al. (2007)

Forced Choice Tests

- Task is to identify/recognize in 2-choice trials stimuli to which patient has been previously exposed (i.e. Warrington)
- Individuals feigning brain impairment do not realize that recognition is substantially easier than free recall, and typically perform more poorly on FC tests than patients with real brain injury
- Sensitive to feigned deficits in memory

Cognitive Domains In Which Symptoms Can Be Faked:

- Memory
- Mental Speed
- Language (including reading)
- Math
- Visual Perceptual/Spatial
- Intelligence
- Motor dexterity/strength
- All of the above

How to select effort indicators….

- They need to be
  - brief
  - nonredundant (tap differing skill domains)

Available “Effort” Tests

- Forced Choice (FC)
  - Word Memory Test (WMT)
  - Test of Memory Malingering (TOMM)
  - Warrington Recognition Memory Test (Words)
  - Computerized Assessment of Response Bias (CARB)
  - Victoria Symptom Validity Test (VSVT)
  - Portland Digit Memory Test
  - Hiscock Memory Test
  - Validity Indicator Profile (VIP)

Test of Memory Malingering: TOMM

DETERMINING EFFORT LEVEL

TEST OF MEMORY MALINGERING [TOMM] - example

LEARNING

FORCED-CHOICE

A

B

Validity Indicator Profile: VIP

DETERMINING EFFORT LEVEL

Validity Indicator Profile [VIP]

Test Author:	Richard Frederick
Publication Date:	1997
Test Format:	Form & Booklet
Scoring:	Computer
General Description:	78 Vocabulary; 100 'Matrix'
Approximate Time:	10-15 min. for Verbal; 20-35 min. for Non-Verbal

Victoria Symptom Validity Test

DETERMINING EFFORT LEVEL

- **Victoria Symptom Validity Test [VSVT]**

  Example:

  764921

  [Brief Delay]

  Then, Which one?

  253172 **764921**

  [Easy]

CARB, Portland Digit Recognition

DETERMINING EFFORT LEVEL

Similar Measures:

- ***Computerized Assessment of Response Bias*** **[CARB]**
- ***Portland Digit Recognition Test*** **[PDRT]**

Road Sign Perception Test (Green, 2008)




This is an experimental test and any user of the WMT, MSVT or NV-MSVT may use it free of charge.

The RSPT tests the ability to perceive up to three road signs which are flashed on the screen very quickly (e.g. 75 msecs.) or more slowly (e.g. 900 msecs.).

Importance of Testing for Effort

- As effort decreases, scores on most neuropsychological tests decrease significantly and systematically
- Memory is particularly sensitive to effort
- Failure to remove poor effort cases is a fatal flaw in studies claiming that condition X is linked with cognitive deficits.
- Condition X may be chronic pain, panic attacks, mild TBI, depression, memory disorder.

Green, 2006

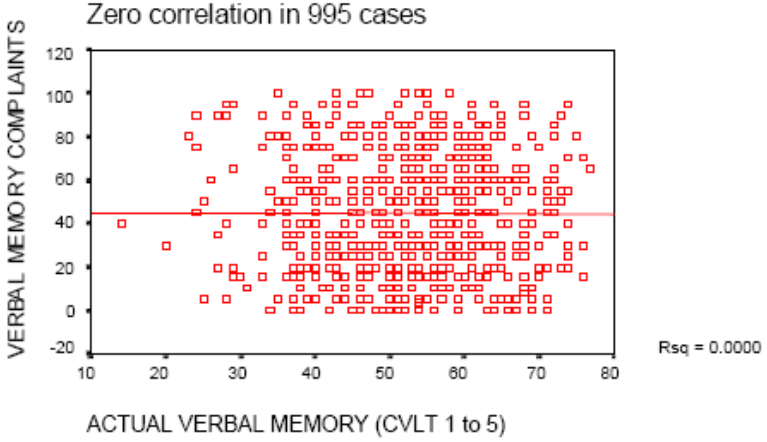
Memory Complaints and Scores

DETERMINING EFFORT LEVEL

Verbal memory complaints versus

verbal memory test scores

Zero correlation in 995 cases



Dr. Bill Lynch

Data collected daily over 8 years by Dr. Paul Green & Dr. Roger Gervais

- Over 2,000 clinical cases & over 100 healthy adults
- Effort test data available in all cases

- We can study **self-rated memory complaints**
- By diagnosis and by level of effort

MCI scales

- GMP general memory problems
- NIP numerical information problems
- VMP verbal memory problems
- VSMP visual-spatial memory problems

- PIM pain interferes with memory
- MIW memory interferes with work

- IRM impairment of remote memory
- ACB amnesia for complex behaviour
- AAB amnesia for antisocial behavior

Memory complaints:-

- A) tell us nothing about brain disease
- B) do not correlate with memory test scores

Subjective memory complaints are widespread among people claiming disability

- On MCI, the greatest memory complaints are not found in cases of brain disease
- Chronic pain cases have more memory complaints than any other group
- 93% say they cannot work because of impaired memory

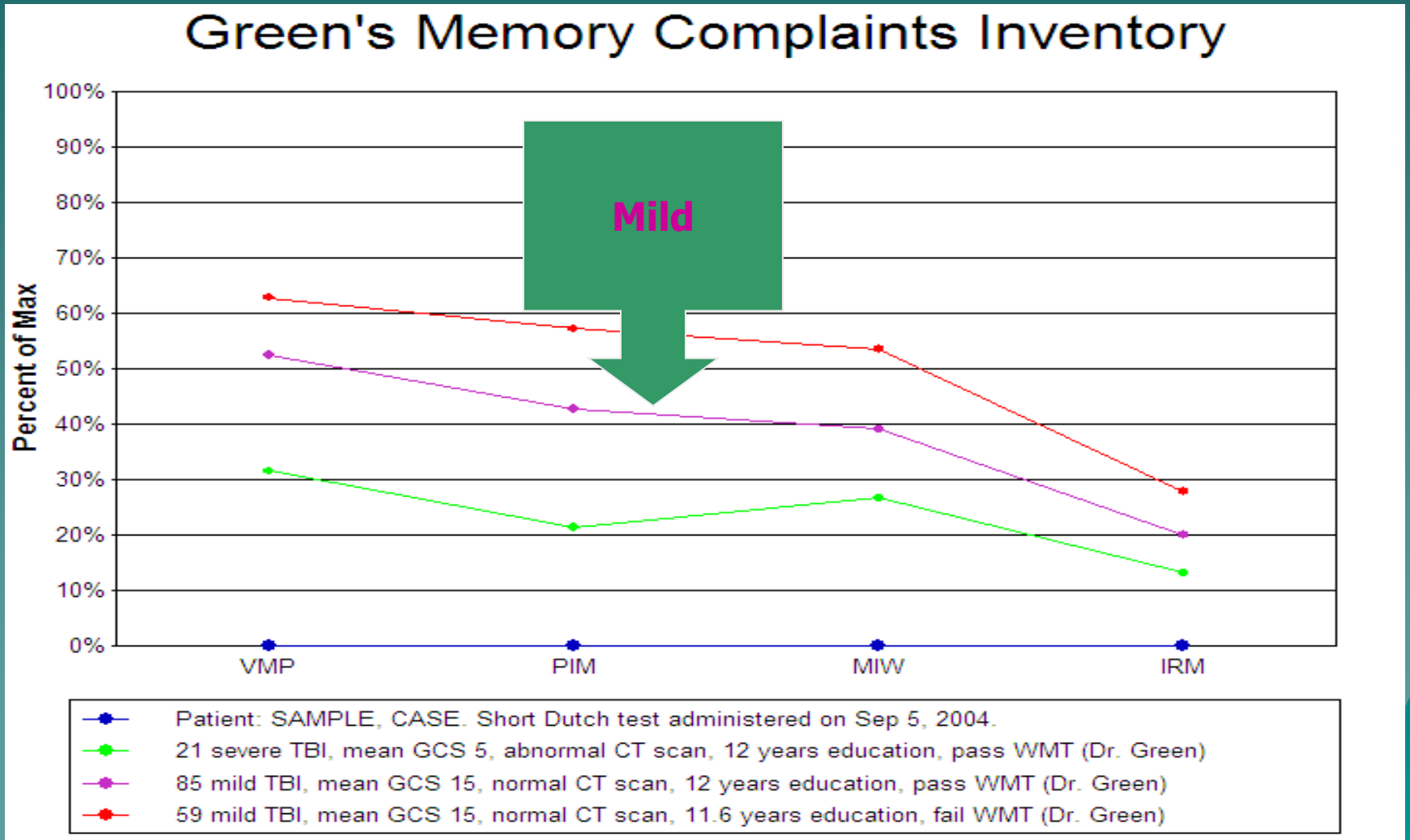
Top 4 memory complainers are:-

CFS / Chronic pain / Fibromyalgia	40% of max
Major depression	36% of max
Mild head injury	34% of max

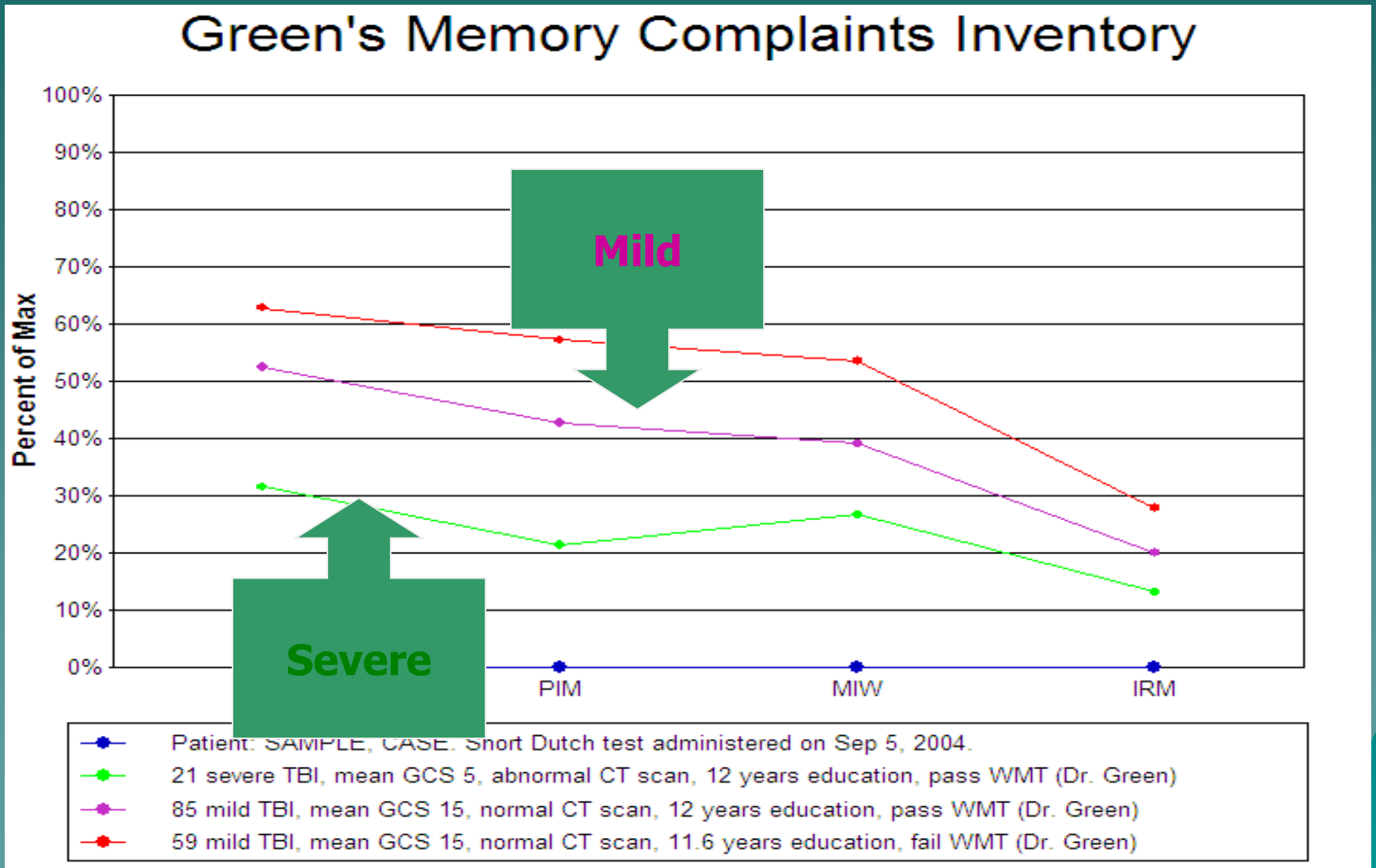
- And the lowest are:-

Neurological	20% of max
Severe brain injury	25% of max
Orthopedic	23% of max

More memory complaints in mild than in severe brain injury

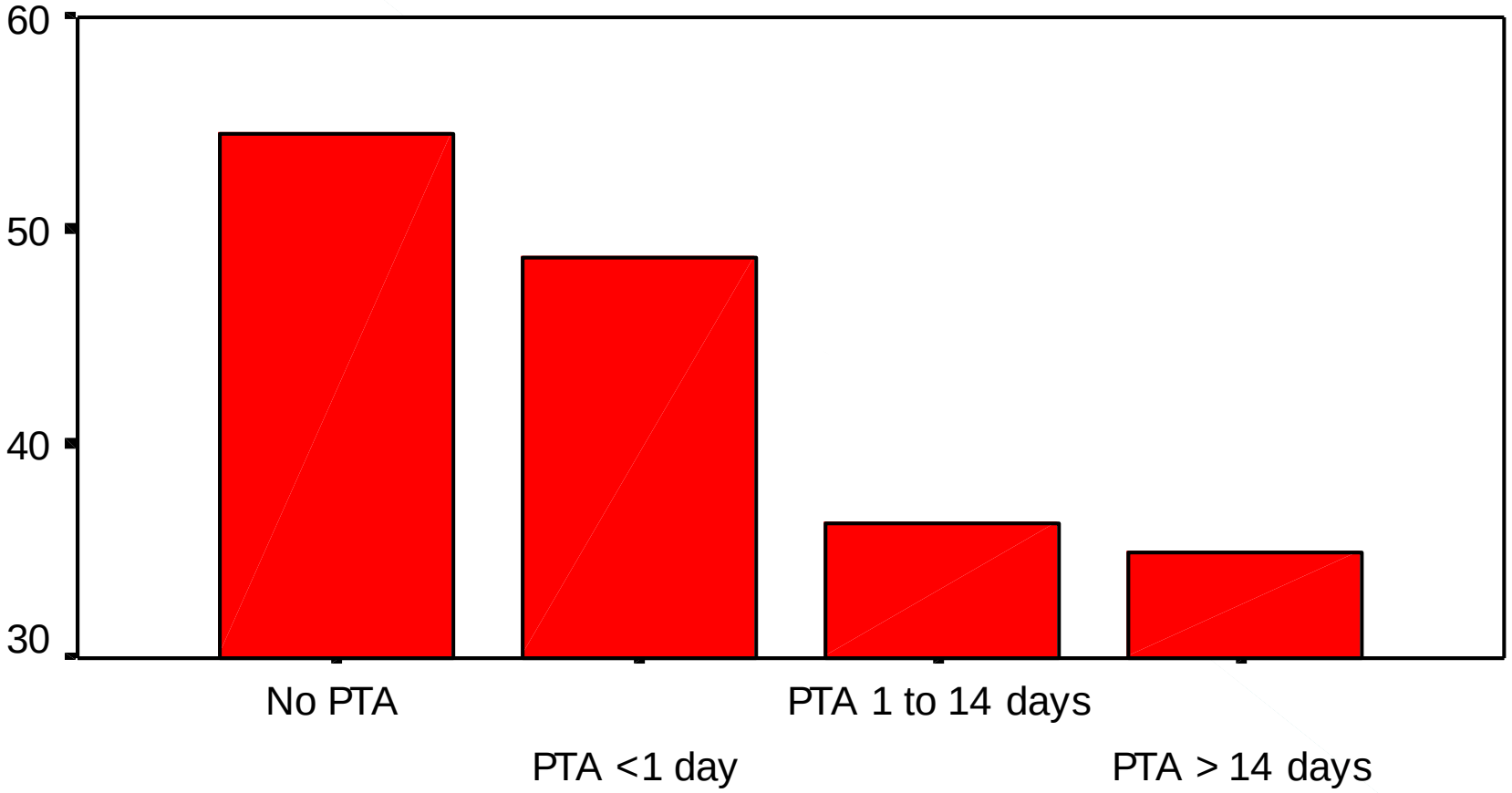


More memory complaints in mild than in severe brain injury (claimants)

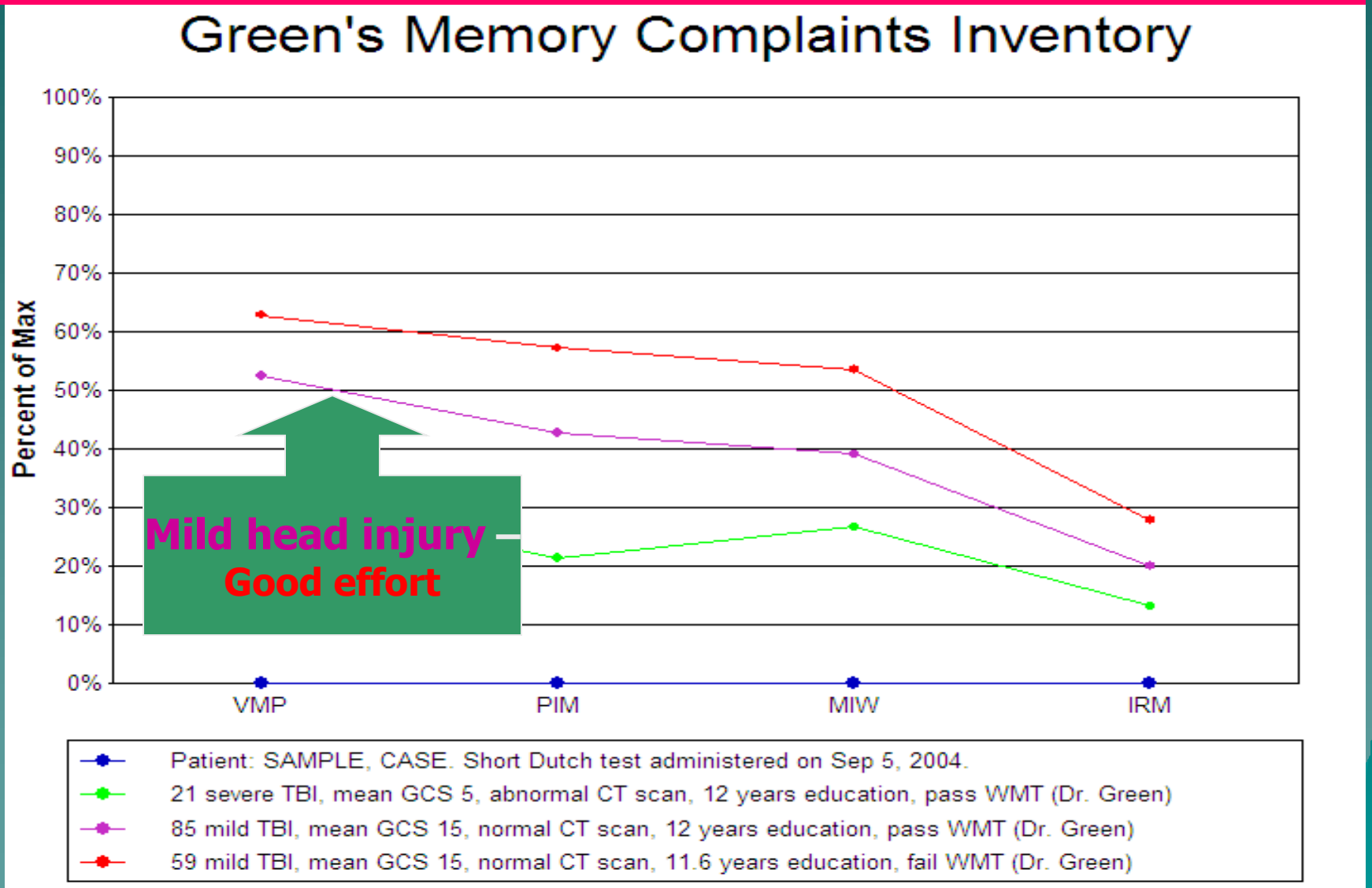


Memory complaints greater in cases with least Post Traumatic Amnesia

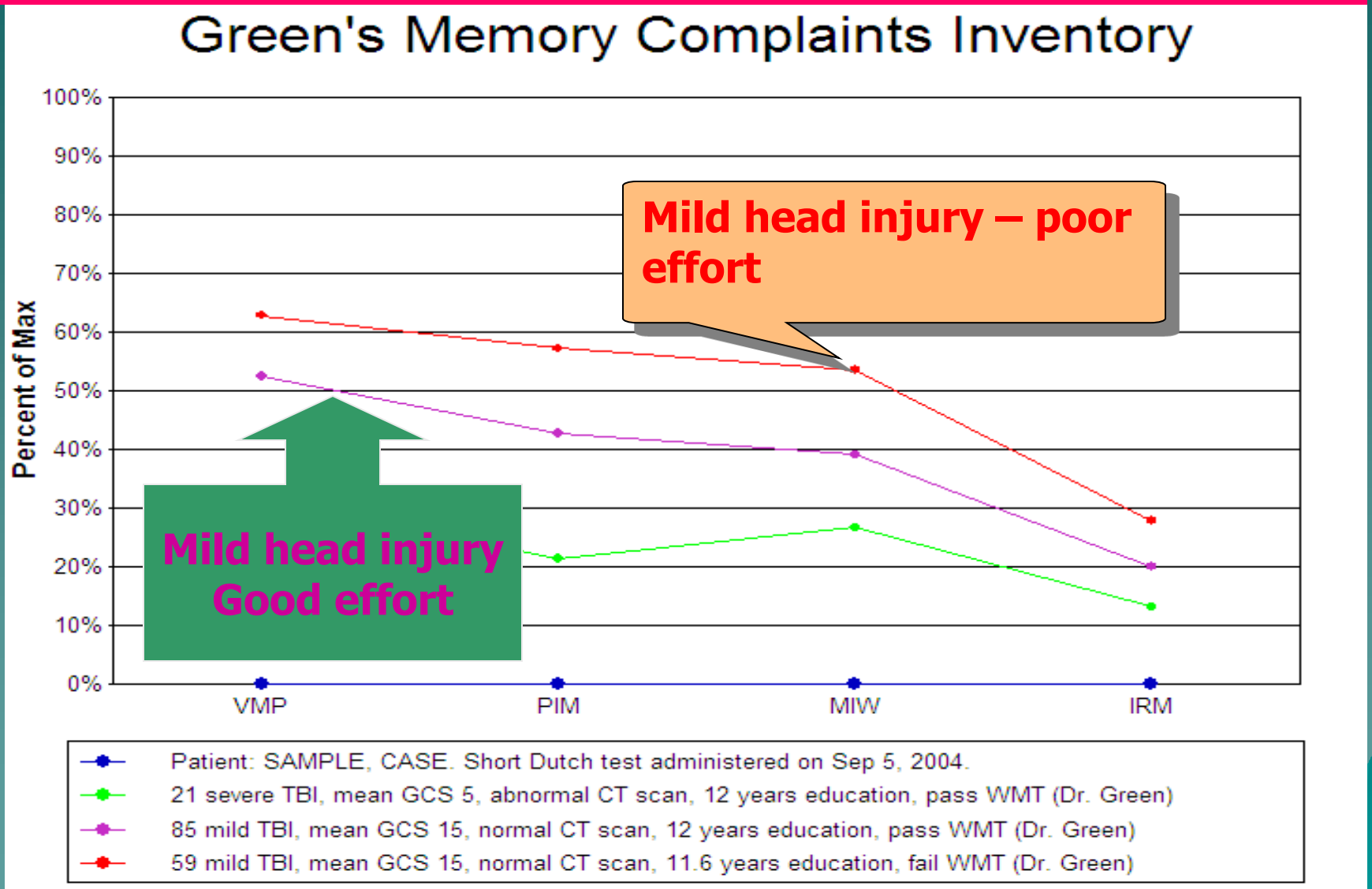
(a function of exaggeration)



More memory complaints in poor effort than in good effort cases



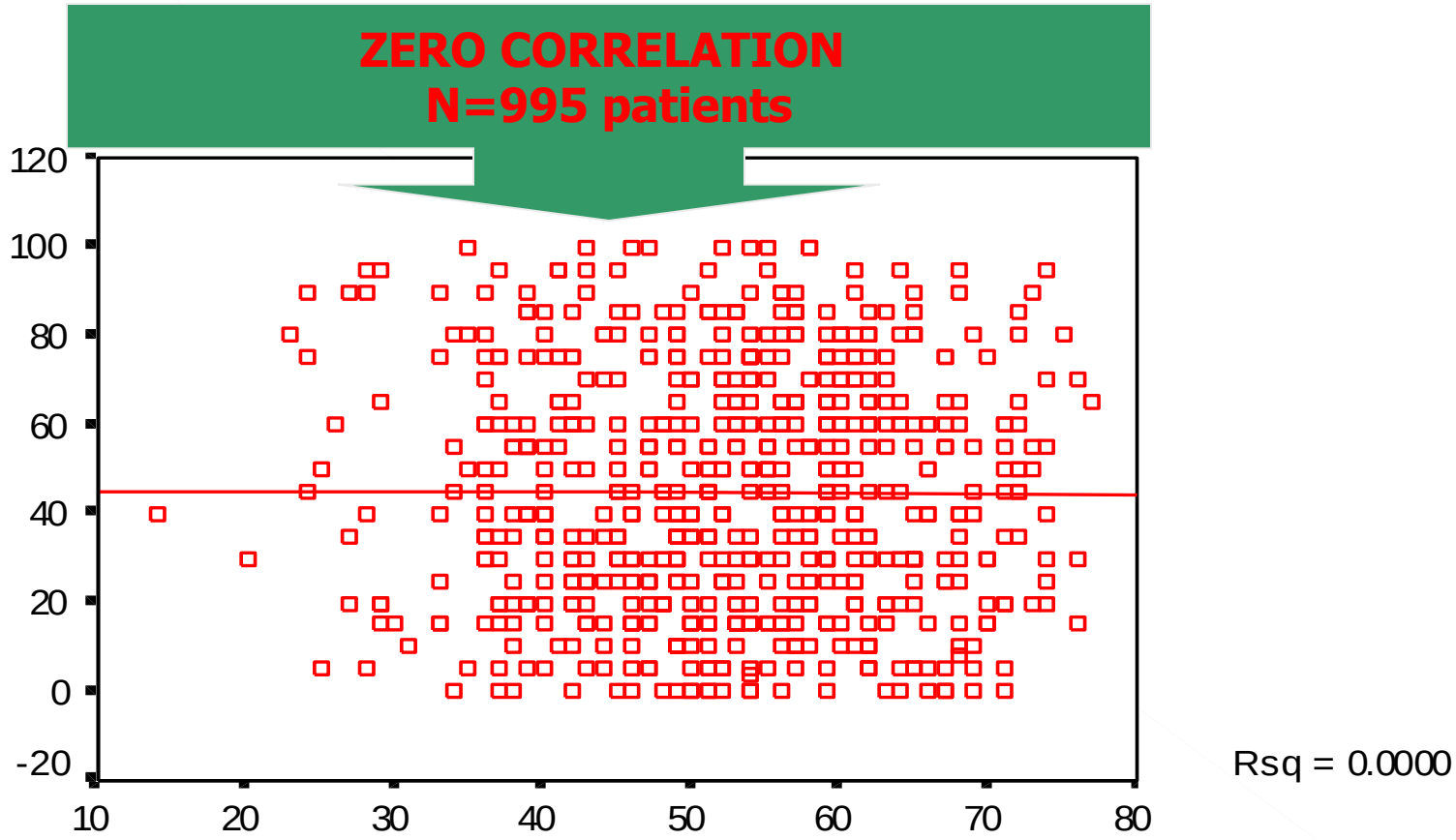
More memory complaints in poor effort than in good effort cases



MEMORY COMPLAINTS ARE NOT INDICATORS OF BRAIN FUNCTION or OBJECTIVE MEMORY ABILITY.

- There is no correlation between memory complaints and performance on actual memory tests
- The next graph shows a perfect non-correlation

Verbal memory complaints versus verbal memory test scores

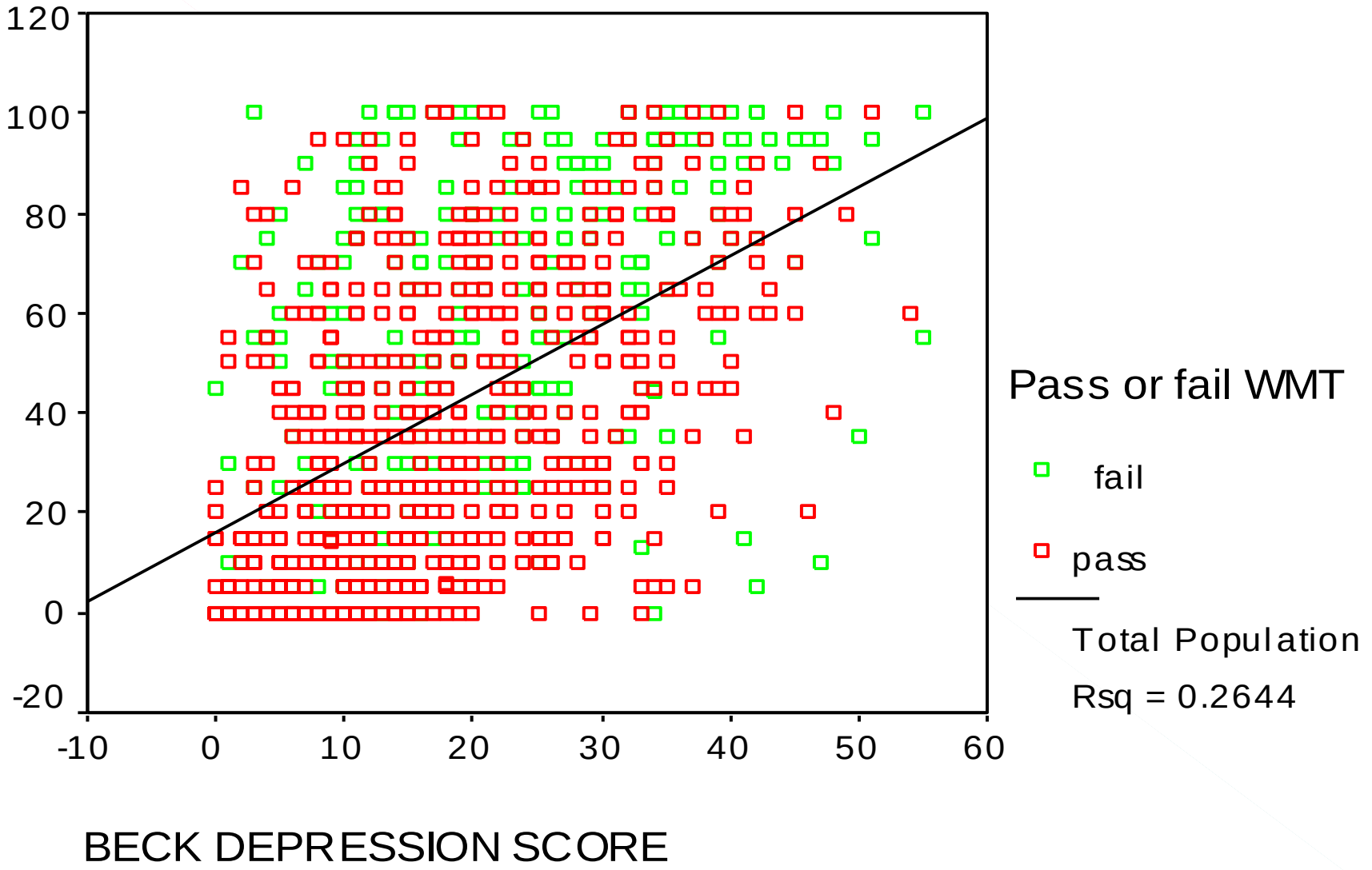


Memory complaints tell us nothing about memory test scores

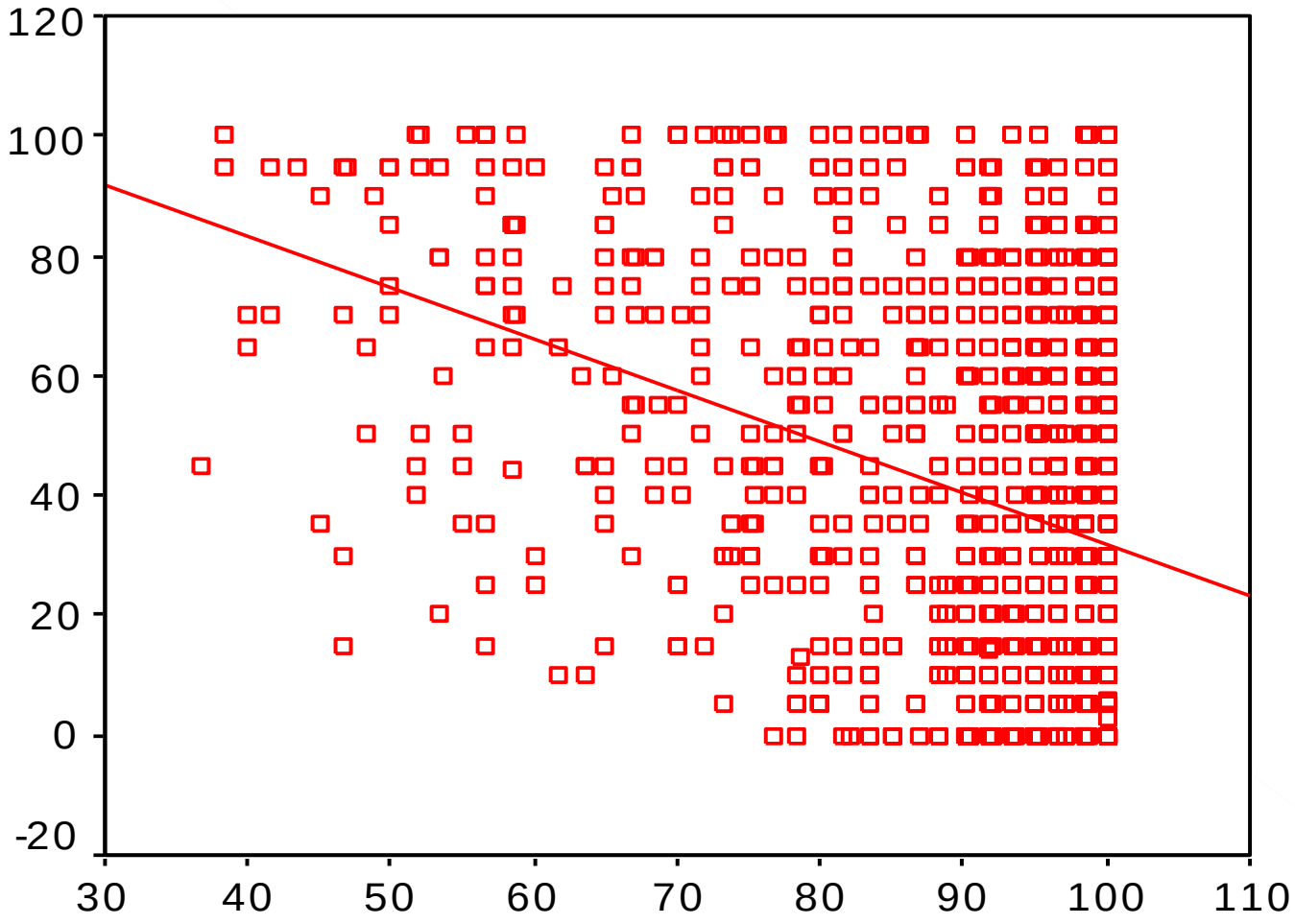
BUT MEMORY COMPLAINTS ARE LINKED WITH

- **A) PSYCHIATRIC SYMPTOMS** (0.4 to 0.6 correlation with Beck Depression scores)
- & **B) SYMPTOM EXAGGERATION** (0.33 correlation with effort test scores).

Memory complaints are a function of depression



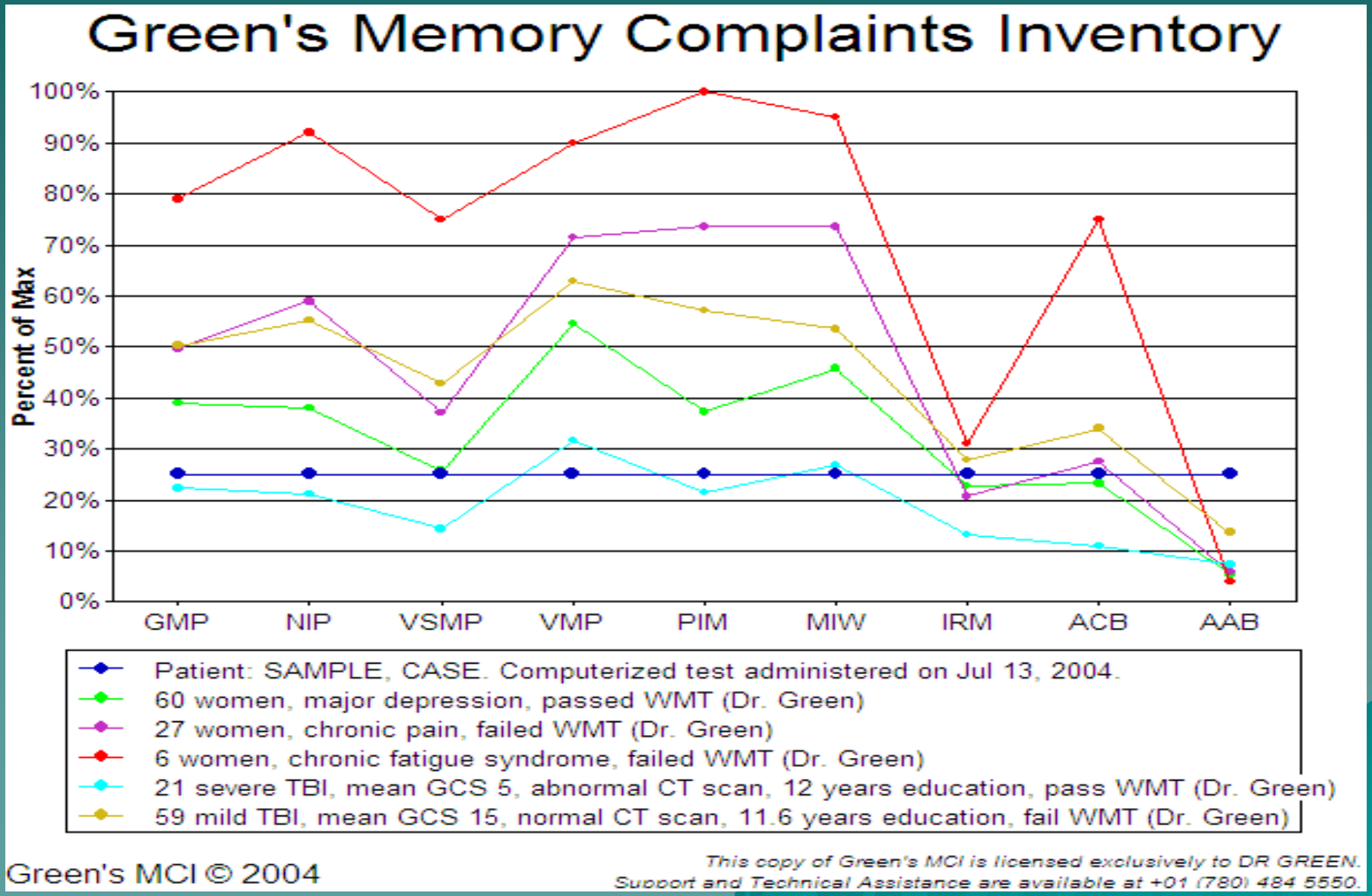
Memory complaints are a function of symptom exaggeration



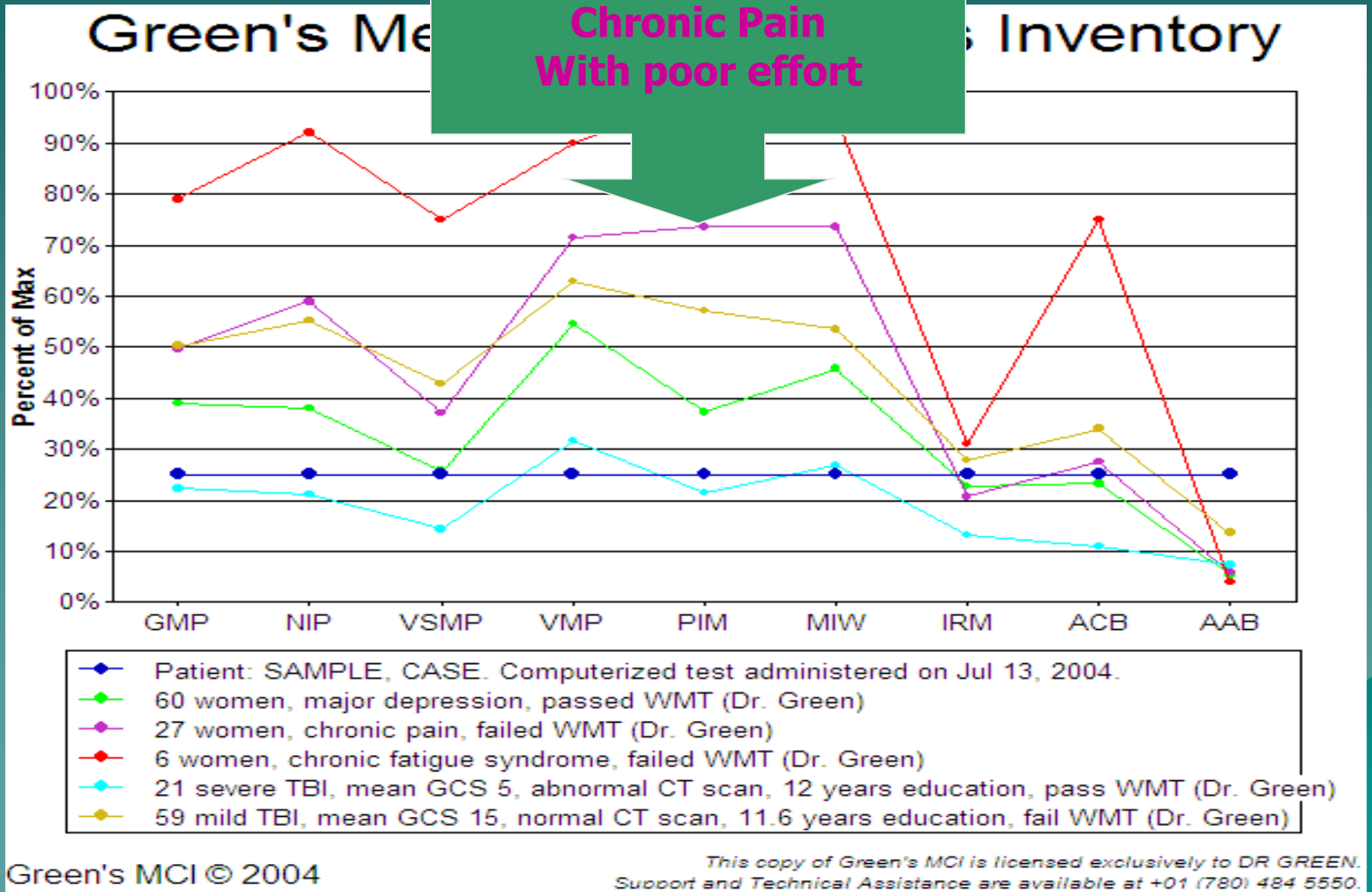
Best predictors of memory complaints in a regression (in order) are:-

- Beck Depression scores
- Reported headaches
- **Effort** on the WMT
- MMPI-2 depression and K scale
- Pain rating on scale 0 to 5
- Gender
- These variables explain 46% of variance in memory complaints
- CT abnormalities or head injury severity add nothing

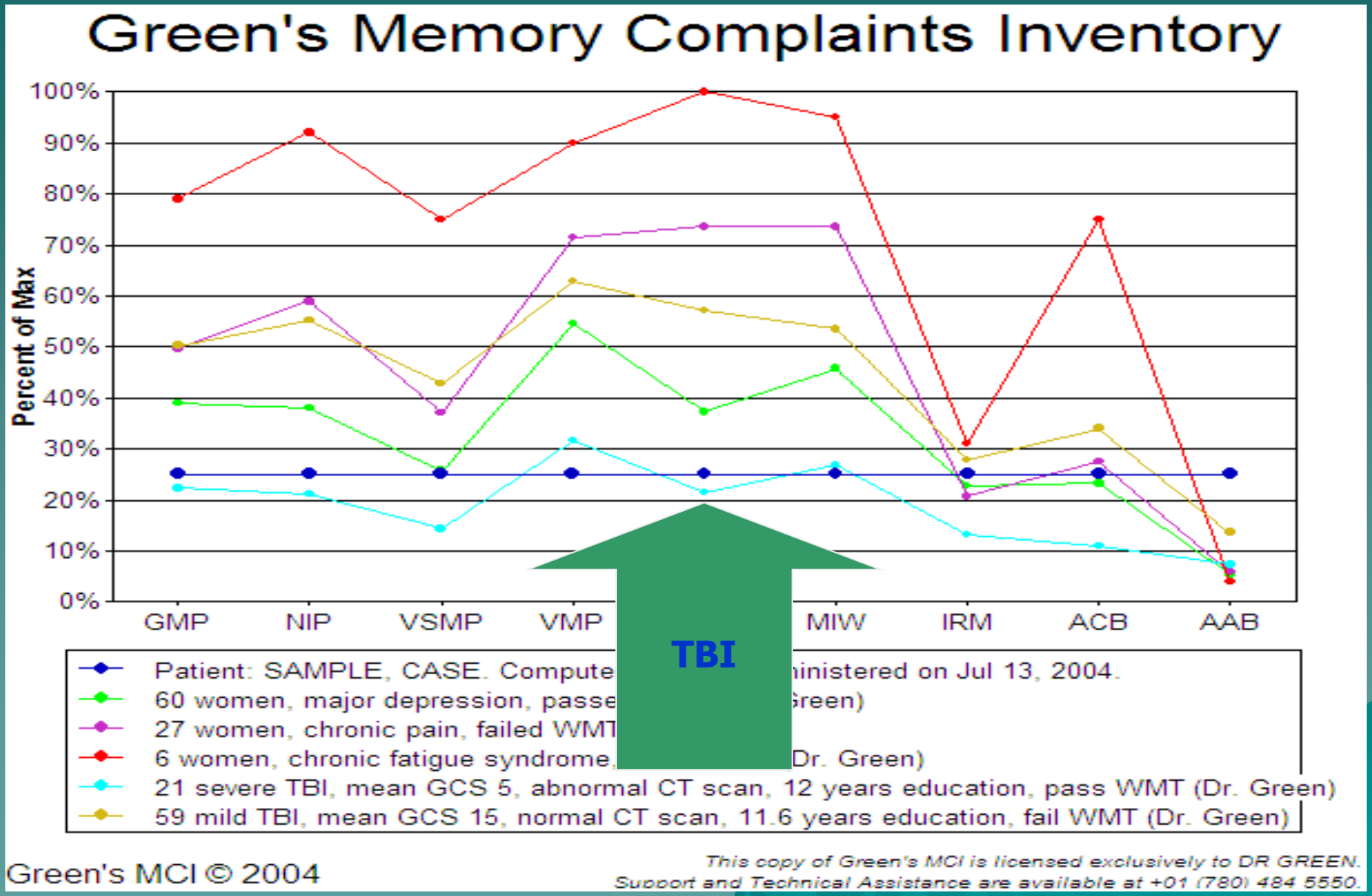
Chronic pain and depression patients have more memory complaints than those with severe TBI



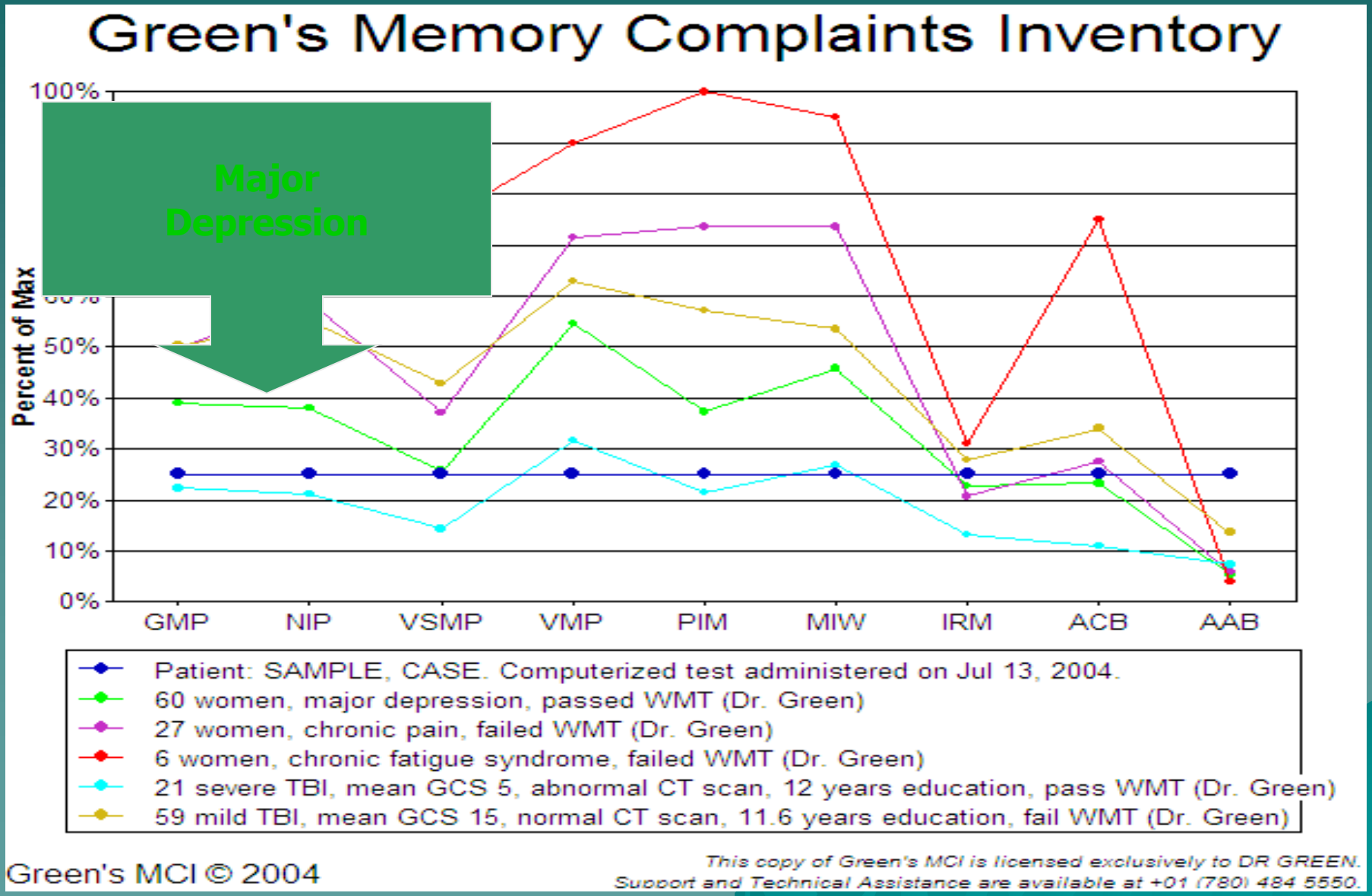
Chronic pain and depression patients have more memory complaints than those with severe TBI



Chronic pain and depression patients have more memory complaints than those with severe TBI



Chronic pain and depression patients have more memory complaints than those with severe TBI

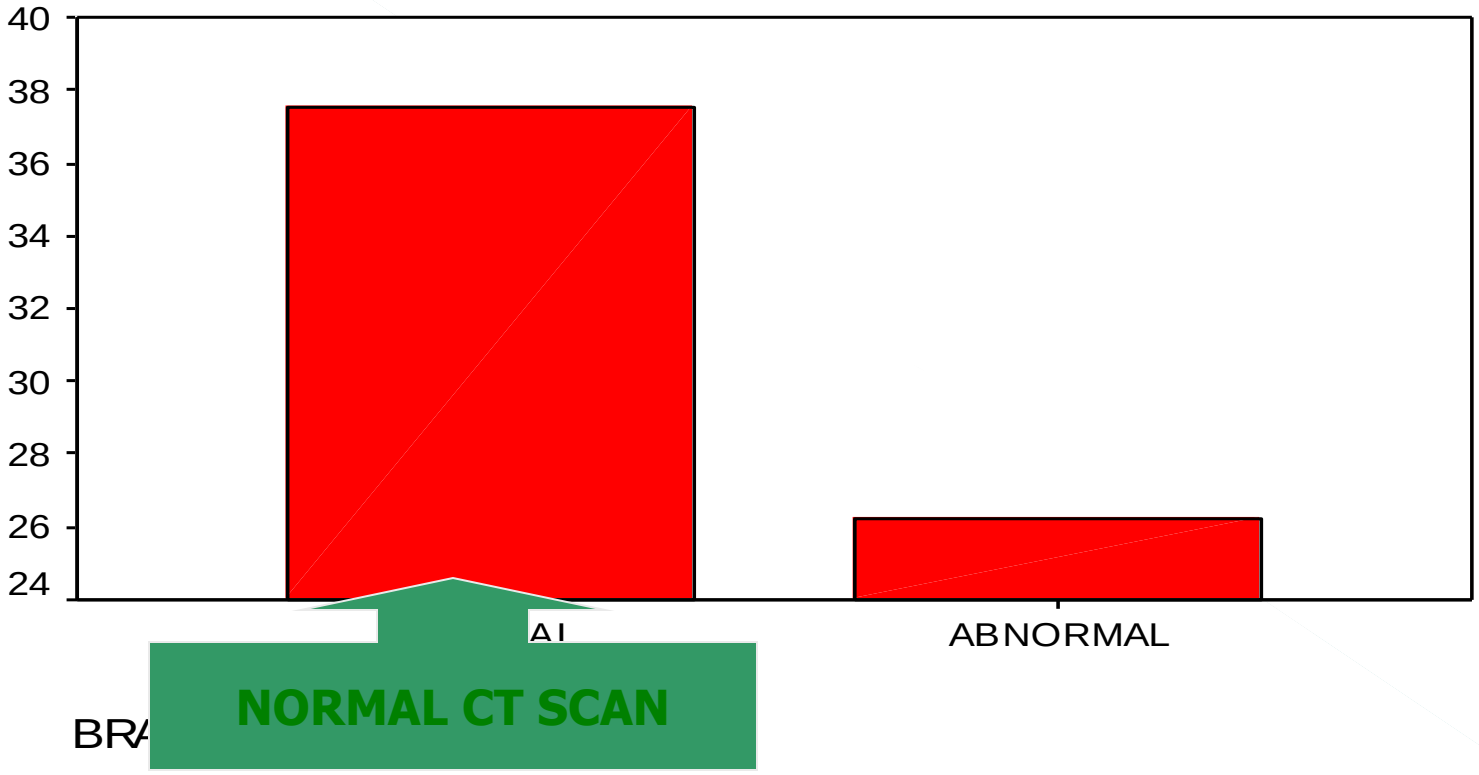


More complaints in normal than in abnormal brain scan (N=568)

Mean MCI memory complaints

by normal/abnormal brain scan

Difference significant at p<.0001

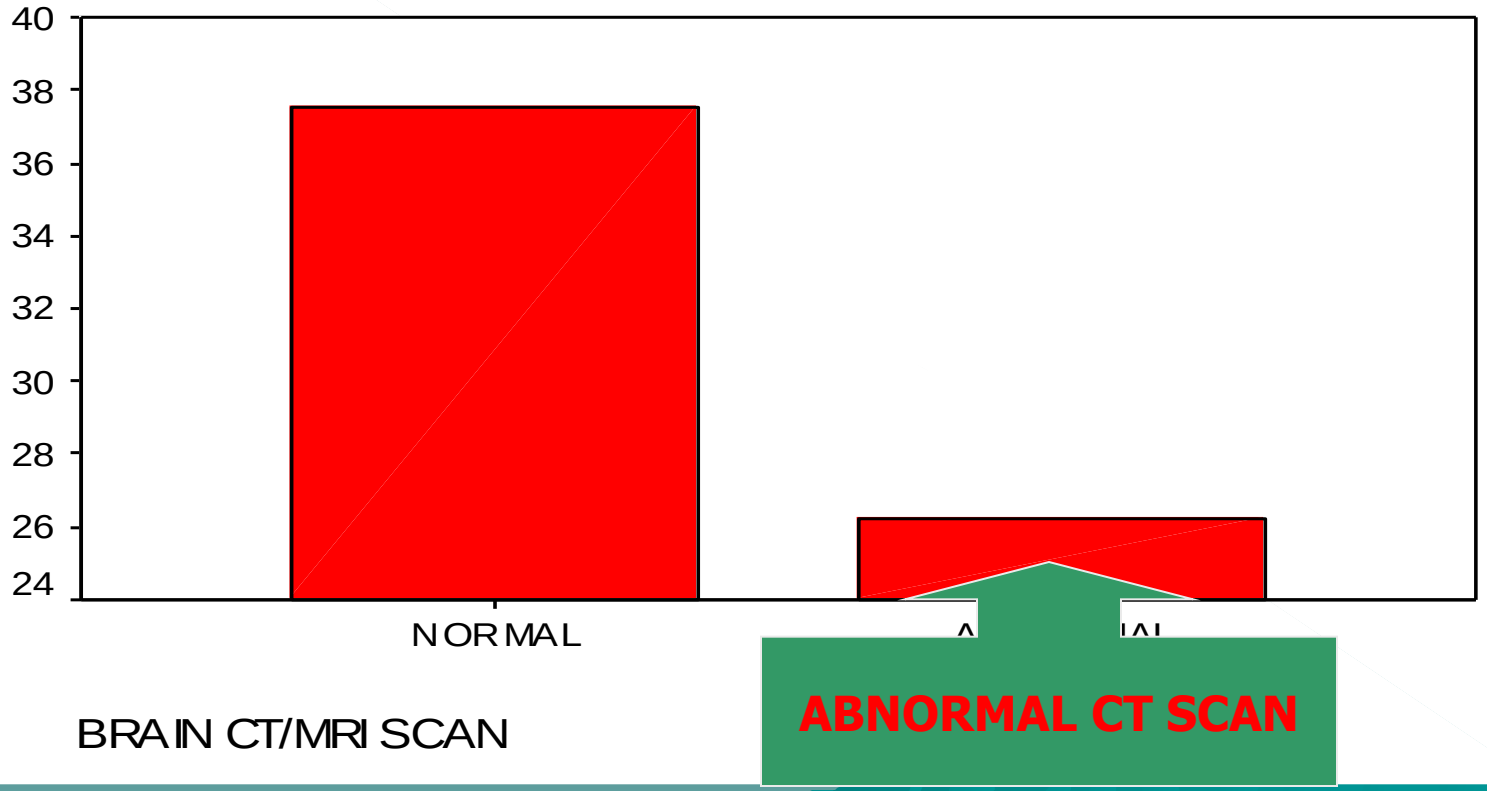


More complaints in normal than in abnormal brain scan (N=568)

Mean MCI memory complaints

by normal/abnormal brain scan

Difference significant at p<.0001



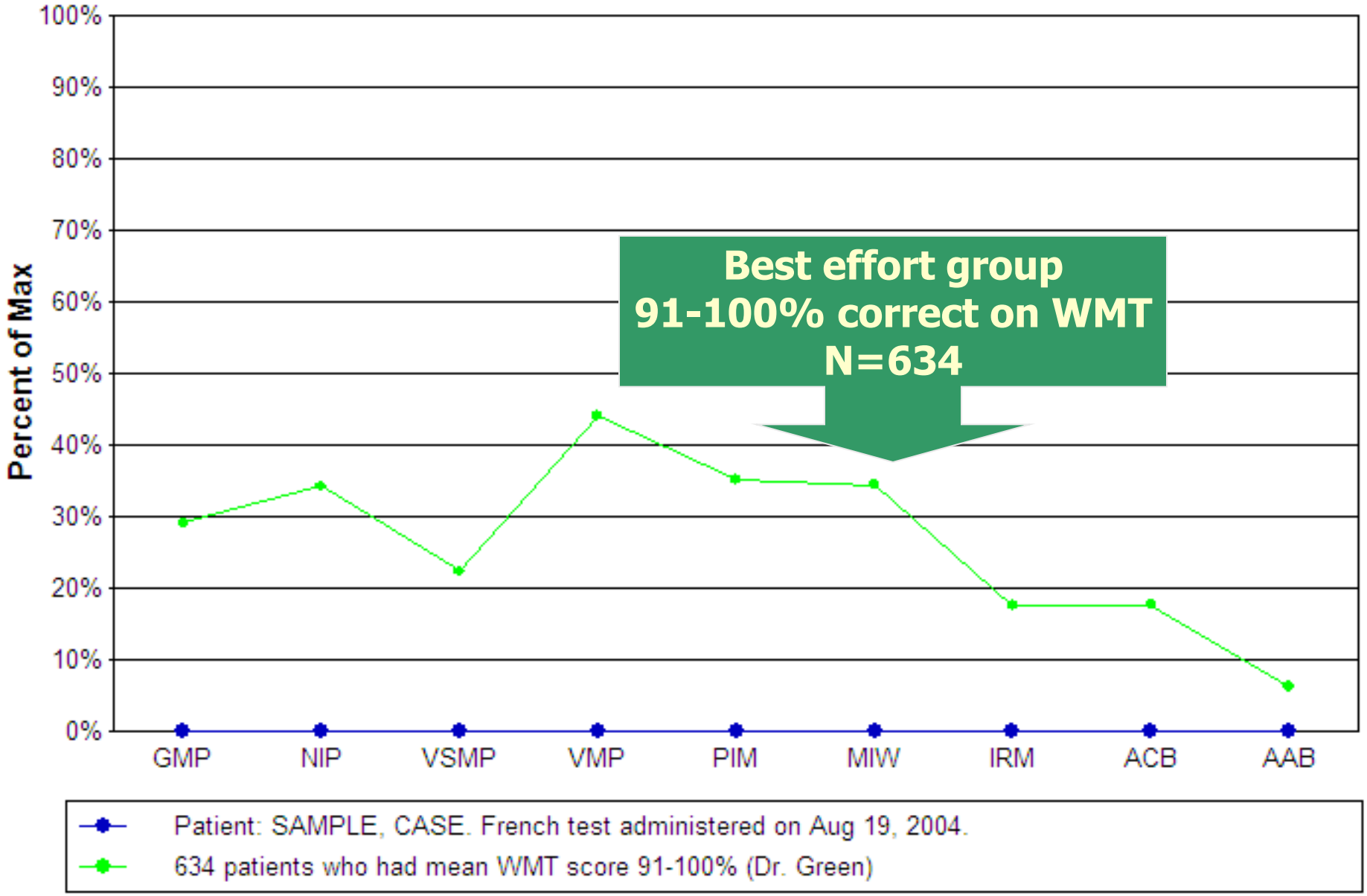
For these reasons, subjective complaints must be viewed with caution

- Memory complaints are most prevalent in people with chronic pain, chronic fatigue syndrome & depression
- They do not correlate with brain damage or memory test scores
- They do correlate with self-rated depression

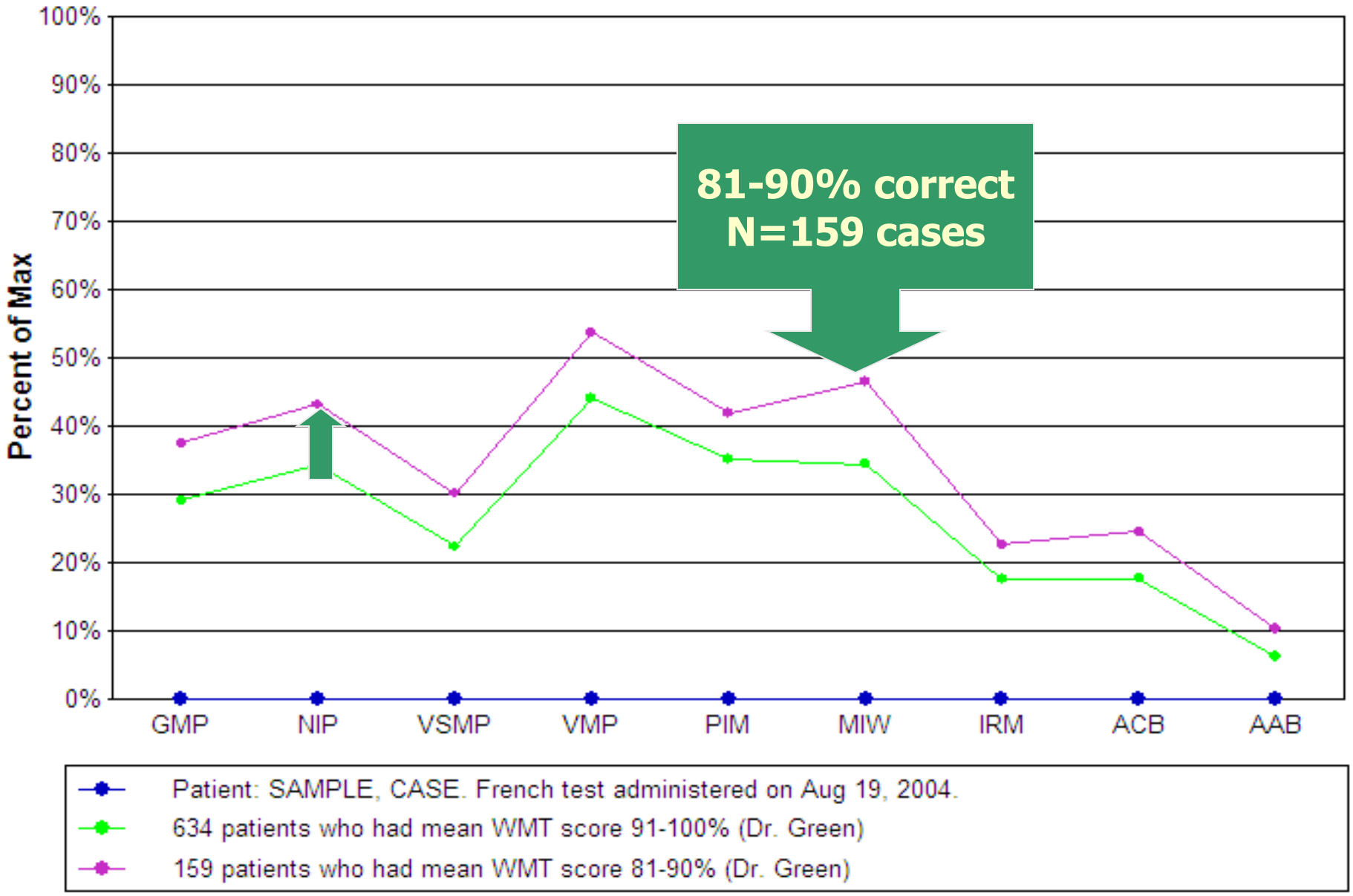
Memory complaints are also strongly linked with effort

- This is evident on all 9 scales of the memory complaints inventory (MCI) in a sample of over 1,000 patients.

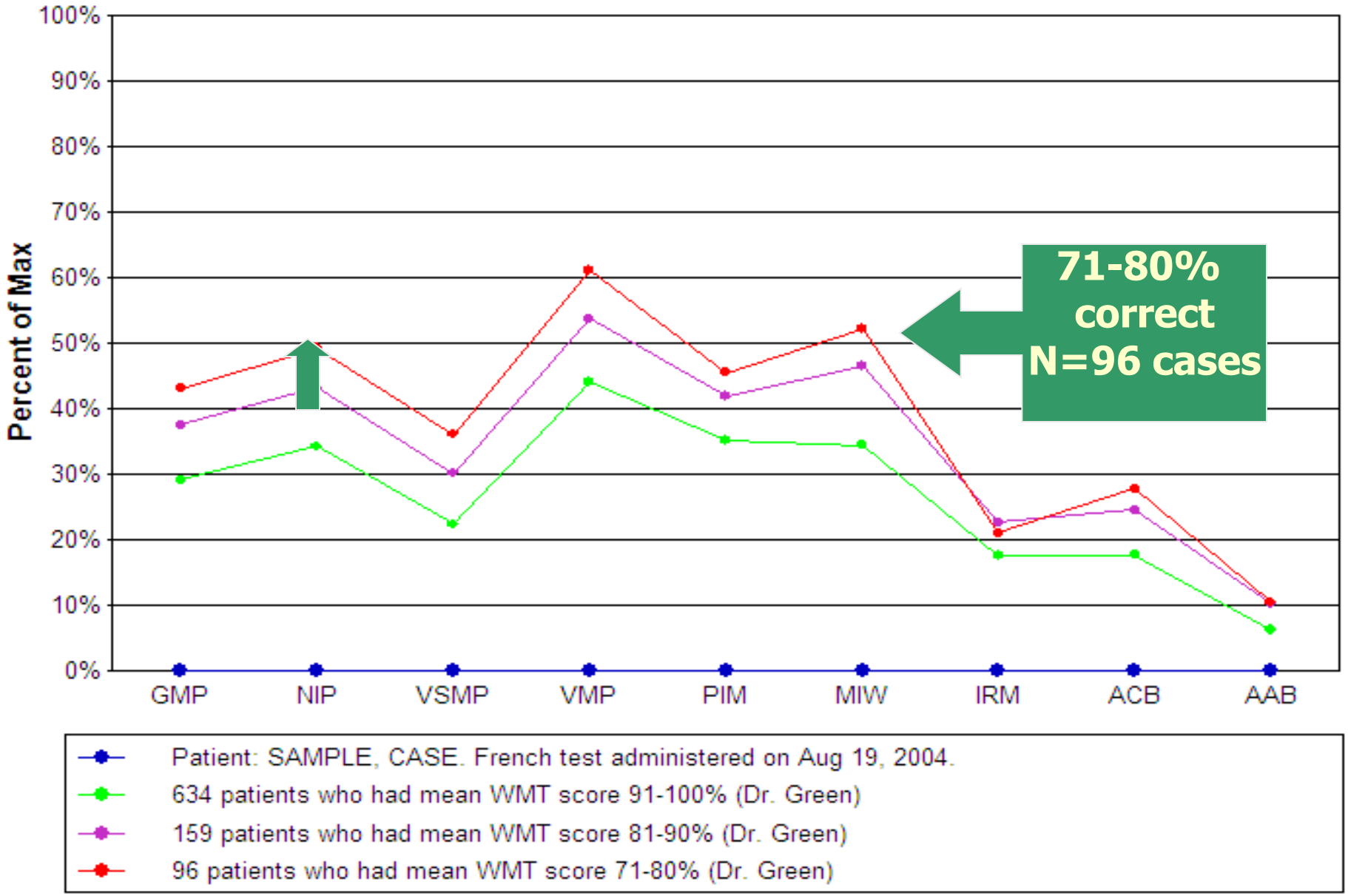
Green's Memory Complaints Inventory



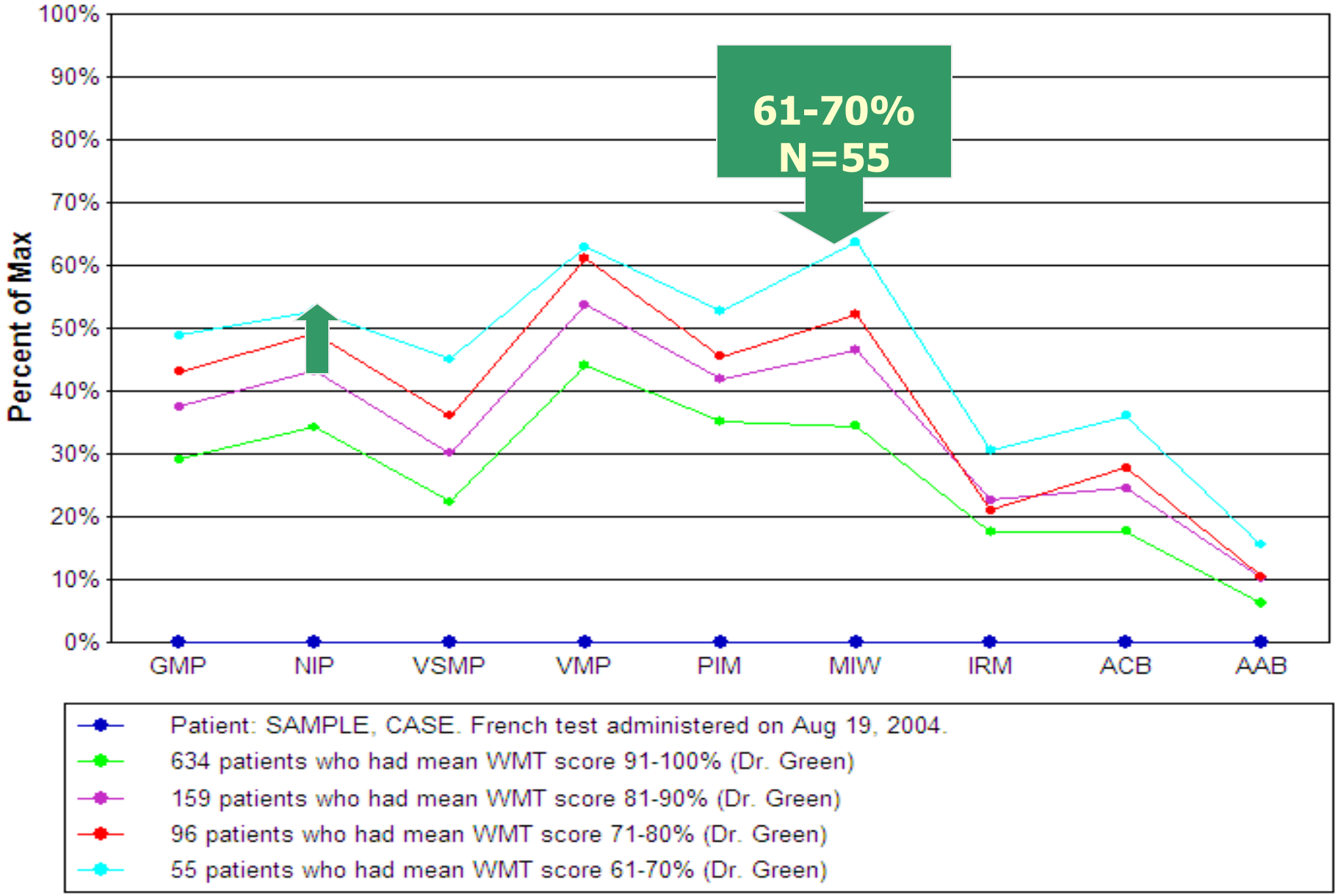
Green's Memory Complaints Inventory



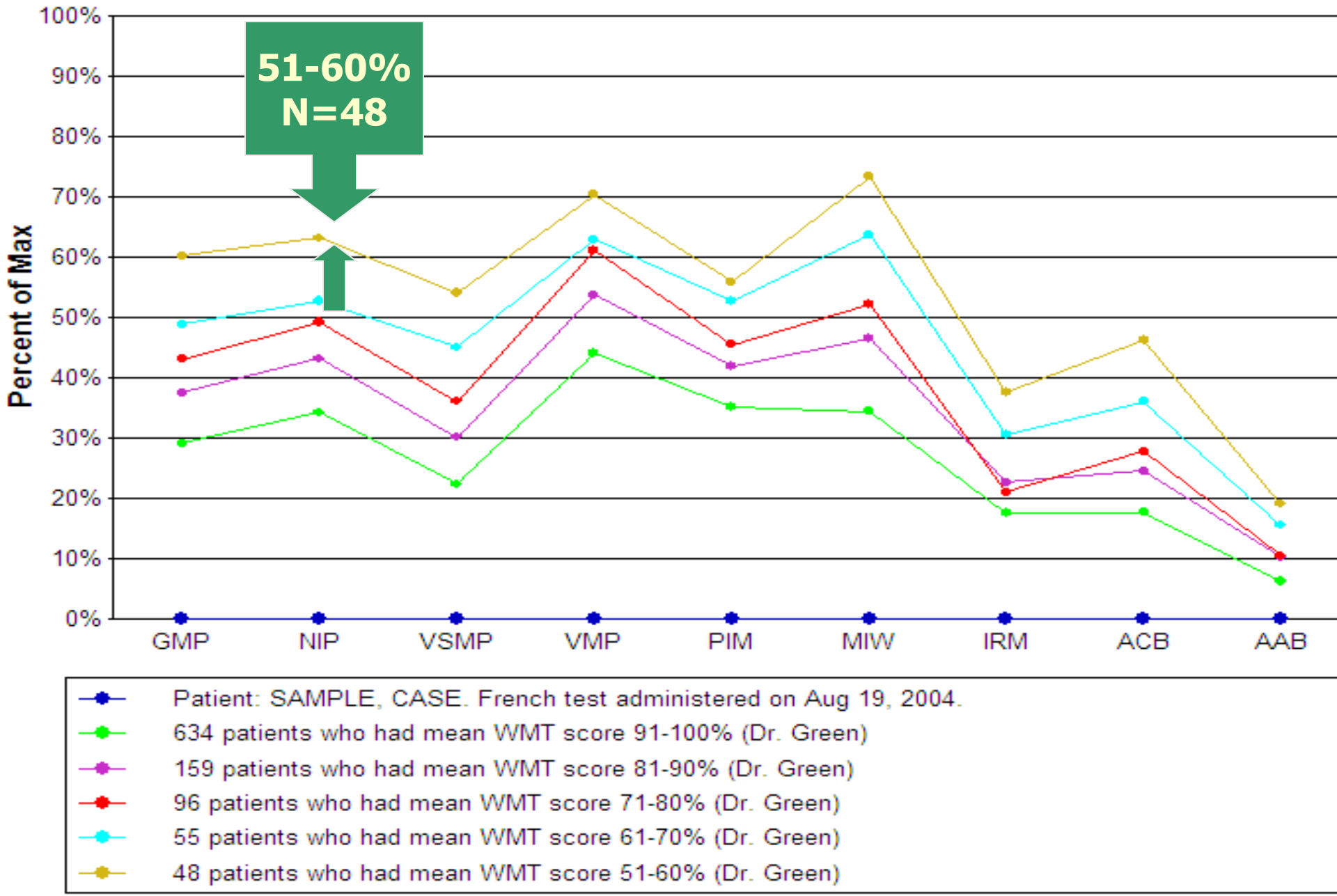
Green's Memory Complaints Inventory



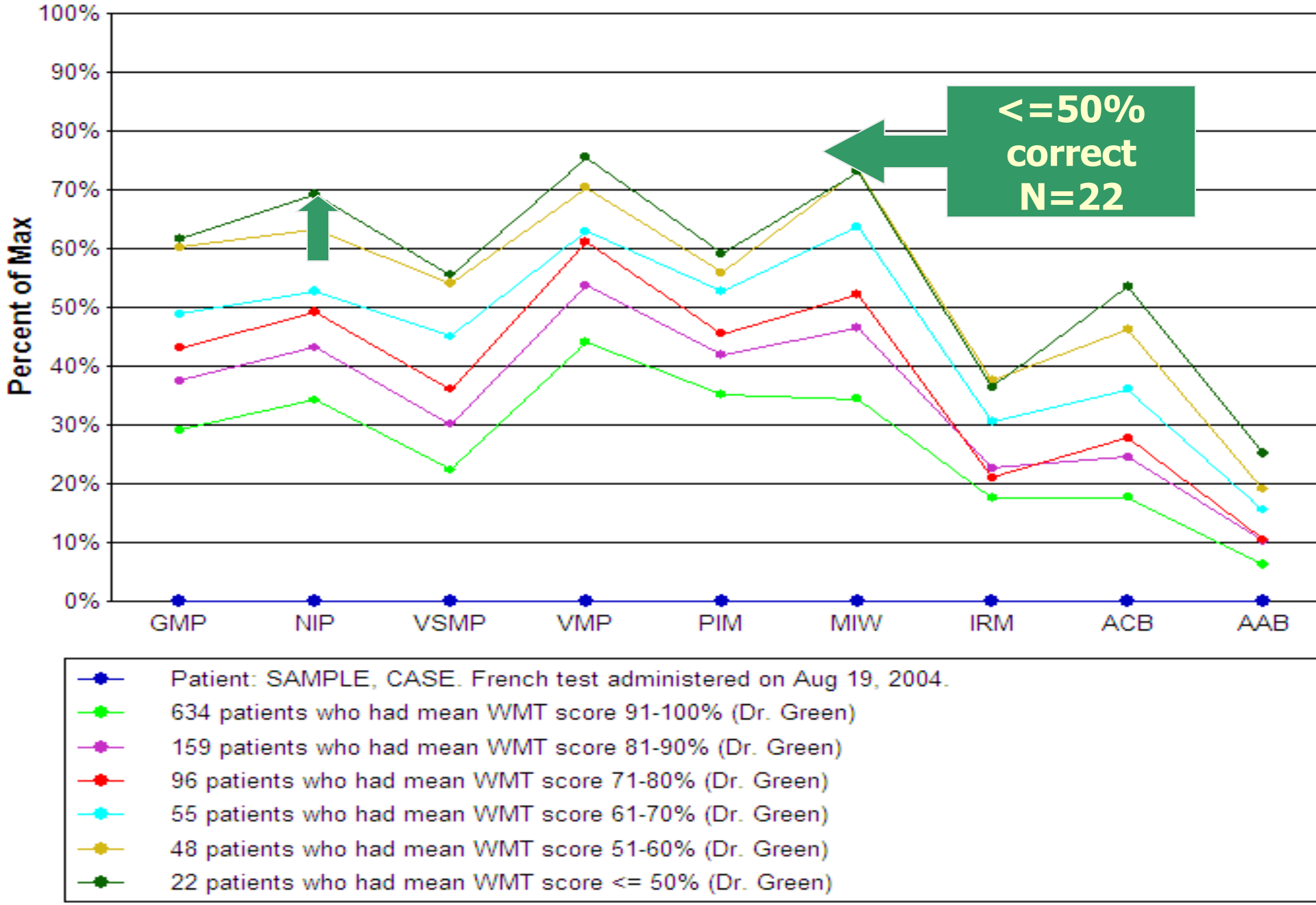
Green's Memory Complaints Inventory



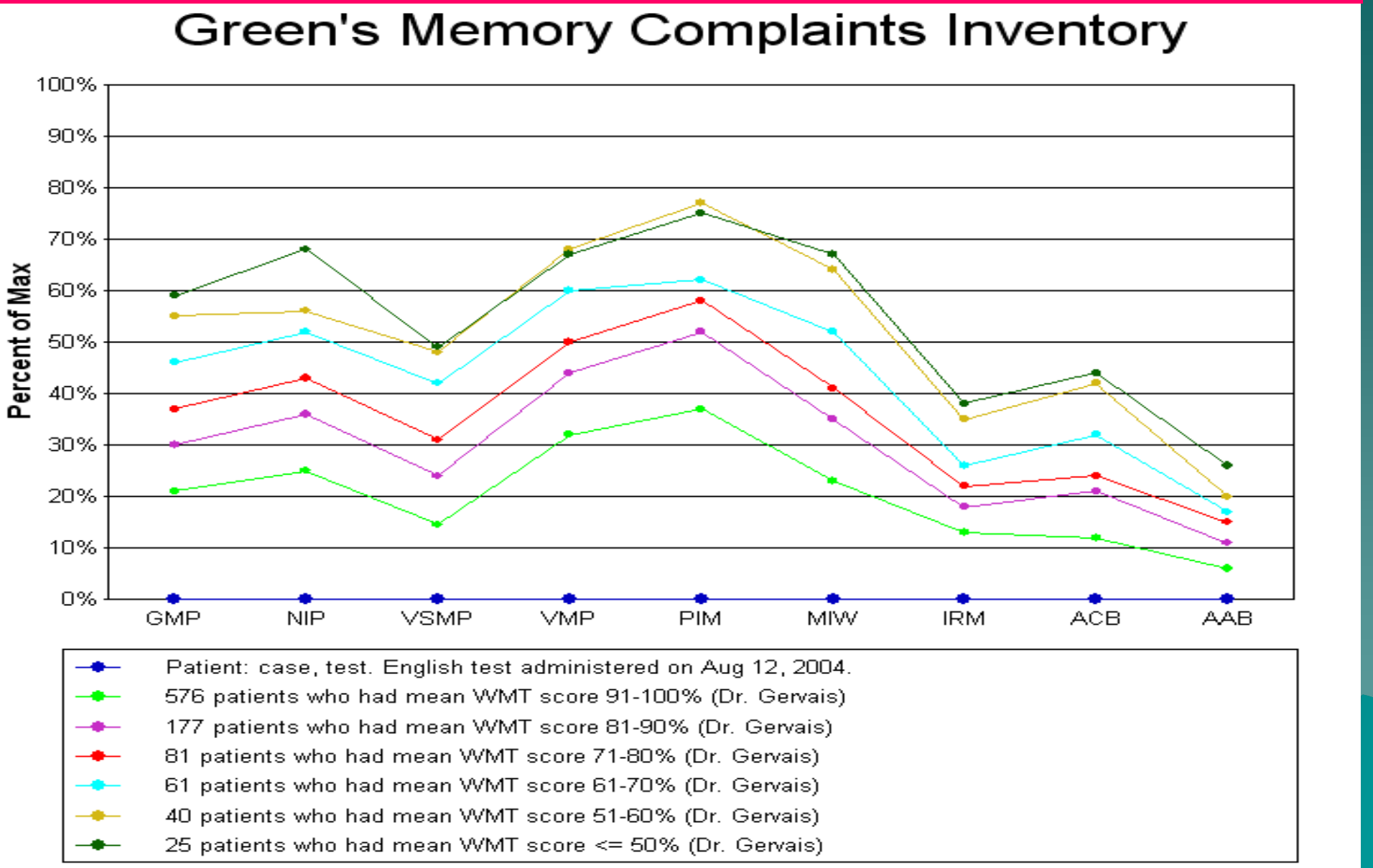
Green's Memory Complaints Inventory



Green's Memory Complaints Inventory



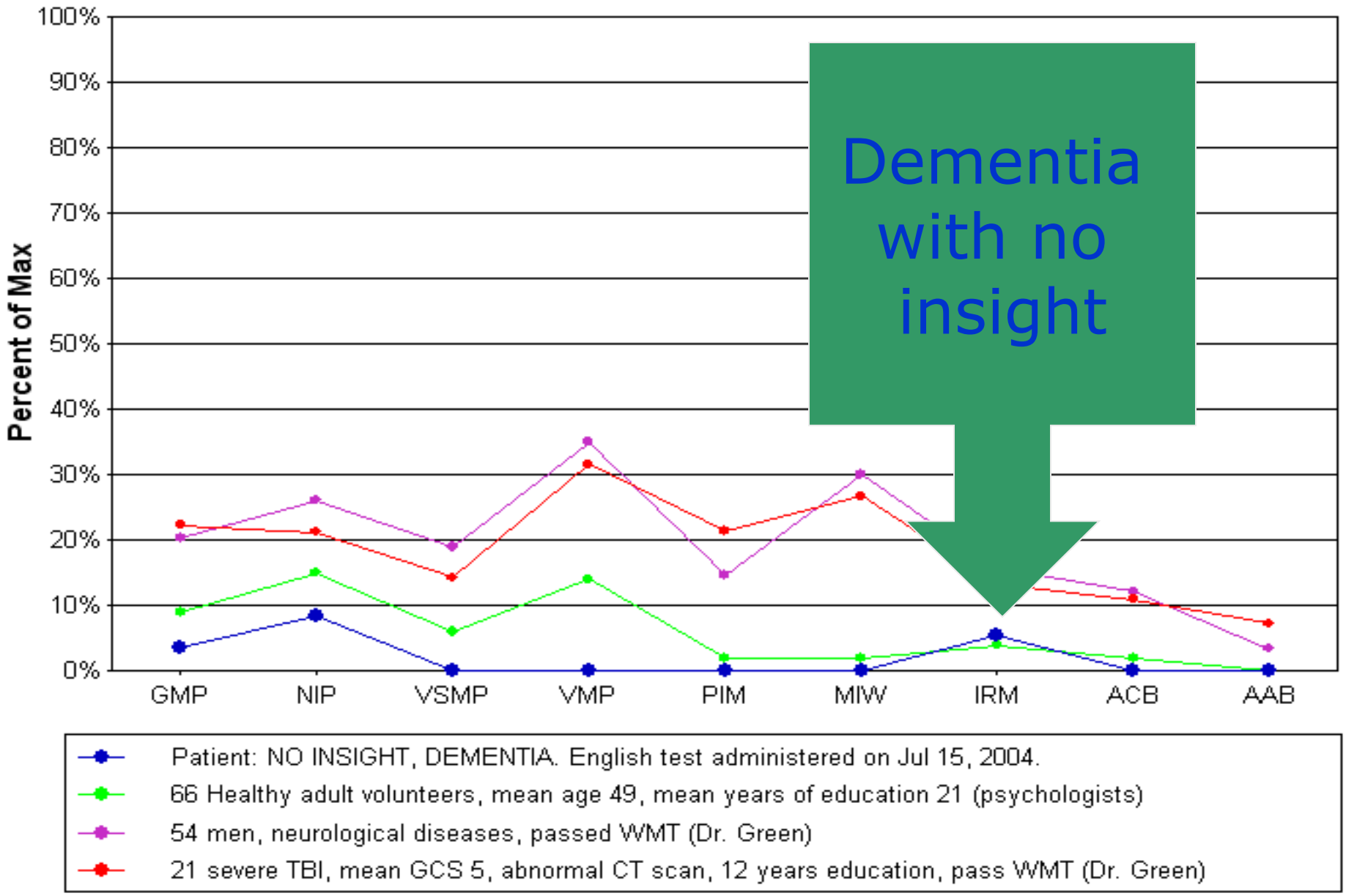
Independent data from R. Gervais



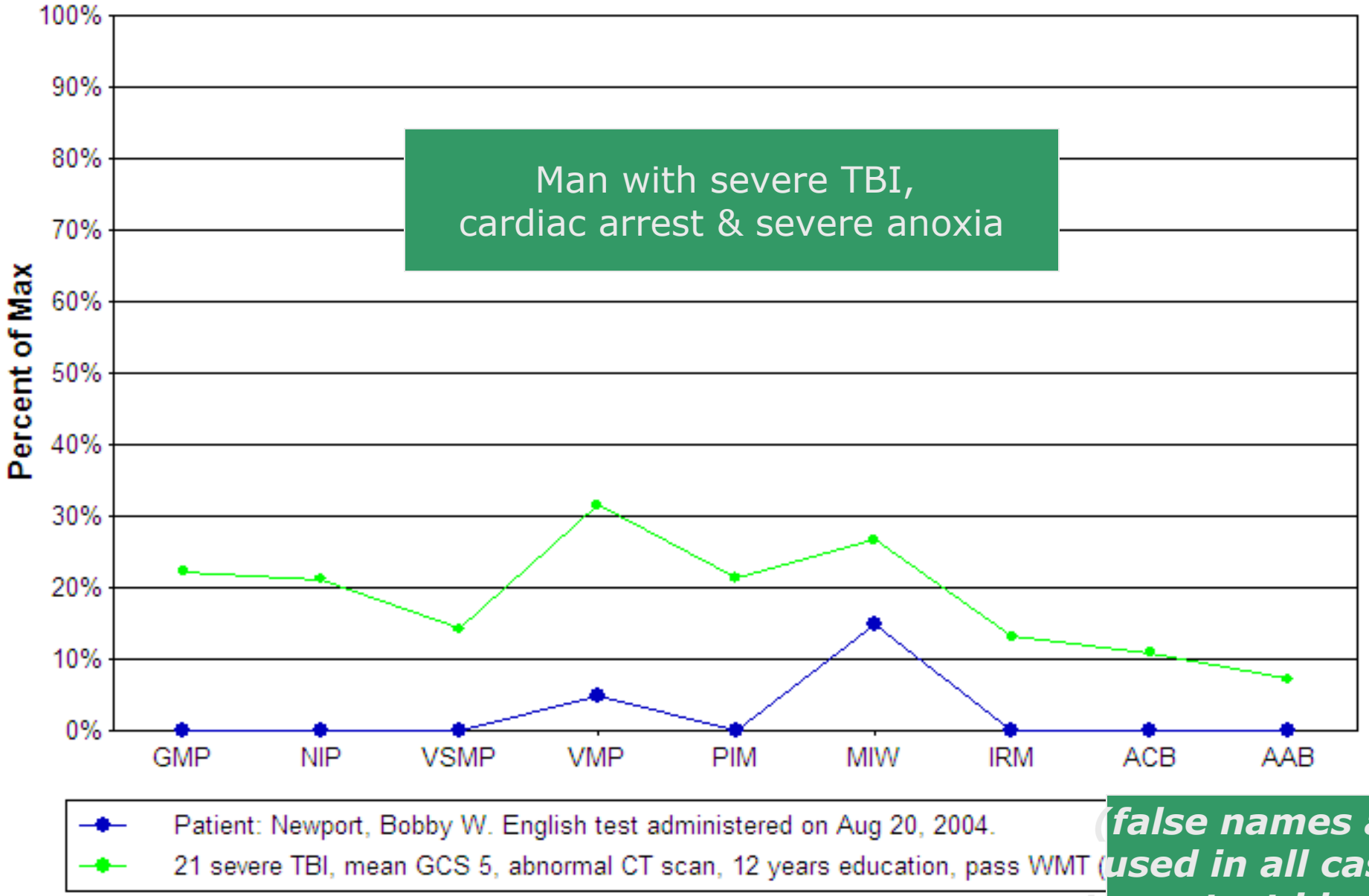
At the end of the day, such a strong link between poor effort and subjective memory complaints can only be explained by one thing:-

symptom exaggeration

Green's Memory Complaints Inventory

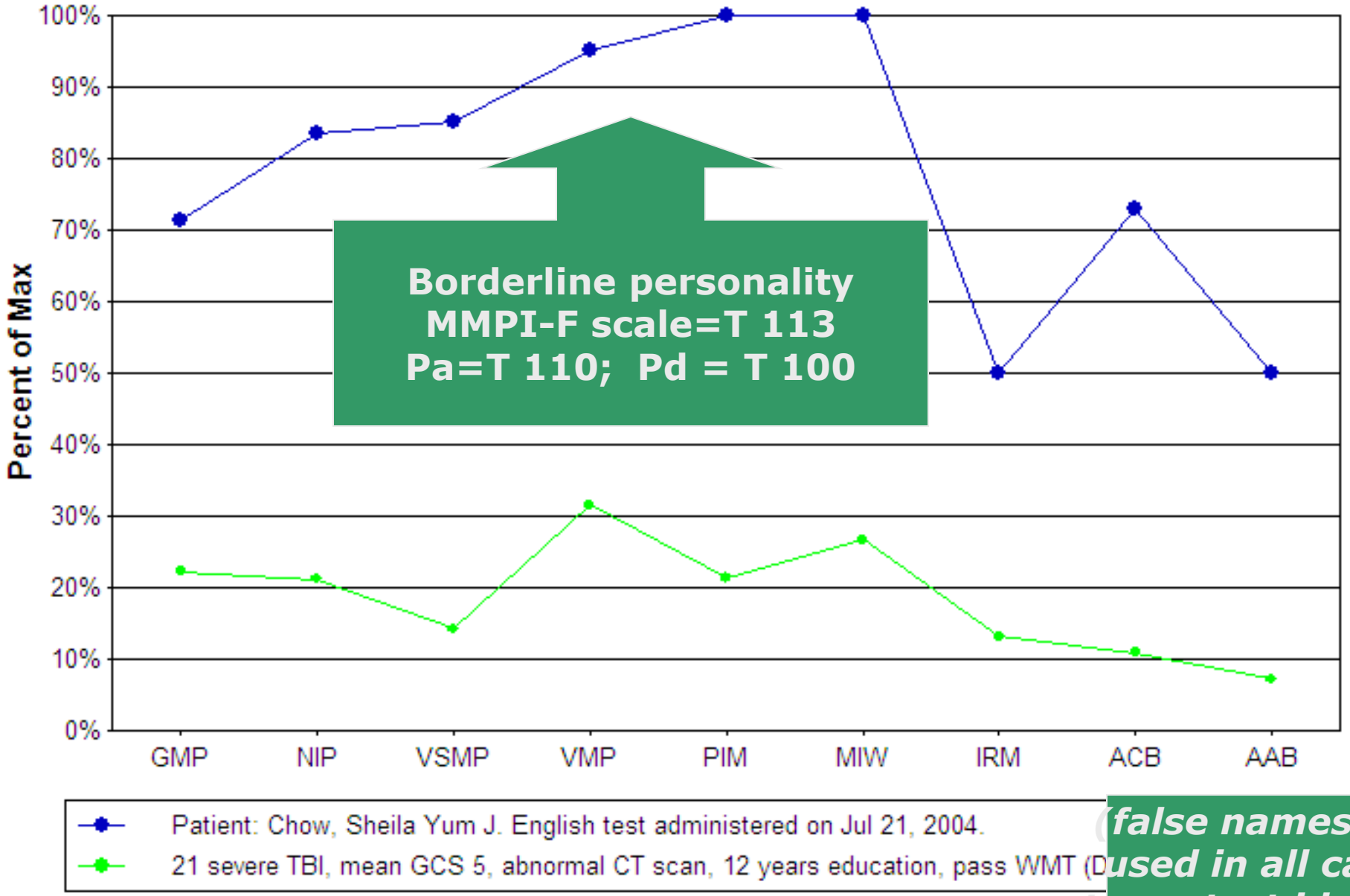


Green's Memory Complaints Inventory



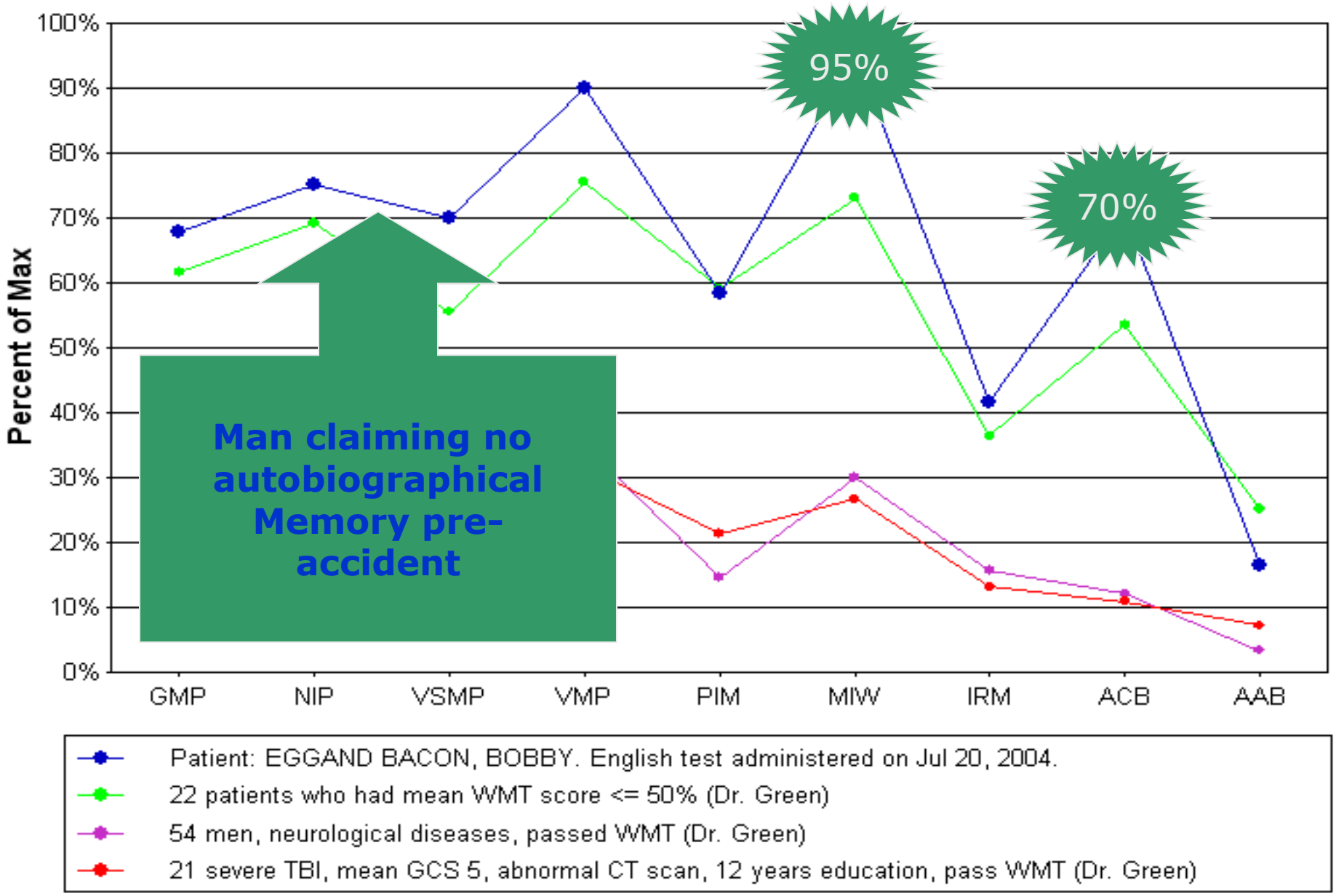
(false names are used in all cases to protect identity)

Green's Memory Complaints Inventory

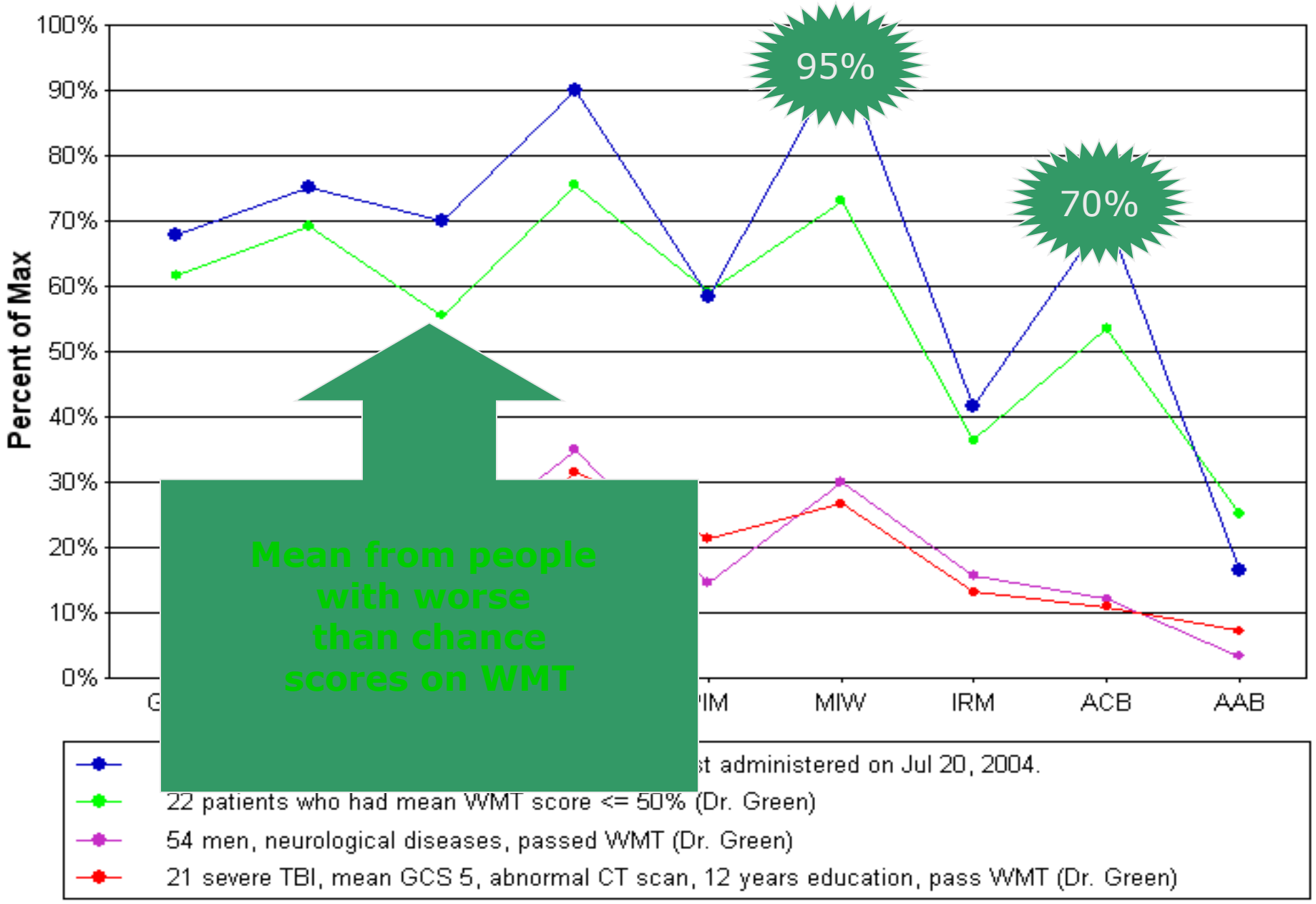


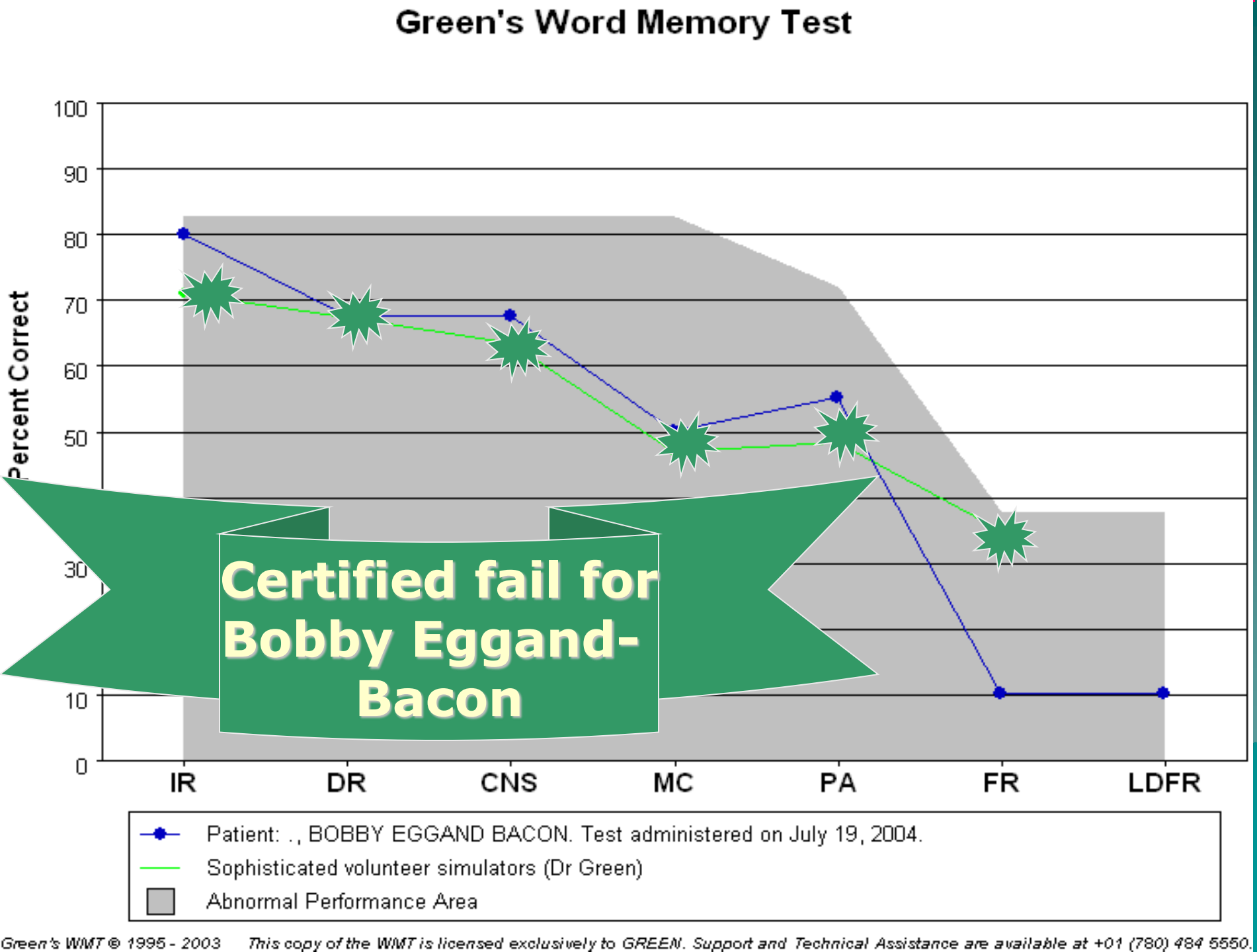
(false names are used in all cases to protect identity

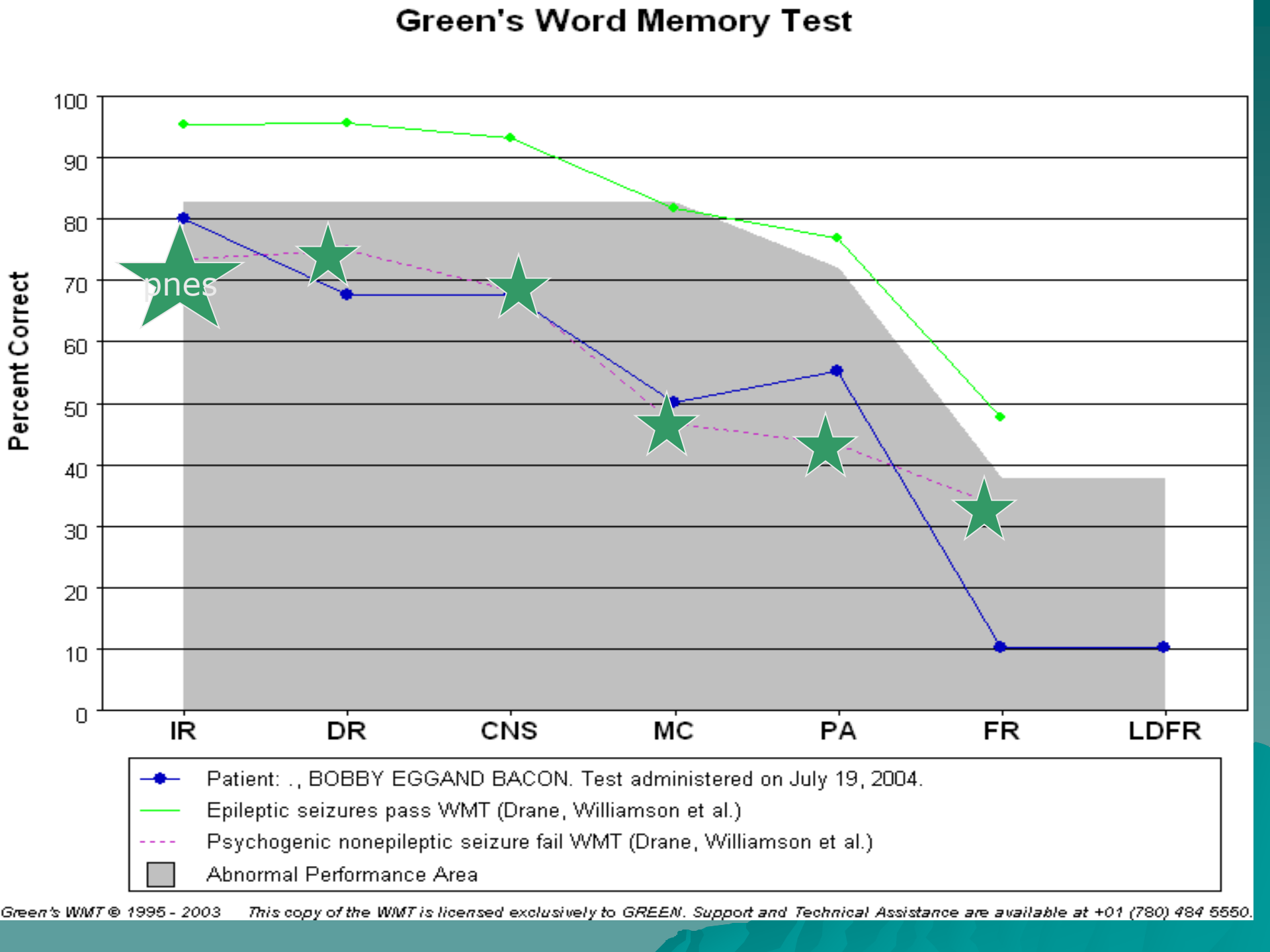
Green's Memory Complaints Inventory

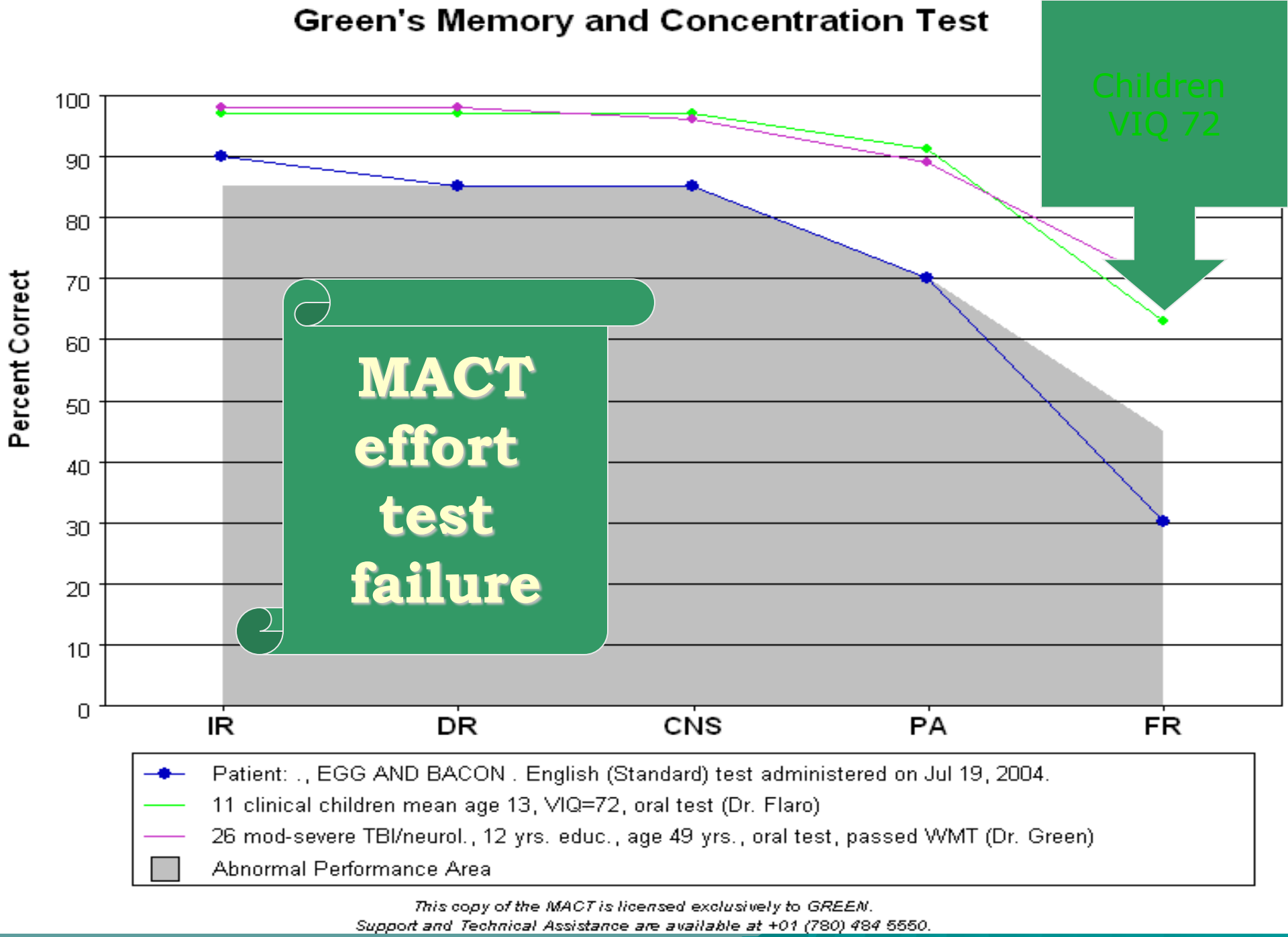


Green's Memory Complaints Inventory



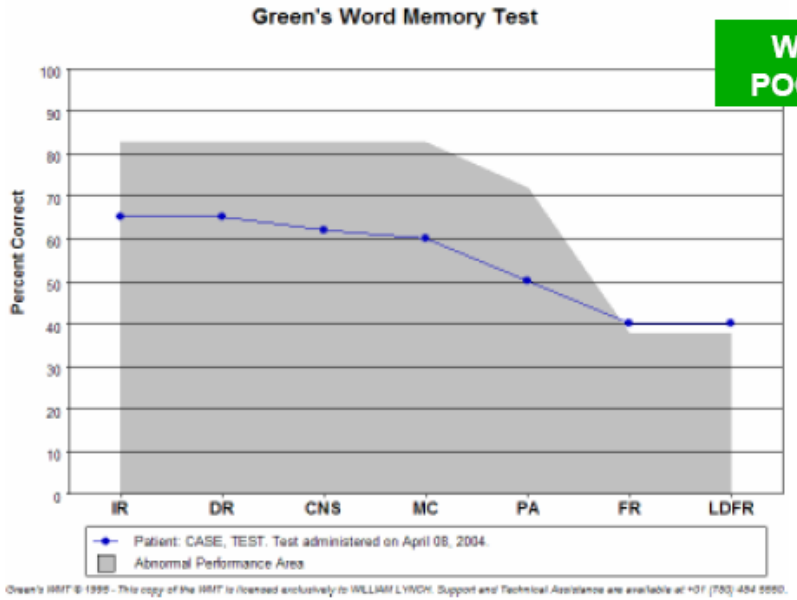






WMT: Poor Effort

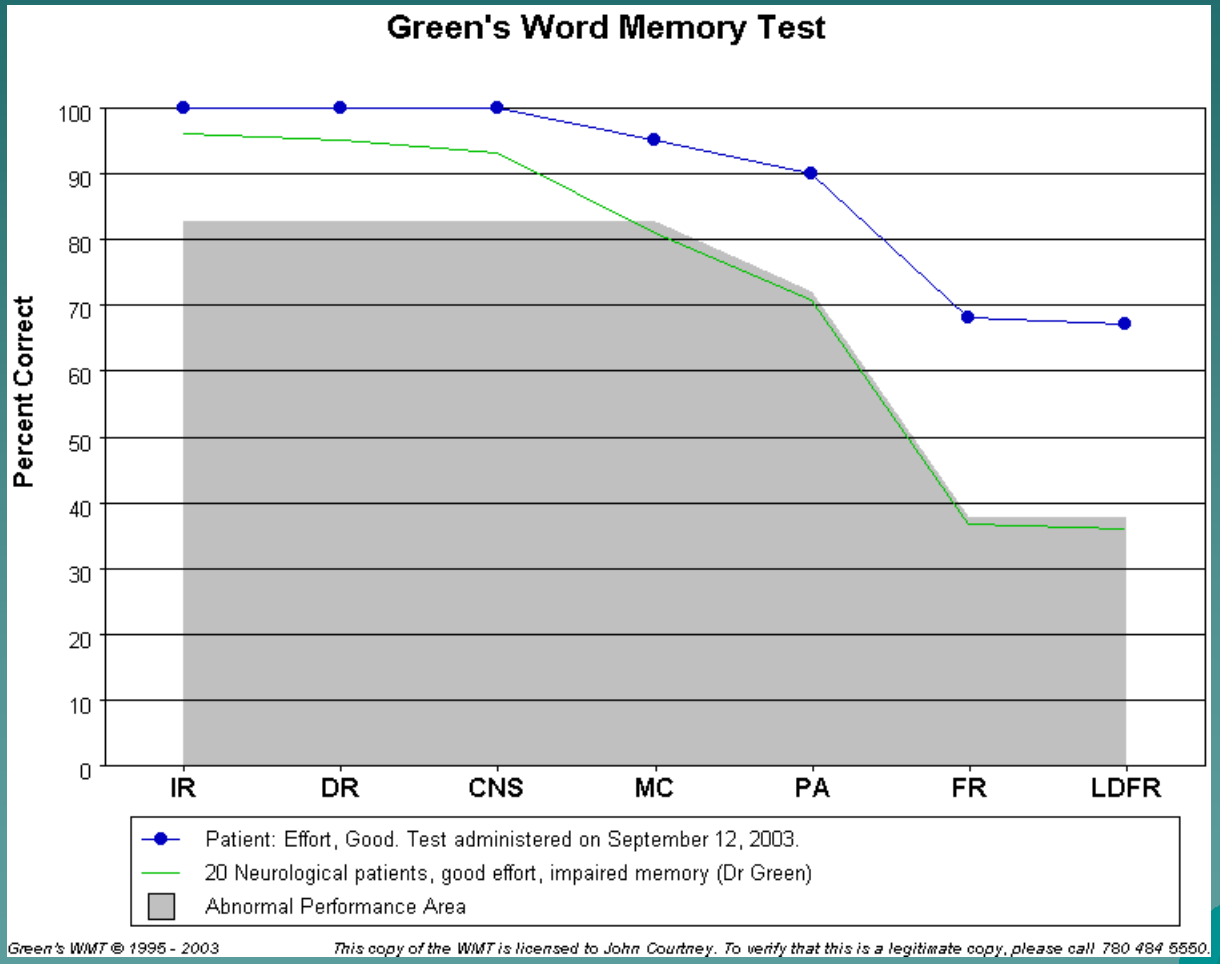
DETERMINING EFFORT LEVEL



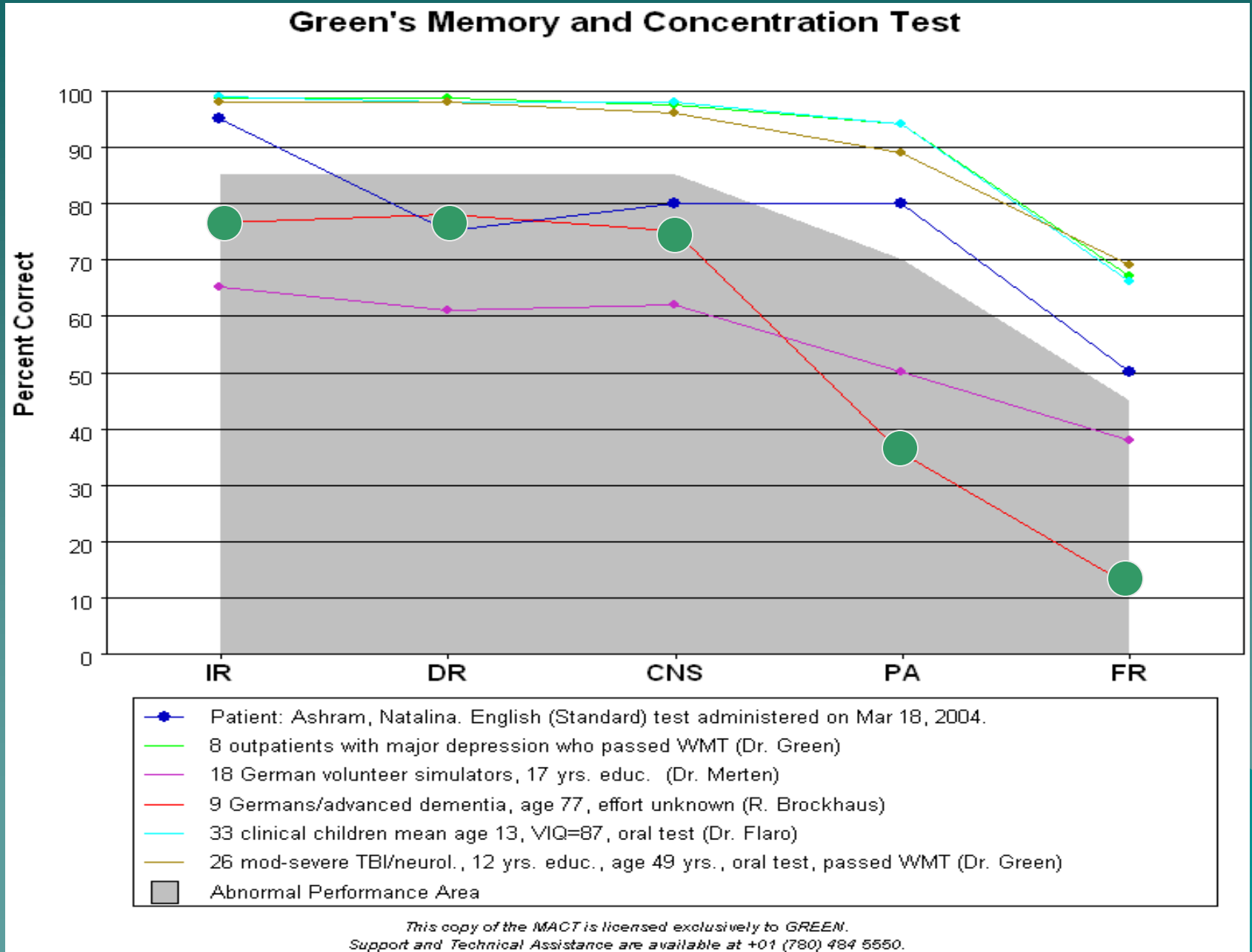
WMT DATA: POOR EFFORT

Dr. Bill Lynch

Good effort, normal & impaired memory



Dementia looks like this



WMT: Dementia, Simulators

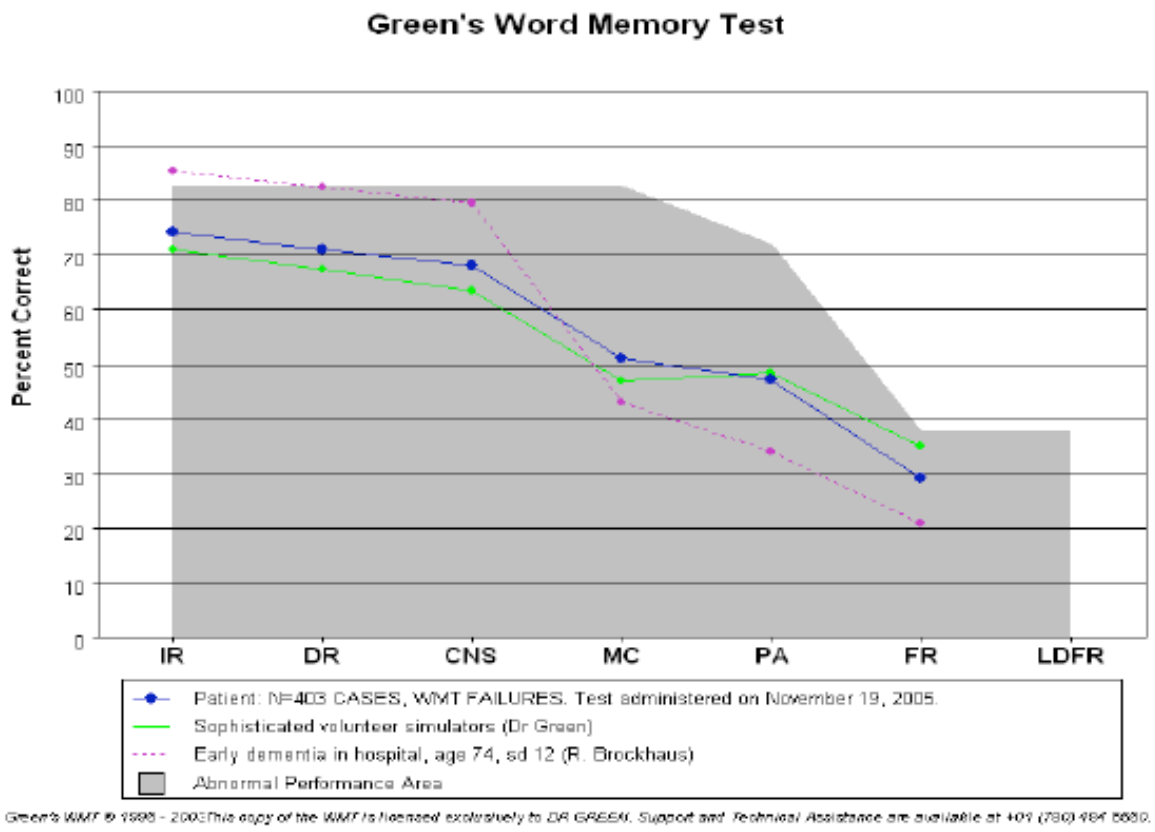
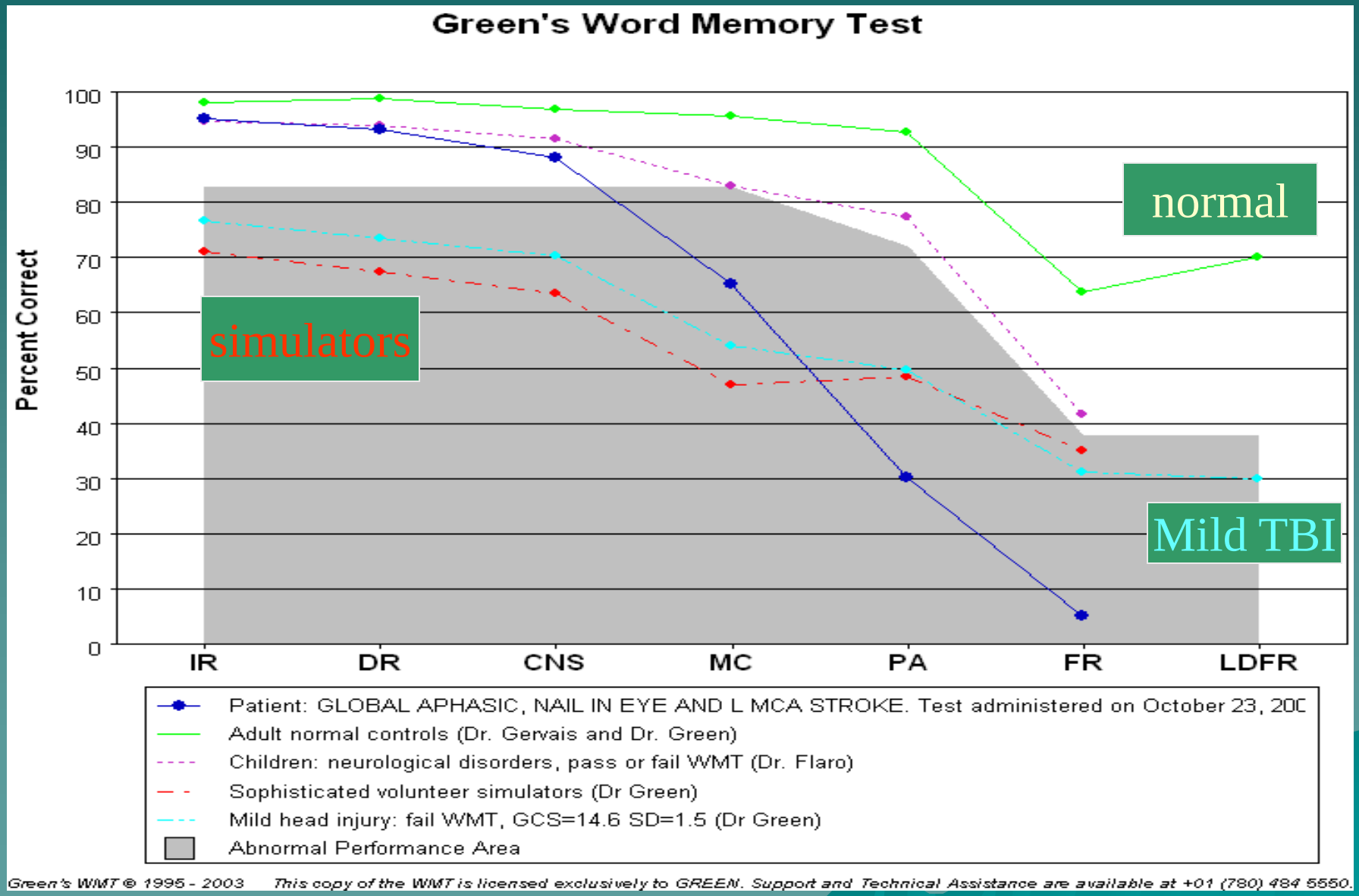
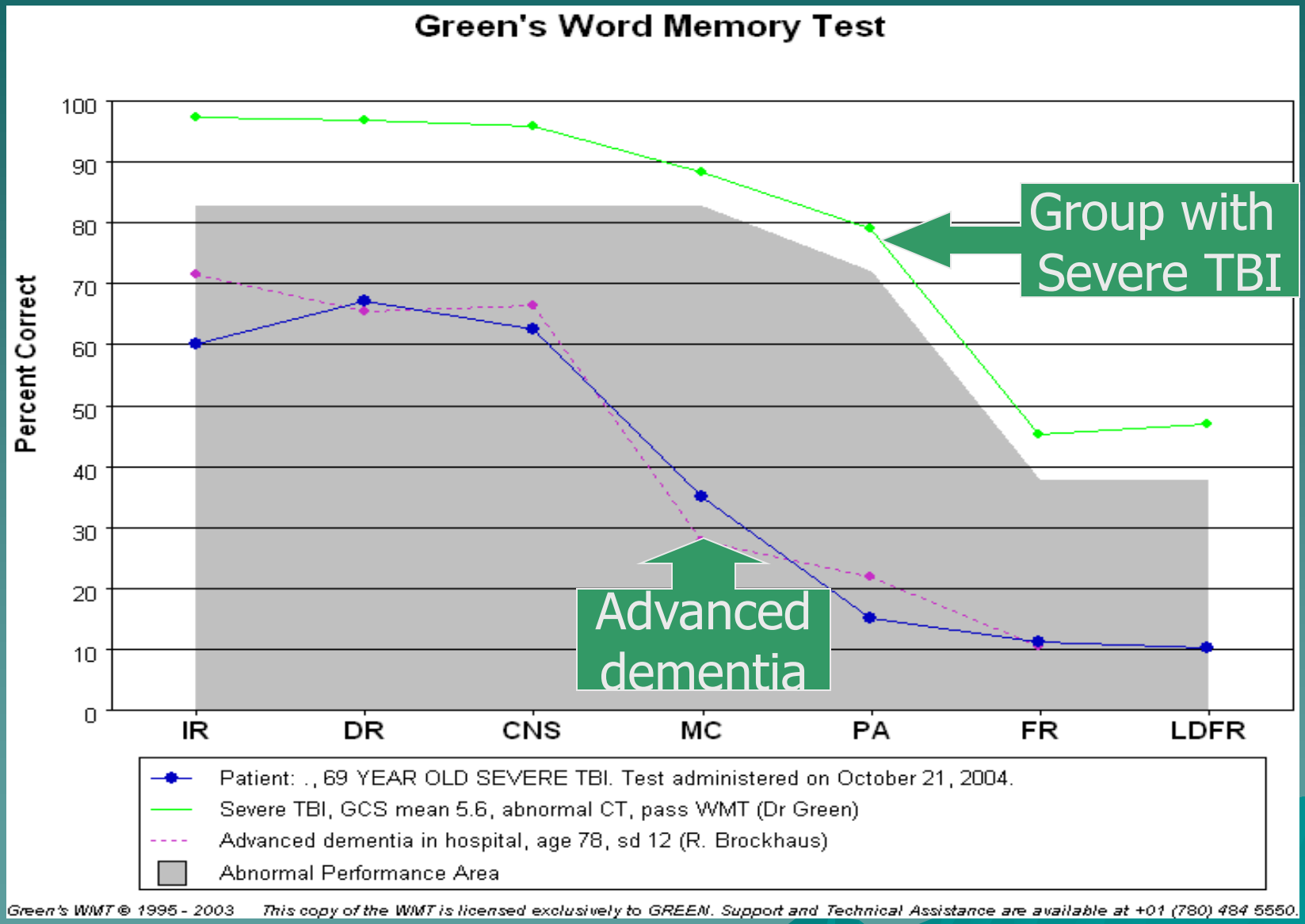


Figure 1. Scores from 403 cases failing WMT when tested clinically resemble those from a group of volunteer simulators: They score lower than dementia patients on easy subtests and higher on harder subtests.

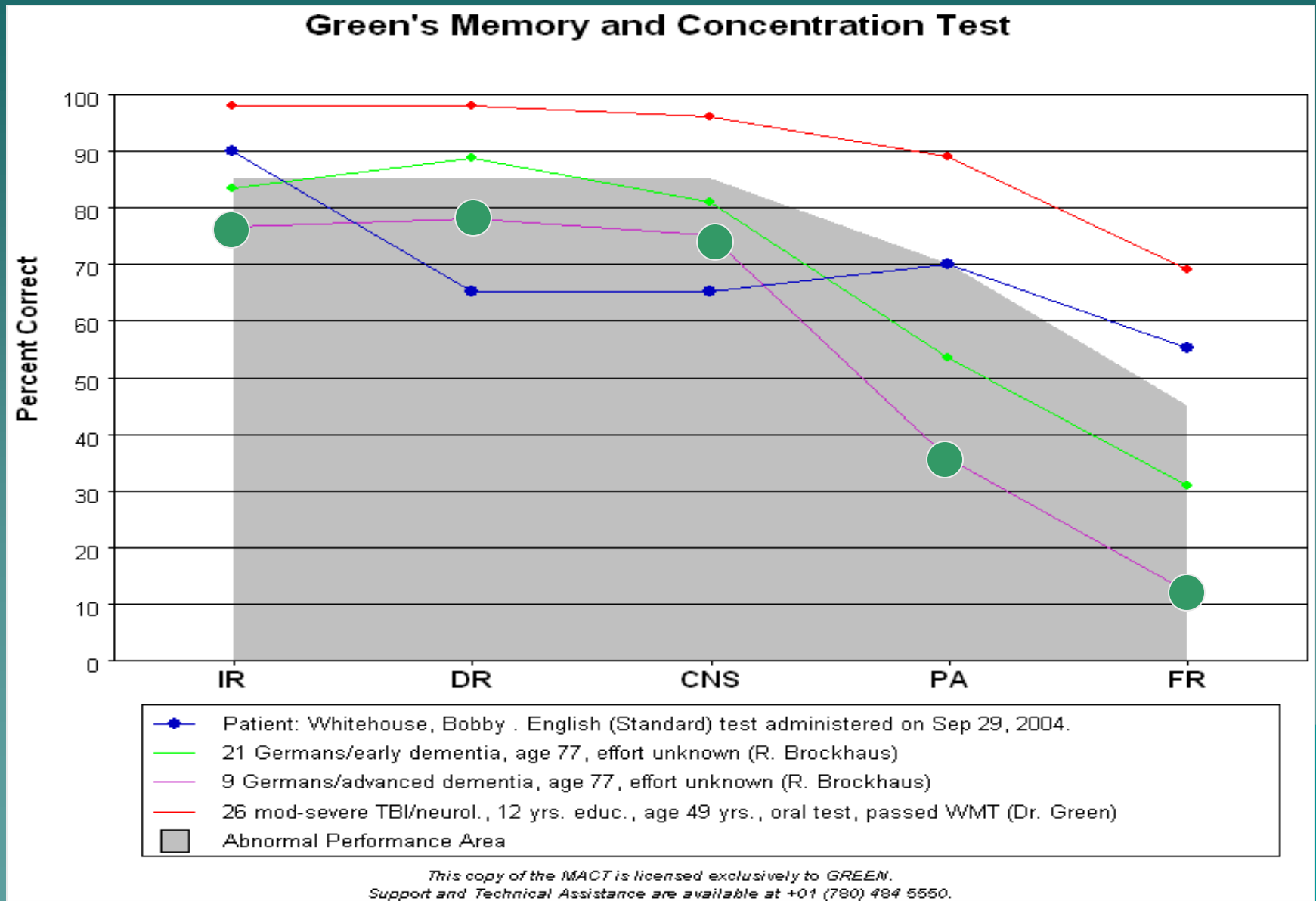
How simulators score compared with mild head injuries failing WMT



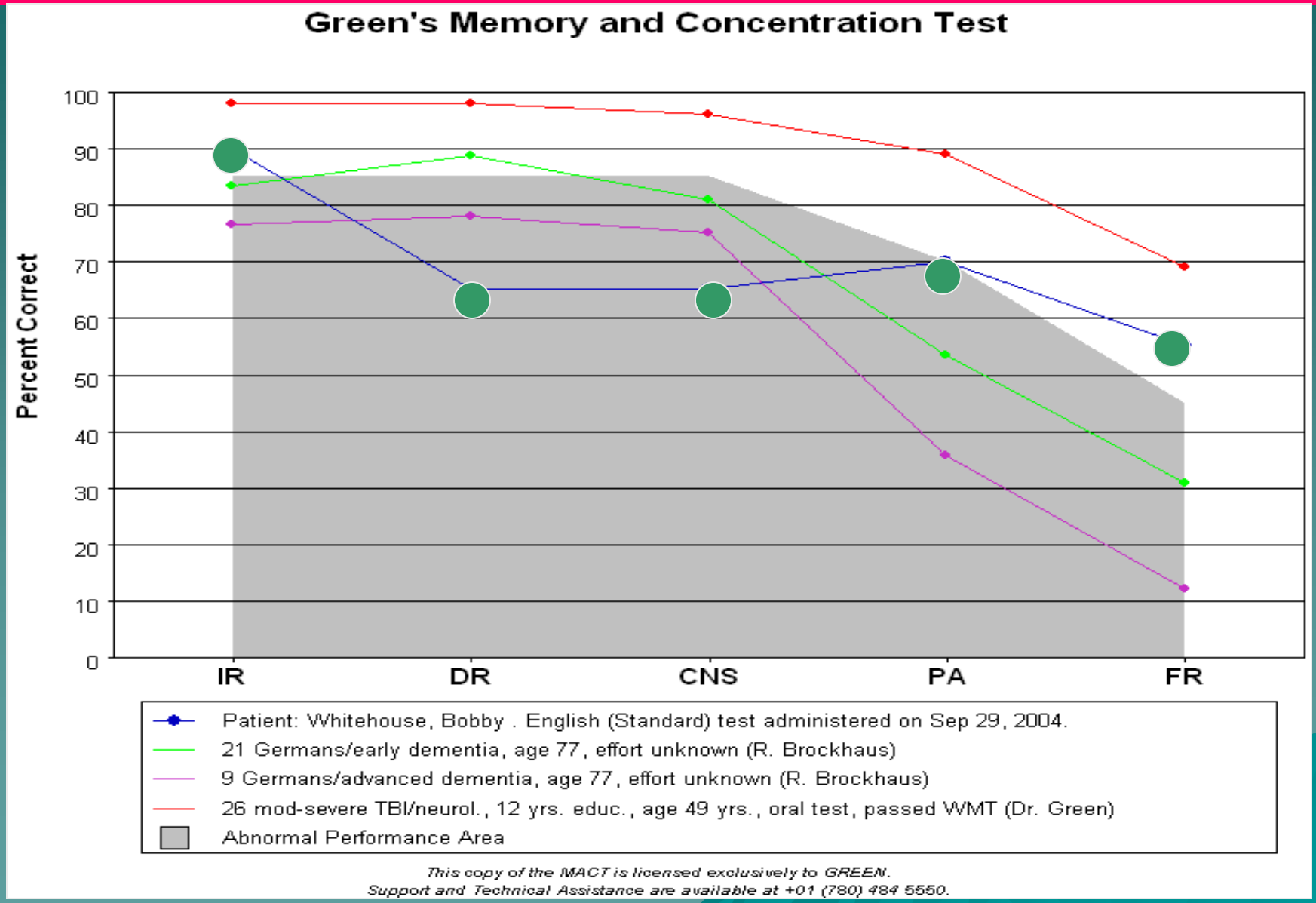
TBI and Dementia



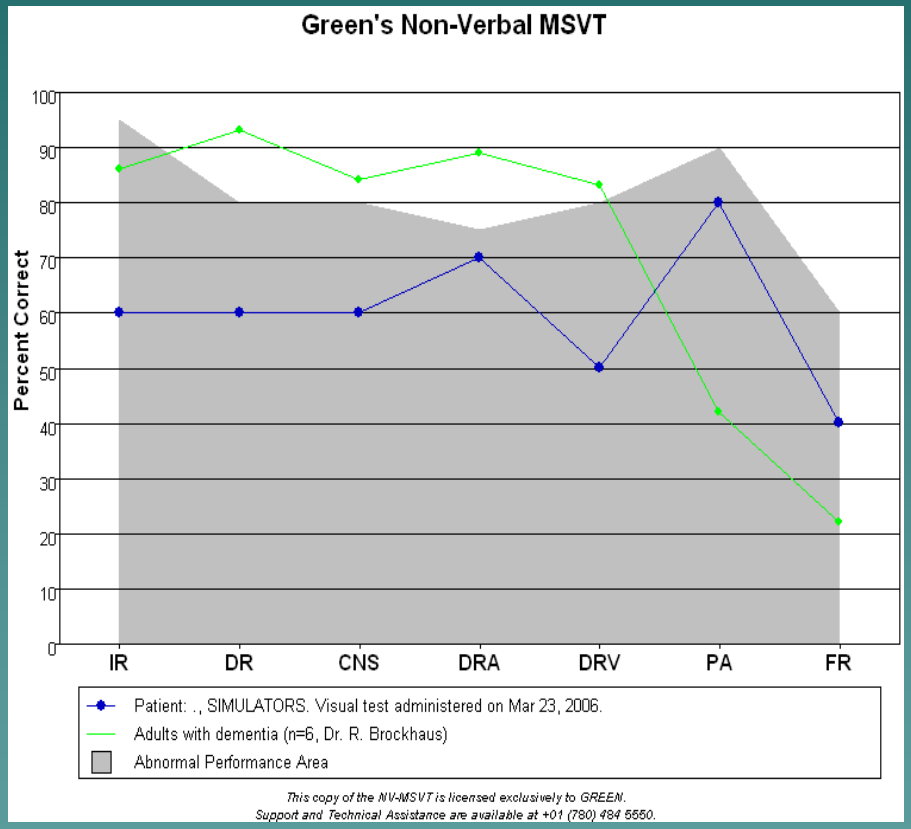
In advanced dementia, gradient reflects objective difficulty of subtests (e.g. IR=DR and DR >PA>FR)



In poor effort, gradient does not reflect objective difficulty of subtests



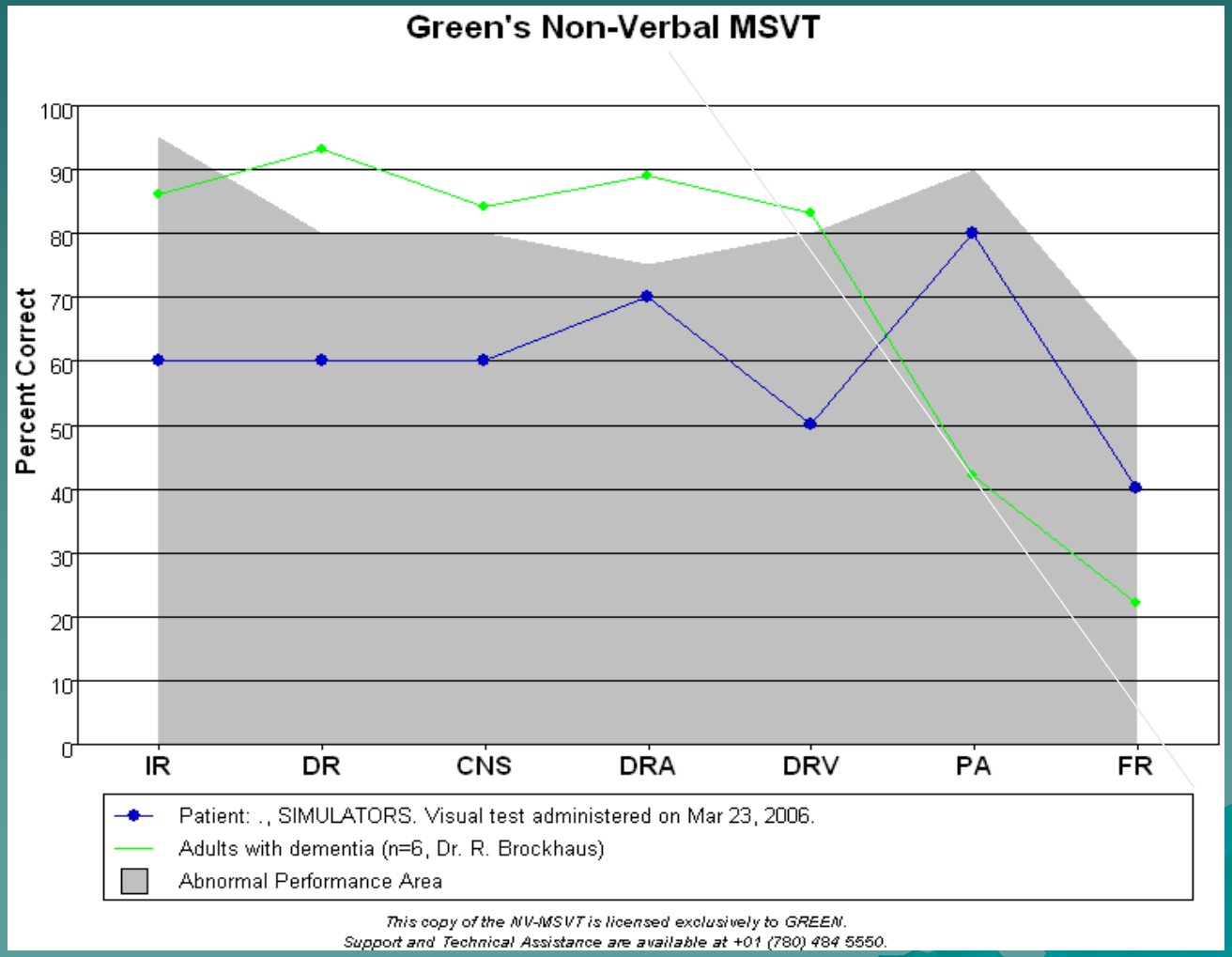
The PA peak in MSVT typical of known simulators (Pinocchio profile).



NV-MSVT in TBI

- On the NV-MSVT, the failure rate in mild TBI cases was 31% but in severe TBI cases it was 0%.
- Is the evidence still out on this fact or does it indicate poor effort in the mild TBI cases?
- Incidentally the pattern in the mild TBI cases failing the test is similar to that which we see in simulators (NV-MSVT test manual, 2008) but not in dementia.

Note down-slope in dementia in MSVT



Mean overall score on the Memory Complaints Inventory by diagnostic group

Condition	N	Mean as % of max possible score	Std. Deviation
Police recruits	20	3.0	3.1
Orthopedic injury	49	19.7	16.9
Neurological disease	137	24.0	18.3
Other	155	22.8	17.4
Probable Early Dementia	14	24.0	13.5
Depression	128	29.4	16.3
Anxiety	30	22.8	15.3
Bipolar	17	29.5	18.7
Psychotic	12	27.5	21.8
Mod-Severe TBI	130	**22.5**	16.6
Mild TBI	276	**27.7**	16.9
Chronic Pain/ Fibromyalgia	47	**32.1**	21.3
CFS	18	**39.4**	**1**3.2

Note that the two groups with the highest overall memory complaints are those with Chronic Fatigue Syndrome and Chronic Pain syndrome, including fibromyalgia.

Pervasive Influence of Effort on NP tests

International Journal of Forensic Psychology
Copyright 2006

Volume 1, No. 3 SEPTEMBER 2006 pp. 1-21

The Pervasive Influence of Effort on Neuropsychological Tests

Paul Green[+]

Neurobehavioural Associates, Edmonton, Canada

Effort and NP tests

- 1307 consecutive outpatients (WC, Insurance cases)
- 31% failed WMT
- As effort decreases, scores on neuropsychological tests decrease significantly and systematically

Correlations

- 4 correlations:
- WMT Pass/Fail with
  - Category test errors (.28)
  - Trails B (.36)
  - Smell R (.09)
  - CVLT Free Recall (-.45)

WMT and CVLT Recall

Mean California Verbal Learning Test Recall Scores at Each Level of Effort

Mean WMT effort	N	CVLT Short Delay Free Recall Mean	CVLT SD Free SD	CVLT Long Delay Free Recall Mean	CVLT LD Free SD
91-100%	745	10.7	3.2	11.2	3.2
81-90%	206	8.3	3.2	8.9	3.2
71-80%	105	7.8	3.5	8.2	3.4
61-70%	61	7.4	3.0	7.3	3.3
51-60%	50	5.8	2.9	5.5	3.1
<=50%	34	4.4	2.5	3.3	3.0

CVLT Cued Recall and Recognition

Table 2

CVLT Cued Recall Scores by Level of WMT Effort

Mean WMT effort	N	CVLT SD Cued Recall Mean	CVLT SD Cued Recall SD	CVLT LD Cued Recall Mean	CVLT LD Cued Recall SD
91-100%	745	11.8	2.8	11.9	2.9
81-90%	206	9.7	3.0	9.8	3.0
71-80%	105	9.5	3.4	9.2	3.2
61-70%	61	8.9	3.0	8.5	3.2
51-60%	50	7.0	3.0	6.5	3.2
<=50%	34	5.4	3.1	4.5	3.1

SD=short delay, LD=long delay

Table 3

CVLT Learning Trial Scores and Recognition Hits by Level of WMT Effort

Mean WMT effort	N	CVLT Trial 1 Mean	CVLT Trial 1 SD	CVLT Trial 5 Mean	CVLT Trial 5 SD	CVLT Rec.Hits Mean	CVLT Rec.Hits SD
91-100%	745	7.4	2.3	12.4	2.5	14.6	1.6
81-90%	206	6.4	2.1	10.6	2.7	13.5	2.2
71-80%	105	6.0	2.2	10.2	3.1	12.8	2.9
61-70%	61	6.0	2.3	9.7	2.6	11.9	2.8
51-60%	50	4.9	1.6	8.7	2.8	10.7	3.2
<=50%	34	4.5	2.1	7.0	2.9	7.6	3.7

CVLT = California Verbal Learning Test, version 1; Rec.Hits = Recognition Hits score

WMT and Warrington

Table 6

Warrington's Recognition Memory Test (RMT) for Words and Faces by Level of WMT Effort

Mean WMT effort	N	WRMT Faces Mean	WRMT Faces SD	N	WRMT Words Mean	WRMT Words SD
91-100%	645	41.8	5.1	301	46.0	4.3
81-90%	176	39.8	5.8	76	41.7	5.7
71-80%	85	37.1	6.6	35	39.8	6.7
61-70%	51	36.2	6.5	25	36.1	7.3
51-60%	37	32.3	8.1	13	29.1	6.9
<=50%	31	26.1	8.8	15	23.0	5.4

WMT & Delayed Story Recall

Table 7

Immediate and 30-minute Delayed Story Recall by Level of WMT Effort

Mean WMT effort	N	Immediate Story Recall Mean	Immediate Story Recall SD	N	Delayed Recall Mean	Delayed Recall SD
91-100%	773	47.1	9.5	766	36.3	12.2
81-90%	207	42.3	9.4	203	27.0	11.7
71-80%	105	40.9	10.3	104	26.9	12.3
61-70%	58	38.6	8.9	57	23.5	9.9
51-60%	45	34.8	11.6	45	17.8	10.7
<=50%	31	28.5	10.6	31	13.8	8.8

Note. Scores are out of a maximum of 80 for the five-story set on immediate and delayed recall, where 50 (SD 7) is the normal mean for immediate recall in adults of average IQ.

Rey Complex Figure and WMT

Rey Complex Figure Test (Meyer method) by Level of WMT Effort

Mean WMT effort	N	Rey CFT Immediate Recall (%ile) Mean	Rey CFT Immediate Recall SD	Rey CFT Delayed Recall (%ile) Mean	Rey CFT Delayed Recall SD	Rey CFT Copy (Raw) Mean	Rey CFT Copy SD
91-100%	813	37.7	30.6	35.9	31.1	33.1	3.0
81-90%	218	26.4	28.2	23.3	26.4	32.1	3.4
71-80%	119	22.6	23.8	20.5	25.2	31.4	4.7
61-70%	63	14.3	18.5	12.4	16.6	30.5	4.1
51-60%	55	10.9	14.9	10.4	15.0	27.5	6.5
41.1%	39	9.5	16.7	5.4	9.4	27.0	5.8

Note. The Rey CFT score is the raw score for the copy trial. Otherwise, the Rey CFT scores and the Digit Span scores are expressed as a percentile rank relative to age, gender and education.

WMT and TMT

The Trail Making Test by Levels of Effort on WMT

Mean WMT effort	N	Trail Making A (seconds) Mean	Trail Making A SD	Trail Making B (seconds) Mean	Trail Making B SD
91-100%	813	30.4	12.6	76.5	48.1
81-90%	218	38.6	17.3	105.0	77.0
71-80%	119	42.1	20.2	112.7	78.5
61-70%	63	44.9	24.9	139.7	96.9
51-60%	55	60.3	46.1	183.1	157.5
<=50%	39	65.2	45.1	159.6	141.1

WMT and WAIS

PIQ and VIQ by Levels of Effort on WMT

Mean WMT effort	N	PIQ Mean	PIQ SD	VIQ Mean	VIQ SD
91-100%	761	104.7	12.9	101.7	13.2
81-90%	199	98.3	14.1	94.8	13.1
71-80%	98	94.2	12.9	95.5	13.4
61-70%	56	93.1	13.3	94.6	13.5
51-60%	43	86.3	13.0	87.7	13.2
<=50%	24	84.0	16.3	86.0	15.8

WMT and Finger Tip Number Writing

Finger Tip Number Writing by Levels of Effort on WMT

Mean WMT effort	N	FTNW Errors Left Mean	FTNW Errors Left SD	FTNW Errors Right Mean	FTNW Errors Right SD
91-100%	272	1.5	2.3	1.8	2.3
81-90%	65	1.7	2.3	2.1	2.2
71-80%	24	2.4	2.8	3.1	2.8
61-70%	16	2.4	2.5	2.1	2.1
51-60%	11	3.3	3.1	4.1	3.5
<=50%	4	7.2	7.9	8.5	8.3

WMT & Grip and Finger Tapping

Grip Strength by Level of WMT Effort

Mean WMT effort	N	Grip strength Right (Kgs) Mean	Grip strength Right SD	Grip strength Left (Kgs) Mean	Grip strength Left SD
91-100%	813	41.4	14.2	38.1	13.6
81-90%	218	39.4	13.6	38.0	13.3
71-80%	119	36.9	16.4	35.3	14.8
61-70%	63	38.0	16.6	36.7	16.0
51-60%	55	33.6	14.6	32.5	14.2
<=50%	39	31.7	14.2	31.6	12.5

Table 13

Finger Tapping Speed by Level of WMT Effort

Mean WMT effort	N	Finger tapping speed Right Mean	FT Right SD	Finger tapping speed Left Mean	FT Left SD
91-100%	529	48.8	8.9	45.3	8.6
81-90%	134	45.3	11.2	42.7	9.3
71-80%	63	43.3	11.1	40.7	9.2
61-70%	32	44.9	10.6	42.8	10.0
51-60%	26	38.0	11.9	37.4	12.0
<=50%	15	34.7	15.4	36.0	13.8

WMT and Grooved Pegboard

Grooved Pegboard by Level of WMT Effort

Mean WMT Effort	N	Grooved Pegboard Right (secs) Mean	Grooved Pegboard Right SD	Grooved Pegboard Left (secs) Mean	Grooved Pegboard Left SD
91-100%	813	72.2	24.4	81.3	32.9
81-90%	218	80.3	22.2	86.2	24.3
71-80%	119	92.6	51.2	90.5	29.6
61-70%	63	82.8	18.9	90.5	18.5
51-60%	55	109.6	55.4	116.2	61.4
<=50%	39	108.4	55.1	123.6	96.3

WMT and WCST

Wisconsin Card Sorting Test and Scores by Levels of Effort on WMT

Mean WMT effort	N	WCST Perseverative Errors Mean (% ile rank)	WCST Categories Mean	WCST Categories SD
91-100%	813	51.3	5.2	1.5
81-90%	218	42.1	4.5	1.9
71-80%	119	38.7	4.4	1.9
61-70%	63	33.4	4.2	2.0
51-60%	55	26.9	3.3	2.3
<=50%	39	11.2	2.1	2.3

Note. The WCST perseverative error score is expressed as a percentile rank relative to age, gender and education, using Heaton's norm tables.

WMT and Category Test Errors

Category Test Errors by Levels of Effort on WMT

Mean WMT effort	N	Category Test Errors Mean	Cat Test Errors SD
91-100%	674	55.1	29.5
81-90%	164	66.2	28.9
71-80%	77	73.3	28.2
61-70%	44	70.7	28.2
51-60%	30	91.1	28.4
<=50%	18	88.4	33.8

WMT and Digit Span

Digit Span and Visual Memory Span by Level of WMT Effort

Mean WMT Effort	N	Digits Forward (%ile) Mean	Digits Backward (%ile) Mean	N	Visual Memory Span Forward (% ile)	Visual Memory Span Backward (% ile)
91-100%	675	49.3	58.7	630	56.8	65.6
81-90%	178	41.4	51.2	166	46.7	53.1
71-80%	95	27.1	37.6	88	38.8	46.9
61-70%	54	31.7	35.9	49	32.8	41.3
51-60%	42	20.6	33.3	36	20.4	32.1
<=50%	30	11.5	14.2	21	17.7	29.4

Note. Digit Span and Visual Memory Span scores are expressed as a percentile rank relative to age, gender and education, using Heaton's norm tables.

Word Fluency and Figural Fluency

Thurstone Word Fluency and Ruff Figural Fluency Test by Level of WMT Effort

Mean WMT Effort	N	Thurstone Word Fluency Mean	Thurstone Word Fluency SD	N	Ruff FFT Total Designs Mean	Ruff FFT Total Designs SD
91-100%	513	51.1	17.8	611	76.8	22.6
81-90%	105	44.0	16.4	145	65.2	23.6
71-80%	49	38.5	15.5	64	61.7	21.9
61-70%	23	43.3	17.9	35	63.2	24.2
51-60%	19	35.5	18.6	27	61.7	17.7
<=50%	10	37.0	17.2	14	52.6	21.5

Note. Thurstone scores are percentile ranks for age, gender and education using Heaton's norms.

Line Orientation & Visual Memory

Benton Judgment of Line Orientation (BJLO) and Continuous Visual Memory Test (CVMT) by Level of WMT Effort

Mean WMT effort	N	BJLO Mean	BJLO SD	CVMT Delay Mean	CVMT Delay Mean
91-100%	653	24.7	4.2	4.2	1.5
81-90%	169	23.8	4.4	3.6	1.6
71-80%	87	22.7	4.4	4.2	1.5
61-70%	49	21.7	5.8	2.8	1.5
51-60%	43	18.5	5.1	2.0	2.0
<=50%	32	15.3	6.7	1.5	1.3

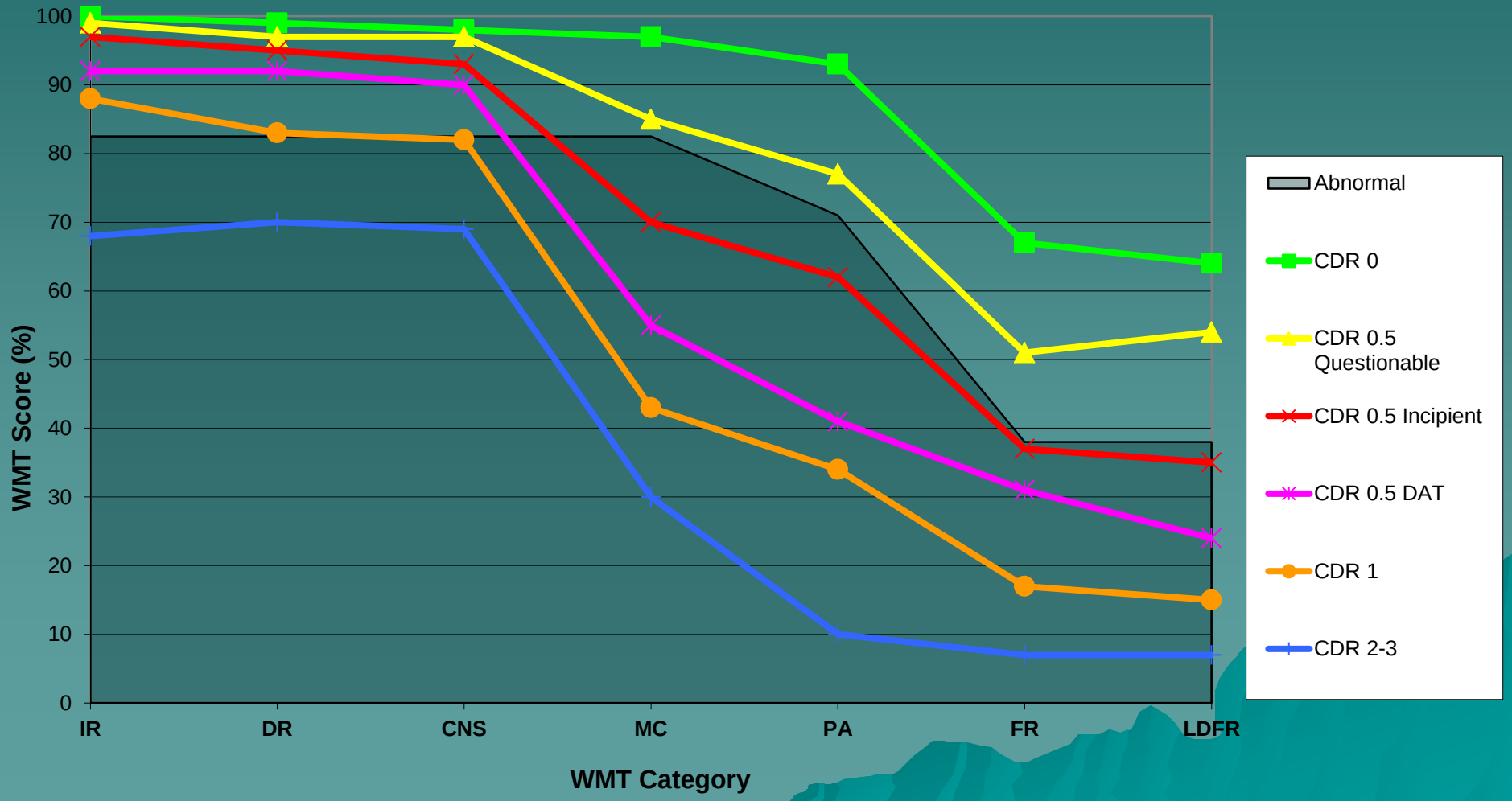
Average Drop in Performance

Average Drop in Performance on Eight Tests for Each Level of Effort on WMT: Scores Expressed in Terms of Standard Deviations Below the Mean for Those in the Top Range of Effort

Mean WMT effort	CVLT SD FREE	WRMT Faces	WCST Categ-ories.	Trail Making A	Ruff FFT Designs	Finger Tap Right	PIQ	Immed. Story Recall	Mean of 8 Tests
91-100%	0	0	0	0	0	0	0	0	0
81-90%	-0.8	-0.4	-0.5	-0.7	-0.5	-0.4	-0.5	-0.5	-0.54
71-80%	-0.9	-0.9	-0.5	-0.9	-0.7	-0.6	-0.8	-0.7	-0.75
61-70%	-1.0	-1.1	-0.7	-1.2	-0.6	-0.4	-0.9	-0.9	-0.85
51-60%	-1.5	-1.9	-1.3	-2.4	-0.7	-1.2	-1.4	-1.3	-1.46
<=50%	-2.0	-3.1	-2.1	-2.8	-1.1	-1.6	-1.6	-1.2	-1.94

Note. For Trail Making A, the signs have been reversed to make the table consistent, because longer times imply poorer performance

WMT Scores by CDR (Clinical Dementia Rating Scale)



Effort is a matter of degree: WMT & TOMM

Pattern of effort test failure	N	Mean WMT	Std. Dev.	Mean TOMM Trial 2 out of 50	Std. Dev.	% of group failing CARB
1) Pass both	698	96%	4	50		0%
2) Fail only TOMM	6	93%	4	40		20%
3) Fail only WMT	240	77%	10	49	1	30%
4) Fail both	102	62%	12	35	8	70%

Good effort

Effort is a matter of degree

Pattern of effort test failure	N	Mean WMT	Std. Dev.	Mean TOMM Trial 2 out of 50	Std. Dev.	% of group failing CARB
1) Pass both	698	96%	4	50	1	0%
2) Fail only TOMM	6	93%	4	40	3	20%
3) Fail only WMT	240	77%	10	49		
4) Fail both	102	62%	12	35	8	70%

Poor effort

Effort is a matter of degree

Pattern of effort test failure	N	Mean WMT	Std. Dev.	Mean TOMM Trial 2 out of 50	Std. Dev.	% of group failing CARB
1) Pass both	698	96%	4	50	1	0%
2) Fail only TOMM	6	93%	4	40	3	20%
3) Fail only WMT	240	77%	10	49	1	30%
4) Fail both	102	62%	12	35		

Extremely poor effort

TOMM vs. MSVT

- For every one case who fails TOMM, another two will fail WMT. They are far from equivalent and what applies to TOMM does not necessarily apply to WMT.
- Everyone who fails the *nonverbal* test called TOMM will also fail the easy subtests of the *nonverbal* MSVT but there are as many more who pass TOMM and fail the nonverbal MSVT.
- In those who pass TOMM and fail the nonverbal MSVT, there is a unique profile that cannot be explained, except by fluctuating effort and unreliable test scores. It involves scoring the same as dementia cases on very easy tests and higher than dementia cases on much harder tasks.

TOMM vs. MSVT

- **Virtually everyone who failed TOMM also failed the NV-MSVT (Green n=15 failing TOMM, Gervais n=32 failing TOMM).**
- **However, there were many people who failed the NV-MSVT and passed the TOMM (Gervais n=37, Green n=21).**
- **Both groups show a clear 'Pinocchio' profile.**
- **They score as low as dementia cases on the very easy NV-MSVT subtests but much higher than dementia on the harder NV-MSVT subtests. This profile indicates unreliable and invalid data.**
- **The data indicate high rates of false negatives for poor effort using the TOMM. The NV-MSVT is more sensitive to poor effort than the TOMM.**

TOMM

- Dominic Carone: the TOMM is obsolete as a primary effort measure, meaning that if you use that test as your only effort measure, you are going to miss a lot of cases of poor effort that more sensitive tests (e.g., WMT, MSVT, NV-MSVT) would pick up.

- TOMM may still have valid supplemental uses in some cases but should not be used as a stand-alone or as a gold standard supplemented only with other insensitive effort tests (e.g., Rey 15).

Objection by Mike Williams: What is effort?

- Calling the recognition memory tests an evaluation of an undefined construct named effort only begs the question of how you are using the term.
- Given the structure of these tests, it appears that the variance on them is caused by inattention.
- That explains why the TOMM loads on a factor that also includes Digit Span and Arithmetic and not Vocabulary.
- What you are calling effort, I call poor sustained attention.
- Poor sustained attention is a malingering strategy that should cause low scores on a number of neuropsych tests and not others.
- Variance on the TOMM, WMT and the PASAT are explained by levels of sustained attention. Variance on tests like Vocabulary are not a product of sustained attention.
- What do the TOMM and WMT measure? - effort. What is effort?

Dominic Carone: Levels of Effort

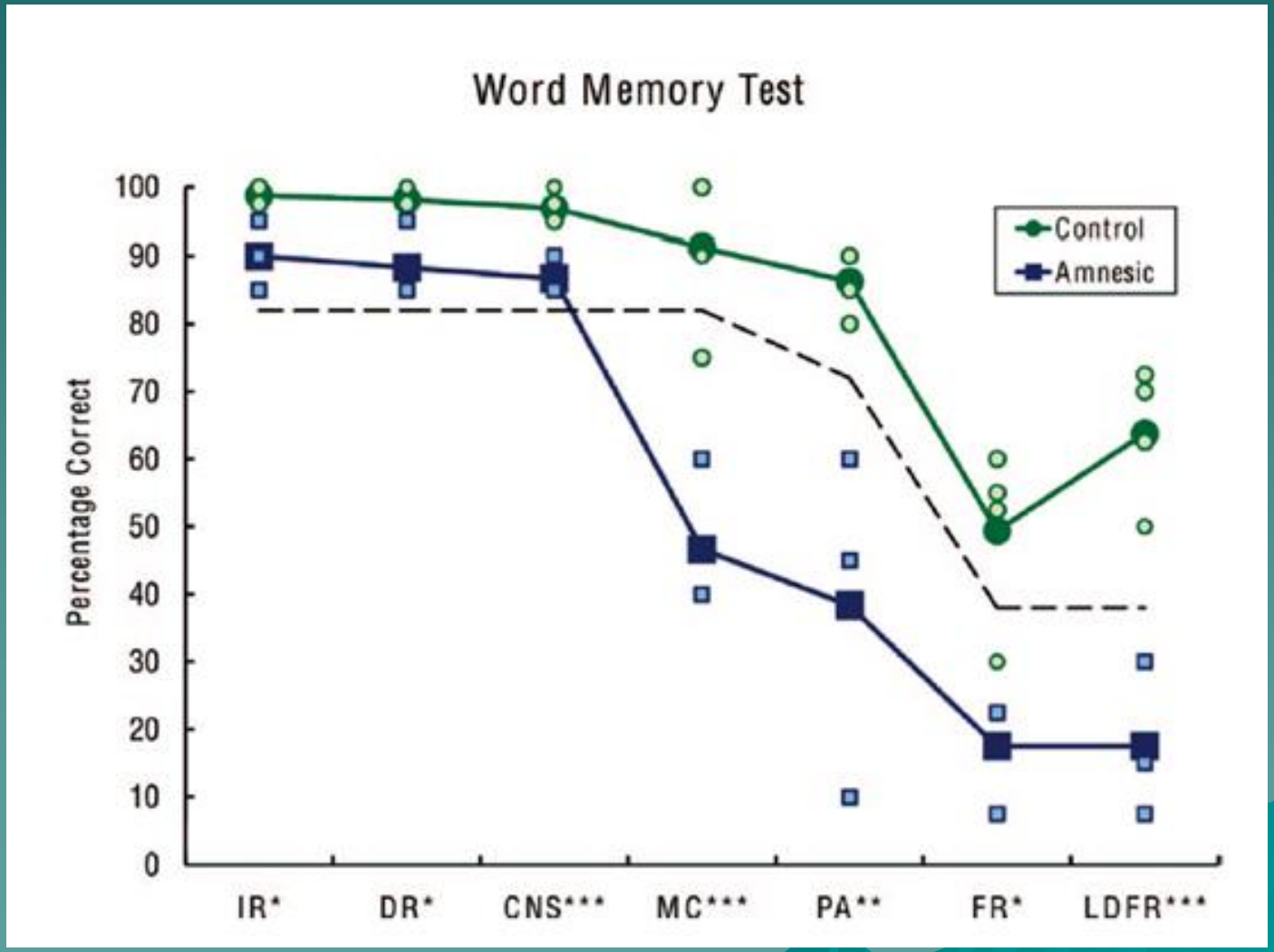
- LEVEL 1: These are the Type A's who try 100% in everything. This is the rare patient who strangely seems to LOVE being tested and wants more testing even when it's over. You practically need to kick them out of the office to stop the evaluation. These people likely have problems.
- LEVEL 2: These are people who try to do well, not always 100%, but fluctuate slightly around this.
- Because it is very difficult to put forth 100% effort 100% of the time, level 2 performance can still be considered capacity performance. We want level 1 or 2 performance on a neuropsych evaluation
- LEVEL 3: People who do not fall in the above categories. They put forth such little effort to do well that they do not perform close to capacity levels. These are the people we are trying to identify because their data will be confounded.

Limitations to exclusive use of Forced Choice Measures

- May be (only or primarily) sensitive to feigned memory deficits
  - Osmon et al. (2006): FC measure underperformed relative to dedicated LD effort test in LD simulation
- May be highly correlated with each other (due to same format) and therefore may not provide non-redundant information
  - Rosenfeld, Sands, and van Gorp (2000)
- Lengthy to give (e.g., 20+ minutes)
  - Rx is to give several within battery – numerous FC measures will substantially lengthen battery
- FC measures are the most "popular" and widely administered
  - Is that a good thing?
- May be easy to coach/educate
  - "Whenever you see a test where you have to pick between two choices, do well on that test" or "Do well on the computer test!"

Browndyke JN, Brain Injury, 2008

People with bilateral hippocampal lesions and severe verbal memory impairment all passed the WMT effort subtests (IR and DR).



FMRI of Effort

- Malingered recognition memory errors were associated with inferior parietal and superior temporal activity relative to normal performance
- Feigned recognition target misses produced additional dorsomedial frontal activation and feigned foil false alarms activated bilateral ventrolateral frontal regions.
- Malingered response times were associated with activity in the dorsomedial frontal, temporal and inferior parietal regions.
- Normal memory responses were associated with greater inferior occipitotemporal and dorsomedial parietal activity, suggesting greater reliance upon visual/attentional networks for proper task performance.

FMRI of Effort

- Task deception requires greater "effort" (PFC engagement) than truth. This is a consistent replication in the functional imaging literature.
- Failing an effort test actually requires greater engagement of dorso- and ventrolateral PFC.

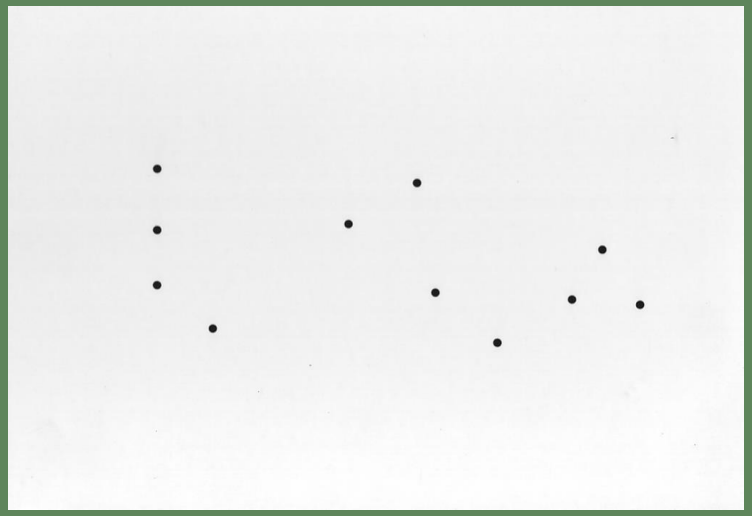
Non-Forced Choice Effort Indices

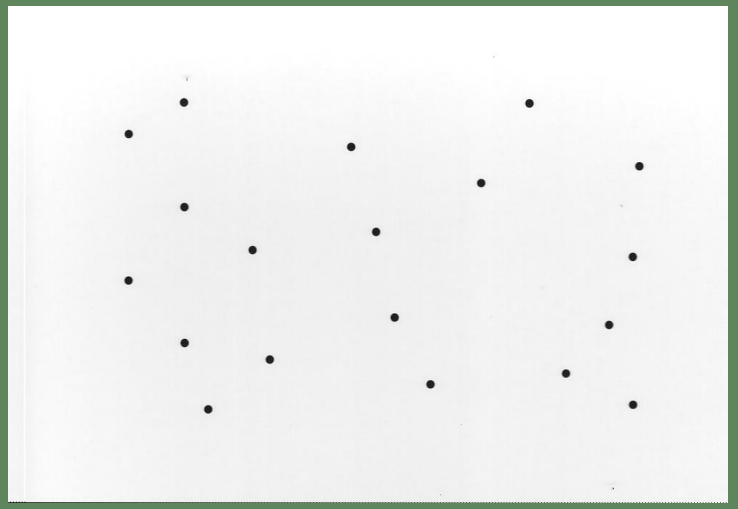
- Brief free-standing effort tests
  - Dot Counting Test
  - b Test
  - Rey 15-item + Recognition Trial
  - Rey Word Recognition Test
    - Total of no more than 30 minutes administration time
  - Harbor-UCLA Medical Center
- Standard cognitive tests already in battery
  - Digit Span
  - Rey-Osterrieth + recognition trial
  - RAVLT recognition equation
  - Rey-Osterrieth/RAVLT discriminant function
  - Finger Tapping
  - CVLT measures

Dot Counting Test

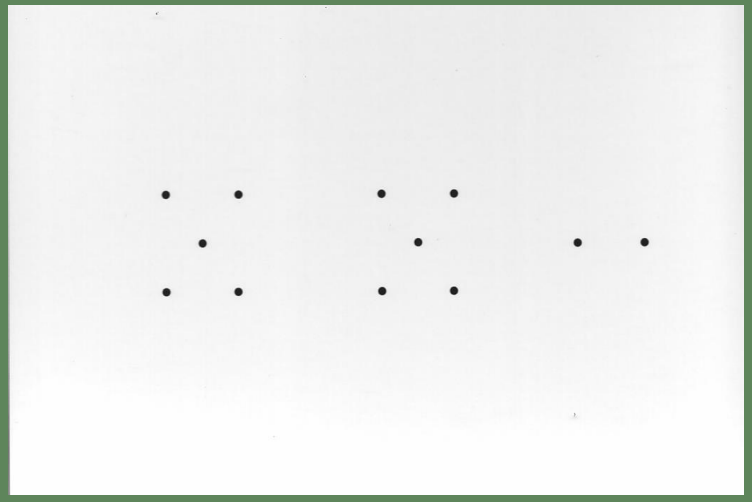
- Developed by Andre Rey over 60 years ago
- Stimuli consist of 12 index cards with varying amounts of dots
  - first 6 cards depict random dots (7–27)
  - last 6 cards contain grouped dots in specific formations (8-28)
- Patients to count dots as quickly as possible; each trial timed

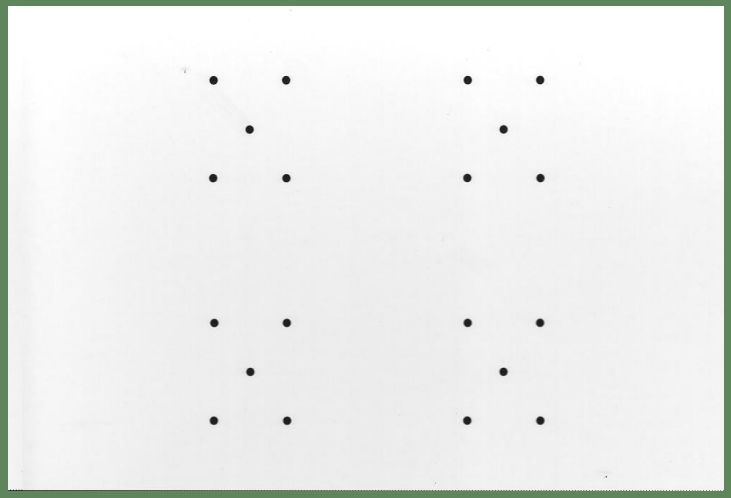
  – Boone, Lu, and Herzberg (2002a)





Ungrouped Dots





Grouped Dots

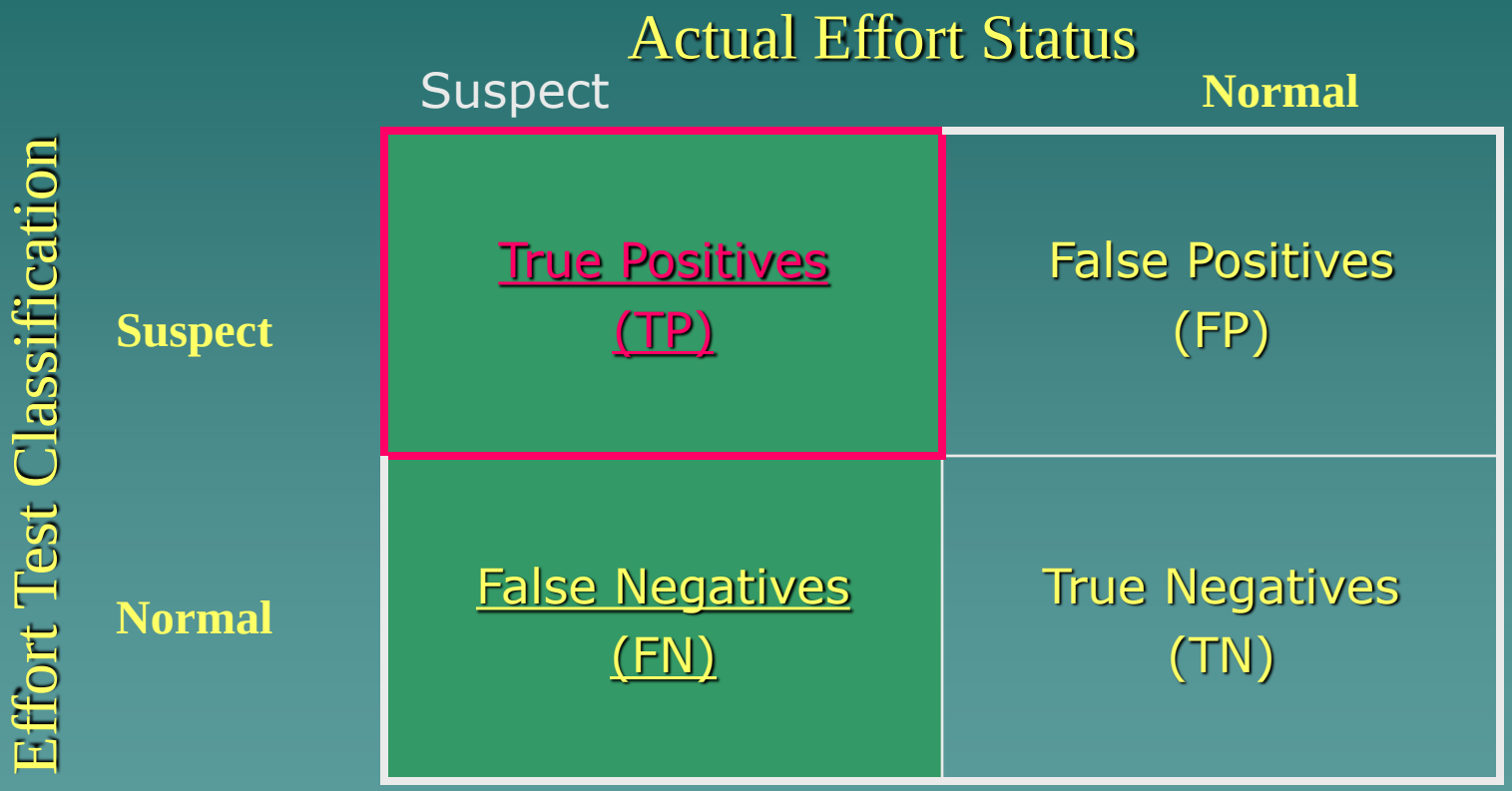
Dot Counting Test

- Sensitive to feigned deficits in:
  - Mental speed
  - Overlearned math skills (simple multiplication)

Dot Counting Test

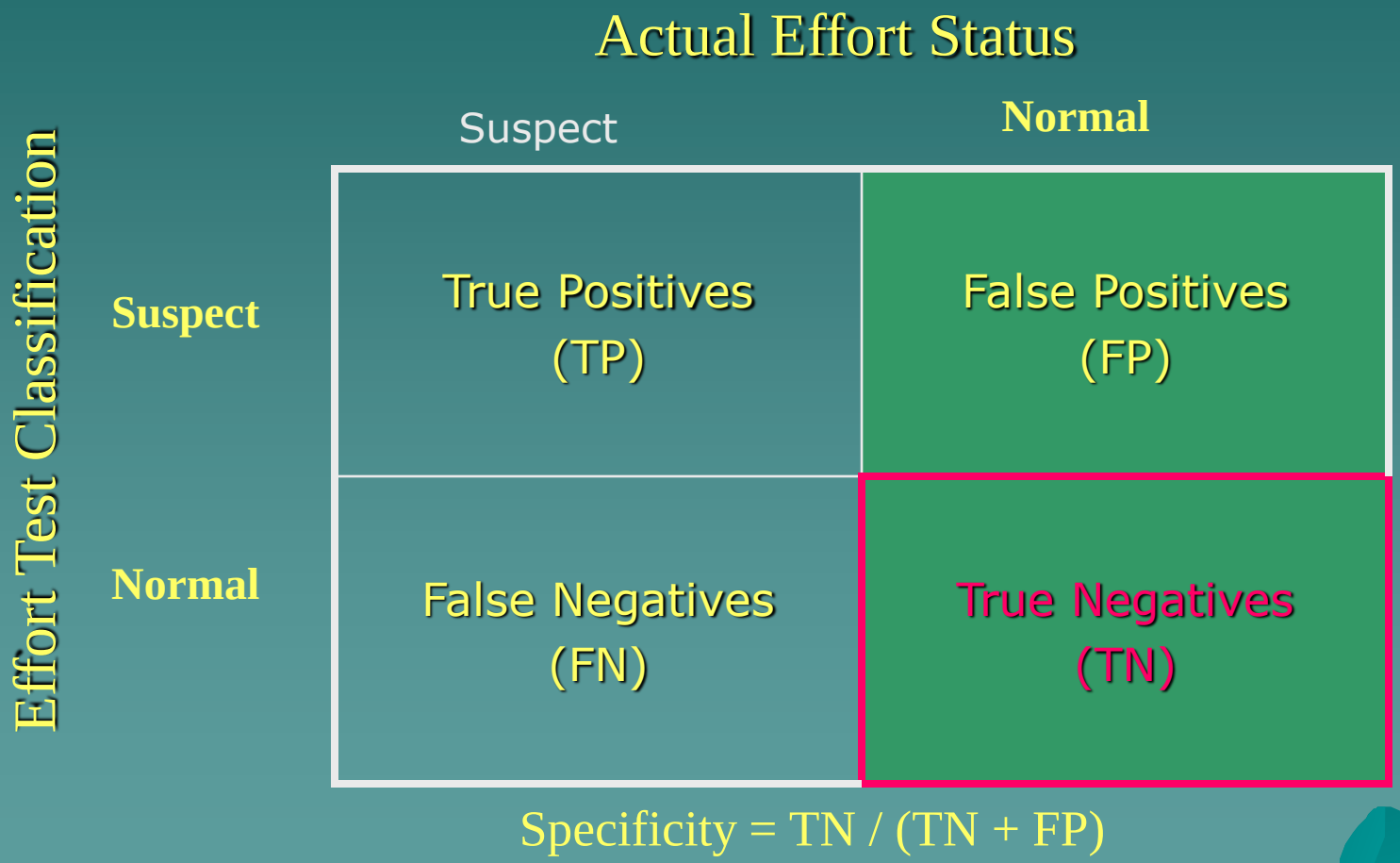
- Most sensitive score (at ≥90% specificity):
  - Mean ungrouped dot counting time + mean grouped dot counting time + number of errors

Sensitivity



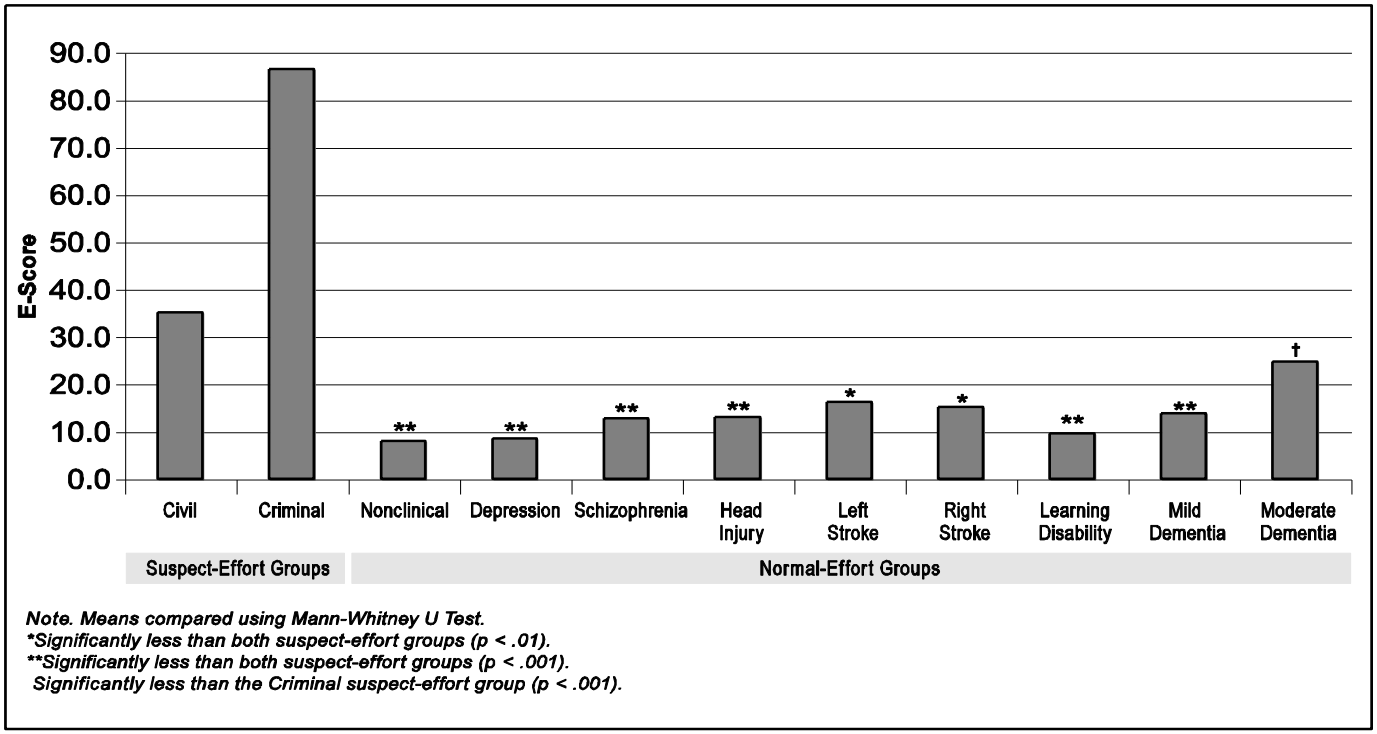
Sensitivity = TP / (TP + FN)

Specificity



Specificity = TN / (TN + FP)

Dot Counting Test Validity Study: Mean E-scores



Note. Means compared using Mann-Whitney U Test.
**Significantly less than both suspect-effort groups ($p < .01$).*
***Significantly less than both suspect-effort groups ($p < .001$).*
† Significantly less than the Criminal suspect-effort group ($p < .001$).

Figure 5
Mean E-Scores in the DCT Validity Study Groups

Dot Counting Test

- "All purpose" cut-off: >17
  - Sensitivity = 78%
  - Specificity = ≥90%
- However, cut-offs can be adjusted for the specific differential diagnosis
  - e.g., actual versus feigned depression

Dot Counting Test: Recommended Cutoff Scores

Table 3
Recommended DCT Cutoffs and Interpretive Ranges

		Interpretive Range				Base Rate Assumption					
						15%[a]		30%[b]		45%[c]	
Normal-Effort Group	E-Score Cutoff	Normal Effort	Suspect Effort	Sensitivity (%)	Specificity (%)	PPA (%)	NPA (%)	PPA (%)	NPA (%)	PPA (%)	NPA (%)
Nonclinical	14	≤13	≥14	88.2	96.1	79.9	97.9	90.6	95.0	94.8	90.9
Depression	14	≤13	≥14	88.2	95.3	76.9	97.9	89.0	95.0	93.9	90.8
Schizophrenia	20	≤19	≥20	68.2	96.4	77.1	94.5	89.1	87.6	94.0	78.8
Head Injury	20	≤19	≥20	68.2	95.0	70.7	94.4	85.4	87.5	91.8	78.5
Stroke	22	≤21	≥22	62.4	88.9	49.8	93.0	70.6	84.6	82.1	74.3
Learning Disability	15	≤14	≥15	85.9	96.8	82.5	97.5	91.9	94.1	95.6	89.3
Mild Dementia	22	≤21	≥22	62.4	93.7	63.8	93.4	81.0	85.3	89.1	75.3
Normal-Effort Groups Combined	19	≤18	≥19	71.8	94.7	70.6	95.0	85.4	88.7	91.8	80.4

Note. See text for further discussion. PPA = positive predictive accuracy, NPA = negative predictive accuracy.
[a]Suggested for use in general clinical settings (i.e., where relatively few patients have external incentives to appear impaired).
[b]Suggested for use in settings where a mixture of compensation-seeking and non-compensation-seeking patients is seen (e.g., assessment practices comprising both medico-legal and clinical referrals).
[c]Suggested for use in settings where most or all patients are expected to have external incentives to appear impaired (e.g., practices specializing primarily in medico-legal neuropsychological assessments).

b Test

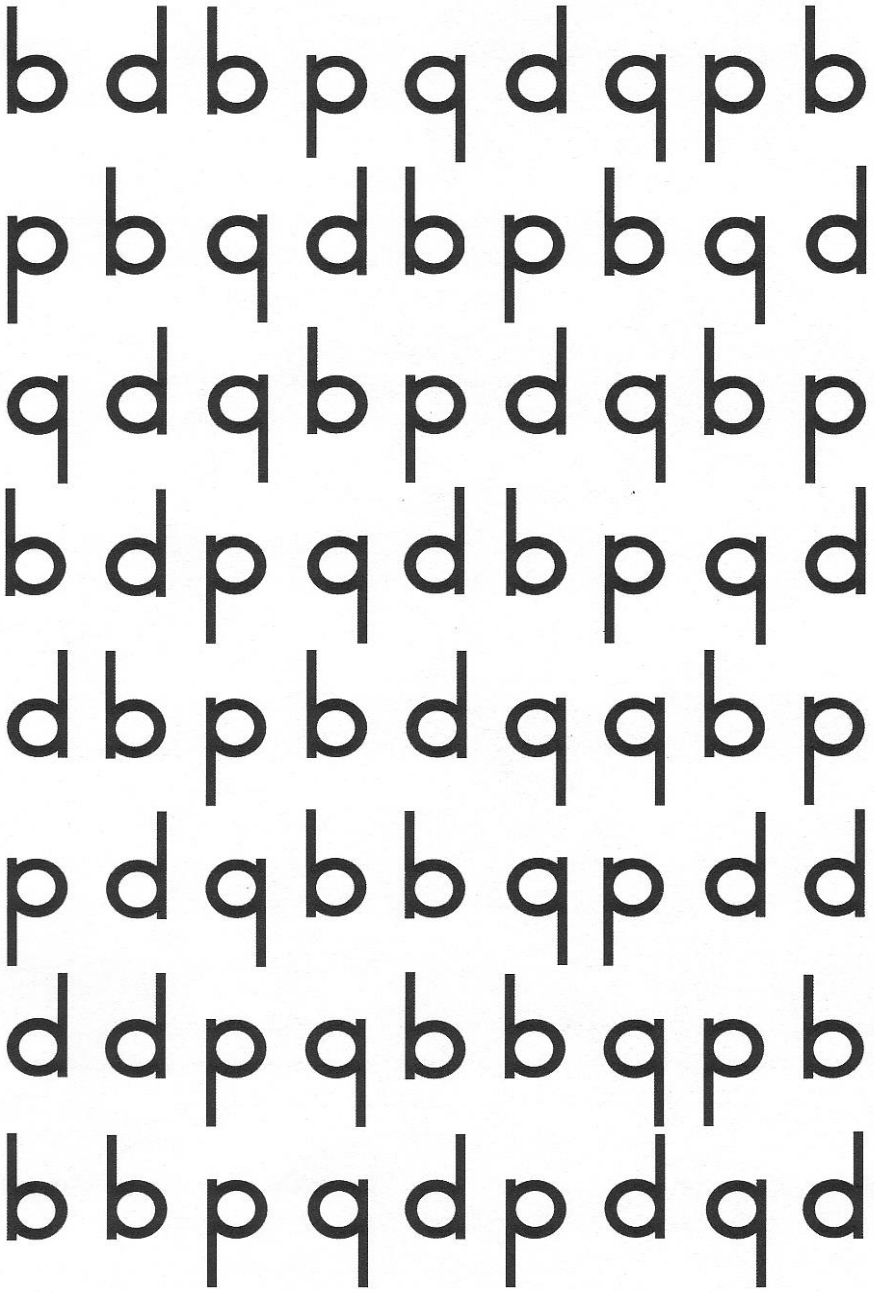
- Originated from the observation that noncredible patients frequently reported that they became dyslexic (i.e., saw letters upside down and backwards) after negligible brain injury
  - However, letter discrimination is a highly overlearned skill resistant to disruption from brain injury
- Consists of 15 pages of "b's" interspersed with "p's", "q's", "d's," and "b's" with diagonal or additional stems
  - the same three pages repeat but become progressively smaller (from letters 7/8 inch to 1/8 inch)
  - Boone et al. (2002b)

b Test

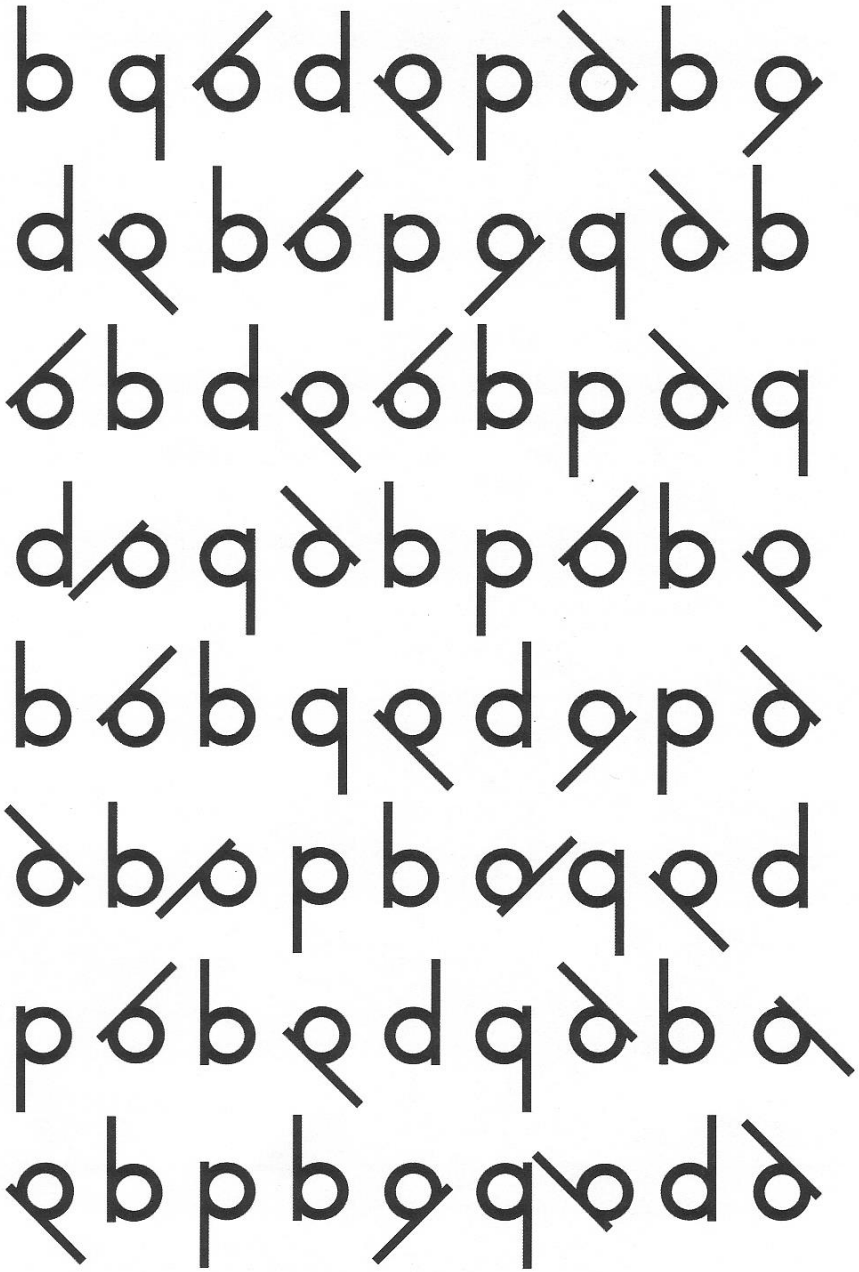
- Patient is to circle all the letter "b's" as quickly as possible; the amount of time spent circling b's is timed

- b Test is sensitive to feigned deficits in:
  - letter discrimination
  - mental speed

Stimulus 1

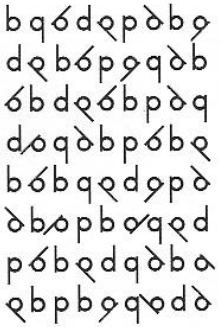

b d b p q d q p b
p b q d b p b q d
q d q b p d q b p
b d p q d b p q d
d b p b d q q b p
p d q b b q p d d
d d p q b b q p b
b b p q d p d q d

Stimulus 2



Stimulus 3

Stimulus 14



b Test

- 3 types of scores:
  - Time
  - Commission Errors
  - Omission Errors

- Most sensitive score (at ≥90% specificity)
  - [(Commission errors + "d" commission errors) x10] + omission errors + mean time per page

b Test Validity Study: Mean E-scores

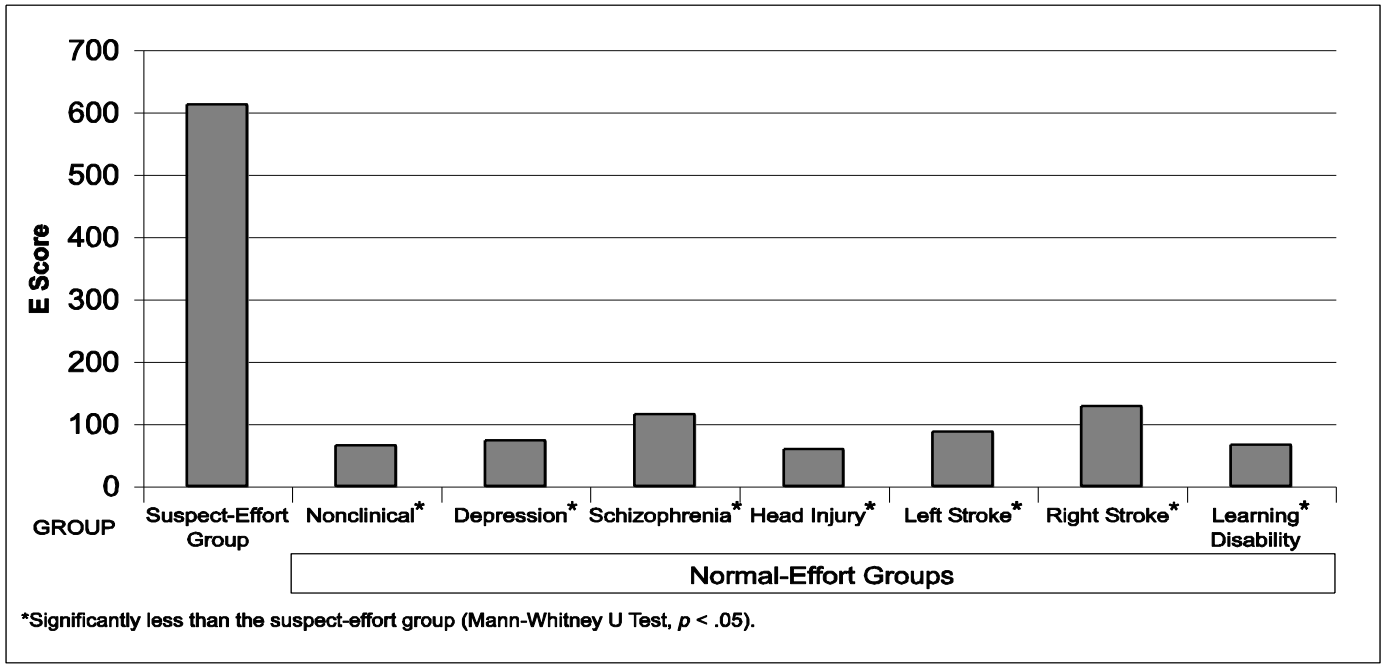


Figure 5
Mean E-Scores in the b Test Validity Study Groups

b Test score ≥ 120

- Sensitivity
  - 74% of "real world" noncredible patients
- Specificity
  - 100% of head injured
    - ≥90 for <10% FP
  - 87% of LD
    - ≥140 for <10% FP
  - 90% of older depressed
  - 61% of stroke
    - ≥170 for <10% FP
  - 75% of schizophrenic
    - ≥190 for <10% FP
  - 96% of elderly normal

b Test: Recommended Cutoff Scores

Table 3
Recommended b Test Cutoffs and Interpretive Ranges

Normal-Effort Group	E-Score Cutoff	Interpretive Range: Normal Effort	Interpretive Range: Suspect Effort	Sensitivity (%)	Specificity (%)	Base Rate Assumption: 15%[a] PPA (%)	15%[a] NPA (%)	30%[b] PPA (%)	30%[b] NPA (%)	45%[c] PPA (%)	45%[c] NPA (%)
Nonclinical	90	≤89	≥90	76.9	88.5	54.1	95.6	74.1	89.9	84.5	82.4
Depression	120	≤119	≥120	73.6	89.5	55.2	95.1	75.0	88.8	85.1	80.6
Schizophrenia	180	≤179	≥180	56.0	85.7	40.9	91.7	62.7	82.0	76.2	70.4
Head Injury	90	≤89	≥90	76.9	90.0	57.6	95.7	76.7	90.1	86.3	82.7
Stroke	170	≤169	≥170	56.0	94.4	64.0	92.4	81.2	83.4	89.2	72.4
Learning Disability	130	≤129	≥130	69.2	87.1	48.6	94.1	69.7	86.9	81.4	77.6
Normal-Effort Groups Combined	120	≤119	≥120	73.6	85.1	46.6	94.8	67.9	88.3	80.2	79.8

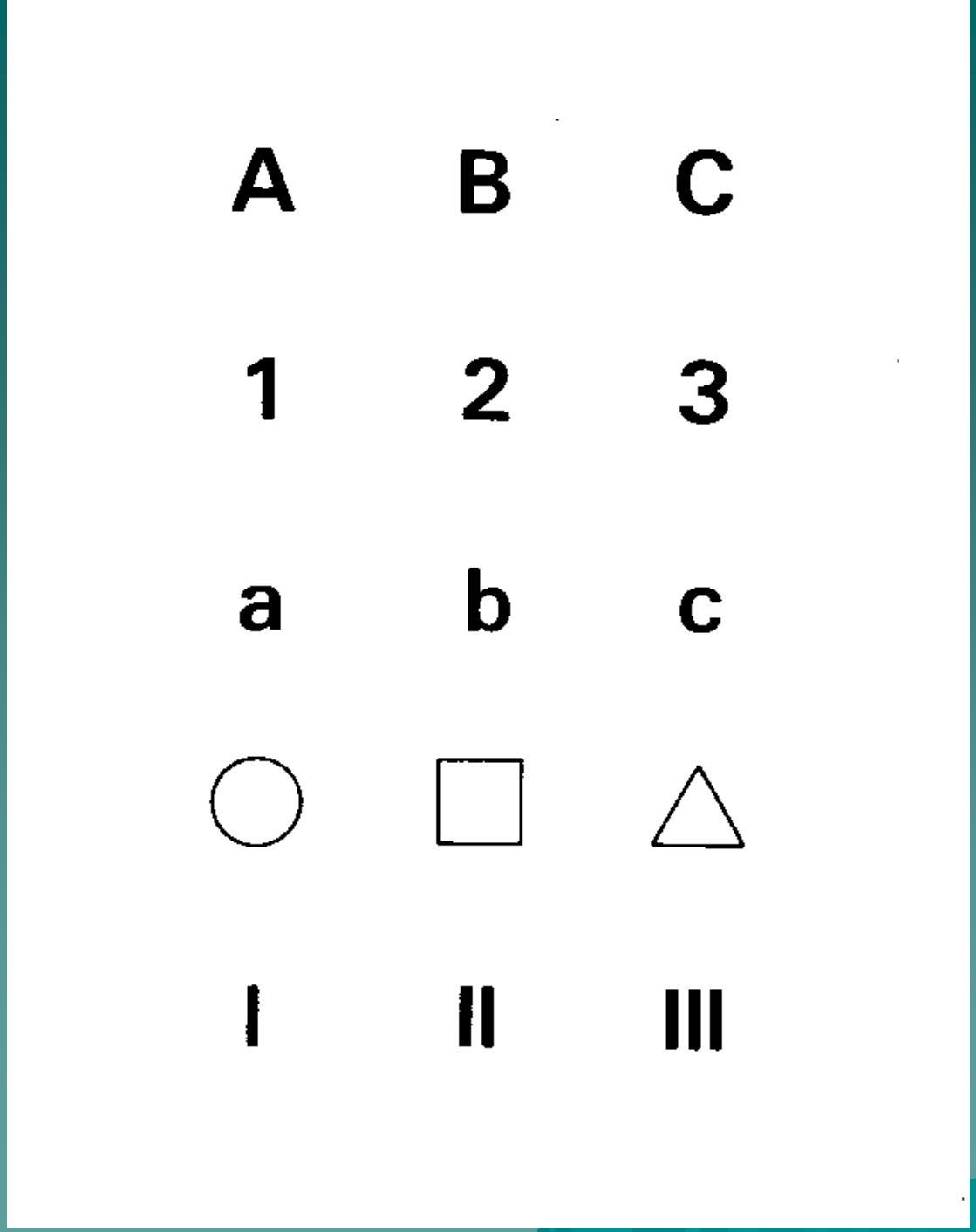
Note. See text for further discussion. PPA = positive predictive accuracy, NPA = negative predictive accuracy.
[a]Suggested for use in general clinical settings (i.e., where relatively few patients have external incentives to appear impaired).
[b]Suggested for use in settings where a mixture of compensation-seeking and non-compensation-seeking patients is seen (e.g., assessment practices comprising both medico-legal and clinical referrals).
[c]Suggested for use in settings where most or all patients are expected to have external incentives to appear impaired (e.g., practices specializing primarily in medico-legal neuropsychological assessments).

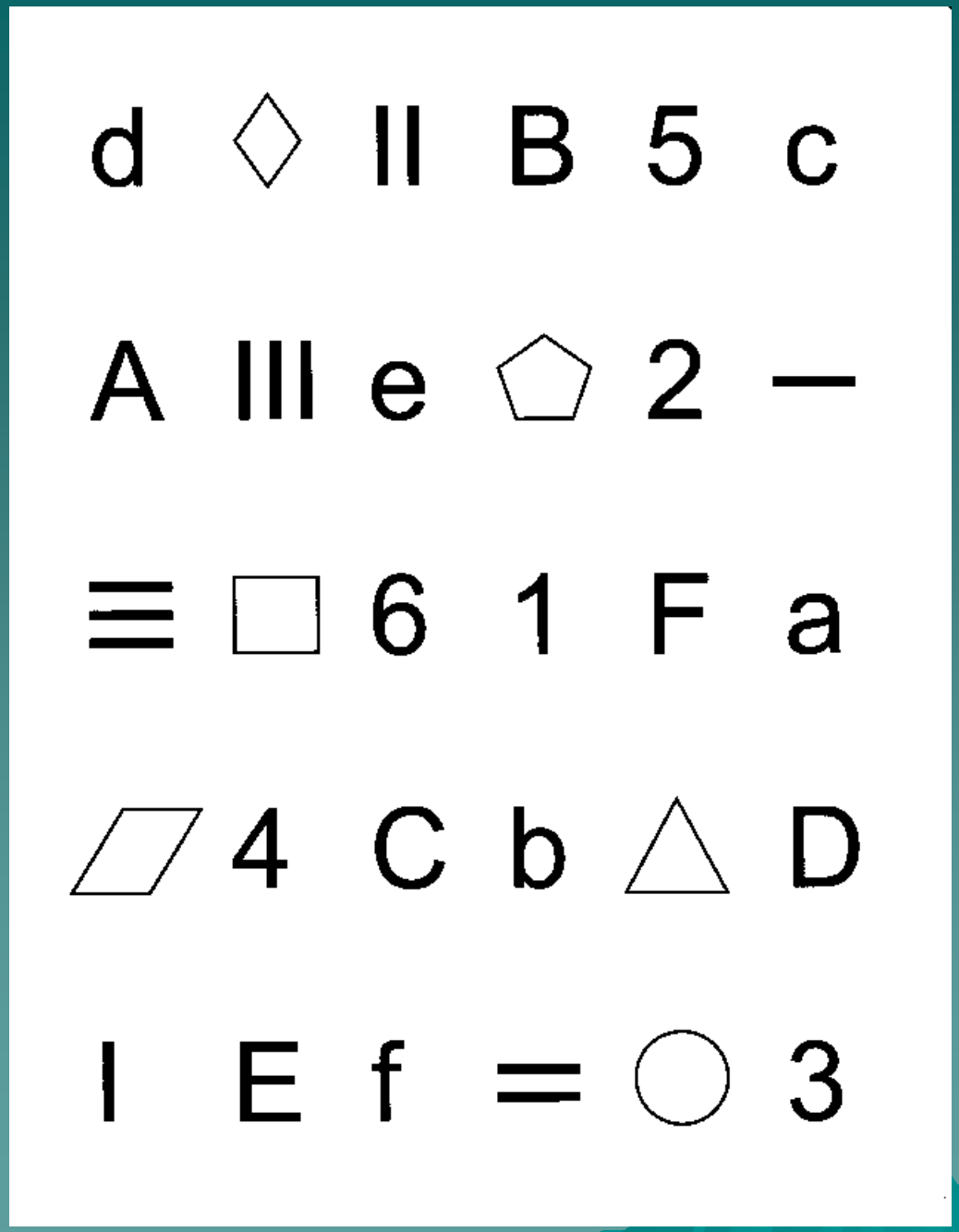
b Test: Conclusions

- Highly sensitive and specific across many clinical diagnoses although problematic for stroke and psychosis
- Brief (cost-effective)

Rey 15-item + Recognition

- Task is to memorize 15 items in a 10-second visual presentation, to draw the items from memory, and then to circle the target items on a recognition trial
- Individuals feigning brain impairment assume that overlearned information, such overlearned sequences (ABC, 123) can be forgotten, and they perform worse than true brain injured patients
- Sensitive to feigned deficits in memory
  - Boone et al. (2002c)


A B C
1 2 3
a b c
I II III


d ◊ II B 5 c
A III e ⬠ 2 —
≡ □ 6 1 F a
▱ 4 C b △ D
I E f = ○ 3

15-item Test + recognition trial

- Score
  - recall correct + (recognition correct minus false positives)
- Compared performed in
  - 49 noncredible subjects
  - 36 heterogeneous neuropsychology clinic patients
  - 33 learning disabled college students
  - 60 older controls
  - Noncredible subjects significantly underperformed relative to other groups which did not differ from each other

15-item Test + recognition trial

- Using a cut-off of <20
  - sensitivity = 71% for noncredible subjects
    - without recognition trial is only 46%
  - specificity =
    - 92% for heterogeneous neuropsychology clinic patients
    - 94% for learning disabled college students
    - 92% for older controls

Rey 15-item Plus Recognition: Conclusions

- Specificity appears relatively stable, however, sensitivity lowered
  - Why?
- Can be used to "rule in" but not "rule out" malingering
- Brief (cost-effective)

Rey Word Recognition Test

- Task is to memorize 15 unrelated words, presented auditorily once, and then to circle the target items on a recognition trial
- Individuals feigning brain impairment do not realize that recognition is easier than free recall, and they perform worse than true brain injured patients
- Sensitive to feigned deficits in memory
  - Lezak (1983)

Rey 15-Word Recognition

DETERMINING EFFORT LEVEL

[Read words at the rate of 1 word per second]

HALF	CAMEL	MISTAKE	TOY	MORNING	HAIR	WAX	GRAIN
COOKIE	FLY	PLACE	CHERRY	DOOR	KNEE	STATE	

MALINGERING: REY 15-WORD RECOGNITION MEASURE

WORD LIST

Rey Word Recognition

DETERMINING EFFORT LEVEL

WORDS IN UPPERCASE and BOLD ARE LIST WORDS

hello	Y ~~N~~		*style*	Y ~~N~~		HAIR	~~Y~~ N
gift	Y ~~N~~		WAX	~~Y~~ N		*cheese*	Y ~~N~~
CAMEL	~~Y~~ N		DOOR	~~Y~~ N		*horse*	Y ~~N~~
MORNING	~~Y~~ N		*concert*	Y ~~N~~		*smile*	Y ~~N~~
HALF	~~Y~~ N		*grass*	Y ~~N~~		*light*	Y ~~N~~
airplane	Y ~~N~~		*bottle*	Y ~~N~~		KNEE	~~Y~~ N
today	Y ~~N~~		COOKIE	~~Y~~ N		FLY	~~Y~~ N
GRAIN	~~Y~~ N		PLACE	~~Y~~ N		MISTAKE	~~Y~~ N
power	Y ~~N~~		*thread*	Y ~~N~~		CHERRY	~~Y~~ N
TOY	~~Y~~ N		*wall*	Y ~~N~~		STATE	~~Y~~ N

MALINGERING: REY WORD RECOGNITION MEASURE

RECOGNITION LIST

Dr. Bill Lynch

Name:

Date:

REY WORD TEST

HELLO	WAX	HORSE
GIFT	DOOR	SMILE
CAMEL	CONCERT	LIGHT
MORNING	GRASS	KNEE
HALF	BOTTLE	FLY
AIRPLANE	COOKIE	MISTAKE
TODAY	PLACE	CHERRY
GRAIN	THREAD	STATE
POWER	WALL	STYLE
TOY	HAIR	CHEESE

Rey Word Recognition Test

- Score
  - Total correctly recognized
    - Subtracting false positives did not increase sensitivity
- Compared performance in
  - 92 noncredible subjects
  - 51 neuropsychology clinic patients
  - 31 learning disabled college students
- Results:
  - Gender effect (women > men)
  - LD and clinic groups did not differ and were collapsed
  - Noncredible men and women underperformed relative to same gender comparison group

Rey Word Recognition Test

- Using a cut-off of
  - <u><7 for women</u>
    - Sensitivity = 80.5%
    - Specificity for female clinic and learning disabled subjects = 90.2%
  - <u><5 for men</u>
    - Sensitivity = 62.7%
    - Specificity for male clinic and learning disabled subjects = 95.1%

Rey Word Recognition Test

- 1st 8 words from list
  - more commonly recognized in credible patients (71.9% versus 59.1% for last 7 words)
  - But not in suspect effort patients (38.3% versus 32.7% for last 7 words)
- Combination score:
  - (Recognition – false positives) + number of words recognized from 1st 8 words of list <9
    - 81.6% sensitivity (with ≥90% specificity) in 38 suspect effort TBI subjects (72% sensitivity in group as a whole)

      – Nitch et al. (2006)

Rey Word Recognition Test: Conclusions

- Test highly sensitive and specific, but must use gender cut-offs
- Very brief (cost-effective)
- Equation in which first 8 words are double-weighted adds to test sensitivity in TBI subgroup

B. Noncredible scores on standard cognitive tests

- Digit Span
- RAVLT Recognition Equation
- RO Effort Equation
- RO/RAVLT Discriminant Function
- Finger Tapping
- CVLT

1 - Digit Span

- Scores
  - ACSS
  - Reliable Digit Span (RDS; the highest number string for both forward and backward in which both trials passed)
  - Time scores for forward span
- Compared performance in
    - 66 noncredible subjects
    - 56 neuropsychological clinic patients
    - 32 controls
  - Noncredible subjects < other groups on all scores
  - Clinic patients < controls on time scores
      - Babikian et al. (2006)

Instructions for Timed Digit Span:

- Administer forward digit span in normal manner, but when you finish saying the number sequence, immediately start the stopwatch. Stop timing when the patient finishes reciting the number sequence. Write the time for each trial and compute an average time for each number string.

Digit Span

- ACSS
  - ≤5
    - Sensitivity = 42%
    - Specificity = 93%
- RDS
  - <6
    - Sensitivity = 45%
    - Specificity = 93%
- Time scores for forward digit span
  - ≥2.0 seconds on average to recite a 3-digit string
    - Sensitivity = 38%
    - Specificity = 93%

Digit Span: Conclusions

- A part of neuropsychological battery, therefore no extra administration time
- Timing of forward digit span can add to test sensitivity
- Test not highly sensitive, but is specific; can be used to "rule in" but not "rule out" malingering

2 - RAVLT Recognition Trial

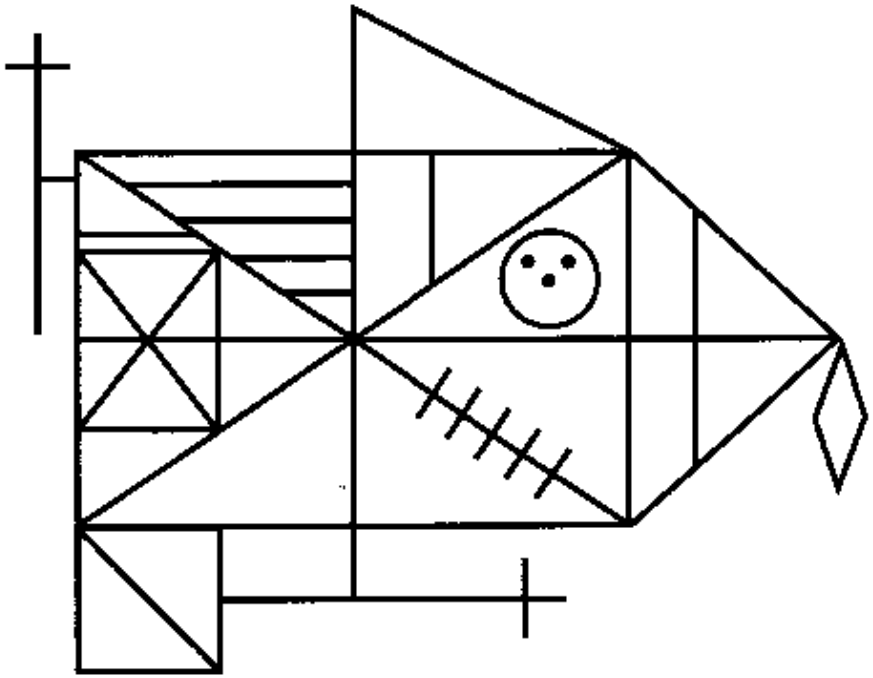
- Ray Auditory Verbal Learning Test
- Scores
  - Recognition
  - Recognition minus false positives
  - Recognition minus false positives + primacy recognition (# of words recognized from the first third of the test)
- Compared performance in
    - 61 noncredible subjects
    - 88 neuropsychology clinic patients
    - 25 controls
  - Noncredible subjects underperformed relative to the two other groups which did not differ from each other
      - Boone, Lu, and Wen (2005)

RAVLT Recognition Trial/Equation

- Recognition <10
  - Sensitivity = 67%
  - Specificity = 90+%
- Recognition <8 (minus false positives)
  - Sensitivity = 64%
  - Specificity – 90+%
- Recognition minus false positives + primacy recognition (# of words recognized from the first third of the test) <12
  - Sensitivity = 74%
  - Specificity = 90+%

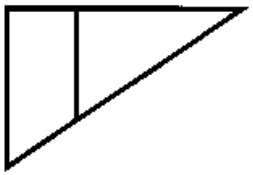
RAVLT Recognition Equation: Conclusions

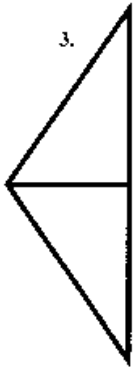
- A part of neuropsychological battery, therefore no extra administration time
- Consideration of primacy effect enhances test effectiveness
- Good sensitivity and specificity



Name_____ ID #_____ Test Date / /

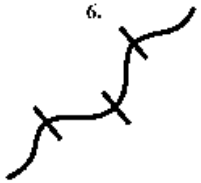
1. 

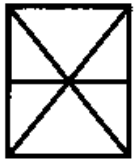
2. 

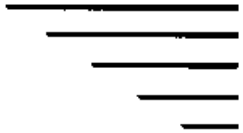
3. 

4. 

5. 

6. 

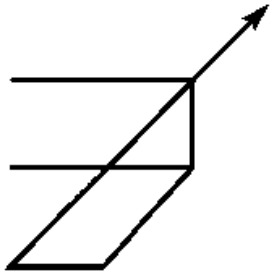
7. 

8. 

9. 

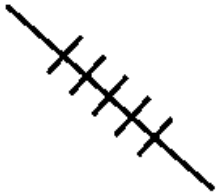
10.

11.

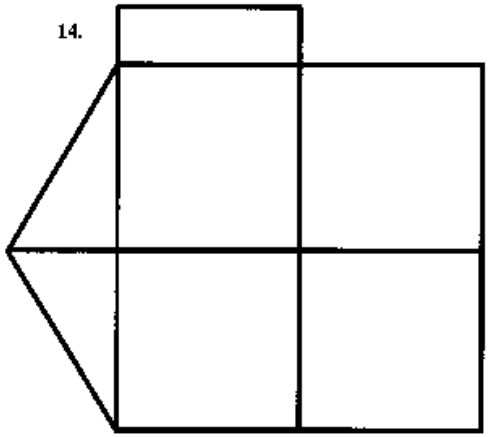


12.

13.

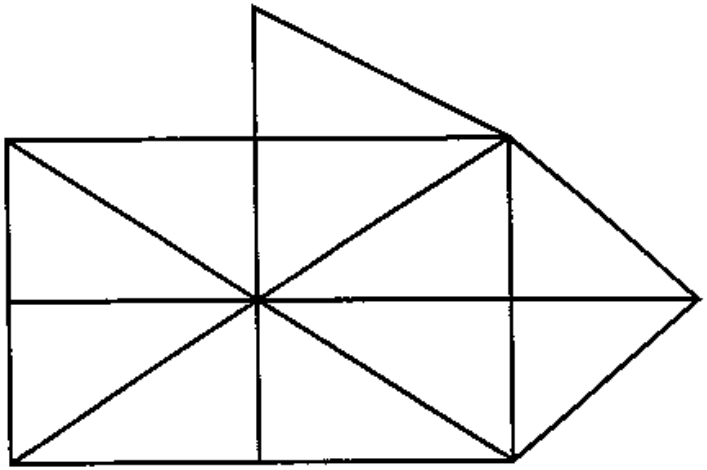


14.



Name_____ ID #_____ Test Date___/___/___

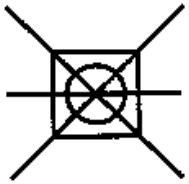
15.



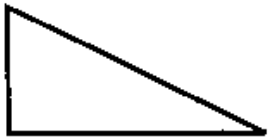
16.

17.

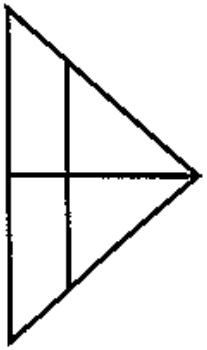
18.



19.

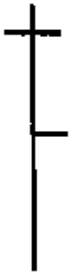


20.



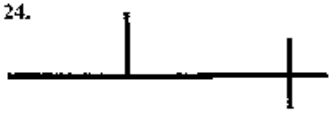
21.

22.



23.

24.



3 - ReyO Effort Equation

- Scores
  - Copy
  - 3-minute recall
  - Recognition (Meyers & Meyers, 1995)
- Compared performance in
  - 58 noncredible subjects
  - 30 clinic patients without memory impairment
  - 23 clinic patients with verbal memory impairment
  - 17 clinic patients with visual memory impairment
  - Noncredible subjects < all groups on recognition and < no memory impairment and verbal memory impairment groups in copy and immediate recall

- Lu et al (2003)

RO Effort Equation

- Following score increased sensitivity
  - copy + [(recognition minus atypical false positive errors [#1,4,6,10,11,16,18,21]) x 3]
- Using a cut-off of <47
  - sensitivity = 76%
  - specificity =
    - 91% for comparison groups combined
      - 93% in non-memory impaired group
      - 96% in verbal memory impaired group
      - 82% in visual memory impaired group

RO Effort Equation: Conclusions

- A part of neuropsychological battery, therefore no extra administration time
- Good sensitivity and specificity for non-memory impaired and verbal memory impaired patients
  - But nearly 20% false positive rate in visual memory impaired patients
- This equation outperformed the Meyers and Volbrecht (1998) Memory Error Patterns for identifying suspect effort
  - Applied to our sample, MEPs had sensitivity of 26% to 50% and specificity of 52% to 100%

4 - Rey-Osterrieth/RAVLT Discriminant Function

- Examined performance in
  - 38 noncredible subjects
  - 34 neuropsychology clinic patients with documented brain injury (15 moderate to severe HI, 9 stroke, 10 tumor/cyst)
  - 33 controls
- Scores
  - Bernard (1990) and Bernard, Houston, and Natoli (1993) Discriminant Function
  - Discriminant Function derived from current sample
    - .006(RAVLT trial 1) - .062(Rey figure delay) + .354(RAVLT recognition) - 2.508

  - Sherman et al. (2002)

Rey-Osterrieth/RAVLT Discriminant Function

- Functions
  - Bernard:
    - Sensitivity = 95%
    - Specificity = 33% for patients, 61% for controls
  - Harbor-UCLA cut-off <-.41:
    - Sensitivity = 71%
    - Specificity =
      - 91% for documented brain injury
      - 100% for controls

RO/AVLT
Discriminant Function: Conclusions

- A part of neuropsychological battery, therefore no extra administration time
- Good sensitivity and specificity
- This discriminant function outperformed a function developed on simulators and controls

5 - Finger Tapping

- Scores
  - Dominant hand mean
  - Nondominant hand mean
  - Sum of the average scores for the two hands
  - Difference between dominant and nondominant hand averages
- Compared performance in
  - 77 noncredible
  - Comparison groups
    - Closed head injury (n = 24)
    - Depression (n = 42)
    - Psychosis (n = 27)
    - Low IQ (FSIQ $\leq$70; n = 18)
    - Dementia (n = 31)
    - Older Controls (n = 18)

- Arnold et al. (2005)

Finger Tapping

- Need separate gender cut-offs
  - Men
    - Dominant hand <35 = 50% sensitivity; 90% specificity in male comparison groups combined
      - 87% in HI (need cut-off of ≤33)
      - 95% in depressed
      - 90% in psychotic
      - 78% in low IQ (need cut-off of ≤33)
      - 87% in dementia (need cut-off of ≤21)
      - 100% in older controls

Finger Tapping

- Need separate gender cut-offs
  - Women
    - Dominant hand <28 = 61% sensitivity; 92% specificity in female comparison groups combined
      - 100% in head injured
      - 95% in depressed
      - 88% in psychotic (need cut-off of ≤15)
      - 87% in low IQ (need cut-off of ≤15)
      - 75% in dementia (need cut-off of ≤15)
      - 100% in older controls

Finger Tapping: Conclusions

- Dominant hand score most effective
- Must use gender-specific cut-offs
- Moderate sensitivity
- A part of neuropsychological battery, therefore no extra administration time

Now what? How to Interpret the tests in concert

Victor, Boone, Serpa, and Buehler ([almost] in press)

- Subjects
  - 37 noncredible (defined by ≥2 or more failures on Rey 15-itme plus recognition, Dot Counting, Warrington Words, b-Test, Rey Word Recognition and motive to feign)
  - 66 credible (defined by failing <2 of the above indicators, no motive to feign, and did not meet criteria for mental retardation or dementia
  - Predictor Variables: Digit Span (RDS), RO equation, RAVLT effort equation, Finger Tapping
- Pairwise failure was superior to use of any one test by itself (sensitivity = 83.8%, specificity = 93.9%, overall hit rate = 90.3%).
- One failure was highly sensitive (94.6%) but with low specificity (53.9%).
- **Failure on three tests was associated with almost perfect specificity (98.5%) but low sensitivity (51.4%).**
- **Thus, failure on two tests was the most accurate and efficient for determining group membership**

Now what? How to Interpret the tests in concert

- Other authors have also reported that 2 failures is most effective in separating groups:
  - Meyer and Volbrecht (2003): 2 of 9 indicators
  - Suhr et al. (1997): 2 of 4 indicators
  - Larrabee (2003): 2 of 5 indicators

Failure on ≥2 indicators:

- Appears to be best cut-off
- However, cannot be used as absolute criterion
  - Some patient subgroups are particularly likely to show false positives on effort tests

How to limit false positive identifications

- First, administer several effort indicators
  - Failure on increasing number of indicators does not increase sensitivity, but does increase specificity
    - e.g., 1/10 x 1/10 x 1/10 x 1/10 x 1/10 x 1/10 = 1/1,000,000
    - 2/10 failures versus ≥5 out of 10 failures (94% specificity versus essentially perfect specificity)

Adjust cut-offs

- Second, adjust cut-offs for conditions that reduce specificity
  - Low IQ?
  - Dementia?
  - ESL and Ethnicity?
  - Learning disability?
  - Psychiatric Condition?

IQ and Effort Test Performance

- 189 neuropsychology clinic outpatients
  - no motive to feign (not in litigation or attempting to obtain disability compensation)
  - Excluded dementia, amnestic disorder and somatoform disorder
  - Mean age = 43.6 (SD = 14.1)
  - Mean education = 12.8 (SD = 2.8)
  - 53% female
    - Dean et al. (2007)

Effort tests failed by IQ band:

FSIQ band	n	Mean failed	range	Mean %
50-59	3	4.0	1-6	60%
60-69	12	2.9	1-6	44%
70-79	48	1.1	0-4	17%
80-89	44	.5	0-4	8%
90-99	39	.3	0-2	7%
100-109	27	.2	0-1	4%
110-119	11	.4	0-2	6%
≥120	5	.2	0-1	5%

Table. Specificity of each indicator by IQ band. As IQ↓↓…

IQ Band	RMT	FTT	DCT	R-O	DSp	DiscFx	Rey15	ReyW
50-59	100	100	**33**	**33**	**33**	**67**	**33**	**0**
60-69	**67**	100	**71**	**64**	**33**	**75**	**45**	**57**
70-79	100	**83**	**85**	**81**	**81**	93	**68**	**85**
80-89	100	**93**	**88**	**85**	98	95	**84**	**71**
90-99	100	97	**88**	93	97	93	90	**100**
100-109	100	**89**	100	**88**	100	**89**	100	100
110-119	100	**89**	100	**80**	100	100	91	100
120-129	100	100	100	100	100	**80**	100	100

Conclusions about IQ and Effort:

- Effort test performance is significantly related to intelligence
- Individuals of MR levels of intelligence fail on average >44% of effort tests in a battery even when putting forth full effort
- Specificities of most indicators drop to inadequate levels with individuals of borderline and MR IQ
- In a subsample of MR Ss, adequate specificity only found for the Warrington and finger tapping, TOMM

Dementia and Effort Test Performance

- 214 patients with dementia
  - no motive to feign (not in litigation or attempting to obtain disability compensation)
  - Excluded patients with delirium and amnestic disorder
  - Mean age = 63.5 (SD = 15.1)
  - Mean education = 13.1 years (SD = 2.9)
  - 49% female
  - Mean MMSE = 18.5 (SD = 6.0)

    - Dean et al. (in press)

Specificity by MMSE 21-30 15-20 <15

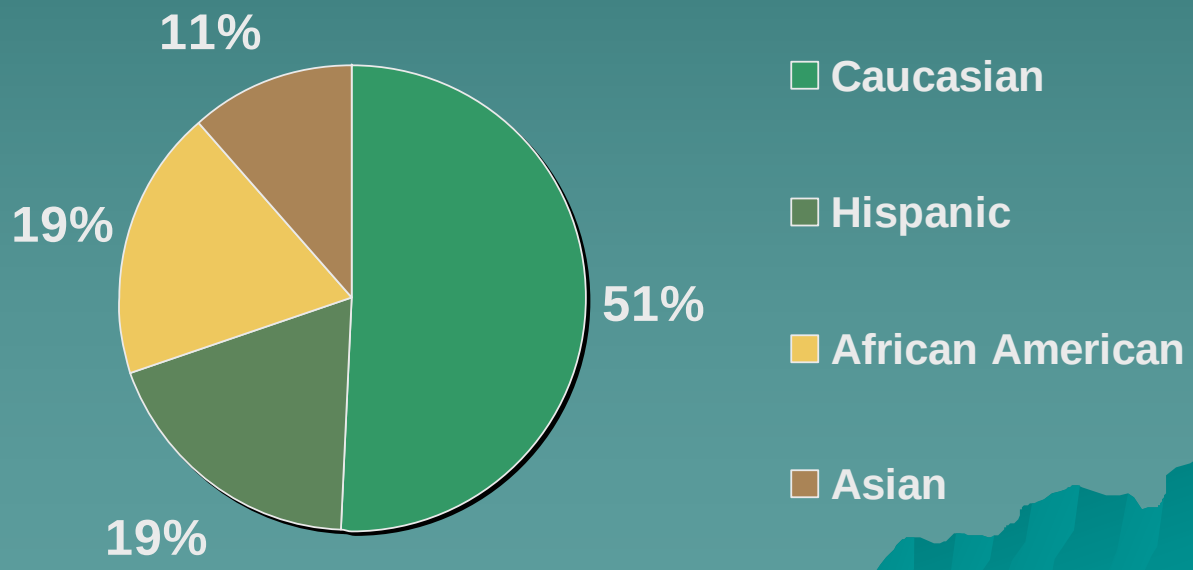
DS ACSS	84%	67%	33%
4-digit x	94%	83%	100%
Voc - DS	94%	100%	100%
Dot Counting	77%	44%	8%
TOMM	63%	33%	0%
Warrington	73%	20%	0%
15-item	21%	0%	0%
tapping	70%	83%	100%
b Test	50%	38%	0%
Rey Word	64%	83%	50%
RAVLT equ.	15%	0%	0%
RO equation	44%	15%	0%
RO/AVLT fx	44%	29%	0%

Conclusions

- Most indicators had specificities 30-70%
  - Although cut-offs for some Digit Span indicators (Vocab minus Digit Span and 4-digit time) maintained ≥90% specificity
  - Finger tapping specificity was preserved in AD and FTD, but not vascular dementia
- Adjusting cut-offs to protect specificity in a dementia population generally lowers sensitivity to unacceptable levels
- New measures need to be developed for the differential diagnosis of actual and feigned dementia

Ethnicity/ESL and Effort Test Performance

- Study 1: 168 fluent English-speaking neuropsychology clinic patients
  - exclusion criteria: in litigation or disability-seeking; FSIQ <70; diagnosis of dementia



Salazar, Wen, Lu, & Boone (2007)

Group Comparisons

- Group comparisons
  - Groups comparable in age, but differed in education (educational level only sig related to Digit Span ACSS and RDS and RO/AVLT fx)
    - Asians and Caucasians>African-Americans and Hispanics; African-Americans>Hispanics
  - Caucasians > Hispanics
    - Digit Span ACSS and Reliable Digit Span
  - Caucasians > African Americans
    - RAVLT recognition, RAVLT recognition equation, RO effort equation, RO/AVLT disc. function

Specificity by Ethnicity/ESL

Test	White	AA	Hisp	Asian	ESL
DS ACSS≤5	98.8	93.5	**87.5**	100	**88.9**
DS RDS≤6	100	93.3	**81.5**	93.7	**84.0**
15+rec<20	**87.5**	**79.2**	**75.9**	**78.6**	**82.6**
AVLT rec≤7	97.5	**86.7**	100	**82.4**	95.8
AVLT eq≤12	**85.1**	**72.4**	92.3	**86.7**	95.7
Warr <33	100	100	100	100	100
DCT ≥17	91.8	92.0	92.6	**81.3**	**81.8**
RO eq ≤47	94.0	**66.7**	**79.2**	93.7	90.9
Disc Fx≤.40	94.0	**81.8**	95.6	**86.7**	95.2

Cut-offs that maintain Specificity ≥90%

Standard	White	AA	Hisp	Asian	ESL
DS ACSS≤5	≤6	≤5	≤4	≤5	≤4
DS RDS≤6	≤6	≤6	≤5	≤6	≤5
15+rec. <20	≤18	≤17	≤15	≤12	≤12
AVLT rec.≤7	≤9	≤4	≤10	≤5	≤11
AVLT eq≤12	≤11	≤1	≤13	≤5	<13
Warr <33	≤39	≤37	≤40	≤35	≤40
DCT ≥17	≥15	≥17	≥16	≥21	≥21
RO eq ≤47	≤47	≤34	≤44	≤47	≤47
Disc Fx≤-.40	≤-.19	≤-1.41	≤.24	≤-.86	≤.30

Study 2: Effort Test Scores in Monolingual Spanish-speakers

- 108 Male Native-Spanish-speaking Day Laborers
  - Mean age of 30.58
  - Mean educational level of 6.11 years
  - Mean residency in US of 44.34 months
    - Exclusion criteria: head injury, DX of cognitive disorder, substance abuse
    - Salazar et al. (2003)

Monolingual Spanish-speakers

- Rey 15-item plus recognition
- Dot Counting Test
  - Cut-off >17:
    - specificity = 95.4%
  - Cut-off <20:
    - Specificity
      - 81% in ≥6 years education
      - 68% in <6 years education
  - Cut-off <17:
    - Specificity
      - 91% in ≥6 years education
      - 82% in <6 years education

Recommendations

- 1) Ultimately, cut-offs will need to be developed for ethnicity/acculturation/ESL
  - However, current data are preliminary (small n's, groups may not have been comparable in diagnosis)
  - Educational level is a confound

Recommendations

- 2) In interim, "de-emphasize" those failed effort indicators on which your patient's ethnic/language group underperforms
  - e.g., note if patient only fails those indicators which are problematic for his/her ethnic/language group; if so, add a caveat indicating that cultural factors may have impacted performance

Learning Disability and Effort Test Performance

- Many effort tests involve reading, letter identification, rote math skills
  - Does this place LD individuals at risk for failure on these measures?

Warner-Chacon & Boone (2007)

Ziegler et al. (under submission)

Learning Disability and Effort Test Performance

- LD population (n = 31) receiving services through Office of Disabled Student Services at Cal State Northridge
  - Dot Counting Test
    - Requires rote math skills/basic multiplication
    - LD scored significantly higher than noncredible subjects
    - Cut-off >13 associated with 90% specificity
  - b Test
    - Requires rapid letter identification
    - LD scored significantly higher than noncredible subjects
    - Cut-off >140 associated with 90% specificity

Impact of Learning Disability on Effort Test Performance

- Rey 15-item + Recognition
  - Requires recall/identification of letters and numbers
  - LD scored significantly higher than noncredible
  - Cut-off <20 associated with 93% specificity
- Rey Word Recognition Test
  - Requires recall/identification (reading) of words
  - LD = controls (2 groups collapsed for determining cut-offs)
  - LD and controls scored significantly higher than noncredible

Impact of low math skills on Digit Span and Dot Counting

- 242 neuropsychology clinic patients with no motive to feign
  - For Digit Span ACSS and RDS, specificity intact with Arithmetic subtest ACSS (AACSS) ≥7, however with AACSS of 5 or 6, specificity for RDS was 78% and Digit Span ACSS was 80%, and with AACSS of ≤4, specificity for RDS was 58% and Digit Span ACSS was 66%
  - For Dot Counting, specificity was maintained at Arithmetic ACSS of ≥9, however, specificity declined to 83% with AACSS of 7 or 8, to 78% with AACSS of 6, to 68% with AACSS of 5, and to 58% for AACSS ≤4.
- Thus, Digit Span and Dot Counting performance is related to math ability and cut-off adjustments as a function of Arithmetic ACSS may be required

Conclusions/Recommendations

- Identified LD does not appear to raise risk of false positive identifications to any significant degree
  - Although the empirical data have been confined to college student LD subjects, who may not be representative of the entire LD population
- Examination of clinic patients with low math skills reveals elevated false positives on Dot Counting and Digit Span effort indicators
- No data for many effort indicators

Psychiatric Conditions and Effort Test Performance

- Many effort tests involve processing speed and attention
  - Does this place depressed or psychotic individuals at risk for failure on these measures?
    - In some neuropsychological reports it has been asserted that patients may have failed effort tests because of depression – is this true?

Goldberg, Back-Madruga, and Boone (2007)

Impact of Depression/Psychosis on Effort Test Performance

- 64 older outpatients meeting criteria for major depression and 28 outpatients with chronic schizophrenia at Harbor-UCLA Medical Center
  - Dot Counting Test
    - Patients with depression and schizophrenia scored significantly higher than noncredible subjects
      - Cut-off >12 associated with 90% specificity in depression
      - Cut-off has to be raised to ≥19 to achieve 90% specificity in schizophrenia
  - b Test
    - Patients with depression and schizophrenia scored significantly higher than noncredible subjects
      - Cut-off >120 associated with 90% specificity in depression
      - Cut-off has to be raised to >190 to achieve 89% specificity in schizophrenia

Boone et al. (2002a, b)

Impact of Depression/Psychosis on Effort Test Performance

- 64 older outpatients meeting criteria for major depression at Harbor-UCLA Medical Center
  - 15-item Test and Dot Counting Test
    - Mild (n = 22), moderate (n = 31), and severe (n = 11) subgroups did not differ in test performance
  - Specificity >90%

- Lee et al. (2000)

Impact of Depression/Psychosis on Effort Test Performance

- 42 outpatients with depressive symptoms and/or diagnosis and 27 outpatients with psychotic symptoms and/or diagnosis at Harbor-UCLA Medical Center
  - Finger Tapping Test (dominant hand score)
    - Female depression and psychosis groups scored significantly higher than noncredible female subjects
      - Cut-off <38 associated with 90% specificity in depression
      - Cut-off has to be lowered to <32 to achieve 88% specificity in psychosis
    - Male depression group scored significantly higher than noncredible male subjects
      - Cut-off <38 associated with 95% specificity in depression
      - Cut-off <40 associated with 90% specificity in schizophrenia
        - Arnold et al. (2005)

Conclusions

- Our data suggest that
  - depression and psychosis do not significantly impact finger tapping performance
  - Depression is not associated with lowered performance on Dot Counting or original Rey 15-item, but is associated with mild lowering of scores on b Test; however, cut-offs can be adjusted to maintain specificity
  - Psychosis is associated with mild lowering of performance on Dot Counting, and to a somewhat greater extent on the b Test; however, cut-offs can be adjusted to maintain specificity
    - The lowered performance in the psychotic group appears to be confined to a subgroup with lowered educational level and cognitive function (i.e., lowered MMSE; Back et al., 1996)

Boone's Checklist regarding measuring Response Bias

- Administer numerous effort indicators
  - Not <5 and preferably more (more confidence in conclusions)
  - Check for performances likely pathognomonic for feigning (e.g., significantly below chance on FC, numerous circled "d's" on b Test)
- Determine if patient is in a high risk group regarding effort test failure
  - If so, adjust cut-offs for specific differential diagnosis
    - E.g., actual versus feigned dementia, mental retardation, psychosis
- Check qualitative behaviors (claim can't identify touched finger because it is numb, claim can't lift finger up from tapper, drawing objects upside down)
- Assess for inconsistencies between test scores and ADLs and between test scores over time
- Assess whether cognitive scores are expected for claimed diagnosis (mild TBI, depression, etc.)

Future Research

- Effort indicators specific to particular differential diagnoses need to be developed
  - E.g., tests effective for normal IQ populations are not necessarily effective in low IQ groups, etc.
- Need to incorporate differential weighting of effort tests based on sensitivity

Suspect Performance in the Context of Mild TBI

- 39-year-old male attorney involved in motor vehicle accident 5 years previous
- No loss of consciousness, head a traumatic on evaluation, alert/fully oriented, and did not complain of head symptoms
- Brain CT and MRI normal

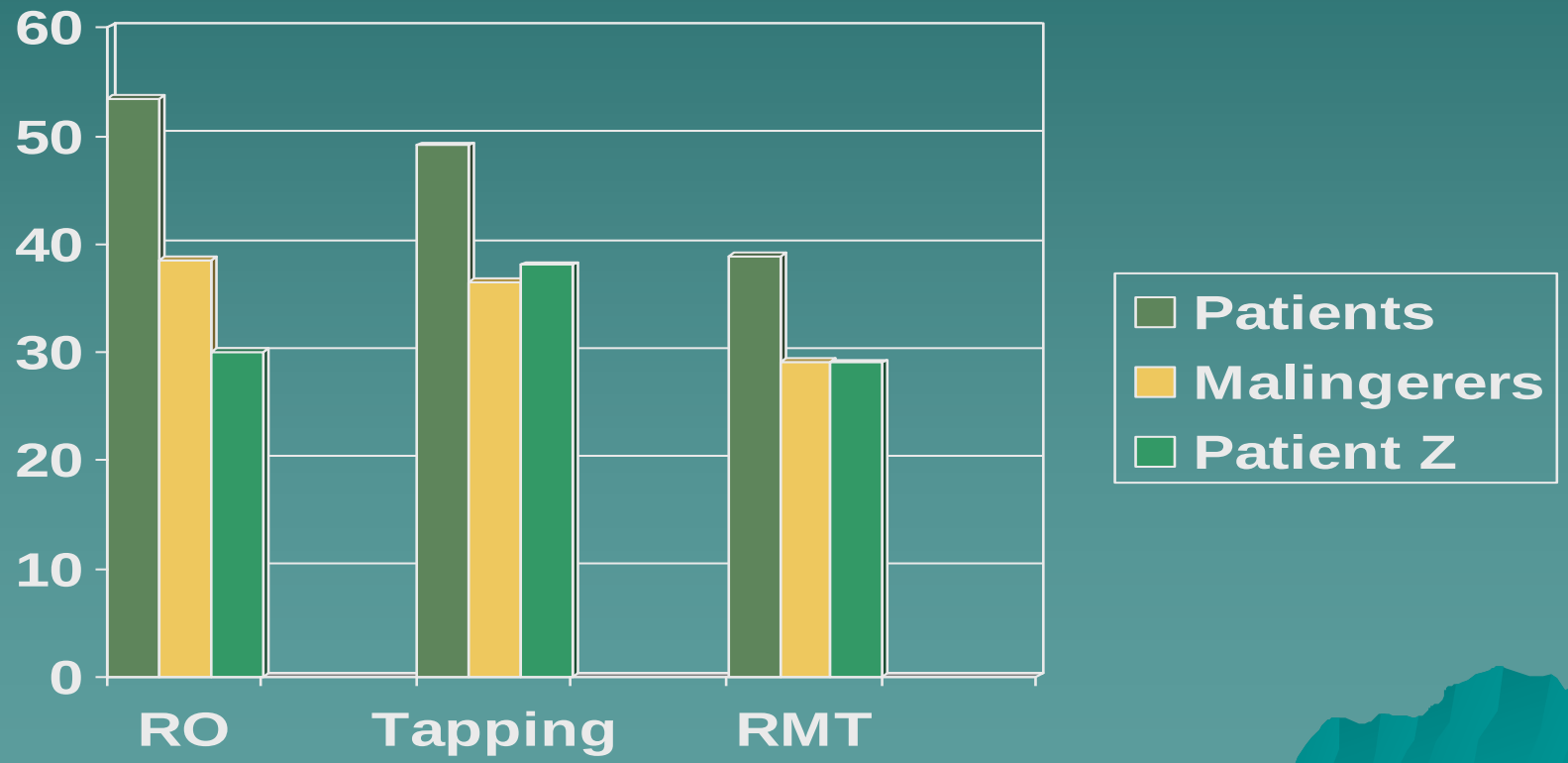
Suspect Performance in the Context of Mild TBI

- Returned to work full-time
- At time of eval, complaining of severe daily headaches, dizziness, neck and back pain, depression, poor sleep, anxiety attacks, "intrusive thoughts", nightmares, poor memory, reduced concentration, and problems in word-finding, math, and "thinking clearly"

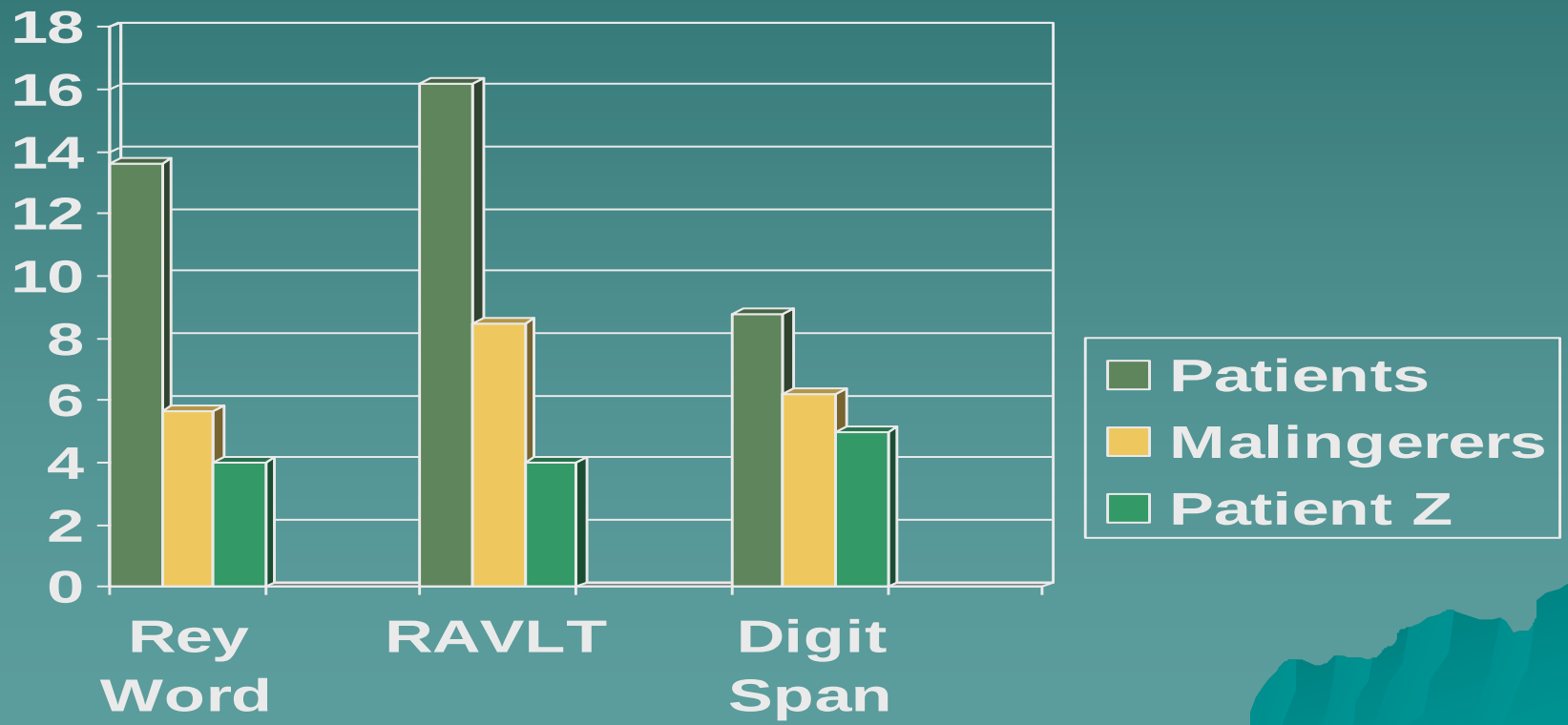
Suspect Performance in the Context of Mild TBI

- However, failed 9 of 13 indicators:
  - Failed 2 dedicated effort tests:
    - Warrington = 29/50 (cut-off <33)
    - Rey Word Recognition Test = 4 (cut-off for men ≤5)
  - Not credible on standard cognitive tests sensitive to feigned performance
    - Finger Tapping: dominant hand = 38 (cut-off for head injured males ≤38)
    - Digit Span ACSS = 5 (cut-off ≤5), and mean time to recite 3 digits = 3" (cut-off >2")
    - RO Equation = 30 (cut-off ≤47)
    - RAVLT Effort Equation = 4 (cut-off ≤12) and RO/RAVLT discriminant function = -1.775 (cut-off ≤-.40)
    - Finger Agnosia errors = 4 (cut-off >3)
      - When middle finger on left hand touched, he paused, and, with eyes closed, said that "that finger is numb"

Patient Compared to Credible Patients and Malingerers



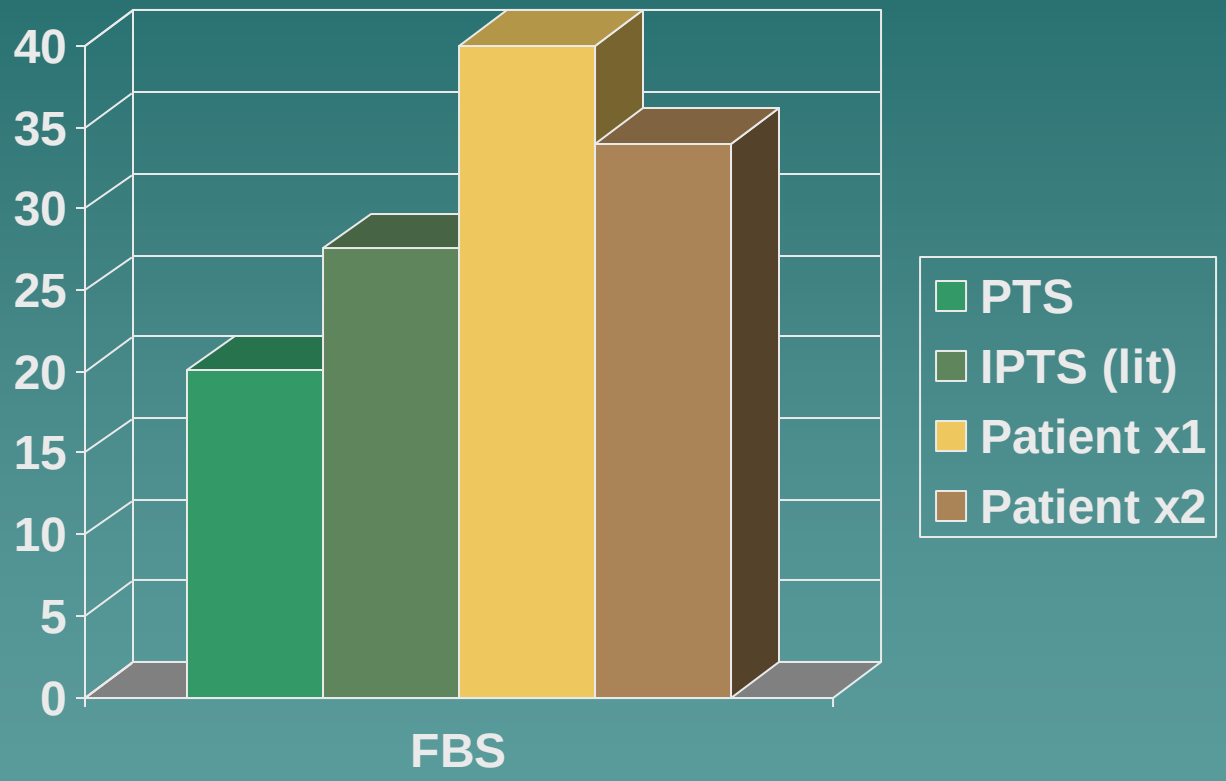
Patient Compared to Credible Patients and Malingerers



Suspect Performance in the Context of Mild TBI

- On personality inventories/pain symptoms questionnaires,
  - MMPI-2
    - Fake Bad Scale of 34 ($\geq$30 associated with 100% probability of malingering)
    - VRIN = 38T (his extreme carefulness in completing MMPI-2 would not be predicted/possible if his low cognitive scores were accurate)
  - MSPQ
    - Score of 34 (cut-off $\geq$14)

Actual and Improbable PTSD and Patient on FBS



Greiffenstein et al. (2004)

PTSD

- PTSD does not affect cognitive abilities and so SVTs should not be affected at all.
- In Roger Gervais 's series, PTSD cases failed effort tests less often than any other diagnostic group.

Suspect Performance in Mild TBI

- Improbable test scores:
  - Test scores did not match those expected for mild traumatic brain injury
    - 5 meta-analyses* show no chronic cognitive sequelae from mild TBI (as defined by loss of consciousness <30 minutes, Glasgow Coma Scale 13 to 15 of 15, anterograde amnesia <24 hours, time to follow commands <1 hour, normal brain imaging)
- Test scores do not match ADLs
  - Claimant was working for a legal temp agency up to 15 hours in one day

*Belanger, et al. (2005); Belanger & Vanderploeg (2005); Carroll et al. (2004); Frencham et al. (2005), Schretlen & Shapiro (2003)

Books

- Boone, K. (Ed). (2007). *Assessment of Feigned Cognitive Impairment: A Neuropsychological Perspective* . New York , NY : Guilford Publications.
- Larrabee, G. J. (Ed.) (2007). *Assessment of Malingered Neuropsychological Deficits* . New York : Oxford University Press.
- Morgan, J. & Sweet, J. (Ed's.) (2008). *Neuropsychology of Malingering Casebook* . New York : Taylor & Francis/American Academy of Clinical Neuropsychology and Psychology Press.
- Rogers, R. (Ed.) (2008). *Clinical assessment of malingering and deception (Third Edition)* New York : Guilford .