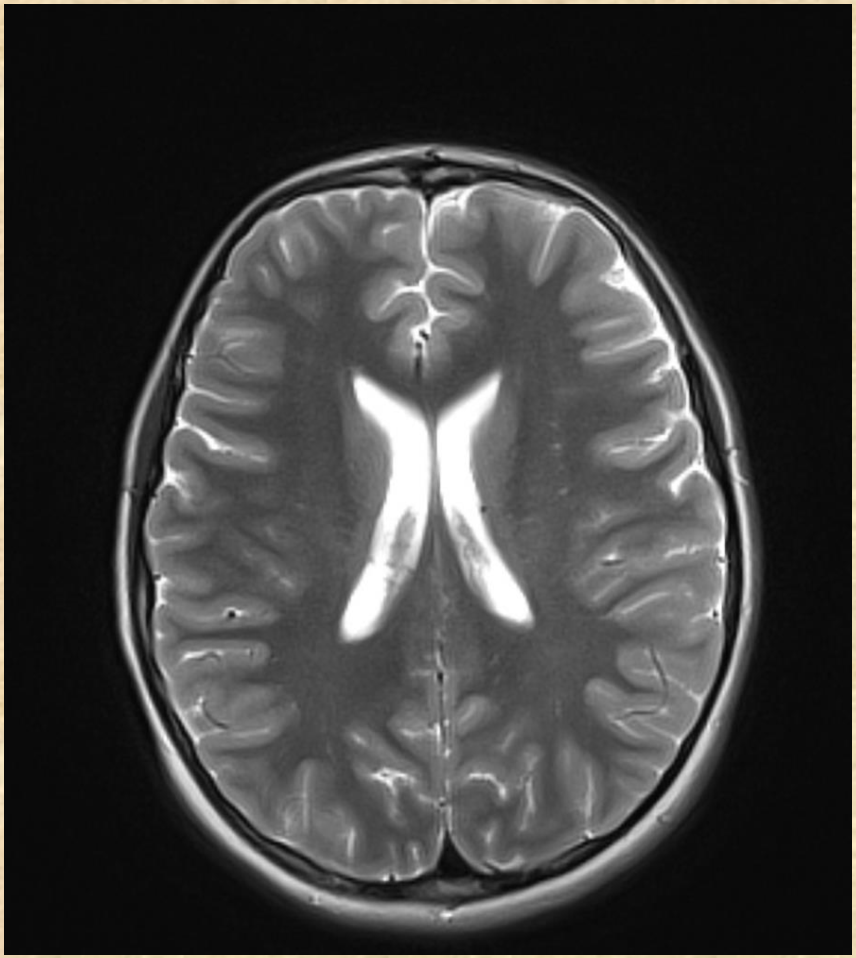
Clinical accuracy in Neuropsychological Testing : normal variation v's abnormal deficits

Dr. Sundeep Sembi.
Consultant Clinical Neuropsychologist
Psychology Chambers Ltd
10 Harley Street, London. W1G 9PF
Tel: 020 7467 8365 **Mobile** 07813873578
Email: Info@psychologychambers.com
www.psychologychambers.com

This presentation assumes that Effort and validity are satisfied



Assessment v's Testing

- Neuropsychological assessment, done properly, is much more than testing alone and critically has to be if it is to be of value to the medico legal process.

- The assessment process places the results of tests in the context of information gathered at interview, evidence in the notes but critically evidence of functioning in the real world

As the previous presentation conveyed

- Self-report alone can however be unreliable.
- Claimants report difficulties to experts, treating clinicians, their GP and lawyer to varying degrees
- Claimants may be poor historians, lack insight as to the full extent of their impairments or be in denial and thus report difficulties inaccurately for genuine reasons
- Claimants may not able to differentiate neuropsychological symptoms and sincerely report e.g., word finding difficulty, slowed processing speed or sequencing difficulties all as *"memory problems"*
- Significant levels of apparent neuropsychological and neurological symptoms can also be reported in a variety of conditions meaning their complaints may be, in part or entirely, a reflection of distress or secondary to non-organic factors and not brain injury

As the previous presentation reported

How many of these symptoms would you assume indicate TBI?
(Lees-Hayley & Brown 1993)

Family Practice Group	Non-TBI Claimants	Symptom
58%	79%	Fatigue (mental or physical)
26%	78%	Concentration problems
38%	77%	Irritability
24%	61%	Feeling disorganised
16%	59%	Confusion
20%	53%	Memory problems
20%	34%	"Word finding problems, not finding the word you want, using the wrong word"*
12%	24%	Trouble reading
16%	18%	Speech problems

Scores v's self-report

- It is critical that test scores are placed in context

- Different patients could fail the same test for different reasons.
  - A demented patient may fail a memory test because of their memory impairment however a frontal lobe patient, despite their memory being intact, may fail the same test because of their inability to choose an appropriate memory strategy. Both these patients might achieve identical scores on the same test.

- Reliance upon scores alone is inappropriate

However it is also recognised that

test results, in isolation, are generally a better indicator of ability than interview responses alone

How are scores interpreted

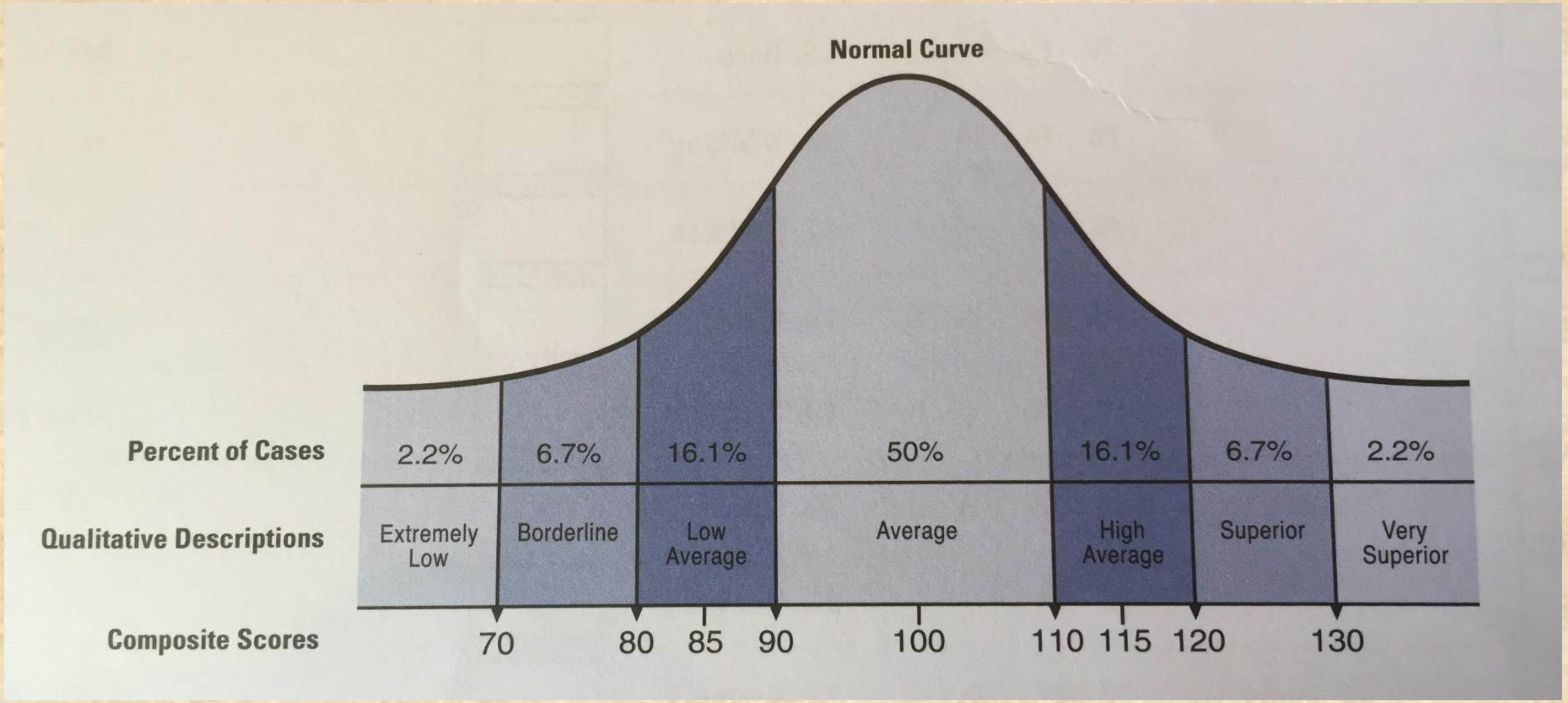
- In practice one looks for evidence of decline in scores from estimated pre-accident ability.
- <u>But</u> in the context of a deterioration in independence and neurobehavioural change.
- Failure to precisely capture the functional consequences of impairment means interpretation will be less accurate. Not placing apparently normal or mildly depressed results in the context of real world functioning may miss upon significant disability in the real world however it may equally magnify disability where there is none.
- A third facet is the failure to distinguish between impairment and disability due to secondary factors such as mental health, fatigue or pain, which may be more amenable to treatment.

As we are all familiar it is within these arenas that defendant-claimant perspectives often vary.

Determining Impairment

- Quantifying and describing cognitive deficits, patterns of deficit and distinguishing between statistical & clinical significance.
- It is standard practice to use normative data to quantify test performance.
- Traditionally it is usual to consider where an individual was compared to where they are now.

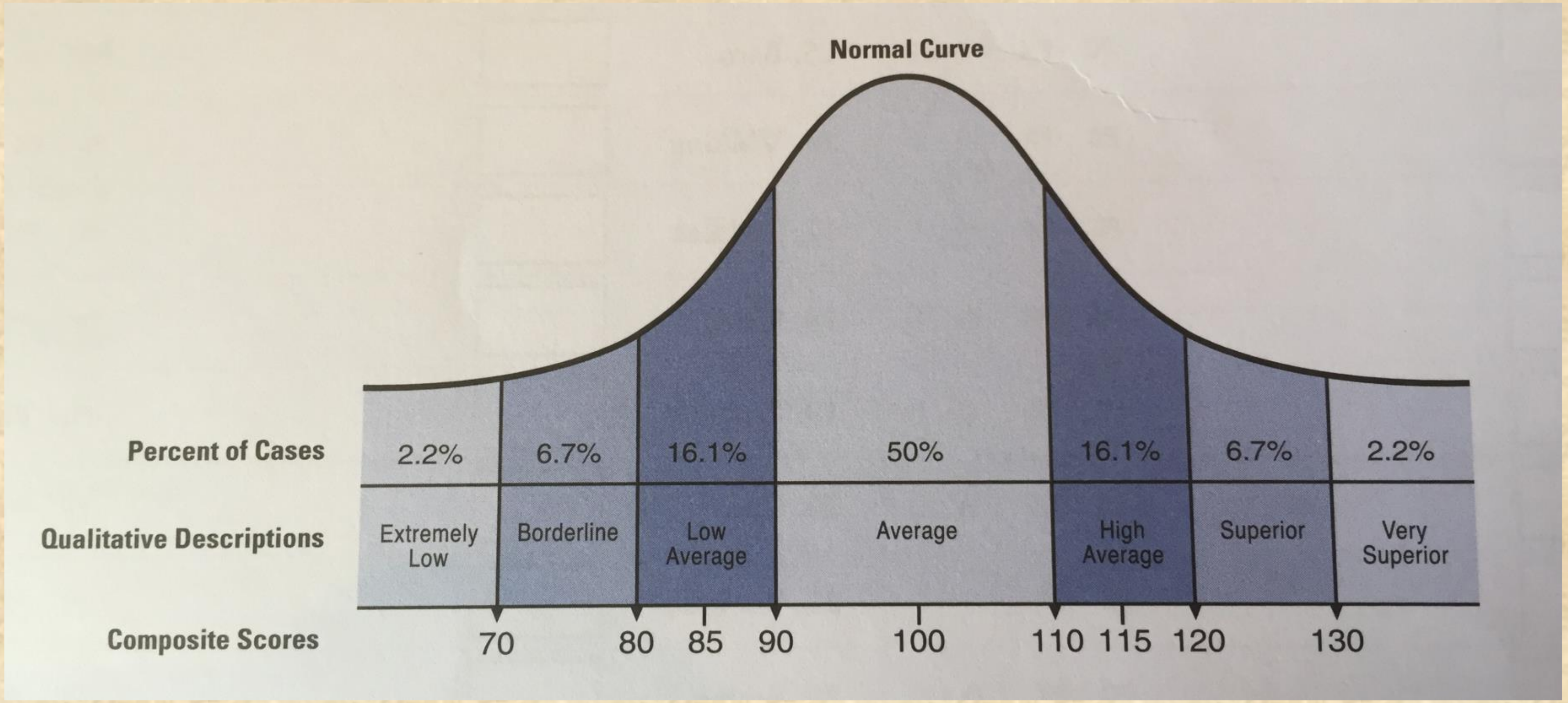
This is usually achieved by consideration of an individual ability in comparison to their peers



Traditional method of interpretation

- Premorbid v’s current intellect
- Pattern analysis
  - Considering whether low scores ‘fit’ with expectations of pre-accident ability but also if they ‘fit’ with the profile of diagnosis.
  - There is objectivity but also subjectivity.
  - Subjectivity is important in clinical decision making however one must be guarded against biases in interpretation and increasing error.

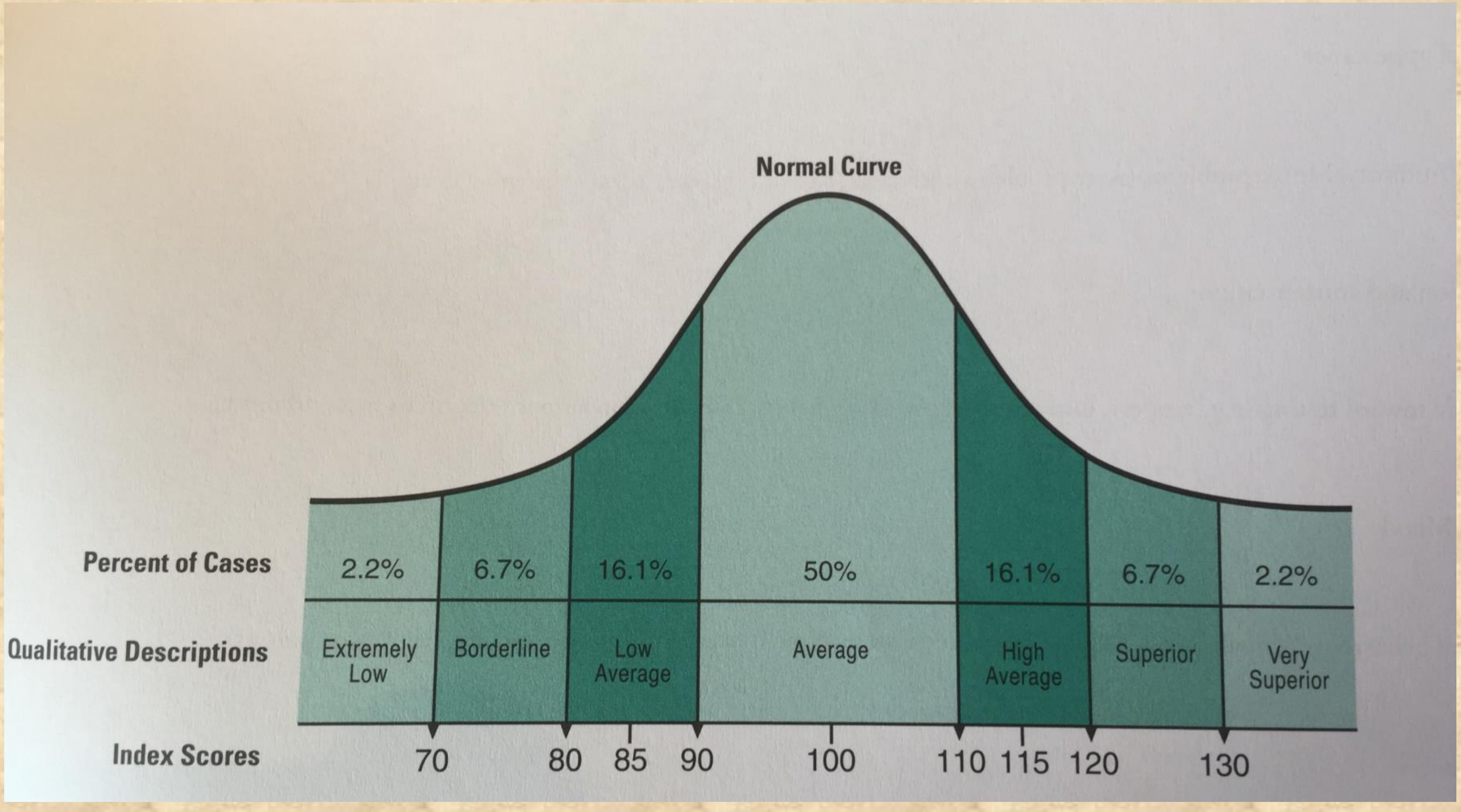
This is usually achieved by consideration of an individual ability in comparison to their peers



Premorbid v's current intellect

TOPF-UK			
Scale	*Obtained Score**	*Estimated Premorbid Score**	*Discrepancy*
Full Scale IQ (FSIQ)	65	93	-28
Verbal Comprehension Index (VCI)	70	92	-22
Perceptual Reasoning Index (PRI)	73	96	-23
Working Memory Index (WMI)	80	94	-14
Processing Speed Index (PSI)	59	96	-37

* Numbers are rounded to wholes for ease of presentation


Normal Curve
Percent of Cases
2.2%
6.7%
16.1%
50%
16.1%
6.7%
2.2%
Qualitative Descriptions
Extremely Low
Borderline
Low Average
Average
High Average
Superior
Very Superior
Index Scores
70
80
85
90
100
110
115
120
130

Base rates tell us how common that difference is allowing a clearer determination of the clinical significance of that discrepancy

TOPF-UK				
Scale	*Obtained Score**	*Estimated Premorbid Score**	*Discrepancy*	*Percentage of normative population with a lower discrepancy score*
Full Scale IQ (FSIQ)	65	93	-28	0.1
Verbal Comprehension Index (VCI)	70	92	-22	1.0
Perceptual Reasoning Index (PRI)	73	96	-23	3.7
Working Memory Index (WMI)	80	94	-14	9.1
Processing Speed Index (PSI)	59	96	-37	0.1

* Numbers are rounded to wholes for ease of presentation

As well as how common the differences between broad cognitive domains is in the general population

WAIS IV Index Level Discrepancy Comparisons

Comparison	*Score 1*	*Score 2*	*Difference*	*Critical Value .05*	*Significant Difference*	*Base Rate*
VCI - PRI	70	73	-3	7.78	N	50.8
VCI – WMI	70	80	-10	8.31	Y	21.7
VCI – PSI	70	59	11	11.76	N	13.8
PRI – WMI	73	80	-7	8.81	N	25.4
PRI – PSI	73	59	14	12.12	Y	12.7
WMI – PSI	80	59	21	12.47	Y	4.8
FSIQ - GAI	65	69	-4	3.29	Y	21.1

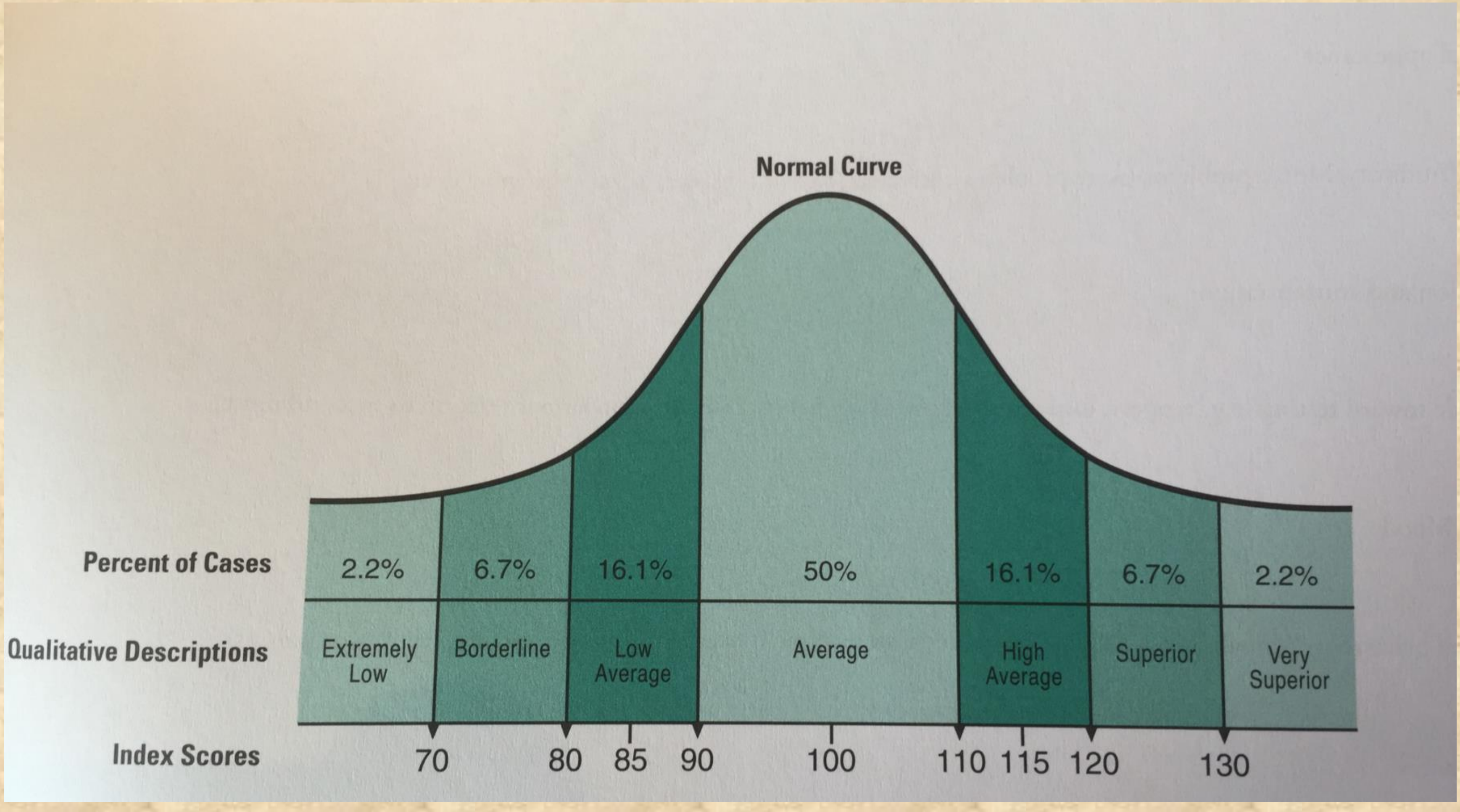
Subjective determinations at the subtest level

TOPF-UK				
Scale	*Obtained Score**	*Estimated Premorbid Score**	*Discrepancy*	*Percentage of normative population with a lower discrepancy score*
Full Scale IQ (FSIQ)	65	93	-28	0.1
Verbal Comprehension Index (VCI)	70	92	-22	1.0
Perceptual Reasoning Index (PRI)	73	96	-23	3.7
Working Memory Index (WMI)	80	94	-14	9.1
Processing Speed Index (PSI)	59	96	-37	0.1

* Numbers are rounded to wholes for ease of presentation

Relating individual subtest scores to pre-accident expectations

WAIS IV Subtests				
Index	*Subtest*	*Raw Score*	*Scaled Score*	*Percentile Rank*
VCI	Similarities	13	4	2
VCI	Vocabulary	22	6	9
VCI	Information	4	4	2
PRI	Block Design	20	6	9
PRI	Matrix Reasoning	6	4	2
PRI	Visual Puzzles	7	6	9
WMI	Digit Span	21	7	16
WMI	Arithmetic	9	6	9
PSI	Symbol Search	16	4	2
PSI	Coding	18	1	0.1


Normal Curve
Percent of Cases
2.2%
6.7%
16.1%
50%
16.1%
6.7%
2.2%
Qualitative Descriptions
Extremely Low
Borderline
Low Average
Average
High Average
Superior
Very Superior
Index Scores
70
80
85
90
100
110
115
120
130

Increasing subjectivity can lead to increased error

- Overreliance on the 'bell curve' for interpretation of multiple scores can become increasingly subjective which can lead to a gradual, insidious pattern of error in the clinical interpretation of neuropsychological test data.

- The reason is that when interpreting test scores, the clinical inference derived from a percentile rank is an accurate indicator of a person's level of functioning on each test in *isolation* but not when all tests are considered *simultaneously.*

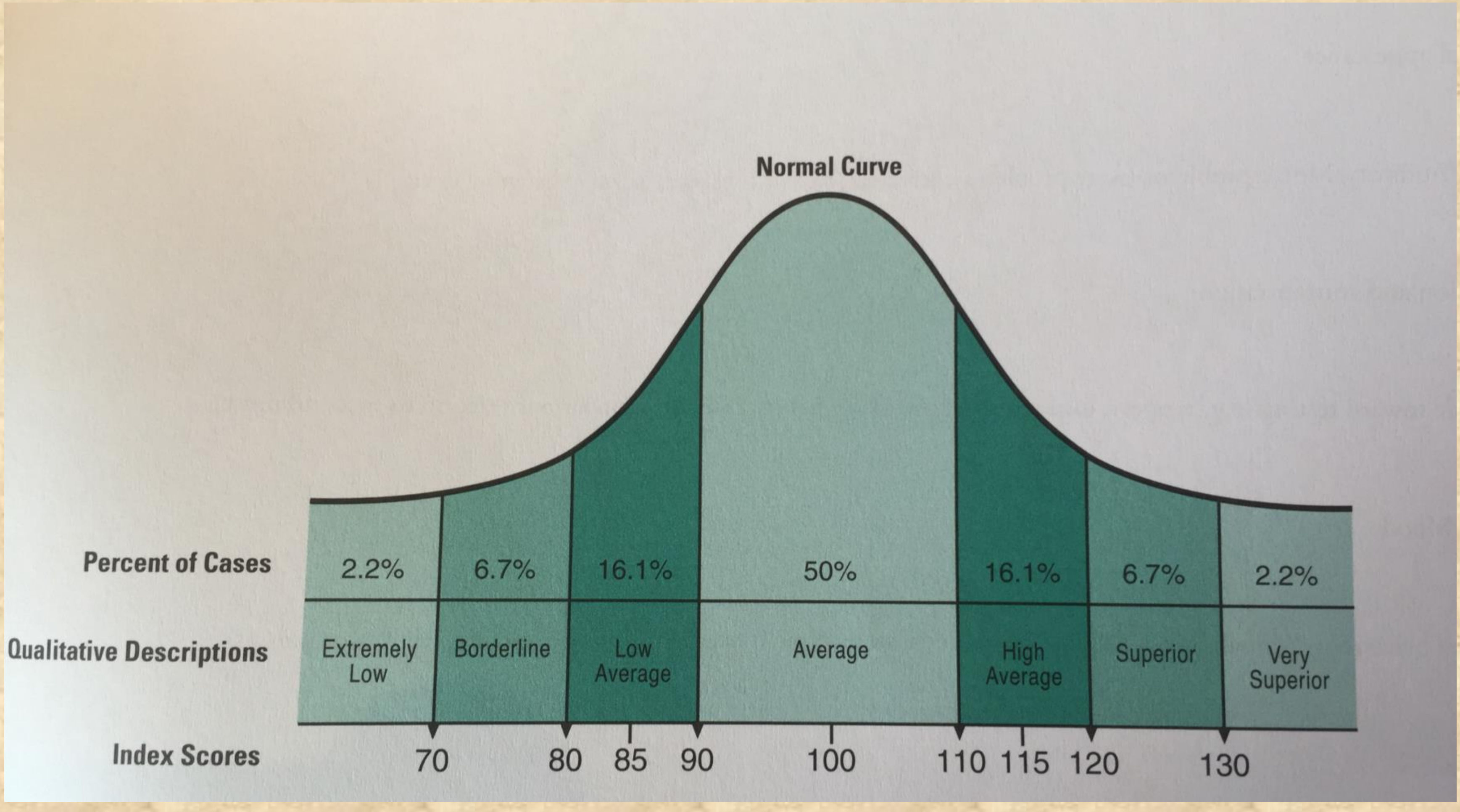
Natural Variation

- A common assumption is that an individual of average (or above average etc..) Intellect must have all their abilities in that range also.
- It is acknowledged that we all have strengths and weaknesses but the breadth of that range is often assumed to be narrower than is actually the case.
- The use of base rate data now allows us to determine more accurately how unusual that profile of low scores is.

Biases and idiosyncratic criteria for impairment

- Clinicians may have idiosyncratic criteria for determining clinical significance.

- Clinicians may have an unconscious bias to subjectively use the same criteria for impairment irrespective of pre-accident ability or education.



Normal Curve
Percent of Cases
2.2%
6.7%
16.1%
50%
16.1%
6.7%
2.2%
Qualitative Descriptions
Extremely Low
Borderline
Low Average
Average
High Average
Superior
Very Superior
Index Scores
70
80
85
90
100
110
115
120
130

Test interpretation, base rates and sources of bias

- Ignoring normal variance in test scores can lead to significant error (e.g., Brooks et. al., 2011, Holdnack et al., 2013).

- A few low scores may simply represent normal variation.

- It is important to consider whether the number of low index and subtest scores are *broadly normal*, *below expected*, or *well below expected* when compared with (a) those with similar level of estimated intelligence and (b) those with a similar number of years of education.

Brooks, Iverson, Holdnack, & Feldman, 2008

- It is commonplace for a person of average ability to have up to 10% of their test scores on intelligence and memory testing to be at or below the 5^{th} percentile, that is, falling in the bottom 5% of the population.

- In healthy older adults from the Wechsler Memory Scale–Third Edition (WMS-III) standardization sample, approximately 42.4% obtained at least one low score (below 9^{th} percentile) when considering all eight scores *simultaneously* .

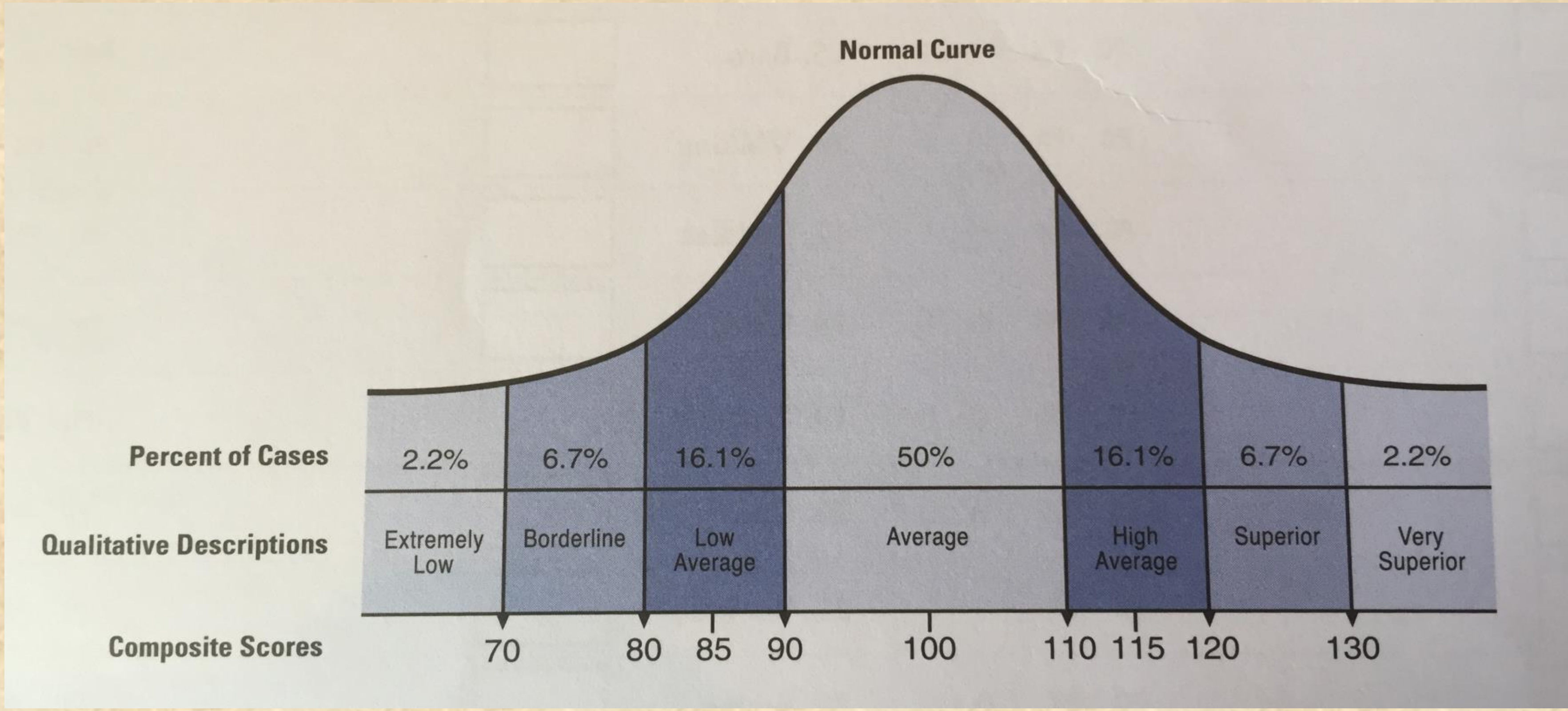
- This means that not taking base rates into account may result in interpreting impairment when there is none.

Increasing studies have come to the same conclusion Base Rate data is increasingly available for different tests

- Prevalence rates of low scores are related to:
  - Test intercorrelations (Crawford, Garthwaite, & Gault, 2007; Ingraham & Aiken, 1996).
  - The number of tests administered (Iverson & Brooks, in press).
  - Demographic variables (Brooks et al., 2008; Schretlen, Testa, Winicki, Pearlson, & Gordon, 2008).
  - Level of intelligence (Binder et al., 2009; Brooks, Iverson, & White, 2007; Brooks, Strauss, et al., 2009).

- Several large batteries of Neuropsychological tests now possess this data
  - The Expanded Halstead– Reitan Neuropsychological Battery (Binder et al., 2009; Heaton, Miller, Taylor, & Grant, 2004).
  - The Neuropsychological Assessment Battery (Brooks et al., 2007; Brooks, Iverson, & White, 2009;Iverson, Brooks, White, & Stern, 2008).
  - WAIS III/ WMS-III (Iverson, Brooks, & Holdnack, 2008).
  - WAIS IV/WMS IV (Brooks, Holdnack & Iverson 2011; Holdnack 2013).
  - DKEFS (Karr et., al. 2016).
  - Children’s memory scale (Brooks et., al. 2009).

How common is the pattern of low scores

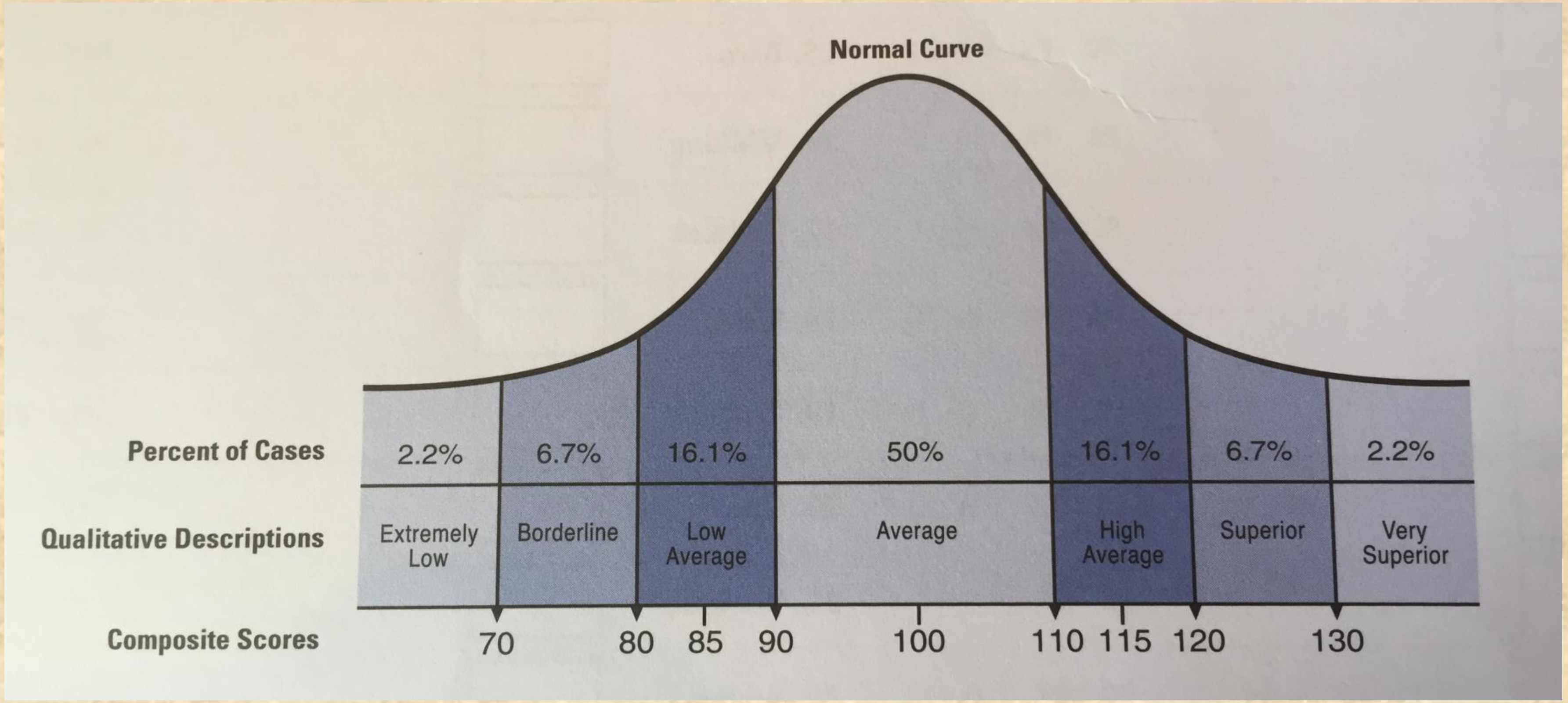


Brooks et. al., (2011) WAIS IV/WMS IV

Table 3. Prevalence of Low WAIS-IV/WMS-IV Subtest Scores

Cutoff Scores and Samples	N	% With 0 Low Scores	% With 1 or More Low Scores	Median Number of Low Scores	Below Expected Number of Low Scores (~10% to 25% of Sample)	Well Below Expected Number of Low Scores (<10% of Sample)
Cutoff: ≤25th percentile (scaled score ≤8)						
Total sample	900	10.9	89.1	5	10-14	15+
Education (years)						
8 or less	41	0.0	100	12	18-19	20
9-11	78	2.6	97.4	11	15-18	19+
12	276	5.1	94.9	6	11-14	15+
13-15	266	10.5	89.5	5	9-12	13+
16+	239	22.6	77.4	2	5-8	9+
TOPF-Demo. FSIQ						
<80	22	0.0	100	17	19	20
80-89	95	1.1	98.9	11	15-17	18+
90-109	478	7.1	92.9	5	10-12	13+
110-119	157	27.4	72.6	1	5-6	7+
120+	22	50.0	50.0	0.5	3	4+

How common is the pattern of low scores

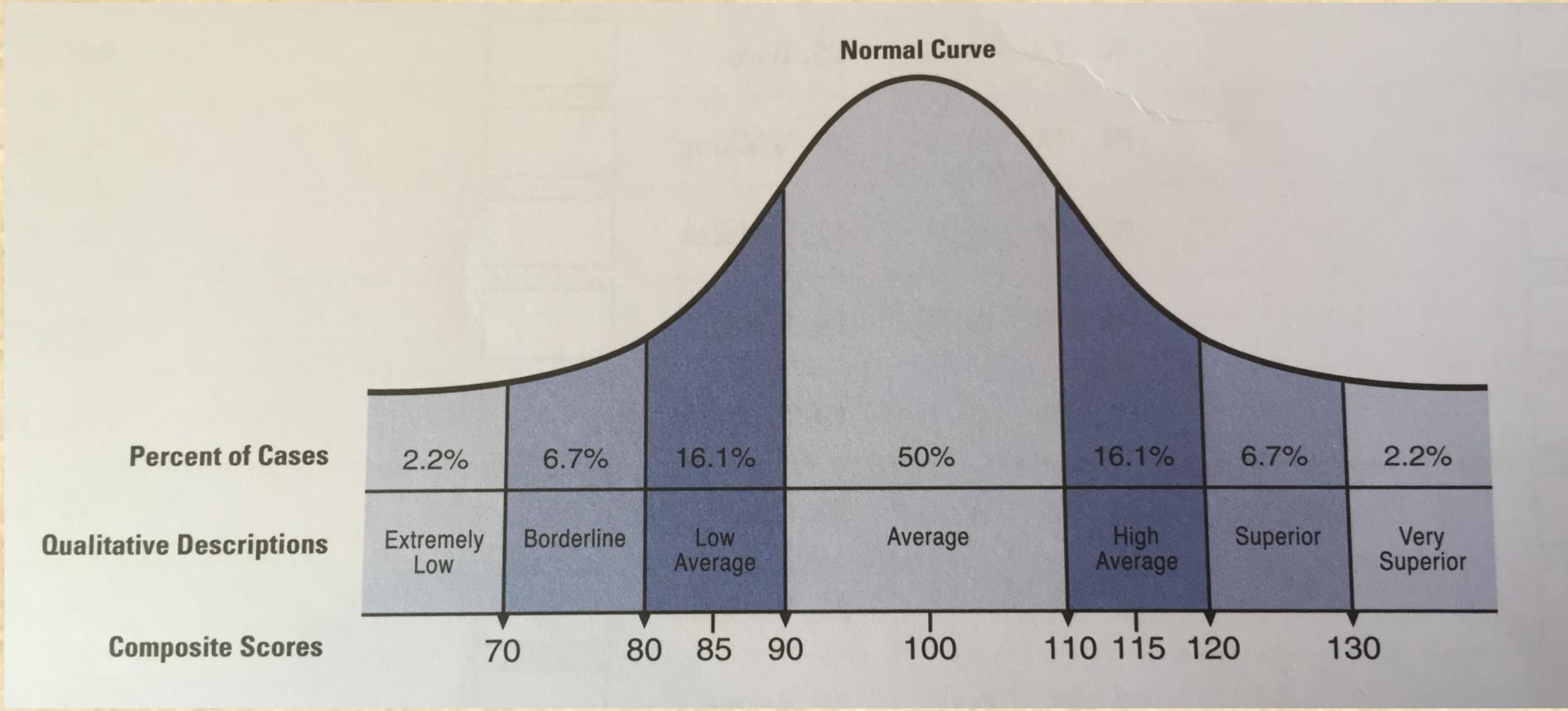


WAIS IV/WMS IV

Table 3. (continued)

Cutoff Scores and Samples	N	% With 0 Low Scores	% With 1 or More Low Scores	Median Number of Low Scores	Below Expected Number of Low Scores (~10% to 25% of Sample)	Well Below Expected Number of Low Scores (<10% of Sample)
TOPF-Demo. FSIQ						
<80	22	0.0	100	6	10-11	12+
80-89	95	28.4	71.6	2	5-8	9+
90-109	478	57.9	42.1	0	2-3	4+
110-119	157	79.6	20.4	0	1	2+
120+	22	86.4	13.6	0	1	2+
Cutoff: ≤2nd percentile (scaled score ≤4)						
Total sample	900	73.7	26.3	0	2	3+
Education (years)						
8 or less	41	34.1	65.9	1	5-6	7+
9-11	78	47.4	52.6	1	3-4	5+
12	276	67.4	32.6	0	2	3+
13-15	266	78.9	21.1	0	1	2+
16+	239	90.4	9.6	0	—	1+
TOPF-Demo. FSIQ						
<80	22	9.1	90.9	3	6-7	8+
80-89	95	48.4	51.6	1	2-4	5+
90-109	478	77.0	23.0	0	1	2+
110-119	157	91.1	8.9	0	—	1+
120+	22	90.9	9.1	0	—	1+

How common is the pattern of low scores



Results of WAIS IV/WMS IV base rate data

- Low scores become more common in those with fewer years of education or lesser intelligence.

- Having one or more Index scores ≤5th percentile is found in 61.0% of those with 8 or fewer years of education but in only 4.6% of those with 16 or more years of education.

- Considering intelligence, it is *well below expected* to have six or more Index scores ≤5th percentile in those with estimated low average intelligence. In contrast, having one or more Index scores ≤5th percentile is *well below expected* in those with estimated high average, superior, or very superior intelligence.

- When considering at or below the 5th percentile as a cut- off score, one or more low subtest scores are found in 44.4% of healthy people.

Karr et, al,. (2017) DKEFS

- The Trail Making Test
  - Condition 4: Number-Letter Switching
- Verbal Fluency Test
  - Condition 1: Letter Fluency – Total Correct
  - Condition 2: Category Fluency – Total Correct
  - Condition 3: Category Switching – Total Correct Responses
  - Condition 4: Category Switching – Total Switching Accuracy),
- Color-Word Interference Test
  - Condition 3: Inhibition – Time-to-completion
  - Condition 4: Inhibition/Switching – Time-to-completion).

DKEFS

Number of low scores	Total sample	WASI FSIQ ≤89	90–99	100–109	110+
Sample size	1,028	160	163	243	241
≤25th percentile					
7 low scores	0.3	1.3	—	—	—
6 or more	2.9	7.5	1.8	1.2	0.4
5 or more	8.4	22.5	6.1	5.3	0.8
4 or more	18.2	44.4	18.4	11.9	2.9
3 or more	32.3	63.8	33.7	25.9	10.8
2 or more	50.9	83.1	58.3	42.0	27.8
1 or more	76.6	95.0	85.9	71.2	60.2
No low scores	23.4	5.0	14.1	28.8	39.8
≤16th percentile					
7 low scores	0.2	1.3	—	—	—
6 or more	1.5	4.4	1.2	0.4	—
5 or more	2.8	9.4	1.2	1.6	0.4
4 or more	8.9	23.1	8.6	4.1	0.4
3 or more	18.7	44.4	17.8	9.9	2.9
2 or more	35.8	67.5	37.4	28.0	14.9
1 or more	62.8	88.1	67.5	58.4	38.6
No low scores	37.2	11.9	32.5	41.6	61.4
≤9th percentile					
7 low scores	—	—	—	—	—
6 or more	0.5	1.3	—	—	—
5 or more	1.6	5.6	0.6	0.8	—
4 or more	5.2	15.0	4.9	2.1	—
3 or more	11.4	28.8	12.3	5.3	1.2
2 or more	24.4	51.9	25.2	18.5	6.6
1 or more	48.1	78.1	51.5	41.6	23.7
No low scores	51.9	21.9	48.5	58.4	76.3
≤5th percentile					
7 low scores	—	—	—	—	—
6 or more	0.1	—	—	—	—
5 or more	0.3	—	—	—	—
4 or more	1.9	6.3	1.2	0.8	—
3 or more	6.0	16.3	7.4	2.5	—
2 or more	14.0	33.8	13.5	7.8	2.5
1 or more	36.1	66.9	41.1	25.5	16.2
No low scores	63.9	33.1	58.9	74.5	83.8

Base Rates I

- Base rate information is therefore a more psychometrically robust method for establishing the significance of low cognitive abilities.
- Clinicians can use the base rates of low scores in healthy people to reduce the likelihood of misdiagnosing cognitive impairment.
- Healthy people obtain some low scores with prevalence rates increasing with fewer years of education and lower predicted intelligence.

Base Rates II

- Take account of pre-accident ability.
- Take account of years of education.
- Differentiate between ‘relatively uncommon’ (the bottom 25%) or ‘unusual’ (bottom 10%).
- ‘Unusual’ would routinely be taken as indicating clinically significant impairment and ‘relatively uncommon’ may indicate a specific problem or significant suppression/inefficiency.

However clinical opinion and context remain important

- The presence of more low scores than would be expected in healthy people is not diagnostic of a specific clinical condition (eg dementia or a TBI); rather, it indicates that the number of low scores obtained is atypical in healthy people.

- It remains incumbent on the clinician to determine if the low scores reflect the consequences of injury or disease or if other factors (e.g., level of effort, medications, other medical conditions, mental health or normal variability) account for the number of obtained low scores.

This is relevant to both Defendant and Claiment solicitors

- It prevents misinterpretation of natural variability as impairment BUT equally prevents genuine impairment being missed.

- It contributes to the minimisation of subjectivity, erroneous bias and differing criteria for ‘impairment’ and provides a framework and criteria for determining impairment

- It guards against unconscious biases in interpretation

- When applying the base rates data to a clinical sample, patients with moderate-to-severe TBIs are 13 times more likely to be identified as having a low cognitive profile compared with healthy controls

Key references

- Brooks et. al., Advanced Clinical Interpretation of the WAIS-IV and WMS-IV: Prevalence of Low Scores Varies by Level of Intelligence and Years of Education. Assessment 18 (1) 2011

- Holdnack et. al., WAIS IV, WMS IV and ACS: Advanced clinical interpretation. Academic Press, 2013.

- Karr et. al., Using Multivariate Base Rates to Interpret Low Scores on an Abbreviated Battery of the Delis–Kaplan Executive Function System. Archives of Clinical Neuropsychology 32 (297-305) 2017.

- Richards et. al., The Dirty Dozen: 12 Sources of Bias in Forensic Neuropsychology with Ways to Mitigate. Psychol. Inj. and Law (2015) 8:265–280

Brain Injury Claims
van den Broek & Sembi. Sweet and Maxwell 2017

GENERAL EDITORS: Dr Martin D van den Broek and Dr Sundeep Sembi

---

Brain Injury Claims examines the legal and clinical aspects of brain injury litigation with a particular emphasis on exploring the mechanisms of traumatic brain injuries, the neurological and neuropsychological sequelae, treatment and rehabilitation, and employment and care issues. The text reviews the central question of judicial reasoning when considering evidence, and incorporates illustrative case studies. Brain Injury Claims is a key text in civil litigation for both claimant and defendant lawyers and all of those involved in compensation claims.

Summary I

- The process of giving multiple tests results in a pattern of variation that is a reflection of strength and weakness in abilities.
- Clinicians must guard against the incorrect clinical inference that obtaining a low test score is necessarily unusual when multiple tests are considered simultaneously.
- But equally must place the results in context and consider the overall clinical picture.
- The use of Base Rates does however contribute to the minimisation of subjectivity, erroneous bias and differing criteria for ‘impairment’ and provides a framework and criteria for determining impairment.

Summary II

- The base rates of low scores vary by the cutoff score used (e.g., 16th percentile vs. the 2nd percentile), demographic characteristics (e.g., years of formal education), and predicted intellectual functioning (e.g., TOPF-FSIQ scores).

- Low scores are very common in healthy people with low formal educational attainment (e.g., 8 or fewer years of education) and low estimated intellectual functioning (e.g., TOPF-FSIQ < 80).

- Therefore, obtaining multiple low scores may or may not be atypical, so the clinician must consider the prevalence rates relative to education or predicted intelligence to determine whether the findings are uncommon.

- Interpretation of low scores as indicating a loss in cognitive functioning secondary to an injury or disease must take into account whether non-injured/ healthy people would also obtain the same number of low scores.

Summary III

- However having several low scores across the WAIS-IV and WMS-IV is uncommon in high functioning people.

- Similar to the concern expressed about misdiagnosing cognitive problems based on a few isolated low scores, there should also be concern about failing to detect impairment in those persons who have many years of education and/or high intelligence.

- For example, having one or more Index scores 1 *SD* below the mean (i.e., 16th percentile), which might be considered a fairly liberal cut- off score for identifying cognitive problems, is uncommon in healthy people with predicted superior intelligence.

- If a clinician applies the same cutoff score from low functioning patients to those who are high functioning, then there is an increased likelihood of making an assumption that the obtained profile across the battery of tests is broadly normal in a high functioning person.