

Rating Scales and Safety Measurements in Bipolar Disorder and Schizophrenia – *A Reference Guide*

INTRODUCTION

Psychopharmacology Bulletin and its sister journal, *Schizophrenia Bulletin*, were published by the National Institute of Mental Health until they were acquired by private publishers in 1999 and 2004, respectively. Both journals—which predate the *Diagnostic and Statistical Manuals of Mental Disorders* (e.g., DSM-I)—have a long and storied history of publishing rating scales and psychiatric guidelines for researchers and clinicians. For example, both the Q-LES-Q-SF¹ and the PANSS², originally published in the “Bulletins” are represented within this guide.

The goal of this reference guide is to assist mental health experts in understanding valid and reliable³ rating scales in schizophrenia and bipolar disorder. The end result, we hope, is to provide the clinician with a valuable tool to evaluate the use of psychotropics for the treatment of these two diseases. — The Editors

A Note on The Rating Scales

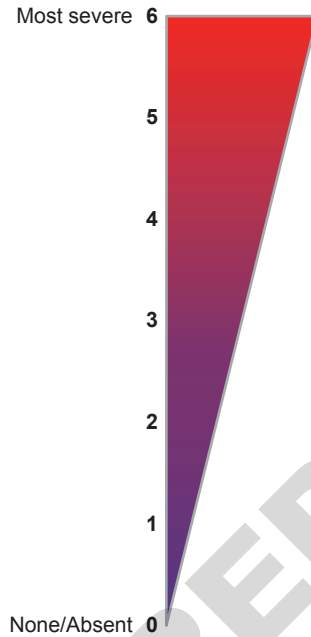
The rating scales described in this booklet have been tested for validity and reliability in clinical trials. In this context, validity refers to the clinical appropriateness of the measure and how adequately the questions reflect the aims that were specified within the scope of the evaluation. Reliability describes the ability of the scale to convey consistent and reproducible information.

Armed with this information, the healthcare professional may be better able to understand and evaluate the clinical evidence for the use of medications for the treatment of bipolar disorder and schizophrenia.

Presented at the American Psychiatric Association Annual Meeting, May 20–24, 2017, San Diego, CA., and was originally sponsored by Sunovion Pharmaceuticals Inc., a U.S. subsidiary of Daiippon Sumitomo Pharma Co., Ltd.

To whom correspondence should be addressed: James M. La Rossa Jr., Publisher, 2205 Rockefeller Lane, Bldg. B, Redondo Beach, CA 90278. Phone: 310.374.1300; Fax: 424.398.0067; E-mail: editorial@medworksmmedia.com

MONTGOMERY-ÅSBERG DEPRESSION RATING SCALE (MADRS)⁴



78

Items

1. Apparent sadness
2. Reported sadness
3. Inner tension
4. Reduced sleep
5. Reduced appetite
6. Concentration difficulties
7. Lassitude
8. Inability to feel
9. Pessimistic thoughts
10. Suicidal thoughts

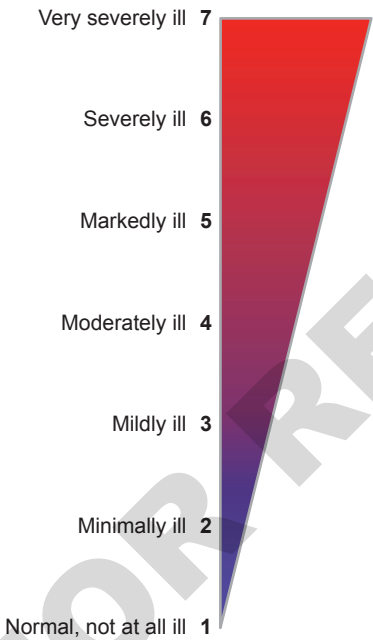
Notes

- Consists of 10 items that represent core emotional and depressive symptoms
- Items are rated from 0–6
- Clinician rated
- Frequently used in clinical trials
- **Total score range:** 0 (none/absent) to 60 (most severe)

CLINICAL GLOBAL IMPRESSION—BIPOLAR VERSION—SEVERITY OF ILLNESS (CGI-BP-S)⁵

Severity of Illness

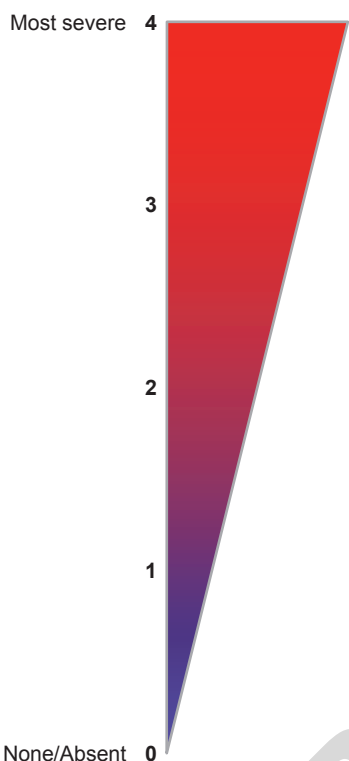
Considering your total clinical experience with patients with bipolar disorder, how severely ill has the patient been during the assessment period?



Notes

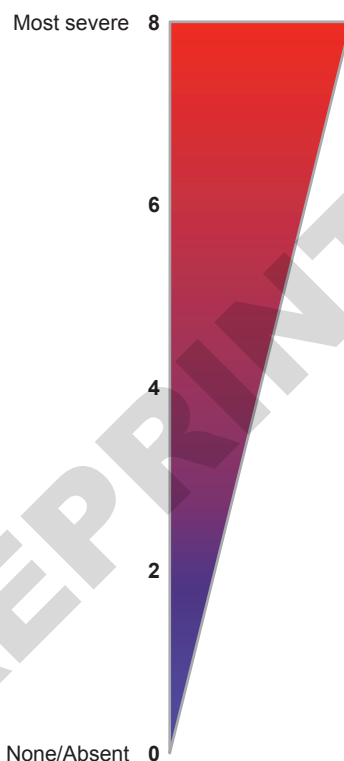
- A modification of the Clinical Global Impressions (CGI) scale, designed to measure the severity of bipolar disorder
- In applying this scale, the rater is asked to draw upon his or her clinical experience and compare the patient with other patients with bipolar disorder
- Measures manic and depressive episodes

YOUNG MANIA RATING SCALE (YMRS)⁶



Seven items rated on a scale of 0 to 4:

1. Elevated mood
2. Increased motor activity/energy
3. Sexual interest
4. Reduction in sleep
7. Language/thought disorder
8. Poor appearance
9. Lack of insight



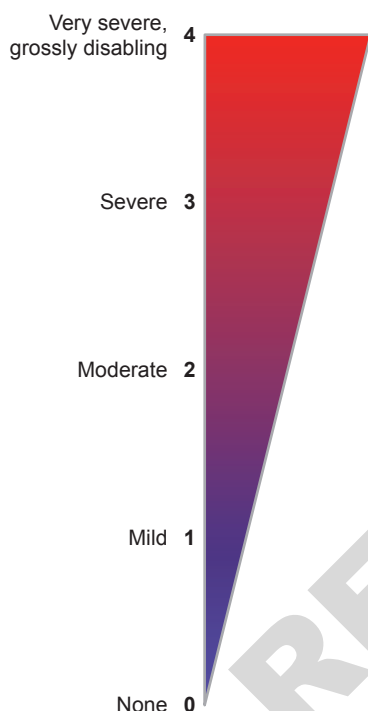
Four items rated on a scale of 0 to 8:

5. Irritability
6. Rate and amount of speech
7. Thought content
8. Disruptive/aggressive behavior

Notes

- 11-item scale designed to evaluate severity of manic symptoms
- Clinician rated
- Most frequently used scale for mania
- **Total score range:** 0 (none) to 60 (most severe)

HAMILTON RATING SCALE FOR ANXIETY (HAM-A)⁷



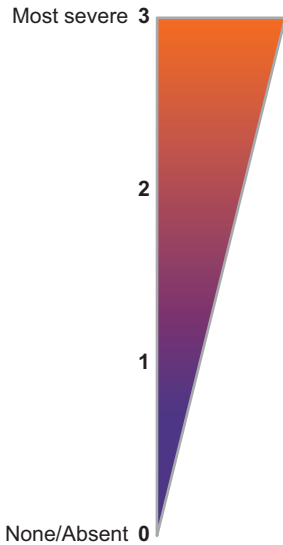
81

1. Anxious mood
2. Tension
3. Fears
4. Insomnia
5. Intellect
6. Depressed mood
7. Somatic general (muscular)
8. Somatic general (sensory)
9. Cardiovascular system
10. Respiratory system
11. Gastrointestinal system
12. Genitourinary system
13. Autonomic system
14. Behavior at interview

Notes

- 14-item scale relies on patient report
- Each item rated 0–4
- Most frequently used scale for anxiety
- **Total score range:** 0 (none) to 56 (most severe)

QUICK INVENTORY OF DEPRESSIVE SYMPTATOLOGY – SELF-REPORT (QIDS-SR₁₆)⁸



Most severe 3

2

1

None/Absent 0

Items

Four items pertaining to sleep disturbances

1. Falling Asleep
2. Sleeping During the Night
3. Waking Up Too Early
4. Sleeping Too Much

Enter the highest individual score from items 1 to 4

5. Feeling Sad

Four items pertaining to weight/appetite

6. Decreased Appetite
7. Increased Appetite
8. Decreased Weight (Within the Last Two Weeks)
9. Increased Weight (Within the Last Two Weeks)

Enter the highest individual score from items 6 to 9

10. Concentration/Decision Making
11. View of Myself
12. Thoughts of Death or Suicide
13. General Interest
14. Energy Level

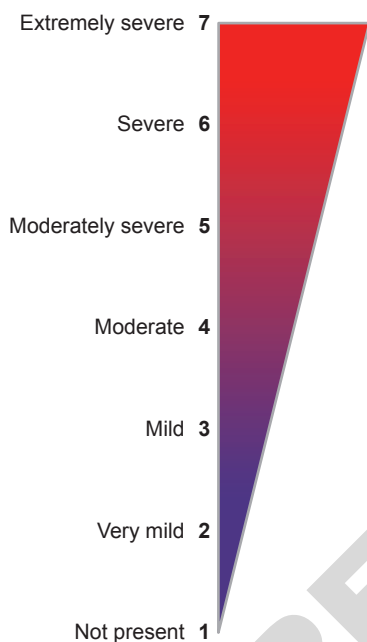
Two items pertaining to psychomotor disturbances

15. Feeling Slowed Down
16. Feeling Restless

Enter the highest individual score from items 15 to 16

Notes

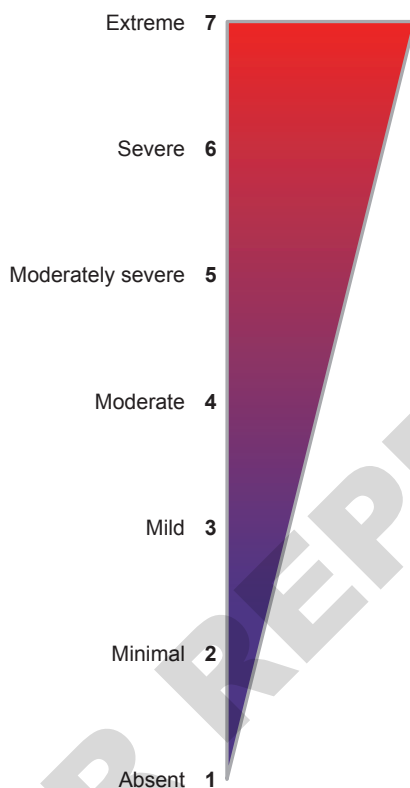
- 16 separate self-reported items that correspond to the 9 core symptom domains of DSM-IV MDD
- Patient is asked to select the score for each item that best describes him/her for the past 7 days.
- **Total score range:** 0 (none) to 27 (most severe)

BRIEF PSYCHIATRIC RATING SCALE (BPRS)⁹**83***Symptom Constructs*

- | | |
|-------------------------------|-----------------------------|
| 1. Somatic concern | 9. Depressive mood |
| 2. Anxiety | 10. Hostility |
| 3. Emotional withdrawal | 11. Suspiciousness |
| 4. Conceptual disorganization | 12. Hallucinatory behavior |
| 5. Feelings of guilt | 13. Motor retardation |
| 6. Tension | 14. Uncooperativeness |
| 7. Mannerisms and posturing | 15. Unusual thought content |
| 8. Grandiosity | 16. Blunted affect |

Notes

- Widely used, relatively brief scale that measures major psychotic and nonpsychotic symptoms
- Consists of 16 symptom constructs, or items
- Each item is rated on a 7-point scale by a psychiatrist, a psychologist, or other trained rater
- **Total score range:** 16 (not present) to 112 (extremely severe)

POSITIVE AND NEGATIVE SYNDROME SCALE (PANSS)^{2,10}

84

Positive Scale

1. Delusions
2. Conceptual disorganization
3. Hallucinatory behavior
4. Excitement
5. Grandiosity
6. Suspiciousness
7. Hostility

Negative Scale

1. Blunted affect
2. Emotional withdrawal
3. Poor rapport
4. Passive-apathetic social withdrawal
5. Difficulty in abstract thinking
6. Lack of spontaneity & flow of conversation
7. Stereotyped thinking

General Psychopathology Scale

1. Somatic concern
2. Anxiety
3. Feelings of guilt
4. Tension
5. Mannerisms & posturing
6. Depression
7. Motor retardation
8. Uncooperativeness
9. Unusual thought content
10. Disorientation
11. Poor attention
12. Lack of judgment & insight
13. Disturbance of volition
14. Poor impulse control
15. Preoccupation
16. Active social avoidance

85**Notes**

- The PANSS is an adaptation of 2 earlier psychopathology scales: the BPRS and the Psychopathology Rating Scale (PRS)
- Consists of a clinical interview with the patient and any available supporting information (such as that from family members or hospital staff)
- The 30 items are symptoms associated with schizophrenia
- The severity of each symptom is rated by an interviewer, usually a clinician or other qualified professional
- **Total score range:** 30 to 210
 - 58 = Mildly ill
 - 75 = Moderately ill
 - 95 = Markedly ill
 - 116 = Severely ill

BRIEF PSYCHIATRIC RATING SCALE – DERIVED (BPRSd)

Items of the BPRS are embedded in the PANSS; therefore, the BPRS can be scored from the PANSS interview. This is described as the Brief Psychiatric Rating Scale—derived, or BPRSd.

CLINICAL GLOBAL IMPRESSIONS – SEVERITY OF ILLNESS (CGI-S)¹¹**Severity of Illness**

Considering your total clinical experience with this particular population, how mentally ill is the patient at this time?

Among the most extremely ill patients 7

Severely ill 6

Markedly ill 5

Moderately ill 4

Mildly ill 3

Borderline mentally ill 2

Normal, not at all ill 1

Not assessed 0

86

Notes

- Developed at the National Institute of Mental Health
- One of the most widely used brief assessment tools in psychiatry
- A rating of 1 is normal and 7 indicates the patient falls among the most extremely ill
- In applying this scale, the rater is asked to draw upon his or her clinical experience and, on a scale of 1 to 7, compare the patient with other patients with the same diagnosis

SHEEHAN DISABILITY SCALE (SDS)¹²**WORK**

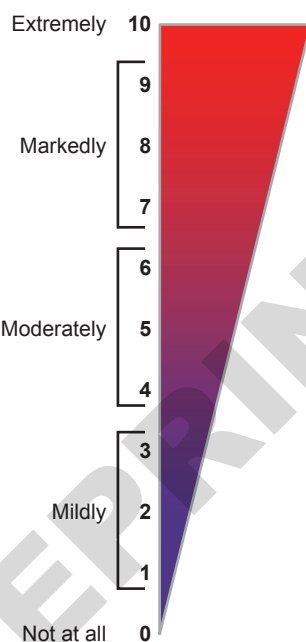
The symptoms have disrupted your work...

SOCIAL LIFE

The symptoms have disrupted your social life...

FAMILY LIFE

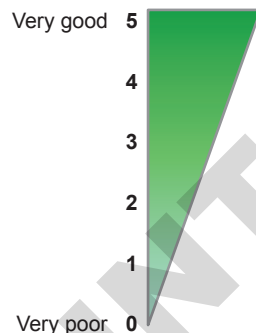
The symptoms have disrupted your family life/home responsibilities...

**Notes**

- Self-report measure; patient asked to rate how their symptoms have disrupted their WORK, SOCIAL LIFE, and FAMILY LIFE
- Each domain is rated on a scale from 0–10
- Very brief and simple
- Has demonstrated sensitivity to the effects of treatment

QUALITY OF LIFE ENJOYMENT AND SATISFACTION QUESTIONNAIRE – SHORT FORM (Q-LES-Q-SF)¹

*Taking everything into consideration,
during the past week how satisfied have
you been with your...*



88

...physical health?
...mood?
...work?
...household activities?
...social relationships?
...family relationships?
...leisure-time activities?
...ability to function in daily life?
...sexual drive, interest, and/or
performance?
...economic status?
...living/housing situation?
...ability to get around physically
without feeling dizzy or falling?
...your vision in terms of ability
to do work or hobbies?
...overall sense of well-being?

Summed to yield
raw score (range
of 14 to 70)

...medication? (If not taking any,
leave blank)

How would you rate your overall
life satisfaction and contentment
during the past week?

Standalone items,
not included in
total score

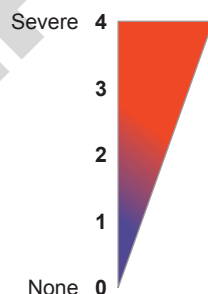
Notes

- The full version (Q-LES-Q) consists of 93 items grouped into 8 summary scales
- The short form of this questionnaire, known as the Q-LES-Q-SF, consists of 16 items
- Each item is rated on scale of 1 to 5, with 1 signifying “very poor” and 5 signifying “very good”
- The first 14 items are summed to yield a raw score, ranging from 14 to 70; **the raw total score is converted into a percentage maximum possible score, ranging from 0% to 100%**
- The last 2 items stand alone and are not included in the total score
- Higher scores on the Q-LES-Q-SF indicate better quality of life as perceived by the patient

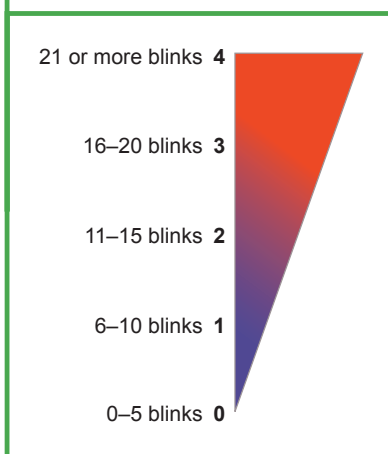
SIMPSON-ANGUS SCALE (SAS)^{13,14}

1. Gait
2. Arm dropping
3. Shoulder shaking
4. Elbow rigidity
5. Fixation of position or wrist rigidity
6. Leg pendulousness (ability to swing freely in a hanging position)
7. Head dropping
8. Glabella* tap
9. Tremor
10. Salivation

*The region between the eyebrows and above the nose.



89



Notes

- Ten-item instrument used to evaluate symptoms of parkinsonism related to the use of antipsychotic medications
- The eighth item on the scale, glabella tap, is evaluated after the clinician administers the following test: the patient is told to open his or her eyes and not blink; the clinician then taps at a steady, rapid speed on the region between the patient's eyebrows and above the nose; the number of blinks in succession is noted and the rating is applied, with higher ratings corresponding to higher blink counts
- The global score on the SAS is the sum of all scores divided by the total number of items (10); final scores of up to 0.3 are considered within the normal range
- **Global score range:** 0 to 4

BARNES AKATHISIA RATING SCALE (BAS, BARS)¹⁵

90

1. Objective

- 0 = normal, occasional fidgety movement of limbs
- 1 = presence of characteristic restless movements
- 2 = observed phenomena, present for at least half the observation period
- 3 = constantly engaged in characteristic restless movements

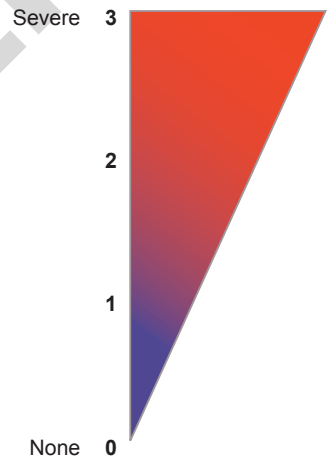
2. Subjective

Awareness of Restlessness

- 0 = absence
- 1 = nonspecific sense
- 2 = aware and/or complains of inner restlessness aggravated specifically by being required to stand still
- 3 = awareness of intense compulsion to move most of the time and/or reports strong desire to walk or pace most of the time

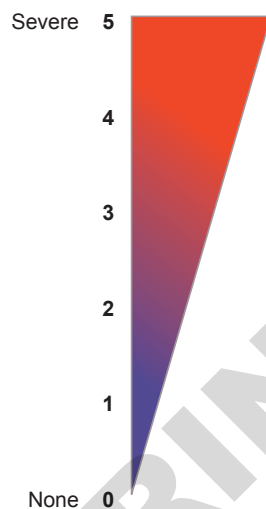
Distress Related to Restlessness

- 0 = no distress
- 1 = mild
- 2 = moderate
- 3 = severe



3. Global Clinical Assessment of Akathisia

- 0 = absent
- 1 = questionable
- 2 = mild
- 3 = moderate
- 4 = marked
- 5 = severe



Notes

- Evaluates the severity of akathisia, which is characterized by a feeling of inner restlessness and the urge to move the limbs, especially the legs
- Rated by a clinician following observation and interview of the patient
- Three components of the scale
 - The *Objective* component assesses on a scale of 0 to 3 the visible behavior of the patient during examination; a rating of 0 signifies normal behavior with only occasional fidgety movement, while 3 indicates that the patient is constantly engaged in restless movements
 - The *Subjective* component is elicited by direct questioning of the patient and assesses the patient's awareness of, and distress arising from, the akathisia; ratings are from 0 to 3, with 3 indicating greater awareness or distress
 - The *Global Clinical Assessment* provides an overall evaluation of the severity of akathisia, rated on a scale of 0 to 5; a score of 0 corresponds to normal or absent, while 5 indicates severe akathisia

ABNORMAL INVOLUNTARY MOVEMENT SCALE (AIMS)^{11,16}

92

*FACIAL
AND ORAL
MOVEMENTS*

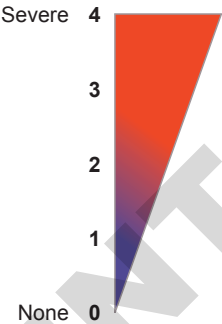
*EXTREMITY
MOVEMENTS*

*TRUNK
MOVEMENTS*

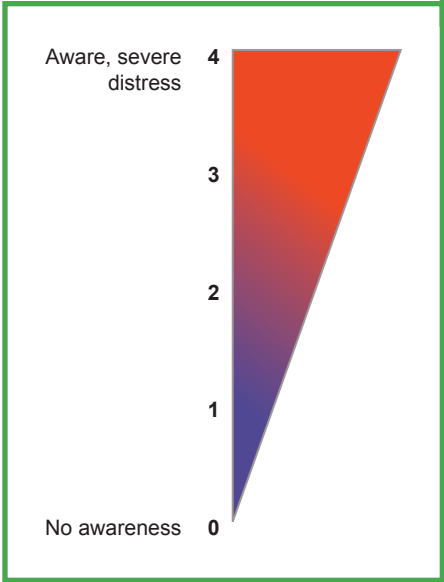
*GLOBAL
JUDGMENTS*

*DENTAL
STATUS*

- 1. Muscles of facial expression
- 2. Lips and perioral area
- 3. Jaw
- 4. Tongue
- 5. Upper (arms, wrists, hands, fingers)
- 6. Lower (legs, knees, ankles, toes)
- 7. Neck, shoulders, hips
- 8. Severity of abnormal movements
- 9. Incapacitation due to abnormal movements
- 10. Patient's awareness of abnormal movements
- 11. Current problems with teeth and/or dentures
- 12. Does patient usually wear dentures?



No (0)/Yes (1)



Notes

- Developed by the National Institute of Mental Health
- Administered by a clinician who conducts an examination during which the patient is asked to perform certain tests of body movement
- The performance on each of these tests is rated on a scale of 0 to 4, with 0 corresponding to normal and 4 corresponding to severe
- Item 10 on the scale evaluates patient awareness of his or her own abnormal movements on a scale of 0 to 4, with 0 indicating no awareness and 4 indicating awareness with severe distress
- Two items relate to dental status and are answered in a yes-no fashion
- Total scores range from 0 to 42; higher total scores on the AIMS correspond to greater severity of dyskinetic movements

COLUMBIA-SUICIDE SEVERITY RATING SCALE (C-SSRS)

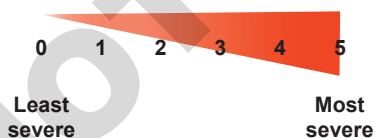
SUICIDAL IDEATION

SEVERITY SUBSCALE

5-point scale (1 = wish to be dead, 5 = suicidal intent with plan)

INTENSITY SUBSCALE

- 5 items (frequency, duration, controllability, deterrents, reason)
- Each item is evaluated on the following scale:



SUICIDAL BEHAVIOR

BEHAVIOR SUBSCALE

For each of the following that apply, record the number of attempts/instances:

- Actual, interrupted, and aborted attempts
- Preparatory acts or behavior
- Nonsuicidal self-injurious behavior
- Each item is rated on a nominal scale

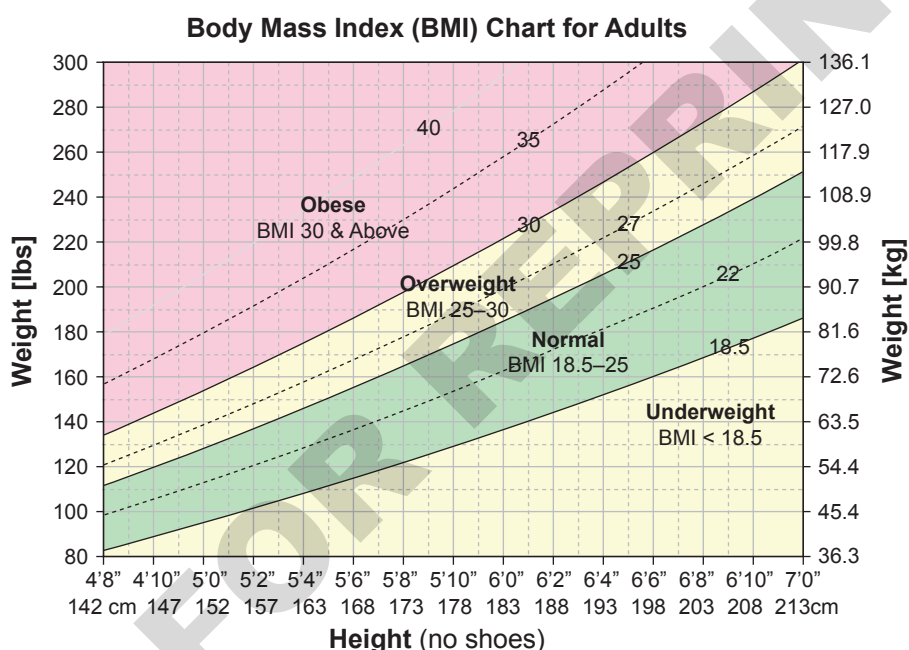
LETHALITY SUBSCALE

Assesses actual attempts and lethality on a 6-point scale (if actual lethality is 0, potential lethality or attempts is rated on a 3-point scale)

Notes

- Designed to assess suicide and distinguish between *ideation* (the capacity for suicidal thoughts) and *behavior* (the propensity for acting on suicidal thoughts).^{17,18}
- Four constructs or subscales; scores are not summed.¹⁸
- Uses different assessment periods, depending on research or clinical need (e.g., lifetime period assesses worst-point ideation).¹⁸

BODY MASS INDEX (BMI)



Body mass index (BMI) is a number calculated from a person's weight and height. It is a fairly reliable indicator of body fat for most people and is considered an alternative for direct measures of body fat.¹⁹

The formula = weight (kg)/[height (m)]

Using pounds and inches = weight (lb)/[height (in)] × 703

>7% increase in weight = 1-point increase in BMI

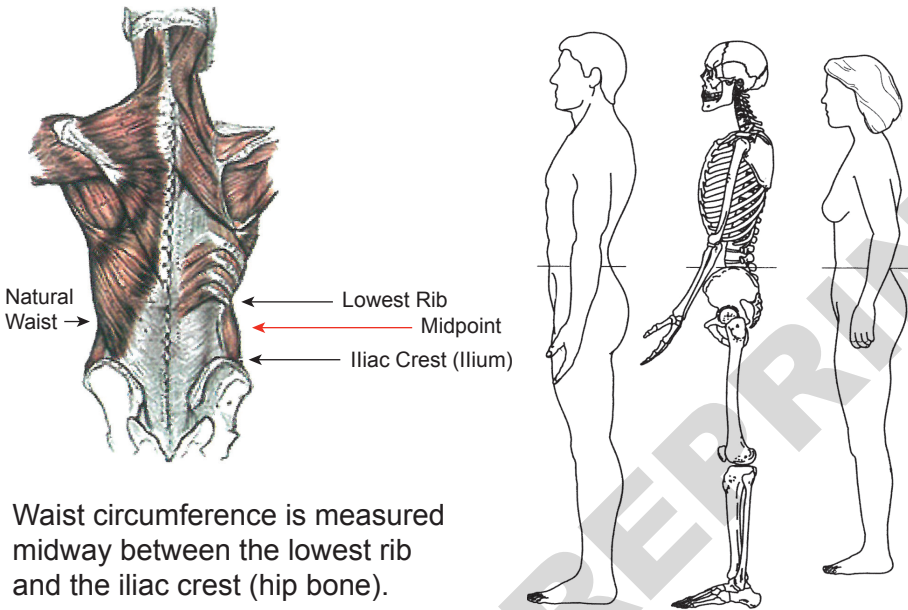
WAIST CIRCUMFERENCE

A high waist circumference and too much abdominal fat is a high risk factor for type 2 diabetes, hypertension, dyslipidemia, metabolic syndrome, and cardiovascular disease.¹⁹

A high-risk waist circumference is:

Men = waist measurement greater than 40 inches (102 cm)

Women = waist measurement greater than 35 inches (88 cm)



Waist circumference is measured midway between the lowest rib and the iliac crest (hip bone).

GLUCOSE, GLYCOSYLATED HEMOGLOBIN (HbA_{1c})

	HbA _{1c} (%)	FASTING PLASMA GLUCOSE (MG/DL)	ORAL GLUCOSE TOLERANCE TEST (MG/DL)
Diabetes	≥6.5	≥126	≥200
Prediabetes	5.7–6.4	100–125	140–199
Normal	<5.7	<100	<140

Manu et al. *J Clin Psychiatry*. 2012;73:460–466.²⁰
American Diabetes Association. *Diabetes Care*. 2013;36(Suppl 1):S13, Tables 2 and 3.²¹

Glycosylated hemoglobin is a form of hemoglobin measured to identify the average plasma glucose concentration over prolonged periods (i.e., 1 to 3 months).²¹

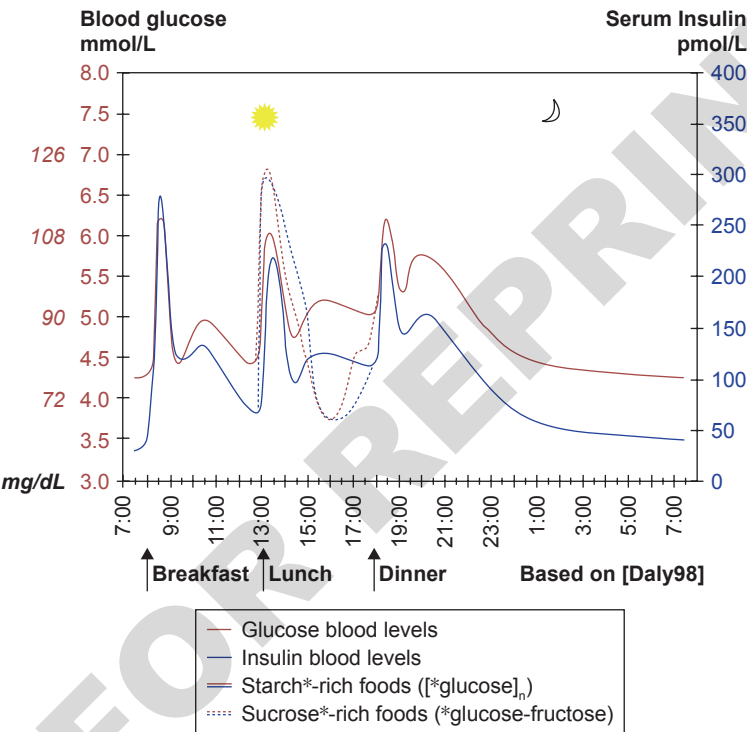
When blood glucose levels rise, glucose molecules attach to the hemoglobin in red blood cells.²¹

The more glucose that binds to red blood cells, the higher the glycosylated hemoglobin.²¹

INSULIN

Insulin helps control blood sugar levels. Insulin resistance describes inefficient use of insulin and is one of the 2 most important risk factors for metabolic syndrome (predictors of coronary artery disease, stroke, and type 2 diabetes) and may be an important predictor of cardiovascular disease.²²

96

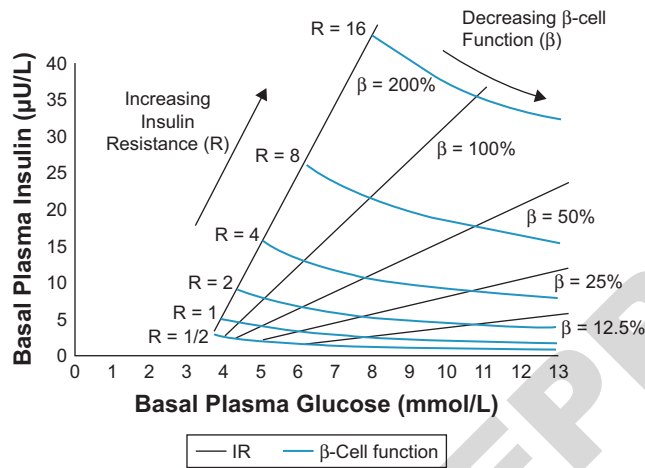


Insulin levels can be impractical to measure in a clinical setting; therefore, assessment is through indirect measures related to blood glucose levels.²⁴

GLUCOSE TOLERANCE TEST WITH INSULIN (GTT/IGTT)

TIME	NORMAL GLUCOSE VALUES (MG/DL)	NORMAL INSULIN VALUES	INSULIN RESISTANCE
Fasting	<126	<10 mIU/mL	>10 mIU/mL
1/2 hour	<200	40–70 mIU/mL	>80 mIU/mL
1 hour	<200	50–90 mIU/mL	5 times fasting level
2 hours	<140	6–50 mIU/mL	>60 IU/mL

The Homeostasis Model Assessment of Insulin Resistance (HOMA-IR) is used in large epidemiological studies and in clinical practice to estimate insulin based on a combination of (β -cell deficiency and insulin resistance).²⁵



HOMA-IR = fasting serum insulin (μ U/mL) \times fasting plasma glucose (mmol/L)/22.5

Metabolic characteristics and lifestyle habits influence HOMA-IR in nondiabetic individuals.²²

Nonlinear correlations between HOMA-IR, metabolic characteristics, and lifestyle habits (Generalized Additive Model analyses, adjusted for age and gender)²²

METABOLIC CHARACTERISTICS	P VALUE	LIFESTYLE HABITS	P VALUE
BMI	<0.001	Alcohol consumption*	<0.001
Waist circumference	<0.001	Physical activity**	<0.001
Triglycerides	<0.001		
Systolic blood pressure	<0.001		

*Negative association in men only; HOMA-IR levels in heavy drinkers (>280 g/wk) were lower than in abstainers.

**Positive association in women only; HOMA-IR levels were lower with intense physical activity than in sedentary women.

CHOLESTEROL (LDL, TOTAL, HDL)

LDL CHOLESTEROL	
<100	Optimal
100–129	Near or above optimal
130–159	Borderline high
160–189	High
≥190	Very high
TOTAL CHOLESTEROL	
<200	Desirable
200–239	Borderline high
≥240	High
HDL CHOLESTEROL CHOLESTEROL	
<40	Low
≥60	High

American Medical Association. *JAMA*. 2001;285:2486–2497.²⁶

98

Cholesterol is a lipid that is transported to and from cells in the blood by carriers called lipoproteins. Low-density lipoprotein, or LDL, is known as “bad” cholesterol. High-density lipoprotein, or HDL, is known as “good” cholesterol. These 2 types of lipids, along with other lipid components, make up the total cholesterol count.²⁶

TRIGLYCERIDES

NCEP GUIDELINE	TRIGLYCERIDE (MG/DL)
Normal	<150
Borderline high	150–199
High	200–499
Very high	>500

American Medical Association. *JAMA*. 2001;285:2486–2497.²⁶

Triglyceride is a form of fat made in the body that can be influenced by diet, cigarette smoking, alcohol consumption, and physical activity. People with high triglyceride levels often have a high total cholesterol level, including a high LDL and low HDL level.²⁶

PROLACTIN

SUBJECT	NORMAL PROLACTIN RANGE (NG/ML OR MCG/L)
Nonpregnant women	4–23
Men	3–15
Pregnant women	34–386

Prolactin levels can vary depending on the time of day they are measured, the age and gender of the patient, and even in response to stress. In normal individuals, prolactin levels can rise in response to²⁷:

- Sleep
- Exercise
- Nipple stimulation
- Sexual intercourse
- Hypoglycemia
- Postpartum period
- Medications, especially CNS dopamine antagonists

99

Upper normal limits of prolactin are considered in the range of 18–23 ng/mL.²⁷

High levels of prolactin are usually ≥ 200 ng/mL and can indicate the presence of a pituitary tumor, excessive pituitary production of prolactin, pregnancy, liver disease, or hypothyroidism.

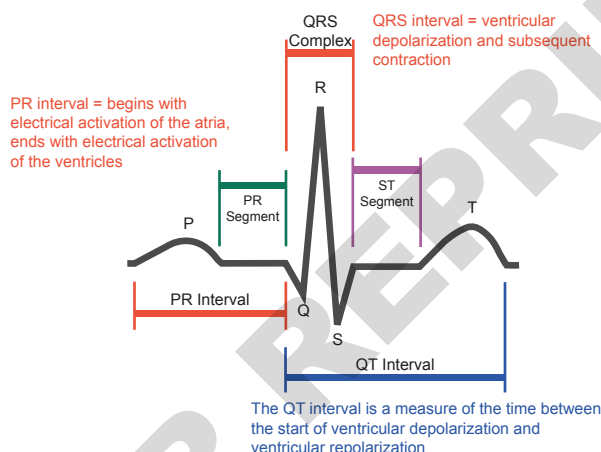
Excess serum prolactin is associated with²⁷:

- hypoenestrogenism (low estrogen levels) anovulatory infertility (lack of ovulation)
- oligomenorrhea (infrequent or very light menstruation)
- amenorrhea (absence of menstruation during reproductive age)
- unexpected lactation
- loss of libido in women
- erectile dysfunction in men

ELECTROCARDIOGRAM (ECG) AND QT

	ECG RESULTS
Normal	Regular rhythm, usually 60–100 bpm Tracing looks normal
Abnormal	Too slow: <60 bpm Too fast: >100 bpm Heart rhythm is not regular Tracing does not look normal

Kadish et al. *J Am Coll Cardiol.* 2001;104:3169–3178.²⁸



The faster the heart rate, the shorter the QT interval.

Evaluation of QT/QTc prolongation is required in drug development, because it can create a physiological environment that predisposes the cardiac muscle (myocardium) to ventricular tachyarrhythmia that can ultimately progress to sudden cardiac death.

Strnadova. *Drug Inf J.* 2005;39:407–433.²⁹

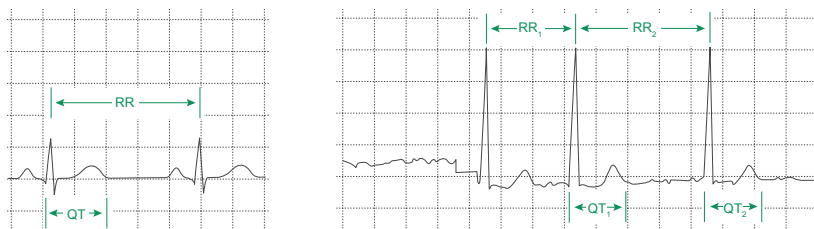
QTc INTERVAL

The QT interval is dependent on the heart rate and may be adjusted to improve the detection of patients at increased risk for ventricular arrhythmia.

Common formulas used to correct for heart rate include:

$$\text{Bazett: } QT_c = \frac{QT}{\sqrt{RR}}$$

$$\text{Atrial Fibrillation: } QT_c = \frac{QT_{c1} + QT_{c2}}{2}$$



A QTc of ≥ 500 milliseconds generally correlates with higher risk of torsades de pointes; however, there is no definitive consensus on the degree of drug-induced QT prolongation that should require drug discontinuation.³⁰

Viskin (2009) proposed the following QT interval scale to aid diagnosis of patients with short and long QT syndromes:

QT scale		
Male		Female
470	Very Long QT LQTS even if asymptomatic. Exclude 2° causes	480
	Long QT LQTS when supported by symptoms, family history, or additional tests.	
450		460
	Long QT possible Additional tests when indicated: Repeated ECG, holter, T-wave morphology, exercise, epinephrine-challenge, adenosine-challenge	
390		400
360	Normal QT	370
	Short QT SQTS when supported by symptoms or family history. Additional test: Repeated ECG, Holter, T-wave morphology, electrophysiologic studies	
330		340
	Very short QT SQTS even when asymptomatic. Exclude 2° causes.	

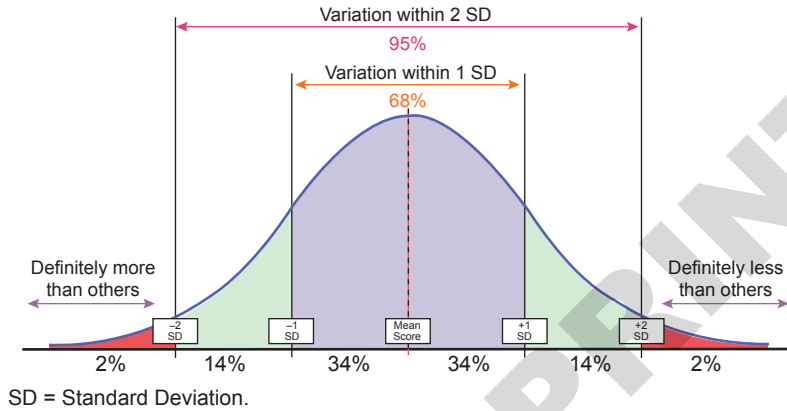
LQTS = long QT syndrome; SQTS = short QT syndrome.
Viskin S. *Heart Rhythm*. 2009;6:711–715.³¹

P VALUES

When comparing measurements between study treatment groups:

- Compare active treatment group with a “control” group like a placebo or other treatment group
- Have at least one measurement/assessment for the comparison(s)
- Use *P* values to explain whether the difference between the groups of interest is due to chance
- *P* value provides a probability of the observed data (e.g., measured difference) when the null hypothesis is true

Within a population, measurements are compared with the “average” or mean. The *P* value quantifies the probability that the result falls within a range of variability in results around the mean, assuming the null hypothesis is true.



102

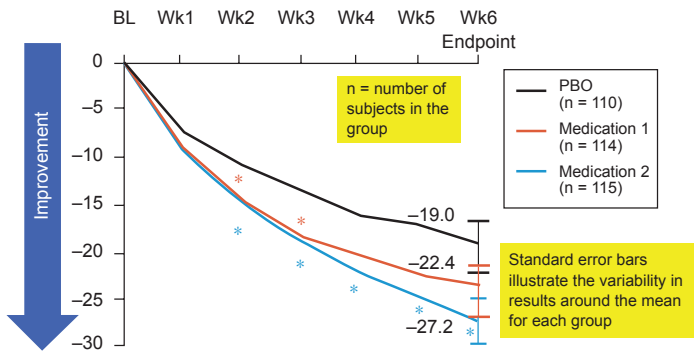
This figure represents a standard bell curve, around which we assume most populations, or results, are distributed. For example, in a class with an average score on an examination of 70%, we assume that is the average and that most students have earned a score close to that value.

Statistical methods help us account for the variability of the observed measurements.

A few details about *P* values³²:

- “*P*” is governed by sample size (*n*) and effect size (i.e., a measurement of the magnitude of a treatment effect when comparing one group to the other)
- “*P*” describes the significance of study treatment based on a statistical test
 - $P < 0.05$ indicates significance if the type I error (level of significance) was predefined as 0.05
 - Researchers set the level of significance, or α , before a study is conducted
 - A *P* value lower than the α means that the researcher can reject the null hypothesis (H_0) of no effect of treatment (i.e., no difference between the 2 groups.)

Example: Hypothetical Drug Study



Mean baseline PANSS score was 98.9 for all treatment groups (n = 339).
* $P \leq 0.05$ compared with placebo.

At $\alpha = 0.05$, you can reject the H_0 of no difference of PANSS score at endpoint between each medication group compared with placebo because you have demonstrated a statistically significant difference in mean score between the groups.

- Statisticians don't prove that a drug works
- They disprove that a drug doesn't work (the "null hypothesis")

103

INTENT-TO-TREAT (ITT)

- The ITT population includes all data from all randomly allocated subjects to study treatment groups, whether or not they completed the study^{32,33}
- It is sometimes defined as all randomized patients who received at least one dose of study drug and had at least one postbaseline assessment^{32,33}

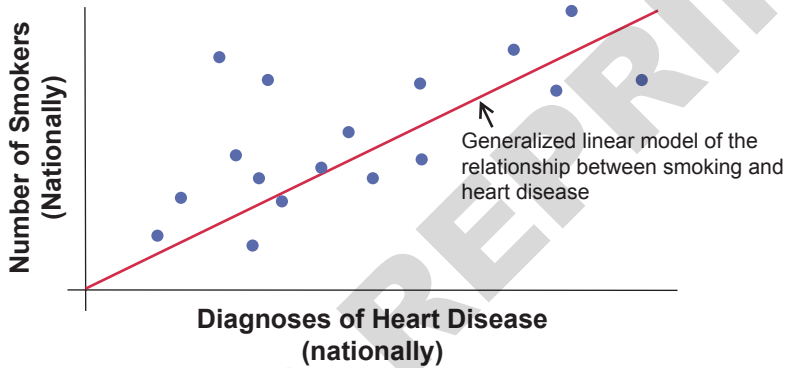
PATIENT #	OBSERVED DATA						LAST VISIT
	Visit 1	Visit 2	Visit 3	Visit 4	Visit 5	Visit 6	Analysis
1							All data used in analysis
2							
3							
4							

LEAST SQUARES (LS) MEAN

Sometimes it is important to detect the relationship between 2 variables in addition to comparing the difference between study treatment groups.

In this case, consider:

- Do the trends associated with 2 variables change at the same time? If they do, they are “correlated.”
e.g.: Smoking and heart disease; wearing seatbelts and automobile deaths; antipsychotic medication and PANSS scores
- How can we describe the relationship? Is it linear?



104





























When analyzing correlations, the relationship can be tested by determining how well the data fit the equation of the line (the model) if only 2 variables are considered.³⁴

Least squares mean is calculated based on the differences in the observed measurement and the predicted value, obtained by fitting the data to the linear model.³⁴

It is important to remember: “correlated” or “associated” “caused by.”

LAST OBSERVATION CARRIED FORWARD (LOCF)

- The LOCF (last observation carried forward) is a method to impute the last time point assessment value by the last available assessed value^{32,33}
- The LOCF approach assumes that a subject's response would remain invariable from the time of the last observation to the last study visit^{32,33}

PATIENT #	OBSERVED DATA						LAST VISIT
	Visit 1	Visit 2	Visit 3	Visit 4	Visit 5	Visit 6	Analysis
1							
2							
3							
4							

Imputed visit 2 data

























Imputed visit 3 data

105

MIXED MODEL FOR REPEATED MEASURES (MMRM)

- In MMRM analysis, all observed data from each visit are used in the model³³
- The model considers the correlation of measurements between visits of each subject³⁵
- Model parameters such as treatment effect are estimated by a maximum likelihood method
- MMRM is thought to be less biased in estimation of treatment effect than LOCF and OC (Observed Cases) when missing data have occurred and missing pattern is appropriate for using MMRM³⁶

























MIXED MODEL FOR REPEATED MEASURES (MMRM)

PATIENT #	OBSERVED DATA					
	Visit 1	Visit 2	Visit 3	Visit 4	Visit 5	Visit 6
1						
2						
3						
4						

106

OBSERVED CASE (OC)

- OC analysis includes ONLY those subjects who completed the trial, ignoring any and all dropouts³⁶

PATIENT #	OBSERVED DATA					
	Visit 1	Visit 2	Visit 3	Visit 4	Visit 5	Visit 6
1						
2						
3						
4						

ANALYSIS OF COVARIANCE (ANCOVA)

Analysis of covariance (ANCOVA) is an analytical method that allows analysis of complex data sets that involves many interacting variables, or covariates.^{33,37}

ANCOVA is an extension of analysis of variance (ANOVA) and allows for the possible effects of covariates on the response measurements in addition to the effects of the study treatment.^{33,37}

In ANCOVA (also ANOVA), the total variance in the observed measurements is divided into a part due to between group means and a part due to intergroup means, and then the mean difference of the groups can be assessed.^{33,37}

EFFECT SIZE

Effect Size is a quantification of size of a difference between two groups.^{38,39} There are many ways of measuring effect size. The most straightforward is to simply calculate the difference between the mean values of two groups (for example, a population receiving an experimental compound and a population receiving placebo), and divide by the pooled standard deviation. This method is referred to as *Cohen's d*:

$$d = \frac{m_2 - m_1}{S_{pooled}}$$

where m_1 and m_2 are the respective mean values of a measurement made within two groups, and S_{pooled} is the pooled standard deviation.³⁸ The larger the Cohen's d , the larger the effect size. A Cohen's d of 0.2 signifies a small effect, 0.5 signifies a medium effect, and 0.8 signifies a large effect.⁴⁰

Effect Size is broadly applicable to many types of measurements, and can be used to:

- Investigate the effectiveness of a particular intervention for a defined group
- Compare the magnitude of effectiveness between different interventions

NUMBER NEEDED TO TREAT (NNT) AND NUMBER NEEDED TO HARM (NNH)⁴¹

The Number Needed to Treat (NNT) is the number of patients who must be treated with one intervention versus another to see a difference in an outcome (e.g., response).

Example: with an NNT of 2, one would expect to encounter an additional outcome of interest for every 2 patients treated with one treatment versus another. An NNT of 100 would mean that you would need to treat 100 patients before expecting to encounter an additional outcome of interest.

NNT is calculated by first subtracting the rates of the outcome in question that are associated with the two interventions, and then calculating the reciprocal of this difference. In other words,

$$\text{NNT} = 1/(\text{Rate}_1 - \text{Rate}_2)$$

The concept of NNT can be used to compare treatments in terms of potential adverse events. By convention, the term in this case is **Number Needed to Harm (NNH)**. NNH is calculated in the same manner as NNT but is used to describe the number of patients we would need to treat with Treatment A versus Treatment B before we would expect to encounter one additional adverse event of interest.

REFERENCES

1. Endicott J, Nee J, Harrison W, Blumenthal R. Quality of Life Enjoyment and Satisfaction Questionnaire: a new measure. *Psychopharmacol Bull.* 1993;29:321–326.
2. Kay SR, et al. The Positive and Negative Syndrome Scale (PANSS) for schizophrenia. *Schizophr Bull.* 1987;13:263–276.
3. Sajatovic M, Ramirez LF. *Rating Scales in Mental Health*. 2nd ed. Hudson, OH: Lexi-Comp, Inc; 2003.
4. Montgomery SA, Åsberg M. A new depression scale designed to be sensitive to change. *Br J Psychiat.* 1979;134:382–389.
5. Spearing MK, Post RM, Leverich GS, Brandt D, Nolen W. Modification of the Clinical Global Impressions (CGI) scale for use in bipolar illness (BP): the CGI BP. *Psychiat Res.* 1997;73:159–171.
6. Young RC, Biggs JT, Ziegler VE, Meyer DA. A rating scale for mania: reliability, validity, and sensitivity. *Br J Psychiat.* 1978;133:429–435.
7. Hamilton M. The assessment of anxiety states by rating. *Br J Med Psychol.* 1959;32:50–55.
8. Rush AJ, Trivedi MH, Ibrahim HM, et al. The 16-item Quick Inventory of Depressive Symptomatology (QIDS), Clinician Rating (QIDS-C), and Self-Report (QIDS-SR): a psychometric evaluation in patients with chronic major depression. *Biol Psychiatry.* 2003;54:573–583.
9. Overall JE, Gorham DR. The Brief Psychiatric Rating Scale. *Psychol Rep.* 1962;10:799–812.
10. Leucht S, et al. What does the PANSS mean? *Schizophr Res.* 2005;79:231–238.
11. Guy W. ECDEU Assessment Manual for Psychopharmacology. 1976. US Department of Health, Education, and Welfare. Public Health Service: Alcohol, Drug Abuse, and Mental Health Administration.
12. Sheehan DV, Harnett-Sheehan K, Raj BA. The measurement of disability. *Int Clin Psychopharmacol.* 1996;11(suppl 3):89–95.
13. Simpson GM, Angus JWS. A rating scale for extrapyramidal side effects. *Acta Psychiatr Scand Suppl.* 1970;212:11–19.

14. Hawley CJ, et al. The use of the Simpson Angus Scale for the assessment of movement disorder: a training guide. *Int J Psych Clin Pract.* 2003;7:249–257.
15. Barnes TRE. A rating scale for drug-induced akathisia. *Br J Psychiatry.* 1989;154:672–676.
16. Lane RD, et al. Assessment of tardive dyskinesia using the abnormal involuntary movement scale. *J Nerv Ment Dis.* 1985;173:353–357.
17. Posner K, Brent D, Lucas C, et al. Columbia-Suicide Severity Rating Scale (C-SSRS). 2008. The Research Foundation for Mental Hygiene, Inc.
18. Posner K, Brown GK, Stanley B, et al. The Columbia-Suicide Severity Rating Scale: initial validity and internal consistency findings from three multisite studies with adolescents and adults. *Am J Psychiatry.* 2011;168:1266–1277.
19. National Heart, Lung, and Blood Institute. Obesity Education Initiative: “The Practical Guide: Identification, Evaluation, and Treatment of Overweight and Obesity in Adults.”
20. Manu P, et al. Prediabetes in patients treated with antipsychotic drugs. *J Clin Psychiatry.* 2012;73:460–466.
21. American Diabetes Association. Standards of medical care in diabetes–2013. *Diabetes Care.* 2013; 36(Suppl 1):S11–S66.
22. Gayoso-Diz P, et al. Insulin resistance index (HOMA-IR) levels in a general adult population: curves percentile by gender and age. The EPIRCE study. *Diabetes Res Clin Pract.* 2011;94:146–155.
23. Daly ME, et al. Acute effects on insulin sensitivity and diurnal metabolic profiles of a high-sucrose compared with a high-starch diet. *Am J Clin Nutr.* 1998;67:1186–1196.
24. Grundy SM, et al. Diagnosis and management of the metabolic syndrome: an American Heart Association/National Heart, Lung, and Blood Institute Scientific Statement. *Circulation.* 2005;112: 2735–2752.
25. Matthews DR, et al. Homeostasis model assessment: insulin resistance and beta-cell function from fasting plasma glucose and insulin concentrations in man. *Diabetologia.* 1985;28:412–419.
26. American Medical Association. Executive Summary of the Third Report of the National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III). *JAMA.* 2001;285:2486–2497.
27. Schmidt M, et al. Increased prolactin concentrations in a patient with bipolar disorder. *Clin Chem.* 2013;59:473–475.
28. Kadish A, et al. ACC/AHA clinical competence statement on electrocardiography and ambulatory electrocardiography: A report of the ACC/AHA/ACP-ASIM task force on clinical competence (ACC/ AHA Committee to Develop a Clinical Competence Statement on Electrocardiography and Ambulatory Electrocardiography). *J Am Coll Cardiol.* 2001;104:3169–3178.
29. Strnadova C. The assessment of QT/QTc interval prolongation in clinical trials: a regulatory perspective. *Drug Inf J.* 2005;39:407–433.
30. Al-Khatib SM, et al. What clinicians should know about the QT interval. *JAMA.* 2003;289:2120–2127.
31. Viskin S. The QT interval: too long, too short or just right. *Heart Rhythm.* 2009;6:711–715.
32. Lachin JM. Statistical considerations in the intent-to-treat principle. *Control Clin Trials.* 2000; 21:167–189.
33. DeSouza CM, et al. An overview of practical approaches for handling missing data in clinical trials. *J Biopharm Stat.* 2009;19:1055–1073.
34. Ugrinowitsch C, et al. Limitations of ordinary least squares models in analyzing repeated measures data. *Med Sci Sports Exerc.* 2004;36:2144–2148.
35. Fieuws S, et al. Random-effects models for multivariate repeated measures. *Stat Methods Med Res.* 2007;16:387–397.
36. Prakash A, et al. The impact of analytic method on interpretation of outcomes in longitudinal clinical trials. *Int J Clin Pract.* 2008;62:1147–1158.
37. Van Breukelen GJ, Van Dijk KR. Use of covariates in randomized controlled trials. *J Int Neuropsychol Soc.* 2007;13:903–904.
38. Nakagawa S, Cuthill IC. Effect size, confidence interval and statistical significance: a practical guide for biologists. *Biol Rev.* 2007;82:591–605.
39. Kelley K, Preacher KJ. On effect size. *Psychol Methods.* 2012;17:137–152.
40. Kraemer HC, Kupfer DJ. Size of treatment effects and their importance to clinical research and practice. *Biol Psychiatry.* 2006;59:990–996.
41. Citrome L. Number needed to treat: what it is and what it isn't, and why every clinician should know how to calculate it. *J Clin Psychiatry.* 2011;72(3):412–413.