Glossary

Numbers in parentheses indicate the chapter in which the term is introduced.

Terms

absolute reliability. Indicates how much of a measured value, expressed in the original units, is likely to be due to error. (9,32)

absolute risk increase (ARI). The increase in risk associated with an intervention as compared to the risk without the intervention (or the control condition); the absolute difference between the control event rate (CER) and the experimental event rate (EER). (34)

absolute risk reduction (ARR). The reduction in risk associated with an intervention as compared to the risk without the intervention (or the control condition); the absolute difference between the experimental event rate (EER) and the control event rate (CER). (34)

accessible population. The actual population of subjects available to be chosen for a study. This group is usually a nonrandom subset of the target population. (13)

active variable. An independent variable with levels that can be manipulated and assigned by the researcher. (14)

adjusted means. Means that have been adjusted based on the value of a covariate in an analysis of covariance. (30)

agreement. (see *percent agreement*)

allocation concealment. Implementation of a process of random assignment where those involved in the trial are shielded from knowing the upcoming participant group assignment. (14)

alpha coefficient. (see *Cronbach's alpha*)

alpha (α). Level of statistical significance, or risk of Type I error; maximum probability level that can be achieved in a statistical test to reject the null hypothesis. (23) (see also *Cronbach's alpha*)

alternate forms reliability. Reliability of two equivalent forms of a measuring instrument. (32)

alternating treatment design. A single-case design in which two (or more) treatments are compared by alternating them within a session or in alternate sessions. (18)

alternative hypothesis (H_1). Hypothesis stating the expected relationship between independent and dependent variables; considered the negation of the null hypothesis. The alternative hypothesis is accepted when the null hypothesis is rejected. (23)

analysis of covariance (ANCOVA). Statistical procedure used to compare two or more conditions while controlling for the effect of one or more covariates. (15,30)

analysis of variance (ANOVA). Statistical procedure appropriate for comparison of three or more treatment groups or conditions, or the simultaneous manipulation of two or more independent variables; based on the *F* statistic. (25)

a priori comparisons. (see *planned comparisons*)

arm. (see *treatment arm*)

attributable risk. An estimate used to quantify the risk of disease in an exposed group that is attributable to the exposure, by removing the risk that would have occurred as a result of other causes (risk in the unexposed group). (34)

attribute variable. An independent variable with levels that cannot be manipulated or assigned by the researcher but that represent subject characteristics (such as age and sex). (14)

attrition (experimental mortality). A threat to internal validity, referring to the differential loss of participants during the course of data collection, potentially introducing bias by changing the composition of the sample. (15)

audit trail. Comprehensive process of documenting interpretation of qualitative data. (21)

autonomy. The capacity of individuals to make decisions affecting their own lives and to act on those decisions. (7)

background question. Question related to etiology or general knowledge about a patient's condition, referring to the cause of a disease or condition, its natural history, signs and symptoms, general management, or the anatomic or physiological mechanisms that relate to pathophysiology. (5)

backward selection. Form of stepwise multiple regression. The equation starts out with all independent variables in the model, and variables are eliminated one at a time if they do not contribute significantly to the model. (30)

basic research (preclinical research). Research that contributes to basic knowledge but that does not have immediate practical goals. (2,14)

Bayes' theorem. The calculation of the probability of an event based on the prior probability of another event; used to estimate posttest probabilities based on pretest probabilities of a diagnostic outcome. (33)

beneficence. Obligation to attend to the well-being of individuals engaged as research subjects. (7)

beta (β). 1. Probability of making a Type II error. (23) 2. Used to represent coefficients for the relationship between two endogenous variables in structural equation modeling. (31)

beta weight. In a multiple regression equation, the standardized coefficient for each independent variable. (30)

between-groups variance. That portion of the total variance in a set of scores that is attributed to the difference between groups. (24,25)

between-subjects design. A design that compares independent groups. (16)

bias. Any influence that may interfere with the valid relationship between variables, potentially resulting in misleading interpretation of outcomes. (15)

binomial test. A significance test of the deviations of observations from a theoretically expected deviation for a dichotomous variable. 1. Used to evaluate data points above or below the split middle line in single-subject designs. (18) 2. Used to compare two sets of data in a repeated sample with the nonparametric sign test. (27)

Bland-Altman plot. A plot analyzing the agreement between two measurements. The mean of the two measures for each subject is plotted against the difference between the means. (see also *limits of agreement*). (32)

blinding. Techniques to reduce experimental bias by keeping participants and/or investigators ignorant of group assignments and research hypotheses. (14)

blocking variable. Attribute variable in which subjects are divided into groups or blocks that are homogeneous on a particular characteristic. (15,16)

block randomization. Distributes subjects evenly among treatment groups within small, even numbered subgroups or "blocks." (14)

Bonferroni correction (adjustment). A correction often used when multiple statistical tests are performed on the same data set to reduce Type I error. The desired level of significance (α) is divided by the number of comparisons. The resulting value is then used as the level of significance for each comparison to reject the null hypothesis. (26)

Boolean logic. In literature searches, the terms AND, NOT, and OR used to expand or narrow search terms. (6)

box plot (box-and-whisker plot). A graphic display of a distribution, showing the median, 25th and 75th percentiles (interquartile range), and highest and lowest scores. (22)

canonical correlation. A multivariate correlation procedure, whereby two sets of variables are correlated. (31)

carryover effect. A temporary or permanent change in behavior resulting from prior treatments. (10)

case-control study. A design in analytic epidemiology in which the investigator selects subjects on the basis of their having or not having a particular disease or condition and then determines their previous exposure to a risk factor. (1,19)

case report (case series). Detailed report of the symptoms, signs, diagnosis, treatment, and follow-up of one or more individual patients, usually describing interesting treatment options or uncommon conditions. (20)

case study. A qualitative research design used to study complex phenomena within their social context, investigating individuals or organizations. Can involve evaluation of interventions, policies, or theoretical premises. (21)

causal modeling. Statistical technique that examines patterns of intercorrelations among variables to determine if they fit an underlying theory of which variables cause others. (see also *structural equation modeling*) (31)

ceiling effect. A measurement limitation of an instrument whereby the scale cannot determine increased performance beyond a certain level. (10)

celeration line. (see *split middle line*)
censored observation. An observation whose value is unknown because the subject has not been in the study long enough for the outcome to have occurred; used to estimate survival curves. (31)
central tendency. Descriptive statistics that represent "averages" or scores that are representative of a distribution; includes mean, median, and mode. (22)
centroid. A point determined from the intersection of two or more means of two dependent variables (*X*, *Y*), used in multivariate analysis. (31)
chi-square (χ^2). A nonparametric test applied to categorical data, comparing observed frequencies within categories to frequencies expected by chance. (28)
classical measurement theory (CMT). Concept of measurement whereby a measured value is considered to be a function of an underlying true score and measurement error. (9, 12)
clinical practice guideline (CPG). Systematically developed statement, based on available evidence, that provide recommendations to assist practitioner and patient decisions about appropriate healthcare for specific clinical circumstances. (5,37)
clinical prediction rule. A combination of clinical findings that have been shown to predict a diagnosis, prognosis, or treatment outcome. (33)
clinical trial. (see *randomized controlled trial*)
clinical trials registry. A database that provides listing of trials in planning and implementation phases. (14)
closed-ended question. A question on a survey (interview or questionnaire) that offers a set of specific response choices that are mutually exclusive and exhaustive. (11)
cluster analysis. A multivariate statistical procedure that classifies subjects into sets based on defined characteristics. (31)
cluster random assignment. The site, representing a "cluster" of subjects, is randomly assigned to an intervention, and all individuals at that site receive that treatment. (14)
cluster sampling. A form of probability sampling in which large subgroups (clusters) are randomly selected first, and then smaller units from these clusters are successively chosen; also called multistage sampling. (13)
coefficient alpha. (see *Cronbach's alpha*)
coefficient of determination (r^2). Coefficient representing the amount of variance in one variable (*Y*) that can be explained (accounted for) by a second variable (*X*). (30)
coefficient of variation (CV). A measure of relative variation; based on the standard deviation divided by the mean, expressed as a percentage. Can be used to describe data measured on the interval or ratio scale. (22)
Cohen's *d*. The effect size index for comparing two groups, equals the difference between means divided by a pooled standard deviation. (see also *standardized mean difference*). (24)
cohort effects. Variations of effects of a given group as they move through time, such as individuals born in a particular era. (19,20)
cohort study. An observational study design in which a specific group is followed over time. Subjects are classified according to whether they do or do not have a particular risk factor or exposure and followed to determine disease outcomes. (19)
collinearity. The correlation between independent variables in a multiple regression equation, causing them to provide redundant information. Also called multicollinearity. (30)
Common Rule. Codification of policies related to informed consent and ethical conduct in research. (7)
communality. In factor analysis, the extent to which an item correlates with other variables. (31)
completer analysis (complete case analysis). Analysis of data in a clinical trial only for those subjects who complete the study. (15)
complex contrasts. A multiple comparison strategy in which means from two or more groups are combined as a subset and compared with other individual means or subsets of means. (26)
complex hypothesis. Contains more than one independent or dependent variable. (3)
computer adaptive testing (CAT). A computer-based test or scale in which items are adapted to the subject's ability level. (12)
concurrent validity. A form of criterion-related validity; the degree to which the outcomes of one test correlate with outcomes on a criterion test, when both tests are given at relatively the same time. (10,32)
confirmability. In qualitative research, ensuring, as much as possible, that findings are due to the experiences and ideas of the participants, rather than the characteristics and preferences of the researcher. (21)
confirmatory factor analysis (CFA). Used to support the theoretical structure of an instrument, to determine if it fits with current empirical understanding of a construct. (10, 31)
confidence interval (CI). The range of values within which a population parameter is estimated

to fall, with a specific level of confidence, usually 95%. (23)

confounding. The contaminating effect of extraneous variables on interpretation of the relationship between independent and dependent variables. (15,19,34)

consecutive sampling. A form of nonprobability sampling, where subjects are recruited as they become available. (13)

constant comparison. In qualitative research, the process by which each new piece of information is compared with data already collected to determine where the data agree or conflict. (21)

construct. An abstract concept that is unobservable, that can only be measured by observing related behaviors. (see also *latent trait*) (4,8)

construct validity. 1. A type of measurement validity indicating the degree to which a theoretical construct is measured by an instrument. (10) 2. Design validity related to operational definitions of independent and dependent variables. (15)

content analysis. A procedure for analyzing and coding narrative data in a systematic way. (21)

content validity. A type of measurement validity indicating the degree to which the items in an instrument adequately reflect the content domain being measured. (10)

contingency table (crosstabulation). A two-dimensional table displaying frequencies or counts, with rows (R) and columns (C) representing categories of nominal or ordinal variables. (28)

continuity correction. (see *Yates' continuity correction*)

continuous variable. A quantitative variable that can theoretically take on values along a continuum. (8)

control event rate (CER). The number of subjects in the control group who develop the outcome of interest. (34)

control group. In an experiment, a group of subjects who resemble the experimental group but who do not receive the experimental treatment (assigned a placebo or control condition), providing a baseline of comparison to interpret effects of treatment. (14)

convenience sampling. A nonprobability sampling procedure, involving selection of the most available subjects for a study. (13)

convergent validity. An approach in construct validation, assessing the degree to which two different instruments or methods are able to measure the same construct. (10)

correlation. The tendency for variation in one variable to be related to variation in a second variable; those statistical procedures used to assess the degree of covariation between two variables. (29)

covariate. An extraneous variable that is statistically controlled in an analysis of covariance or regression analysis, so that the relationship between the independent and dependent variables is analyzed with the effect of the extraneous factor removed. (15,30)

Cox proportional hazards regression. A regression procedure used to measure the effect of one or more predictor variables on the rate (hazard) at which an outcome occurs, accounting for differing lengths of follow-up among subjects. Used in survival analysis. (31)

credibility. In qualitative research, a criterion for integrity of data, confidence in the truth or validity of findings. (21)

criterion-referenced test. A fixed standard that represents an acceptable level of performance. (10)

criterion-related validity. A type of measurement validity indicating the degree to which the outcomes of one test correlate with outcomes on a criterion test; can be assessed as concurrent validity or predictive validity. (10)

critically appraised topic (CAT). A short summary and appraisal of evidence focused on a clinical question. (36)

critical region. Area of a sample probability curve that corresponds to rejection of the null hypothesis based on the critical value of the test statistic. (23)

critical value. The value of a test statistic that must be exceeded for the null hypothesis to be rejected; the value of a statistic that defines the critical region. (23)

Cronbach's alpha (α). Reliability index of internal consistency of items on a scale. (9,32)

crossover design. A repeated measures design used to control order effects when comparing two treatments, where half of the sample receives treatment A first followed by treatment B, and the other half receives treatment B first followed by treatment A. May include a washout period. (16)

cross-sectional study. A study based on data collected at one point in time. Observations of different age or developmental groups may provide the basis for inferring trends over time. (19)

crosstabulation. (see *contingency table*)

cumulative incidence (CI). The number of new cases of a disease during a specified time period divided by the total number of people at risk; the proportion of new cases of a disease in a population. (34)

cumulative scale. A scale designed so that agreement with higher-level responses assumes agreement with all lower-level responses. Also called a Guttman scale. (12)

curvilinear relationship. The relationship between two variables that does not follow a linear proportional relationship. (29,30)

cut-off score. Score used as the demarcation of a positive or negative continuous test outcome. (12,33)

deductive reasoning. The logical process of developing specific hypotheses based on general principles. (4)

degrees of freedom (*df*). Statistical concept indicating the number of values within a distribution that are free to vary, given restrictions on the data set; usually n-1. (23)

Delphi survey. Survey method whereby decisions on items are based on consensus of a panel over several rounds. (11 Supplement)

dependability. In qualitative research, the stability of data over time, the degree to which the study could be repeated with similar results. (21)

dependent variable. A response variable that is assumed to depend on or be caused by another (independent) variable. (3)

descriptive research. Research studies that are designed to describe the characteristics of individuals in specific populations. (20)

descriptive statistics. Statistics that are used to characterize the shape, central tendency, and variability within a set of data, often with the intent to describe a sample or population. (22)

developmental research. A descriptive research approach designed to document how certain groups change over time on specific variables. (20)

deviation score. The distance of a single data point from the mean of the distribution. The sum of the deviation scores for a given distribution will always equal zero. (22)

dichotomy (dichotomous variable). A nominal variable having only two categories, such as yes/no and male/female; a binomial variable. (8)

difference score (*d*). The difference between two scores taken on the same individual. (19)

differential item functioning (DIF). Potential item bias in the fit of data to the Rash model, showing how an item may be measuring different abilities for subgroups. (12)

digital object identifier (DOI). A unique number assigned to scholarly articles that can be used to locate particular articles in a search. (6,38)

directional hypothesis. A research hypothesis (or alternative hypothesis) that predicts the direction of a relationship between two variables. (3)

discrete variable. A variable that can only be measured in separate units and that cannot be measured in intervals of less than 1. (8)

discriminant analysis. A multivariate statistical technique used to determine if a set of variables can predict group membership. (31)

discriminant validity. An approach in construct validation assessing the degree to which an instrument yields different results when measuring two different constructs; that is, the ability to discriminate between the constructs. (10)

disproportional sample. A sample stratified on a particular variable, when the number of subjects within a stratum is not proportional to the population size of that stratum. (13)

distribution. The total set of scores for a particular variable. (22)

divergent validity. (see *discriminant validity*)

DOI. (see *digital object identifier*)

dose-response relationship. An outcome in which risk varies, not only with the presence or absence of an exposure, but also with varying levels of the exposure's presence. (19)

double-blind study. An experiment in which both the investigator and the subject are kept ignorant of group assignment. (14)

dummy variable (coding). In regression procedures, the assignment of codes (0 and 1) to a nominal variable, reflecting the presence or absence of certain traits. (30)

ecological validity. The generalizability of study findings to real-word conditions, societal norms, and health of populations. (15)

effectiveness. Benefits of an intervention as tested under "real world" conditions. (2)

effect modification. A situation in which the strength of an association between two variables is affected by a third variable that differs across subgroups in the population. (34)

effect size. The magnitude of the difference between treatments or the magnitude of a relationship between variables. (23)

effect size index. A statistical value used to measure the effect size in standardized units based on the proportional relationship of the difference to variance. Different indices are used with various statistical tests. (23)

efficacy. Benefit of an intervention as tested under controlled experimental conditions, usually with a control group in a randomized controlled trial. (2)

eigenvalue. A measure of the proportion of the total variance accounted for by a factor in a factor analysis. (31)

endogenous variable. The variable being explained by a causal model, a dependent variable. (31)

epidemiology. That branch of research dedicated to exploring the frequency and determinants of disease or other health outcomes in populations. (34)

equipoise. The ethics of a clinical research situation in which there is genuine uncertainty regarding the comparative therapeutic merits of each arm of a trial. (14)

equivalence trial. An intervention study focused on showing that two treatments are not different, no better, no worse. (14)

error variance. That portion of the total variance in a data set that cannot be attributed to treatment effects but that is due to differences between subjects. (23)

eta squared (η^2). Effect size index for analysis of variance. (25)

ethnography. An approach to qualitative research in which the experiences of a specific cultural group are studied. (21)

exclusion criteria. Specific criteria that are used to determine who is not eligible for a study. (8,13)

exogenous variable. A variable in a causal model that is not explained by the model, with variance that is accounted for by variables outside of the model, an independent variable. (31)

expected frequencies. In a contingency table, the frequencies that would be expected if the null hypothesis is true; frequencies that are expected just by chance. (28)

experimental design. A design in which the investigator manipulates the independent variables and randomly assigns subjects to groups, and in which a control group or comparison group is used. (16)

experimental event rate (EER). The number of subjects in the experimental or treatment group who develop the outcome of interest. (34)

explained variance. Between-groups variance; that portion of the total variance in a data set that can be attributed to the differences between groups or treatment conditions. (25)

explanatory research. Utilizes various types of experimental designs to compare two or more conditions or interventions. (1,3)

exploratory factor analysis (EFA). Used to study a set of items, when the purpose is to determine how the variables cluster, or to establish what underlying concepts may be present in the construct. (10,31)

exploratory research. Observational research that has as its purpose the exploration of data to determine relationships among variables. (3,19)

exposure. An experience or condition that may be responsible for a health outcome. (19)

external criticism. Assessment of generalizability in historical research. (20)

external validity. The degree to which results of a study can be generalized to persons or settings outside the experimental situation. (15)

fabrication. Reporting data or results that have been made up. (7)

face validity. The assumption of validity of a measuring instrument based on its appearance as a reasonable measure of a given variable. (10)

factor. 1. A variable. (3) 2. An independent variable in an experimental study. (25) 3. A set of interrelated variables in a factor analysis. (31)

factor analysis. An exploratory multivariate statistical technique used to examine the structure of a latent variable within a large set of variables and to determine the underlying dimensions that exist within that set of variables. (see also *exploratory* and *confirmatory factor analysis*). (10,31)

factorial design. A design that incorporates two or more independent variables, with independent groups of subjects randomly assigned to various combinations of levels of the variables. (10)

false negative. A test result that is negative in a person who has the disease or condition of interest. (34)

falsification. Manipulating, changing, or omitting data or results such that the research is not accurately represented in the research record. (7)

false positive. A test result that is positive in a person who does not have the disease or condition of interest. (34)

familywise error rate (α_{FW}). The probability of at least one false significant finding in a series of hypothesis tests, Type I error. (26)

field notes. Notes recorded during participation or observation as part of qualitative research. (21)

Fisher's exact test. A nonparametric procedure applied to nominal data in a 2 x 2 contingency table, comparing observed frequencies within categories to frequencies expected by chance. Used when samples are too small to use the chi-square test. (28)

fixed effect. Variable levels that are nonrandom, representing the only qualities of interest. (32)

floor effect. A measurement limitation of an instrument whereby the scale cannot determine decreased performance beyond a certain level. (10)

foreground question. A clinical question that focuses on evidence to inform decisions about a particular patient's management. (5)

forest plot. Graphic display of effect sizes and confidence intervals for individual studies and pooled data in a meta-analysis. (37)

forward selection. A process used in multiple regression that enters variables one at a time into the equation based on the strength of their

association with the outcome variable, until all statistically significant variables are included. (30)

frequency distribution. A table of rank ordered scores that shows the number of times each value occurs in a distribution. (22)

Friedman two-way analysis of variance by ranks (χ_r^2). A nonparametric statistical procedure for repeated measures, comparing more than two treatment conditions of one independent variable; analogous to the one-way repeated measures analysis of variance. (27)

funnel plot. Scatterplot of treatment effect size against a measure of study precision such as the standard error, used as a visual aid for detecting publication bias or study heterogeneity in a meta-analysis. (37)

gamma (γ). A coefficient representing the relationship between exogenous and endogenous variables in structural equation modeling. (31)

generalizability. 1. The quality of research that justifies application of outcomes to groups or situations other than those directly involved in the investigation, or external validity. (15) 2. The concept of reliability theory in which measurement error is viewed as multidimensional and must be interpreted under specific measurement conditions. (9)

gold standard. A measurement that defines the true value of a variable. 1. In criterion-related validity, an instrument that is considered a valid measure and that can be used as the standard for assessing validity of other instruments. (10) 2. In diagnostic testing, a procedure that accurately identifies the true disease condition (negative or positive) of the subject. (33)

goodness of fit. Use of a test to determine if an observed distribution of variables fits a given theoretical distribution. (28,31)

grand theory. A comprehensive idea that tries to explain phenomena at the societal level. (4)

Greenhouse-Geiser correction. A correction for variance differences in a repeated measures analysis of variance. (25)

grey literature. Written materials not produced by a commercial publisher, including government documents, reports of all types, fact sheets, practice guidelines, conference proceedings, and theses or dissertations. (6,37)

grounded theory. An approach to collecting and analyzing data in qualitative research, with the goal of developing theories to explain observations and experience. (21)

Guttman scale. (see *cumulative scale*)

Hawthorne effect. The effect of participants' knowledge that they are part of a study on their performance. (15)

hazard function. The probability that a subject will achieve a specific outcome in a certain time interval. (31)

hierarchical regression. Multiple regression model where sets of variables are entered in an iterative fashion. (30)

histogram. A type of bar graph, composed of a series of columns, each representing the frequency of one score or group interval. (22)

historical controls. Subjects from previous research studies that serve as controls for experimental subjects in a subsequent study. (17)

historical research. Research that seeks to examine relationships and facts based on documentation of past events. (20)

history effect. A threat to internal validity, referring to the occurrence of extraneous events prior to a posttest that can affect the dependent variable. (15)

homogeneity of slopes. Assumption in analysis of covariance that regression slopes of covariates are parallel. (30)

homogeneity of variance. An underlying assumption in parametric statistics that variances of samples are not significantly different. (23,24,25)

homogeneous sample. A sample where participants are similar on specific characteristics, such as all the same gender or age group. (15)

homoscedasticity. (see *homogeneity of variance*)

hospital-based study. An observational study that recruits cases from patients in a medical institution. (19)

Huynh-Feldt correction. A correction for variance differences in a repeated measures analysis of variance. (25)

hypothesis. A statement of the expected relationship between variables. (3)

impact factor. A rating of journal reputation based on the frequency with which articles from that journal are cited in reports in other journals. (38)

implementation science. Approach to research that focuses on understanding the influence of environment, attitudes, and resources on whether research findings are actually translated to practice. (2)

imputation. Replacement of missing data points with estimated values that are based on observed data. (15)

incidence. The proportion of people who develop a given disease or condition within a specified time period. (34)

inclusion criteria. Specific criteria that are used to determine who is eligible for a study. (13)

independent factor. An independent variable in which the levels represent independent groups of subjects. (15)

independent variable. The variable that is presumed to cause, explain, or influence a dependent variable; a variable that is manipulated or controlled by the researcher, who sets its "values" or levels. (3)

index test. A diagnostic test evaluated against a reference standard to determine if it accurately identifies those with and without the condition of interest. (33)

inductive reasoning. The logical process of developing generalizations based on specific observations or facts. (4)

inductive theories. Data-based theories that evolve through a process of inductive reasoning. (4)

inferential statistics. That branch of statistics concerned with testing hypotheses and using sample data to make generalizations concerning populations. (23)

informed consent. An ethical principle that requires obtaining the consent of the individual to participate in a study based on full prior disclosure of risks and benefits. (7)

institutional review board (IRB). That group in an institution that is responsible for reviewing research proposals that will involve human subjects to determine adherence to ethical principles. (7)

instrumentation effect. A threat to internal validity in which bias is introduced by an unreliable or inaccurate measurement system. (15)

intention-to-treat (ITT). In a randomized trial, the principle whereby data are analyzed according to original group assignments, regardless of how subjects actually received treatment. (15)

interaction effect. The differential effect of levels of one independent variable on a second independent variable. (25)

internal consistency. A form of reliability, assessing the degree to which a set of items in an instrument all measure the same trait. Typically measured using Cronbach's alpha. (9,26)

internal criticism. Assessment of the accuracy of sources in historical research. (20)

internal validity. The degree to which the relationship between the independent and dependent variables is free from the effects of extraneous factors. (15)

International Classification of Functioning, Disability and Health (ICF). A model developed by the World Health Organization that posits the relationship among a health condition, body structures/functions, activities, and participation. (1)

interprofessional. Members of multiple professions working together, contributing their various skills in an integrative fashion, sharing perspectives to inform decision making. (1)

interquartile range (IQR). The difference between the first and third quartiles in a distribution, often expressed graphically in a boxplot. (22)

inter-rater reliability. The degree to which two or more raters can obtain the same ratings for a given variable. (9,32)

interrupted time-series design (ITS). A quasi-experimental design involving a series of measurements over time, interrupted by one or more treatment occasions. (17)

interval scale. Level of measurement in which values have equal intervals, but no true zero point. (8)

intraclass correlation coefficient (ICC). A reliability coefficient used to assess test-retest and rater reliability based on an analysis of variance; a generalizability coefficient. (9,32)

intrarater reliability. The degree to which one rater can obtain the same rating on multiple occasions of measuring the same unchanging variable. (9)

item response theory (IRT). A model that examines error in a series of items that measure a latent trait, usually on a questionnaire, to determine how well the items differentiate individuals based on ability and difficulty of items. Also called *latent trait theory*. (12)

item-to-total correlation. Correlation of individual items in a scale with the total scale score; an indication of internal consistency. (9,32)

justice. Fairness in all aspects of the research process. (7)

Kaplan-Meier estimate. A nonparametric statistic used to estimate survival based on that portion of individuals living for a certain amount of time who have experienced a treatment or disease. (31)

kappa (κ). A correction factor for percent agreement measures of reliability, accounting for the potential effect of chance agreements. (9,32)

Kendall's tau (τ). A nonparametric correlation procedure for use with ordinal data. (29)

knowledge translation. Describes the process of accelerating the application of knowledge to improve outcomes and change behavior for those involved in providing care. (5)

known groups method. A technique for construct validation, in which validity is determined by the degree to which an instrument can demonstrate different scores for groups known to vary on the variable being measured. (10)

Kolmogorov-Smirnov test of normality. A test of the null hypothesis that an observed distribution does not differ from the pattern of a normal distribution. If the test is not significant ($p>.05$), the distribution does not differ from a normal distribution. (see also *Shapiro-Wilks test of normality*) (23)

Kruskal–Wallis one-way analysis of variance by ranks (*H*). A nonparametric statistical procedure for comparing more than two independent groups representing levels of one independent variable; analogous to the one-way analysis of variance. (27)

last observation carried forward (LOCF). A method of data imputation for missing scores in which a subject's last data point before dropping out is used as the outcome score. (15)

latent trait. A multidimensional construct or abstract characteristic of individuals that cannot be observed, and must be measured through a proxy. (10) (see also *item response theory*)

Latin square. A matrix of columns and rows used to assign sequences of treatments to control for order effects. (16)

least squares method. A method of fitting a regression line to a set of bivariate data so as to minimize the sum of the squared vertical deviations of *Y* values around that line. (30)

level. 1. The "value" or classification of an independent variable. (3) 2. In single-subject research, the magnitude of the target behavior; changes in level are associated with differences in magnitude between the end of one phase and the beginning of the following phase. (18)

level of measurement. The precision of a scale based on how a characteristic is measured: nominal, ordinal, interval, and ratio levels. (8)

level of significance (α). (see *alpha*).

levels of evidence. A classification system whereby the level of confidence placed in study findings is based on the type of research and control of bias in the design. (5)

Levene's test. A test of the equality of variances, used with the independent *t* test and the analysis of variance. (24,25)

likelihood ratio (LR). 1. In diagnostic testing, the ratio indicating the usefulness of the test for ruling in or ruling out a condition. A positive likelihood ratio (LR+) indicates how much the odds of a disease are increased if the test is positive. A negative likelihood ratio (LR–) indicates how much the odds of a disease are decreased if a diagnostic test is negative. (33) 2. In logistic regression, a test of the overall relationship among variables, analogous to the change in R^2 in multiple regression. (31)

Likert scale. A summative scale based on responses to a set of statements for which respondents are asked to rate their degree of agreement or disagreement. (11,12)

limits of agreement. Index of reliability between alternate forms of an instrument. (9,32)

linear regression. The process of determining a regression equation to predict values of *Y* based on a linear relationship with values of *X*. (30)

line of best fit. The regression line, representing the relationship between two variables, usually plotted on a scatter diagram. (see also *least squares method*) (30)

listwise deletion. Elimination of all cases with missing data on any variable in data analysis. (15, Appendix C)

logical positivism. Philosophical approach to inquiry based on the assumption that values can be confirmed by observation or experimentation, contrasted with naturalistic inquiry. (21)

logistic regression. Multiple regression procedure to test the effect of one or more predictor variables on a dichotomous dependent variable; predicts odds associated with presence or absence of the dependent variable based on the independent variables. (31)

logit. Standardized value that is the natural logarithm of odds. (12,31)

log rank test. A test of the null hypothesis that two survival curves are identical in a Kaplan-Meier procedure. (31)

longitudinal study. A study designed to collect data over time. (19,20)

main effect. The separate effect of one independent variable in a multifactor design. (16)

Mann–Whitney *U* test. A nonparametric statistical test for comparing two independent groups; analogous to the unpaired *t*-test. Equivalent to the Wilcoxon rank sum test. (27)

marginal means. Means for levels of main effects in a factorial design. (25)

matching. Pairing of subjects on the basis of similarities on one or more variables to make two groups more homogenous and to avoid confounding. (15)

maturation effect. A threat to internal validity, in which changes occur in the dependent variable as a result of the passing of time. (15)

Mauchley's test of sphericity. Statistical measure of sphericity in a repeated measures analysis of variance (see *sphericity*). (25)

maximum likelihood estimation. Method of estimating parameters in a statistical model that are the best fit to observed data. Used with logistic regression. (31)

McNemar test. A nonparametric statistical test to assess the relationship between correlated nominal level measures; related to the chi-square test. (28)

mean ($\bar{X}$). A measure of central tendency, computed by summing the values of several observations and dividing by the number of observations; the value that is typically called the "average." (22)

mean square (*MS*). A value representing the variance; calculated by dividing the sum of squares for a particular effect by the degrees of freedom for that effect. (22,25)

measurement error. The difference between an observed value for a measurement and the theoretical true score; may be the result of systematic or random effects. (9,32)

median. A measure of central tendency representing the 50th percentile in a ranked distribution of scores; that is, that point at which 50% of the scores fall below and 50% fall above. (22)

median survival time. In a survival analysis, the time when half the patients are expected to survive, or when the chance of surviving beyond that time is 50%. (31)

Medical Subject Headings (MeSH). Hierarchical structure of search terms developed by the National Library of Medicine. (6)

MEDLINE. Database of bibliographic references supported by the National Library of Medicine. (6)

member checking. In qualitative research, the process of sharing preliminary findings and interpretations with research participants, allowing them to offer validation, feedback, critique, and alternative explanations that ultimately contribute to the credibility of findings. (21)

meta-analysis. Use of statistical techniques to pool effect sizes from several studies with similar independent and dependent variables, resulting an overall summary effect. (5,37)

meta-theory. An overarching theory that attempts to reconcile several theoretical perspectives in the explanation of sociological, psychological, and physiological phenomena. (4)

method error. A form of reliability testing for assessing response stability based on the discrepancy between two sets of repeated scores. (32 Supplement)

methodological research. Research designed to develop or refine procedures or instruments for measuring variables, generally focusing on reliability, validity, and change. (9,10)

middle range theories. Theories that sit between basic hypotheses that guide everyday practice and the systematic efforts to develop a unified theory to explain a set of social behaviors. (4)

minimal clinically important difference (MCID). The smallest difference in a measured variable that signifies an important rather than trivial difference a measurement. (10,32)

minimal detectable change (MDC). That amount of change in a variable that must be achieved to reflect a true difference; the smallest amount of change that passes the threshold of error. (9,32)

misclassification. In cohort or case-control studies, classifying participants to an exposure or outcome category that is inaccurate. (19)

missing at random (MAR). Missing data that may be related to the methodology of the study but not to the treatment variable. (15)

missing completely at random (MCAR). Assumption that missing data are missing because of unpredictable circumstances that have no connection to the variables of interest or the study design. (15)

missing not at random (MNAR). Missing data related to group membership, creating bias in data. (15)

mixed design. A design that incorporates independent variables that are independent (between-subjects) and repeated (within-subjects) factors. Also called a split-plot design. (16)

mixed methods research. A study that incorporates both quantitative and qualitative research methods. (1,21)

mode. A measure of central tendency representing the most commonly occurring score in a distribution. (22)

model. Symbolic representation of reality delineating concepts or variables and their relationships, often demonstrating the structural components of a theory or process. (4)

multicollinearity. (see *collinearity*)

multiple baseline design. In single-subject research, a design for collecting data for more than one subject, behavior, or treatment condition wherein baseline phases are staggered to provide control. (18)

multiple comparison test. A test of differences between individual means following analysis of variance, used to control for Type I error. (26)

multiple imputation. A method of dealing with missing data by creating a random data set using the available data, predicting plausible values derived from observed data. (15)

multiple regression. A multivariate statistical technique for establishing the predictive relationship between one dependent variable and a set of independent variables. (31)

multistage sampling. (see *cluster sampling*)

multitrait-multimethod matrix (MTMM). An approach to validity testing to examine the

relationship between two or more traits measured by two or more methods. (10)

multivariate analysis. A set of statistical procedures designed to analyze the relationship among three or more variables; includes techniques such as multiple regression, discriminant analysis, factor analysis, and multivariate analysis of variance. (31)

multivariate analysis of variance (MANOVA). An advanced multivariate procedure that provides a global test of significance for multiple dependent variables using an analysis of variance. (31)

N-of-1 trial. A form of single-subject investigation that applies a randomized crossover design to an individual patient. (18)

natural history. Longitudinal study of a disease or disorder, demonstrating the typical progress of the condition. (20)

naturalistic inquiry. Qualitative observation and interaction with subjects in their own natural environment. (21)

negative case analysis. In qualitative research, a method of validation by searching for concepts that do not support or may contradict themes emerging from the data. (21)

negative likelihood ratio (LR–). (see *likelihood ratio*)

negative predictive value (PV–), (see *predictive value*)

Newman-Keuls (NK) multiple comparison test. (see *Student-Newman-Keuls multiple comparison*)

nominal scale. Level of measurement for classification variables; assignment of "values" based on mutually exclusive and exhaustive categories with no inherent rank order. (8)

nondirectional hypothesis. A research hypothesis (or alternative hypothesis) that does not indicate the expected direction of the relationship between independent and dependent variables. (3)

non-inferiority margin. In a non-inferiority study, the biggest difference that would be acceptable to consider the new treatment a reasonable substitute for the standard therapy. (14)

non-inferiority trial. A clinical trial designed to show that a new treatment is no better and no worse than standard care. (14)

nonparametric statistics. A set of statistical procedures that are not based on assumptions about population parameters, or the shape of the underlying population distribution; most often used when data are measured on the nominal or ordinal scales. (8,27)

nonprobability sample. A sample that was not selected using random selection. (11,13)

normal curve. (see *normal distribution*)

normal distribution. A symmetrical bell-shaped theoretical distribution with defined properties, where most of the scores fall in the middle of the scale and progressively fewer fall at the extremes. (22)

normative research. A descriptive research approach designed to determine normal values for specific variables within a population. (20)

norm referencing. Interpretation of a score based on its value relative to a standardized score. (10)

null hypothesis (H_0). A statement of no difference or no relationship between variables; the statistical hypothesis. (3,23)

number needed to harm (NNH). The number of patients that need to be treated to observe one adverse outcome; the reciprocal of absolute risk increase (ARI). (34)

number needed to treat (NNT). The number of patients that need to be treated to prevent one adverse outcome or achieve one successful outcome; the reciprocal of absolute risk reduction (ARR). (34)

oblique rotation. In factor analysis, the rotation of factors that are correlated with each other. (31)

observational study. A study that does not involve an intervention or manipulation of an independent variable. (19)

observed frequencies. Frequencies that occur in a study of categorical variables. (28)

odds ratio (OR). An estimate of risk representing the odds that an outcome will occur given a particular exposure compared to the odds of the outcome occurring in the absence of the exposure; used as a risk estimate in case-control studies and logistic regression. (30,34)

one-tailed test. A statistical test based on a directional alternative hypothesis, in which critical values are obtained for only one tail of a distribution. (23)

one-way analysis of variance. An analysis of variance with one independent variable. (25)

one-way design. An experimental or quasi-experimental design that involves one independent variable. (16)

on-protocol analysis. Analysis of data in an experiment based only on subjects who completed the study according to assigned groups. Also called *completer analysis* or *on-treatment analysis*. (15)

open access. A term used to describe articles and journals that are free of restrictions on access. (6,38)

open-ended question. A question on a survey (interview or questionnaire) that does not restrict the respondent to specific choices but allows for a free response. (11)

open-label trial. A trial design wherein both the researchers and participants know which treatment is being administered. (14)

operational definition. Definition of a variable based on how it will be used in a particular study; how a dependent variable will be measured, how an independent variable will be manipulated. (3)

order effects. The sequential effect of one subject being exposed to several treatments in the same order; potentially manifested as carryover or practice effects. (16)

ordinal scale. Level of measurement in which scores are ranks. (8)

orthogonal. A model in which variables are independent or uncorrelated. Literally means "perpendicular" in mathematics. (26,31)

outcome variable. (see *dependent variable*)

outlier. A numeric value that does not fall within the range of most scores in a distribution. (30)

paired *t*-test. A parametric test for comparing two means for correlated samples or repeated measures; also called a correlated *t*-test. (24)

pairwise deletion. Elimination of cases with missing data on particular variables in a specific analysis only. (15, Appendix C)

paradigm shift. Fundamental transition in the way a discipline thinks about priorities and relationships, stimulating change in perspectives, and fostering preferences for varied approaches to research. (1)

parallel group design. A design comparing two independent groups that are similar on all characteristics other than the intervention. (14)

parameter. A measured characteristic of a population. (13,22)

parametric statistics. Statistical procedures for estimating population parameters and for testing hypotheses based on population parameters, with assumptions about the distribution of variables, and for use with interval or ratio measures. (8,18)

partial correlation. The correlation between two variables, with the effect of a third variable removed; also called a first-order correlation. (29)

participant observation. A method of data collection in qualitative research in which the researcher is embedded as a participant in the group that is being observed. (21)

path diagram. A flow chart that shows interconnections among variables in a causal model. (31)

patient-centered outcomes research (PCOR). An approach that has the distinct goal of engaging patients and other stakeholders in the development of questions and outcomes measures, encouraging them to become integral members of the research process. (2)

patient-oriented evidence that matters (POEM). Refers to outcomes that measure things that a patient would care about, such as symptoms, quality of life, function, cost of care, length of stay. (2)

patient-reported outcome measures (PROM). Any report of the status of a patient's health condition that comes directly from the patient, without interpretation of the patient's response by a clinician or anyone else. (2)

Pearson product–moment coefficient of correlation (*r*). A parametric statistical technique for determining the relationship between two variables. (29)

peer review. Process of review of submitted manuscripts or research proposals by subject experts to determine quality. (7,38).

percent agreement. A reliability test for categorical variables, estimating the ability of researchers to agree on category ratings. (32)

per comparison error rate (α_{PC}). Probability of making a Type I error in a single comparison. (26)

percentile. The percentage of a distribution that is below a specified value. A distribution is divided into 99 equal ranks, or percentiles, with 1% of the scores in each rank. (22)

performance bias. A source of bias related to differences in the provision of care to comparison groups. (15, 37)

per-protocol analysis. Eliminating subjects who did not get or complete their assigned treatment, and including only those subjects who sufficiently complied with the trial's protocol. (15)

person-item map. Output of a Rasch analysis that shows the distribution of persons according to ability and of items according to difficulty. (12)

person-time. The total amount of time each case in a study is at risk, used to calculate incidence rates. (34)

phase I trial. Researchers work to show that a new therapy is safe. (2,14)

phase II trial. Studies that explore efficacy and dosage effects of an intervention by measuring relevant outcomes. (2,14)

phase III trial. Builds on prior research to establish efficacy through randomized controlled studies that may include thousands of patients in multiple sites over many years. (2,14)

phase IV trial. Completed after a treatment has been approved with the purpose of gathering information on the treatment's effect in various populations or subgroups, under different clinical conditions to explore side effects associated with long-term use, and to learn about risk factors, benefits, and optimal use patterns. (2,14)

phenomenology. An approach to qualitative research involving the study of complex human experience as it is actually lived. (21)

phi coefficient (ϕ). A nonparametric correlation statistic for estimating the relationship between two dichotomous variables. (28,29)

PICO. An acronym used to represent identification of components of clinical and research questions: **P** = population, **I** = intervention, **C** = comparison, **O** = outcome. (3,5)

placebo. A control that is similar in every way to the experimental treatment except that it does not contain the active component that comprises the intervention's actions. (14)

plagiarism. The appropriation of another person's ideas, processes, results, or words without giving appropriate credit or attribution, and representing the work as one's own. (7)

planned comparisons. Multiple comparison tests that are designated prior to running a study. (26)

point biserial correlation (r_{pb}). A correlation statistic for estimating the relationship between a dichotomy and a continuous variable on the interval or ratio scale. (29)

point estimate. A single sample statistic that serves as an estimate of a population parameter. (22)

polynomial regression. Regression procedure for nonlinear data. (30)

polytomous. Variables that can have multiple values, such as a 5-point scale. (8,11)

population. The entire set of individuals or units to which data will be generalized. (11,13)

population-based study. A study in which participants are recruited from the general population. (19)

positive likelihood ratio (LR+). (see *likelihood ratio*)

positive predictive value (PV+). (see *predictive value*)

posterior probability. (see *posttest probability*)

post hoc comparisons. Multiple comparison tests that follow a significant analysis of variance. (26)

posttest probability (posterior probability). The probability of a condition existing after performing a diagnostic test; predictive value of a diagnostic test. Depends on the prior probability, and the test's sensitivity and specificity. (33)

power (1-β). The ability of a statistical test to find a significant difference that really does exist; the probability that a test will lead to rejection of the null hypothesis, based on sample size, variance, level of significance, and effect size. (23)

practice-based evidence (PBE). The type of evidence that is derived from real patient care problems, identifying gaps between recommended and actual practice. (2)

practice effects. The effect of learning with repeated tasks in a repeated measures design. (16)

pragmatic clinical trial (practical clinical trial) (PCT). An effectiveness trial carried out in real-world settings. Participants represent the patients who would typically receive treatment, and testing takes place in clinical practice settings. Outcomes focus on issues of importance to patients and stakeholders. (2,14)

precision. 1. The number of decimal places to which a calculation is taken. (8) 2. The acceptable degree of error in estimating sample size for relative risk. (34)

preclinical research. (see *basic research*)

predictive validity. A form of measurement validity in which an instrument is used to predict future performance. (10)

predictive value (PV). In diagnostic testing, a positive predictive value (PV+) indicates the probability that individuals with a positive test truly have the disease. A negative predictive value (PV–) indicates the probability that individuals with a negative test truly do not have the disease. (33)

predictor variable. (see *independent variable)*

pretest probability (prior probability). The probability that a condition exists prior to performing a diagnostic test. Can be estimated based on the prevalence of the condition in a specified group of subjects. (33)

prevalence. The number of cases of a disease at a given point in time, expressed as a proportion of the total population at risk. (34)

primary outcome. The measure that will be used to arrive at a decision on the overall result of the study, and which represents the greatest therapeutic benefit. (3)

primary source. Reference source that represents the original document by the original author. (6)

principal axis factoring. A method of extraction in exploratory factor analysis. (31)

principal components analysis (PCA). Multivariate analysis that converts a set of possibly correlated variables into a set of uncorrelated principal components based on variance within the data. Differentiated from factor analysis, which uses a different model to account for variance. (31)

principal investigator. That person designated as the lead investigator who will have primary responsibility for overseeing the study. (35)

prior probability. (see *pretest probability*)

probability (*p*). The likelihood that an event will occur, given all possible events. (23)

probability sample. A sample chosen using random selection methods. (11,13)

propensity score. A single score based on a set of characteristics generated through logistic regression that can be used in matching subjects

in observational or nonrandomized studies to reduce baseline confounding. (15,31)

proportional hazards model. (see *Cox's regression*)

proposition. Statement of the relationship between variables. (4)

prospective study. A study designed to collect data following development of the research question. (19)

publication bias. Tendency for researchers and editors to publish studies finding significant effects, to the exclusion of studies finding no effect. (6,37)

purposive (purposeful) sample. A nonprobability sample in which subjects are specifically selected by the researcher on the basis of subjective judgment that they will be the most representative. (13)

Q-sort. A research methodology based on ranking variables as to their importance. (11 Supplement)

quadratic trend. A nonlinear trend, with one turn in direction. (26,30)

qualitative research. Research that derives data from observation, interviews, or verbal interactions and focuses on the meaning of experience of the participants. (1,21)

quantitative research. Measurement of outcomes using numerical data under standardized conditions. (1)

quartile (*Q*). Three quartiles divide a distribution of ranked data into four equal groups, each containing 25% of the scores. (22)

quasi-experimental research. Comparative research approach in which subjects cannot be randomly assigned to groups or control groups are not used. (15,17)

quota sampling. Nonprobability sampling method in which stratification is used to obtain representative proportions of specific subgroups. (13)

random assignment (random allocation). Assignment of subjects to groups using probability methods, where every subject has an equal chance of being assigned to a group. (14)

random effect. Variable levels that are chosen at random. (32)

random error. A measurement error that occurs by chance, potentially increasing or decreasing the true score value to varying degrees. (8)

random sampling. Probability method of selecting subjects for a sample, where every subject in the population has an equal chance of being chosen. (13)

random selection. (see *random sampling*)

randomization. (see *random assignment*)

randomized block design. An experimental design in which one independent variable is an attribute variable (blocking variable), creating homogeneous blocks of subjects who are then randomly assigned to levels of the other independent variable. (16)

randomized consent design (Zelen design). Involves randomizing participants to experimental treatment or standard care prior to seeking consent, and then only approaching those who will be assigned to the experimental intervention. (14)

randomized controlled trial (RCT). An experimental study in which a clinical treatment is compared with a control condition, where subjects are randomly assigned to groups. Also called a randomized clinical trial. (1,14)

range. A measure of dispersion equal to the difference between the largest and smallest scores in a distribution. (22)

rank biserial correlation (r_{rb}). A correlation procedure for estimating the degree of relationship between a dichotomy and an ordinal variable. (29)

Rasch analysis. Transformation of items on an ordinal scale to an interval scale, demonstrating the unidimensional nature of a scale. (12)

ratio scale. The highest level of measurement, in which there are equal intervals between score units and a true zero point. (8)

reactive measurement. A measurement that distorts the variable being measured, either by the subject's awareness of being measured or by influence of the measurement process. (15)

recall bias. The possible inaccuracy of participants recalling medical history or previous exposures; of particular concern in retrospective studies. (11,19)

receiver operating characteristic (ROC) curve. In diagnostic testing, a plot of sensitivity (true positive rate) against 1-specificity (false positive rate), demonstrating strength of diagnostic accuracy, used to determine the most effective cut-off score. (33)

reference standard. A value used as a standard against which to judge a criterion; may or may not be a gold standard. Used to judge criterion-related validity or diagnostic accuracy. (10)

reflexivity. In qualitative research, the critical self-examination by researchers regarding their own biases or preferences and how they might influence interpretation of data. (21)

regression analysis. A statistical procedure for examining the predictive relationship between a dependent (criterion) variable and an independent (predictor) variable. (30)

regression coefficient. In a regression equation, the weight (*b*) assigned to the independent variable; the slope of the regression line. (30)

regression line. The straight line that is drawn on a scatter plot for bivariate data from the regression

equation, summarizing the relationship between variables. (see also *least squares method*). (30)

regression toward the mean (RTM). A statistical phenomenon in which scores on a pretest are likely to move toward the group mean on a posttest because of inherent positive or negative measurement error; also called statistical regression. (15)

relative reliability. Reliability measures that reflect true variance as a proportion of the total variance in a set of scores. (9)

relative risk (RR). Estimate of the magnitude of the association between an exposure and disease, indicating the likelihood that the exposed group will develop the disease relative to those who are not exposed. (34)

relative risk reduction (RRR). The reduction in risk associated with an intervention relative to the risk without the intervention (control); the absolute difference between the experimental event rate and the control event rate divided by the control event rate. (34)

reliability. The degree of consistency with which an instrument or rater measures a variable; the degree to which a measurement is free from error. (9,32)

reliability coefficient. Value used to quantify the degree of consistency in repeated measurements. (9)

repeated measure (repeated factor). An independent variable for which subjects act as their own control. Also called a within-subjects factor. (15,16)

research hypothesis. A statement of the researcher's expectations about the relationship between variables under study. (3,23)

reverse causation. A situation wherein the variable designated as the "outcome" may actually cause the "exposure." (19)

residual ($Y - \hat{Y}$). The difference between an observed value and a predicted value. (30,31)

response rate. Percentage of people who receive a survey who actually complete it. (11)

response stability. Consistency with which a response is manifested over repeated trials. (32)

responsiveness. The ability of a test to demonstrate change. (10,32)

retrospective study. A study that analyzes observations that were collected in the past. (13,19)

risk. Consideration of physical, psychological, or social harm that goes beyond expected experiences in daily life. (19,34)

risk–benefit ratio. An ethical principle that is an element of informed consent, in which the risks of a research study to the participant are evaluated in relation to the potential benefits of the study's outcomes. (7)

risk factor. A characteristic or exposure that potentially increases the likelihood of having a disease or condition. (34)

risk ratio. (see *relative risk)*

ROC curve. (see *receiver operating characteristic curve*)

run-in period. A time during which all eligible participants receive a placebo prior to implementing an intervention, and only those who are adherent to the protocol are eligible for the formal trial. (14)

sample. Subset of a population chosen for study. (13)

sampling bias. Bias that occurs when individuals who are selected for a sample overrepresent or underrepresent the underlying population characteristics. (13)

sampling distribution. A theoretical frequency distribution of a statistic based on the value of the statistic over an infinite number of samples. (18)

sampling error. The difference between an observed statistic from a sample and the population parameter. (11,13)

saturation. In the ongoing analysis of qualitative data, the point at which no new knowledge contributes to the emerging theory. (21)

scale. An ordered system based on a series of questions or items, resulting in a score that represents the degree to which a respondent possesses a particular attitude, value, or characteristic. (12)

scale of measurement. (see *level of measurement*)

scatter plot. A graphic representation of the relationship between two variables. (29)

Scheffé's multiple comparison test. A multiple comparison procedure for comparing means following a significant analysis of variance based on the *F* distribution. Considered the most conservative of the multiple comparison methods. (26)

scoping review. An exploratory review of a broad question about a clinical topic, including a comprehensive synthesis of evidence with the aim of informing practice, programs, and policy, and providing direction to future research. (5,37)

secondary analysis. An approach to research involving the use of data that were collected for another purpose. (13)

secondary outcome. Other endpoint measures, besides the primary outcome, that may be used to assess the effectiveness of the intervention, as well as side effects, costs, or other outcomes of interest. (3)

secondary source. Reference source that represents a review or report of another's work. (6)

selection bias. A threat to internal validity in which bias is introduced by initial differences between groups, when these differences are not random. (15)

self-report measures. Data based on participants' reports of their own attitudes or conditions. (12)

sensitivity. 1. A measure of validity of a screening procedure, based on the probability that someone with a disease will test positive; true positive rate. (33) 2. In a literature search, the proportion of relevant articles identified out of all relevant articles on that topic; ability to identify all pertinent citations. (6)

sensitivity analysis. A procedure in decision making to determine how decisions change as values are systematically varied. (37)

sequential analysis. In mixed methods research, the design of studies that follow each other, whereby the first study informs the second. (21)

sequential clinical trial. Experimental research design that allows consecutive entrance to a clinical trial and continuous analysis of data, permitting stopping of the trial when data are sufficient to show a significant effect. (16)

serial dependency. Correlation in a set of data collected over time, in which one observation can be predicted based on previous observations. (18)

sham treatment. Analogous to a placebo, involving a "fake" treatment as a control condition. (14)

Shapiro-Wilks test of normality. A test of the null hypothesis that states that an observed distribution does not differ from the pattern of a normal distribution. If the test is not significant ($p>.05$), the distribution does not differ from a normal distribution. (see also *Kolmogorov-Smirnov test of normality*) (23)

sign test. A nonparametric statistical procedure for comparing two correlated samples, based on comparison of positive or negative outcomes. (27)

significance. (see *statistical significance*)

significance level (α). (see *alpha level*)

simple contrast. In a multiple comparison, the contrast of two means to determine if they are significantly different from one another. (26)

simple hypothesis. A hypothesis statement that includes one independent variable and one dependent variable. (3)

single-blind study. An experiment in which either the investigator or the subject is kept ignorant of group assignment, but not both. (14)

single-subject design (SSD). An experimental design based on time-series data from one or more subjects, with data compared across baseline and intervention phases. Also called single-case design. (18)

skewed distribution. A distribution of scores that is asymmetrical, with more scores to one extreme. (22)

slope. 1. In regression analysis, the rate of change in values of Y for one unit of change in X. (30) 2. In single-study research, the rate of change in the magnitude of the target behavior over time. (18)

SnNout. Pneumonic to remember that with high sensitivity, a negative test rules out the diagnosis. (33)

snowball sampling. A nonprobability sampling method in which subjects are successively recruited by referrals from other subjects. (13)

Spearman's rank correlation coefficient (r_s). A nonparametric correlation procedure for ordinal data. Also called Spearman's rho. (29)

specificity. 1. A measure of validity of a screening procedure, based on the probability that someone who does not have a disease will test negative; true negative rate. (34) 2. In a literature search, the proportion of relevant citations retrieved, the ability to exclude irrelevant citations. (6)

sphericity. An assumption in a repeated-measures ANOVA that the variances of the differences between all possible pairs of within-subject conditions (levels of the independent variable) are equal. (25)

split-half reliability. A reliability measure of internal consistency based on dividing the items on an instrument into two halves and correlating the results. (9)

split middle line. In single-subject research, a line used to separate data points within one phase into equal halves, reflecting the trend of the data within that phase. (18)

SpPin. Pneumonic to remember that with high specificity, a positive test rules in the diagnosis. (33)

standard deviation. A descriptive statistic reflecting the variability or dispersion of scores around the mean. (22)

standard error of measurement (SEM). A reliability measure of response stability, estimating the standard error in a set of repeated scores. (9,26)

standard error of the estimate (SEE). In regression analysis, an estimate of prediction accuracy; a measure of the spread of scores around the regression line. (24)

standard error of the mean ($s_{\bar{X}}$). The standard deviation of a distribution of sample means; an estimate of the population standard deviation. (18)

standardized mean difference (SMD). Difference between group means divided by their common standard deviation (analogous to the effect size index, d). (24,37)

standardized residual. Difference between observed and expected values in a chi-square test divided by

the expected frequency, indicating the contribution of each cell to the overall statistic. (28)

standardized response mean (SRM). One approach to evaluating effect size with change scores. Calculated as the difference between pretest and posttest scores, divided by the standard deviation of the change scores. (32)

standardized score. (see *z-score*)

statistic. A measured characteristic of a sample. (13,22)

statistical conclusion validity. The validity of conclusions drawn from statistical analyses, based on the proper application of statistical tests and principles. (15)

statistical hypothesis. (see *null hypothesis*)

statistical process control (SPC). A method of charting production outcomes over time to identify and monitor variances; can be used as a method of analysis for single-subject designs. (18)

stem-and-leaf plot. A graphic display for numerical data in a frequency distribution showing each value in the distribution. (22)

stepwise multiple regression. An approach to multiple regression that involves a sequential process of selecting variables for inclusion in the prediction equation. (30)

stopping rule. In a sequential clinical trial, the threshold for stopping a study based on crossing a boundary that indicates a difference or no difference between treatments. (16)

stratified random assignment. Accomplished by first dividing subjects into strata, and then within each stratum randomly assigning them to groups. (14)

stratified random sampling. Identifying relevant population characteristics, and partitioning members of a population into homogeneous, nonoverlapping subsets, or strata, based on these characteristics. (13)

structural equation modeling (SEM). Multivariate method of modeling theoretical causal relationships of latent variables. (31)

studentized range (*q*). Critical values used to determine significant differences in Tukey and Student-Newman-Keuls multiple comparisons.

Student-Newman-Keuls (SNK) multiple comparison. A stepwise multiple comparison procedure used to determine if means are significantly different following an analysis of variance. Also called the Newman-Keuls (NK) test. (26)

Student's *t*-test. (see *t*-test)

summative scale. A scale that results in a total score by adding values across a set of items. (12)

sum of squares (*SS*). A measure of variability in a set of data, equal to the sum of squared deviation scores for a distribution $[\Sigma(X - \bar{X})^2]$; the numerator in the formula for variance. Used in analysis of variance and other procedures as the basis for partitioning between-groups and within-groups variance components. (22)

superiority trials. Clinical trials that seek evidence in favor of a new treatment. (14)

survival analysis. Methods to analyze data where the variable of interest is the time until occurrence of an outcome event. (31)

systematic error. A form of measurement error, where error is constant across trials. (9)

systematic review. Review of a clearly formulated question that uses systematic and explicit methods to identify, select, and critically appraise relevant research. (1,5,37)

systematic sampling. A sampling method in which persons are randomly chosen from unordered lists using a fixed sampling interval, such as every tenth person. (13)

***t*-test.** A parametric test for comparing two means; also called Student's *t*-test (see *paired t-test* and *unpaired t-test*). (24)

target behavior. In single-subject research, the response behavior that is monitored over time. (18)

target population. The larger population to which results of a study will be generalized, defined by clinical and demographic characteristics. (3,13)

testing effect. The effect that occurs when a test itself is responsible for observed changes in the measured variable. (15)

test–retest reliability. The degree to which an instrument is stable, based on repeated administrations of the test to the same individuals over a specified time interval. (9)

theoretical sampling. In qualitative research, the process of data collection for generating theory by continuous analysis and recruitment of participants to add to themes. (21)

time-series design. (see *interrupted time-series design*)

tolerance. In multiple regression, a measure of collinearity among independent variables. Lower values of tolerance (<.2) indicate greater collinearity. (30)

transferability. In qualitative research, the extent to which findings can be generalized beyond the study. (21)

translational research. Clinical investigations with human subjects in which knowledge obtained from basic research is translated into diagnostic

or therapeutic interventions that can be applied to treatment or prevention. (2)

treatment arm. Another term for independent groups in a clinical trial. (14)

trend. 1. The shape of a distribution of scores taken over time, reflecting the distribution's linearity or lack of linearity. (26) 2. In single-subject research, the direction of change in the target behavior within a phase or across phases. (18)

trend analysis. Using an analysis of variance, a test to assess trend within data taken over ordered intervals; can express data as linear, quadratic, or cubic, reflecting the number of changes in direction in the data over time. (26)

triangulation. The use of multiple methods to document and confirm observations related to a phenomenon. (21)

true negative. A test result that is negative for those who do not have the disease or condition of interest. (33)

true positive. A test result that is positive for those who do have the disease or condition of interest. (33)

truncation. Usually an * used to substitute for word endings that may have varied forms in a literature search. (6)

Tukey's honestly significant difference (HSD). A multiple comparison test for comparing multiple means following a significant analysis of variance. (26)

two standard deviation band method. A method of data analysis in single-subject research; involves calculating the mean and standard deviation of data points within the baseline phase and extending these values into the intervention phase. If two or more consecutive points in the intervention phase fall outside these bands, the change from baseline to intervention is considered significant. (18)

two-tailed test. A statistical test based on a nondirectional alternative hypothesis, in which critical values represent both positive and negative tails of a distribution. (22)

two-way analysis of variance. An analysis of variance with two independent variables. (25)

two-way design. An experimental or quasi-experimental study that involves two independent variables. (16)

Type I error. An incorrect decision to reject the null hypothesis, concluding that a relationship exists when in fact it does not. (23)

Type II error. An incorrect decision to not reject the null hypothesis, concluding that no relationship exists when in fact it does. (23)

unpaired *t*-test. A parametric test for comparing two means for independent samples; also called an independent *t*-test. (24)

unplanned comparisons. Multiple comparison tests that explore all possible comparisons following a significant analysis of variance. (26)

validity. 1. The degree to which an instrument measures what it is intended to measure. (10) 2. The degree to which a research design allows for reasonable interpretations from the data, based on control of bias and confounding (internal validity), appropriate operational definitions (construct validity), appropriate analysis procedures (statistical conclusion validity), and generalizability (external validity). (15)

variable. A characteristic that can be manipulated or observed and that can take on different values, either quantitatively or qualitatively. (3)

variance. A measure of variability in a distribution, equal to the square of the standard deviation. (22)

variance inflation factor (VIF). In multiple regression, a measure of collinearity, the inverse of tolerance. Higher values indicate greater collinearity. (30)

vector. In MANOVA, the combined mean of several dependent variables for each group. (31)

visual analog scale (VAS). An instrument used to quantify a subjective experience by marking a line between two anchors that represent extremes of the experience. A commonly used visual analog scale is a 10 cm line labeled with "no pain" as the left anchor, and "worst pain imaginable" as the right anchor. (11,12)

wait list control group. A control group composed of subjects for whom the experimental treatment is delayed, scheduled to receive the treatment following completion of the study. (14)

Wald statistic. Statistic used to assess significance of coefficients in a logistic regression. (31)

washout period. In a crossover design, that period of time between administration of the two treatments, allowing effects of the experimental treatment to dissipate. (16)

weighted kappa (κ_w). An estimate of percentage agreement, corrected for chance, based on weights reflecting levels of seriousness of disagreements. (32)

wildcard. Usually an * used in a literature search to substitute for several letters within a search term, to allow for alternate spellings. (6)

Wilcoxon signed-ranks test (*T*). A nonparametric statistical procedure, comparing two correlated samples (repeated measures); analogous to the paired *t*-test. (27)

Wilk's lambda (λ). Statistical test of significance in multivariate analysis of variance. (31)

withdrawal design. In single-subject research, a design that involves withdrawal of the intervention. (18)

within-groups variance. (see *error variance*)
within-subjects design (repeated measures design). A design in which subjects act as their own control. (16)
Yates' continuity correction. In the chi-square test, a correction factor applied when expected frequencies are too small, effectively reducing the chi-square statistic. (28)
***z* distribution.** The standardized normal distribution, with a mean of 0 and a standard deviation of 1. (22)
Zelen design. (see *randomized consent design*)
***z*-score.** The number of standard deviations that a given value is above or below the mean of the distribution; also called a standardized score. (22)
zero-order correlation. A bivariate correlation. (29)

Statistical Symbols and Abbreviations

Greek Letters

α	*alpha*	1. Level of significance, denotes risk of Type I error; α_1 and α_2 represent one-tailed and two-tailed levels of significance (23) 2. Also used in Cronbach's alpha. (32)
α_{FW}		Familywise level of significance for a set of multiple comparisons. (26)
α_{PC}		Per comparison level of significance for individual comparisons when multiple comparisons are made. (26)
β	*beta*	1. Probability associated with Type II error. (23) 2. Coefficient in structural equation. (31 Supplement)
γ	*gamma*	Coefficient in structural equation. (31 Supplement)
ε	*epsilon*	Statistic for repeated measures ANOVA, used to adjust degrees of freedom to correct for unequal variances. (25)
η^2	*eta*	Eta-squared, effect size in one-way ANOVA. (25)
η_p^2		Partial eta-squared, effect size for two-way ANOVA. (25)
θ	*theta*	Effect size for proportions, used in sequential clinical trials. (16 Supplement)
κ	*kappa*	Chance-corrected measure of agreement. (32)
κ_w		Weighted kappa. (32 Supplement)
λ	*lambda*	1. A measure of association between categorical variables. (28) 2. Wilks' lambda, multivariate measure of effect. (31)
μ	*mu*	Population mean. (22)
ρ	*rho*	Population measure of correlation. (29)
σ	*sigma*	(lower case) population standard deviation. (22)
σ^2		Sigma-squared, population variance. (22)
$\sigma_{\bar{X}}$		Population standard error of the mean. (22)
Σ	*sigma*	(upper case) used to mean "the sum of." (22)
τ	*tau*	Kendall's tau. (29)
ϕ	*phi*	Phi coefficient measure of association. (28,29)
χ^2	*chi*	Chi-square test of significance for association between categorical variables. (28)
χ_r^2		Chi-square r, test statistic for the Friedman two-way ANOVA by ranks. (27)
ω^2	*omega*	Omega squared, effect size in ANOVA. (25)

Abbreviations

α	Constant (Y-intercept) in a regression equation. (30)
ANOVA	Analysis of variance. (25)
ANCOVA	Analysis of covariance. (30)
ARI	Absolute risk increase. (34)
ARR	Absolute risk reduction. (34)
b	Regression coefficient. (30)
c	Used to denote the number of comparisons in a set of comparisons. (26)
C	1. Contingency coefficient. (28) 2. Number of columns in a contingency table. (28)
CAT	1. Computer adaptive testing. (12) 2. Critically appraised topic. (36)
CER	Control event rate. (34)
CFA	Confirmatory factor analysis. (10,31)
CI	1. Confidence interval. (23) 2. Cumulative incidence. (34)
CV	Coefficient of variation. (22)
d	1. Difference score. (24) 2. Effect size index for the *t*-test, also Cohen's *d*. (24)
$\bar{d}$	Mean of a set of difference scores. (24)
df	Degrees of freedom. (23)
df_b	Degrees of freedom for between-groups source of variance in ANOVA. (25)
df_e	Degrees of freedom for error term (within subjects) in ANOVA. (25)
df_t	Total degrees of freedom. (25)
DOI	Digital object identifier, used to identify research publications. (6,38)
EER	Experimental event rate. (34)
EFA	Exploratory factor analysis. (10,31)
f	1. Frequency of scores. (22) 2. Effect size index for ANOVA. (25)
F	Test statistic for the ANOVA. (25)
H	Test statistic for the Kruskall-Wallis two-way ANOVA by ranks. (27)
H_0	Null hypothesis. (23)
H_1	Alternative hypothesis. (23)
HSD	Tukey's honestly significant difference multiple comparison test. (26)
I^2	Measure of heterogeneity in meta-analysis. (37)
ICC	1. Intraclass correlation coefficient. (32) 2. Item characteristic curve. (12 Supplement)
IQR	Interquartile range. (22)
IR	Incidence rate. (34)
IRT	Item response theory. (12)
ITS	Interrupted time series design. (17)
ITT	Intention-to-treat analysis. (15)
k	Number of groups, means, or predictor variables in an analysis. (25)

LR	Likelihood ratio, can be positive (LR+) or negative (LR–). (33)
MANOVA	Multivariate analysis of variance. (31)
MAR	Missing at random. (15)
MCAR	Missing completely at random. (15)
MCID	Minimal clinically important change. (10,32)
MDC	Minimal detectable change. (9,32)
MeSH	Medical subject headings. (6)
MNAR	Missing not at random. (15)
MS	Mean square. (25)
MS_b	Between-groups mean square. (25)
MS_e	Error mean square. (25)
MTMM	Multitrait-multimethod matrix. (10)
n	Number of subjects in a single sample or group. (22)
N	Number of subjects in a total sample. (22)
NNH	Number needed to harm. (34)
NNT	Number needed to treat. (34)
OR	Odds ratio. (31,34)
p	Probability (23)
P	1. Percentile. (22) 2. Prevalence. (33,34)
PCA	Principal components analysis. (31)
PCOR	Patient-centered outcomes research. (2)
POEM	Patient-oriented evidence that matters. (2)
PROM	Patient-reported outcome measures. (2)
PCT	Pragmatic clinical trial (also *practical clinical trial*). (2,14)
PICO	Elements of an evidence-based question: Population, Intervention, Comparison, Outcome. (3,5)
PV	Predictive value, can be positive (PV+) or negative (PV–). (33)
q	Studentized range statistic. (26)
Q	1. Quartile. (22) 2. Cochrane's Q, a measure of heterogeneity in a meta-analysis. (37)
r	1. Correlation coefficient. (29) 2. Range of ordered means in a multiple comparison test. (26)
r^2	Coefficient of determination. (30)
r_s	Spearman rank correlation coefficient. (29)
$r_{XY \cdot Z}$	First order partial correlation. (29)
r_{pb}	Point biserial correlation coefficient. (29)
r_{rb}	Rank biserial correlation coefficient. (29)
R	1. Multiple regression coefficient. (30) 2. Number of rows in a contingency table. (28) 3. Rank sum in a nonparametric test. (27)
R^2	Multiple correlation coefficient squared. (30)
RCT	Randomized controlled trial. (14)

ROC	Receiver operating characteristic. (33)
RR	Relative risk. (34)
RRR	Relative risk reduction. (34)
s	Sample standard deviation. (22)
s^2	Sample variance. (22)
s_{d}	Standard deviation of a set of difference scores. (24)
$s_{\bar{d}}$	Standard error of the mean of a set of difference scores. (24)
$s_{\bar{X}}$	Standard error of the mean for a sample. (23)
$s_{\bar{X}_1-\bar{X}_2}$	Standard error of the difference between means. (24)
SEE	Standard error of the estimate. (30)
SEM	1. Standard error of measurement. (9,32) 2. Structural equation modeling. (31)
SMD	Standardized mean difference. (24,37)
Sn	Sensitivity. (33)
SNK	Student-Newman-Keuls multiple comparison test. (26)
Sp	Specificity. (33)
SPC	Statistical process control. (18)
SRM	Standardized response mean. (32)
SS	Sum of squares. (22,25)
SS_{b}	Between-groups sum of squares. (25)
SS_{e}	Error sum of squares. (25)
SS_{t}	Total sum of squares. (25)
SSD	Single-subject design. (18)
t	Test statistic for comparison of two means. (24)
T	Test statistic for the Wilcoxon signed-ranks test. (27)
U	Test statistic for the Mann-Whitney U test. (27)
V	Cramer's V coefficient. (28)
$\bar{V}$	Mean vector in MANOVA. (31)
VAS	Visual analog scale. (12)
VIF	Variance inflation factor. (30)
x	Test statistic for the sign test and binomial test. (18,27)
X	1. Single score. (22) 2. Symbol for an independent variable in a regression equation. (30)
$\bar{X}$	Sample mean. (22)
Y	Symbol for a dependent variable in a regression equation. (30)
$\bar{Y}$	Mean for a sample of scores on variable Y. (30)
$\hat{Y}$	Predicted score in a regression equation (Y-hat). (30)
z	Standardized score. (22,23)