



The Grammar of Science: Lost in Statistics? A Practical Roadmap for Biomedical and Health Research

Jaranit Kaewkungwal¹^{*}

Faculty of Tropical Medicine, Mahidol University, Thailand

^{*}Corresponding author, email address: jaranitk@biophics.org

Received: 15 Aug 2026; Revised 9 Sep 2026; Accepted 12 Sep 2026

<https://doi.org/10.59096/osir.v19i3.284849>

A regular reader of this column came up to me the other day and said, “I’ve learned so many statistical methods and models, but now I feel lost in the forest of statistics.” One of my students also told me that statistics had been taught in bits and pieces across different classes, and he could not quite put everything together. They asked, “Could you summarize the common statistical methods in one place and at least guide us on when to use which one?” I have heard similar questions from my other students before: “Statistics is fun to learn, but is there one book that brings all these methods together?” That is a big question. I may not be able to cover everything, but I’ll try my best.

To answer a research question, we collect, manage, and analyze data. Data are generally quantitative or qualitative. Quantitative data are numerical and may be continuous, such as temperature or height, or discrete, such as number of children or heartbeats.^{1,2} Count data are non-negative discrete values representing how often an event occurs or how many objects are observed within a specified time or space.³ Qualitative, or categorical, data may be binary (e.g., no/yes), multinomial with unordered categories (e.g., heart attack/death/stroke), or ordinal with ordered categories (e.g., good/fair/poor/no change). Numbers assigned to categorical data are codes, not numerical measurements.^{2,4,5} Another important kind of data is time-to-event data, which record the time until an event occurs while accounting for participants who remain event-free during follow-up.¹

Research data are called variables because their values vary across individuals or within the same individual over time. The dependent variable is usually the outcome of interest, whereas independent

variables are factors that may influence or predict it and may be called exposures, risk factors, predictors, inputs, or interventions.⁶ Studies may also include confounding variables (confounders), which can distort the observed relationship between an independent variable and an outcome, making the association appear stronger, weaker, absent, or even reversed.⁶

We analyze data using different statistical methods depending on the research question and study objective. Common purposes include describing the main features of the data, making inferences from a sample to a population, comparing outcomes between groups, assessing relationships between variables, identifying risk factors, predicting future events or trends, and supporting clinical decision-making.⁷

But there are many statistical methods to choose from, and sometimes several methods can be used for the same purpose. This article aims to bring these methods together and provide a practical guide to selecting the method that best fits the analytic objective. I group the methods into eight broad analytic purposes. For each group, we will look at the objective, statistical methods, typical predictors (if any), key assumptions, and main statistics to report.

Descriptive Statistics

Descriptive statistics summarize the characteristics of a study sample, whereas inferential statistics generalize findings from the sample to a broader population.^{2,8} A numerical measure from a sample is a statistic, while the corresponding population measure is a parameter.⁹ Confidence intervals estimate a range likely to contain the true population parameter.² Each statistical method has specific assumptions.



Descriptive statistics for continuous data typically include measures of central tendency and spread. Central tendency represents the center of a distribution, whereas spread indicates the variability among observations.² The choice of statistics depends on whether the data follow a normal distribution, also known as a Gaussian distribution. For normally distributed data, the mean and standard deviation are commonly reported; for non-normally distributed data, the median and interquartile range (IQR) are preferred. The range may also be reported and does not require a specific distributional assumption.^{2,5}

Categorical variables are typically described using frequencies and proportions. Frequency is the number of observations in a category, while proportion is the number in that category divided by the total number of observations in the study sample.² Depending on the study design and time frame, related epidemiological

measures include prevalence, incidence proportion, incidence rate, and odds. A rate describes how quickly an event occurs in a defined study group over time, often using person-time as the denominator.^{10,11}

Count data may be described similarly to continuous data, using measures such as the mean and standard deviation or the median and IQR, or as a rate when the number of events is observed over varying periods of exposure, using person-time as the denominator.³ Time-to-event data include individuals who experience the event and those who do not experience it during the follow-up period, referred to as censored observations. The Kaplan–Meier method describes the probability of remaining event-free or experiencing the event over time.¹¹ Median survival time, the time point at which the survival probability falls to 50%, is often reported alongside the Kaplan–Meier curve, provided sufficient events have occurred to reach that threshold (Table 1).

Table 1. Descriptive analysis.

Objective	Statistical method	Typical predictor(s)	Key assumptions	Main statistics to report
CONTINUOUS OUTCOME				
Describe normal/symmetric distribution data	Mean, SD, Variance	None	Approximately symmetric; no extreme outliers	Mean, SD
Describe non-normal/skewed distribution data	Median, IQR	None	No distributional assumption	Median, IQR
BINARY OUTCOME				
Describe a binary outcome	Proportion	None	Clearly defined denominator; independent observations	Proportion, Percentage
Describe a binary outcome (existing cases)	Proportion	None	Defined population and time point/period	Prevalence
Describe a binary outcome (new cases)	Proportion	None	Population at risk; defined follow-up period	Incidence proportion
Describe a binary outcome (new cases) per follow-up time	Incidence rate	Exposure time	Population at risk; accurately measured person-time	Incidence rate
Describe odds of having outcome vs. not having outcome	Odds	None	Clearly defined binary outcome; independent observations	Odds
ORDINAL OUTCOME				
Describe an ordinal outcome	Frequency	None	Mutually exclusive, ordered categories	Frequencies, Percentages
NOMINAL (MULTINOMIAL) OUTCOME				
Describe nominal outcome categories	Frequency	None	Mutually exclusive categories	Frequencies, Percentages
COUNT OUTCOME				
Describe counts of events (approximately symmetric data)	Mean, SD, Variance	None	Approximately symmetric; no extreme outliers	Mean count, SD
Describe counts of events (skewed data)	Median, IQR	None	No distributional assumption	Median count, IQR
Describe counts of events per follow-up time	Event rate	Exposure time	Accurately measured event counts and person-time	Event rate

Table 1. Descriptive analysis (cont.).

Objective	Statistical method	Typical predictor(s)	Key assumptions	Main statistics to report
TIME-TO-EVENT (SURVIVAL TIME)				
Describe time to event, accounting for censoring	Kaplan-Meier (KM) estimate/curve	None	Events/censoring times accurately recorded; censoring is non-informative (independent of event risk)	Survival probabilities at time points; survival curve
Summarize typical time to event	Median survival time (from KM)	None	Sufficient events observed for median cross 0.5 survival)	Median survival time

SD: Standard deviation. IQR: interquartile range.

Comparisons between Groups

When comparing an outcome among groups, two hypotheses are specified: the null hypothesis (e.g., H_0 : the outcomes in Groups 1 and 2 are equal) and the alternative hypothesis (e.g., H_a : the outcomes differ). Statistical hypothesis testing uses observed data to determine whether to reject or not reject H_0 . A Type I error occurs when a true H_0 is rejected, whereas a Type II error occurs when a false H_0 is not rejected. The probability of a Type I error is controlled by the prespecified significance level (α), and the p -value is compared with α to make the decision. Among tests using the same α , the preferred test is the one more likely to detect a true difference; that is, the test with greater statistical power.^{5,9}

For continuous outcomes, the choice of statistical method depends partly on assumptions about the distribution of the parameter or outcome in the underlying population. Parametric methods make assumptions about population parameters and their distributions, commonly including an approximately normal distribution, homogeneity of variance across groups, and independence of observations. Classical parametric tests for comparing independent groups include the independent-samples t -test and analysis of variance (ANOVA).^{5,12–14}

Nonparametric methods, sometimes called distribution-free methods, make fewer assumptions about population parameters and their distributions. They are useful for ordinal or ranked data, skewed continuous data, small samples, or data with influential outliers. These methods generally compare ranks or distributions rather than means; examples include the Mann–Whitney U test, Wilcoxon signed-rank test, Kruskal–Wallis test, and Friedman test. Data distributions may be assessed using Q – Q plots, histograms, and tests such as the Shapiro–Wilk test.^{15,16}

The appropriate test also depends on the study design. Independent groups require between-subject comparisons, whereas pre–post and repeated-measures designs require methods that account for correlated observations from the same individuals. Examples include the paired-samples t -test for two time points and repeated-measures ANOVA for multiple time points. Designs involving both group and time comparisons may use two-way repeated-measures ANOVA.^{17,18} Repeated-measures ANOVA assumes sphericity, meaning that the variances of the differences between all pairs of repeated outcome measurements over time are equal. If this assumption is violated, corrections such as Greenhouse–Geisser or Huynh–Feldt are applied to the degrees of freedom, thereby adjusting the p -value of the F -test.¹⁷

Researchers may also compare outcomes between groups while adjusting for other variables that could influence the outcome, such as confounders. These variables are included in the analysis as covariates. When one or more quantitative continuous are included as covariates in an ANOVA model, the method is called analysis of covariance (ANCOVA).¹⁹ ANCOVA is itself a form of regression model. When several covariates or more flexible adjustments are needed, multiple regression is commonly used, as discussed in the next section.

When groups are compared across multiple related outcomes measured at the same time (such as physical, mental, and social health), multivariate analysis of variance (MANOVA) may be used. MANOVA extends ANOVA by analyzing the outcomes jointly. Its extension that also adjusts for covariates is called multivariate analysis of covariance (MANCOVA).²⁰

For categorical outcomes, the chi-square test is commonly used to compare proportions among groups. It is based on a cross-tabulation of the group and outcome variables and assesses whether the observed cell frequencies differ from those expected under the

null hypothesis of no association. Pearson's chi-square test is commonly used for independent groups. When expected cell frequencies are small, Fisher's exact test is preferred. For paired binary outcomes, such as measurements before and after an intervention in the same individuals, McNemar's test is used.^{21,22}

For count outcomes, comparisons between groups are commonly based on the Poisson distribution. Poisson methods compare observed event counts or rates with those expected under the model, rather than comparing raw counts alone. A key assumption is that the mean and variance of the count outcome are equal. When this assumption is violated and the variance exceeds the mean (overdispersion), negative binomial methods are used instead. When observation times differ, event rates

are compared by accounting for each group's time at risk. Methods include the Poisson rate test and related Wald or exact tests. When adjustment for covariates is required, Poisson regression may be applied.²³

Regardless of the outcome type, comparisons among three or more groups usually begin with an overall (omnibus) test. A significant result indicates only that at least one group differs from another, not that all groups differ. Post-hoc pairwise comparisons may then identify the differing groups, with adjustment for multiplicity (the increased risk of false-positive results from multiple tests). Common approaches include the Bonferroni method for controlling the family-wise error rate and the Benjamini–Hochberg method for controlling the false discovery rate (Table 2).^{15,16,18}

Table 2. Comparisons between groups or conditions.

Objective	Statistical method	Typical predictor(s)	Key assumptions	Main statistics to report
CONTINUOUS OUTCOME				
Compare one sample with a reference value	One-sample t-test	None	Independent observations; approximately normal outcome	Mean difference
Compare two independent groups	Independent t-test	Binary	Approximately normal; independent; equal variance	Mean difference
Compare two independent groups nonparametrically	Mann–Whitney U, or Wilcoxon rank-sum test	Binary	Independent groups; similarly shaped distributions	Rank difference (Median difference only if distributions equally shaped)
Compare paired observations	Paired t-test	Time or paired condition	Approximately normal paired differences	Mean difference
Compare paired observations nonparametrically	Wilcoxon signed-rank test	Time or paired condition	Paired observations; approximately symmetric differences	Rank or median difference
Compare more than two independent groups	One-way ANOVA	Categorical	Approximately normal; independent; equal variance	Mean differences, F statistic
Compare more than two groups nonparametrically	Kruskal–Wallis test	Categorical	Independent observations; similar distribution shapes	Rank differences
Compare repeated measurements	Repeated-measures ANOVA	Time	Approximately normal residuals; sphericity (apply Greenhouse-Geisser/ Huynh-Feldt correction if violated)	Mean differences, F statistic
Compare repeated measurements nonparametrically	Friedman test	Time	Matched observations; ordinal or continuous outcome	Rank differences
Compare groups while adjusting for covariates	ANCOVA	Group and covariates	Linear covariate effect; equal slopes; normal residuals; equal variance	Adjusted mean difference
Compare groups across multiple correlated outcomes	MANOVA	Group	Multivariate normality; equal covariance matrices; independence	Wilks' λ , Pillai's trace
Compare multiple outcomes while adjusting for covariates	MANCOVA	Group and covariates	MANOVA assumptions; linear covariate effects; equal slopes	Adjusted multivariate effect

Table 2. Comparisons between groups or conditions (cont.).

Objective	Statistical method	Typical predictor(s)	Key assumptions	Main statistics to report
BINARY OUTCOME				
Compare one proportion with a reference value	One-sample proportion test	None	Binomial distribution; independent observations	Difference in proportion
Compare two independent groups	Chi-square test	Categorical	Independent observations; Expected cell counts ≥ 5 in $\geq 80\%$ of cells	Association; crude RD, PR, RR, or OR can be calculated from the crosstabulation table
Compare two independent groups with small samples	Fisher's exact test	Categorical	Independent observations; suitable for small expected counts	Association; crude RD, PR, RR, or OR can be calculated from the crosstabulation table
Compare paired binary outcomes	McNemar test	Time or paired condition	Paired observations; binary outcome	Matched proportion difference
Compare more than two independent groups	Chi-square test	Categorical	Independent observations; expected cell counts ≥ 5 in $\geq 80\%$ of cells	Overall association
Compare groups across strata while controlling for a categorical confounder	Mantel–Haenszel test	Stratified categorical predictors	Similar effect across strata	Common OR or RR
Assess trend across ordered groups	Cochran–Armitage trend test	Ordered categorical predictor	Ordered groups; independent observations	Trend statistic
ORDINAL OUTCOME				
Compare two groups	Mann–Whitney U test	Binary group	Independent observations; ordinal outcome; similar distribution shapes	Rank difference
Compare more than two groups	Kruskal–Wallis test	Categorical group	Independent observations; similar distribution shapes	Rank difference
Compare paired ordinal observations	Wilcoxon signed-rank test	Time or paired condition	Paired observations; approximately symmetric differences	Rank difference
Compare repeated ordinal observations	Friedman test	Time or repeated condition	Matched repeated observations; ordinal outcome	Rank difference
NOMINAL (MULTINOMIAL) OUTCOME				
Compare proportions across nominal categories	Chi-square test	Categorical group	Independent observations; expected cell counts ≥ 5 in $\geq 80\%$ of cells	Overall association
COUNT OUTCOME				
Compare counts or event rates	Poisson test (or Poisson regression if adjusting for covariates)	Group or exposure-time variable	Independent Poisson events; mean=variance (use negative binomial if overdispersed)	Rate ratio or rate difference

ANOVA: analysis of variance. ANCOVA: analysis of covariance. MANOVA: multivariate analysis of variance. MANCOVA: multivariate analysis of covariance. RD: risk difference. PR: prevalence ratio. RR: risk ratio (relative risk). OR: odds ratio.

Association and Regression Analysis

Two variables are associated when variability in one can be partly explained by the other.⁵ Correlation measures the strength and direction of association between two variables: positive values indicate movement in the same direction, whereas negative values indicate movement in opposite directions.^{5,9,24} However, correlation does not establish causality. A causal relationship occurs when changes in an independent variable cause changes in a dependent variable.^{5,25,26} Variables may be strongly correlated without being causally related, or causally related without being correlated.⁵

Several correlation coefficients can be selected according to the nature of the data.²⁷ Pearson's correlation (r) measures the linear relationship between two continuous variables that are approximately bivariate normally distributed. Spearman's rank correlation (ρ or r_s) is a nonparametric alternative that measures a monotonic (consistently increasing or decreasing), but not necessarily linear, relationship using ranked data. Other coefficients include Kendall's tau (τ) and the phi coefficient (ϕ). These coefficients range from -1 to $+1$: 0 indicates no association, while $+1$ and -1 indicate perfect positive and negative associations, respectively.

Regression analysis examines the relationship between a dependent variable and one independent variable (simple regression) or several independent variables (multiple regression).^{9,25,26} The same regression equation can serve different purposes, depending on the research objective and how its results are interpreted: (1) predicting average or future outcome values from specified predictors; (2) exploring associations between an outcome and potential risk factors; or (3) testing whether a specific risk factor is associated with the outcome after adjusting for potential confounders.

Linear, logistic, and Poisson regression belong to the generalized linear model (GLM) family and share a common linear predictor, $(\beta_0 + \beta_1 X_1 + \dots)$.^{24,28,29} However, they link this predictor to the outcome differently. Predictors may be continuous or categorical, whereas the regression model is selected mainly according to the outcome type: continuous for linear, binary or categorical for logistic, and count for Poisson regression.

When the outcome is continuous and approximately normally distributed, linear regression can be used to fit a straight-line relationship: $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \epsilon$. Here, β_0 is the intercept, each β represents the expected change in Y for a one-unit increase in its predictor while holding the others constant, and ϵ is the error term.²⁴ However, relationships between variables are not always straight and may follow curved or nonlinear patterns. These patterns can be identified using scatterplots, smoothers such as LOESS, and residual-versus-fitted plots.³⁰ They may be modeled using polynomial terms, including quadratic terms for U-shaped or inverted-U relationships; a log-linear transformation (regressing the logarithm of Y on X) for growth or decline at a constant percentage rate; or logarithmic functions for rapid initial change that gradually slows. Although polynomial regression models curves, it remains a form of linear regression because it is linear in its coefficients.

When the dependent variable is categorical, logistic regression may be used. Binary logistic regression is appropriate for an outcome with two categories, usually coded 0 and 1 , and models the natural logarithm of the odds that $Y=1$: $\text{Logit}(P_y) = \ln(P_y / (1 - P_y)) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n$, where P_y is the probability that the outcome occurs and $\ln$ denotes the natural logarithm. Each β represents the change in the log-odds for a one-unit increase in the predictor, holding other predictors constant. Exponentiating the coefficient (e^β) gives the corresponding odds ratio (OR).^{28,31} Multinomial logistic regression is used for three or more unordered categories, whereas ordinal logistic regression is used for three or more ordered categories.

When the dependent variable is count data (non-negative integers) assumed to follow a Poisson distribution, Poisson regression models the natural logarithm of the expected value of the outcome: $\ln(\text{expected value of } Y) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n$. Exponentiating β (e^β) gives an expected count ratio or, when an exposure-time offset is included, an incidence rate ratio (IRR). Standard Poisson regression assumes equal mean and variance; negative binomial regression accommodates overdispersion; zero-inflated Poisson regression handles excess zeros; zero-truncated Poisson regression applies when zeros cannot occur or are excluded; and hurdle regression separately models zero versus positive counts and the distribution of positive counts (Table 3).^{29,32,33}

Table 3. Association and regression modelling

Objective	Statistical method	Typical predictor(s)	Key assumptions	Main statistics to report
CONTINUOUS OUTCOME				
Assess association between two continuous variables	Pearson correlation	Continuous	Linear relationship; approximate normality; no major outliers	Pearson's r
Assess a non-normal or monotonic association	Spearman correlation	Continuous or ordinal	Monotonic relationship; paired observations	Spearman's ρ
Predict a continuous outcome using one predictor	Simple linear regression	One predictor	Linearity; independent errors; constant variance; normal residuals	Unadjusted β
Predict a continuous outcome using multiple predictors	Multiple linear regression	Multiple predictors	Linearity; independent errors; constant variance; normal residuals; limited collinearity	Adjusted β
Determine whether an effect differs by another variable	Linear regression with interaction	Multiple predictors	Linear-model assumptions; correctly specified interaction	Interaction β
Assess mediation	Linear mediation analysis	Exposure, mediator and covariates	Correct causal model; no unmeasured confounding	Direct and indirect effects (β)
Model nonlinear relationships	Polynomial regression, (or other nonlinear models)	Continuous	Appropriate polynomial form; linear-model assumptions	Polynomial β coefficients
Model flexible nonlinear effects	Restricted cubic spline regression	Continuous	Sufficient data; appropriate number and placement of knots	Spline estimates or predicted effect curve
BINARY OUTCOME				
Estimate the effect of one predictor	Simple logistic regression	Continuous or categorical	Independent observations; linearity of the logit	Odds ratio (OR)
Estimate adjusted effects of multiple predictors	Multiple logistic regression	Multiple	Linear logit; limited collinearity; enough outcome events	Adjusted OR
Analyse matched case-control data	Conditional logistic regression	Matched case-control predictors	Correct matching; independent matched sets	Matched OR
Analyse rare outcomes or sparse data	Exact logistic regression	Multiple	Sparse data; correctly specified model	OR
Reduce overfitting from correlated predictors	Penalized logistic regression: LASSO, Ridge or Elastic Net	Multiple	Appropriate penalty tuning; predictor scaling when needed	Penalized OR
Model nonlinear predictor effects	Logistic regression with splines	Continuous	Adequate observations across the predictor range; appropriate knots	Adjusted OR or predicted probabilities
Assess mediation	Logistic mediation analysis	Exposure, mediator and covariates	Correct causal model; no unmeasured confounding	Direct and indirect effects (OR)
Determine whether an effect differs by another variable	Logistic regression with interaction	Multiple	Correct interaction specification; logistic-model assumptions	Interaction OR

Table 3. Association and regression modelling (cont.).

Objective	Statistical method	Typical predictor(s)	Key assumptions	Main statistics to report
ORDINAL OUTCOME				
Assess monotonic association	Spearman correlation	Ordinal or continuous	Monotonic relationship; paired observations	Spearman's ρ
Model predictors of an ordinal outcome	Ordinal logistic regression	Continuous and/or categorical	Proportional odds; independent observations; no severe multicollinearity	Common OR
Handle partial proportional-odds violations	Partial proportional odds model	Continuous and/or categorical	Proportional odds holds for some but not all predictors	Predictor-specific OR
NOMINAL (MULTINOMIAL) OUTCOME				
Predict nominal outcome categories	Multinomial logistic regression	Continuous and/or categorical	Independent observations; limited collinearity; IIA when applicable	Relative risk ratio (RRR)
Classify nominal outcome groups	Linear Discriminant Analysis	Multiple continuous predictors	Multivariate normality; equal covariance matrices; independence	Classification accuracy; posterior probability
Classify groups with unequal covariance matrices	Quadratic Discriminant Analysis	Multiple continuous predictors	Multivariate normality; group-specific covariance matrices; adequate sample size	Classification accuracy; posterior probability
COUNT OUTCOME				
Model count outcomes	Poisson regression	Continuous and/or categorical; exposure offset if needed	Equidispersion; independent observations; correct log-linear relationship	Incidence rate ratio (IRR)
Model overdispersed counts	Negative binomial regression	Continuous and/or categorical; exposure offset if needed	Overdispersion; independent observations; correct mean structure	IRR
Model excess zero counts	Zero-inflated Poisson regression, or Negative binomial (if overdispersion)	Predictors for count and zero components	Two latent processes; structural zeros; Poisson or negative binomial count component (depending on choice)	Count-model IRR; zero-model OR
Model no zero count	Zero-truncated Poisson, or Negative binomial (if overdispersion)	Predictors for count outcome (zero values excluded)	No zero values in the data; correct mean structure for $y > 0$	Count-model IRR
Model zero and positive counts separately	Hurdle Poisson or negative binomial model	Predictors for hurdle and positive-count components	Separate zero and positive-count processes	Hurdle-model OR; positive-count IRR

OR: odds ratio. LASSO: least absolute shrinkage and selection operator. RRR: relative risk ratio. IRR: incidence rate ratio.

Repeated-measures, Longitudinal or Clustered Data Analysis

The terms multivariate and multivariable are often used interchangeably in the literature, although this distinction is commonly overlooked. Strictly, multivariate analysis involves multiple outcome variables, whereas multivariable analysis examines one outcome using multiple predictor variables.³⁴ Longitudinal data analysis applies to data collected over time. When observations are repeatedly collected from the same individuals, it is called repeated-measures data analysis.^{35,36} Because observations from the same individual may be correlated, it may also be called correlated data analysis.

In repeated-measures data, measurement times or conditions are nested within each individual. In clustered data, individuals are nested within clusters, such as patients within hospitals or students within schools. Both have hierarchical structures in which observations are not independent and are therefore forms of correlated data.^{37,38} Consequently, repeated-measures and clustered data can often be analyzed using similar statistical methods. Several techniques are available for analyzing these data structures.

As previously mentioned in the section on comparisons between groups, repeated-measures ANOVA is used for a normally distributed continuous outcome measured in the same individuals under three or more conditions or at multiple time points. Its sphericity assumption should be checked and, if violated, an appropriate correction should be applied. Significant overall results may be followed by post hoc tests, such as Bonferroni-adjusted pairwise comparisons.³⁹

Multilevel mixed models analyze repeated-measures and clustered data and can accommodate multiple clustering levels, such as repeated measurements within patients nested within clinics. They include fixed effects, representing the general relationship between predictors or exposures and outcomes, and random effects, representing subject- or cluster-specific deviations. Thus, they provide both a general equation and specific equations for individual subjects or clusters. The two classic forms are random-intercept models, allowing different starting outcome levels, and random-slope models, allowing the relationship between predictors or exposures and outcomes to differ across subjects or clusters. Mixed models extend linear, logistic, and Poisson regression from one independent outcome observation per individual to multiple correlated outcome observations.^{40–42}

Generalized estimating equations (GEE) similarly extend linear, logistic, and Poisson regression to multiple correlated outcome observations. Unlike mixed models, GEE uses a working correlation matrix (an assumed pattern of correlation among repeated outcomes) to account for correlations among outcomes rather than modeling fixed and random effects. It provides one population-average equation, rather than subject- or cluster-specific equations, and generally accommodates one clustering level. Common working correlation structures include independence, exchangeable, autoregressive, and unstructured. Correct specification improves efficiency, while robust standard errors can support valid inference when the correlation structure is misspecified.^{43–45}

While mixed models and GEE examine predictor–outcome relationships, other methods for repeated categorical outcomes focus on states or groups and how individuals move between them over time. A Markov model estimates the probability of moving between observed states, such as from mild to severe, or from severe to recovered. A first-order Markov model assumes that the next state depends only on the current state, not on earlier history.⁴⁶ Latent transition analysis (LTA) estimates how individuals remain in or move between unobserved groups over time.⁴⁷ It is based on latent class analysis (LCA, discussed later), which identifies these unobserved groups from patterns across several categorical indicators.⁴⁸

Time-series data are observations recorded sequentially over time, often aggregated into regular intervals (e.g., weekly dengue case counts). Because nearby observations are often correlated (autocorrelation), specialized methods are needed to examine trends, forecast future values, or evaluate interventions.^{2,49} Exponential smoothing forecasts by giving more weight to recent values and can account for trend and seasonality. Interrupted time-series regression estimates whether an intervention produces an immediate level change or a change in trend. ARIMA combines autoregression (AR: prediction from past values), integration (I: differencing the data to remove trends to achieve stationarity), and moving average (MA: prediction using past forecast errors). SARIMA extends ARIMA to model seasonality (regular patterns that repeat daily, monthly, or yearly), whereas ARIMAX adds external predictors (X variables) that may influence the outcome; for example, rainfall and temperature may help explain or forecast weekly dengue case counts (Table 4).^{49,50}

Table 4. Repeated-measures, longitudinal or clustered data.

Objective	Statistical method	Typical predictor(s)	Key assumptions	Main statistics to report
CONTINUOUS OUTCOME				
Compare changes over time within and/or between groups	Repeated-measures ANOVA	Time; group; (Time × Group interaction)	Continuous outcome; independent subjects; approximately normal residuals; sphericity	F statistic, partial η^2 ; adjusted means
Estimate subject- or cluster-specific effects	Linear mixed-effects model	Fixed and random effects	Correct random-effects/covariance structure; approximate normality	Subject-specific β
Estimate population-average effects	GEE with Gaussian distribution	Time, treatment and covariates	Correct mean model; appropriate working correlation; independent clusters	Population-average β
BINARY OUTCOME				
Estimate subject- or cluster-specific effects	Mixed-effects logistic regression	Multiple predictors and random effects	Correct random-effects structure; conditional independence	Subject-specific OR
Estimate population-average effects	GEE logistic regression	Time, treatment, time × treatment and covariates	Appropriate correlation structure; correct mean model; adequate clusters	Population-average OR
Estimate subject- or cluster-specific incidence-rate effects	Mixed-effects Poisson regression with log (person-time) offset	Time, treatment, covariates, exposure time and random effects	Correctly specified person-time offset; conditional equidispersion; correct random-effects structure	Subject- or cluster-specific IRR
Estimate cluster-specific risk or prevalence ratios (for cross-sectional study)	Mixed-effects modified Poisson regression with robust SE, or mixed-effects log-binomial regression	Individual- and cluster-level predictors plus random effects; no time offset	Binary outcome; correct random-effects structure; conditional independence; robust SE for modified Poisson	Cluster-specific RR or PR
Estimate population-average incidence-rate effects	GEE; Poisson regression with log(person-time) offset	Time, treatment, covariates and exposure time	Correct mean model and offset; appropriate working correlation; independent clusters; adequate number of clusters	Population-average IRR
Estimate population-average risk or prevalence ratios	Modified Poisson GEE with robust SE, or GEE log-binomial regression	Time, treatment and covariates; no time offset	Binary outcome; correct mean model; appropriate working correlation; independent clusters; adequate number of clusters	Population-average RR or PR
ORDINAL OUTCOME				
Estimate subject-specific effects for repeated ordinal outcomes	Mixed ordinal logistic regression	Time, treatment, covariates and random effects	Proportional odds; correct random-effects structure; conditional independence	Subject-specific OR
Estimate population-average effects for correlated ordinal outcomes	GEE ordinal logistic regression	Time, treatment and covariates	Proportional odds; appropriate working correlation; sufficiently large sample	Population-average OR

Table 4. Repeated-measures, longitudinal or clustered data (cont.).

Objective	Statistical method	Typical predictor(s)	Key assumptions	Main statistics to report
NOMINAL (MULTINOMIAL) OUTCOME				
Estimate subject-specific effects for clustered nominal outcomes	Mixed multinomial logistic regression	Individual- and cluster-level predictors plus random effects	Correct random-effects structure; conditional independence within clusters	Subject-specific RRR
Estimate population-average effects for correlated nominal outcomes	GEE multinomial regression	Time, treatment and covariates	Appropriate working correlation; sufficiently large sample	Population-average RRR
COUNT OUTCOME				
Model clustered counts	Mixed-effects Poisson regression	Individual- and cluster-level predictors plus random effects	Equidispersion conditional on random effects; correct random-effects structure	Subject-specific IRR
Model clustered overdispersed counts	Mixed-effects negative binomial regression	Individual- and cluster-level predictors plus random effects	Overdispersion; correct random-effects structure; conditional independence	Subject-specific IRR
Model correlated counts	GEE Poisson or negative binomial regression	Time, treatment, exposure and covariates	Appropriate working correlation; correct mean model; independent clusters; robust standard errors	Population-average IRR
Model clustered counts with excess zeros (mixture of structural and sampling zeros)	Zero-inflated mixed-effects Poisson/negative binomial model	Predictors for the zero and count components plus random effects	Defensible mixture of structural and count-generating zero sources; correct count distribution and random-effects structure	Count-component IRR; zero-component OR or probability
Model clustered counts with excess zeros (single zero-generating process)	Hurdle mixed-effects Poisson/negative binomial model	Predictors for the zero and count components plus random effects	All zeros arise from one binary process; positive counts follow a zero-truncated distribution; correct random-effects structure	Count-component IRR (zero-truncated); zero-component OR or probability
GENERAL LONGITUDINAL CATEGORICAL OUTCOMES				
Model transitions between states	Markov model	Current state, previous state, time, treatment and covariates	Markov property; correctly specified transition structure	Transition probabilities or transition OR/RR
Identify longitudinal latent classes	Latent Transition Analysis	Repeated categorical indicators and covariates	Local independence; measurement invariance when assumed; correct number of classes	Class and transition probabilities
TIME-SERIES MODELS				
Estimate a linear change in the outcome over time	Linear trend regression	Time	Correct functional form; independent or appropriately modeled errors	β coefficient for time
Estimate individual or group growth trajectories over time	Growth curve model	Time, treatment, covariates, and a subject/cluster ID for repeated measures	Correct mixed-model and covariance structure; appropriate growth form; normal residuals/random effects	Initial status, growth-rate parameters, variance components
Forecast future values using past values and errors	ARIMA	Time-ordered outcome series (equally spaced)	Stationary after differencing; correct AR/MA orders; residuals near white noise	AR coefficients, MA coefficients, forecasts
Forecast time-series values while adjusting for external predictors	ARIMAX	Time-ordered outcome series and external predictor(s), equally spaced	Stationary after differencing; correct ARIMA form; valid external predictors	β coefficients, AR/MA parameters, forecasts

ANOVA: analysis of variance. SE: standard error. GEE: generalized estimating equation. PR: prevalence ratio. RR: risk ratio. IRR: incidence rate ratio. OR: odds ratio. RRR: relative risk ratio. AR: autoregression. MA: moving average.

Time-to-event Analysis

Time-to-event data differ from continuous data because they consider both whether and when an event occurs and may include censored observations. Censored individuals remain event-free at their last follow-up but still contribute information until that time. Therefore, specialized methods are needed to analyze event timing and censoring.^{51,52} As a descriptive statistic, the Kaplan–Meier method estimates the survival function, $S(t)$, the probability of remaining event-free beyond time t , while the log-rank test compares survival functions between groups.^{11,53}

The Cox proportional hazards model is a classic semiparametric regression method for time-to-event data. It models the hazard function, $h(t)$, the instantaneous risk of an event at time t among individuals who remain event-free: $h(t|X) = h_0(t) \times \exp(\beta_1 X_1 + \beta_2 X_2 + \dots)$. The baseline hazard, $h_0(t)$, is unspecified, while exponentiated coefficients (e^b) are interpreted as hazard ratios (HRs). For example, an HR of two for treatment means that, at any given time, the treated group has twice the hazard of experiencing the event compared with the reference group, among individuals who are still event-free and holding other covariates constant. This does not mean twice the overall probability of experiencing the event. The key assumption is proportional hazards: the HR remains constant over time.^{54,55}

In contrast, parametric survival models assume a specific distribution for survival times: exponential for a constant hazard; Weibull for a hazard that increases or decreases over time; Gompertz for a hazard that increases or decreases exponentially with time; and log-normal or log-logistic for a non-monotonic hazard that may rise and then fall. These models are useful for extrapolating survival beyond observed data and describing the hazard shape. Many can also be formulated as accelerated failure time (AFT) models, in which the effect measure is a time ratio indicating whether a predictor accelerates or decelerates time to the event, rather than a hazard ratio.⁵⁶ However, parametric models are used less often than the Cox model.^{51,55}

Standard time-to-event analysis generally considers only one type of event. A competing risk is an event that prevents the primary event from occurring; for example, non-cardiovascular death prevents subsequent cardiovascular death.⁵⁷ Two common regression approaches are the cause-specific hazard model, which examines factors affecting the event rate among individuals currently event-free, and the Fine–Gray subdistribution hazard model, which examines factors affecting the cumulative incidence, the actual probability of the event over time in the presence of competing events. The Fine–Gray model is generally recommended when the goal is individual risk prediction or prognosis, whereas the cause-specific hazard model is preferred for understanding etiology, since it more directly reflects a covariate's biological effect on the event rate.^{57–59}

When an individual experiences the same event repeatedly over time, the data are called recurrent-event data. Ignoring these events loses information, while events within the same individual are correlated. Several models, often called multiple-failure-time models, are available depending on the research assumptions.^{60–62} The Andersen–Gill model extends Cox regression and assumes that recurrent events are conditionally independent given the covariates. The Prentice–Williams–Peterson (PWP) model accounts for event order and may reset time after each event. The Wei–Lin–Weissfeld (WLW) model analyzes each event number as a separate marginal outcome. Frailty models include person-specific random effects to capture differences in event susceptibility. Rate or mean-function models, such as Poisson, negative binomial, and Lin–Wei–Yang–Ying models, focus on event frequency or burden.

Multistate survival analysis extends standard time-to-event analysis by modeling transitions through a sequence of health or disease states over time. Instead of one endpoint, individuals may follow state sequences such as Healthy → Relapse → Death or Diagnosis → Remission → Relapse → Death. The model estimates transition probabilities between states and state-occupation probabilities, such as the probability of being alive and in remission at time t (Table 5).⁶³

Table 5. Time-to-event analysis.

Objective	Statistical method	Typical predictor(s)	Key assumptions	Main statistics to report
DESCRIPTIVE SURVIVAL ANALYSIS				
Describe survival over time	Kaplan–Meier estimator	None	Independent, non-informative censoring; clear time origin and event	Survival probability; median survival
COMPARISON BETWEEN GROUPS				
Compare survival between two or more groups	Log-rank test	Categorical group	Independent censoring; similar censoring patterns; greatest power under proportional hazards	χ^2 ; p -value
Compare groups when proportional hazards assumption is violated	Restricted mean survival time (RMST)	Categorical group	Independent, non-informative censoring; pre-specified truncation time τ	RMST difference or ratio; p -value
Compare groups with greater emphasis on early events	Gehan–Breslow–Wilcoxon test	Categorical group	Independent censoring; valid group comparison	χ^2 ; p -value
Compare groups with intermediate weighting of earlier events	Tarone–Ware test	Categorical group	Independent censoring; valid group comparison	χ^2 ; p -value
COX'S PROPORTIONAL HAZARD REGRESSION MODELS				
Estimate the effect of predictor(s) on time to event	Cox proportional hazards regression	Continuous or categorical predictor(s)	Proportional hazards; independent censoring; correct functional form	Hazard ratio (HR)
Estimate subject- or cluster-specific effects	Mixed-effects or frailty Cox model	Multiple predictors and random effects	Proportional hazards conditional on frailty; appropriate frailty distribution and random-effects structure	Subject-specific HR; frailty variance
Estimate marginal effects for correlated survival outcomes	Marginal Cox model with robust sandwich SE	Individual- and cluster-level predictors	Proportional hazards; independent clusters; correct model specification; adequate number of clusters	Marginal or population-average HR
Allow different baseline hazards across strata	Stratified Cox regression	Multiple predictors and stratification variable	Proportional hazards for estimated predictors within strata; unrestricted baseline hazard across strata	HR
Include predictors that change over time	Extended Cox model with time-varying covariates	Fixed and time-dependent predictors	Correctly measured time-dependent predictors; proportional hazards for fixed effects	HR
Model effects that change over follow-up time	Cox model with predictor $\times$ time interaction	Multiple predictors and time interactions	Correct specification of time interaction; independent censoring	Time-varying HR
PARAMETRIC AND FLEXIBLE SURVIVAL MODELS				
Model survival assuming a constant hazard	Exponential regression	Multiple predictors	Constant baseline hazard; independent censoring; correct model specification	HR or time ratio
Model monotonic increasing or decreasing hazards	Weibull regression	Multiple predictors	Weibull survival distribution; independent censoring	HR or time ratio
Model hazards that change exponentially with time	Gompertz regression	Multiple predictors	Gompertz hazard distribution; independent censoring	HR
Model survival times with a log-normal distribution	Log-normal regression	Multiple predictors	Log-normal survival-time distribution; independent censoring	Time ratio

Table 5. Time-to-event analysis (cont.).

Objective	Statistical method	Typical predictor(s)	Key assumptions	Main statistics to report
Model survival with a non-monotonic hazard pattern	Log-logistic regression	Multiple predictors	Log-logistic survival-time distribution; independent censoring	Time ratio
Model complex baseline hazards flexibly	Royston–Parmar flexible parametric model	Multiple predictors	Smooth baseline log cumulative hazard or related scale; appropriate spline knots	HR, survival difference, or time-varying effect
Estimate how predictors accelerate or delay event time	Accelerated Failure Time model	Multiple predictors	Correct survival-time distribution; multiplicative effect on survival time	Time ratio or acceleration factor
Estimate the probability of long-term cure or immunity from the event	Cure model	Multiple predictors	Mixture or non-mixture cure structure; identifiable cured fraction; sufficient follow-up	Cure probability and HR or time ratio
Model hazards that are constant within intervals but vary across intervals	Piecewise exponential regression	Multiple predictors	Constant hazard within pre-specified intervals; independent censoring	HR
RECURRENT-EVENT ANALYSIS (MULTIPLE FAILURES MODELS)				
Model repeated events using a counting-process approach	Andersen–Gill model	Multiple predictors	Proportional hazards; no unmeasured within-subject dependence beyond covariates; robust SE for within-subject correlation	HR
Model ordered recurrent events conditionally on prior events	Prentice–Williams–Peterson model	Multiple predictors	Ordered events; subjects enter risk set after prior event; proportional hazards	Event-specific HR
Estimate marginal effects for each recurrent-event order	Wei–Lin–Weissfeld model	Multiple predictors	Marginal event-specific processes; robust variance for within-subject dependence	Event-specific marginal HR
Model the total number of events during follow-up	Poisson or negative binomial regression	Multiple predictors and exposure time	Correct count distribution; appropriate exposure offset; independent or appropriately clustered observations	Incidence rate ratio (IRR)
COMPETING-RISK ANALYSIS				
Estimate the effect of predictors on a specific cause of failure	Cause-specific Cox regression	Multiple predictors	Proportional cause-specific hazards; competing events treated as censored for cause-specific modelling	Cause-specific HR
Estimate the effect of predictors on cumulative incidence	Fine–Gray subdistribution hazards model	Multiple predictors	Proportional subdistribution hazards; independent censoring	Subdistribution HR
MULTI-STATE AND CENSORED-DATA MODELS				
Model movement through multiple clinical or disease states	Multi-state model	Multiple predictors	Correct transition structure; Markov or semi-Markov assumption as specified	Transition probability; transition HR
Analyse events known only to occur within a time interval	Interval-censored survival regression	Multiple predictors	Correct interval definition; appropriate parametric or semiparametric model; independent censoring mechanism	HR or time ratio

SE: standard error.

Factor Analysis and Path/Causal Models

In research, variables may be observed (manifest) or latent. Observed variables are measured directly, such as test scores, survey responses, or ratings. Latent variables are unobservable theoretical constructs (e.g., quality of life, anxiety, intelligence, attitude, or

motivation) that are inferred from patterns across several related observed indicators rather than measured by a single question.⁶⁴

Factor analysis is a multivariate technique that identifies latent factors explaining the correlations among observed variables. There are two main types.

Exploratory factor analysis (EFA) allows the data to reveal the underlying structure; for example, 10 medication-nonadherence questionnaire items may form two factors: intentional and unintentional nonadherence. Confirmatory factor analysis (CFA) tests a prespecified structure; for example, whether 15 items represent three health domains (physical, mental, and social health) with five items per domain. CFA can also be used to assess the construct validity of research instruments.^{65–67}

Path analysis and structural equation modeling (SEM) examine networks of relationships among multiple variables simultaneously, extending beyond simple regression with one outcome and its predictors.⁶⁸ Predictors are called exogenous variables, whereas variables predicted within the model are endogenous variables. For example, demographics such as sex, age, and illness may predict anxiety and quality of life (QoL), while anxiety may also predict QoL. The model therefore estimates direct paths: demographics → anxiety, demographics → QoL, and anxiety → QoL, as well as an indirect path: demographics → anxiety → QoL. Note that a variable can be endogenous with respect to one path and exogenous with respect to

another: here, anxiety is endogenous relative to demographics (since demographics predict it) but exogenous relative to QoL (since it predicts QoL), meaning anxiety also functions as a mediating variable between demographics and QoL.

Path analysis uses only observed (measured) variables, without latent constructs, and therefore does not separate measurement error from the constructs of interest. In the example above, questionnaire items are summed to obtain anxiety and QoL scores, and all variables in the model are treated as observed variables.

While path analysis uses only directly measured (observed) variables, SEM includes both observed variables and latent variables, the latter inferred from multiple observed indicators. SEM has two components: a measurement model, which uses CFA to link latent constructs to their observed indicators, and a structural model, which uses path analysis to specify the hypothesized relationships among the latent constructs. In the example above, anxiety and QoL are treated as latent constructs measured by questionnaire items, while their relationships are examined through the structural model (Table 6).^{68–70}

Table 6. Factor analysis, path/causal models, and time-series analysis.

Objective	Statistical method	Typical data, or predictor(s)	Key assumptions	Main statistics to report
FACTOR ANALYSIS				
Reduce multiple correlated variables into fewer components	Principal component analysis (PCA)	Multiple variables	Adequate correlation structure; variables measured on compatible scales	Components, component loadings, eigenvalues
Identify underlying latent constructs	Exploratory factor analysis (EFA)	Multiple variables	Adequate correlations; sufficient sample size; appropriate extraction and rotation	Factor loadings, communalities, eigenvalues
Confirm a hypothesized latent structure	Confirmatory factor analysis (CFA)	Multiple observed indicators	Prespecified model; adequate sample size; appropriate estimator; acceptable model fit	Standardized loadings, CFI, TLI, RMSEA, SRMR
PATH AND CAUSAL MODELS				
Estimate linked direct and indirect effects among observed variables	Path analysis (observed variables only)	Observed exposures, mediators, outcomes and covariates	Prespecified causal order; linear relations; independent errors; adequate sample	Standardized path coefficients; direct, indirect and total effects
Estimate structural relationships among observed and latent variables	Structural equation modeling (SEM)	Multiple observed and latent variables	Correctly specified and identified model; adequate sample; suitable distribution	Path coefficients, direct and indirect effects, model-fit indices
Estimate how an exposure affects an outcome through a mediator	Linear mediation or path analysis	Exposure, mediator, outcome, and covariates	Correct time order; no major unmeasured confounding; correct models	Direct effect, indirect effect, total effect
Determine whether a predictor effect differs by a moderator	Linear regression with interaction	Predictor, moderator, and predictor × moderator	Linear-model assumptions; correctly specified interaction; adequate data across moderator levels	Interaction β coefficient

CFI: comparative fit index. TLI: Tucker–Lewis index. RMSEA: root mean square error of approximation. SRMR: standardized root mean square residual.

Classification and Clustering Analysis Methods

Researchers may use several variables to determine which group an individual belongs to. When groups are predefined (e.g., infected vs. uninfected or treatment success vs. failure), supervised methods predict group membership. When groups are not defined in advance (e.g., identifying patient subtypes from clinical or laboratory profiles), unsupervised methods uncover hidden subgroups based on patterns in the data.

Supervised classification methods include logistic regression, discriminant analysis, K-nearest neighbors (KNN), classification trees, and random forests. Logistic regression can estimate the probability that an individual belongs to a binary or categorical group from one or more predictors. For binary outcomes, the individual is then assigned to a group using a chosen probability threshold, commonly 0.5, although another threshold may be selected. For outcomes with more than two categories, individuals are typically assigned to the group with the highest predicted probability.⁷¹

Discriminant analysis separates known groups using the predictor variables. Parametric approaches include linear discriminant analysis (LDA) and quadratic discriminant analysis (QDA), which generally assume multivariate normality within groups. LDA finds the linear combination of variables that best separates groups, while QDA allows a more flexible, curved decision boundary by permitting each group its own

covariance structure. Nonparametric classifiers do not require this assumption of multivariate normality. The KNN method classifies an individual based on the majority class among its k nearest neighbors, where k is the specified number of neighbors considered. Classification trees split observations into groups using a series of if-then decision rules on the predictors. Random forests aggregate predictions from many classification trees to improve accuracy and stability.^{72–74}

Unsupervised approaches include cluster analysis and model-based mixture methods. Hierarchical clustering progressively combines or separates observations and displays the resulting structure in a dendrogram. K-means clustering requires the number of clusters, k , to be specified and assigns each observation to the nearest cluster center. Both methods seek high similarity within clusters and low similarity between clusters.⁷⁵

Model-based clustering assumes that the data arise from a mixture of probability distributions and estimates each individual's probability of belonging to each subgroup. The number of subgroups is commonly selected using model-fit criteria such as the Akaike information criterion (AIC) and Bayesian information criterion (BIC). Latent class analysis (LCA) uses categorical indicators, latent profile analysis (LPA) uses continuous indicators, and mixed-indicator mixture models accommodate both categorical and continuous indicators (Table 7).^{76–78}

Table 7. Classification and clustering analysis methods.

Objective	Statistical method	Typical data, or predictor(s)	Key assumptions	Main statistics to report
GROUP MEMBERSHIP KNOWN: CLASSIFICATION				
Predict membership in predefined groups	Logistic regression: binary, multinomial, or ordinal	Continuous and/or categorical predictors; known outcome groups	Independent observations; correct model form; no severe multicollinearity	ORs with 95% CIs, predicted probabilities, accuracy, sensitivity, specificity, and AUC
Classify observations into predefined groups (parametric)	LDA; QDA	Continuous and/or categorical predictors; known group membership	LDA: multivariate normality and equal covariance matrices across groups; QDA: multivariate normality (covariance matrices allowed to differ across groups)	Confusion matrix, accuracy, sensitivity, specificity, AUC, and cross-validated performance
Classify observations into predefined groups (non-parametric)	KNN; classification tree; random forest	Continuous and/or categorical predictors; known group membership	KNN: appropriate scaling and distance metric; trees/RF: appropriate complexity control (e.g., pruning, depth, number of trees) to avoid overfitting	Confusion matrix, accuracy, sensitivity, specificity, AUC, and cross-validated performance

Table 7. Classification and clustering analysis methods (cont.).

Objective	Statistical method	Typical data, or predictor(s)	Key assumptions	Main statistics to report
GROUP MEMBERSHIP UNKNOWN: CLUSTERING				
Identify naturally occurring groups	Cluster analysis: hierarchical, k-means	Continuous, categorical, or mixed variables	Appropriate scaling and similarity measure; suitable number of clusters	Cluster number and size, centroids, dendrogram, and silhouette coefficient
Identify unobserved groups from response patterns	LCA	Mainly categorical indicators; optional covariates	Local independence; appropriate number of classes	Class proportions, item-response probabilities, entropy, AIC, and BIC
Identify unobserved profiles from continuous response patterns	LPA	Continuous indicators; optional class-membership covariates	Conditional normality and local independence; appropriate number of profiles	Profile proportions, indicator means and variances, posterior probabilities, entropy, AIC, and BIC
Identify unobserved groups from mixed response patterns	Mixed-indicator mixture model	Both categorical and continuous indicators; optional covariates	Appropriate distribution for each indicator; local independence; appropriate number of groups	Group proportions, category-response probabilities, indicator means and variances, posterior probabilities, entropy, AIC, and BIC

OR: odds ratio. CI: confidence interval. AUC: area under the curve. LDA: linear discriminant analysis. QDA: quadratic discriminant analysis. KNN: K-nearest neighbors. AIC: Akaike information criterion. BIC: Bayesian information criterion (also called the Schwarz information criterion). LCA: latent class analysis. LPA: latent profile analysis.

Reliability and Validity of Research Instruments

Because measurement tools (e.g., scales, questionnaires, and diagnostic tests) generate study data, unreliable or invalid tools can directly compromise study findings.⁷⁹ Measurement quality depends mainly on reliability and validity, and instrument development aims to minimize measurement error.^{80,81} Reliability refers to the consistency or reproducibility of measurements and includes internal consistency, stability over time and equivalence across raters. Validity refers to whether an instrument measures its intended construct and includes face, content, criterion (concurrent or predictive), and construct validity (convergent or discriminant).

Internal consistency, also called homogeneity reliability, can be assessed using Cronbach's alpha or the Kuder–Richardson formulas. Cronbach's alpha is the most widely used measure of internal consistency for multi-item scales, particularly those with Likert-type or continuous items. It is based on the variance of individual items relative to the variance of the total score. Alpha increases when items are more strongly correlated, but may also increase simply by adding more items. Kuder–Richardson Formula 20 (KR-20) is a special case of Cronbach's alpha for dichotomous items, such as right/wrong or yes/no, and allows item difficulty to vary. Kuder–Richardson Formula 21 (KR-21) is a simpler version that assumes

approximately equal item difficulty but is less accurate when this assumption is violated.^{82,83}

For the reliability of continuous measurements, scores may be obtained by measuring the same individuals on different occasions (test–retest stability) or by having different raters assess the same individuals (inter-rater equivalence). Pearson's correlation or the intraclass correlation coefficient (ICC) may be used. Pearson's correlation assesses only whether two sets of scores vary together linearly, whereas the ICC assesses their consistency or agreement and is therefore generally preferred. A one-way model is used when individuals are assessed by different randomly selected raters, whereas a two-way model is used when all individuals are assessed by the same raters or undergo the same measurement on different occasions. A two-way random-effects model allows generalization beyond the raters studied, whereas a two-way mixed-effects model applies only to the specific raters or measurement occasions included. For two-way models, ICC forms further differ according to whether reliability is based on a single measurement or the average of several measurements and whether consistency or absolute agreement is required.⁸⁴

Correlation can be misleading when assessing agreement because two methods or measurements may be highly correlated despite constant or proportional systematic bias. For example, if a new device always

reads 10 mmHg higher than the reference device at every level of blood pressure, the two would correlate almost perfectly ($r \approx 1$) even though they never agree. Correlation shows whether measurements vary together, whereas Bland-Altman analysis quantifies how much they differ and whether the disagreement is consistent, making it commonly used to determine whether a new measurement method agrees sufficiently with an established method to replace it, such as two laboratory assays measuring the same substance in blood.⁸⁵

For inter-rater reliability of categorical measurements, commonly used statistics include percent agreement and the kappa statistic. Kappa assesses the extent to which raters assign observations to the same categories. Unlike simple percent agreement, which may be inflated by chance agreement, kappa adjusts the observed agreement for agreement expected by chance.⁸⁶ For continuous measurements, inter-rater reliability is generally assessed using the ICC, as described above.

Content validity is commonly assessed during questionnaire development to ensure that the items adequately cover the intended content. It uses expert judgment rather than statistical patterns in respondent data. The item-level content validity index (I-CVI) evaluates the relevance of each item, while the scale-level content validity index (S-CVI) evaluates the overall scale. The content validity ratio (CVR) uses a different calculation to assess whether an individual item is considered essential.⁸⁷

Construct validity assesses whether the items represent the intended abstract theoretical construct, such as anxiety or quality of life. It can be evaluated using confirmatory factor analysis (CFA), which uses respondent data to test the hypothesized factor structure mentioned in the previous section. Confirmation of the proposed construct should be reported using standardized factor loadings, composite reliability, and model-fit indices, including chi-square (χ^2), comparative fit index (CFI), Tucker–Lewis index (TLI), root mean square error of approximation (RMSEA), and standardized root mean square residual (SRMR).

For criterion validity in diagnostic testing, a new test is compared with a gold standard or valid reference standard, which may not be perfect but is the best practicable method available.⁸⁸ The comparison produces true-positive (TP), true-negative (TN), false-positive (FP), and false-negative (FN) results. Sensitivity and specificity measure diagnostic accuracy relative to the reference standard, while positive and negative predictive values indicate the probability of having or not having the condition given the test result.^{89,90} When a new test yields a continuous score rather than a binary result, a cut-off must be chosen to classify individuals as positive or negative. For quantitative tests, a receiver operating characteristic (ROC) curve plots sensitivity against (1–specificity) across possible cut-offs, and the area under the curve (AUC) indicates how well the test distinguishes individuals with and without the condition (Table 8).⁹¹

Table 8. Reliability and validity of research instruments.

Objective	Statistical method	Typical data	Key assumptions	Main statistics to report
RELIABILITY				
Assess internal consistency	Cronbach's alpha	Multi-item scale with continuous or Likert-type items	Items measure the same construct; errors are not strongly correlated	Alpha coefficient
Assess internal consistency of binary items	KR-20	Dichotomous items scored 0/1	Items measure one construct; item difficulties may differ	KR-20 coefficient
Estimate binary-item consistency using mean difficulty	KR-21	Dichotomous items scored 0/1	Items have roughly similar difficulty; stronger assumptions than KR-20	KR-21 coefficient
Assess reliability of continuous measurements	ICC	Continuous scores from repeated measurements or raters	Appropriate ICC model selected (e.g., two-way random effects, absolute agreement vs. consistency); stable subjects; similar measurement conditions	ICC
Assess agreement between two continuous measurements	Bland–Altman analysis	Two methods or repeated continuous measurements	Differences are approximately normal; no trend in differences	Mean bias and limits of agreement

Table 8. Reliability and validity of research instruments (cont.).

Objective	Statistical method	Typical data	Key assumptions	Main statistics to report
Assess agreement between two raters for categories	Cohen's kappa (weighted kappa for ordinal categories)	Two raters; nominal or ordinal categories	Independent subjects; same categories used by both raters	Kappa coefficient
Assess agreement among multiple raters	Fleiss' kappa	Three or more raters; categorical ratings	Independent subjects; comparable rating process	Kappa coefficient
VALIDITY / DIAGNOSTIC ACCURACY				
Assess content validity of questionnaire	Content validity index (CVI); CVR	Expert panel ratings of item relevance/ clarity (typically 3–10 experts, ordinal relevance scale (e.g., 1–4))	Experts are qualified in the content domain; adequate panel size; items rated independently	I-CVI, S-CVI; CVR per item
Confirm construct of the questionnaire	CFA	Multi-item scale with continuous, ordinal, or Likert-type items; adequate sample size (commonly ≥ 200 or ≥ 10 cases per item)	Prespecified factor structure based on theory; adequate model fit; no major cross-loadings or correlated errors; sufficient sample size relative to model complexity	Model fit indices (χ^2 , CFI, TLI, RMSEA, SRMR); standardized factor loadings; composite
Assess ability to detect true positives	Sensitivity	Test result compared with a reference standard	Valid reference standard; representative participants	Sensitivity = $TP/(TP+FN)$
Assess ability to detect true negatives	Specificity	Test result compared with a reference standard	Valid reference standard; representative participants	Specificity = $TN/(TN+FP)$
Assess probability that a positive test is correct	Positive predictive value (PPV)	Positive test results	Depends on disease prevalence in the study population	$PPV = TP/(TP+FP)$
Assess probability that a negative test is correct	Negative predictive value (NPV)	Negative test results	Depends on disease prevalence in the study population	$NPV = TN/(TN+FN)$
Assess overall discrimination across cut-offs	ROC curve and AUC	Continuous or ordinal test compared with a reference standard	Independent observations; valid reference standard	AUC, optimal cut-off, sensitivity and specificity

KR-20: Kuder-Richardson formula 20. KR-21: Kuder-Richardson formula 21. ICC: intraclass correlation coefficient. CVR: content validity ratio. I-CVI: item-level content validity index. S-CVI: scale-level content validity index. CFA: confirmatory factor analysis. CFI: comparative fit index. TLI: Tucker–Lewis index. RMSEA: root mean square error of approximation. SRMR: Standardized root mean square residual. TP: true-positive. TN: true-negative. FP: false-positive. FN: false-negative. PPV: positive predictive value. ROC: receiver operating characteristic curve. AUC: area under the curve.

Key Take-away Message

In this article, I have summarized the classic approach to statistical methods and mapped out some paths through this statistical forest. However, the wider forest also includes other approaches, such as Bayesian statistics and machine learning. Each group of methods serves a different purpose. To choose the right one, we first need to know what we want to answer and what kind of data we have. Nowadays, AI can help us select a method, perform the analysis, and even interpret the results, but statistical knowledge remains essential. We must still decide whether the AI-suggested method is appropriate, understand why

it was selected, and interpret and use the results correctly. AI can suggest a path, but we must understand where it leads.

Acknowledgements

An artificial intelligence tool, ChatGPT (OpenAI, 2026), Claude (Anthropic, 2026), Gemini (Google, 2026) were used for reference searching and language support during the preparation of this manuscript.^{92–94} All generated content was carefully reviewed for accuracy and appropriateness. The author revised and edited the material and assumes full responsibility for the final content.

References

- Hefni AE. Types of Data used in Medical Statistics [Internet]. Sunnyvale (CA): LinkedIn; 2023 Aug 9 [cited 2026 Aug 10]. <<https://www.linkedin.com/pulse/types-data-used-medical-statistics-abdallah-el-hefni>>
- Australia Bureau of Statistics. Statistical terms and concepts glossary [Internet]. Canberra: Australia Bureau of Statistics; [cited 2026 Aug 10]. <<https://www.abs.gov.au/statistics/understanding-statistics/statistical-terms-and-concepts/statistical-terms-and-concepts-glossary>>
- Too MA, Alter U, Flora DB. Modelling count data in psychological research: an applied tutorial. *Int J Psychol*. 2025 Apr;60(2):e70018. doi:10.1002/ijop.70018.
- Cheatham ML. A practical guide to biostatistics [Internet]. Orlando (FL): Surgicalcriticalcare.net; c2024. Chapter two: describing and using data; [cited 2026 Aug 10]; p. 9–12. <<https://www.surgicalcriticalcare.net/Statistics/data.pdf>>
- Stark PB. Glossary of statistical terms [Internet]. Berkeley: University of California; [cited 2026 Aug 10]. <<https://www.stat.berkeley.edu/~stark/SticiGui/Text/gloss.htm>>
- Network of the National Library of Medicine. Common Terms and Equations [Internet]. Bethesda (MD): National Library of Medicine; [cited 2026 Aug 10]. <<https://www.nlm.nih.gov/oet/ed/stats/02-200.html>>
- Sterrantino AF. Observational studies: practical tips for avoiding common statistical pitfalls. *Lancet Reg Health Southeast Asia*. 2024 May 9;25:100415. doi:10.1016/j.lansea.2024.100415.
- Coursera Staff. Descriptive vs. inferential statistics: what's the difference? [Internet]. Mountain View (CA): Coursera Inc.; [updated 2026 Apr 25; cited 2026 Aug 10]. <<https://www.coursera.org/articles/descriptive-vs-inferential-statistics>>
- Otto T. 100 english terms for statistics [Internet]. London: Speak Tech English; 2025 Dec 8 [cited 2026 Aug 10]. <<https://speaktechenglish.com/100-english-terms-for-statistics-2026/>>
- Centers for Disease Control and Prevention (US). Lesson 3: measures of risk [Internet]. Atlanta: Centers for Disease Control and Prevention; 2012 May 18 [cited 2026 Aug 10]. <https://archive.cdc.gov/www_cdc_gov/csels/ds_epd/ss1978/lesson3/section1.html>
- Kaplan EL, Meier P. Nonparametric Estimation from Incomplete Observations. *J Am Stat Assoc*. 1958;53(282):457–81. doi:10.1080/01621459.1958.10501452.
- Hoskin T. Parametric and nonparametric: demystifying the terms [Internet]. Scottsdale (AZ): Mayo Clinic; [cited 2026 Aug 10]. 5 p. <<https://www.mayo.edu/research/documents/p-parametric-and-nonparametric-demystifying-the-terms/doc-20408960>>
- du Prel JB, Rohrig B, Hommel G, Blettner M. Choosing statistical tests: part 12 of a series on evaluation of scientific publications. *Dtsch Arztebl Int*. 2010 May;107(19):343–8. doi:10.3238/arztebl.2010.0343.
- Ranganathan P. An Introduction to Statistics: Choosing the Correct Statistical Test. *Indian J Crit Care Med*. 2021 May;25(Suppl 2):S184–6. doi:10.5005/jp-journals-10071-23815.
- Nahm FS. Nonparametric statistical tests for the continuous data: the basic concept and the practical use. *Korean J Anesthesiol*. 2016 Feb;69(1):8–14. doi:10.4097/kjae.2016.69.1.8.
- Ledolter J, Gramlich OW, Kardon RH. Parametric statistical inference for comparing means and variances. *Invest Ophthalmol Vis Sci*. 2020 Jul 1;61(8):25. doi:10.1167/iovs.61.8.25.
- Muhammad LN. Guidelines for repeated measures statistical analysis approaches with basic science research considerations. *J Clin Invest*. 2023 Jun 1;133(11):e171058. doi:10.1172/JCI171058.
- Langenberg B, Janczyk M, Koob V, Kliegl R, Mayer A. A tutorial on using the paired t test for power calculations in repeated measures ANOVA with interactions. *Behav Res Methods*. 2023 Aug;55(5):2467–84. doi:10.3758/s13428-022-01902-8.
- Schwarz W. The Ancova model for comparing two groups: a tutorial emphasizing statistical distribution theory. *Front Psychol*. 2025 May 15;16:1600764. doi:10.3389/fpsyg.2025.1600764.
- Buderer NM, Brannan GD. Comparing the Means of Independent Groups: ANOVA, ANCOVA, MANOVA, and MANCOVA [Internet]. Treasure

- Island (FL): StatPearls Publishing; [updated 2024 May 19. cited 2026 Aug 10]. <<https://www.ncbi.nlm.nih.gov/books/NBK606084/>>
21. Siegel S, Castellan NJ. Nonparametric statistics for the behavioral sciences. 2nd ed. New York: McGraw-Hill; 1988.
 22. Agresti A. An introduction to categorical data analysis. 3rd ed. Chichester: Wiley; 2018. 400 p.
 23. Luo R. Hypothesis testing of Poisson rates in COVID-19 offspring distributions. *Infect Dis Model*. 2023 Aug 9;8(4):980–1001.
 24. Schober P, Vetter TR. Linear Regression in Medical Research. *Anesth Analg*. 2021 Jan;132(1):108–9. doi:10.1213/ANE.00000000000005206.
 25. Australian Bureau of Statistics. Correlation and causation [Internet]. Canberra: Australian Bureau of Statistics; [cited 2026 Aug 10]. <<https://www.abs.gov.au/statistics/understanding-statistics/statistical-terms-and-concepts/correlation-and-causation>>
 26. Cote C. What is regression analysis in business analytics? [Internet]. Boston: Harvard Business School; 2021 [cited 2026 Aug 10]. <<https://online.hbs.edu/blog/post/what-is-regression-analysis>>
 27. Schober P, Boer C, Schwarte LA. Correlation coefficients: appropriate use and interpretation. *Anesth Analg*. 2018 May;126(5):1763–8. doi:10.1213/ANE.0000000000002864.
 28. Zabor EC, Reddy CA, Tendulkar RD, Patil S. Logistic Regression in Clinical Studies. *Int J Radiat Oncol Biol Phys*. 2022;112(2):271–7. doi:10.1016/j.ijrobp.2021.08.007.
 29. Green JA. Too many zeros and/or highly skewed? A tutorial on modelling health behaviour as count data with Poisson and negative binomial regression. *Health Psychol Behav Med*. 2021; 9(1):436–55. doi:10.1080/21642850.2021.1920416.
 30. Jarantow SW, Pisors ED, Chiu ML. Introduction to the use of linear and nonlinear regression analysis in quantitative biological assays. *Curr Protoc*. 2023;3(6):e801. doi:10.1002/cpz1.801.
 31. Wallisch C, Bach P, Hafermann L, Klein N, Sauerbrei W, Steyerberg EW, et al. Review of guidance papers on regression modeling in statistical series of medical journals. *PLoS One*. 2022 Jan 24;17(1):e0262918. doi:10.1371/journal.pone.0262918.
 32. Cox S, West SG, Aiken LS. The analysis of count data: a gentle introduction to poisson regression and its alternatives. *J Pers Assess*. 2009 Mar;91(2):121–36. doi:10.1080/00223890.802634175.
 33. Feng CX. A comparison of zero-inflated and hurdle models for modeling zero-inflated count data. *J Stat Distrib Appl*. 2021;8(1):8. doi:10.1186/s40488-021-00121-4.
 34. Hidalgo B, Goodman M. Multivariate or multivariable regression? *Am J Public Health*. 2013 Jan;103(1):39–40. doi:10.2105/AJPH.2012.300897.
 35. Diggle P, Heagerty P, Liang KY, Zeger S. Analysis of Longitudinal Data. 2nd ed. New York: Oxford University Press; 2013.
 36. Muhammad LN. Guidelines for repeated measures statistical analysis approaches with basic science research considerations. *J Clin Invest*. 2023 Jun 1;133(11):e171058. doi:10.1172/JCI171058.
 37. Hair JF, Favero LP. Multilevel modeling for longitudinal data: concepts and applications. *RAUSP Management Journal*. 2019;54(4):459–89. doi:10.1108/RAUSP-04-2019-0059.
 38. Chang YH, Buras MR, Davis III JM, Crowson CS. Avoiding blunders when analyzing correlated data, clustered data, or repeated measures. *J Rheumatol*. 2023;50(10):1269–71. doi:10.3899/jrheum.2022-1109.
 39. Blanca MJ, Arnau J, Garcia-Castro FJ, Alarcon R, Bono R. Repeated measures ANOVA and adjusted *F*-tests when sphericity is violated: which procedure is best? *Front Psychol*. 2023 Aug 30;14:1192453. doi:10.3389/fpsyg.2023.1192453.
 40. Madden LV, Ojiambo PS. The value of generalized linear mixed models for data analysis in the plant sciences. *Front Hortic*. 2024;3:1423462. doi:10.3389/fhort.2024.1423462.
 41. Silveira LTYD, Ferreira JC, Patino CM. Mixed-effects model: a useful statistical tool for longitudinal and cluster studies. *J Bras Pneumol*. 2023 May 15;49(2):e20230137. doi:10.36416/1806-3756/e20230137.
 42. Schober P, Vetter TR. Linear Mixed-Effects Models in Medical Research. *Anesth Analg*.

- 2021 Jun 1;132(6):1592–3. doi:10.1213/ANE.0000000000005541.
43. Liang KY, Zeger SL. Longitudinal data analysis using generalized linear models. *Biometrika*. 1986 Apr;73(1):13–22. doi:10.1093/biomet/73.1.13.
 44. Wang M. Generalized estimating equations in longitudinal data analysis: a review and recent developments. *Advances in Statistics*. 2014;2014(1):303728. doi:10.1155/2014/303728.
 45. Mahalingam V, Singh R, Santhanakrishnan RK, Bhaskar A, Chinnaiyan P. Generalized estimating equations for modeling cluster randomized trial data on smoking cessation among tuberculosis patients. *PLoS One*. 2025 Oct 10;20(10):e0333992. doi:10.1371/journal.pone.0333992.
 46. Ooi QX, Plan E, Bergstrand M. A tutorial on pharmacometric Markov models. *CPT Pharmacometrics Syst Pharmacol*. 2025 Feb;14(2):197–216. doi:10.1002/psp4.13278.
 47. Moore EWG, Quartiroli A, Little TD. Introduction to the best practice recommendations for longitudinal latent transition analysis. *Int J Psychol*. 2025 Apr;60(2):e70021. doi:10.1002/ijop.70021.
 48. Sorgente A, Caliciuri R, Robba M, Lanz M, Zumbo BD. A systematic review of latent class analysis in psychology: examining the gap between guidelines and research practice. *Behav Res*. 2025;57(301). doi:10.3758/s13428-025-02812-1.
 49. Jebb AT, Tay L, Wang W, Huang Q. Time series analysis for psychological research: examining and forecasting change. *Front Psychol*. 2015 Jun 9;6:727. doi:10.3389/fpsyg.2015.00727.
 50. Bernal JL, Cummins S, Gasparrini A. Interrupted time series regression for the evaluation of public health interventions: a tutorial. *Int J Epidemiol*. 2017 Feb 1;46(1):348–355. doi: 10.1093/ije/dyw098.
 51. Schober P, Vetter TR. Survival Analysis and Interpretation of Time-to-Event Data: The Tortoise and the Hare. *Anesth Analg*. 2018 Sep;127(3):792–8. doi:10.1213/ANE.0000000000003653.
 52. Alligon M, Mahlaoui N, Bouaziz O. Pitfalls in time-to-event analysis of registry data: a tutorial based on simulated and real cases. *Front Epidemiol*. 2024 Jul 11;4:1386922. doi:10.3389/fepid.2024.1386922.
 53. Bland M. *Biostatistics in research practice: time to event data*. York: University of York; [cited 2026 Aug 10]. 18 p. <https://www-users.york.ac.uk/~mb55/msc/resprac/survplus_text.pdf>
 54. Cox DR. Regression models and life-tables. *J R Stat Soc Series B Stat Methodol*. 1972;34(2):187–220. doi:10.1111/j.2517-6161.1972.tb00899.x.
 55. Andrade C. Survival analysis, Kaplan-Meier curves, and cox regression: basic concepts. *Indian J Psychol Med*. 2023 Jul;45(4):434–5. doi:10.1177/02537176231176986.
 56. Sweeting MJ, Rutherford MJ, Jackson D, Lee S, Latimer NR, Hettle R, et al. Survival Extrapolation incorporating general population mortality using excess hazard and cure models: a tutorial. *Med Decis Making*. 2023 Aug;43(6):737–48. doi:10.1177/0272989X231184247.
 57. Austin PC, Lee DS, Fine JP. Introduction to the analysis of survival data in the presence of competing risks. *Circulation*. 2016 Feb 9;133(6):601–9. doi:10.1161/CIRCULATIONAHA.115.017719.
 58. Fine JP, Gray RJ. A proportional hazards model for the subdistribution of a competing risk. *J Am Stat Assoc*. 1999;94(446):496–509. doi:10.1080/01621459.1999.10474144.
 59. Gray RJ. A class of K-sample tests for comparing the cumulative incidence of a competing risk. *The Annals of Statistics*. 1988;16(3):1141–54. doi:10.1214/aos/1176350951.
 60. Andersen PK, Gill RD. Cox's regression model for counting processes: a large sample study. *The Annals of Statistics*. 1982;10(4):1100–20. doi:10.1214/aos/1176345976.
 61. Prentice RL, Williams BJ, Peterson AV. On the regression analysis of multivariate failure time data. *Biometrika*. 1981;68(2):373–9. doi:10.1093/biomet/68.2.373.
 62. Amorim LD, Cai J. Modelling recurrent events: a tutorial for analysis in epidemiology. *Int J*

- Epidemiol. 2015 Feb;44(1):324–33. doi:10.1093/ije/dyu222.
63. Le-Rademacher JG, Therneau TM, Ou FS. The utility of multistate models: a flexible framework for time-to-event data. *Curr Epidemiol Rep*. 2022;9(3):183–9. doi:10.1007/s40471-022-00291-y.
 64. Ropovik I. A cautionary note on testing latent variable models. *Front Psychol*. 2015 Nov 6;6:1715. doi:10.3389/fpsyg.2015.01715.
 65. Watkins MW. Exploratory factor analysis: a guide to best practice. *The Journal of Black Psychology*. 2018;44(3):219–46. doi:10.1177/0095798418771807.
 66. Goretzko D, Pham TTH, Buhner M. Exploratory factor analysis: current use, methodological developments and recommendations for good practice. *Current Psychology*. 2019;40(7):3510–21. doi:10.1007/s12144-019-00300-2.
 67. Goretzko D, Siemund K, Sterner P. Evaluating model fit of measurement models in confirmatory factor analysis. *Educ Psychol Meas*. 2024 Feb;84(1):123–44. doi:10.1177/00131644231163813.
 68. Harris JE, Gleason PM. Application of path analysis and structural equation modeling in nutrition and dietetics. *J Acad Nutr Diet*. 2022 Nov;122(11):2023–35. doi:10.1016/j.jand.2022.07.007.
 69. Jenatabadi HS. A tutorial for analyzing structural equation modelling. *arXiv [Preprint]*. 2015 [cited 2026 Aug 10]. <<https://arxiv.org/pdf/1504.03430>>
 70. Jenatabadi HS. An overview of path analysis: Mediation analysis concept in structural equation modeling. *arXiv [Preprint]*. 2015 [cited 2026 Aug 10]. <<https://arxiv.org/abs/1504.03441>>
 71. Schober P, Vetter TR. Logistic regression in medical research. *Anesth Analg*. 2021;132(2):365–6. doi:10.1213/ANE.0000000000005247.
 72. Shariatnia S, Ziaratban M, Rajabi A, Salehi A, Abdi Zarrini K, Vakili M. Modeling the diagnosis of coronary artery disease by discriminant analysis and logistic regression: a cross-sectional study. *BMC Med Inform Decis Mak*. 2022 Mar 29;22(1):85. doi:10.1186/s12911-022-01823-8.
 73. Dhamnetiya D, Goel MK, Jha RP, Shalini S, Bhattacharyya K. How to perform discriminant analysis in medical research? explained with illustrations. *J Lab Physicians*. 2022 Jun 1;14(4):511–20. doi:10.1055/s-0042-1747675.
 74. Song YY, Lu Y. Decision tree methods: applications for classification and prediction. *Shanghai Arch Psychiatry*. 2015 Apr 25;27(2):130–5. doi:10.11919/j.issn.1002-0829.215044.
 75. Dou B, Vandenberg B, Moons P. Cluster analysis: an overview of traditional and novel approaches for health researchers. *Eur J Cardiovasc Nurs*. 2026 May 26:zvag100. doi:10.1093/eurjcn/zvag100.
 76. Sorgente A, Caliciuri R, Robba M, Lanz M, Zumbo BD. A systematic review of latent class analysis in psychology: Examining the gap between guidelines and research practice. *Behav Res Methods*. 2025 Oct 3;57(11):301. doi:10.3758/s13428-025-02812-1.
 77. Mathew A, Doorenbos AZ. Latent profile analysis – An emerging advanced statistical clustering approach. *Indian Journal of Continuing Nursing Education*. 2022;23(2):127–33. doi:10.4103/ijcn.ijcn_24_22.
 78. Bailly S, Dey T. Latent Class Analysis to Identify Subgroups in Clinical Research: A Data-Driven Approach. *Chest*. 2026 Jan;169(1):37–40. doi:10.1016/j.chest.2025.07.4085.
 79. International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use. ICH E6(R3) guideline for good clinical practice. Geneva: International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use; 2025 Jan 6 [cited 2026 Aug 10]. 79 p. <[https://database.ich.org/sites/default/files/ICH_E6\(R3\)_Step4_FinalGuideline_2025_0106.pdf](https://database.ich.org/sites/default/files/ICH_E6(R3)_Step4_FinalGuideline_2025_0106.pdf)>
 80. Heale R, Twycross A. Validity and reliability in quantitative studies. *Evid Based Nurs*. 2015 Jul;18(3):66–7. doi:10.1136/eb-2015-102129.
 81. Kimberlin CL, Winterstein AG. Validity and reliability of measurement instruments used in research. *Am J Health Syst Pharm*. 2008 Dec 1;65(23):2276–84. doi:10.2146/ajhp070364.
 82. Madadzadeh F, Bahariniya S. Tutorial on internal consistency assessment by Cronbach's alpha and McDonald's omega. *Perioperative Care and Operating Room Management*. 2025; 41:100568. doi:10.1016/j.pcorm.2025.100568.

83. Tavakol M, Dennick R. Making sense of Cronbach's alpha. *Int J Med Educ*. 2011 Jun 27;2:53–55. doi:10.5116/ijme.4dfb.8dfd.
84. Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *J Chiropr Med*. 2016 Jun;15(2):155–63. doi:10.1016/j.jcm.2016.02.012.
85. University of York Department of Health Science. Measuring health and disease: assessing agreement between methods of clinical measurement [Internet]. York: University of York Department of Health Science; [cited 2026 Aug 10]. 10 p. <<https://www-users.york.ac.uk/~mb55/msc/clinimet/week3/limagree.pdf>>
86. McHugh ML. Interrater reliability: The kappa statistic. *Biochem Med (Zagreb)*. 2012;22(3): 276–82. doi:10.11613/BM.2012.031.
87. Yusoff MSB. ABC of content validation and content validity index calculation. *Education in Medicine Journal*. 2019;11(2):49–54. doi:10.21315/eimj2019.11.2.6
88. Reitsma JB, Rutjes AW, Khan KS, Coomarasamy A, Bossuyt PM. A review of solutions for diagnostic accuracy studies with an imperfect or missing reference standard. *J Clin Epidemiol*. 2009 Aug;62(8):797–806. doi:10.1016/j.jclinepi.2009.02.005.
89. Vrbic CM. Recommendations for reporting measures of diagnostic accuracy. *Cytopathology*. 2023 May;34(3):185–90. doi:10.1111/cyt.13208.
90. Trevethan R. Sensitivity, specificity, and predictive values: foundations, plausibilities, and pitfalls in research and practice. *Front Public Health*. 2017 Nov 20;5:307. doi:10.3389/fpubh.2017.00307.
91. Hajian-Tilaki K. Receiver operating characteristic (ROC) curve analysis for medical diagnostic test evaluation. *Caspian J Intern Med*. 2013 [cited 2026 Aug 10];4(2):627–35. <<https://pmc.ncbi.nlm.nih.gov/articles/PMC3755824/>>
92. OpenAI. ChatGPT (GPT-5.6 Thinking) [Internet]. San Francisco: OpenAI; 2026 [cited 2026 Aug 10]. <<https://chatgpt.com/>>
93. Anthropic. Claude (Sonnet 5) [Internet]. San Francisco: Anthropic; 2026 [cited 2026 Aug 10]. <<https://claude.ai>>
94. Google. Gemini (Aug 2 version) [Internet]. Mountain View: Google; 2026 [2026 Aug 10]. <<https://gemini.google.com/>>