

Glossary of Study Design and Statistical Terms

Absolute risk reduction: The absolute difference in rates of an outcome between treatment and control groups.¹ Example: A hypothetical clinical trial compares the effect of a new statin and placebo on the incidence of stroke. Over the course of the study, the incidence of stroke is 4% with the statin and 6% with placebo. The absolute risk reduction with the statin is $6\% - 4\% = 2\%$.

Bias: Flaws in the design or operation of a study that lead to results that are not truth.² Bias can occur at any point of a study. Some types include:⁵

- *Information bias:* systematic differences in collection and handling of data.
- *Confounding bias:* additional factors distort the interaction between the treatment and the outcome.
- *Selection bias:* Differences between treatment and control groups that result from the way patients were selected. Randomization and blinding should help prevent selection bias.

Blinding or Masking: May help reduce bias in a study. Blinding can be double-blind (where neither the investigator nor the patient knows who is assigned to which treatment group), single-blinded (one party is aware of treatment groups), or open-label. In an open-label study, patients and investigators know the treatment group assignments.^{3,4}

Case-control study: Compares patients who have the outcome of interest (cases; often rare conditions) and patients without that outcome (controls) to look for causes or characteristics that are linked to the outcome. Case-control studies are retrospective and observational.⁶

Cohort study: Compares two groups (cohorts) of patients, one with exposure or risk factor and one which does not (control group). Cohort studies are observational studies.⁶

Composite endpoint: A combination of endpoints that each have a low incidence but together provide a single measure of effect. Individual outcomes within a composite endpoint should have similar value (e.g., cardiovascular death, myocardial infarction, and stroke.) Each component should be analyzed individually. Improvement in a single component can be responsible for statistical significance of a composite endpoint. A negative outcome for one component can negate or dilute the positive outcome for another component.⁷

Confidence Interval (CI): An estimate of the range within which the true treatment effect lies. The 95% CI is the range of values within which we are 95% certain that the true value lies. If the confidence interval for the difference in efficacy (a difference in means or proportions) between two treatments includes zero, then you cannot exclude the possibility that there is no difference in efficacy between treatments. The width of the CI is determined by the number of patients studied, the variability of the data, and the pre-set confidence level. The confidence level is usually set at 95%, but can range from 90% to 99%.²

Confounding factor: An additional factor (other than the treatment/intervention) in a study that affects the statistical outcome of a treatment/intervention. A confounding factor can make it appear that there is a direct relationship between two factors when, in reality, the confounder is responsible for the relationship.²

Crossover study: Two groups receive both interventions/treatments. Each group serves as their own control. Reduced variability means a smaller sample size is needed than for a parallel-group trial. The two treatments/interventions are usually separated by a washout period.^{3,6}

Cross-sectional study: This type of study looks at a defined population at a single point in time; it is a snapshot of what is happening at that moment in time.⁶

Effectiveness: How well a drug or intervention works in every-day real-world use.⁸

Efficacy: How well a drug or intervention works under ideal circumstances, such as in a randomized controlled trial.⁸

Endpoint: Pre-determined outcome used to measure an outcome of benefit or safety. The primary endpoint addresses the most important question the study attempts to answer. Secondary endpoints address supportive questions.⁸

Heterogeneity: Measure of variability among studies in a meta-analysis or systematic review. Heterogeneity occurs when there is more variation between the study results than would be expected to occur by chance alone. Testing for heterogeneity helps determine if it's appropriate to combine studies.²

Incidence: The rate of newly diagnosed cases of a disease occurring in the population at risk during a specified period of time.⁸

Intention-to-treat analysis: A statistical analysis for randomized trials that includes all of the patients who were randomized to a treatment arm regardless of whether or not they finished the study. An intention-to-treat analysis is considered to mimic clinical practice more closely than an analysis that includes just the patients who completed the study.²

Meta-analysis: An analysis of pooled data from several studies, which all address the same clinical question. Criteria for study inclusion in the analysis are established beforehand. Meta-analysis can be used to increase sample size, statistical power, and/or allow for increased subgroup analysis. Analysis is considered reliable with rigorous methods and appropriate study inclusion.⁸

Non-inferiority study: Looks to see if a new treatment works not worse than an existing treatment based on prespecified outcomes. Often used when it would be considered unethical to randomize patients take a placebo. This type of study requires fewer patients to show a significant difference compared to a superiority study.⁸

Null hypothesis: Hypothesis that there is no statistical difference between treatment groups in a study.⁹

Number needed to harm (NNH): The number of patients treated with a specified therapy/intervention in order for one of them to have a bad outcome (i.e., harm), over a specified time period. Round calculation down to a whole number.⁸

Number needed to treat (NNT): The number of patients that need to be treated with a specified therapy/intervention in order for one patient to benefit from treatment over a specified time.⁸ The NNT is the inverse of the absolute risk reduction (1 divided by absolute risk reduction expressed as a decimal or 100 divided by the absolute risk reduction expressed as a percentage). Must be calculated with statistically significant results. Takes into account the relative risks as well as the absolute risk of no treatment. $NNT = 100/(\% \text{ in control group} - \% \text{ in intervention group})$. Round the NNT up to a whole number.¹⁰

Observational study: Patients are not randomized to intervention or treatment groups. The investigators observe patients with a disease or outcome to assess outcomes. Examples are case-control studies, cross-sectional studies, and longitudinal studies (cohort study, panel study).⁵

Odds ratio (OR): The odds of an event occurring with or without a treatment/intervention. Odds ratios and relative risk are comparable when the outcome is rare. But the odds ratio can make risk appear greater when the disease or outcome is more common. In case-control studies evaluating the risk of an adverse effect, an odds ratio of 1 indicates that exposure to the drug is equally likely in cases and controls. If the odds ratio is greater than 1, the risk of exposure is greater in cases than controls. If the odds ratio is less than 1, the risk of exposure is smaller in cases than controls.^{1,2}

p-value (probability-value): The level of statistical significance (i.e., the alpha). A value of $p < 0.05$ means that the probability that the result is due to chance is less than 1 in 20. The smaller the p-value, the greater the statistical significance. The p-value does not provide any information about the size of an effect. It only describes the strength of the result.¹¹

Point estimate: The result of a clinical trial or meta-analysis which is used as a best estimate of what the true value is in the population that the study sample came from.¹²

Positive predictive value: Proportion of people who actually have the disease when a diagnostic test is positive. Positive predictive value = $(100 \times \text{true positive}) / (\text{true positive} + \text{false positive})$.⁸

Power: The ability of a study to detect a significant difference between treatment groups (i.e., the probability that a study will have a statistically significant result [$p < 0.05$]). Accepted study power is usually set at 0.8 (80%). Power increases as sample size increases.²

Prevalence: The proportion of existing cases of a disease or condition in the population at risk at a given time. Prevalence = incidence/population or people at risk.¹³

Prospective study: Studies that begin in the present and will evaluate events as they occur in the future.⁸

Randomization: Provides each patient an equal chance of being assigned to any of the groups in a study, to avoid selection bias.²

Randomized controlled trial (RCT): A prospective study in which patients are randomized into treatment or control groups. These groups are followed up for the variables/outcomes of interest.²

Relative risk: The risk of an event in individuals in an exposure group compared with the risk of that event in a non-exposed group. In a clinical trial, this is the probability of an event in the treatment group divided by the probability of that event in the placebo group.⁸

Relative risk ratio: Ratio of relative risk rates with treatment vs. control group. A relative risk ratio of 1 indicates no association between treatment and outcome. A relative risk greater than 1 indicates a positive association between treatment and outcome. A relative risk less than 1 indicates a negative association between treatment and outcome.¹

Relative risk reduction: Relative risk subtracted from 1.¹

Retrospective study: Observational studies that look back in time to evaluate events that occurred in the past.⁸

Sample size: Calculated prior to initiation of a study based on the number of patients required for a study to have valid results. An increased sample size is required when differences between treatment groups are small, if the power of the study is set higher (e.g., 90% power instead of the standard 80%), as statistical significance increases (as in $p < 0.001$ instead of $p < 0.05$), and if there is more variability in the outcome being measured. The larger the sample size, the more narrow the confidence interval.²

Sensitivity: The true positive rate. Measure of how many people who test positive are actually positive. Sensitivity = $(100 \times \text{true positives}) / (\text{true positives} + \text{false negatives})$.⁸

Sensitivity analysis: A statistical method to determine how sensitive the results of a study or systematic review are to changes in the data or methodology. This is particularly important to perform in meta-analyses. Looks at the main outcome with alternative assumptions (e.g., may exclude weaker studies).²

Significance: Results in a study are statistically significant if the p-value is less than the predetermined value (often $p < 0.05$). Statistical significance is a measure of the probability that the observed results are an actual difference and are not due to chance. Study results are clinically significant if they are important enough to implement in clinical practice.⁹

Specificity: The ability of a diagnostic test to reliably rule out a disease. The proportion of patients without the target disease who have a negative test. Specificity = $(100 \times \text{true negatives}) / (\text{true negatives} + \text{false positives})$.⁸

Subgroup analysis: Looks at outcomes in specific groups of patients within a study to determine if the observed outcome is consistent across groups. Subgroups can be by age, sex, concomitant medical conditions, or other characteristics.²

Superiority study: This type of study tests to see if a new treatment/intervention is better than an active treatment (e.g., standard of care) or control (e.g., placebo).⁸

Surrogate Endpoint: A surrogate endpoint is an endpoint that stands in for another endpoint. Examples include measurement of blood pressure as a surrogate for reducing cardiovascular events in patients with hypertension, or measurement of CD4 cell counts as surrogate for reducing mortality with antiretroviral therapy. Surrogate endpoints should measure an effect more quickly or easily and should be highly correlated to the clinical outcome it stands in for.⁸

Systematic review: Collection and review of all available studies addressing a particular clinical question. Pre-determined specific criteria and methods are used. A systematic review may include meta-analysis as a method of analyzing and quantifying the results.²

Type I error: To conclude there is a difference between treatments when there is really no difference between them; rejection of the null hypothesis when it is actually true.

Type II error: To conclude there is no difference between treatments when there really is a difference between them; accepting the null hypothesis when it is actually false. This type of error is common in clinical trials, often because they don't enroll enough patients.²

Users of this resource are cautioned to use their own professional judgment and consult any other necessary or appropriate sources prior to making clinical judgments based on the content of this document. Our editors have researched the information with input from experts, government agencies, and national organizations. Information and internet links in this article were current as of the date of publication.

Cite this document as follows: Clinical Resource, Glossary of Study Design and Statistical Terms. Pharmacist's Letter/Pharmacy Technician's Letter/Prescriber Insights. April 2025. [410466]

References

1. Darzi AJ, Busse JW, Phillips M, et al. Interpreting results from randomized controlled trials: What measures to focus on in clinical practice. *Eye (Lond)*. 2023 Oct;37(15):3055-3058.
2. Nagendrababu V, Dilokthornsakul P, Jinatongthai P, et al. Glossary for systematic reviews and meta-analyses. *Int Endod J*. 2020 Feb;53(2):232-249.
3. National Library of Medicine. Clinicaltrials.gov glossary terms. December 9, 2024. <https://clinicaltrials.gov/study-basics/glossary>. (Accessed March 20, 2025).
4. Higgins KM, Levin G, FDA. Considerations for open-label clinical trials: design, conduct, and analysis. <https://www.fda.gov/media/168664/download>. (Accessed March 19, 2025).
5. University of Texas Libraries. Systematic reviews & evidence synthesis methods – glossary of terms. March 19, 2025. <https://guides.lib.utexas.edu/systematicreviews/glossary>. (Accessed March 20, 2025).
6. Chidambaram AG, Josephson M. Clinical research study designs: The essentials. *Pediatr Investig*. 2019 Dec 21;3(4):245-252.
7. High-powered database. Quick review of biostatistics. <https://highpoweredmedicine.com/biostatsPage.html>. (Accessed March 20, 2025).
8. Association of Health Care Journalists. Health journalism glossary. <https://healthjournalism.org/glossary/>. (Accessed March 20, 2025).
9. Shreffler J, Huecker MR. Hypothesis testing, p values, confidence intervals, and significance. March 13, 2023. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2025 Jan-.
10. Ranganathan P, Pramesh CS, Aggarwal R. Common pitfalls in statistical analysis: Absolute risk reduction, relative risk reduction, and number needed to treat. *Perspect Clin Res*. 2016 Jan-Mar;7(1):51-3.
11. Tenny S, Abdelgawad I. Statistical Significance. November 23, 2023. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2025 Jan-.
12. eCQI Resource Center. Point estimate. September 7, 2022. <https://ecqi.healthit.gov/glossary/point-estimate>. (Accessed March 20, 2025).
13. Tenny S, Hoffman MR. Prevalence. May 22, 2023. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2025 Jan-.

—To access hundreds more clinical resources like this one, visit trchealthcare.com to log in or subscribe—