Overview of Statistical Concepts for Interpreting Data and Clinical Trial Results

Amy Baum Jones, MPAS, PA-C

Keywords: Statistics, p-value, confidence interval, median, mean, relative ratio

Statistical concepts such as p-value, confidence interval, measures of central tendency, and relative ratio are commonly used by clinicians for data interpretation. This article will serve as a refresher for these concepts and give readers an idea of their practical application. Healthcare providers can use this knowledge to identify clinical relevance as well as any misinterpretation or misrepresentation of data.

CASE REPORT

In the ever-changing world of dermatology, treatments and clinical developments emerge daily. An evidence-based approach to these changes by statistical analysis generates clinically relevant information that healthcare providers can act upon. However, weeding through clinical trials and articles can prove challenging with the statistical jargon used to quantify results. Terms such as null/alternative hypothesis, confidence interval (CI), p-value, mean vs. median, and relative ratio can be overwhelming if out of practice or unfamiliar with these concepts. Familiarization and proficiency with these concepts can help clinicians identify misinterpretation and/or misrepresentation of data when interpreting the reported outcomes of clinical trials as well as determine results that are fair, accurate, and applicable to actual patient care. The goals of this article are to provide readers with a review of some of the elements of data interpretation and increase practitioners’ confidence in their ability to critically evaluate and apply findings from clinical research studies.

NULL HYPOTHESIS AND ALTERNATIVE HYPOTHESIS FORMATION

For starters, let’s look at hypothesis formation. When new therapies (B) want to assert superior efficacy or safety over the current standard of care (A), a null hypothesis (e.g., Hypothesizing that there is NO DIFFERENCE between the effect/safety of Drug A vs. Drug B) must be proven as false. Instead, an alternative hypothesis is proposed. Alternative hypotheses can be one-tailed or two-tailed in nature. For example, “Drug B is SUPERIOR to Drug A” is a one-tailed hypothesis. Stating that Drug B is 1) NOT THE SAME in efficacy as Drug A and 2) better OR worse is a two-tailed hypothesis. Understanding how hypotheses are formed is critical to preventing type I errors (i.e., false positive; effect detected when there actually is none), and type 2 errors (false negative; failure to detect an effect that truly exists).

CONFIDENCE INTERVALS VS. P-VALUES

The next set of concepts go hand in hand: confidence intervals and p-values. Before a study begins, an "Alpha level" is assigned to determine the significance of the trial’s findings. This Alpha level is the standard to which the p-value will be compared to determine statistical significance. The p-value helps determine the power of a particular study, or how likely a test is to detect an effect if one truly exists. P-values determine the strength of a probability on a scale of 0 to 1. General concepts for p-values include the following:

- A small value (<0.05 or 5%) indicates strong evidence against the null hypothesis (resulting in a rejection of the null hypothesis).
- A large value (>0.05) implies weak evidence, and the null hypothesis remains likely true.

Conversely, the CIs identify a range in which the parameter being investigated will likely fall. For example, let’s say fasting blood glucose is the parameter being investigated by our study. A 95% CI states that 95% times out of 100, we can expect a fasting blood glucose result that falls in the identified interval range of this particular study; however, 5% of the time (5/100) the result will fall out of the interval range due to chance alone.
KEY CONSIDERATIONS FOR CONFIDENCE INTERVALS INCLUDE:

- A larger CI (99%) vs smaller CI (95%) implies a wider interval range of potential numerical results to ensure a higher probability of including the true value of the parameter being evaluated.
- A larger CI indicates a high dispersion of parameter values; meaning less confidence that the study has determined the true value of the parameter being studied.
- Smaller CI occurs with larger sample sizes. This results in a narrower interval range and greater confidence that the study has determined the true value of the parameter being studied.
- When Drug A and Drug B have overlapping CIs, it suggests their results are comparable and not significantly different, thus claiming superiority/inferiority is not substantiated.
- When a CI contains the value "0," the findings of this parameter are not statistically significant or meaningful. Similarly, when a CI crosses the value "1," this implies there is no difference between the study arms being compared.

When comparing CIs and p-values, remember that CIs provide information about statistical significance and direction of the effect. CIs also provide a range of plausible values for the target effect, and a narrow CI results in higher strength in the predictability of the effect. On the contrary, p-values only allow for the binary interpretation of a result as statistically significant or not, which can be misleading (think statistical significance vs. clinical relevance). P-values require an attempt to draw conclusions for a target population based on a single value.

MEASURES OF CENTRAL TENDENCY

Another consideration for the interpretation of data involves reporting of mean vs. median. These terms are likely familiar to most as representations of the middle for a given data set. Reporting of both values together is most transparent, but not always incorporated in data displayed. Additionally, it may not always be appropriate to report a mean or median value based on the parameter being investigated; one may be a more correct evaluation of the data than the other. The mode is the third measure of central tendency. In reporting scientific data, the mode is rarely used, since it is more appropriately applied to nominal data sets such as yes or no, names, country of birth, etc. Beyond these considerations, the differences between these two terms are as follows:

- Mean is the average of all the values in a data set and is thought to be representative of a typical result. This works well when results follow a classic bell-shaped curve, evenly distributed about the center.
- Median is the middle number of a data set. This is more representative of a typical result when the overall data set is skewed in either direction (positively or negatively) and has been influenced by outliers (very high and/or very low scores).

RELATIVE RATIO

Finally, let’s delve into relative ratio (RR), sometimes seen as "relative risk" or "risk ratio." RR allows the evaluation of a particular outcome in the drug arm of a trial in comparison to the same outcome in the placebo arm. RR is portrayed in its relation to the number 1. For example, if we are evaluating efficacy of Drug A when compared to placebo, a value greater than 1 (e.g., 1.5) favors the drug for the parameter being discussed. If the RR for Drug A is 1, then we conclude it is equal to placebo for this parameter. For an RR number less than 1 (e.g., 0.95), we can conclude this in favor of placebo for the parameter being discussed. So, if adverse events (AEs) are the parameter, a value of 1.5 says the AE is more likely in Drug A; a value of 1 says the AE happens equally between placebo and Drug A (think 1:1 ratio); and a value of 0.95 says the AE is more likely in the placebo group than the drug arm. This system holds true for any parameter being assessed in a trial, whether in terms of safety or efficacy.

CONCLUSION

These are several of the concepts used in the presentation of data from clinical trials, articles, and meta-analyses. Reviewing these statistical concepts can help clinicians identify strengths and weaknesses of studies and integrate therapies with compelling data into their treatment repertoire. Familiarity with key statistical tools can help providers better discern clinical trial data and interpret results of research analyses.

RESOURCES

The author found the following resources helpful in the preparation of this article:


DISCLOSURES

Amy Baum Jones, MPAS, PA-C, is a Medical Science Liaison at Bristol Myers Squibb. She is also a former member of the Dermatology Physician Assistant Foundation (DPAF) Board of Trustees and serves on the Advisory Council for Foundation for Research and Education in Dermatology (FRED).