Statistics in Clinical Research – Part II

Chyon-Hwa Yeh
Statistician, Cook Research Incorporated, inc.

Continuing the discussion from Part I, the key roles of statistics for clinical research will be discussed here. As we mentioned in the last article, clinical research is a complex subject as it involves many different scientific disciplines. Statistics plays an important role in study designs, execution, clinical operation, data analyses, and interpretation of study results. Conducting data analyses is the most critical function that statistics discipline offers for a clinical research. It is important that correct and appropriate statistical methods be used for various data.

Subject Disposition

After a clinical study is completed, it is imperative to report subject accountability of each study subject: did the subject complete the study or exit study early; if not, why; if death occurs, what caused it; if a subject withdrew early or lost-to-follow-up, when was the last time the patient was contacted. The summary of subject disposition also serves as a tool to assess the soundness of study conduct. Also, the distribution of subject accountability between study groups (if applicable) should be assessed for comparability. If not, further investigation may be warranted.

Data Analyses – Analysis Datasets

It is common that there are missing data or incomplete data after the study is completed. Depending on the composition of the data, there are various types of analysis datasets that statistical analyses should be performed on to evaluate the robustness and validity of the results. Intent-to-treat (ITT), analyzable data, per-protocol, and safety analysis datasets consist of different patient populations; they may provide different interpretations of the results. Statistical analyses should be performed on different types of analysis datasets to ensure the results are consist regardless of the datasets used.

An intent-to-treat dataset is defined as all patients who are randomized and analyzed as such, regardless of any circumstances. However, occasionally, one can further define ITT dataset consists only patients who are randomized and receive the study treatment. Patients who do not receive treatment for various reasons (patients decide not to participate the study after being randomized, patients die prior to the 1st dose, etc.) cannot be evaluated for safety or efficacy since there is no treatment effect to be assessed.

An analyzable dataset refers to the dataset that composes of patients with study outcome. Patients whose study outcome is not available for evaluation or cannot be determined due to various reasons (lost to follow up, early withdraw, patients refuse scheduled clinic visits, etc.) are excluded in this dataset. If there are many patients with missing study outcome, imputation for missing data may be required for this dataset.

A per-protocol dataset includes only patients who have no protocol violations. This dataset is usually what a clinical study sets out to study. However, during the conduct of the study many protocol violations result in patients that are not appropriate for analyses and therefore they will be excluded from the analyses. If there are many patients that are excluded from this dataset, it indicates that the conduct of the study may be flawed. Therefore, the results from this analysis dataset are usually not being accepted.

A safety dataset includes patients who has taken at least 1 study dose. The purpose of this dataset is to evaluate the safety of the study treatments.

Efficacy Analyses

The purpose of efficacy analyses is to assess whether the study treatment of interests meets the proposed hypothesis of superiority, inferiority, or no difference. Depending on the hypothesis, an endpoint can be analyzed with several different analytical methods and each method will lead to different interpretations of the data. We will explore various analytical methods that are commonly used in categorical and continuous variables, as well as for time to event endpoints.

Efficacy Analyses - Categorical Variables

For categorical data, the endpoints can be a binary endpoint (yes/no, improved/not improved), or endpoints with ordinal scales (mild, moderate, severe). The Fisher's chisq test is often used to assess whether there's association between the row and column variables (treatment and outcome). The Cochran–Mantel–Haenszel test is used to assess the associations between the row and column variables among various strata (for example, diabetes with prior MI history, diabetes without prior MI, without diabetes with prior MI, without diabetes without prior MI). Additionally, the difference in percentage between 2 treatment groups and the 95% confidence intervals are usually provided. Also, the efficacy of a study treatment can also be assessed by odds ratios using the logistics regression method. The magnitude of odd ratios is used to evaluate the likelihood that the outcome of interest will/will not occur.

For endpoints with ordinal scales, one should analyze the data as ordinal variables, instead of treating them as continuous variables. The reasons are that (1) the scales of ordinal variables are not scaled to represent what it means to measure. For example, when a patient improved from ‘2’ (moderate) to ‘1’ (mild), it may not represent the same magnitude for a patient that improved from ‘3’ (severe) to ‘2’ (moderate). (2) the non-integer changes in scale does not provide appropriate interpretations. For example, for an illness measured by ordinal scales of where 0 = Asymptomatic, 1 =Mild claudication, 2 = Moderate claudication, 3 = Severe claudication, 4 = Rest pain, 5 = Ischemic ulceration not exceeding ulcer of the digits of the foot, and 6 = Severe ischemic ulcers or frank gangrene. An improvement of non-integer (for example, 2.32, 1.08, etc.) provides no clinical meaning and interpretation of the data.

For ordinal categorical data, one can use cumulative logit method for multinomial variables, depending on the objectives of the study. The logit of ordinal categorical data can be expressed as

$$logit\left(\frac{category\ k}{category\ j}\right) = \alpha_k + \beta_k,$$

or the cumulative logit can be expressed as

$$\frac{\frac{p(response \le j)}{1-(response \le j)}}{\frac{p(response \le k)}{1-(response \le k)}} = \exp(x_j \ - \ x_k\ )\,\beta.$$

It provides results in terms of odds ratio where the researchers can assess the efficacy of study treatments by the magnitude of likelihood

For example, a study evaluates the clinical benefit of 2 new treatments against placebo for relieving the symptoms on lower legs due to the occlusion of the peripheral vessels. The Rutherford scoring system is used to assess the severity. The Rutherford scores are Stage 0 – Asymptomatic, Stage 1 – Mild claudication, Stage 2 – Moderate claudication, Stage 3 – Severe claudication, Stage 4 – Rest pain, Stage 5 – Ischemic ulceration not exceeding ulcer of the digits of the foot, Stage 6 – Severe ischemic ulcers or frank gangrene.

The clinical assessment after the surgery is tabulated in the following table.

Table of treatment by outcome							
Treatment	outcome						
	Asymptomatic (0)	Mild claudication (1)	Moderate claudication (2)	Severe claudication (3)	Rest pain (4)	Severe ischemic ulcers or frank gangrene (5 /6)	Total
New Drug A	33	12	10	5	3	3	66
New Drug B	10	16	22	8	10	2	68
Placebo	30	16	12	5	4	2	69
Total	73	44	44	18	17	7	203

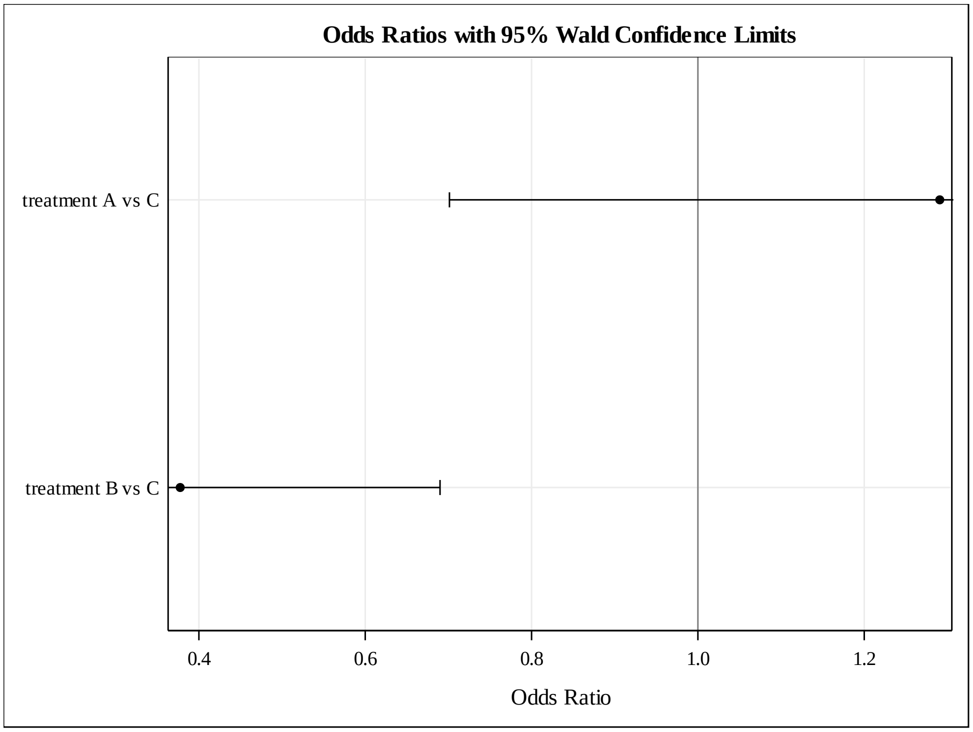
If one uses general liner model to analyze the results, the average least square

means for each treatment is 2.12, 2.97, and 2.17. Since the Rutherford score is an ordinal scoring system, it is not clinically meaningful to quantify the treatment effect as a continuous endpoint. For example, a score of 2.12 is in between moderate and severe claudication; it is difficult to interpret and conclude the treatment effect.

If one analyzes the results using logistic regression method for multinomial outcome, the clinical benefit can be assessed by the magnitude of odds ratio. The outcome using logistic regression method is summarized in the following table.

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
treatment A vs placebo	1.291	0.701	2.376
treatment B vs placebo	0.378	0.207	0.690

The following figure displays the relative benefit of Treatment A vs. Placebo and Treatment B vs. Placebo.



Efficacy Analyses: Continuous Variables

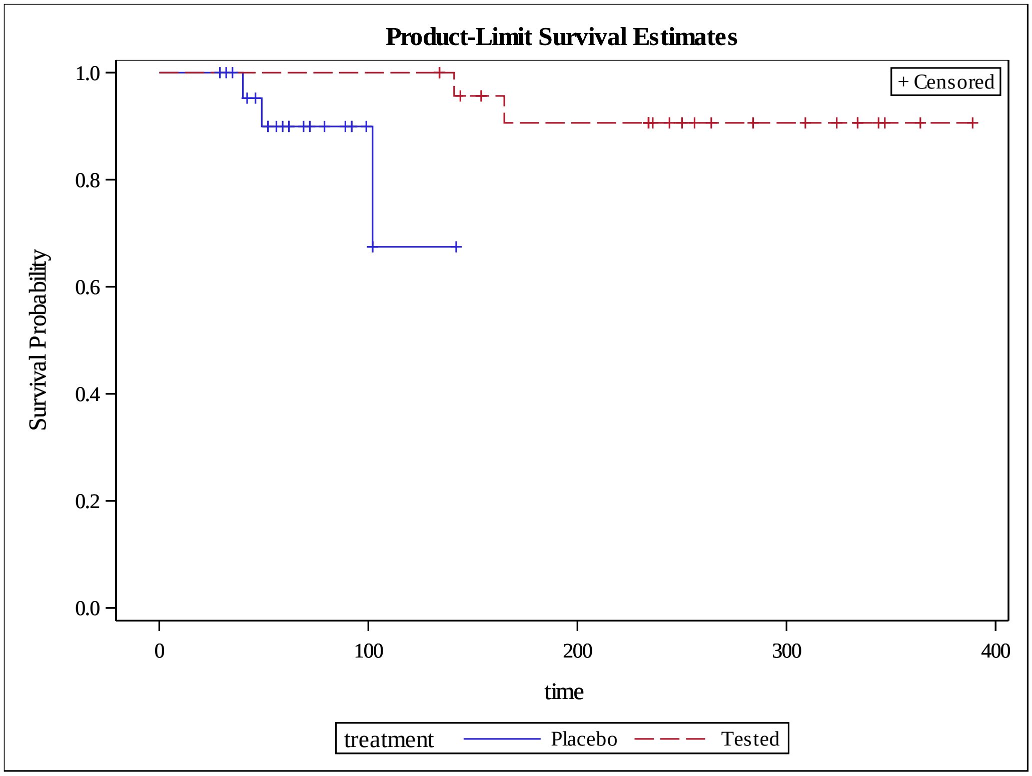
For continuous endpoints, the purposes of research often are to evaluate the improvement (change from baseline), the outcome at each planned time points (for example, blood pressure at 12 Months, tumor size at 6 months), or clinical benefit over a period (over 24 months). Statistical methods such as analysis of variance (ANOVA), GLM, non-parametric Wilcoxon method, or mixed model for longitudinal data are appropriate methods for analyses.

Efficacy Analyses: Time-to-Event Endpoints

In research areas where time to the event of interest, such as death, MI, tumor size reaching a pre-specified diameter, blood vessel occlusion, survival analyses provide as appropriate statistical methods for the data. Time-to-event endpoints relay on accurate data collection on timing of the event. Therefore, the data collection for survival endpoints should be carefully planned and executed to ensure the correct data are being collected.

For one-time-point event (for example, time to death, time to first MI, time to tumor growth >=50%), Kaplan-Meier product limited estimate or Cox-proportional hazard model is used to assess clinical outcome. For recurrent/multiple events (for example, time to 1st, 2nd, … atrial fibrillations within 12 Months), the Anderson Gill model is used to evaluate the treatment benefit.

For example, a new treatment was tested vs. placebo to determine if the new treatment can prolong or prevent atrial fibrillations (AF). The time to 1st occurrence of AF over 12months was the endpoint of interest. Patients who had no AF occurrence during the 12 months were censored at 12 months. The KM product-limit survival estimate (median time to event) was calculated and used to compare the treatment effect between the tested treatment and placebo. The figure below displays the survival probability over 12 months between 2 treatments. The tested treatment demonstrates that it has a higher survival probability than the placebo at any given time during the 12 months, particularly as time progresses, the tested treatment had a higher probability of free from not having AF.



Data Analyses – Safety Analyses

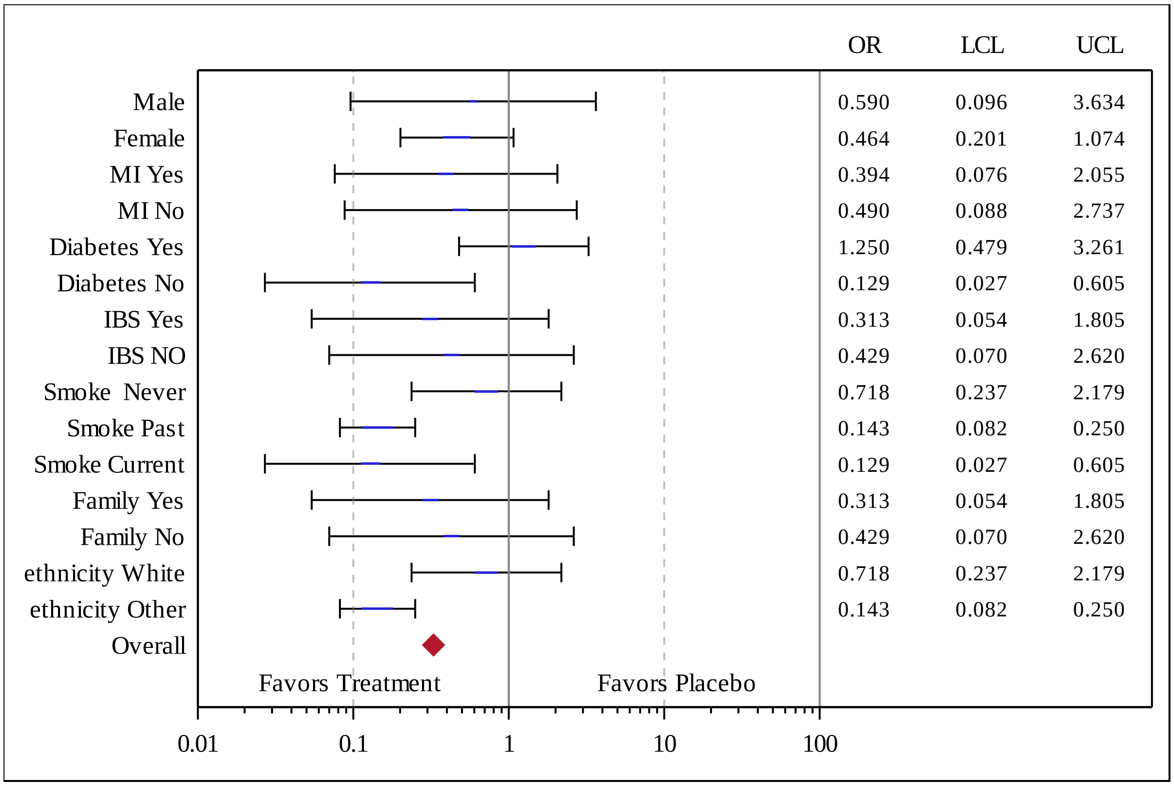
Self-reported and physician-reported adverse events are used to assess the safety profile for clinical studies. For pharmaceuticals, adverse events can be summarized by dose and exposure level; for medical devices, adverse events can be summarized by procedure or device related adverse event. Commonly, the types of AEs, related-ness of the AEs to the tested products, the severity of the AEs, the types of serious AEs, and the treatments for the AEs are summarized to provide the safety profile of the tested products.

Statistical analyses (hypothesis testing, p-value, confidence intervals) should not be the focus of the safety endpoints, unless the study is designed to investigate the safety of the tested treatment. Multiple comparisons, low/rare adverse events, particularly “no event”, present different statistical challenges to the analyses and may result in misleading conclusions. Therefore, unless a study is specially designed to assess the safety endpoint of interest, statistical analyses for safety endpoints should be avoided.

Subgroup Analyses

The purposes of subgroup analyses are to assess the consistent therapeutic effect across relevant clinical parameters, and to explore whether there are differences in therapeutic effects. Clinically relevant factors such as demographics, medical history, or baseline disease severity are commonly used to evaluate the consistency of the therapeutic effect of test treatments. Forest plots are often used to present the treatment effect among the subgroup factors of interests. Statistical methods such as Breslow-Day test for categorical data or interaction terms for continuous variables can be used to determine the significance of subgroup factors.

For example, the following forest plot shows the odds ratio of a tested treatment vs. placebo on mortality. Researchers can assess the treatment effect visually by examining the odds ratio of subgroup factors where it is greater or less than 1 (where treatment effect between a tested treatment and placebo is not different).



	OR	LCL	UCL
Male	0.590	0.096	3.634
Female	0.464	0.201	1.074
MI Yes	0.394	0.076	2.055
MI No	0.490	0.088	2.737
Diabetes Yes	1.250	0.479	3.261
Diabetes No	0.129	0.027	0.605
IBS Yes	0.313	0.054	1.805
IBS NO	0.429	0.070	2.620
Smoke Never	0.718	0.237	2.179
Smoke Past	0.143	0.082	0.250
Smoke Current	0.129	0.027	0.605
Family Yes	0.313	0.054	1.805
Family No	0.429	0.070	2.620
ethnicity White	0.718	0.237	2.179
ethnicity Other	0.143	0.082	0.250

Missing data

Missing data occurs commonly in clinical research. When the endpoints of interest are missing, imputation is performed to estimate the “maybe” value for missing data. There are several types of missing data: missing completely at random, missing at random, and missing not at random. Techniques that are often used to handle missing data include single imputation method, model-base method, or multiple imputation method.

Final Notes

Various statistical analyses can be applied to a set of clinical data and obtain various interpretations. It is important that the clinical scientists and statisticians work closely on the data analyses such that the statisticians provide appropriate analytical methods to the data based on the clinical input and the clinical scientists provide clinical insight to the results based on the statistical results.