Statistics in Clinical Research

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

Clinical research is a complex subject as it involves many different scientific disciplines. The success of clinical research requires sound study designs, well-executed study operation, appropriate data analyses, and scientific-based conclusions. Statistics plays an important role in study designs, execution, clinical operation and interpretation of study results. This article discusses the roles that statistics contributes in the study design and clinical operation of clinical research.

Statistics and Study Design in Clinical Research

Clinical research is motivated by scientific matters with many unknowns that are of interest to scientists. Once the research theme has been selected, the scientists carry out the research from designing the experiment, executing the study, then analyzing the data collected from the study to understand the outcome and ultimately answering the questions that initiate the research. There are multiple elements in the study design of clinical research that are related to statistics, we will focus in the following subjects: study objectives, statistical hypothesis, study population, selection of study groups, efficacy and safety endpoints, study procedures and schedules, sample size determination, and applications of meta-analyses in clinical studies.

Study Objectives

A newly proposed clinical research has many unknowns. The study objectives should be clearly defined as the objectives will determine the rest of the study elements, such as the study design, study endpoint, study size, or data collections, etc. For example, if the purpose of a clinical research is to assess whether a new pharmaceutical provides better efficacy in treating an illness than a current treatment, a superiority study will provide the answer to the unknown. If the study objective is to determine whether Drug A is bioequivalent to Drug B then a cross-over study with an equivalent hypothesis testing will provide the insight to the question.

Statistical Hypothesis

Once the study objective is established, the statistical hypothesis can be formulated based on the objective. Depending on the study objective the statistical hypothesis should be set up accordingly as a 2-sided (no different), 1-sided (better, worse), equivalent (no more than xx units and no less than xx units), non-inferior (no worse by xx units). The type of hypothesis will

then determine the magnitude of type I error. A study can have multiple hypotheses; however, it should be kept in mind that the type I error will need to be adjusted for multiple testing in order to claim statistical significance.

Study Population

Study population is the patients or subjects that the study intends to understand the underlying research of interest. The patient characteristics, such as the demographics, medical history and illness severity at baseline should be clearly defined. It is critical that the study enrolls correct patients to meet the research objectives. If the study enrolls inappropriate patients, which is defined as protocol violations, it may result in inconclusive outcome for the study conclusions. Also, some of the patient characteristics may have confounding effects on the study outcome, it should be called out in the study design phase so that the confounding effects can be mitigated by stratified randomization or by various statistical analyses to detangle treatment effect and the confounding factors.

Selection of Study Groups

There are several choices when it comes to the selection of study groups. For most of clinical research, a double-blinded, placebo controlled study is a gold standard. However, depending on the research contents and objectives, the scientists can choose different study groups, such as active-controlled or external-controlled (historical data). For a placebo controlled study design, it minimizes the study biases and establishes safety profiles. However, in many illness areas, it is not ethical to conduct a placebo controlled study, such as cancer, diabetes, or diarrhea. For an active controlled study, it compares the current treatment with the new treatment and has less ethical concerns, but requires establishing the non-inferiority margin. The researchers should weight each study design's pros and cons to determine the best study groups to meet the study objectives.

Study Endpoint

Study endpoint is one of the most important elements in clinical research. The researchers should carefully select an endpoint that will most represent and address the study objectives. The definition of an endpoint, such as the data collection method, calculation of the endpoint, or the time frame that the endpoint will contain should be clearly specified. A study can have multiple endpoints, such as primary endpoints, secondary endpoints, etc. Also, a study may define the endpoints using single measurement or composite measurements (death + MI+ stroke).

Study Procedure and Schedule

The study intervention should be clearly laid out, such as what study products will be utilized, what study procedures will be performed at what time points and by whom, how and where and what data will be collected, and when the study starts and when it ends. It will facilitate the conduct of the research and avoid confusion and misconduct of the research.

Sample Size Determination

Sample size for clinical research depends on several parameters: type I error, study power, treatment benefit, and variation of study. Conventionally, type I error is set at 5% for a 2-sided hypothesis testing or at 2.5% for a 1-sided testing; the study power is set at either 80% or 90%. The most difficult task to determine the sample size is the estimate of the treatment effect and the study variation since they are unknown. The common practice is to estimate them from the historically relevant data or published literature. Researchers would search thru the historical data from the experiments or published literature that were conducted with similar study characteristics. The magnitude of the treatment benefit should be clinically relevant for the management of patient illness. The variance of the study parameters (endpoints) can be estimated with similar matter with the assumptions that the new research would have similar variation as the historical data. Sometimes the variance will be inflated to prepare for unforeseen factors that might increase the variation of the new study.

Meta-Analyses

In most of the new clinical research, the information regarding the treatment benefit or study variation is often unknown. In order to establish close-to-accurate estimate for treatment benefit or parameter variation for the purpose of determining study size and optimizing study design, most researchers would seek information from historical data. One of the approaches is to aggregate the relevant historical data to obtain the estimate given a set of assumptions via meta-analysis. A meta-analysis is a statistical method that systematically combines pertinent qualitative and quantitative study data from several selected studies to assess and develop a single conclusion that has greater statistical power. This conclusion is statistically stronger than the analysis of a single study, due to increased numbers of subjects, greater diversity among subjects, or accumulated effects and results. A Meta-analysis is useful in developing and searching for the optimal estimates. However, it is technically challenging and time consuming. The inherited heterogeneity of studies included for meta analyses needs to be carefully reviewed and managed.

Prior to conducting a meta-analysis, the reviewer should first collect empirical evidence that meets the prespecified eligibility criteria that will answer the specific research questions. Once the relevant studies or data have been determined, the reviewers evaluate the similarities and dis-similarities between the new study and the historical studies. The study elements that should be carefully reviewed include study design, such as the type of study (cross-over, parallel, active or placebo controlled), length of study period, treatment groups, study endpoints, study population, study procedures or schedules, and other elements that would have potential impact on the study outcome. After the statistician analyzes the data from the historical data, he/she should discuss the results with the clinical scientist to adjust the estimates accordingly based on the differences between the historical data and new study and to evaluate the robustness of the estimates.

Meta-analysis can be very helpful if it is carried out appropriately. It requires the collaboration between clinical scientists and statisticians to reach an estimate of interest. For further information, please refer to Meta-Analysis Of Controlled Clinical Trials by Anne Whitehead (2003, John Wiley & Sons).