John Del Valle, Section Editor

Analyzing and Interpreting Clinical Trials

Cathie Spino

Department of Biostatistics, University of Michigan, Ann Arbor, Michigan

The set of skills necessary for clinicians and researchers to analyze and interpret clinical trials has necessarily grown over the last 25 years. The sophistication of integrated biomedical platforms that inform our understanding of disease and therapeutics, the complexity of study designs and statistical methods, and the possibility for greater bioethical dilemmas are the driving force behind this need. Thus, it is nearly impossible to be expert or even conversant in all aspects of analysis and interpretation of clinical trials. Lest you lose heart, oh earnest reader, reliance on some fundamental principles will serve you well. In this article, I discuss some of the key constructs required for appropriate analysis and interpretation of any clinical trial—whether a first-in-human phase I study of a new therapeutic agent for hepatocellular carcinoma or a cluster randomized trial of an innovative cognitive behavioral method in inflammatory bowel disease. In addition, a few specific topics that I find trip up new clinical researchers are also reviewed.

Clinical Trials

The National Institutes of Health defines a clinical trial as

> *A research study in which one or more human subjects are prospectively assigned to one or more interventions (which may include placebo or other control) to evaluate the effects of those interventions on health-related biomedical or behavioral outcomes.*[1]

Randomization is the process that prospectively assigns an individual or a group of individuals to $\geq$1 interventions. The interventions can be within a single arm (eg, dose-ranging study of a novel drug) or multiple arms that include placebo or other control groups (eg, standard of care). Interventions can also include biologics (eg, erlotinib), devices (eg, esophageal stents), procedures (eg, endoscopic retrograde cholangiopancreatography), delivery systems (eg, telemedicine), behavioral interventions (eg, cognitive behavior therapy or diet), and other health-related strategies.[1]

Design

It is impossible to interpret and analyze a study that is not designed properly. Thus, any discourse on analysis and interpretation of clinical trials must emphasize adequate design principles, such as randomization, control, and blinding. If you select the wrong analytic technique or if the assumptions underlying the chosen statistical analytic approach are not valid, then your inferences and conclusions are necessarily flawed. The Consolidated Standards of Reporting Trials (CONSORT) guidance[2] on reporting clinical trials is helpful during the design stage and well-designed studies should be able to address all of the reporting requirements of the CONSORT checklist. During my 10-year statistical career in the pharmaceutical industry, there was a push to have statisticians write the statistical analysis plan concurrently with the development of the protocol (which was not heartily endorsed by the ranks of clinical statisticians!). However, the premise was sound: careful consideration of how you will analyze, report, and interpret your clinical trial during the design stage sharpens your thinking about the primary questions you wish to answer and whether the design allows you to answer them.

Analysis

The analysis of a clinical trial begins with matching statistical method to study aims, design and outcome (Table 1). For example, in a single-center single-arm phase 1 trial of AZD6094 (Volitinib) in combination with docetaxel for advanced gastric adenocarcinoma patients (NCT02447406), the goal is to estimate the dose based on toxicity. Hence, estimation of the maximum tolerated dose according to the study design (eg, classical 3+3 design[3] or the more contemporary continual reassessment method, which uses Bayesian methods[4]) and 95% confidence or prediction intervals would be reported and interpreted to inform the next study. There would be no need for statistical tests and inferential conclusions. For a randomized, placebo-controlled phase II crossover trial of GSK962040 on esophageal function (NCT01366560), we

Most current article

© 2016 by the AGA Institute
0016-5085/$36.00
http://dx.doi.org/10.1053/j.gastro.2016.02.065

Table 1. Analysis Paradigm

1. Think backward: consider your manuscript—analysis, reporting, interpretation—while you are designing the study
2. Mix and match: select statistical methods appropriate to study aims, design and outcomes

Study Aims	Design		Outcome
Estimation	Randomization		Continuous
Superiority	Control group	Single-arm	Dichotomous
Equivalence		Crossover	Categorical
Noninferiority		Parallel group	Ordered categorical
	Blinding		Time to event

may seek to assess preliminary activity of the change from baseline in lower esophageal sphincter pressure, a continuous outcome, by using a mixed effects linear model.[5] Formal statistical tests provide evidence whether the new agent improves esophageal function, with estimates of the treatment differences, 95% confidence intervals and *P* values. Determining noninferiority of simeprevir versus telaprevir with peginterferon and ribavirin to improve a dichotomous outcome such as sustained virologic response in patients with chronic hepatitis C infection requires the assessment of whether the 1-sided 95% lower confidence bound includes or excludes the noninferiority margin of 12%.[6]

The statistical techniques used in the analysis of clinical trials can be similar to those used in other experiments; however, I emphasize 2 areas that I have found to be important and yet often misunderstood or misused in analysis of clinical trials by clinical researchers: (1) methods to handle missing outcomes and (2) subgroup analyses.

Methods for Missing Outcomes

Simple methods to deal with missing data, such as analysis of complete cases and last-observation-carried-forward imputation, are valid only under a very restrictive assumption that the missing outcome data are unrelated to the study variables (ie, missing completely at random).[7] A more realistic assumption of missing at random (where other variables from the study can account for differences in the primary outcome for observed and missing subjects) requires more complicated statistical methods, yet the software is readily available and the interpretation is relatively straightforward. For situations where the data are not missing at random, collaboration with a statistician is necessary to properly analyze the data and provide valid conclusions. A clear and concise description of the problem of missing data in clinical trials and how to address it is provided by Little et al.[8] They recommend methods that are model based (eg, multiple imputation and weighted estimating equations) that provide standard errors and *P* values that incorporate the uncertainty about the missing data. Because assumptions about the missing data mechanism cannot ever be definitively proven by the observed data from the clinical trial, sensitivity analyses to assess the robustness of the conclusions to different assumptions and methods are recommended.

Subgroup Analyses

A subgroup analysis is an analysis based on fewer subjects than actually were assigned randomly. A proper subgroup is based on prerandomization characteristics (eg, genotype), whereas an improper subgroup is based on characteristics determined after randomization (eg, postoperative infection). Analyses based on proper subgroups are more defensible than those based on improper subgroups.[9] Subgroups analyses can be wrong in 2 ways: (1) falsely indicating treatment is beneficial in a particular subgroup when the trial shows no overall effect (the most common subgroup analysis) and (2) falsely indicating there is no treatment effect in a particular subgroup when the trial shows a benefit overall.

Separate tests of treatment effect in different subgroups do not provide evidence of treatment differences because they are generally underpowered and the many statistical comparisons performed increase the likelihood that an unusual result arises purely by chance (inflation of the type I error). Evidence of treatment effect differences is better supported with a "global" test of heterogeneity (eg, treatment–subgroup interaction test). However, even the interaction test suffers from lack of power and type I errors (false-positive findings). One of my favorite examples of this phenomenon is reported in the International Studies of Infarct Survival (ISIS-2), a placebo-controlled study of streptokinase and aspirin in 17,187 patients with acute myocardial infarction.[10] The overall treatment effect for aspirin was highly significant ($P < .0001$) with a 23% reduction in the odds of vascular mortality. In addition, the authors considered the effect of aspirin across 26 different subgroups. Interestingly, patients born under the astrological sign of Gemini or Libra and patients with prior myocardial infarction showed no evidence of benefit with aspirin in a subgroup-specific analysis. The test of interaction in both cases also showed a significant test of heterogeneity ($P < .05$). In this case, the result should be interpreted as a chance finding, or otherwise treatment would need to be administered according to astrological birth sign! This example underscores a key principle of subgroup analyses: a plausible biological mechanism, in addition to appropriate statistical significance testing, is needed before treatment heterogeneity can be concluded.

Careful planning of the trial with a few predefined, justified subgroups and transparent reporting of when a post hoc analysis is performed so that the reader can

Download English Version:

https://daneshyari.com/en/article/3292169

Download Persian Version:

https://daneshyari.com/article/3292169

Daneshyari.com