



Keynote Lecture 5

Clinical study design in evaluating antibacterial agents

Michael R. Jacobs, MD, PhD

Case Western Reserve University and University Hospitals
Cleveland, OH, USA

Treatment recommendations for infectious diseases should be based on scientifically valid, evidence-based conclusions, and it is therefore worthwhile to examine the underpinnings of the clinical trial study designs that contribute to treatment guidelines and approval of antimicrobial agents by regulatory agencies. The randomized, double-blind, controlled clinical trial, with sample sizes providing adequate statistical power, is often considered the “gold standard” of objective evidence wherein the study agent is demonstrated to be superior to a placebo or an active control. In these studies the statistical hypothesis being tested is “agent A is superior to placebo or agent B,” and is referred to as H1. The default state of “no difference” is called the “null” hypothesis, referenced as H0. After data are gathered, statistical tests appropriate to the test model are done and if the difference between the groups (referred to as “delta” or δ) is above a certain point it can be considered to have reached statistical significance, usually, but not necessarily, when the p-value is less than 0.05. An important factor in statistical significance is the sample size, which is a major part of determining the statistical power (or robustness of the conclusions) of a given trial. An inadequate sample size may not provide the ability to discern a true difference between two treatments from the differences that result from random chance. The sample size needed to show a difference in superiority outcome studies is calculated from a complex formula which includes expected outcomes of the two arms, study power (typically 80% or 90%), and p-value (typically 0.05). As an example, if standard therapy of a disease produces a 75% rate of successful clinical outcome and the new therapy is expected to produce a 85% rate of successful clinical outcome, 251 patients per arm are needed to show this difference at 80% power and a double-sided p-value of 0.05. A more stringent study, using 90% power, would require 335 patients per arm.

Alternatively, when testing new a clinical treatment, rather than using placebo, which is in many cases considered unethical, one of the standard of care agents is used as a comparator in comparative or “non-inferiority” trials. In these, the alternate hypothesis (H1) is that the test drug will perform “as well as” the standard drug. In this study design, sample size determinations are made using similar methods as for the standard superiority trial logic, and statistical tests that fail to find a difference are used to reject the null hypothesis and claim that the agents are equivalent in activity. Statistical tests typically used are a p-value of greater than 0.05, or that the upper and lower limits of the 95% confidence interval of the difference in outcome between the two arms should not include 0. The sample size needed to show a difference in noninferiority outcome studies is calculated from a formula similar to that used for superiority outcome studies, and includes expected outcomes of the two arms (which are set at the same value), difference between the two arms that will be considered noninferior (e.g., a 5% difference), study power (typically 80% or 90%), and p-value (typically 0.05). As an example, if standard therapy of a disease produces a 85% rate of successful clinical outcome and the new therapy will be considered noninferior if it produces a rate of successful clinical outcome varying by no more than 5%, 804 patients per arm are needed at 80% power and a double-sided p-value of 0.05. A more stringent study, using 90% power, would require 1073 patients per arm. If the variation in outcome allowed is increased to 10%, then the sample size is much lower - only 201 or 269 patients per arm are needed at 80% or 90% power, respectively. If the variation allowed is even more liberal at 15%, then only 90 or 120 patients per arm are needed at 80% or 90% power, respectively. The variation allowed by regulatory agencies is a subject of considerable debate as this has a major impact on study size and, therefore, cost. There are, additionally, several assumptions are made when using a noninferiority study design as the trial itself is not able to distinguish active from inactive therapy. It must be assumed, based on historical experience, that the standard agent is superior to placebo, and, furthermore, that the study population is comparable to study populations used when the standard agent was shown to be superior to placebo. Unless its true efficacy is known, an agent cannot be used as a valid comparator. Comparative studies generally fulfill these criteria when severe or life-threatening

infections are studied, such as bacteremia, endocarditis and meningitis, but not when self-limiting diseases such as acute sinusitis and acute otitis media are studied. These problems have become evident to drug approval agencies such as the US Food and Drug Administration (FDA) and the European Medicines Agency (EMA) Committee for Medicinal Products for Human use (CHMP), both of which are currently re-evaluating their non-inferiority testing guidelines. Until these issues have been resolved, bacteriologic outcome studies, such as double tympanocentesis acute otitis media studies, and application of pharmacokinetic/pharmacodynamic parameters from animal models to human infections are currently considered the best parameters available for assessing the clinical utility of antimicrobials in self-limiting diseases.