Overview of Recent Studies of Community-Acquired Pneumonia

K. Higgins, M. Singer, T. Valappil, S. Nambiar, D. Lin, and E. Cox
US Food and Drug Administration, Center for Drug Evaluation and Research, Silver Spring, Maryland

All recent studies of antibacterial drugs for the indication of community-acquired pneumonia submitted to the US Food and Drug Administration have been designed as noninferiority studies. We provide a summary of results of 7 recent clinical studies of oral antibacterial drugs for treatment of community-acquired pneumonia. In these 7 studies, the majority of patients enrolled had Pneumonia Patient Outcomes Research Team scores of I or II. The percentage of randomized subjects with pathogens identified at baseline ranged from 47% to 76%, and the percentage of subjects with *Streptococcus pneumoniae* isolated at baseline ranged from ~6% to 20%. The primary end point in these studies was clinical cure, assessed 7–21 days after completion of therapy. Clinical cure rates were >80% in the intent-to-treat populations and >90% in the per-protocol populations. We also briefly summarize the results from several recently submitted clinical studies of intravenously administered antibacterial drugs for treatment of community-acquired pneumonia, in which we found similar results.

All studies recently submitted to the US Food and Drug Administration (FDA) for review of drugs for the indication of community-acquired pneumonia (CAP) have been designed as noninferiority studies. We provide an overview of several CAP studies that have been submitted recently to the Office of Antimicrobial Products, Center for Drug Evaluation and Research, FDA. The objective of these studies was to demonstrate the efficacy of a test drug by comparing it with a drug currently approved for treatment of CAP; efficacy of the test drug is inferred by showing that it is “noninferior” to, or not a specified degree worse than, the control drug. For noninferiority studies to be informative, they need to be well designed and conducted and supported by adequate historical information, especially about how effective the control drug is in the treatment of CAP. Before giving the overview of CAP studies, we briefly summarize issues to be considered in the interpretation of noninferiority studies.

NONINFERIORITY STUDIES OF CAP

The goal of a noninferiority study is to infer the clinical effect of a test drug by demonstrating that the test drug is similar enough to (or not a specified degree worse than) the control drug. For the finding of noninferiority of the test drug to the control drug to be informative, one needs to know whether the control drug would have been better than placebo (and by how much) if a placebo been included in the trial. This point can be illustrated by examining some hypothetical examples.

For a mild disease that spontaneously resolves without any treatment in most patients within 2 weeks, the demonstration that a test drug performs similar to a control drug at the 2-week time point is not informative; a placebo could have also performed similar to the control drug in this situation. If the disease being studied is one that is more severe and spontaneous resolution occurs infrequently within 2 weeks, the finding of similarity of a test drug and a control drug at the 2-week time point is informative, because a placebo would not have performed as well as the control drug in this situation. This fundamental limitation and cont-
cern regarding noninferiority studies—that is, whether the finding of similarity of a test drug and a control drug is informative—is explicitly stated in the Code of Federal Regulations: “If the intent of the trial is to show similarity of the test and control drugs, the report of the study should assess the ability of the study to have detected a difference between treatments. Similarity of test drug and active control can mean either that both drugs were effective or that neither was effective. The analysis of the study should explain why the drugs should be considered effective in the study, for example, by reference to results in previous placebo-controlled studies of the active control drug” [1, p. 148]. If this cannot be done, one cannot conclude efficacy from an active-controlled study that claims to show noninferiority.

Data from placebo-controlled studies of the active control are the ideal means of understanding how the control drug performs compared with a placebo. One can use data from placebo-controlled trials of the control drug to quantitatively estimate the amount by which the control drug is more effective than a placebo. Knowing the improved effect of the control drug over placebo allows one to select a quantitative margin of how “similar” the test drug and the control drug should be to reliably conclude that the test drug is efficacious. However, for CAP, results from contemporary placebo-controlled studies do not exist, because antibacterial drug therapy has been accepted as the standard of care for CAP since the late 1930s or 1940s [2]. When recent placebo-controlled trials are not available, it may be possible to estimate what the effect of a control drug is compared with what the effect of placebo would be, by examining other historical information on the disease of interest. For example, historical information on the natural history of disease (information from the pre–antibacterial drug era on the course of disease in the absence of treatment) may help to inform plausible estimates of the effect of a control drug (i.e., the effect of treatment), compared with a placebo or no treatment.

It is also important for a noninferiority study to be well designed and rigorously conducted. Errors or any factors in the conduct of the study that result in carelessness or loss of accuracy could cause the 2 treatment arms to be more similar than they actually are. In noninferiority studies, this can result in false conclusions about efficacy. Unlike noninferiority studies, superiority studies have their own built-in quality control. For example, a superiority study of an antibacterial drug that enrolls subjects with viral diseases, rather than bacterial diseases, will be unlikely to show superiority of an antibacterial drug over placebo, because subjects do not have a condition that responds to the antibacterial drug therapy. In this situation, a placebo would perform similar to the antibacterial drug, because antibacterials have no effect on viral disease. However, if a study is conducted under the same circumstances (i.e., a study of an antibacterial drug that enrolls patients with viral illness) but has a noninferiority design, the finding that the test drug performs similar to (not a specified degree worse than) the control drug may lead to the erroneous conclusion that the test drug has demonstrated efficacy. If a placebo had been included in the study that enrolled patients with viral disease, one could have observed that the test drug, control drug, and placebo all performed similarly. A noninferiority study that enrolls patients without the disease of interest is uninformative because the study does not have the capacity to distinguish an active therapy from an inactive therapy. Other errors in the conduct of a noninferiority study could also lead one to falsely conclude that efficacy has been demonstrated.

The challenges in the design and interpretation of noninferiority studies have been widely discussed in many publications [3–9], and this article does not review these issues in detail. More information on issues regarding noninferiority studies is also available in the International Conference on Harmonisation guidance documents E9, “Statistical Principles for Clinical Trials” [10], and E10, “Choice of Control Group and Related Issues in Clinical Trials” [11].

**OVERVIEW OF RECENT STUDIES OF ORAL ANTIBACTERIAL DRUGS FOR TREATMENT OF CAP**

This section reviews 7 studies recently submitted to the Office of Antimicrobial Products to support indications for CAP. These are all comparative studies that were conducted within the past 8 years and ranged in size from ∼300 to 500 subjects. The control drug varied among studies and included clarithromycin, amoxicillin-clavulanate, and levofloxacin. All studies were randomized, double-blind studies designed to show noninferiority to an approved product, and all closely followed the FDA’s 1998 draft guidance document “Community-Acquired Pneumonia—Developing Antimicrobial Drugs for Treatment” [12].

In general, the diagnosis of CAP in these studies was based on the presence of a new infiltrate on chest radiograph and at least 2 of the following signs and symptoms: cough, sputum production, auscultatory findings, dyspnea or tachypnea, fever, elevated WBC count, or hypoxemia. The inclusion and/or exclusion criteria varied slightly among studies, and some studies limited enrollment to patients with Pneumonia Patient Outcomes Research Team (PORT) scores of I–II or I–III [13]. Microbiological evaluation was performed for each patient, although isolation of a pathogen was not required for overall evaluable.

Patients were assessed for outcome at the test-of-cure visit, which, in most studies, occurred 7–21 days after completion of therapy. The primary end point was clinical outcome. A clinical outcome of “cure” was defined as complete resolution
or improvement of all signs and symptoms of pneumonia and improvement or lack of progression of all abnormalities seen on chest radiograph, such that no additional antibacterial therapy was required. All the studies allowed for improvement of symptoms to be considered a “cure.”

Although the test-of-cure visit occurred 7–21 days after completion of the study treatment, the protocol, in most cases, specified that subjects should be seen or contacted earlier, typically at the end of study therapy. All patients determined to have a treatment failure at the end-of-therapy visit were automatically considered to have had a treatment failure at the test-of-cure visit. However, the requirement of an earlier visit (i.e., an end-of-therapy visit) was not consistent among studies, and the degree of follow-up for this visit was difficult to determine for most studies.

Microbiological response outcome measures included eradication (defined as the absence of the original pathogen in a culture specimen obtained at the test-of-cure visit), presumed eradication (defined as clinical cure without a culture specimen obtained at the test-of-cure visit), persistence (defined as presence of the original pathogen in the culture specimen obtained at the test-of-cure visit), or presumed persistence (defined as clinical failure without culture of a specimen obtained at the test-of-cure visit). Microbiological success was defined as either eradication or presumed eradication, and microbiological failure was defined as either persistence or presumed persistence.

The following 4 analysis populations were often defined in the protocol or discussed in the FDA reviews.

1. Intent-to-treat (ITT) population: all randomized subjects
2. Per-protocol population: all subjects in the ITT population who do not have any major protocol violations (sometimes called the “clinically evaluable” population)
3. Modified ITT population: all ITT subjects with a pretreatment pathogen isolated
4. Microbiologically evaluable population: all subjects in the modified ITT population who do not have any major protocol violations

Because not all subjects in these studies were required or expected to have a pretreatment pathogen isolated, analyses using the modified ITT and microbiologically evaluable populations were usually considered as sensitivity analyses only.

Regarding which population should be considered the primary analysis population, many believe that the per-protocol population is the most relevant for noninferiority studies, because it excludes subjects who may otherwise cloud the ability to see a treatment effect, if one exists. For example, if some subjects in each arm did not receive a minimally effective dose of therapy, the inclusion of these subjects in the analysis may have the effect of making the 2 treatment arms look more similar than they actually are. Therefore, for a noninferiority study, many argue that the per-protocol population is the more valid population, whereas the ITT population, which does not exclude these subjects, may be a “less conservative” population by comparison. On the other hand, there have been objections to use of the per-protocol population as the primary analysis population. This population excludes subjects after randomization for events—typically protocol violations—that occur during the trial itself. Some of these exclusion events may be drug related, thereby causing the per-protocol population to become a biased population and potentially losing much of the benefit of randomization. (Note that events that occur after randomization cannot be ruled out as being possibly related to the randomized treatment. For example, subjects could be lost to follow-up, and the reason they did not return for follow-up could be lack of drug effect.)

Overall, for noninferiority studies, there are drawbacks to analyses based on either the per-protocol population or the ITT population, which is why the per-protocol and ITT populations are often considered equally important in the assessment of the results of a noninferiority study. Note that, for any study, there should be consistency in results of analyses using each of these 2 populations, and, if large differences in outcome are seen between the 2 analyses, this would likely lead to concerns about the interpretation of the overall study results. When there are differences in the outcomes of analyses of the per-protocol and ITT populations in a noninferiority study, the reasons for these differences should be evaluated.

The percentage of ITT subjects who were excluded from the per-protocol data set in the 7 studies we reviewed varied from $0\%$ to $\sim20\%$ (figure 1). The differences in the percentage of subjects excluded could have been due to differences in how
strictly investigators followed the protocols, differences in the patient populations, or differences in the strictness of criteria for entry into a per-protocol population. Reasons for exclusion varied slightly among studies but included insufficient signs and symptoms, insufficient radiographic findings, withdrawal or loss of subjects, adverse events leading to discontinuation of treatment, inadequate dosing, test-of-cure visit occurring outside the time window, indeterminate clinical outcome, use of concomitant antimicrobial (used not because of failure), and, in some cases, death not due to CAP.

The primary method of assessing noninferiority in these 7 studies was to construct a 2-sided 95% CI for the difference in cure rates (for the test drug vs. the control drug) for both the ITT and the per-protocol populations. To conclude that there is noninferiority, the lower bounds of both 95% CIs would need to be greater than −10% or −15%. The noninferiority margins for these studies, −10% and −15%, were typically prespecified, and their acceptability was determined mainly from “clinical judgment,” in the absence of a justification for the noninferiority margin on the basis of data. After a reexamination in 2006 of the scientific basis for the noninferiority-margin selection, sponsors studying CAP in 2006 or subsequently were asked by the Office of Antimicrobial Products to provide a data-driven justification for the noninferiority margin proposed for each specific study. This approach is outlined in the draft guidance document “Antibacterial Drug Products: Use of Noninferiority Studies to Support Approval” [14], which asks sponsors to provide adequate evidence to support proposed noninferiority margins for any indication for which active-controlled studies are designed to show noninferiority and to support drug approval.

All 7 studies were multinational studies, with 5 studies including subjects from the United States. Subjects enrolled from the United States constituted ≥50% of the study population in 2 studies. The number of countries per study ranged from 3 to 14. The number of sites per study ranged from ∼40 to 80. The age range of subjects was 18–98 years. The mean and median ages were 46 and 45 years, respectively, with the age of the middle 50% of the population ranging from 35 to 55 years. Most studies limited enrollment to subjects aged ≥18 years.

The PORT scores for subjects enrolled into these studies are summarized in figure 2. These studies all evaluated oral antibacterial drugs for treatment of CAP, and, accordingly, many studies limited enrollment to subjects with certain PORT scores (e.g., PORT scores of I–II or I–III). As a result, most enrolled subjects had PORT scores of I or II. The percentage of subjects with PORT scores of III or higher ranged from ∼5% to 10%. In some studies, PORT scores were calculated on the basis of data that might have been incomplete; calculation of a PORT score when data are missing may result in a subject being assigned to a category lower than the category that would have been assigned if a complete data set had been available.

For baseline signs and symptoms, 97%–100% of subjects in the 7 studies had cough at baseline, and 75%–100% had sputum production, as was expected because cough and sputum production were requirements for entry in some studies. The percentage of subjects with fever ranged from 19% to 98%, although 98% value is an outlier reflecting that fever was an inclusion criterion for that particular study. The percentage of subjects with chills ranged from <2% to 69%, with 69% an outlier—all other studies had rates of <6% for chills at baseline. The percentage of subjects with dyspnea varied greatly among studies, from 18% to 100%, and the percentage with chest pain varied from 41% to 76%. Multilobe involvement, determined on the basis of chest radiograph, was seen in ∼20% of subjects in all studies. Note that all studies required a new infiltrate on chest radiograph as an inclusion criterion. Bacteremia was uncommon; 0%–8% of subjects were bacteremic, with 0%–2% of subjects having Streptococcus pneumoniae bacteremia.

The percentage of all randomized subjects with pathogens identified at baseline ranged from ∼45% to 75% (figure 3). All patients should have been screened for a pathogen at entry, but, as stated above, isolation of a pathogen at entry was not required for overall evaluable. Note that the subpopulation of patients with a baseline pathogen (including those with a serological diagnosis of a baseline pathogen) comprises the population of subjects included in the modified ITT population.

Examination of the 5 most common pathogens isolated at baseline shows considerable variation among studies (figure 4). Mycoplasma pneumoniae, S. pneumoniae, and Chlamydia pneumoniae were usually the more frequently identified baseline pathogens. If we examine rates of isolation of S. pneumoniae
among these studies, we find that 6%–20% of patients enrolled in these studies had *S. pneumoniae* as a baseline pathogen (data not shown).

The results of the studies in terms of clinical cure were very similar. Clinical response at the test-of-cure visit in the ITT population, with missing data imputed as failures, showed success rates of $>80\%$ for all treatment arms. Most studies showed similar results between the 2 arms in the study. The cure rates for the per-protocol populations were higher than for the ITT populations, with all studies having cure rates for the per-protocol population of $\geq 90\%$.

The comparative results for the difference in clinical cure rates between the test drug and the control drug in each of the studies for the per-protocol and ITT populations are very similar (figure 5). All these studies would have been able to claim noninferiority with a 15% margin, and 5 studies would have been able to claim noninferiority with a 10% margin. Note that there is no clear pattern as to which analysis population, the per-protocol or ITT population, leads to more-conservative results (i.e., a lower value for the lower bound of the 95% CI).

Microbiological response is also usually evaluated in CAP studies. However, culture specimens are infrequently obtained at the test-of-cure visits, because of the nature of the disease. When a culture specimen either is not obtained or cannot be obtained at the test-of-cure visit and the patient is classified as having experienced a clinical cure, the case is classified as a presumed microbiological eradication. In all these studies, the vast majority of microbiological responses were either presumed eradication or presumed persistence, on the basis of whether a subject experienced clinical cure or clinical failure, respectively.

Mortality rates in these studies were low (<2%), as would be expected for studies of outpatients with CAP. Although death was the outcome assessed in most of the historical studies of outcomes in CAP from the 1930s and 1940s, mortality is not a plausible end point to use in present-day studies of mild-to-moderate CAP. The additional measures taken when therapy is failing (e.g., a switch to another, nonstudy antibacterial therapy) make the rates of death in current studies not comparable to the rates of death in historical studies. Additionally, issues arise in determining how the end point of mortality would be analyzed when death is used as the primary outcome. It is reasonable to expect that any subject whose therapy was failing would receive additional, nonstudy antibacterial therapy. If any of these subjects subsequently survived, one would then need to determine how the subjects should be analyzed in the primary analysis of mortality. If the analysis considered only death as failure, and given that these patients survived, then the analysis would capture efficacy of any rescue therapy used along with the randomized treatment (i.e., falsely attributing the beneficial effect of the rescue therapy to the efficacy of the randomized therapy). Alternatively, if the analysis considered receipt of additional rescue therapy as failure, as well a death, the analysis would be very similar to the analysis of a clinical-response end point.

**BRIEF OVERVIEW OF RECENT STUDIES OF INTRAVENOUS ANTIBACTERIAL DRUGS FOR TREATMENT OF CAP**

Fewer studies of intravenous antibacterial drugs for treatment of CAP than studies of oral antibacterial drugs have been sub-
The clinical response rates in studies of intravenous drugs for treatment of CAP were ~80% for the ITT populations and 90% for the per-protocol populations. These rates were, in general, slightly lower than those in studies of oral drugs for treatment of CAP. Mortality rates in the studies of intravenous drugs ranged from ~2% to 4%.

**SUMMARY**

All the studies in this overview of trials recently submitted to the FDA were active-controlled trials that showed noninferiority. In general, the studies enrolled patients with signs and symptoms consistent with CAP. As would be expected, the severity of illness of subjects enrolled in trials of oral antibacterial drugs for treatment of CAP was generally less than that of subjects in trials of intravenous drugs. The percentage of subjects with a baseline pathogen identified ranged from 45% to 75% in studies of oral antibacterial drugs and from 30% to 55% in studies of intravenous antibacterial drugs. Although each of these studies showed that the test drug is “similar” to the control drug, uncertainty remains regarding what conclusions can be drawn from a noninferiority study when evidence to support a noninferiority margin has not been provided. For studies of CAP, this is particularly challenging, given the absence of contemporary placebo-controlled studies. We hope that the information provided in this overview will be informative to those interested in previously conducted studies of CAP and that it might be helpful in the design and conduct of future trials of treatments for CAP.

**Acknowledgments**

We thank Steven Gitterman and Robert O’Neill for comments on the manuscript. We also thank Thomas Fleming and David Gilbert, the co-chairs, along with E.C., of the workshop cosponsored by the Infectious Diseases Society of America and the US Food and Drug Administration (17–18 January 2008).

**Supplement sponsorship.** This article was published as part of a supplement entitled “Workshop on Issues in the Design and Conduct of Clinical Trials of Antibacterial Drugs for the Treatment of Community-Acquired Pneumonia,” sponsored by the US Food and Drug Administration and the Infectious Diseases Society of America.

**Potential conflicts of interest.** All authors: no conflicts.

**References**