



Published in final edited form as:

Clin Infect Dis. 2010 August 1; 51(Suppl 1): S18–S28. doi:10.1086/653036.

Recommendations for Improving the Design, Conduct, and Analysis of Clinical Trials in Hospital-Acquired Pneumonia and Ventilator-Associated Pneumonia

John H. Powers

Scientific Applications International Corporation, Collaborative Clinical Research Branch, National Institute of Allergy and Infectious Diseases, National Institutes of Health; George Washington University School of Medicine; and University of Maryland School of Medicine, Bethesda, Maryland

Abstract

Overall decisions on the clinical use of new antimicrobials depend on the validity and reliability of the evidence from appropriately designed, conducted, and analyzed clinical trials. Because pneumonia is the sixth leading cause of death in the United States and the leading cause of infectious disease–related death, appropriate design of trials in hospital-acquired pneumonia and ventilator-associated pneumonia are an important public health issue. Several issues with the current design of trials in hospital-acquired pneumonia and/or ventilator-associated pneumonia potentially bias their results and raise questions about their validity. These issues are magnified in the context of noninferiority trials, in which bias can make interventions appear more similar, giving false-positive results of safety and effectiveness. The goal of this article is to provide a scientific basis for improving the validity, reliability, and efficiency of clinical trials in hospital-acquired pneumonia and/or ventilator-associated pneumonia to provide better information for decision making for patients, clinicians, regulators, and other stakeholders.

In 1943, Hopkins [1] published the results of a trial of treatment of the common cold that tested a new agent, patulin, derived from the mold *Penicillium patulinum*. Hopkins alternately assigned 180 participants to receive patulin or placebo. The trial used no specific definition of disease, no blinding, and no control for concomitant medications and used an outcome measurement of clinicians' judgment of complete resolution of signs and symptoms (not specifically defined). The results of the trial showed that 55 (58%) of 96 persons were clinically cured at 48 h with patulin, compared with 8 (9%) of 85 persons with placebo, for a treatment difference of 48% (95% confidence interval, 35%–60%; $P < .002$)

© 2010 by the Infectious Diseases Society of America. All rights reserved.

Reprints or correspondence: Dr John H Powers, SAIC-Frederick, 6700B Rockledge, Dr, Rm 1123, Bethesda, MD 20892 (powersjohn@mail.nih.gov).

The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the US Government.

Potential conflicts of interest. J.H.P. has received consulting fees from Acureon, Advanced Life Sciences, Astellas, Astra-Zeneca, Basilea, Centegen, Cerexa, Concert, Cubist, Destiny, Forest, Gilead, Great Lakes, Johnson and Johnson, LEO, Merck, Methylgene, MPEX, Pharming, Octopus, Takeda, Theravance, and Wyeth.

Supplement sponsorship. This article was published as part of a supplement entitled “Workshop on Issues in the Design of Clinical Trials for Antibacterial Drugs for Hospital-Acquired Pneumonia and Ventilator-Associated Pneumonia,” sponsored by the US Food and Drug Administration, Infectious Diseases Society of America, American College of Chest Physicians, American Thoracic Society, and the Society of Critical Care Medicine, with financial support from the Pharmaceutical Research and Manufacturers of America, AstraZeneca Pharmaceuticals, and Forest Pharmaceuticals.

in favor of patulin. Hopkins described the results as “dramatic.” However, during a time of war, the British government was unwilling to spend scarce resources to purchase and administer the drug for treatment of a very common illness without confirmatory evidence of its safety and effectiveness. The Medical Research Council (MRC), in noting the potential biases in the Hopkins trial, designed a double-blinded, placebo-controlled trial, which was one of the first to use random numbers in assigning the larger sample size of 1449 participants to study interventions [2]. The trial also included specific definitions of disease and outcomes. The MRC trial showed rates of success at 48 h of 13% (87 of 668 patients) for patulin and 13% (88 of 680 patients) for placebo, for a treatment difference of 0% (95% confidence interval, -4.0% to 4.0%; $P = .96$).

The dramatic lack of effect of patulin in the MRC trial, compared with the results observed in the Hopkins trial, are in large part attributable to the differences in methodology of the 2 studies. The MRC trial used methods to control for various biases that could obscure a causal relationship between the effect of the drug and outcomes. The importance of proper evaluation of a drug such as patulin becomes clearer today because of the evidence that shows the potential carcinogenicity of the compound, which is included on the World Health Organization list of biological foodborne hazards [3]. Despite the small P value, the public health impact could have been large because of the preventable malignancies without benefit if investigators in the 1940s had accepted the “dramatic” results of the Hopkins study as evidence of a large benefit for patients. Imagine if the next agent for the common cold had been compared with patulin in a noninferiority trial that showed similarity of the new agent to patulin. The result could have been a succession of ineffective and potentially toxic agents in a very common disease. This highlights the issue that appropriate trial design protects patients by providing clinicians and patients with the best information on which to base therapeutic decisions. It also shows not only the impact on the drug being studied, but also the downstream impact that inappropriately designed trials might have on an entire therapeutic area. Of importance, it highlights that appropriate trial design and scientific validity are necessary considerations in the ethics of clinical research.

The same concerns with methodology of clinical trials hold true today. Clinicians need valid and reliable results from clinical trials that minimize as much as possible the influence of bias on those results. Validity refers to the capacity of a trial to measure what it purports to measure [4]. Validity is not measured only by publication in peer-reviewed journals or consensus, because studies document the issues with publication bias and design flaws in many published clinical studies. Reliability is the ability to obtain similar results with confirmatory studies [4]. However, it is not surprising to obtain confirmatory results from studies with similar biases. This results only in trials that provide reliably incorrect inferences. The importance of obtaining valid and reliable results of effectiveness and harm is magnified in serious and life-threatening diseases, such as hospital-acquired pneumonia (HAP) and ventilator-associated pneumonia (VAP), for which ineffective or less effective drugs for treatment or drugs that cause excess harm can result in avoidable deaths.

Over the past several years, the US Food and Drug Administration (FDA) has acknowledged the need to readdress clinical trial designs to improve the ability of clinical trials of various infectious diseases to provide valid and reliable results. There is also a need to harmonize the recommendations in previous FDA anti-infective guidance with subsequent general FDA and international guidance that has superseded it [5,6]. Concerns about the influence of bias on trial results are magnified in the context of noninferiority trials [7,8]. The same biases that would result in false-negative results in superiority trials make interventions appear to be more similar and show false-positive results in noninferiority trials. Several recent drugs studied for treatment of HAP and VAP have had unclear evidence of effectiveness, especially for VAP. At a July 2008 FDA advisory committee meeting, participants discussed

data that showed increased mortality associated with doripenem treatment for VAP [9]. In other studies, tigecycline and ceftobiprole had lower success rates for VAP [10,11]. The overall results of these trials still showed noninferiority, compared with the control regimens, in the combined populations of persons with HAP and VAP with an end point of clinical success; however, this raises questions about whether the design of these trials was incapable of showing differences in the population with HAP if differences existed, whereas the population with VAP was more capable of demonstrating any differences. In addition, several aspects of current clinical trial design in HAP and/or VAP raise questions about the validity of the results of trials, especially noninferiority trials [12–21]. These aspects include unclear study objectives, unduly large noninferiority margins, prior and concomitant active antimicrobial therapy, unclear outcome definitions, and large numbers of participants excluded from analyses. Because of these issues, addressing clinical trial design in HAP and VAP is a timely concern.

This article will outline the legal and scientific standards for FDA approval of drugs and apply those criteria to clinical trials in HAP and VAP, focusing on issues in the design of appropriate noninferiority trials. This article also will make recommendations for changes in the clinical trial design, conduct, and analysis of HAP and VAP trials. The goal of this article is to provide a scientific basis for improving the validity, reliability, and efficiency of clinical trials in HAP and VAP to provide better information for decision making for patients, clinicians, regulators, and other stakeholders.

BACKGROUND AND STANDARDS FOR FDA DRUG APPROVAL

The overall decision about clinical usefulness of a medical intervention and FDA approval is the balance of harms and benefits of that intervention under the proposed conditions of use. FDA analysis of drug safety is based on “adequate tests by all methods reasonably applicable to show whether or not such drug is safe for use under the conditions prescribed, recommended, or suggested in the proposed labeling” [22]. In 1962, the US Congress passed amendments to the Food Drug and Cosmetic Act that required drug sponsors to show evidence of effectiveness to balance the harms associated with all medical interventions. The basis for drug effectiveness is substantial evidence from adequate and well-controlled trials. Congressional intent was that the FDA base drug approvals on scientific data rather than on testimonials, clinical impressions, practice experience, or data on drug sales [23]. The FDA published criteria in the Code of Federal Regulations in 1970 to outline specific definitions for adequate and well-controlled trials, which are still in force today [24]. The 7 criteria are (Table 1) (1) a clear statement of the objectives of the trial is present, (2) the study permits a valid quantitative comparison with a control regimen, (3) the study selects participants with the disease in treatment trials or at risk of the disease in prevention trials, (4) there is baseline comparability between the test and control groups, (5) the study minimizes the influences of bias on the results, (6) the study uses well-defined and reliable assessments of outcomes, and (7) the study uses appropriate methods of analysis of the data. Court cases have clarified that these criteria are a minimal standard for drug approval and that these criteria are not meant for prospective application only [23]. In other words, as new scientific information becomes available, the FDA should apply those criteria regardless of prior agreements with drug sponsors. The Food Drug and Cosmetic Act states that the FDA can change the parameters of agreement on trial design if “a substantial scientific issue essential to determining the safety or effectiveness of the drug has been identified after the testing has begun” [25].

The idea of applying current science was exemplified in a court case in 1970 involving antimicrobials, when Upjohn sued the FDA over withdrawal of approval for the antibiotic Panalba (a combination of tetracycline and novobiocin) [26]. Upjohn held that in vitro data,

animal studies, articles in the peer-reviewed medical literature (many of which were not controlled or poorly controlled), and clinician testimonials should suffice for continued approval because that was the standard at the time of initial FDA approval of the drugs. The court indicated that the “in vitro studies are suggestive of some effectiveness in laboratory experiments utilizing artificially cultured microorganisms or test systems, but because the studies are not at all correlated with clinical experience they cannot be used as a basis for concluding the drug will have the effectiveness claimed for them used to treat naturally occurring disease in man” [26, Appendix A, p12]. Indeed, although the 28% approval rate of investigational new drugs for antimicrobials exceeds that of any other therapeutic area, the lack of effectiveness and/or safety of the 72% of antimicrobials not approved shows the limitations of preclinical data, because all those drugs had promising in vitro and animal studies as the basis for filing an investigational new drug [27]. Therefore the “Bayesian prior” probability of antimicrobial safety and effectiveness is not as large as might be assumed.

The criteria for a valid study apply similarly to serious and life-threatening diseases as they do to self-resolving illnesses. In a landmark Supreme Court case related to the drug Laetrile, Justice Thurgood Marshall indicated that “The [Food Drug and Cosmetic] Act makes no express exception for drugs used by the terminally ill and no implied exemption is necessary in order to attain congressional objectives or to avert an unreasonable reading of the terms ‘safe’ and ‘effective.’ Nothing in the legislative history suggests that Congress intended protection only for persons suffering from curable diseases....To the contrary, in deliberations preceding the 1938 Act, Congress expressed concern that individuals with fatal illnesses such as cancer, should be shielded from fraudulent cures” [28, p 2].

Another point of confusion in clinical research that impacts on clinical trial design in HAP and/or VAP is separating clinical research from clinical practice. Some clinicians worry that changes in clinical trial design would deviate from accepted clinical practice. First, much of what is done in practice is based on currently available evidence, not evidence that is valid on the basis of the results of appropriately designed, conducted, and analyzed trials. A recent published survey of clinical practice guidelines in cardiology showed that guideline authors based 11% of recommendations on level A evidence (a standard commensurate with that of FDA approvals based on adequate and well-controlled trials), and 48% of recommendations were based on expert opinion [29]. Two similar analyses of infectious diseases guidelines noted that ~15% of recommendations were based on level I (at least 1 randomized trial) evidence. More than half of recommendations overall were based on expert opinion or case studies [30,31]. To determine whether expert opinion is indeed correct, it is incumbent on investigators to view areas on which treatment decisions are based on opinion as areas on which future research is needed, rather than barring future research based on recommendations in treatment guidelines. Second, treatment guidelines are based on the use of interventions already shown to be safe and effective in some setting, because they have already been approved by regulatory agencies (although not necessarily for the indications noted in guidelines). Treatment guidelines do not necessarily provide the evidence needed for designing valid trials of experimental agents. Last, use of treatment guidelines as a basis for clinical trial design reinforces the therapeutic misconception that research participants who volunteer to participate in a trial through informed consent are receiving treatment [32]. The Belmont Report in 1979 clearly separated clinical practice and clinical research [33]. The goal of clinical research is to develop or contribute to generalizable knowledge. The purpose of clinical practice is to enhance the well-being of an individual patient. The safety and effectiveness of new interventions is not clear (the basis for doing the study in the first place) and research participants are subject to some risk, thus the need for informed consent. Appropriate design of trials, including a thorough review of prior evidence, is meant to

minimize that risk, but referring to practice guidelines for which the evidentiary basis is unclear does not minimize risks for research participants or future patients.

Some clinicians propose that investigators or research volunteers will not choose to participate in clinical trials that are not designed according to current practice. There are several issues with this line of thinking. First, it seems to obviate scientific advances and places expert opinion as the standard for clinical trials and ethics, contrary to the definition of substantial evidence, as noted above, and trial ethics, as formulated in the Belmont Report. Second, the concern that individuals will not choose to enroll is not borne out in other therapeutic areas or by available evidence from patients themselves. The same concerns regarding participation and ethics occurred in trials of hormone replacement therapy in women, and that randomized trial was completed and showed that prevailing opinions based on observational data were not supported [34]. A recent survey of patients' willingness to participate in clinical trials showed that ~60% of those surveyed were willing to participate in a trial that included a placebo [35]. One hundred percent of those who would participate stated that their reason for participation was to support the growth of new treatments and to help other patients. Of those who would decline to participate in a trial, >75% said they were not opposed to the use of placebos in general but wanted to know the interventions that they were receiving. This wish to know the intervention that one is receiving applies to any kind of trial, not just one that uses accepted therapies or placebos. Last, a trial based on weak evidence and with a potentially biased trial design does not become ethical or valid merely because persons are willing to participate. Authors have indicated that such trials are merely precise measures of prevailing biases rather than contributions to generalizable knowledge [36]. Investigators and volunteers are always free to not participate in trials on the basis of informed consent and their personal beliefs. The concept of equipoise, however, does not mean that all persons agree that a trial is needed or that all persons would agree to participate, but that there is disagreement regarding the evidence of the benefits and harms of interventions [37].

APPLYING SUBSTANTIAL EVIDENCE CRITERIA TO HAP AND VAP TRIALS

Clear statement of study objectives

Current trials in HAP and VAP pool participants with these 2 diseases in the same trials. However, patients with HAP and VAP make up 2 different populations with different risks factors for mortality. There is also a different epidemiology of causative pathogens and different exposures to drugs in the 2 populations. As noted above, several recent trials of new antimicrobials showed differing results for the subsets of patients with HAP and VAP, with worse results in the group with VAP. Merging the populations can mask differences between the groups and make a drug appear to be noninferior overall, while clouding differences between drugs in the different populations. Because the FDA requires 2 clinical trials to support approval of treatment for a given disease in most circumstances, it makes logical sense to design one trial in HAP and another trial in VAP. Enrollment of participants in a single trial and analyses of the results for patients with HAP and VAP separately would still require adequate power in each subset, which would essentially be the same as conducting 2 separate trials; therefore, there seems to be little gain in efficiency of performing a single trial with 2 adequately powered subsets of patients with HAP and VAP.

Although clinical trials should be clinically relevant, it is important to separate explanatory trials from strategy trials. Explanatory trials evaluate the effectiveness of an agent, and a strategy trial evaluates overall outcomes in a setting where clinicians may use other drugs that make it challenging if not impossible to separate the effect of the study drug from other concomitant medications. Although strategy trials are useful, the first step is to ensure that a new intervention has an effect, and then perform future studies on how to use the drug

correctly in clinical practice. FDA approval is a first step in the lifecycle of a new antimicrobial, and clinicians will still need evidence on how a new antimicrobial fits into overall practice. Future trials should clearly evaluate the contribution of a new agent to overall effectiveness.

Quantitative comparison with a control group

Several issues arise with regard to choice of control groups. These include the decision on a superiority or noninferiority trial design and the drugs chosen. First, a decision must be made with regard to whether the trial will evaluate superiority or noninferiority of the test and control interventions. Noninferiority trials are only valid under the following conditions: (1) there is reliable and reproducible evidence of the effect of the control regimen, compared with no specific therapy, on the basis of historical studies; (2) the planned noninferiority trial conforms as closely as possible to the design of the studies that showed the effect of the control regimen in terms of disease definition, study populations, concomitant medication, and end point definitions and timing; and (3) the selected margin of potential inferiority of the test intervention compared with the control intervention is smaller than the effect of the control compared with no specific intervention and rules out a clinically meaningful difference between the test and control regimens [6,38].

There is historical evidence for a treatment effect of antimicrobials, compared with no specific therapy, in community-acquired pneumonia (CAP) [39]. Because of the same pathophysiology of both CAP and HAP and/or VAP, it is reasonable to apply the treatment effect sizes from CAP to HAP and/or VAP. The data from observational studies showing that inappropriate therapy for HAP and/or VAP is associated with increased mortality are affected by bias in terms of lack of baseline comparability. Patients who receive inappropriate therapy may differ from those who receive appropriate therapy in terms of important measured and unmeasured variables that might affect outcomes independent of treatment [40]. However, these data seem to support the treatment effect sizes shown in historical studies of CAP and reinforce the idea that the data from CAP are applicable to HAP and/or VAP. The treatment effect size for pneumonia from the historical data varies widely depending on the baseline characteristics of patients. Patients who are older, have bacteremia, and have comorbid illness have higher mortality rates independent of treatment (confounding) and benefit most from treatment, because the effect size for effective drugs is greater (effect modification) [39].

The data from CAP studies are only applicable for HAP and/or VAP when studies meet the second criteria for valid noninferiority trials (ie, that their design is similar in important aspects to the trials that showed the effects of the control regimens). Trials in HAP and/or VAP should use all-cause mortality as an end point and enroll individuals whose baseline risk of mortality is sufficient to justify the chosen margin of inferiority for the trial. The lower the baseline risk of mortality, the smaller the absolute margin of inferiority must be to obtain valid results. For instance, it is not logical that 10% absolute loss of effect (10% greater mortality) could be allowed in a young, nonbacteremic population (age, 30–49 years) in which the point estimate of the effect of antimicrobials, compared with placebo, is 10% with a 95% confidence interval of 5%–14% [39]. This strategy would potentially lose of all the effect of the control drug and not ensure that the test drug was any more effective than placebo. To justify an absolute margin of inferiority of 10%, the control population should have a mortality rate of ~15%. Traditionally, trials in infectious diseases have used an absolute margin of inferiority. However, the data on pneumonia show that a relative margin based on an odds ratio of 1.67 would preserve the benefit of the control drug regardless of the baseline characteristics of patients [39].

Even in contexts in which the treatment effect of antimicrobials is great (older patients with or without bacteremia), what loss of effect is clinically acceptable should be addressed. This issue is not one of sample size or practicality of trials but one of patient safety. An absolute loss of effect on an all-cause mortality end point of >10% is hard to justify clinically. A loss of effect of 15%–20% on mortality would mean that for every 5–7 patients treated with the new drug, compared with the older agent, 1 more person might die. An absolute margin of inferiority for any trial in pneumonia should not exceed 10%. At a recent FDA advisory committee meeting, the majority of advisors voted that margins of inferiority as large as 20% were not clinically acceptable [9]. Again, use of a margin based on odds ratios would be helpful, because there is no one absolute margin that applies to all populations with HAP and/or VAP [39].

Of importance, there are situations in which superiority trials are still needed and ethical in HAP and/or VAP. For instance, in trials of combinations of drugs, investigators should design trials to show the added benefit of the combination compared with monotherapy. Superiority trials are also needed in the clinical context of disease caused by highly drug-resistant pathogens for which there may be no active drugs. This is an area in which new drugs are most urgently needed, but demonstration of noninferiority to drugs for which effects are unknown is not meaningful. It seems incongruous to state the urgent public health need of decreasing effectiveness of older drugs but hold that investigators cannot perform superiority trials to demonstrate that the new agents are truly more effective than older agents.

Selecting patients with disease

Separation of patients who have HAP and VAP from patients with other disease that can mimic the clinical presentation of HAP and VAP is challenging. Enrollment of participants without pneumonia in a superiority trial will result in false-negative results but will result in false conclusions of similarity of drugs in noninferiority trials. Individual findings, such as fever, white blood cell count, sputum purulence, hypoxia, and new infiltrates on chest radiograph, have low positive likelihood ratios (1.2–1.7) for predicting the presence of pneumonia [41]. Studies usually consider likelihood ratios ≥ 10 as useful in distinguishing one disease from another. Even combinations of findings still have relatively low likelihood ratios, with combinations of fever, white blood cell count, infiltrates on chest radiograph, and purulent sputum showing positive likelihood ratios of 1.2–2.5 [41]. In addition, cultures of specimens obtained from various methods (endotracheal aspirate vs bronchoalveolar lavage specimens) may have different predictive values, and pooling of microbiological results may not be justifiable. Additional research is needed to develop rapid point-of-care diagnostics that help select participants for research trials to avoid exposing them to unnecessary harm from experimental agents. These same diagnostics could then be applied in clinical practice to appropriately select patients who might benefit from therapy. At present, inclusion criteria should consist of protocol-defined criteria for signs, symptoms, radiography, and standardized collection of microbiology specimens to provide the best evidence that the participant truly has bacterial pneumonia.

As noted above, investigators should select participants with disease of sufficient baseline severity such that all-cause mortality in the control group is at least 15%–20% [39]. There is a need for better natural history data to select such participants, and the current criteria available for defining severity of disease may miss important variables that affect outcomes. Trials in pneumonia have used the Pneumonia Severity Index (also called the PORT score), CURB-65, and APACHE scores [42,43]. The exact scoring system used is less important than ensuring that the resultant population has mortality in the control group commensurate with that in the historical data that justifies the use of a noninferiority trial (15%–20%).

Baseline comparability

The process of randomization is used in modern trials to give a similar probability that both measured and unmeasured baseline variables are distributed similarly between study groups. This provides some assurance that the observed drug effects are attributable to drug effect and not to lack of baseline comparability between groups. It is well accepted that randomized trials provide the best protection from selection bias and the rationale for robust statistical testing. Observational studies cannot account for unmeasured biases that may influence trial results.

It is equally important, however, to provide follow-up for as many patients as possible and include them in the analysis, to not subvert the protection that randomization provides from selection bias. In 1946, Margaret Rennels, one of the investigators on the earliest trials of penicillin in syphilis, noted the importance of follow-up of participants. She stated, "It is less important to get very large numbers of patients on a particular schedule and then not pay much attention to following them, than it is to get a smaller number who are followed through. There is balance between the two problems of getting large numbers and devoting enough energy to following them up, so that conclusions are not based primarily on pure assumptions" [44, p 128]. If investigators wish to increase the efficiency of clinical trials and decrease sample sizes, decreasing the number of persons excluded from analyses is one way to accomplish this. Clinical trials in HAP and/or VAP exclude large numbers of individuals on the basis of events that occur after randomization [12–21]. For instance, these trials exclude patients on the basis of not receiving a sufficient amount of therapy. There is no scientific reason to exclude patients on the basis of receipt of a given amount of therapy, and even a single dose may have an effect on outcomes and can affect statistical analyses by undermining randomization's protection against selection bias [45]. Recent data show that even a single dose of antimicrobial can affect outcomes, and these data are supported by historical data on treatment of pneumonia [46,47]. In the initial studies of penicillin (a short-acting drug), participants received 2–4 days of therapy, and many recovered during the first 1–2 days [47]. The idea that only long-acting drugs can affect outcomes is not supported by these historical data. Other postrandomization exclusions include confounding intercurrent illness. Patients with intercurrent illness will be randomized in similar proportions to each study group, and those with intercurrent illness are exactly those patients in whom antimicrobials have the greatest treatment effect based on the historical evidence, even though success rates may be lower. Exclusion because of indeterminate outcome is also not an appropriate exclusion criterion. Clear outcome measures not based on clinician judgment would solve this issue. Finally, receipt of concomitant antimicrobials is an issue of study conduct. If a patient receives additional antimicrobials for spread of disease or disease at another site, this should be considered to be a failure of therapy. It is important to know whether an antimicrobial cures one disease but causes another.

The number of persons excluded from HAP and/or VAP trials approaches 50% in some trials, seriously affecting the validity of conclusions and potentially changing a randomized trial into a large observational case series [19]. Authors have noted that exclusions of >5% to 10% of enrolled participants should raise questions about trial validity [45]. A solution for this problem is clear outcome criteria (discussed below), not excluding persons on the basis of postrandomization events, and making every effort to follow-up with all participants enrolled.

Stratified randomization can help ensure that similar numbers of participants in each group possess baseline factors of interest that are associated with outcome and provide confidence in unadjusted analyses. Most multicenter studies stratify by study center. However, confirmatory analyses cannot be performed on each stratum unless investigators specify a hypothesis in advance, ensure an appropriate sample size for each stratum, and make some

adjustment for the increased rate of false-positive results due to multiple comparisons. Merely pre-specifying multiple strata is insufficient if more than exploratory analyses are performed. The numerous analyses by causative pathogen are usually exploratory, because there is no predefined hypothesis and most are underpowered. The idea that 10 patients with disease due to each type of organism are sufficient to make confirmatory conclusions regarding effectiveness in disease due to that organism has no scientific basis.

Minimizing bias

Several kinds of biases can affect clinical trials. Bias can be divided into 3 categories: (1) selection bias, (2) misclassification and/or information bias, and (3) confounding [48]. Randomization with appropriate follow-up (as discussed above) helps control for selection bias. Misclassification and/or information bias occurs when investigators misclassify exposures or outcomes in clinical trials. As noted above, lack of clear diagnostic criteria in noninferiority HAP and/or VAP trials can misclassify patients who have received a diagnosis of a disease that they do not have, resulting in potential false-positive conclusions in noninferiority trials. Misclassification of outcomes can occur in HAP and VAP trials because of lack of specific criteria for assessing outcomes and leaving definitions of outcome to clinician judgment. As noted above, investigators classify substantial proportions of participants in current trials as having indeterminate outcomes, which shows that even clinicians have trouble using current definitions. In addition, different clinicians have different criteria for cure, resulting in inter- and intrarater variability and lack of reliability in outcome assessments.

Misclassification bias is also more likely to occur in trials in which treatment assignment is not blinded. Many current trials in HAP and/or VAP are not blinded, usually because of receipt of concomitant antimicrobials (discussed below) [14,18]. Unclear outcome definitions in open-label trials increase the risk of misclassification bias. Conversely, use of clear outcome definitions, such as all-cause mortality, may lessen concerns regarding misclassification of outcomes, but lack of blinding may still result in operational biases related to how investigators treat patients during the study if they are aware of treatment assignment.

One of the major issues affecting clinical trials in HAP and/or VAP is confounding because of use of prior antimicrobial or combination therapy for disease due to specific pathogens, such as *Pseudomonas aeruginosa*. As noted above, prior antimicrobial therapy, even a single dose of a short-acting agent, can minimize differences between interventions. When the spectrum of activity of the concomitant medication given during the trial overlaps with that of the study medication, it is not possible to separate the effect of the study medication from that of the concomitant antimicrobial. This is of less concern in a superiority trial but raises serious concerns about validity in noninferiority trials. The data on combination therapy are based on in vitro microbiological data, and several studies do not show an added benefit of combination therapy, compared with monotherapy, for HAP and/or VAP [15,49]. In addition, combination therapy is associated with increased risk of adverse events among both research participants and patients, without evidence of clear benefit. A solution for this problem is to study monotherapy in noninferiority trials acknowledging the equipoise that exists regarding combination therapy at the present time in clinical trials. Superiority trials of combination therapy compared with monotherapy would help clarify whether there is benefit for combination therapy. This is an area in which issues related to clinical practice and clinical trials may diverge. Of note, clinicians are poor at selecting which persons are infected with *Pseudomonas* at baseline (eg, the κ coefficient in a clinical trial of doripenem was 0.28) [14]. This means that clinicians are wrong as often as they are correct in choosing patients who might benefit from combination therapy; many participants who are not infected with *Pseudomonas* receive combination therapy, and many participants who have

cultures positive for *Pseudomonas* do not receive combination therapy. Unclear diagnosis is also an issue here, because a culture positive for *Pseudomonas* does not necessarily imply that the organism is causative, especially when isolated from endotracheal aspirate specimens, further complicating assignment to combination therapy.

Well-defined and reliable outcome measures

When choosing end points for clinical trials, investigators should choose end points that directly measure factors that are important to patients (ie, how the patient feels, functions, and survives) [50]. Studies should measure those end points with use of a timing that is relevant to the disease being study and in a well-defined, standardized, and reliable way, while minimizing bias in the assessments. Finally, investigators should provide data on how clinically meaningful an end point is for patients with the disease.

Current outcome measures in clinical trials of HAP and/or VAP are a poorly defined composite of signs and symptoms. There are several issues with this end point. First, there is no basis for using this end point in noninferiority trials. The only evidence for a treatment effect on which to base noninferiority trials is all-cause mortality [39]. Second, although an effect on symptoms is a direct measure of patient benefit, signs of disease are biomarkers used as surrogate end points [5,50–53]. There is no need to use a surrogate end point for an acute short-term disease in which clinical end points can be measured directly. FDA regulations state that studies can use surrogate end points in a setting where an intervention provides “meaningful therapeutic benefit to patients over existing therapies” [52,53]. This statement describes superiority trials; thus, the use of surrogate end points in noninferiority trials is questionable. Third, the lack of specific criteria for clinicians to use means that the outcome measures are neither well defined nor reliable.

The outcomes measures for noninferiority hypotheses should be all-cause mortality measured 14–21 days after randomization and initiation of study interventions. The concept of pneumonia-related mortality has no scientific foundation, because pneumonia can worsen the function of other body systems and cause death by mechanisms other than direct respiratory failure. William Osler stated in his 1892 textbook that “death rarely occurs from direct interference with the function of respiration, though it may happen in cases of extensive double-pneumonia. In a majority of cases the fatal result is brought about by gradual heart failure” [54, p 571]. Thus, the causes of death from various types of organ failure in the context of pneumonia are directly related to each other. Because the older persons and persons with comorbidities are exactly the persons in whom antimicrobials showed the largest treatment effect in historical studies, the data show that antimicrobials, in fact, decrease the rate of death from heart failure [39]. In addition, several studies showed that investigators could not accurately judge the cause of death, compared with autopsy findings, which shows that misclassification bias is inherent in clinicians judging a specific cause of death [55,56]. Also, attributable mortality excludes deaths that may be directly due to harms of the intervention, which can negate any benefits of treatment of pneumonia. The claim that patients do not die of pneumonia in the current era is not supported by evidence. Pneumonia is still the sixth leading cause of death, and data from current studies show a mortality rate of ~15% [57]. Anecdotal descriptions of individual cases from historical studies with claims that such patients would not die today is not evidence that overall mortality rates are different today than they were in the past. If mortality rates are lower today, there is no basis for noninferiority trials in pneumonia, because the historical data do not apply to current trials or clinical settings.

In superiority hypotheses, investigators could test other clinically relevant end points, either singly or as part of a composite end point [58]. Other end points could include nonfatal clinical events (eg, extension of disease, such as empyema, or protocol-defined disease at

another body site, such as meningitis; endocarditis; acute respiratory distress syndrome; and respiratory failure) or direct measures of patient symptoms (eg, cough, chest pain, dyspnea, warmth, and chills). Investigators could use appropriately developed and validated patient-reported outcome measures to evaluate symptoms in a standardized way in persons who are not receiving mechanical ventilation. Patient-reported outcomes measure the same symptoms as measured by clinicians, albeit in a standardized and reliable way, thereby decreasing misclassification bias and random error and increasing the efficiency of trials, allowing for a smaller sample size because of decreased variability [59].

Appropriate analysis

In analyzing trial results, investigators should address several points, including (1) the choice of analysis population, (2) how to analyze missing data, (3) issues related to subgroup analyses, and (4) issues related to multiple comparisons. The modified intent-to-treat analysis of all participants randomized who have the disease being studied and who have received at least 1 dose of study medication preserves the protections of randomization from bias as long as the number of exclusions is relatively small (5%–10% of those randomized) [45]. Some authors express concern that modified intent-to-treat analysis may make drugs appear more similar in noninferiority trials, however recent studies have shown that this is not always the case [60]. Conversely, per protocol analysis (also called the clinically evaluable or as treated group) is a subgroup analysis that may subvert the protections of randomization from bias. This highlights the problem that there is no optimal analysis population in noninferiority trials and that confirmation of results across various population analyses should be studied. For this reason, the FDA has asked drug sponsors to analyze results in both the modified intent-to-treat and per protocol populations.

The aforementioned issues related to inappropriate postrandomization exclusions from HAP and/or VAP trials relate to how to analyze missing data. All methods to analyze missing data make usually unverifiable assumptions about the nature of the missing data [61,62]. Therefore, it is most rational to look for robustness across a number of sensitivity analyses and conduct the trial in such a way as to minimize loss to follow-up. Analysis of all persons who are excluded from the per protocol population because of failure of treatment in the modified intent-to-treat analysis is not the only sensitivity analysis that can be performed, nor is it the most conservative analysis in noninferiority trials, because this may indeed make 2 interventions appear more similar. Participants who stop study medication because of adverse events do not necessarily withdraw from the study. Investigators should continue to follow-up with such persons, and because of information that short courses of therapy may be effective in treating pneumonia, the clinical outcomes in these persons should be assessed at the time of discontinuation of study medication. If persons meet the definition of clinical success in the trial (survival in a noninferiority trial), the outcome of their condition should be considered to be a success. Because the outcome in noninferiority trials in pneumonia should be all-cause mortality, survival is more important to patients than adverse events, such as nausea or headache. Treatment should not be considered to be a failure in persons who switch to another antimicrobial, because historical data showed high mortality among these patients in the pre-antibiotic era (ie, treatment is successful if patients live long enough to receive another therapy). The idea that there were no therapies to which clinicians could switch in early studies is incorrect, because both sulfa drugs and serum therapy were available at the time of the introduction of penicillin [39,63]. Investigators in 1943 specifically noted that “patients with pneumococcal pneumonia with negative blood cultures who showed no improvement over 12–18 hours” received serum therapy [63, p 25]. Current investigators should use protocol-defined parameters for defining switches of antimicrobial therapy. The data on short courses of therapy for pneumonia may obviate use of oral switch

therapy, because prolonged courses of antimicrobial therapy may result in more adverse events and increased antimicrobial resistance.

Subgroup analyses and multiple comparisons raise similar analytical concerns [64,65]. Evaluation of multiple end points, the same end point at multiple times, or numerous subgroup analyses increases the chance of false-positive results in clinical trials. To evaluate confirmatory results from subgroup analyses, investigators should specify hypotheses for those subgroups in advance, including an appropriate definition for entry into the subgroup with an appropriate sample size, evaluation of the complementary groups to the designated subgroups to ensure that there is an absence of harm in the complementary groups, and awareness of the effect of multiple comparisons. Evaluation of posthoc subgroups from noninferiority trials to claim superiority is an inappropriate use of subgroup analyses. Such exploratory analyses need confirmation in future trials [66,67].

EVALUATING HARMS IN HAP AND VAP TRIALS

As noted above, it is necessary to obtain data on the effectiveness of medical interventions to ensure that there is some benefit to balance against the harms inherent in all interventions. However, when measuring harm in clinical trials, investigators most often are not testing specific hypotheses but are actually searching for hypotheses to test. The analysis of harms is composed of a complete analysis of preclinical testing, testing in healthy volunteers, and data from early- and late-phase clinical trials. Because hypotheses are most often not tested, lack of statistical significance does not equate to absence of harm or absence of differences in harm between medical interventions. Most clinical trials have an insufficient sample size or too short of a follow-up period to critically evaluate adverse events. A good rule of thumb is the rule of threes, which states that if one observes no adverse events of a given type, one can rule out, with 95% confidence, a rate of 3 divided by the sample size [68]. For instance, a study in which there are no cases of hepatic failure in a database of 300 participants rules out a rate of 3 divided by 300, or 1%. By realizing that the rate of hepatic failure in the general population is ~1 in 1 million, this does not rule out a substantial risk above background. The limitations of preapproval data highlight the need for postmarketing follow-up and further assessment of medical interventions after regulatory approval. It is important to realize that passive reporting of adverse events, however, results in reporting of only 1%–10% of actual adverse events; thus, the estimate of rates of adverse events may be inaccurate by several orders of magnitude and makes comparisons of rates between drugs problematic.

An appropriate risk-benefit analysis takes into account the conditions of use of the intervention and the specific benefits and risks, in association with estimates of their magnitude. In clinical trials in which the outcome measure is all-cause mortality and the benefit to patients is saving lives, one can accept a greater risk of adverse events. When the outcome measure is symptom relief or a surrogate end point that does not measure direct benefit to patients, there is less margin for error in accepting adverse events, especially those that might be serious and life-threatening. Use of all-cause mortality as an end point in clinical trials of HAP and/or VAP allows a better justification of risk to benefit, given demonstration of preservation of a meaningful benefit to patients in appropriately designed noninferiority trials. A database of 300–500 persons who received the dose and duration of study medication planned for use in patients with HAP and/or VAP is usually adequate for analysis as long as no safety signals emerge from that database. The absence of serious adverse events in such a database would allow one to rule out, with 95% confidence, a rate of 1% for these events.

CONCLUSIONS

The current design of HAP and VAP trials contains several potential biases, as noted in the preceding sections. Table 2 contains a list of recommendations for improving the clinical trial design of HAP and/or VAP trials to decrease the effects of bias and improve their validity and reliability. Improving the design of HAP and/or VAP trials entails not only addressing issues related to the design and analysis of trials but also addressing issues related to the conduct of trials. Ensuring that participants receive informed consent, randomization and assignment to study medication in a timely fashion, and minimizing loss to follow-up, are issues of study conduct. Trials in other therapeutic areas, such as thrombolytic therapy, have enrolled persons in a narrow time window, and this is achievable in infectious diseases trials. Addressing prompt enrollment also entails forming relationships with other care providers in emergency departments and intensive care units. Although it is important to make trials as practical as possible, issues of practicality should not subvert the validity of the trial. Indeed, subverting the validity of a trial calls into question the ethics of the trial, because the purpose of obtaining the data is to add to generalizable knowledge. If the data are flawed, inaccurate, or even frankly incorrect, researchers are not advancing public health or contributing to generalizable knowledge. All measurements in science are associated with some amount of error, so the goal is not perfection. However, the inability to achieve perfect results is not an excuse for mediocrity. Researchers should still try to obtain as accurate and valid results as possible. Science is an ever-changing field, and researchers should incorporate new knowledge into the way that they perform clinical research. Unfortunately, noninferiority trials contain an inherent and unbreakable link with the past that limits the ability to change facets of a trial using this design. Greater use of superiority hypotheses built into noninferiority trials may help us answer new questions while addressing the potential biases that threaten noninferiority trials [69]. The results of our efforts will be more reliable and accurate information on which to base decisions now and in the future.

Acknowledgments

Financial support. National Cancer Institute, National Institutes of Health (contract HHSN261200800001E) and the National Institute of Allergy and Infectious Disease.

References

1. Hopkins WA. Patulin. Biological properties: extended trial in the common cold. *Lancet* 1943;i:631–634.
2. Medical Research Council. Clinical trial in the common cold. *Lancet* 1944;i:373–375.
3. World Health Organization. Food-borne hazards. 2009 [Accessed 15 May 2010]. <http://www.who.int/foodsafety/publications/capacity/en/2.pdf>
4. Anastasi, A.; Urbina, S. Validity: basic concepts. In: Anastasi, A.; Urbina, S., editors. *Psychological testing*. Upper Saddle River, NJ: Prentice Hall; 1997. p. 113-139.
5. International Conference on Harmonisation. *Statistical Principles for Clinical Trials (ICH E-9)*. 1998 [Accessed 15 May 2010.]. <http://www.ich.org/LOB/media/MEDIA485.pdf>
6. International Conference on Harmonisation. *Choice of Control Group and Related Issues in Clinical Trials (ICH E-10)*. 2000 [Accessed 15 May 2010]. <http://www.ich.org/LOB/media/MEDIA486.pdf>
7. Fleming TR. Current issues in non-inferiority trials. *Stat Med* 2008;27(3):317–332. [PubMed: 17340597]
8. Pocock SJ. The pros and cons of noninferiority trials. *Fundam Clin Pharmacol* 2003;17(4):483–490. [PubMed: 12914552]

9. US Food and Drug Administration. Transcripts of the Anti-Infective Drugs Advisory Committee: doripenem for hospital-acquired pneumonia. 2008 [Accessed 15 May 2010]. <http://www.fda.gov/ohrms/dockets/ac/08/minutes/2008-4364m1-Final.pdf>
10. Basilea Pharmaceuticals. Ceftobiprole press release. 2009 [Accessed 15 May 2010]. <http://hugin.info/134390/R/1158619/224195.pdf>
11. Wyeth Pharmaceuticals. Tigecycline. 2007 [Accessed 15 May 2010]. http://www.wyeth.com/news/archive?nav=display&navTo=/wyeth_html/home/news/pressreleases/2007-1184074360162.html
12. Alvarez-Lerma F, Insausti-Ordenana J, Jorda-Marcos R, et al. Efficacy and tolerability of piperacillin/tazobactam versus ceftazidime in association with amikacin for treating nosocomial pneumonia in intensive care patients: a prospective randomized multicenter trial. *Intensive Care Med* 2001;27(3):493–502. [PubMed: 11355117]
13. Brun-Buisson C, Sollet JP, Schweich H, Briere S, Petit C. Treatment of ventilator-associated pneumonia with piperacillin-tazobactam/amikacin versus ceftazidime/amikacin: a multicenter, randomized controlled trial. VAP Study Group. *Clin Infect Dis* 1998;26(2):346–354. [PubMed: 9502454]
14. Chastre J, Wunderink R, Prokocimer P, Lee M, Kaniga K, Friedland I. Efficacy and safety of intravenous infusion of doripenem versus imipenem in ventilator-associated pneumonia: a multicenter, randomized study. *Crit Care Med* 2008;36(4):1089–1096. [PubMed: 18379232]
15. Cometta A, Baumgartner JD, Lew D, et al. Prospective randomized comparison of imipenem monotherapy with imipenem plus netilmicin for treatment of severe infections in nonneutropenic patients. *Antimicrob Agents Chemother* 1994;38(6):1309–1313. [PubMed: 8092830]
16. Fagon J, Patrick H, Haas DW, et al. Treatment of gram-positive nosocomial pneumonia. Prospective randomized comparison of quinupristin/dalfopristin versus vancomycin. Nosocomial Pneumonia Group. *Am J Respir Crit Care Med* 2000;161(3 Pt 1):753–762. [PubMed: 10712318]
17. Fink MP, Snyderman DR, Niederman MS, et al. Treatment of severe pneumonia in hospitalized patients: results of a multicenter, randomized, double-blind trial comparing intravenous ciprofloxacin with imipenem-cilastatin. The Severe Pneumonia Study Group. *Antimicrob Agents Chemother* 1994;38(3):547–557. [PubMed: 8203853]
18. Rea-Neto A, Niederman M, Lobo SM, et al. Efficacy and safety of doripenem versus piperacillin/tazobactam in nosocomial pneumonia: a randomized, open-label, multicenter study. *Curr Med Res Opin* 2008;24(7):2113–2126. [PubMed: 18549664]
19. Rubinstein E, Cammarata S, Oliphant T, Wunderink R. Linezolid (PNU-100766) versus vancomycin in the treatment of hospitalized patients with nosocomial pneumonia: a randomized, double-blind, multicenter study. *Clin Infect Dis* 2001;32(3):402–412. [PubMed: 11170948]
20. West M, Boulanger BR, Fogarty C, et al. Levofloxacin compared with imipenem/cilastatin followed by ciprofloxacin in adult patients with nosocomial pneumonia: a multicenter, prospective, randomized, open-label study. *Clin Ther* 2003;25(2):485–506. [PubMed: 12749509]
21. Wunderink RG, Cammarata SK, Oliphant TH, Kollef MH. Continuation of a randomized, double-blind, multicenter study of linezolid versus vancomycin in the treatment of patients with nosocomial pneumonia. *Clin Ther* 2003;25(3):980–992. [PubMed: 12852712]
22. US Government Printing Office. Federal Food Drug and Cosmetic Act, Section 505(d). 2007 [Accessed 15 May 2010]. <http://www.fda.gov/opacom/laws/fdcact/fdcact5a.htm>
23. *Pharmaceutical Manufacturers Association v Richardson* 318 F Supp 301. 1970.
24. US Government Printing Office. US Code of Federal Regulations, Title 21, Part 314.126. [Accessed 15 May 2010]. http://a257.g.akamaitech.net/7/257/2422/10apr2006-1500/edocket.access.gpo.gov/cfr_2006/aprqr/21cfr314.126.htm
25. US Government Printing Office. Federal Food Drug and Cosmetic Act, Section 505(b)(5)(C)(ii). [Accessed 15 May 2010]. <http://www.fda.gov/opacom/laws/fdcact/fdcact5a.htm>
26. *Upjohn v Finch*, 422 F 2d 944. 1970.
27. Dimasi JA. Risks in new drug development: approval success rates for investigational drugs. *Clin Pharmacol Ther* 2001;69(5):297–307. [PubMed: 11371997]
28. *United States v Rutherford*, 442 US 544, 78, 605. 1979.

29. Tricoci P, Allen JM, Kramer JM, Califf RM, Smith SC Jr. Scientific evidence underlying the ACC/AHA clinical practice guidelines. *JAMA* 2009;301(8):831–841. [PubMed: 19244190]
30. Lee, DH.; Vielmeyer, O.; Solari, P.; Chowdhury, M. IDSA guidelines: what evidence are they based on?. Program and abstracts of the 47th Meeting of the Infectious Diseases Society of America (Philadelphia); Infectious Diseases Society of American; Arlington, VA. 2009. Abstract 1324
31. Khan, AR.; Khan, S.; Baddour, LM.; Tleyjeh, IM. The quality and strength of evidence of the Infectious Diseases Society of America clinical practice guidelines. (abstract LB-31). Program and abstracts of the 47th Meeting of the Infectious Diseases Society of America (Philadelphia); Infectious Diseases Society of American; Arlington, VA. 2009. p. 79
32. Miller FG, Rosenstein DL. The therapeutic orientation to clinical trials. *N Engl J Med* 2003;348(14):1383–1386. [PubMed: 12672867]
33. US Department of Health and Human Services. The Belmont Report: ethical principles and guideline for the protection of human subjects of research. [Accessed 15 May 2010]. <http://ohsr.od.nih.gov/guidelines/belmont.html>
34. Rossouw JE, Anderson GL, Prentice RL, et al. Risks and benefits of estrogen plus progestin in healthy postmenopausal women: principal results From the Women's Health Initiative randomized controlled trial. *JAMA* 2002;288(3):321–333. [PubMed: 12117397]
35. Chen GF, Johnson MH. Patients' attitudes to the use of placebos: results from a New Zealand survey. *N Z Med J* 2009;122(1296):35–46. [PubMed: 19652679]
36. Ioannidis JP. Why most published research findings are false: author's reply to Goodman and Greenland. *PLoS Med* 2007;4(6):e215. [PubMed: 17593900]
37. Freedman B. Scientific value and validity as ethical requirements for research: a proposed explication. *IRB* 1987;9(6):7–10. [PubMed: 11650779]
38. Powers JH. Noninferiority and equivalence trials: deciphering 'similarity' of medical interventions. *Stat Med* 2008;27(3):343–352. [PubMed: 18186148]
39. Fleming TR, Powers JH. Issues in noninferiority trials: the evidence in community-acquired pneumonia. *Clin Infect Dis* 2008;47 Suppl 3:S108–S120. [PubMed: 18986275]
40. Luna CM, Aruj P, Niederman MS, et al. Appropriateness and delay to initiate therapy in ventilator-associated pneumonia. *Eur Respir J* 2006;27(1):158–164. [PubMed: 16387949]
41. Klompas M. Does this patient have ventilator-associated pneumonia? *JAMA* 2007;297(14):1583–1593. [PubMed: 17426278]
42. Fine MJ, Auble TE, Yealy DM, et al. A prediction rule to identify low-risk patients with community-acquired pneumonia. *N Engl J Med* 1997;336(4):243–250. [PubMed: 8995086]
43. Lim WS, van der Eerden MM, Laing R, et al. Defining community acquired pneumonia severity on presentation to hospital: an international derivation and validation study. *Thorax* 2003;58(5):377–382. [PubMed: 12728155]
44. Marks, HM. The progress of experiment. 1 ed.. Cambridge: Cambridge University Press; 2000.
45. Gillings D, Koch G. The application of the principle of intention-to-treat to the analysis of clinical trials. *Drug Information J* 1991;25:411–424.
46. Pertel PE, Bernardo P, Fogarty C, et al. Effects of prior effective therapy on the efficacy of daptomycin and ceftriaxone for the treatment of community-acquired pneumonia. *Clin Infect Dis* 2008;46(8):1142–1151. [PubMed: 18444848]
47. Powers JH. Reassessing the design, conduct, and analysis of clinical trials of therapy for community-acquired pneumonia. *Clin Infect Dis* 2008;46(8):1152–1156. [PubMed: 18444849]
48. Miettinen OS, Cook EF. Confounding: essence and detection. *Am J Epidemiol* 1981;114(4):593–603. [PubMed: 7304589]
49. LaForce FM. Systemic antimicrobial therapy of nosocomial pneumonia: monotherapy versus combination therapy. *Eur J Clin Microbiol Infect Dis* 1989;8(1):61–68. [PubMed: 2495953]
50. Temple, RJ. A regulatory authority's opinion about surrogate endpoints. In: Nimmo, WS.; Tucker, GT., editors. *Clinical Measurement in Drug Evaluation*. New York: John Wiley and Sons; 1995. p. 3-22.

51. Biomarkers Definitions Working Group. Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. *Clin Pharmacol Ther* 2001;69(3):89–95. [PubMed: 11240971]
52. US Government Printing Office. US Code of Federal Regulations, Title 21, Part 314.500 Subpart H. [Accessed 15 May 2010].
http://a257.g.akamaitech.net/7/257/2422/10apr20061500/edocket.access.gpo.gov/cfr_2006/aprqr/21cfr314.126.htm
53. Federal Register. New drug, antibiotic, and biological drug regulations; accelerated approval. Docket No. 91N-0278 1992;57(73):13234–13242.
54. Osler, W. Principles and Practice of Medicine. 1 ed.. Baltimore: Johns Hopkins University Press; 1892.
55. Kirch W, Schaffii C. Misdiagnosis at a university hospital in 4 medical eras. *Medicine (Baltimore)* 1996;75(1):29–40. [PubMed: 8569468]
56. Sharma S, Nadrous HF, Peters SG, et al. Pulmonary complications in adult blood and marrow transplant recipients: autopsy findings. *Chest* 2005;128(3):1385–1392. [PubMed: 16162733]
57. Ochoa-Gondar O, Vila-Corcoles A, de DC, et al. The burden of community-acquired pneumonia in the elderly: the Spanish EVAN-65 study. *BMC Public Health* 2008;8:222. [PubMed: 18582392]
58. Lubsen J, Kirwan BA. Combined endpoints: can we use them? *Stat Med* 2002;21(19):2959–2970. [PubMed: 12325112]
59. Patrick DL, Burke LB, Powers JH, et al. Patient-reported outcomes to support medical product labeling claims: FDA perspective. *Value Health* 2007;10 Suppl 2:S125–S137. [PubMed: 17995471]
60. Sheng D, Kim MY. The effects of non-compliance on intent-to-treat analysis of equivalence trials. *Stat Med* 2006;25(7):1183–1199. [PubMed: 16220491]
61. Altman DG, Bland JM. Missing data. *BMJ* 2007;334(7590):424. [PubMed: 17322261]
62. Altman DG. Missing outcomes in randomized trials. *Open Med* 2009;3:e51–e53. [PubMed: 19946393]
63. Collen MF, Anderson E. Perspective Kaiser Permanente Medicine 50 years ago. *Permanente J* 1998;2(1):22–28. p25.
64. Moye LA, Deswal A. Trials within trials: confirmatory subgroup analyses in controlled clinical experiments. *Control Clin Trials* 2001;22(6):605–619. [PubMed: 11738119]
65. Pocock SJ. Clinical trials with multiple outcomes: a statistical perspective on their design, analysis, and interpretation. *Control Clin Trials* 1997;18(6):530–545. [PubMed: 9408716]
66. Powers JH, Lin D, Ross D. FDA evaluation of antimicrobials: subgroup analysis. *Chest* 2005;127(6):2298–2299. [PubMed: 15947370]
67. Wunderink RG, Rello J, Cammarata SK, Croos-Dabrera RV, Kollef MH. Linezolid vs vancomycin: analysis of two double-blind studies of patients with methicillin-resistant *Staphylococcus aureus* nosocomial pneumonia. *Chest* 2003;124(5):1789–1797. [PubMed: 14605050]
68. Powers, JH. Interpreting the results of clinical trials on antimicrobial agents. In: Mandell, GL.; Bennett, JE.; Dolin, R., editors. Principles and practice of infectious diseases. Philadelphia: Elsevier Churchill Livingstone; 2005. p. 619-628.
69. Powers JH, Fleming TR. Design, conduct, and analysis of clinical trials in disease due to methicillin-resistant *Staphylococcus aureus*. *Clin Pharmacol Ther* 2009;86(3):244–247. [PubMed: 19707216]

Table 1

Criteria for Effectiveness Based on Substantial Evidence from Adequate and Well-Controlled Trials

Criteria
Clear objective of the study
Study allows a quantitative comparison with a control group
Appropriate selection of participants with the disease in treatment trials or at risk of disease in prevention trials
Baseline comparability between study groups
Minimizing bias
Well-defined and reliable outcome measures
Appropriate analysis of results

NOTE. Evidence of effectiveness must be balanced against potential harms to obtain overall evaluation of risk-benefit.

Table 2

Recommendations for Improving the Validity and Reliability of Clinical Trials in Hospital-Acquired Pneumonia (HAP) and Ventilator-Associated Pneumonia (VAP)

Recommendation
Explanatory trials with separate hypotheses for HAP and VAP preferably in separate clinical trials with hypotheses (dosing, duration of treatment) supported by appropriate preclinical and early clinical data
Noninferiority hypotheses designed with all-cause mortality end point for population in which control group has mortality of 15%–20%; absolute margin of inferiority no greater than 10% in this population and preferably use of odds ratio margin of 1.67; superiority hypotheses in some settings, such as drug combinations, disease due to drug-resistant pathogens, or use of novel end points, such as patient-reported outcome measures
Selection of participants based on combination of signs, symptoms, and laboratory and radiological parameters, with appropriate microbiological confirmation of disease in appropriate specimens
Randomization with stratification by study center and possibly by other important factors that affect outcome independent of treatment (eg, bacteremia and age >50), with minimization of exclusions from analysis to preserve protection from bias afforded by randomization
Minimize bias by eliminating prior and concomitant antimicrobial therapies, decreasing loss to follow-up and postrandomization exclusions, and use of clearly defined outcome measures assessed in a double-blinded manner when possible.
End point of all-cause mortality measured at 14–21 days after initiation of study drugs for noninferiority hypotheses; testing of other end points, such as nonfatal clinical events and resolution of symptoms, evaluated by appropriately designed and validated patient-reported outcome measurements in superiority hypotheses
Appropriate analysis based on evaluating robustness of conclusions in modified intent-to-treat and per protocol populations, with sensitivity analyses for missing data; limited and appropriate use of subgroup analyses with accounting for multiple comparisons at time of design of trial