Increasing the Efficiency of Clinical Trials of Antimicrobials: The Scientific Basis of Substantial Evidence of Effectiveness of Drugs

John H. Powers
Science Applications International Corporation (SAIC) in support of the Collaborative Clinical Research Branch, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, and University of Maryland School of Medicine, Baltimore, Maryland; and George Washington University School of Medicine and Health Sciences, Washington, DC

In the United States, drug sponsors must obtain approval from the US Food and Drug Administration before licensure and widespread clinical use of drugs. In this article, I discuss the definition and history of the regulatory requirement for “substantial evidence” of effectiveness from “adequate and well-controlled” clinical trials of drugs. These requirements apply to antimicrobials as they do to other therapeutic drug classes, and they may be even more important in their application to antimicrobials, given issues of antimicrobial resistance. I will discuss the evidence requirements, using examples from clinical trials in diseases such as acute otitis media, acute bacterial sinusitis, and acute exacerbations of chronic bronchitis. Examination of the principles of substantial evidence also points to opportunities to improve the efficiency of confirmatory clinical trials of antimicrobials to obtain more clinically relevant and useful information without increasing the uncertainty regarding the safety and efficacy of these drugs.

Across the globe, various regulatory agencies review the data on the effectiveness and safety of medical interventions before their approval for general clinical use. In the United States, the Food and Drug Administration (FDA) performs the function of reviewing the data on effectiveness and potential harms submitted by drug sponsors in new drug applications for new drugs or new uses for older drugs. The FDA also performs continuing review even after approval of drugs for clinical use, to evaluate changes in both the safety profile and effectiveness of drugs.

Unfortunately, some view the regulatory process as a “hurdle” or “obstacle” to overcome rather than a scientifically based process. Some seem to imply that “registrational” trials have little clinical importance or relevance. Scientific worthiness is a prerequisite for ethical clinical research, so registrational trials can and should answer clinically relevant questions [1].

History shows that regulation of drug products is needed and beneficial to the public by setting an appropriate standard to both advance and promote public health. Scientifically appropriate standards also are useful for regulated drug sponsors, because they allow planning of drug development programs and ensure that their competitors must follow the same standards. Regulatory standards, however, must change as science and our understanding of it change. Some have implied that, once they reach an agreement on a trial design with the FDA, the trial design cannot be changed. However, the federal Food, Drug, and Cosmetic (FD&C) Act, which authorizes the FDA, states that a trial’s design and its interpretation can be changed when a “substantial scientific issue essential to determining the safety or effectiveness of the drug has been identified after the testing has begun” [2].

The history of clinical trials closely follows the history of drug regulation. Of note, investigators first used many of the modern methods that allow evaluation of the causal relationship between medical interventions...
and outcomes, such as randomization and blinding, in trials evaluating infectious diseases at approximately the same time as the passage of the FD&C Act, in 1938. The British Medical Research Council’s trial of patulin for the common cold, published in 1944, was one of the first examples of the use of a placebo in a multicenter trial [3]. This same group, led by Austin Bradford Hill and others, first used random numbers as a method of randomization in testing streptomycin in pulmonary tuberculosis [4]. Regulatory agencies have adopted these principles in the past and helped to advance their use, and they can still use their influence today to advance the scientific process of evaluation of new medical interventions. The mandate of advancing public health points to the need for innovative but scientifically valid approaches to clinical trials and a need for leadership from regulators, not merely passive adherence to “precedent.”

The adage of “might makes right” should be reversed in the case of regulatory evaluation of medical interventions. Instead, “right” should make “might.” That is, instead of merely following rules for rules’ sake, the authority to review new medical interventions should derive from appropriate scientific principles. An examination of the history of US drug regulation shows that this is, indeed, the case, and legislation authorizing the existence and function of the FDA has been and still is based on appropriate science.

One of the reasons for the misunderstandings regarding regulatory issues may be a lack of understanding regarding their scientific basis. Therefore, in the present article, I will review the scientific and legal foundations for drug evaluation in the United States, which are based on “substantial evidence” of drug effectiveness from “adequate and well-controlled trials” balanced against an adequate assessment of the potential harms of a drug. I will then examine how clinical trials may fulfill the criteria of substantial evidence of effectiveness and evaluate potential harms utilizing novel trial designs in infectious diseases. Such trials may provide more clinically relevant information to practicing clinicians, streamlining the drug development process without increasing uncertainty regarding safety and effectiveness.

THE SCIENTIFIC AND LEGAL BASIS OF DRUG REGULATION: DEFINING “SUBSTANTIAL EVIDENCE”

The scientific method, first described in the 13th century and based on the principles of inductive reasoning explored by Aristotle, is based on 4 steps: (1) observation of events in nature, (2) formation of a hypothesis based on those observations, (3) testing the hypothesis through experimentation, and (4) confirmation of the results of the experiment, to evaluate the probability that the results could have occurred by chance alone [5]. The process of evaluating medical interventions follows these same steps. The initial impetus for evaluating new therapies comes from observations in clinical practice. Clinicians may note the medical need for treatment of a particular disease and may even note that some patients appear to respond favorably when administered an experimental agent. However, although observations may provide associations with outcomes, associations are not direct evidence of a causal relationship between outcomes and administration of a drug [6]. A medical need is a reason to perform a study but does not de facto imply that a drug is effective and safe, nor does it imply that less evidence is needed. Therefore, investigators form a hypothesis, aided by early evaluations of the drug in vitro, in animal models, and in healthy volunteers, as well as by preliminary evaluations of drug safety and effectiveness in early clinical trials. These evaluations of mechanism of action and early safety signals form the ethical and scientific basis for exposing larger numbers of subjects to the potential harms of experimental agents. Confirmatory trials form the experiments on which one bases the evidence of the effectiveness and safety of a drug. When clinical trials show a net positive effect of a medical intervention on clinical outcomes important to patients, it is still necessary to confirm those results with evidence from other trials, to ascertain whether they might have occurred by chance.

To discuss more efficient and more informative trial designs, one must first understand the scientific and regulatory bases for evaluating the evidence of drug effectiveness and safety. These form the standards to which any novel designs still must adhere. In 1906, the US Congress passed the Pure Food and Drug Act. At that time, there was no requirement for evaluation of safety and effectiveness before the use of a drug in clinical practice, and the law was mainly related to adulterated foods; the government merely responded to crises involving drugs. It is interesting to note that further changes in legislation were associated with issues related to drugs used to treat infectious diseases. Spurred on by >100 deaths among children administered elixir of sulfanilamide, Congress passed the FD&C Act in 1938. At that time, the act required an evaluation of only the safety of a drug before licensure. Phocomelia associated with the use of thalidomide (a drug now indicated for the treatment of leprosy) resulted in the passage of the Kefauver-Harris amendments to the FD&C Act in 1962. These amendments required an evaluation of the effectiveness, as well as the safety, of a drug before approval [7].

The Senate hearings related to the passage of the 1962 amendments and subsequent court cases made several important points. First, they made it clear that evidence of effectiveness is necessary to balance any potential harms of drugs. Without evidence of effectiveness, any adverse events in patients, no matter how rare or minor, are not justifiable [8]. Second, testimony at the hearings and subsequent court cases made it clear that the impressions of clinicians, use of a drug in clinical
practice, or robust sales did not form an adequate basis for drug approval [8–10]. This reflects the scientific principle that observations alone, without experimentation and confirmation, do not constitute evidence of effectiveness. The amendments also pointed out that poorly controlled experiments were not adequate evidence, highlighting that the experimentation must be of sufficient rigor to distinguish the effect of the drug from random error, confounding factors, and bias [10]. Third, the amendments provided a clear definition of “substantial evidence.” During Senate hearings, President Kennedy conveyed to Congress that an undefined standard of “substantial evidence” was not adequate to ensure that drugs are effective for the claims made for them [8]. This means that the definition of substantial evidence for drugs is not the same as that used for legal cases. The definition of substantial evidence for legal cases is “more than a mere scintilla of evidence such that a reasonable mind might accept as adequate to support a conclusion” [11]. Obviously, such a definition would leave a good deal of vagueness as to what a “reasonable mind” might conclude regarding the effects of a drug. Rather, the amendments to the FD&C Act provided that the only evidence acceptable to support the effectiveness of drugs would be evidence from “adequate and well-controlled trials” [8–10, 12].

Congress tasked the FDA with defining the terms of “adequate and well-controlled” trials in the Code of Federal Regulations. In 1970, the FDA put forth regulations that defined the criteria for “adequate and well-controlled trials,” that, with few changes, are still in place today [13]. The Pharmaceutical Manufacturers of America sued the FDA in the same year, holding that the criteria were too onerous. The Pharmaceutical Manufacturers of America held that Congress intended that any drug “believed by a substantial number of experts” to be effective could be approved even if the view of the majority of experts was that the drug was ineffective. The courts rejected this opinion and reiterated that experts’ beliefs are not the basis for evidence of drug effectiveness [8]. The courts also found that the criteria described in the regulations were in no sense unduly rigid or narrow, allowed substantial flexibility for investigators, and were entirely reasonable in describing the scientific content of an investigation. The courts also pointed out that these criteria were minimal requirements for the evaluation of drug effectiveness. Compliance with these criteria was necessary for any study to have a chance of scientific validity. However, trials using these criteria may still fail to provide evidence of effectiveness if, for instance, investigators fail to appropriately select patients with the disease or do not measure outcomes appropriately [8]. In other words, merely performing a trial is not sufficient in itself, but the trial must meet these criteria for one to consider it adequate and well controlled.

Importantly, the courts also held that the FDA was not to apply these criteria for adequate and well-controlled trials only prospectively [8, 10]. With the advent of new information or a new understanding of clinical trials, the FDA should reassess the evidence of previously approved drugs. Indeed, in their suit, the Pharmaceutical Manufacturers of America themselves stated, “Undoubtedly, even more advanced principles will be developed as testing know-how improves. They, too, should be applied to trials commenced after their development, when appropriate” [8]. The idea of reassessment of previous theories and prior evidence is a fundamental principle in science. Science is an ever-changing field, and, to advance and promote public health, regulatory agencies must keep current with the most up-to-date understanding of clinical trial design and analysis. If we know now that previous approvals were based on evidence that today is known not to meet the standard of adequate and well-controlled trials, the FDA can and should reassess those drugs to protect the public’s health. This may require a reevaluation of the evidence regarding safety, effectiveness, or both. This is even more important for antimicrobials, because an antimicrobial that lacks substantial evidence of effectiveness is still capable of spreading resistance not only to that drug but to other antimicrobials as well. In this sense, antimicrobial resistance is clearly a safety issue, because ineffective drugs may harm not only the person who takes the drug but also those who do not take it. The use of drugs for less serious diseases for which effectiveness is unclear may obviate the use of that same drug and other drugs for more serious diseases for which there may be substantial evidence of effectiveness. In addition, there seems to be a widely held misunderstanding that, once FDA and drug sponsors reach an “agreement” on a study design, regulatory agencies should accept the results of such trials as substantial evidence of effectiveness even though it may be clear at the time of study’s completion that changes in the science or understanding of clinical trials no longer support this conclusion [14]. As stated above, the FD&C Act specifically states that the FDA should reassess trials when the science has changed. The idea of “fairness” should apply first and foremost to the patients who would be exposed to the drug without adequate evidence of effectiveness. Regulators and investigators should employ the most up-to-date principles in evaluating the results of clinical trials, even if science has advanced since the inception of the trial. It is not appropriate for public health for clinicians and regulators to address these issues “when the next drug comes along” or to avoid addressing these issues because drugs in the past received approvals through the use of older methodology now known to lack scientific credibility or not to meet the FDA’s own standards.

MOVING FORWARD

Seven criteria define adequate and well-controlled trials in US regulations that, in turn, form the basis for the evidence needed to evaluate drug effectiveness [13]. In addition, these same trials
also provide the evidence for evaluating the preliminary safety of a drug. The 7 criteria plus the requirement for evaluating potential harms with "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" [2] form 8 criteria by which to assess the balance of risks and benefit of drugs. As stated above, this is an iterative process, because one must first demonstrate evidence of effectiveness to justify any risks and reassess the balance of risks and benefits when new evidence becomes available.

FDA guidances, including those published under the auspices of the International Conference on Harmonization, provide more detail on the application of these criteria [15, 16]. These same criteria apply to all therapeutic areas, including the study of antimicrobials in infectious diseases. The only differences in trial design are those related to the natural history of the diseases under study. For instance, the definitions and timing of outcomes measurements would differ between an acute self-resolving disease whose natural history lasts a few days and a serious and life-threatening chronic disease lasting several months or years. In 1998, the FDA published draft guidances regarding the trial design for several infectious diseases [17]. Since that time, several other FDA guidances, including those published by the International Conference on Harmonization, have outlined advances in thinking about trial design [15, 16].

The 1998 guidances, at present, are inconsistent with some of these newer principles and, in some places, are frankly at odds with the requirements of adequate and well-controlled trials. Therefore, there are several opportunities for advancement in antimicrobial trial design that may help to provide better evidence related to both effectiveness and safety while acquiring data in a more efficient manner and fulfilling FDA requirements. These principles can serve as a guide for future trials of antimicrobials in infectious diseases and are described below.

1. Clinical trials should have a clear objective. Antimicrobial trials are based on the evaluation of the treatment, prevention, or diagnosis of a recognized disease, not on the evaluation of "response" to specific pathogens [18]. The goal of administering antimicrobials is to decrease the risk of death, improve function, or cure symptoms in patients with a given disease. It is important to realize that clinical trials measure average effects in groups of subjects. This is a major difference from clinical practice, in which one is concerned with the outcome in an individual patient. In most cases, it is challenging, if not impossible, to ascribe causality of drug effect in an individual case. This is the point of randomizing trials and studying groups of subjects to control for confounding variables and to evaluate the effect of the drug compared with spontaneous changes in the course of the disease, placebo effect, or biased observations. There are differences in the population, enrollment and outcome criteria, natural history, concomitant medications and interventions, and other factors across various infectious diseases, even if they are caused by the same pathogen. Including patients with various sites of infection in a single trial makes it challenging to interpret the results in terms of causally relating outcomes to the drug administered. Relating outcomes to the drug prescribed is the very point of conducting a trial. In addition, the sample size for various types of infections may be insufficient to draw conclusions about drug efficacy in various diseases. For instance, it would be challenging to interpret the effect of a drug in individual diseases in a trial that included patients with pneumonia, sinusitis, and meningitis. These diseases have widely different natural histories, and it is likely that there would be few patients with meningitis included in such trials.

Another issue in trials in infectious diseases is deciding whether the trial is evaluating treatment or prevention of disease. Confusing treatment and prevention makes it challenging to define a population to enroll, to define outcomes, and to ascertain the causality of the effect of the drug. For instance, with regard to empirical antifungal therapy in persistently neutropenic patients, investigators have debated whether the goal is to treat occult fungal infections or to prevent fungal infections [19]. The distinction is important in defining enrollment and outcomes, because one cannot prevent a disease that is already present in the patient.

It is also important to distinguish between "explanatory" and "strategy" trials [20]. Explanatory trials attempt to evaluate the causal relationship between the drug and measured outcomes to determine the drug’s effectiveness. Strategy trials measure the various conditions under which the drug might be useful in clinical practice, after determining that the drug is effective. For instance, one might determine that a drug is effective in the setting in which diagnosis of a disease is clear on the basis of laboratory testing in an explanatory trial, but then a strategy trial might show that a drug lacks effectiveness as used in clinical practice in the setting of empirical therapy, in which the diagnosis is less certain. Confirmatory explanatory clinical trials evaluating drug effectiveness should precede strategy trials. It is unlikely that a drug that lacks effectiveness as used in an explanatory trial can have benefits for patients in a strategy trial. There is also the ethical issue of exposing patients to potential harms in a strategy trial without knowing that the drug is beneficial, because patients without the disease who cannot benefit are exposed to potential harms. It is often challenging to determine causal relationships between drug administration and outcomes in strategy trials, and, therefore, it may be difficult to provide generalizable knowledge for clinical practice. For instance, many trials of antiseptics combine educational campaigns with the topical antiseptic under study. If the trial notes a decrease in the number of infections with this strategy, it will not be clear whether the antiseptic, the educational campaign, or both...
were responsible for the benefits. Once one determines that a drug has effectiveness for treatment of a defined disease, then trials can evaluate various strategies for use.

2. Trials should have a quantitative comparison with a control. All clinical trials are comparative [21]. That is, they compare the measured outcomes in a group of subjects who receive a medical intervention with what would have happened had they not received the intervention. Because, in most cases, it is not possible to ascertain with any degree of certainty the outcomes in subjects who would not receive the intervention, trials often use concurrent controls. Investigators may design trials with 1 of 5 types of controls: (1) no-treatment concurrent controls, (2) placebo concurrent controls, (3) exposure response controls, (4) active concurrent controls, or (5) external (historical) controls [13].

Investigators design trials to evaluate whether a test drug may be more effective than a control (superiority trials) or to rule out that the test intervention may be less effective than the control by a chosen amount (noninferiority trials). Many trials in infectious diseases are designed as noninferiority trials. Noninferiority trials cannot show that 2 interventions are “equal,” and the conclusion that 2 interventions are “as good as” each other or “equivalent” is not justifiable unless the test intervention is statistically superior to the control. It is clear that there is a need for a better understanding of the issues related to the design, conduct, and analysis of noninferiority trials, as well as the situations in which noninferiority trials are rational and the situations in which they are not capable of providing the evidence necessary to demonstrate drug effectiveness [22]. To design a credible noninferiority trial, investigators must (1) ascertain the reliable and reproducible magnitude of the benefit of the control drug compared with placebo from previous historical evidence, such as placebo-controlled or historically controlled trials; (2) evaluate the constancy of the effect of the control drug in the planned noninferiority trial by keeping the major design features of the current trial (enrollment criteria, timing and definitions of outcomes, concomitant medications, etc.) essentially the same as in the previous placebo-controlled trials; and (3) choose a margin of inferiority that preserves some of the benefits of the control and is smaller than the amount of benefit of the control compared with placebo [15, 16]. It is not enough merely to know that a control drug “works,” but, as noted above, because the comparison with a control is quantitative, one must know the amount by which the control’s effect exceeds that of the placebo in the setting of a planned noninferiority trial.

It is now clear that previous placebo-controlled trials did not provide evidence of a reliable and reproducible margin of benefit of antimicrobials compared with placebo in some infectious diseases, such as acute otitis media, acute bacterial sinusitis, and acute exacerbations of chronic bronchitis [15, 16, 23–27]. Without this evidence, it is not possible to design a credible noninferiority trial. Noninferiority margins should not be based on clinician opinion alone but must be based on adequate data. The issues regarding noninferiority trials in infectious diseases have been discussed at numerous advisory committees and workshops over the past half decade [15, 16]. Regulators need to clearly specify when noninferiority trials are not scientifically justifiable on the basis of data, as in the study of sinusitis, bronchitis, and otitis. Regulators should also clearly outline the data that sponsors should provide to justify the use of noninferiority trials in diseases in which this trial design is acceptable. It seems incongruous to claim that the need for new antimicrobials is based on the lack of efficacy of older agents in diseases as a result of resistant pathogens and then to design trials to show how much less effective a new drug might be compared with an older drug whose effectiveness is in doubt. Labeling claims for disease due to resistant pathogens are tacit superiority claims for the new drug, compared with the drug to which the organism is resistant [18]. Therefore, one needs to clearly show superiority on the basis of the data in clinical trials. Superiority trials can and should be performed in various situations, such as when the benefit of antimicrobials compared with placebo is not known with certainty for a given disease, for the evaluation of combination therapies, when there are no available therapies for given disease, or for the evaluation of the superiority of newer drugs in a disease caused by resistant pathogens. It seems ironic that the field of infectious diseases introduced the use of placebos in clinical trials [3], yet now investigators in this field argue most strongly against their use. It is not unethical to perform placebo-controlled trials when the efficacy of the control drug is uncertain, when the disease is not serious and life threatening, and when subjects are adequately informed of the trial’s design. There are advantages to receiving placebo, because subjects are less likely to experience adverse events. Indeed, the meaning of the Latin word “placebo” is “to please.” Conversely, it is unethical to enroll subjects in trials that have no chance of providing generalizable knowledge [1]. Finally, in many situations, superiority trials require a smaller sample size than noninferiority trials, which may increase the efficiency of obtaining generalizable data and expose fewer subjects to potential harm.

3. Trials should ensure that patients have the disease under study. The development and use of rapid diagnostics would aid both clinical trials and clinical practice. A better ability to diagnose infections at the point of care would streamline enrollment by allowing enrichment of trials for patients whose disease is due to resistant pathogens and would allow the study
of narrow-spectrum drugs. The use of rapid diagnostics in clinical practice would decrease the inappropriate use of antimicrobials for viral disease, help to preserve the utility of older drugs, and decrease adverse events in patients who received a drug that they did not need. The use of biomarkers to help define the disease under study could revolutionize both clinical practice and infectious diseases trials. Again, there is an important distinction here between clinical trials and clinical practice. Whereas patients in practice may receive drugs empirically, in clinical trials, subjects are exposed to experimental agents whose toxicity is not known. Therefore, it is even more important to ensure they have the disease under study. Exposing subjects to the potential toxicities of experimental agents when they do not have the disease under study in treatment trials exposes them to potential risks for no benefits, which raises ethical issues, as well.

4. Trials should ensure baseline comparability of patients. The process of randomization and blinded allocation concealment helps to give an equal probability of baseline comparability in trials and guards against selection bias. However, randomization is not foolproof, and baseline imbalances still can occur. Imbalances are more likely when the population under study is highly heterogeneous, such as when patients with a wide array of diseases are enrolled in a single trial, as noted above. Choosing the appropriate population is a matter of balance between choosing a population that is heterogeneous enough to extrapolate the results of the trial to general practice and choosing one homogeneous enough to obtain useful results. A trial that lacks internal validity cannot have external validity in clinical practice. One of the first reported randomized trials evaluated streptomycin in pulmonary tuberculosis. However, some investigators have suggested that observational studies should become the standard in some situations for evaluating drug effectiveness [28, 29]. Referring back to the scientific method, observational studies provide what their name implies—namely, observations. These observations measure associations, but it is challenging to imply that these associations provide evidence of a causal relationship between drug effect and clinical outcomes. Austin Bradford Hill's landmark work on evaluating associations and causality [30] points out that experimentation is one of the clearest ways in which to adequately evaluate causality. Observational studies can be useful to generate the hypotheses for future randomized clinical trials.

5. Trials should attempt to minimize bias. Investigators should double blind trials whenever possible. Microbiological test results should also be blinded (or partially blinded) whether the administration of drugs is blinded or not. Clinicians often make clinical decisions on the basis of microbiological test results that are surrogate end points for the clinical outcomes of patient symptoms, function, and survival. In some cases, there is a poor correlation between clinical and microbiological outcomes [28, 31, 32]. If clinicians deem a patient as having experienced treatment failure in clinical trials on the basis of microbiological testing when the patient is clinically cured, this inserts a bias into the trial. This also actually can make it more difficult for a given drug to show effectiveness and decreases the efficiency of the trial. The utility of "provisional breakpoints" to define susceptibility of organisms is not clear in most cases and may introduce bias into trials. Because investigators will follow patients closely for clinical outcome and, in many cases, will not know the susceptibility of organisms until several days into therapy, investigators should use the data from clinical trials to measure the correlation of organism "susceptibility" with clinical outcomes. Assigning outcomes according to microbiological criteria alone indicates an assumption that one's hypothesis is true before it has been tested.

6. Trials should have well-defined and reliable end points. Developing appropriate end points is perhaps one of the most intriguing and complex areas in drug development. A detailed discussion is beyond the scope of this article; however, several basic principles apply. End points should measure outcomes that are clinically relevant to patients and capture the net harms and benefits of an intervention on those outcomes. Trials should measure these end points without bias. Biomarkers do not directly measure outcomes that are clinically relevant to patients and can provide misleading results if they are not developed appropriately [33]. Investigators need to compare outcomes with biomarkers and clinical outcomes, but, in short-term diseases, there is less of a need to measure biomarkers as surrogate end points, because investigators can measure the actual clinical outcome directly [32]. Biomarkers are most useful in evaluating chronic diseases in which the clinical end points can take months or years to measure. Use of a combination end point of clinical outcomes plus biomarkers can actually make it more difficult to demonstrate the effect of a drug in some situations. For instance, patients with pneumonia are often clinically cured long before the chest radiograph findings normalize. If one used a combination end point of clinical cure plus normalization of chest radiograph findings, the true effect of the drug on clinical outcomes would be underestimated.

Much of the confusion regarding surrogate end points in infectious diseases stems from a belief that a negative culture result "should" predict a successful clinical outcome. This confuses mechanism of action with outcome and confuses measurement of risk factors with measurement of drug effect [32]. Biomarkers may be useful in diagnosing a disease, but this does not mean that they will be similarly useful in evaluating outcomes. The term "presumed eradication" should be removed from clinical trials, because it presumes what one is trying to measure in the first place—namely, the correlation between clinical and microbiological outcomes—and illogically substi-
tutes the clinical outcome for the surrogate outcome, obviating the need for a surrogate. Use of patient-reported outcomes (PROs) can provide a more patient-centered and, perhaps, more sensitive way of measuring clinically relevant patient outcomes. PROs can also provide information in addition to the more conventional end points that measure mortality. Investigators have used PRO instruments in other therapeutic areas, such as analgesia, to measure patient-centered outcomes, and their use in infectious diseases like sinusitis, in which the findings of the disease in patients are related primarily to symptoms, is entirely rational. Infectious disease investigators were leaders in this field; the trial of patulin to treat the common cold, published in 1944, used PROs as the primary end point [3].

End points based on investigator discretion or on “cause-specific” outcomes related to infection often cannot be measured without bias. Another way of evaluating the effect of antimicrobials, especially in short-term diseases, would be to evaluate the time to clinical cure. PROs would allow collection of data on a continuous basis, and such continuous data would allow greater power to determine differences between drugs and allow a more rational consideration of the benefits of a drug, in terms of shortening the disease, balanced against the risks. For instance, clinicians may consider the balance of risks to benefits to be clinically relevant if an antimicrobial shortens the course of illness by 5 days but not if it shortens it by 5 h. It is hard to justify serious adverse events like anaphylaxis and liver failure for small benefits in self-resolving diseases. If we are willing to tolerate symptomatic adverse events like nausea and diarrhea to prevent death due to diseases like pneumonia by administering a drug, should we not also consider it beneficial to prevent death due to adverse events and to tolerate uncomfortable symptoms in self-resolving diseases by withholding drugs? This is a point that needs further discussion. Finally, time-to-event analyses may allow us to more accurately measure the duration of antimicrobial therapy necessary to result in a cure and to individualize the duration of antimicrobial therapy. Prescribing less drug for the same effect may decrease adverse events, decrease resistance, decrease costs, and increase adherence to medication. Patients have already performed the “observational study” for us, because many patients discontinue therapy after resolution of their symptoms, and most do not experience relapse.

7. Trials should have an appropriate analysis of the data. Investigators should evaluate the information obtained from a trial in a way that allows an estimation of the uncertainty regarding the conclusions. Statistical testing is a tool that investigators use to evaluate the variability in the data and the probability that the results may occur by chance alone. There are a number of issues in clinical trials in infectious diseases that would benefit from increased discussion and research, such as the appropriate analysis populations (intent to treat, per protocol, or both) to use in various trial designs and the appropriate analysis of secondary end points and subgroup analyses. For instance, the 1998 FDA guidances point to the “per protocol” population as the primary analysis population in noninferiority trials. However, such analyses are really post hoc subgroup analyses that do not include all randomized patients and may not preserve the integrity of randomization, which, in turn, may result in selection bias and misleading findings [34]. There are issues with the intent-to-treat population in noninferiority trials as well, because this population may bias toward a finding of noninferiority. Consistent findings across various analysis populations provide the most convincing evidence of drug effectiveness. Investigators in infectious diseases trials have given little discussion to the topics of adjustment of the type I error for multiple comparisons of different end points, comparisons of the same end point over time, or multiple subgroup analyses with the attendant increased risk of false-positive results. One should address these issues in addition to specifying any secondary or subgroup analyses before initiation of the trial.

8. Trials should provide an adequate evaluation of the safety of a medical intervention. Unlike the evaluation of drug effectiveness, confirmatory trials of drugs are not evaluating a hypothesis related to drug safety. One often does not know the adverse events caused by a drug before testing in confirmatory trials to form a hypothesis. Rather, confirmatory trials may provide safety signals for further evaluation. The sample size for confirmatory trials often is based on an evaluation of effectiveness and is insufficient to draw statistically based conclusions regarding potential toxicities of a drug. Therefore, the absence of a statistically significant difference in adverse events in a trial does not imply that there is no difference between interventions. If no serious adverse events are noted, a database of 300–500 patients is required to rule out a risk of 1.0% with 95% confidence [35]. When the evaluation of safety in this number of patients does show a potential signal for a particular type of toxicity, sponsors should perform follow-up studies to more clearly evaluate that toxicity in a more rigorous way. Approval of a drug, usually based on findings in a few thousand subjects, is only the beginning of evaluating the safety of a drug, and vigilant postapproval studies should become standard practice. The principles of differentiating associations and causality espoused by Hill [30] are helpful in evaluating safety signals.

**BALANCING RISKS AND BENEFITS**

There has been substantial discussion regarding the need for new antimicrobial therapies at a time when major pharmaceutical firms are choosing to exit research and development in this field [36]. Despite the gloomy news, there have been more antimicrobials with novel mechanisms of action approved in the past few years than in the previous 40 years combined.
Addressing the medical needs related to antimicrobial resistance will require the concerted efforts of clinicians, researchers, regulators, patients, and policy makers. It is important to keep in mind, however, that a medical need is exactly why we need adequately designed trials to address important public health questions. As stated above, a medical need provides the rationale for why one performs a trial, but it should not imply a lower standard for evaluating such therapies. The story of laetrile, a drug made from apricot pits, which many claimed to have effectiveness in a variety of cancers, is illustrative of this problem. Despite the claims of many advocates, investigators subsequently showed the lack of effectiveness in randomized trials, despite the claims of medical need [12].

One of the issues for which there is an urgent need for further discussion is the overall assessment of the risks and benefits of antimicrobials. It is clear that trials attempting to evaluate the effectiveness of antimicrobials for self-resolving diseases like acute otitis media, acute bacterial sinusitis, and acute exacerbations of chronic bronchitis have not met the FDA’s own standards of substantial evidence of effectiveness from adequate and well-controlled trials. In September and December of 2006, 2 separate FDA advisory committees voted that noninferiority trials did not provide substantial evidence of effectiveness for various drugs in acute bacterial sinusitis and acute exacerbations of chronic bronchitis [26, 27]. The same issues apply in acute otitis media. If we do not have adequate evidence of effectiveness for antimicrobials, we risk promoting the very problem we are trying to solve. The widespread use of antimicrobials for self-resolving infections without evidence of effectiveness will promote antimicrobial resistance. Resistance will then limit the effectiveness of that same drug as well as that of other drugs in serious and life-threatening diseases for which there may be substantial evidence of effectiveness. In addition, because self-resolving infections are more common than serious and life-threatening diseases, the absolute numbers of patients experiencing adverse events will be greatest for these diseases. However, the benefits for these patients will be the least, in terms of both the magnitude of the effect and the qualitative nature of that effect, which is usually only shorter symptom duration. This is in contrast to the large decrease in mortality among patients receiving effective antimicrobials, compared with those receiving placebo, for serious and life-threatening diseases. The inherent mind-set that antimicrobials are highly effective in all diseases regardless of their natural history and that prescribing them “won’t hurt” has led to widespread antimicrobial resistance and adverse events that are not justifiable, especially in patients with viral illness. Even if antimicrobials are effective in self-resolving bacterial diseases, appropriate trials that measure the magnitude of the benefit in terms of time to resolution of illness and that compare the number needed to treat to benefit with the number needed to treat to harm are urgently needed to address the important societal question of when the risks are justified by the benefits. The medical community needs to have a serious discussion about these important issues of balancing risks and benefits. For instance, if antimicrobials indeed shorten the duration of ear pain associated with otitis by a day but cause as many or more cases of diarrhea and less commonly cause serious adverse events like liver failure and death, is this treatment justifiable in small children? Several studies show that there may be long-term consequences of treatment with antimicrobials, as well [37]. Superinfections such as *Clostridium difficile*-associated disease, which can occur as a result of antimicrobial administration, are also becoming more common. Traditionally, clinicians have overemphasized the benefits while minimizing the adverse effects of antimicrobial therapies. The societal problem of antimicrobial resistance is well known, but the issue of unwarranted adverse effects in patients in whom the benefits are small or unmeasurable should receive equal discussion. Clinicians need relevant information on situations in which new drugs may be superior to older drugs; therefore, there is a need to perform superiority trials to provide this information. Even in trials in serious diseases like severe pneumonia, in which the end point should be all-cause mortality, investigators can evaluate appropriately designed secondary superiority hypotheses to evaluate outcomes like time to clinical success with one drug compared with another. It is long past time that we adequately evaluated the hypothesis of whether increased in vitro potency has important clinically relevant benefits for patients or whether the host immune response makes such differences in vitro irrelevant to patient outcomes. This is especially important when more potent drugs in vitro have more adverse effects in vivo. For instance, many authors have pointed to vancomycin as a suboptimal alternative for serious and life-threatening diseases due to *Staphylococcus aureus*, but randomized controlled trials have not shown new drugs to be superior to vancomycin in terms of clinical outcomes. On the other hand, some investigators have hinted at the superiority of newer drugs on the basis of subgroup analyses, but these analyses leave considerable uncertainty as to the true effect of these drugs [38]. Such subgroups can form the basis for further randomized trials.

The use of meta-analysis and pooling of results of very different trials may give misleading results as well [39]. Recent meta-analyses quoted as showing an “effect” of antimicrobials in various infections combine trials involving very different populations, different definitions of disease, and different definitions of outcome [40, 41]. In addition, these meta-analyses often do not include all of the available data and instead include only the trials producing the data the authors wish to evaluate [40]. Studies show that, almost half the time, subsequent randomized trials do not confirm the results of previous meta-
analyses [42]. Although properly conducted meta-analyses may be useful, meta-analyses combining flawed trials do not eliminate the inherent biases and confounding factors in the trials included in the analysis and actually increase the potential bias and resultant inaccuracy of pooled results despite apparently acceptable statistical tests for homogeneity. The increased sample size in meta-analyses allows more precision, but this is of little utility if the answers at which we arrive are just more precisely incorrect.

The various sectors of the medical community need to work together to answer the important questions regarding the treatment and prevention of infectious diseases. Problems that seem “too hard” to solve individually can be solved with our combined efforts. Patients need to understand when antimicrobials are indicated and when they are not. Clinicians need to make appropriate diagnoses and to demand the data that they need to make accurate decisions. Researchers need to study rapid, point-of-care diagnostics that clinicians can use to accurately diagnose infections, as well as study the natural history of disease to design trials most appropriately. Clinical trialists need to design infectious disease trials to answer these important questions and provide relevant information to clinicians. Finally, regulatory agencies need to provide timely, accurate, and scientifically based advice on how to design clinical trials that are based on the best information we have today and not on outdated guidance that does not follow the FDA’s own regulations. There is an urgent need to remove the outdated 1998 guidances from the FDA Web site, because they provide misleading information to sponsors. There is an even more urgent need to replace these older guidances with updated guidances based on appropriate scientific principles, which are the true basis of drug regulation. These principles are already outlined in general FDA guidances, so applying them to the setting of infectious diseases trials should be not be an insurmountable task [15, 16]. Drug sponsors should also realize that it is time to move beyond the noninferiority trial design and provide clinicians with the evidence that they need to make appropriate clinical decisions and show the added benefits of drugs. Clinicians do not “reserve” drugs that truly are proven to be clinically superior, so there are advantages to sponsors in providing evidence of the clinical superiority of one drug compared with another. Implying superiority on the basis of in vitro testing does not provide clinicians with the clinical evidence of superior outcomes in patients that they so desperately need. These important public health goals are eminently achievable if we all pool our resources for the common good.

Acknowledgments

If we all pool our resources for the common good. These important public health goals are eminently achievable if we all pool our resources for the common good.

Supplement sponsorship. This article was published as part of a supplement entitled “Annual Conference on Antimicrobial Resistance,” sponsored by the National Foundation for Infectious Diseases.

Potential conflicts of interest. J.H.P. is a consultant for AstraZeneca, Cerexa, Centegen, Johnson & Johnson, Merck, Methylgene, Octoplus, Takeda, Theravance, and Wyeth.

References