Tribulations of trials for antibacterial drugs: interview with Brad Spellberg

Brad Spellberg is an Associate Professor of Medicine at the University of California, Los Angeles and is based at the Division of General Internal Medicine at the Los Angeles Biomedical Research Institute at Harbor-University of California, Los Angeles Medical Center. His research is diverse, ranging from NIH-funded basic immunology and vaccinology to pure clinical and outcomes research. He is the co-founder and Medical Director of Clinical Research Solutions, a clinical trial unit at his research institute. He is also a Fellow in the Infectious Diseases Society of America (IDSA) and is a member of the IDSA’s Antimicrobial Availability Task Force, formed to bring attention to the problems of increasing drug resistance and decreasing new antibiotics. His research regarding new drug development has been a cornerstone of the IDSA’s white paper, ‘Bad Bugs, No Drugs’, and has been cited extensively in medical literature and on Capitol Hill. He has lead authored numerous IDSA position papers and review articles relating to public policy of antibiotic resistance and antibiotic development. Brad Spellberg is also the author of Rising Plague, which he wrote to inform and educate the public about the crisis in antibiotic resistant infections and lack of antibiotic development. Brad Spellberg, a previous contributor to Clinical Investigation, speaks to Laura Harvey at the journal, giving his personal opinion on the difficulties facing antibacterial drug trials today.

Q What would you say were the most pressing motivations for the recent US FDA re-evaluation (and following workshops) of non-inferiority trials as a means of assessing drug efficacy for bacterial diseases?

The primary driver of the recent revisions regarding antibacterial development is the realization that establishing non-inferiority of an experimental to a comparator drug results in two equally possible statistical interpretations:

- Both drugs are better than placebo, and thus the experimental drug should be approved;
- Neither drug is better than placebo, and the reason why the experimental drug appears to be non-inferior to the comparator drug is that an equivalent placebo effect occurred in both arms.

Hence, the statistical concern of non-inferiority trials is that a positive trial result could accidentally result in regulatory approval of a drug which is in fact not more effective than placebo, leading to public harm. This is a general concern about all drugs, not specific to antibacterial agents, but the proposed solution creates a problem that is far greater for antibacterial agents than other drug classes.
Could you expand on this proposed solution and why it impacts more greatly on antibacterial agents?

The resolution of this concern is to require that the comparator drug, used in a current non-inferiority trial, has been previously shown to be superior to placebo in an older clinical trial. Using the following principle, there can then be regulatory comfort that the experimental drug is superior to placebo:

- If the comparator drug was previously shown to be superior to placebo; and
- The experimental drug is now shown to be non-inferior to the comparator drug; then
- The experimental drug can be safely inferred to be superior in efficacy to placebo.

This resolution of the concern about non-inferiority trials is logical and acceptable for most drugs. Unfortunately, antibacterial agents were among the first effective drugs used in medicine. The first sulfonamide hit the market in the USA in late 1936. That drug predated the use of randomized, placebo-controlled trials by at least a decade, and their routine use by 15–20 years. Thus, in contrast to virtually all other classes of drugs, placebo-controlled trials of antibacterial agents for the treatment of serious and life-threatening infections were never done and never will be done, because they were unethical to conduct by the time the technology to conduct them became available.

In the absence of randomized, placebo-controlled trials for antibacterial agents, would you say that non-inferiority trials of these agents place the public at risk?

It has been hypothesized that non-inferiority trials of antibacterial agents have placed the public at risk by failing to detect inferior therapy, resulting in approval of inferior or ineffective drugs. The question is, is this hypothesis correct? There are at least three testable implications of this hypothesis:

- First, after decades of use, there should be recognizable instances where antibacterials approved based on non-inferiority studies were less effective, or ineffective, for the treatment of serious/life threatening infections;
- Second, there should be few or no examples of non-inferiority studies that detected inferiority of the experimental therapy;
- Third, there should be little to no alternative data (i.e., not randomized placebo controlled) documenting by how much antibacterial agents are superior to placebo/background medical therapy using specific end points.

Could you expand on this proposed solution and why it impacts more greatly on antibacterial agents?

Third, there should be little to no alternative data (i.e., not randomized placebo controlled) documenting by how much antibacterial agents are superior to placebo/background medical therapy using specific end points.

In your experience have you observed any of these indications that non-inferiority trials place the public at risk of inferior antibacterial therapy?

After years of public debate, I have only encountered one example offered of an antibacterial agent approved based on non-inferiority which was claimed to be inferior for the treatment of a serious or life-threatening infection. At an FDA advisory committee meeting in 2009, the approval of ciprofloxacin for the treatment of community-acquired pneumonia (CAP) was referenced as an example of non-inferiority studies leading to approval of an agent that everyone now agrees is not optimal for treating CAP [101]. However, this is a bad example. When ciprofloxacin was approved, it had very good Gram-positive activity, and was an effective therapy for CAP. Indeed, much of the Medicare data demonstrating that delayed initiation of antibiotic therapy for CAP resulted in higher mortality was based on delayed vs. early initiation of ciprofloxacin [John Bartlett, Pers. Comm.]. The problem with ciprofloxacin is that Gram-positive bacteria rapidly developed resistance to it. Hence, it is no longer useful for Gram-positive infections, and second and third generation fluoroquinolones had to be developed which had superior Gram-positive activity and are now standard of care for CAP. Those subsequent fluoroquinolones became available only because standard non-inferiority trials using standard end points were available. Hence, I think this example actually underscores why non-inferiority trials are necessary and why they have worked well for antibacterial products, generating over many decades what is in reality an awesome arsenal of extremely effective drugs that save lives.

What about the second testable implication?

This implication is that non-inferiority studies of antibacterial agents are poor at detecting inferior or ineffective therapy. In just the past few years, there have been multiple examples of antibacterial drugs which have been found to be inferior in pivotal non-inferiority studies, and science is available to explain why the drugs were inferior (confirming that these were not false-negative studies). For example, tigecycline and ceftobiprole were inferior for the treatment of nosocomial pneumonia, and the scientific explanation was that the drug levels achieved in patients with ventilator-associated pneumonia (VAP) were half what was predicted, while minimum
Evidence of therapy that was only moderately inferior to comparator regimens.

Q So that’s the first two indicators dealt with, what about data documenting how much antibacterial agents are superior to placebo or background medical therapy?

The question is, are there alternative data to randomized placebo-controlled trials for antibacterial agents that enable us to confirm by how much antibacterial efficacy is superior to placebo using specific end points? Indeed there are! And they come in several forms, and have been summarized by numerous publications over the past several years for multiple disease areas:

- First, in the 1930s and 1940s, physician scientists conducted historically controlled examinations of antibacterial effectiveness, comparing outcomes in patients treated in the period before antibacterial agents with those of patients treated in the period after antibacterial agents became available;

- Second, in the 1930s concurrent controlled studies were conducted in which patients with bacterial infections were alternated to antibacterial therapy or background medical therapy;

- Third, natural history studies are available spanning the pre- and post-antibiotic era documenting mortality of patients with bacterial infections on an annual basis;

- Fourth, modern dose-escalation studies are available (e.g., for complicated skin and skin structure infections) which document the difference in outcomes in patients treated with lower versus higher doses using the standard ‘clinical cure’ end point;

- Fifth, modern pharmacokinetic/pharmacodynamic data are available in which outcomes in modern, randomized controlled studies are compared between patients who had higher drug exposure levels (i.e., higher drug levels in blood) versus lower drug exposure levels (i.e., lower drug-exposure levels in blood) using the standard ‘clinical cure’ end point;

Finally, data are available documenting outcomes of patients after a delay in initiation of effective therapy (either because therapy was started late or because initial therapy was ineffective) versus rapid administration of effective therapy.

All of these types of data confirm that antibacterial agents are enormously more effective than background medical therapy for various types of bacterial infections as defined by specific end points. Thus placebo-controlled superiority studies cannot be conducted for these diseases, underscoring the absolute public health need for feasible non-inferiority studies. These data also define specific end points and non-inferiority margins for modern non-inferiority studies.

So, three testable implications fail to demonstrate any evidence that non-inferiority trials using previous end points have ever or are likely ever to lead to approval of ineffective antibacterial agents for serious or life threatening infections. The fact is, the antibacterial arsenal now available to modern medicine is awesome in its power, and this arsenal would not have been possible without the very non-inferiority studies that have been called into doubt by a hypothesis with no data to support it.

Q If, judging by the indications you set out, non-inferiority studies are capable of assessing new antibacterial therapies, what steps would you say are necessary to address current concerns and what would you say is the biggest concern for patients, if not the emergence of inferior therapies?

While there is no evidence that patients have been harmed by treatment with inferior antibacterial therapy for serious and life threatening infections resulting from non-inferiority studies, it is unequivocal that patients with antibiotic-resistant infections are being harmed by lack of antibacterial development. A massive overreaction to the hypothetical concern of non-inferiority studies is at least partly responsible for this lack of antibacterial development. Proportionate, targeted measures to modernize, reform, and improve the conduct of non-inferiority trials are adequate to address the hypothetical concern, and would not have the unfortunate side effect of paralyzing critically needed development. Such measures could include (but are not limited to) converting standard subjective ‘clinical response’ end points
into objective, check-box end points so that auditable, objective end points are used for all trials, and insisting that a higher percentage of patients enrolled have confirmation of bacterial infection than previously. For example, previous identification of the bacterial etiology of infection in pneumonia studies has generally been in the 25–35% range.

Q: What effect do you see the new regulatory guidance having on drug development?

The FDA should be applauded for the rapid pace at which they recently have been working on new guidances to provide clarity to clinical trial pathways. After many years of asking, we hope to have final guidances soon for many diseases, including skin infections, bacterial CAP, hospital-acquired pneumonia and VAP, and possibly for complicated urinary tract infections. Of course, we don't have final guidances yet for any of these diseases, so exactly how the studies will look in the future is not yet known.

What is clear after years of debate is that however things settle out, trial sizes in all disease areas will be much larger than previously due to smaller margins and stricter evaluability criteria. Furthermore, due to unfamiliarity with the biology of acute bacterial infections, proposed end points seem to have become either totally irrelevant or questionably relevant, to patients and doctors. Finally, certain elements of trial conduct, such as the desire to enroll patients before a single dose of nonstudy antibiotics has been administered, and the desire to have 100% microbiological confirmation of bacterial infection for community-acquired bacterial pneumonia or hospital-acquired bacterial pneumonia/ventilator-associated bacterial pneumonia, could well make trials impossible to conduct. No one knows what the final guidances will look like. But if they do not advance beyond the pure statistical desire for perfection, and incorporate clinical reality, the trials may well be impossible to conduct and irrelevant to people who need the drugs.

Q: You say that the biological nature of acute bacterial infections has made some of the proposed end points either totally irrelevant or questionably relevant. What do you think these end points should reflect?

There is a standard set of rules one must obey when setting end points for pivotal non-inferiority trials:

- First, the end points chosen must be based on ‘historical evidence of sensitivity to drug effect’, or HESDE (as described in International Conference on Harmonisation [ICH] E9 and E10). That is, the end points must be based on those previously used to establish efficacy of comparator antibacterial therapy relative to placebo/background medical therapy. This limitation does not apply to superiority studies, only to non-inferiority studies. I’ve already mentioned that there is indeed HESDE for antibacterial efficacy for defined end points, which provides reassurance that non-inferiority trials for these agents are extraordinarily unlikely to lead to approval of ineffective therapy for serious and life-threatening infections;

- Second is the need to establish ‘constancy’, that is, that the data used to justify that modern trial end points are still relevant to medicine in the 21st Century. An example of the concern raised by lack of constancy is the justification for the newly proposed primary end point for non-inferiority trials of skin infections, which derives from two studies conducted by Snodgrass and Anderson in 1937 [4]. In those studies, comparing sulfonamide therapy with UV lamp therapy, background medical therapy for all patients consisted of a liquid diet, including Horlick’s malted milk, arrowroot, junket, the forbidding of eggs and onion intake and treatment with mandatory liquid paraffin enemas. Can such studies, by themselves, and absent any modern equivalents, possibly be considered relevant to clinical trials or medical practice in the 21st Century? The bottom line is that historical data must be validated by modern data comparing effective vs. ineffective antibacterial therapy for the same end point. In the absence of concordance across modern and historical datasets, the use of single studies from the 1930s cannot meet the need to establish constancy across datasets which is a cornerstone of justifying non-inferiority end points and margins, according to ICH standards (discussed in E9 and E10 guidances);

- Third, once end points and margins are chosen for non-inferiority studies that meet HESDE and fulfill constancy, the next issue becomes selecting end points that are most likely to detect efficacy of drugs. The term for this is assay sensitivity. A good example of a problem with assay sensitivity again derives from the proposed new end point for skin infections, for which treatment success is defined as stabilization of lesion size by day three of therapy. As mentioned, this end point derives explicitly from the two Snodgrass and Anderson studies of sulfa drugs versus UV lamp therapy in the British Medical Journal from 1937 [4,5]. Sulfonamides had a >99% efficacy for this end point in 1937. We know from numerous sources of data that sulfonamides were substantially less effective than...
β-lactam therapy for the treatment of infections, including complicated skin infections [6]. Given that sulfonamides were clearly less effective than modern antibiotics and yet had a >95% success rate with this end point, how can this end point possibly distinguish less effective drugs from more effective drugs? This end point has no assay sensitivity according to the very data used to justify the end point. It is ironic that the statistical concern that non-inferiority studies could lead to approval of inferior therapies has led to selection of a new end point for future studies which, according to the very historical data used to justify it, can’t distinguish inferior from superior therapy.

Finally, recent experience suggests that the cessation of lesion spread end point has much lower success rates (e.g., 70%) in modern studies than in the Snodgrass and Anderson data. This is another example of lack of constancy. If the end point resulted in 95% success rates in 1937 and 70% success rates now, clearly there is no constancy for this end point, and it therefore violates the fundamental ICH principles upon which the FDA has been saying for years non-inferiority end points must be based.

Q: So using these criteria, what do you think is key in choosing an end point for antibacterial therapy trials?

For antibacterial therapy for acute infections, the end point should reflect the concept that the disease is eradicated. Antibacterial agents must not be judged by the same standards as those for treatment of chronic illnesses. Antibiotics eradicate their target disease. The goal is cure. The signs and symptoms caused by the infection should be gone when the treatment is stopped, or shortly thereafter. We all know that some signs and symptoms can linger for a period even after eradication of the etiologic organism. Hence, use of a test-of-cure end point, which occurs 1–2 weeks after end of therapy, is important not just to evaluate for relapse, but also to allow enough time for host response to infection to resolve before adjudicating success or failure.

Q: Practically speaking, how would you go about designing such an end point?

The key is to make an objective list of signs and symptoms caused by the infection which were not present before the infection, are present during the infection, and are assessed for resolution at test-of-cure. In contrast to the subjective ‘clinical cure’ which has predominated in trial designs previously, this new, auditable, objective list should become the means to determine treatment success or not. The mere fact that symptoms are more rapidly better cannot be defined as treatment success if the symptoms do not resolve in the end as a result of therapy. For acute bacterial infections, objective end points which document abject cure are the only end points that are clinically acceptable to patients or providers.

Q: One of the main criticisms of antibacterial trials is that end point assessment, often based on investigator assessment, is very subjective. What difficulties are there in constructing an end point such that it is more objective and implementing specific defining criteria? Do you have any suggestions for an approach to choosing specific criteria of particular relevance for a given trial?

If you ask, what are the signs and symptoms associated with an infection, which are not present before the infection, are present during the infection, and resolve at end-of-therapy, these can be objectively defined. For example, complicated skin infections cause redness, pain, swelling, loss of mobility of the infected area, occasionally bacteremia, fever, elevated white count, tachycardia and so forth. Pneumonia causes cough, chest pain, hypoxia, tachypnea, tachycardia, fatigue, fever, elevated white count and so forth. Not all patients have all of these signs and symptoms. But one can create a master list of objective signs and symptoms, document during the patient’s baseline evaluation which of those signs and symptoms are present from the master list, and then document resolution of those signs and symptoms at test-of-cure. It’s conceptually very simple, auditable, objective, and clinically relevant. This simple solution deserves to be discussed at a national level because it could rapidly resolve what has been many years of ongoing contentious debate, which in some cases seem to have led to bizarre clinical trial design proposals.

Q: Moving on to the issue of mortality as an end point – although this is historically well established, how would you assess the advantages and disadvantages of this as an end point given that it is only high for very small specific subgroups?

I think that mortality is a reasonable end point for diseases that have very high fatality rates despite effective antibiotic therapy. There are very few bacterial infections that fit into this category. Septic shock is one. Some critics have suggested that mortality is the preferred end point because it is objective and ‘most meaningful’ to patients. That is not necessarily true. If you asked most people if it would be meaningful to survive in a vegetative state, they would say absolutely not. So...
yes, people always want to survive, but they want to survive in the same physiological condition they were in before the infection. To put it another way, survival is a key component of success, but survival by itself is not enough to be a success.

Another problem with mortality for severe infections is that many things aside from antibacterial therapy can affect mortality of some infections. For example, it has been pointed out repeatedly that patients with VAP treated with antibacterial agents frequently die of their underlying diseases, not of their pneumonia. In contrast to superiority studies, for non-inferiority studies, confounders which drive the end point aside from drug therapy bias the study towards success (make it easier to show non-inferiority). Again, ironically, the angst over the ability of non-inferiority studies to distinguish effective from ineffective therapy has resulted in selection of an end point for VAP studies that is likely to be driven by nontherapy variables which could obscure actual differences in drug efficacy (i.e., make it easier to show non-inferiority).

Finally, for many infections, mortality is impractical as an end point because the death rate with effective therapy is so low that many thousands of patients would have to be included to adequately power the study (e.g., community-acquired bacterial pneumonia).

The bottom line is that mortality is a critical component of clinical cure, and should be included as one objective component of the definition of cure. Mortality is insensitive (it is not enough to survive with marked decline in function) but highly specific (if the patient dies, the patient is a true failure) as an end point for non-inferiority studies. It should be included as a composite in the primary end point, but rarely relied on as the sole end point.

Q: How would you recommend that the results of composite end point analysis be treated such that trends in mortality are not obscured and that some of the individual components of the composite will be more important to the patient than others, for example 'death' versus 'fever'?

There are two ways to address this issue. The first is to simply require that non-inferiority be met by clinical composite end point, but still reserve the right to reject the drug if there is a concerning trend in the wrong direction for a mortality end point. Because the study is not powered to detect a mortality signal, the difference in mortality as a component subset does not have to be statistically significant. It merely has to be a trend in the wrong direction. This concept is frightening to industry because it means a drug could be rejected for a stochastic trend in the wrong direction. But, the fact is, this is already standard practice. The FDA already has authority to approve or reject a drug on any grounds, irrespective of meeting the primary end point. And certainly a trend in the wrong direction on mortality would already lead FDA to reject drugs that otherwise met their clinical primary end point.

The second, more complex way is to use a hierarchical end point. In the hierarchical end point, you specify which element(s) must be achieved first in order to proceed to the second (or more) layer of the hierarchy. So, for example, one could say that the patient must have survived. If not, the patient has failed, period, irrespective of other factors, and there is no further testing. But, if the study meets non-inferiority for survival, then the second layer of the hierarchy, for example, a composite of erythema, swelling, and pain resolution, must be resolved. The drug must be non-inferior for both layers of the hierarchy to be considered non-inferior to the comparator drug. However, because the testing is sequential and does not proceed to the second layer if the first layer fails, this approach does not cost alpha and does not raise multiple comparisons issues.

Note that for this hierarchical approach to be valid, the study must be adequately powered for all elements of the hierarchy. So, if the first layer of the hierarchy is mortality, the study must be powered to detect a difference in mortality. Such an approach would likely be infeasible for many infections treated with antibacterial therapy, given the low resulting death rates. Exceptions could include VAP, bacteremia, and severe sepsis.

Q: Timing of end point assessment is also an issue, as it is now possible to microbiologically assess disease development/progression/regression during a trial, improvement can be seen as early as 3–5 days with some diseases, raising the questions as to the timing of end point assessment. What is your opinion on the timing of end-point assessment and what difficulties must be considered when deciding when to assess?

Rapidity of resolution could be a secondary end point, or perhaps included as a component of a hierarchical end point. But it is not acceptable as a sole primary end point. Acute bacterial infections must be cured to be considered a treatment success. So, primary efficacy analysis must occur at end-of-therapy or test-of-cure.

Q: What is your opinion on the use of multiple assessments? What cost/ease of implementation implications would this have?

Assessments always occur at multiple time points in studies. John Rex (AstraZeneca) has repeatedly, publicly, pointed out that in all clinical trials, patients can fail at
any point but can only succeed at the end. So, this notion that previous trials failed to consider early events is wrong. If patients are doing poorly during daily assessments they are withdrawn from the study and considered treatment failures. So, multiple ‘assessments’ do and should occur. But, assessment for cure can only be done at the end, and inherently incorporates all of those early assessments (because the patient could have failed at any one of them).

Q What effect do you see the regulatory uncertainty in antibacterial trial design having on further research into developing effective treatments?

This is an easy one! Companies have voted with their feet. There are perhaps only two big pharmaceutical companies left with active discovery programs: GlaxoSmithKline and AstraZeneca. That’s down from perhaps 20 companies two–three decades ago. All others have gotten out of this business. We need to fix the economics of antibacterial R&D. But even if we fix the economics, if we don’t fix the regulatory problem, we won’t get new drugs. And our patients with antibiotic-resistant infections are the ones who suffer.

Q The need for new antibacterial agents is becoming more urgent – what would you say the vital next steps are in trial design to ensure that much-needed effective treatments are brought to the market?

I think the key to this question is ‘much-needed effective treatments’. The medical need currently is for new drugs to treat extreme or multidrug resistant Gram-negative bacilli. This need can be addressed by creating new pathways to approval, such as superiority, historically controlled, and organism-specific pathways, combined with economic incentives that make companies want to develop these drugs despite a small intrinsic market size. VAP and urinary tract infection non-inferiority pathways could also lead to new agents from Gram-negative bacilli.

To increase general activity in the area, pathways for non-inferiority must be opened in a way that is scientific, rigorous, clinically meaningful and feasible. All of those elements must be true. If the pathways are scientific and rigorous, but not feasible, or not clinically meaningful, they will not address the public health need.

Financial & competing interests disclosure
Brad Spellberg receives grant and contract support from the NIH, Gilead, Astellas, Novartis and Cubist. He also owns shares in NovaDigm Therapeutics and has received compensation from the following companies in respect of consultancy services: Pfizer, Baseia, The Medicines Company, Novartis, Achogen, Trius, GlaxoSmithKline, Eisai, Anacor, Zenech, Theravance, Meiji and Sevident. Brad Spellberg has no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

Bibliography

3 Silverman JA, Mortin LI, Vanpraagh AD, Li T, Alder J. Inhibition of daptomycin by pulmonar...