Published in final edited form as:

Clin Trials. 2013 ; 10(6): . doi:10.1177/1740774513507503.

Discordant MIC Analysis: A New Path to Licensure for Anti-infective Drugs

Dean Follmann¹, Erica Brittain¹, and John H. Powers²

¹DF, EB Biostatistics Research Branch National Institute of Allergy and Infectious Diseases

²JHP SAIC in support of the Division of Clinical Research National Institute of Allergy and Infectious Diseases

Summary

Background—Evaluation of anti-infective drugs for licensure often relies on a noninferiority (NI) design where new drug B is noninferior to comparator drug A if the difference in success rates is reliably not worse than some fixed margin. The margin must be based on historical studies that estimate the magnitude of the benefit of drug A over placebo. This approach hampers drug development because the obligatory evidence for margin determination is often nonexistent.

Purpose—To develop a new method for licensure of anti-infective drugs when there is no historical evidence for reliable construction of the NI margin.

Methods—The MIC measures the minimum amount of drug that it takes to inhibit growth of bacteria *in vitro*. Patients who are infected with bacteria that have a low MIC to a given drug are expected to have good outcome when treated with that drug. Thus a differential effect of drug B versus drug A, if it exists, is likely to occur in patients whose pathogens have discordant MICs (e.g. low MIC for drug B, high MIC for drug A, or vice versa). A new paradigm for licensure of anti-infective drugs is proposed where a clinically acceptable NI margin is selected and licensure supported if the NI margin is met and B is reliably demonstrated superior to A in a subset of patients whose paired MICs favor B. The requirement for some evidence of superiority encourages a study that is carefully designed and executed.

Results—Simulations indicate the approach shows promise in realistic settings provided adequate data are available. A simulated example illustrates use of the methods.

Limitations—If the data have small sample size, weak MIC/success relationship, or high correlation between MIC-A, MIC-B, this procedure will have poor power.

Conclusion—Discordant MIC analysis may offer a novel path to licensure for certain anti-infective drugs.

Keywords

AUC:MIC ratio; Interaction Test; Logistic Regression; Licensure

1 Introduction

Contemporary anti-infective drug trials typically involve the comparison of two active drugs using a noninferiority design with the primary endpoint defined as successful treatment of the infection by a specific time. Often, part of the design involves determination of the minimum inhibitory concentration (MIC) for each drug using a baseline sample of the pathogen that was extracted from the patient prior to randomization. The MIC measures the smallest amount of drug that is needed to inhibit growth of a sample pathogen in the test tube. Intuitively, the probability of clinical success in the patient should vary inversely with the effectiveness of the drug to the infecting organism as measured *in vitro* by the MIC. Many studies have demonstrated a relationship between success and the MIC, sometimes combined with pharmacometric parameters that describe the time course of drug in the blood[1-3].

While suggestive, a significant inverse relationship between MIC and success does not necessarily prove that the drug is better than a placebo. It could be that organisms with high MICs are inherently less fragile, or that sicker hospitalized patients get preferentially infected with high MIC organisms[4-7]. Such possibilities could explain the analyses of Holmes et al. [8]. Baseline samples on patients with *staphylococcus aureus* bacteremia were obtained and the MIC to the antibiotic drug flucloxacillin was determined. The patients were subsequently all treated with vancomycin and those with high MICs to flucloxacillin had worse outcomes than those with low MICs to flucloxacillin. Thus flucloxacillin MICs were not independent of patient characteristics and/or vancomycin MIC. Typically, the FDA requires randomized trials to license a new anti-infective drug, many of which are designed to show noninferiority (NI) of the new drug to a comparator[9,10]. Such NI trials require that there be historical evidence of the magnitude of the effect of the comparator drug versus placebo in a relevant setting to ensure that the comparator has benefit. Practically speaking, drug development is restricted to settings where the luck of history has provided such studies[11].

This paper proposes a new approach for the analysis of contemporary randomized trials of anti-infective drugs that blend the intuitive MIC information with the rigor of a randomized trial. Our thinking is similar to that of uplift modeling [12]. Patients with simultaneously low MICs to drug B and high MICs to drug A are most likely to show a benefit of drug B relative to drug A. We thus test for superiority of B to A in pre-selected patients with maximally discordant MICs favoring B. This type of analysis has the strength of evidence of a randomized trial, in sharp contrast to a one arm observational study of drug B which compares patients with high MICs to drug B to patients with low MICs to drug B. This analysis suggests a new paradigm where licensure is supported if the difference in success rates on A and B reliably exceeds a sensible margin based on a clinically acceptable loss of effect, and superiority of drug B to drug A in pre-selected patients is demonstrated.

The traditional NI design requires specification of a margin that ensures the following dual goals: 1) that the new drug has absolute efficacy (i.e. relative to placebo) and 2) that its efficacy is similar to existing therapy. Addressing the first goal requires reliable historical data about the magnitude of A's benefit over placebo in the context of the current trial. This magnitude is referred to as M1 in the FDA Guidance on NI designs [9]. Addressing the second goal requires a subjective assessment of the clinically acceptable loss in efficacy, termed M2. The NI margin is the minimum of M1 and M2. When there are relevant data to set M1, a NI design can meet the dual goal.

However, it is common in anti-infective settings to have little basis to establish a rigorous value for M1. That is, FDA will have approved a drug A, but this process may not provide

reliable evidence for the magnitude of the efficacy over placebo and M1 cannot be set. Our approach is proposed as an alternate pathway to licensure when M1 cannot be reliably estimated. We achieve the dual goal by requiring two tests. We do the traditional NI analysis using the M2 margin, that is, the subjective clinically acceptable difference. To address the other goal, that is, whether B has efficacy over placebo, we look to its advantage over A in the subpopulation where its advantage should be greatest. If we observe superiority of B over A, the licensed drug, we have indirect evidence that B is better than placebo as it beat a licensed drug in the current trial.

Our approach requires an untestable assumption, that drug A is no worse than placebo ($M1 > 0$). Untestable assumptions are undesirable, but it is worthwhile to put this into context. First, drug A is licensed so from the FDA's perspective, it is better than placebo, it is just that the exact M1 is hard to quantify. Second, the current approach effectively requires a different untestable assumption that the M1 that was estimated in the past is an accurate estimate of the advantage of drug A over placebo in the current trial. These past studies may be observational, have different (e.g. less hardy) pathogens, different types of patients, and a different standard of care. We feel that in many settings, our assumption that M1 is greater than zero will be acceptable in the difficult world of anti-infective drug development.

2 Discordant MIC Analysis

Consider a randomized trial of two anti-infective drugs termed A (i.e. comparator) and B (i.e. investigational), respectively. Prior to randomization, sample pathogens are obtained from each patient and drug susceptibility tests are performed for both drugs resulting in baseline variables MIC-A and MIC-B for each patient. This determination generally cannot be made until a few days after randomization/drug initiation as it takes time to determine the MICs and patients often receive therapy immediately after randomization. Such drug susceptibility testing is routine in clinical trials and in clinical practice. Once randomized, patients should receive their intended therapy for a specified period of time. At some fixed time post-randomization, the primary endpoint of clinical success is determined.

In clinical trials it is common to perform subgroup analysis where randomized volunteers are grouped on the basis of baseline characteristics such as type of baseline pathogen or history of prior antibiotic use. Subgroup analyses are typically undertaken to confirm that the drug effects are consistent across different subpopulations, or to identify a gradient of benefit along a baseline variable[13,14].

Intuitively, the two drugs should produce similar outcomes for patients with concordant values for MIC-A and MIC-B. Conversely, the two drugs should produce dissimilar outcomes in patients for whom MIC-A and MIC-B are discordant. The idea of this paper is to use both MICs together to construct subgroups where we are most likely to detect differential drug effects. For illustration, suppose we categorize patients based on whether their MICs for each drug lie above or below the median MIC for each drug. The four classes of patients can be randomized and their success rate on each drug estimated. This approach is illustrated in Figure 1 and Table 1.

The lower right cell of Table 1 is comprised of patients with a pathogen that has a low MIC for drug B and a high MIC for drug A. Such low/high patients have a success rate of 90% if randomized to drug B and a success rate of 70% if randomized to drug A. In the upper right cell, with concordantly poor MICs, we see a 70% success rate for both Drug A and Drug B. Table 1 suggests that even if drug B is equivalent or not better than drug A overall, we may be able to demonstrate superiority of B to A in the likeliest subgroup.

Another possibility for success rates is given in Table 2. Here the MICs identify patients with differential prognosis independent of the drug used, but for every patient, drugs A and B have the same success rate. Such a result could obtain if the MIC identified the inherent fragility of the pathogen, or if sicker patients tended to get pathogens with higher MICs. This possibility was suggested by the analysis of Holmes et al. [8]. Note that both Tables 1 and 2 have average success rates on drug B of 70% and 90% for high and low drug B MICs respectively. Obviously having higher success rates on drug B for low MIC-B compared to high MIC-B cannot tell us if drug B has clinical benefit. Indeed in Table 2, drug A and drug B might be no better than placebo in terms of success.

The categories used in Tables 1 and 2 are made from baseline variables and thus each cell of the table defines a valid subgroup for which hypotheses are pre-specified. Thus hidden within an ordinary noninferiority trial are precious subtrials well placed to show superiority. An intriguing analysis would be to test for a statistically significant subgroup effect in the discordant cells.

This suggests a new paradigm for NI trials of anti-infective drugs to demonstrate efficacy. Current FDA guidance requires that the margin for the comparison of B to A be based on historical evidence of the magnitude (M1) of the superiority of A compared to placebo[9], e.g., a previous study where the success rate of A was reliably 20% larger than placebo. The NI margin can be no larger than this magnitude. Then a smaller margin (M2) is often set based on clinical judgment about how much worse B could be compared to A (e.g., at most 10% worse, preserving half the advantage). That is, the goal of the current NI approach is two-fold: i) indirectly demonstrate B is better than placebo and ii) demonstrate that B is not clinically unacceptably worse than A. Without such historical evidence there is no path forward. However, the approach described in this paper could meet these two goals without external data. One could clinically choose a margin (M2) of, say, 10% and then obtain licensure if this margin were met for the difference in overall success rates and the *pre-selected* subgroup showed superiority of B to A. This latter condition would be used as evidence that B is better than placebo. Strictly speaking, superiority is identified only in a subgroup with discordant MICs, and one must assume drug A is not worse than placebo. But it is direct evidence of superiority based on a randomized trial of contemporary patients in exactly the subgroup where we should have the best ability to detect superiority. The current practice of demonstrating drug B is superior to placebo, when possible, is indirect and may rely on old nonrandomized studies with different pathogens, patients, and standard of care from the current NI trial. An important consequence of the ability to show superiority is that it also provides evidence the trial was well-run and sensitive enough to show benefit, (i.e. it has so-called assay sensitivity)[9].

A key parameter in our approach is the specification of the clinically acceptable margin or M2. Anti-infective drug trials often use a margin of .10 for the difference in success rates where success rates of 80% or greater are expected. While there has been some discussion of using a margin based on relative risk, this is rare in the anti-infective world. When using our method, one might use .10 as a subjective margin of what loss in efficacy could be tolerated, but preferably the disease context would be considered for a more nuanced approach.

As in any clinical trial, it is prudent to examine for consistency of the treatment effect across different subgroups. Logically, one would also want to test for superiority of drug A over drug B in the upper left subgroup of Table 1 in addition to the test in the lower right subgroup. The results of such subgroup testing could provide important nuance to the overall pronouncement of noninferiority of drug B over drug A and might result in specific language in the package insert for the licensed drug. Importantly, such labeling would only

be relevant for infections where one could delay initiating treatment until after the MIC readouts were developed. Many infections are acute and delay is unacceptable.

3 Discordant Regression Analysis

While transparent, categorizing patient into groups based on the median MIC may often result in a very underpowered analyses with few patients in the discordant cells. We are thus led to develop a more efficient statistical approach that exploits the conceptual framework of the previous section. We assume that the probability of success follows a logistic regression model where the effect of MIC differs in each drug group:

$$\text{logit} \{P(\text{success})\} = \theta_0 + \theta_1 Z + \theta_2 \text{MIC} - A + \theta_3 \text{MIC} - B + \theta_4 \text{MIC} - AZ + \theta_5 \text{MIC} - BZ, \quad (1)$$

where $Z = 1(0)$ for patients randomized to drug B (drug A) and $\text{logit}(a) = \log\{a/(1-a)\}$. The parameters are statistically estimated using trial data.

If $\theta_4 = \theta_5 = 0$ then, as in Table 2, we have success curves that depend on MIC-A, MIC-B in exactly the same way in both drug groups and the performance of drug B relative to drug A is unaffected by MIC-A or MIC-B. This relationship suggests that low MIC is merely a proxy for pathogen fragility or intrinsic patient health. If $\theta_4 > 0$ and $\theta_5 < 0$ we have a smoothed version of the relationship in Table 1 where the performance of drug B relative to drug A increases with both larger MIC-A and smaller MIC-B. A model of conditional independence obtains when the log of the odds for success in B divided by the odds for success on A is $\theta_1 + \theta_4 \text{MIC} - A + \theta_5 \text{MIC} - B$. Thus only MIC-z impacts success on drug z and that $\theta_4 = \theta_5$. Other parsimonious parameterizations of (1) may provide more efficient inference if known parameter constraints are incorporated.

A test of interaction, say the likelihood ratio test of $\theta_4 = \theta_5 = 0$ can reveal whether the effect of drug B relative to drug A varies with MIC-A, MIC-B. While interesting, this can be significant even if drug B is worse than drug A for all MIC-A, MIC-B. A different tack is required to identify if drug B is superior to drug A for some subpopulation. An appealing test for superiority of B versus A in a small subpopulation is to identify the (MIC-A, MIC-B) point that is maximally discordant in favor of B and then see if the model based test favors B at that point. This provides direct evidence of the drug's clinical benefit which is generally not possible with either a noninferiority trial or with an observational study.

More specifically, we identify the maximally discordant point by identifying, among the observed (MIC-A, MIC-B)s, the one that is closest in terms of Euclidean distance to the point $(\max(\text{MIC-A}), 0)$, where $\max(\text{MIC-A})$ is the largest observed value for MIC-A. Call this point (a, b) . As a and b are selected without looking at outcome data, we can regard them as fixed constants. To formally test superiority we form

$$T(a, b) = \frac{\hat{\theta}_1 + \hat{\theta}_4 a + \hat{\theta}_5 b}{\sqrt{\widehat{\text{var}}(\hat{\theta}_1 + \hat{\theta}_4 a + \hat{\theta}_5 b)}}. \quad (2)$$

This test statistic can be used for different specific hypothesis tests. Asymptotically, $T(a, b)$ has a standard normal distribution under the null hypothesis that model (1) holds with $H_0: \theta_1 = \theta_4 = \theta_5 = 0$. So for large $T(a, b)$ we reject the null hypothesis that drug A and drug B follow the same logistic regression model in favor of drug B being superior to drug A at the point (a, b) . One can also use a permutation null distribution where we repeatedly calculate $T(a, b)$ for all permutations of the vector of drug indicators. Formally this approach tests a less restrictive null hypothesis that $p_A(\text{MIC-A}, \text{MIC-B}) = p_B(\text{MIC-A}, \text{MIC-B})$, where

$p_Z(\text{MIC-A}, \text{MIC-B})$ is the probability of success on drug Z as a function of MIC-A, MIC-B. In other words, the permutation null does not require that $p_Z(\text{MIC-A}, \text{MIC-B})$ follow a logistic regression form. Of course, it is essential that the explicit test for superiority be specified *a priori* so the type I error rate is rigorously controlled.

4 AUC:MIC ratio

A measure of *in vivo* drug exposure such as the area under the curve of drug concentration (AUC) when coupled with the MIC (e.g. AUC:MIC ratio) can provide a stronger relationship with success than the MIC itself[1]. It is tempting to use AUC:MIC in lieu of MIC in the methods of this paper. However for patients randomized to drug A, AUC-B is not obtained. One way to proceed is to cross-over patients a bit after the end of therapy to the opposite drug and discontinue drug exposure once the AUC to the opposite drug was measured. Even so, the AUCs are measured post-baseline and trial-quality inference requires that the AUCs behave like baseline variables; that is, for each individual, the AUC for drug B measured *after* treatment of infection by drug A has about the same value as the AUC that would have been obtained during *initial* treatment of infection by drug B. This might not hold if, for example, the infectious process modified kidney function and drug metabolism. A better approach might be to obtain AUCs for both A and B after their initial therapy is completed. AUCs measured in steady state well after initial therapy might be similar in both randomized groups. Still, the use of AUC:MIC in place of MIC may lead to bias because it is measured post-baseline and the bivariate distribution of (AUC-A:MIC-A, AUC-B:MIC-B) is not assured to be the same in the two arms.

A different tack that eliminates concern about bias would be to predict AUC using baseline variables such as age, weight, race, or measures of kidney function. For drug A we could postulate a linear regression:

$$\text{AUC} - A = \mathbf{X}'_A \theta_A + e$$

where e is normal ($0, \sigma^2$). We could then estimate θ_A and also θ_B from the analogous model for AUC-B. For each patient we could then form the predicted AUCs: $\mathbf{X}'_A \hat{\theta}_A, \mathbf{X}'_B \hat{\theta}_B$ and then create the ratios $\mathbf{X}'_A \hat{\theta}_A / \text{MIC} - A, \mathbf{X}'_B \hat{\theta}_B / \text{MIC} - B$. These ratios are essentially a pair of (complicated) baseline variables and could be used in lieu of the MICs in (1). Both the crossover and prediction maneuvers have been proposed for use in vaccine trials in order to recover the immune response to the vaccine, a variable that is not directly measurable in the placebo arm [14], rather akin to not directly measuring AUC-A in arm B and *vice versa*. While MICs are typically collected in clinical trials and practice, AUCs are generally only collected in specialized studies. AUC based approaches would require either augmentation of standard anti-infective drug clinical trials, or the development of prediction models based on external data.

5 Evaluation

To explore performance of the discordant regression method in practice, simulations were conducted using the test $T(a, b)$ of (2) for a data set of two anti-infective drugs with 200 patients randomized per arm with success as the primary endpoint. In practice this data might be the result of several clinical trials that form part of the evidence required for licensure. The distribution of MIC-A, MIC-B was bivariate normal with a common mean of 1.63 and a common standard deviation of .45. These (equal) marginal distributions approximate the distribution of vancomycin MICs as measured by Etest in g/mL among 532

patients with *Staphylococcus aureus* bacteriemia as seen in [8]. We varied the correlation between the two MIC-s from 0 to .9.

For the “Modest” scenario, successes were generated using $(\mu_0, \mu_1, \mu_2, \mu_3, \mu_4, \mu_5) = (3.68, 0, -1.24, 0, +1.24, -1.24)$, this was meant to approximate a moderate relationship between MIC and success as seen in Holmes et al. [8]. The overall success probability was set at .82. For the “Strong” scenario, (μ_2, μ_4, μ_5) were 50% larger which was meant to approximate a stronger relationship that might be seen with use of the AUC:MIC ratio instead of the MIC. Again the overall success probability was set at .82 leading to $(\mu_0, \mu_1, \mu_2, \mu_3, \mu_4, \mu_5) = (4.80, 0, -1.86, 0, +1.86, -1.86)$. Figure 2 below graphs the (marginal) probability of success on drug as a function of MIC in the modest and strong scenarios. These curves apply to both drug A and B.

For each setting, 10,000 clinical trials were simulated. Each trial had 200 patients randomized to Drug A and 200 patients randomized to Drug B. For each simulated trial we performed a test of superiority at the identified maximally discordant (MIC-A, MIC-B) point that favored B. That is, for each dataset, we identified the (MIC-A, MIC-B) point that was closest in Euclidean distance to $(\max(\text{MIC-A}), 0)$, where $\max(\text{MIC-A})$ was the largest MIC-A value over all 400 patients. We reject if $T(a; b)$ exceeds a critical value of 1.96.

Table 3 presents the results. The power increases substantially with decreasing correlation, and for the strong scenarios compared to the weak. Of course, different settings may have different power profiles. Remedies for low power include larger trials, combining multiple trials of the same drugs, the use of a more sensitive endpoint such as the average patient reported outcome over a fixed period of time, recruiting sufficient patients with discordant MICs, having drug B be at least slightly better than drug A, or the construction of more sensitive statistical procedures. In practice, one would be interested in the overall probability of both concluding superiority with $T(a, b)$ and passing the NI margin. Evaluation would be context specific and depend on factors such as the chosen margin (e.g. 10%) and whether one or more trials formed the data for (2).

For completeness, we evaluated the type I error rate under different null settings where $\mu_1 = \mu_4 = \mu_5 = 0$. We evaluated 4 scenarios $(\mu_2, \mu_3) = (0, 0), (-1.24, -1.24), (-1.24, -1.86)$, and $(-1.86, -1.86)$ with μ_0 chosen so the overall success rate was .82. Table 4 provides simulated rejection probabilities, P , each based on 100,000 simulated trials. For all scenarios, the lower confidence interval $P - 1.96 \sqrt{P(1-P)/100000}$ is below .025.

Some specific cases are worth considering. In some settings drug B will be chemically similar to drug A. Here, the MICs to A and B might be similar and the methods of this paper might not be very useful. The most promising setting is where the MICs do not have strong correlation, perhaps when drug A and B have quite different mechanisms of action.

Suppose that *a priori* one agreed to use the new paradigm for licensure and selected a clinically acceptable NI margin of .10 based on an anticipated success rate of about .80. To illustrate how to proceed, we selected a typical simulated trial from the simulations of Table 3 with $\rho = .5$. Figure 3 presents a scatterplot of the data from this trial. Overall the A and B success rates were .83 and .82 with a 95% confidence interval for the B minus A difference $(-.08, .07)$. This lies entirely to the right of the margin of $-.10$ so that the goal of demonstrating B is not meaningfully worse than A is met. To test for evidence of superiority, we identify the maximally discordant (MIC-A, MIC-B) = (2.37, 1.18) value that should favor B, the boxed point in Figure 3. The test statistic is given by

$$\frac{\hat{\theta}_1 + \hat{\theta}_4 2.37 + \hat{\theta}_5 1.18}{\sqrt{\widehat{\text{var}}(\hat{\theta}_1 + \hat{\theta}_4 2.37 + \hat{\theta}_5 1.18)}} = 2.66$$

which has a p-value of .0078. Thus this trial meets the proposed dual requirements for licensure.

Once a drug passes the requirements for a NI trial, it is prudent to examine for homogeneity of treatment effect. For anti-infective drugs it is especially relevant to look across different regions defined by MIC-A, MIC-B. Figure 3 thus also provides curves[15] identifying the (a,b) points where the associated pointwise confidence interval favors A (upper left region) favors B (lower right) and favors neither (the middle). These require a correctly specified model for reliable interpretation, the global statistical fit of the model should be evaluated, and multiplicity might be an additional concern as these are pointwise confidence intervals. Nonetheless, if regions of clear superiority were reliably identified, it might make sense for this information to be incorporated in the package labeling to optimize drug use. Importantly, such labeling would only be useful for chronic infections such as tuberculosis where it would be ethical to delay initiation of therapy for several days while the MIC read-outs were developed. Most infections are acute and typically therapy will need to be started before the MIC values are known. Thus often, the regions of Figure 3 will not be clinically useful for choice of initial drug.

6 Discussion

The evidence necessary for the design and analysis of NI trials of anti-infective drugs is often absent or of questionable relevance to modern clinical settings. This requirement has stymied the development of new anti-infective drugs. This paper has proposed a new paradigm for the design of NI trials for licensure where historical evidence of the advantage of drug A to placebo is shaky or absent. Evidence for licensure is obtained if the difference in overall success rates for drug B minus drug A reliably exceeds a clinically acceptable margin and there is rigorous evidence of superiority of drug B over drug A at a pre-specified point or region. Effectively this is a potential statistical solution to the conundrum of bad bugs and no drugs [9]. These methods offer a novel path for licensure and address the true effect of MIC on clinical success. They should be applied to extant data and considered for future trial design.

A November 2012 FDA Advisory Committee (AC) meeting illustrates the feasibility of finding superiority in a subgroup, though using marginal MIC subgroups rather than discordant MIC subgroups. Two pivotal NI studies comparing televancin to vancomycin in the treatment of nosocomial pneumonia were presented [17]. The reason for the AC meeting was because after the studies were completed it was felt there was no reliable evidence for the magnitude of M1. Thus even though the pre-specified NI margins were met for both studies, it was unclear that televancin was better than placebo. The AC met to offer advice on this difficult situation. Combining the two pivotal studies, the sponsor presented results showing numerical superiority of televancin to vancomycin on a clinical response endpoint at a test of cure visit within the subgroup of patients with baseline *Staphylococcus aureus* who also had vancomycin MIC values greater than or equal to one (Table 19: 74/85 versus 78/105, $p=.030$ with Fisher's exact test). Under our proposal, if the latter subgroup had been pre-specified as the region for a test of superiority, it could have eliminated the requirement for the historical justification for a NI margin. Superiority in the subgroup coupled with

meeting the clinically acceptable NI margins for the two pivotal studies would have supported licensure under our proposed criteria.

Acknowledgments

FUNDING This work was supported by the NIH. The authors are either employees or contractors of the NIAID.

This work was conducted as part of the official duties of the authors who are either NIH employees (DF EB) or NIH contractors (JHP).

REFERENCES

1. Craig WA. Pharmacokinetic/Pharmacodynamic Parameters: Rationale for Antibacterial Dosing of Mice and Men. *Clinical Infectious Diseases*. 1998; 26(1):1–10. [PubMed: 9455502]
2. Ambrose PG, Hammel JP, Bhavnani SM, Rubino CM, Ellis-Grosse EJ, Drusano GL. Frequentist and Bayesian Pharmacometric-Based Approaches To Facilitate Critically Needed New Antibiotic Development: Overcoming Lies, Damn Lies, and Statistics. *Antimicrobial Agents and Chemotherapy*. 2012; 56(3):1466–1470. [PubMed: 22155834]
3. Gehanno P, Lenoir G, Berche P. In Vivo Correlates for *Streptococcus pneumoniae* Penicillin Resistance in Acute Otitis Media. *Antimicrobial Agents and Chemotherapy*. 1995; 39(1):271–272. [PubMed: 7695322]
4. Holland TL, Fowler VG. Vancomycin minimum inhibitory concentration and outcome in patients with *staphylococcus aureus* bacteremia: pearl or pellet? *J Infectious Diseases*. 2011; 204:329–331. [PubMed: 21742827]
5. Lambert ML, Suetens C, Savey A, Palomar M, Hiesmayr M, Morales I, Agodi A, Frank U, Mertens K, Schumacher M, Wolkewitz M. Clinical outcomes of health-care-associated infections and antimicrobial resistance in patients admitted to European intensive-care units: a cohort study. *Lancet Infectious Diseases*. 2011; 11:30–38. [PubMed: 21126917]
6. Shorr AF, Zilberberg MD, Reichley R, Kan J, Hoban A, Hoffman J, Micek ST, Kollef MH. Validation of a clinical score for assessing the risk of resistant pathogens in patients with pneumonia presenting to the emergency department. *Clinical Infectious Diseases*. 2012; 54(2):193–198. [PubMed: 22109951]
7. Shorr AF, Zilberberg MD, Micek ST, Kollef MH. Prediction of infection due to antibiotic-resistant bacteria by select risk factors for health care-associated pneumonia. emergency department. *Archives of Internal Medicine*. 2008; 168(20):2205–2210. [PubMed: 19001196]
8. Holmes NE, Turnidge JD, Munckhof WJ, Robinson JO, Korman TM, O'Sullivan MVN, Anderson TL, Roberts SA, Gao W, Christiansen KJ, Coombs GW, Johnson PDR, Howden BP. Antibiotic choice may not explain poorer outcomes in patients with *staphylococcus aureus* bacteremia and high vancomycin minimum inhibitory concentrations. *J Infectious Diseases*. 2011; 204:340–347. [PubMed: 21742831]
9. FDA Guidance for Industry Noninferiority Trials. 2010. www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/10
10. FDA Guidance for Industry Antibacterial Drug Products: Use of Noninferiority Trials to Support Approval. 2010. www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/11
11. IDSA Bad bugs, no drugs: as antibiotic discovery stagnates, a public health crisis brews 2004 IDSA White Paper.
12. Jaskowski, M.; Jaroszewicz, S. Uplift Modeling for Clinical Trial Data; ICML 2012 Workshop on Machine Learning for Clinical Data Analysis; Edinburgh, Scotland. June, 2012; 2012.
13. Follmann, DA. Subgroups and interactions. In: Geller, N., editor. *Advances in Clinical Trial Biostatistics*. Marcel Dekker; 2004.
14. Yusuf S, Wittes J, Probstfield J, Tyroler HA. Analysis and interpretation of treatment effects in subgroups of patients in randomized clinical trials. *Journal of the American Medical Association*. 1991; 266(1):93–98. [PubMed: 2046134]
15. Follmann DA. Augmented designs to assess Design for Vaccine Trials. *Biometrics*. 2006; 62:1161–1169. [PubMed: 17156291]

16. Shuster J, van Eys J. Interaction between prognostic factors and treatment Controlled Clinical. Trials. 1983; 4:209–214.
17. FDA Anti-infective Drugs Advisory Committee Meeting; November 2012; Sponsors briefing document: <http://www.fda.gov/downloads/AdvisoryCommittees/CommitteesMeetingMaterials/InfectiveDrugsAdvisoryCommittee/UCM329482.pdf>

Figure 1.
Schematic of the effect of randomization on a patient population characterized by *in vitro* susceptibility testing of their pathogens to Drugs A and B.

Figure 2.

Visual representation of key features of the simulation parameters used in Table 3. The marginal probability of success following treatment with drug as a function of MIC under the modest scenario is given by the dashed curve; the strong scenario is given by the solid curve. The bottom ticks represent 200 randomly generated MIC values from the assumed Gaussian distribution. The curves apply to both drug A and drug B.

Figure 3.

Data from a simulated randomized clinical trial with 200 patients per arm. The dotted line shows where both drugs have the same estimated success probability. Blue dots identify the MIC pairs for patients randomized to drug B, while red dots are analogous for drug A. Patients who failed are denoted by a solid dot. The maximally discordant (MIC-A, MIC-B) that should favor drug B is identified by a box. The curves define regions of MIC points where the confidence interval at that point reliably favors drug B (lower right) favors drug A (upper left) or neither (the remainder).

Table 1

Drug B: Drug A success percentages as a function of the MICs for Drug A and Drug B in an idealized trial. Bolded entries denote Drug B, italicized entries denote Drug A. The Drug B minus Drug A success percentages are given in the bottom of each cell. For illustration, the entries based on averaging assume each cell has $\frac{1}{4}$ of the patients

		<i>MIC for Drug A</i>		
		<i>Low</i>	<i>High</i>	<i>Average</i>
MIC for	High	70 : 90	70: 70	70: 80
		-20	0	-10
Drug B	Low	90: 90	90: 70	90: 80
		0	+20	+ 10
Average		80: 90	80: 70	80: 80
		-10	+ 10	0

Table 2

Drug B: Drug A success percentages as a function of the MICs for Drug A and Drug B in an idealized trial. Bolded entries denote Drug B, italicized entries denote Drug A. The Drug B minus Drug A success percentages are given in the bottom of each cell. For illustration, the entries based on averaging assume each cell has $\frac{1}{4}$ of the patients

		<i>MIC for Drug A</i>		
		<i>Low</i>	<i>High</i>	<i>Average</i>
MIC for	High	80:80	60:60	70:70
		0	0	0
Drug B	Low	100:100	80:80	90:90
		0	0	0
Average		90:90	70:70	80:80
		0	0	0

Table 3

Simulated probabilities of correctly concluding Drug B is superior to Drug A at the maximally discordant (MIC-A, MIC-B) using model (1). Each arm has 200 patients with evaluable MICs. The correlation between MIC-A and MIC-B is ρ , and the strength of the relationship between MIC and success is similar to Holmes et al. [8] (Modest) or 50% greater (Strong)

Simulated Probability of Correctly Declaring Superiority of Drug B over Drug A When MIC/Success Relationship is			
	Modest for A&B	Strong for B, Modest for A	Strong for A&B
.0	.74	.91	.96
.3	.59	.82	.88
.5	.47	.69	.76
.7	.30	.49	.54
.9	.12	.23	.22

Table 4

Simulated null probabilities of concluding Drug B is superior to Drug A at the maximally discordant (MIC-A,MIC-B) using model (1). Each arm has 200 patients with evaluable MICs. The correlation between MIC-A and MIC-B is p , and the strength of the relationship between (MIC-A,MIC-B) and success varies from nil to similar to Holmes et al. [8] (Modest) to 50% greater (Strong). The effect of MIC-A,MIC-B on success is identical for drug A and drug B for these null scenarios

p	Simulated Probability of Incorrectly Declaring Superiority of Drug B over Drug A When MIC/Success Relationship is			
	Nil	Modest for A&B	Strong for B, Modest for A	Strong for A&B
.0	.0240	.0244	.0242	.0241
.3	.0250	.0251	.0248	.0247
.5	.0249	.0247	.0242	.0250
.7	.0253	.0253	.0250	.0246
.9	.0253	.0236	.0236	.0236