Recommendations to the FDA for Interim Endpoints for Clinical Trials in Acute Bacterial Skin and Skin Structure Infections

Foundation for the National Institutes of Health Biomarkers Consortium Project Team

ABSSSI Docket ID: FDA-2010-D-0433

Contents

0 Executive Summary

During recent decades, the efficacy endpoints for Acute Bacterial Skin and Skin Structure Infections (ABSSSI) registrational studies relied on a clinical assessment of cure requiring "complete resolution of signs and symptoms" based on a combination of non-standardized, physician-based observations and comments collected from the patient by the physician as well as on the investigator's assessment of the need for alternative antibiotic therapy. As non-inferiority clinical trial design advanced during the late 20th and early 21st century, it became apparent to the FDA and others that the development of more readily quantifiable, reproducible, and externally verifiable endpoints would improve the design of present-day non-inferiority clinical trials for ABSSSI.

In developing updated approaches to endpoints, it also was recognized that outcome measures used for studies that support drug registration for ABSSSI must be relevant for clinical practice. Although the level of detail and accuracy in measurement needed in the setting of clinical trials may differ from that needed in clinical practice, a description of the pivotal (Phase 3 or registrational) clinical trials as conducted is an integral part of the prescribing information and must be based directly on the trial data as collected and analyzed. The choice of primary endpoint for a trial may thus need to balance a variety of purposes.

In parallel discussions of the design of studies for Community-acquired Bacterial Pneumonia (CABP), the idea arose that standardized assessments of patient response in ABSSSI in the first few days of therapy might provide key insights into both drug effect and options for trial design (FDA 2010, August). Consequently, and at the request of the FDA, in early May 2010 the Foundation for the National Institutes of Health (FNIH) convened a Project Team for a Biomarkers Consortium Project, entitled "Developing Endpoints for Clinical Trials of Drugs for Treatment of Acute Bacterial Skin and Skin Structure Infections and Community-Acquired Bacterial Pneumonia (Phases 1 and 2)." The Project Team membership included broad participation from the NIH, the FDA, the academic research community (including members of the Infectious Diseases Society of America [IDSA]), and interested biopharmaceutical companies.

This document summarizes the work of the Biomarkers Consortium Project Team. Over a series of meetings, the group reviewed the available historical and modern data and found that control of lesion spread at 48 to 72 hours after randomization was sufficiently well documented that an early response endpoint measure could be proposed. To assess sustained response and other late events, supportive information should be obtained by assessing outcomes at a fixed time point after therapy has been completed. Such information could include a late response endpoint similar to the traditional test-of-cure (TOC) endpoint but more clearly defined. Although incompletely validated under the proposed conditions of use and requiring further research, an early response endpoint can be used to anchor a non-inferiority hypothesis in a trial for this indication. Thus, the Project Team supports a primary endpoint based on early response in review of registrational trials and approval of applications in ABSSSI while further research into outcomes at later time points in this area is conducted.

1 Introduction/Background

1.1 Introduction

Acute Bacterial Skin and Skin Structure Infections (ABSSSI) are common infections most often caused by *Staphylococcus aureus* (approximately 80%) and *Streptococcus pyogenes* (approximately 12%). The recent epidemiological shift to a greater proportion of ABSSSI caused by methicillin-resistant *S. aureus* (MRSA) acquired in the community—approximately 59% in one recent survey (Moran et al. 2006)—has been a cause for concern due to the limited number of safe and effective orally available antibacterial drugs for treatment of disease due to MRSA.

New antibacterial agents for ABSSSI are thus important. In addition, new agents with Gram-positive activity are often initially tested and FDA-approved for use in skin and soft tissue infections prior to testing in more severe infections such as pneumonia. Over the past several decades, the efficacy endpoints for ABSSSI registration studies were determined by resolution of signs and symptoms of the infection at a time point after completion of the antibacterial drug therapy. As understanding of the principles underpinning conduct of non-inferiority clinical trials advanced during the late 20th and early 21st century, it became apparent to the FDA and others that the design of present-day non-inferiority clinical trials for ABSSSI would be improved by better defining outcome measures, reducing the dependence on subjective elements of endpoints to make outcome measures more reliable, and choosing the timing of outcomes at a point where prior evidences shows a reliable and reproducible drug effect to justify the use of the non-inferiority study design. It also became apparent to the FDA that to ensure the constancy of the effect of antimicrobials from prior trials and across current trials, a standard for measurement is needed. In developing updated approaches to endpoints, it was also recognized that outcome measures used for studies that support drug registration for ABSSSI must also be relevant for clinical practice. A description of the pivotal (Phase 3) clinical trials is an integral part of the prescribing information and is based directly on the trial data as collected and analyzed. Thus, the choice of primary endpoint for a trial needs to provide information to meet a range of purposes.

These considerations led to the publication of a new FDA Guidance for studies in ABSSSI (FDA 2010, August). This Guidance includes a focus on assessment of efficacy at the earlier time points. The FDA authors note that published data from the 1930s and 1940s provide historical evidence of substantial treatment effects at time points earlier than those used in recent non-inferiority trials. Specifically, the guidance points to historical data showing convincing and reliable antibacterial drug treatment effects early in the course of treatment at 48 to 72 hours after the initiation of antibiotic therapy for ABSSSI. This observation is of importance as a known treatment effect size on a well-defined and reliable outcome measure at a specific time point is essential for a non-inferiority trial design. However, the historical evidence is limited in that (1) the data are incomplete and cannot be audited; (2) the outcomes measured included recordings of body temperature, pulse, respiratory rate, and other clinical responses that are considered today to be biomarkers; and (3) questions exist as to whether the constancy assumption required for non-inferiority trial design can be met.

Consequently, and at the request of the FDA, in early May 2010, the FNIH convened a Project Team with broad participation from the NIH, the FDA, the academic research community (including members of the IDSA), and interested biopharmaceutical companies to address these issues. The conclusions described within this document represent the work of the Project Team. Over a series of meetings, the group reviewed the historical literature and relevant recent publications, reviewed data from several available modern clinical studies, and discussed the merits and limitations of most of the well-known endpoints in the literature. The Project Team found that it was possible to suggest new endpoints using these data but that these endpoints were incompletely evaluated and further research is needed. Not all members of the team agreed that these novel endpoints should be considered for use as primary endpoints in ABSSSI non-inferiority trials, but the Project Team did agree that the new endpoint provided clarity regarding an objective way to describe the early assessment of clinical response that has been an integral part of the investigator-based decision to continue study therapy.

Given the urgency of the situation with respect to new antibiotic development, the Project Team is submitting this report containing its recommendations to the FDA regarding the definition and timing of interim or "bridging" primary endpoints suggested for immediate use as primary endpoints in registrational clinical trial protocols with the potential to support approval of applications in ABSSSI, concomitant with exploratory research to develop improved endpoints.

This initiative is particularly important at a time when the incidence of treatment-resistant pathogens such as MRSA and many Gram-negative bacilli is increasing, as has been highlighted by the IDSA (Boucher et al. 2009).

1.2 Regulatory Background

In the United States the regulatory standard for approval of drugs is "substantial evidence" from "adequate and well-controlled trials." One of the requirements for adequate and well-controlled trials is that the methods of the assessment of subjects' response must be "well-defined and reliable." U.S. regulations require three components for well-defined and reliable outcomes measures (21CFR314.126(b)(6)): The study should explain (1) the variables measured (what to measure), (2) the methods of observation (how the outcome was measured), and (3) the criteria used to assess response (how the data are analyzed to define a meaningful outcome; e.g., what defines a "responder").

U.S. regulations as well as a recent Institute of Medicine document (Micheel and Ball 2010) point out that a "clinical" outcome measure is a direct measure of how patients feel, function, or survive (21CFR314.500, Subpart H). However, when appropriately developed and evaluated, surrogate endpoints may provide valuable insight and can also be used as outcome measures to support drug approval provided a clinical trial shows that "the drug product has an effect on a surrogate endpoint that is reasonably likely, based on epidemiologic, therapeutic, pathophysiologic, or other evidence, to predict clinical benefit" (21CFR314.510) and followup studies show the benefit on the surrogate reflects the benefit on direct patient-centered outcomes. "Surrogate endpoints" are biomarkers, such as signs of disease, radiological tests, cultures, and laboratory values, that act as substitute and indirect measures for how patients feel, function, or survive. In the case of trials of anti-infective agents, acute symptoms of an infection (warmth,

chills, pain), objective local findings (erythema, swelling), and selected markers of systemic response (leukocytosis, shift in white blood cell population distribution, hypotension) all have a well-established mechanistic link to the triggering infection (triggering of innate immunity cascades by microbial surface components (Dieffenbach, Tramont et al., 2010), release of cytokines by leukocytes activate in response to microbial pathogens (Nauseef and Clark, 2010). Although changes in these surrogate endpoints may have other causes, physicians rely on these measures in clinical practice. The group agreed that in the specific case of ABSSSI there is an observed linkage between control of lesion spread and patient-centered outcomes such as pain. The group agreed that evaluation of the relationship of various outcomes in ABSSSI is a key focus of future planned research.

In the setting of non-inferiority trials, U.S. regulations (21CFR214.126(b)(2)(iv)) point out that if the intent of the trial is to show similarity of the test and control drugs (non-inferiority), the report of the study should assess the ability of the study to have detected a difference between treatments. The analysis of the study should explain why the drugs should be considered effective in the study; for example, by reference to results in previous placebo-controlled studies of the active control drug (21CFR314.126(b)(2)(iv)). To assess the "assay sensitivity" of a trial to detect differences in the setting of non-inferiority trials, international guidance (International Conference on Harmonization Efficacy Guidelines E-9 and E-10) as well as recent FDA general and anti-infective guidances point out that the planned non-inferiority trial should have similar definitions of disease, outcomes measures, and timing of outcome measures to those used to show the effect of the control drug in prior superiority trials. The effect of the control drug compared to placebo or no specific treatment should be quantifiable, as well as reliably and reproducibly shown in prior studies.

2 Summary of Project Team Process

The members of the Project Team convened a series of meetings in May and June 2010 and January and July 2011. Over the course of these meetings, the group discussed the historical literature, recent publications on measurement of skin lesions, and data from several available modern clinical studies. The group reached consensus on a process to identify primary and secondary endpoints for ABSSSI and CABP (the latter discussed separately).

2.1 Review of Historical Evidence and Medical Literature

The past evidence consisted of two trials published in 1937 in the British Medical Journal by Snodgrass and Anderson. These trials compared sulfa drugs to ultraviolet light in patients with erysipelas. The studies were not randomized (used alternate assignment) and not blinded. The investigators evaluated a number of outcome measures including cessation of lesion spread, duration of pyrexia, duration of toxemia, death, relapse, and complications. The greatest treatment effect upon which to base a non-inferiority trial was on cessation of lesion spread at 48 hours. A meta-analysis of the treatment effect of the combined data from the two studies on cessation of spread at 48 hours showed an overall difference of 24% in favor of sulfa drugs with a 95% confidence interval of 18.2% to 30.0%. The Snodgrass and Anderson studies do not present a specific definition of lesion size, how the lesions were measured, or how much change in lesion size was considered meaningful in analyzing outcomes. The study states, "The local lesion was considered with regard to its site, extent, swelling, painfulness and tenderness."

If deaths and patients who failed on initial therapy are included in the analysis, the meta-analysis at 48 hours shows an overall difference of 26% in favor of sulfa drugs with a 95% confidence interval of 19.0% to 33.5%. While there was a treatment difference on duration of pyrexia and although body temperature is a biomarker that is often linked to the infection, these studies were performed before the introduction of antipyretics into clinical practice, and analysis of modern studies might be confounded by the widespread use of drugs or drug combinations with antipyretic properties. Thus, incorporation of the resolution of elevated body temperature into a response measure was not further analyzed. There were no differences in death, relapse, or complications at early or later time points upon which to base the design of future non-inferiority studies.

The Project Team agreed that the data from Snodgrass and Anderson on control of lesion spread are a useful starting point for further investigation and could be assessed by photographic recording. The Project Team performed a review of the medical literature evaluating current measurement techniques used to measure skin lesions and their associated variability. There were no articles specifically addressing the measurement of lesion size in cellulitis. The majority of the articles are focused on measuring the chronic wound healing process. The studies comparing various measurement techniques on wound models including length and width or area measurements with a ruler showed relative biases of between 38% and 60% compared to a coordinate measuring machine (CMM) with an absolute accuracy of 0.0002 inches (Langemo et al. 1998). The studies were performed on plaster of Paris models so error might be greater in the clinical setting where patient position, lesion flexibility, and definitions of lesion boundaries might be encountered. The measurement characteristics and reliability and reproducibility of various techniques in the measurement of ABSSSI lesions remain unknown.

2.2 Retrospective Data Analyses

The Project Team performed retrospective analyses of datasets from existing clinical studies to a) refine currently proposed outcome measures by evaluating their operational characteristics, changes over time, and responsiveness to change at specific time points in a modern clinical trial setting and; b) help identify additional endpoints or biomarkers that might be relevant. The Project Team has identified several sources of data from existing modern industry clinical trials that have been used as an in-kind contribution to the project. These analyses, which have also been contributed in-kind to the project, have been based in each case on a statistical analysis plan (SAP) drafted by qualified biostatisticians who are part of the Project Team; each SAP was shared with the entire Project Team for comment and approval prior to initiating the analyses.

2.3 Pre-existing Data

1. Cerexa, Inc. (a subsidiary of Forest Laboratories) generously provided the FDA-requested analyses from the following two registrational clinical trials:
    a. NCT00633152: A Phase III Study of the Efficacy and Safety of Ceftaroline Versus Vancomycin plus Aztreonam in Subjects With Complicated Skin and Skin Structure Infection

b. NCT00424190: A Phase III Study of the Comparative Study of Ceftaroline vs. Vancomycin plus Aztreonam in Adult Subjects With Complicated Skin Infections (cSSSI)
c. Publication: Corey GR, Wilcox M, Talbot GH, et al. Integrated analysis of CANVAS 1 and 2: phase 3, multicenter, randomized, double-blind studies to evaluate the safety and efficacy of ceftaroline versus vancomycin plus aztreonam in complicated skin and skin-structure infection. Clin Infect Dis 51(6): 641-50.

2. Durata Therapeutics generously provided the primary data tables from the clinical trial referenced in the following publication (note: this clinical trial is not listed on clinicaltrials.gov):
   - Publication of a Phase III Study: Jauregui LE, Babazadeh S, Seltzer E, et al. Randomized, double-blind comparison of once-weekly dalbavancin versus twice-daily linezolid therapy for the treatment of complicated skin and skin structure infections. Clin Infect Dis 41(10): 1407-15.
3. Pfizer generously provided the primary data tables from the following clinical trial:
   - NCT00087490**:** A Phase IV Study of Skin Structure Infections With Suspected or Proven Methicillin-Resistant *Staphylococcus aureus* (MRSA).

The Project Team conducted analyses of these datasets in a post hoc subgroup of all randomized patients defined by criteria similar to those used for inclusion in current clinical trials of antimicrobials for ABSSSI. The purpose of the analysis was to provide descriptive data of the lesion size in skin infections over the course of treatment and to assist in determining the definition of control of lesion spread in skin infections. Analyses provided descriptive statistics (mean, standard deviation, median, minimum and maximum value for continuous data, and number and percentages for dichotomous and ordinal data) and graphical depictions of summary data. No inferential statistical analyses (i.e., hypothesis testing) were completed. and thus, statistical power was not determined for this analysis. The Project Team agreed that the analyses developed from these studies are specific to the specific context of use evaluated in these studies.

3 Core Data Elements Reviewed and Major Project Team Discussion Points

3.1 Major Discussion Points

3.1.1 Clinical Trial Entry Criteria

1. *Lesion size*. An initial lesion with sufficient size to permit demonstration of resolution is required. The Project Team generally agreed that the FDA's proposed criterion of $\geq$ 75 cm^2 of cellulitis (erythema/redness, edema, or swelling) in adults was adequate. However, the Project Team found that defining lesion size based on erythema alone was the simplest way to provide a clear definition at our current state of knowledge. Future work could address inclusion of measures of edema or induration, but defining the edge of these processes is problematic at present for purposes of lesion size definition. Although smaller lesions are relevant in some clinical settings, the group estimated that measuring change with larger lesions would be more accurate. Some studies have included lesions as small as 1 cm^2, but the Project Team agreed that measurement of such small lesions would be challenging in

terms of reliability and reproducibility. The Project Team noted that for small adults and pediatric populations, ABSSSI may produce lesions that are proportionately relevant but that are too small to meet specific size criteria. Suitable age- and site-specific data do not exist on which to recommend a size-based assessment for cellulitis in these populations.

2. *Systemic signs*. There is currently uncertainty and inconsistency on how to define disease severity. The Project Team agreed that use of systemic inflammatory response syndrome (SIRS) criteria (heart rate, respiratory rate, blood pressure, white blood cell count) could provide useful supporting evidence that enrolled subjects have sufficient severity of illness. The Project Team agreed that SIRS criteria should not be a requirement at enrollment, and the group did not decide upon specific definitions for severity. It will be important to examine this in further studies.
    a. *Elevated body temperature*. Consensus was to not include elevated body temperature as either a baseline requirement for enrollment or as part of a primary outcome measure for the following reasons:
        i. There are challenges in measuring body temperature in a well-defined and reliable manner
            (1) There is no consensus on the definition of fever. Hence, it is preferable to use the term "elevated body temperature." When examining elevation in body temperature, there is no single threshold used to define fever. A measurement of 38.0 $^{\circ}$C is often thought acceptable, but there is no evidence or criteria available to support this particular criterion. Further, several methods exist for measuring body temperature at various anatomic sites (mouth, rectum, axilla, tympanic membrane, temporal artery); selecting and standardizing the best method for research studies will require further study.
            (2) The optimal frequency of temperature measurement is uncertain. There are no data in the medical literature on which to base a requirement for the frequency of measurements. Insufficiently frequent measurements may miss key events, but overly frequent measurements would create needless logistical difficulty. Moreover, measurement of temperature during a preset window can be particularly difficult in a clinical setting where most patients are not hospitalized and where measurements are to be recorded by the patient at home in a patient diary, leading to missing data that can translate in a programmatic failure outcome in an intent-to-treat (ITT) analysis set.
            (3) Missing data will contribute unnecessarily to higher failure rates when missing data are coded as failures of therapy. While the FDA draft ABSSSI guidance recommends four body temperature measurements within the 48- to 72-hour visit window, there is a risk that outcomes will be driven by the ability to gather all required measurements rather than by drug effect.
        ii. A requirement for elevated body temperature at enrollment will exclude many vulnerable and important-to-treat populations, such as the elderly, diabetics, and other patients who are immunocompromised. Such patients may not have elevated body temperature at baseline despite documented skin infection.
        iii. Elevated body temperature is not an effect modifier. Analysis of data from the ceftaroline studies showed that elevated body temperature is a predictor of response (i.e., lower success rates in patients with, than in those without, elevated body temperature when this sign is part of the endpoint) but not an effect modifier; i.e., the.

difference between test and control groups was not substantially different in the presence versus the absence of elevated body temperature.

iv. Concomitant antipyretics and anti-inflammatory agents can confound the assessment of the course of elevated body temperature. Many combinations of antipyretics and analgesics are available as over-the-counter medications in multiple presentation forms.

v. Elevated body temperature is not on the causal pathway of the disease.

3. *Local signs and symptoms.* The Project Team agreed that other local symptoms and signs such as assessments of dolor (pain), tumor (swelling and induration), calor (warmth), and drainage would provide useful information but need not be part of the primary outcome measure. Approaches to reliable definition and measurement of these variables would require further study. Also, needed are clear definitions of abscess vs. cellulitis vs. wound infection. When quotas on maximum number or percentage of a given clinical syndrome are to be enrolled, it is essential to have clear definitions of each category of syndrome.

3.1.2 Primary Efficacy Endpoint

1. Two components of the primary endpoint were proposed initially, prior to examination of the clinical trial data.
   a. "Control of lesion spread" at 48 hours relative to the previous measurement at 24 hours; and
   b. Percent change from baseline, with the change not being more than "x" percent, where "x" could be a positive or negative number, to be determined.
2. The first criterion was based on the recognition that a condition for "success" in a "control of lesion spread" endpoint at 48 hours is that the lesion sizes should not still be increasing during the previous 24 hours; i.e., between 24 and 48 hours.
3. The second criterion was included because maintaining the integrity of randomization is of integral importance in causal efficacy assessments. To be specific, an assessment of treatment effect cannot be based solely on a change between 24 hours and 48 hours because the 24-hour value is influenced by treatment assignment. Change from baseline preserves the integrity of randomization and prevents selection bias. Note that by choosing "x" to be a negative number, a case where a modest-to-large increase in lesion size occurs in the first 24 hours could be a success only if an even larger reduction in lesion size occurs between 24 and 48 hours.
4. After "x" was chosen to be -20%, available studies were analyzed regarding the patterns of change at 24 and 48 hours. In these analyses, for subjects in whom the 48-hour lesion size showed a $\geq$20% decrease from baseline (i.e., subjects in whom criterion 1.b is satisfied), there were very few having an increase in lesion size between 24 and 48 hours; i.e., few in whom criterion 1.a is not satisfied. Hence, because validity of the second criterion almost always implies validity of both criteria, it was recognized that it was not necessary to state the first criterion. Because it is sufficient to assess only the second criterion, a criterion based solely on the change from baseline at 48 hours, the need to perform a 24-hour assessment was eliminated. This decision should enhance trial integrity by reducing the level of missing data in the assessment of the primary endpoint.
5. A goal of trials is to maintain the ability to detect clinically meaningful differences between drugs. High success rates (on the order of 90%-95%) may make it more difficult to evaluate

assay sensitivity and may require smaller non-inferiority margins. Therefore, success rates in the 80-85% range better enable the detection of a drug benefit. When the primary endpoint of a clinical trial is a dichotomous outcome of response vs. non-response in settings where the metric used for the primary analysis of the primary endpoint is the absolute difference in response rates (rather than the odds ratio), it usually is difficult to ensure the integrity of a non-inferiority analysis if the response rate on the active control arm is extreme, such as in the range of 95%. This is especially true when there is some subjectivity in the assessment of that endpoint even when outcome assessments are blinded and when clinicians or evaluators are instructed or guided in a manner that patients are called a success. Settings where there is some uncertainty, or where enrollment is focused on those participants highly likely to be called a success, or where one allows prior or concomitant antibiotics or other interventions as ancillary treatments further increase the likelihood of success. In such settings, if absolute non-inferiority margins such as 10% are used, there is an unacceptable likelihood that such margins would be ruled out even for an ineffective or inadequately effective intervention that has a failure rate that truly is two- to four-fold higher than that of the active control. The solution in such settings would be to have endpoint definitions, patient selection, and trial conditions such that the expected success rate on the active control is considerably less than 95%, or by use of an odds ratio metric for the analysis of treatment effect, or by using an absolute difference metric with a properly conservative non-inferiority margin. Defining outcomes such that success rates using a pre-specified definition of success are in the 80% to 85% range in ABSSSI trials would help control for this potential source of error in non-inferiority trials while maintaining feasibility of conducting these trials by avoiding prohibitively large sample sizes, but does have the drawback that each future trial would have a different percent reduction in lesion spread as its response definition.

6. Consensus was that elevated body temperature is not required at baseline (as above), though resolution of elevated body temperature could be evaluated as a safety concern; i.e., as was done in evaluation of inhaled aztreonam in cystic fibrosis patients. Elevated body temperature can have many causes including drug fever, an instance in which the patient may be cured but have an adverse event.
7. The Project Team agreed on the need to assess the "total picture" of the patient at a late endpoint. The group discussed that the variables measured at later time points need to be specifically defined and that the components of this total picture and "overall pattern of response" should be specified. The group did not discuss specific definitions but agreed that drug sponsors should discuss definitions of measurements at later time points with the FDA. Some points of discussion regarding the definition of a late endpoint included the following:
    a. The FDA has pointed out that clinical endpoints, defined as those that directly measure how the patient "feels, functions, or survives," are important and required by regulation. Mortality is an important component of any outcome measure, but it is low in ABSSSI trials. Thus, effects on how patients feel and function would drive the overall analysis.
    b. From the Snodgrass data, there is a concern that if one evaluates response later in the natural history of disease, evaluated in their two studies during or after receipt of an intervention, success can be observed with an ineffective treatment given that the treatment difference between antimicrobial and the control group decrease over time. Thus, success rates that are similar between two active drugs "early" may reflect treatment effects of antimicrobials relative to placebo, whereas success rates that are

similar "late" may be consistent with the natural history of resolution of the illness based on the available data in the Snodgrass trials.

c. The group agreed that outcomes measured at a fixed time point post-randomization are important to evaluate sustained early response as well as other outcomes of interest. These measurements should be specifically defined and noted in the product labeling; e.g., Table 9, Teflaro (ceftaroline fosamil) US Prescribing Information, October 2010.

8. Collection of sufficient pharmacokinetic (PK) data to estimate individual subject drug exposure would allow for more complete exposure-response analyses for both early and late endpoints.

4 Data Not Yet Available and Needed for Final Recommendations/Future Research Needs

The Project Team proposed the suggested outcome measures noted above for use by sponsors as outcomes for registrational trials in ABSSSI. These outcome measures are intended as "bridging" measures for ongoing or currently planned clinical trials while gathering more information on outcome measures in ABSSSI. The Project Team noted that the currently available data leave substantial gaps in the evidence base in terms of issues related to enrollment criteria, determining severity of disease, and which baseline factors are effect modifiers. Some of the other outstanding issues are noted above, including evaluating the quantitative relationship between control of lesion spread and how patients feel and function and the validity, reliability, and reproducibility of the methods used to assess outcomes in ABSSSI. While there is an urgent need to decide on bridging endpoints for the short-term future, it is important to note that these suggestions are a bridge and not final, evidence-based enrollment criteria or outcome measures. There is an equally urgent need for prospective research studies to address these outstanding questions as discussed in the following section. Any final recommendations should be based on evidence from studies and not opinion based.

To address these gaps, a qualitative research phase has been proposed by the Project Team. The proposal includes use of one or more research firms selected through a formal RFP process to complete a qualitative research phase of instrument development—involving both literature searches and patient interviews—to begin the process of developing well-defined and reliable outcome measures that may be used in clinical trials in ABSSSI. The goal of these studies would be to address the research gaps noted in the retrospective data analysis stage of this project. The proposed studies will be conducted by a group of researchers highly experienced in the field of infectious diseases and will be guided by a Project Team that includes academic clinicians, drug development personnel from pharmaceutical companies, and representatives from the NIH and the FDA.

Results from the retrospective clinical trial analyses and qualitative research studies will be used as input to designing prospective clinical studies to be conducted as part of a potential third phase of this project, which would be proposed as a separate Biomarkers Consortium project and be focused on the design and conduct of one or more clinical studies to further test and validate specific endpoints and measurement approaches. While a standalone study cannot be ruled out, it is expected that these later studies will be able to be coordinated as companion or exploratory

studies to current trials being conducted by the National Institute of Allergy and Infectious Diseases or industry.

Remaining challenges to be addressed in future research are listed below:

1. There is a need for evidence on what are the most well-defined, reliable, reproducible, and feasible methods for measuring efficacy outcomes in ABSSSI trials.
    a. Develop an endpoint model for ABSSSI.
    b. Capture data on the outcomes most important to defining treatment benefit (how patients survive, feel, and function) .
    c. Develop and qualify well-defined, reliable, and reproducible outcome assessments for each targeted patient population. The outcome assessments developed should include:
        i. Development and evaluation of patient-reported outcome (PRO) assessments of symptoms and signs that the patient can best observe and report; and
        ii. Development and evaluation of observer-reported outcome assessments of signs for use when patients cannot respond for themselves (e.g., pediatrics).
    d. Develop and evaluate clinician-reported outcome assessments of signs that require clinical examination and interpretation.
2. Improved methods to measure lesion size are needed.
    a. Alternative ways to define the boundaries of the lesion could be considered. As noted above, the Project Team found that visible erythema was the most straightforward approach and that this was representative of the type of data available for analysis. The group agreed that the data analyzed were specific to the context of use in evaluated studies. Other approaches in other setting may be appropriate, and drug sponsors can discuss these with the FDA.
    b. Future studies should evaluate how to measure other variables including edema and induration as well as erythema.
    c. Approaches to computing lesion size based on simple length or width measurements must be clarified.
    d. Variations in measurement methods for obtaining length and width assessments of lesion sizes have not been characterized.
    e. Analyses comparing outcomes based on area vs. those based on separate decrease in length and width analyses would be helpful. This type of data was presented for linezolid but not for the ceftaroline and dalbavancin datasets.
3. Evaluation of lesion size in different populations and anatomic sites is needed. There may be instances where lesions, depending on anatomical sites or type of clinical syndromes, may decrease in size at different rates. This complexity requires further investigation as there were no data presented based on outcome by anatomical site. Moreover, these observations only reflect what is viewed on the surface and not the depth of the lesion that is present.
4. No prospectively collected data are available from pediatric trials. Analyses of enrollment criteria and outcomes based on anatomic site and in various pediatric populations (newborns, infants, and children) would be helpful as minimum lesion sizes of > 75 cm^2 do not apply in these populations.
5. The relationship of how patients feel, function, or survive to a reduction in lesion size from baseline is unknown. An area for future research would be to evaluate the quantitative relationship between effects on measures of patient symptoms and function and effects on signs of disease like control of lesion spread.

6. Currently, there are no valid instruments to measure disease severity (defined as baseline variables that predict outcomes) in ABSSSI. Lesion size may not be the only relevant measure of disease severity in either adults or pediatric populations. Variables such as body temperature may be a component of measurements of severity at baseline but are not suggested as post-baseline measures or as a primary outcome measure. Future research is needed to develop and evaluate severity measures for ABSSSI.
7. Clear definitions of each disease syndrome (cellulitis, abscess, wound) are needed given the overlap among clinical syndromes.
8. Effect of timing window on outcomes (are there differences in outcome within the 48 to 72 hour assessment window?) and evaluation of time to response are needed.
9. Definition and development of outcome measures for late time point assessments are needed, including later measurements of lesion size as well as other measures that reflect how patients feel and function.

5 Interim Recommendations

5.1 Study Design

1. Most studies comparing one active agent with another would utilize a non-inferiority study design due to ethical and feasibility issues. The Project Team's discussion was based on the presumption of outcome measures used in the setting of non-inferiority trials.
2. Superiority trials do not have a requirement for prior evidence of drug effect and therefore have greater choice in outcome measures. These are not discussed here.

5.2 Endpoint Definition

1. *Lesion area.* Success will be defined as a $\geq$ 20% decrease in lesion area (defined as longest head-to-toe length times longest perpendicular width) versus baseline in adults. This primary endpoint was defined based on the available study data, and the Project Team acknowledges that other responder criteria may be relevant depending on the context of use.
2. An alternate endpoint of a variable percent change that is chosen based on a pooled response rate of 80% is an option that requires further research and review.
3. *Body temperature*. Absence of elevated body temperature (fever) should not be a component of the primary outcome measure because it is not on the causal pathway of the disease and temperature measurements cannot be obtained reliably in outpatients.

5.2.1 Timing of Outcome Assessments

Early Assessment

1. Success should be determined at 48-72 hours after randomization, with a strong recommendation that baseline measurement and administration of the first dose of study drugs occur as close to randomization as possible.
2. Notably, the historical data on which the 48- to 72-hour recommendation is based provided the collection day, not the hour; therefore, this proposed timeframe is an area for further investigation.

This assessment is made independent of treatment decisions made by the investigator. The proposed early endpoint serves to link to the historical evidence of drug effect in previous studies

required in the design of non-inferiority trials. The classification of response or failure is also independent of study events other than death—institution of other therapy, unplanned surgical drainage, and adverse events are not counted as part of the early response classification. These measures can be captured as part of supportive information. Patients can continue on study drugs even if classified as an early non-responder in a post hoc early analysis or could be withdrawn from study drugs even if classified as an early responder.

Challenges of the Early Assessment include the following:

- The historical data generated by Snodgrass et al. are not auditable.
- Conclusions are based on post hoc subgroup analyses.
- Substantial amounts of missing data from some of the studies.
- There were differences in the ways patients were treated between trials.
- Comparators for new trials will be drugs approved using the older paradigm
- Analyses accounting for anatomic site comparing outcomes using a decrease in area of the lesion compared to a decrease in both length and width measurements are not available**;** this requires further research.
- While the group agreed there is an observed relationship between control of lesion spread and other measures of disease outcomes (e.g., decrease in pain or other signs of disease), there were no quantitative analyses of the relationship of these outcomes to each other over time; this requires further research.
- There is a presumed link between the biomarker of erythema and the course of the infectious process (or, at least, to the related inflammatory process), but erythema is not the cause of the infectious process, nor is local infection the only cause of erythema.
- The definitions of lesions and methods of measurement in the historical datasets were not defined.

Later Assessment

1. The Project Team agreed that longer term efficacy outcome measures at specified time points in both test and control groups at end of therapy (EOT) and long-term followup are of essential relevance to clinicians and their patients. The FDA recognizes this as well and has included these data in the ceftaroline prescribing information.
2. Evaluation of sustained early control of lesion spread at a later time point is critical, and other clearly defined assessments of how patients feel or function across the entire period from onset to resolution of the process are critical. However, defining components of later time points requires further investigation and development. In the meantime, sponsors should propose to the FDA what and when they plan to measure at later endpoints.

The Project Team suggested later assessments to evaluate sustained response on the initial control of spread achieved at 48-72 hours after the first dose of study medication. This may be measured by assessing change from baseline at later time points on control of lesion spread, assessments of other measures of how patients feel and function, and adverse events.

- Assessments at later time points can provide information on sustained early response. The late endpoint should occur at a fixed time point relative to day 1 and examine sustained response.
- The later time points could be two, preferably fixed, time points; one at the EOT (e.g., 7-10 days following randomization) and the other at an off-therapy time point; e.g., 14-21 days

following randomization. The exact time point post-therapy should be determined depending on the maximum length of treatment and the PK/pharmacodynamic (PD) characteristics of the drug but should be the same time point in both test and control groups.

- This assessment should incorporate a continued measurement of lesion size using the same methods as earlier measures used at baseline and at the 48-72-hour measurement. The lesion should be stable over a period of 2 days with no increase in lesion size that exceeds a threshold percent change pre-specified in the protocol. The finding of sustained response should be accompanied by stability in vital signs including absence of elevated body temperature. If sponsors wish to perform a late assessment, they would examine sustained response in the entire population, not just in the population with initial early response. This late assessment should be based on the primary lesion; a distant site infection is considered a post-baseline event that should be captured and reported. The Project Team notes that more research is needed in this area regarding defining "cure" in skin abscess; in particular, for defining an appropriate threshold for reduction of lesion size defined as primary lesion size $\leq$ x% , the percent change threshold defined in the protocol in relation to the baseline value and no new lesions.

5.2.2 Needs—Future Research Goals

The exact definition of sustained response for a later time point remains to be determined, as a reduction in lesion area of more than 20% may be necessary to be considered successful once subjects have stopped study drug.

1. Later time points could assess other variables in addition to sustained response. The group did not discuss specific definitions for later outcome assessments. The group agreed that sponsors should discuss definitions with the FDA for individual protocols.
2. The Project Team suggested several ways of incorporating the late time point into the overall analyses of study results. These analyses are not conflicting approaches and rely on demonstrating non-inferiority at the early time point as the anchoring assessment for purposes of ensuring assay sensitivity in the setting of non-inferiority study designs. Studies should capture information at both early and later time points to enable performing analyses such as:
   a. Demonstration of non-inferiority at the early time point with descriptive information only on later time points. Approval would be based on non-inferiority at the early time points with descriptive information in labeling related to later time points; or
   b. Demonstration of non-inferiority at the early time point with a pre-specified hypothesis related to superiority at a later time point on the overall response of initial response plus sustained cure. This result could result in specific superiority labeling for the test therapy.

5.2.3 Comments on Key Study Enrollment Criteria

As discussed in Section 3.1.1 of this document, the Project Team recommends the following with regard to enrollment criteria:

1. *Lesion size*. The initial lesion size should be $\geq$75 cm^2 in adults. The Project Team agreed that further study is needed of smaller lesions in regions limited in size by anatomy or populations; e.g., pediatric patients.

2. *Lesion measurement*. Analyses by area and length, width, and length/width presented for linezolid studies indicate area could be used a primary assessment.
3. *Other local and systemic criteria*. The Project Team agreed that a minimum proportion of subjects with SIRS criteria is not necessary. Sponsors are encouraged to capture data on heart rate, blood pressure, respiratory rate, and WBC count and local pain, warmth, erythema, and drainage to ensure that studies address concerns about severity of illness and requirements for different registration authorities.
4. *Body temperature.* The presence of elevated body temperature is not required at enrollment or as an outcome measure (see discussion under Section 3.1.1).

5.2.4 Comments on Key Analysis Issues

1. ***Endpoint models.*** The Project Team did not develop an endpoint model as defined in the draft FDA Guidance on Qualification Process for Drug Development Tools (FDA 2010, October). While the Project Team has focused on the endpoint of early response as measured by control of lesion spread, the team is vitally interested in defining other concepts that should be included in the endpoint model for ABSSSI. This will be evaluated as a component of future research.
2. ***Outcome measure.*** The suggested outcome measure is a comparison of differences in proportions of the numbers of subjects in each study group designated as "responders" based on the definition of control of lesion spread at 48 hours post first dose of study drug as defined above.
3. ***Response assessment time point.*** To avoid the bias that may be produced by following the rule of last observation carried forward (Fleming 2011), response in all study arms should be assessed at fixed time points following initiation of study therapy. This is especially important if study drugs are administered for different lengths of time.
4. ***Inclusion of PK/PD analysis.*** Collection of sufficient PK data to estimate individual subject drug exposure allows for more complete exposure-response analyses at both early and late endpoints and is strongly encouraged.
5. ***Missing data.*** Missing data can increase bias and decrease validity of clinical trials, so sponsors should attempt to minimize missing data as much as possible. Outcome measures that are meaningful and practical to the given clinical setting may help minimize missing data, such as not requiring multiple body temperature measurements within the 48- to 72-hour visit window in outpatients.
6. ***ITT and per-protocol analyses*.** The ITT analysis should be primary (hence, missing data are of critical importance). Per-protocol analyses should be assessed as supportive.
7. ***Proposed non-inferiority margin*.** The group did not discuss this in detail. Recent trials have most often employed a margin of 10%.
8. ***Sample size considerations*.** This was not discussed by the group and would depend on the hypothesis being tested, the specific population under study, the desired study power, and the proposed non-inferiority margin.
9. ***Opportunities for global harmonization.*** It was recognized by the Project Team that the alternative approach to endpoints under development might not be equally acceptable to all regulatory agencies. In minutes from an EMA workshop held on 7-8 Feb 2011, there was suggestion of submitting a separate SAP for ABSSSI trials, in which the primary endpoint of interest is late. These minutes point out that such an approach could satisfy multiple

regulatory agencies. The main issue is that sponsors should capture the data for both early and later assessments. See recent EMA workshop summary atwww.**ema**.europa.eu/docs/en_GB/document.../04/WC500105473.pdf (pg 8-9). Guidance pending. Performing both early and later assessments, as outlined above, will provide a means of developing separate analyses for U.S. and other regulatory agencies. This could allow the same studies to meet different standards and thus become harmonized globally.

5.2.5 Alternative Viewpoints

There are alternative viewpoints within the Project Team regarding the conclusion that the primary measure should be based on a measure of reduction in the size of the associated erythema taken at 48h after initiation of therapy. Although there was agreement among team members that the early measurement provided important information, some concerns were raised and should be addressed in future research.

1. ***There is no evidence that use of test-of-cure (TOC) endpoints has led to approval of inactive therapies for life-threatening infections such as ABSSSI (Spellberg 2011).*** Although there are demonstrated instances of detection of differences in efficacy or safety among agents for life-threatening infections as well as instances of detection of ineffective agents that were not subsequently registered for the given indication (e.g., daptomycin for pneumonia), there are no data to support the contention that genuinely ineffective agents have been approved using traditional TOC endpoints. Older agents may well have been superseded by newer agents for reasons of safety, convenience, or development of resistance, but there are no data to suggest that the older agents were ineffective at the time of approval for their given indication. As discussed below, this is presumably because early endpoints have always been a part of traditional TOC endpoints, albeit not necessarily in a formal manner.
    a. Other members of the group pointed out that failure to show differences between drugs does not mean that differences do not exist in the setting of non-inferiority trials that may lack assay sensitivity.

2. ***Recent pharmacometric analyses show a correlation between drug exposure and TOC endpoints.*** A recently presented observation (European Medicines Agency 2011) is that pharmacometric exposure-response analyses demonstrate a correlation of drug exposure with traditional clinical and microbiological endpoints.
    a. Arguments in favor of the plausibility of these correlations include:
        i. The demonstrated relationships indicate that contemporary clinical endpoints (e.g., success or failure at the TOC) capture a measure of drug effect.
        ii. These analyses produce estimates of treatment effect relative to placebo, which are similar to estimates derived from other sources but that are derived using current data from modern studies and thereby could negate concerns of meeting the constancy assumption.
        iii. The consistency of these observations (similar results can be shown across both multiple indications [HAP, VAP, CAP, ABSSSI, and ABECB] and multiple antibiotic classes), the biological plausibility of the observations (drug effect should decline as exposure declines); the retention of the correlations when the analysis is

controlled for age, severity of illness, or co-morbid disease; and the lack of an hypothesis regarding a host immune factor that would correspondingly alter drug exposure lend support to the need to consider carefully this approach.

iv. In particular, this approach offers the possibility of validating non-inferiority margins using modern trial designs and modern endpoints. Moreover, pharmacometric exposure-response analyses offer the possibility of linking early and contemporary late clinical endpoints.

b. This approach, however, can also be critiqued:

i. Although these analyses are useful for identifying prognostic factors and generating hypotheses regarding plausible doses and schedules to be studied in properly conducted randomized trials, attempts at causal inferences from such analyses are biased due to confounding between treatment effects and prognostic patient characteristics.

ii. Specifically, it is not sufficient that exposure or organisms may be randomly assigned since host factors are not randomly assigned and these latter factors cannot be adequately accounted for by matching. People with differing concentrations or minimum inhibitory concentrations can differ on other factors that affect outcome, like age, severity of illness, co-morbid disease, or many other covariates, and most of these are unidentified or unrecorded. Inherent differences in such patient characteristics are sufficiently influential to lead to substantial differences in concentrations; therefore, it is likely that these inherent differences are also meaningfully predictive of the outcome measures.

iii. The consistency of results across settings may thus be explained by consistency of this same bias across those settings.

3. ***Cessation of spread alone is an insufficient measure.*** Cessation of spread of the erythematous lesions supports a non-inferiority margin to compare treatment effect between two drugs. However, it alone is not necessarily sufficient to indicate clinical response and serve as primary efficacy endpoint. As noted above, erythema is a biomarker for clinical response, albeit a biomarker with a high degree of biological plausibility and the appearance of a good linkage to the course of this disease (or, at least, to the course of the related inflammatory process). As noted by the Project Team during the discussion of the Cerexa (ceftaroline) dataset (see Appendix for further details), "We should be mindful about equating the Day 3 reduction in lesion size to how a patient functions or feels. It is tempting to think that a reduction in lesion size might make a difference to a patient; however, there is no definitive evidence thereof. Thus, one cannot conclude that it has any meaning to the patient, such as earlier discharge, reduced purulent complications, shorter treatment time, etc." The Project Team agreed that defining the components of the endpoint model is a critical part of future research.

4. ***Lesion size requirements may exclude key groups.*** A requirement for $\geq 75 cm^2$ erythema/redness edema and swelling was noted to exclude some patient groups of interest (e.g., adults with infections of hands, face, and genital organs, and children); however, no data were presented on operational characteristics of measurements at these anatomic sites or populations.

5. ***Early time points are already part of the TOC outcome.*** An early measure of response is included in all clinical trials, but the timing and formality of this evaluation may differ from trial to trial, and there is not a systematic requirement for investigators to make a final assessment at this time point. If improvement is not apparent at Day 3 or 4, the patient is generally withdrawn from study medication and the response defined as a failure for effectiveness analyses. These outcomes are carried forward for purposes of analyses at later time points. In some trials, this early assessment has been entirely informal and is captured only by noting whether the physician and patient continued the randomized therapy. In other trials, a formal recording of a decision to continue has been taken. A systematic analysis of early time points with clear definitions of outcomes would help clarify the analysis of trial results. A great strength of the work presented here is that it provides a basis for documenting the reasoning that goes into the decision to continue or discontinue therapy at an early time point. Early and later time point assessments are not mutually exclusive and can both be measured in the setting of clinical trials.

6. ***Later endpoints provide a key overall perspective.***
   a. Overall clinical cure at a late time point following EOT is important to evaluate sustained response and should be noted in the product labeling. Given that this measure thus takes on the role of being the principal measure that is relevant to the use of a drug, it could be argued that this measure best meets the ICH E9 (Section 2.2.2) test: "The primary variable ('target' variable, primary endpoint) should be the variable capable of providing the most clinically relevant and convincing evidence directly related to the primary objective of the trial."
   b. Use of the early endpoint as the primary study endpoint has not, to date, been specifically endorsed by other regulatory agencies. The full implications of the use of dual SAPs have yet to be understood by the community.

7. ***Time to response may provide useful insights.*** Some in the group posed that response rates that are equivalent with a substantial reduction in lesion size at EOT and TOC are what matters most to clinicians and patients; however, there are no systematic studies evaluating this question by surveying patients or clinicians. Since clinicians often choose to change therapy in the absence of response within 2 or 3 days, the early time point must have some value in addition to later time points evaluating sustained response or other variables such as adverse events. Time to response may also be an important measure but not one for which there are data to pose a hypothesis for a non-inferiority trial.

5.2.6 Additional Information Needed To Advance to Final Recommendations

(See Section 4 for more detailed discussion)

It is clear from an analysis of current data that the evidence base for determining outcomes in ABSSSI is still incomplete. In some situations such as pediatrics, the Project Team was not able to analyze any data. It is crucial that future studies address the research gaps in terms of what outcomes are most important to patients and how to reliably and reproducibly assess outcomes in ABSSSI across various disease states (cellulitis, abscess, etc.), in populations (pediatrics, adults), and various anatomic sites.

- Qualitative research phase of project and assessment of various measurement instruments as described above and development of an endpoint model for ABSSSI;
- Analysis of baseline measures and outcomes by anatomic site;
- Analysis of baseline measures and outcomes in pediatric populations;
- Analysis of comparison of area or length by width measurements in two of the datasets;
- A better understanding of the relationship between erythema and how patients feel and function; and
- The use of pharmacometric-based analysis methods should be explored as a possible complementary tool to further validate both early and late endpoints. Methods that evaluate randomized groups would help control for bias in these analyses.

6 Conclusions

In the process outlined, above various stakeholders including members of academia, industry, and Government agencies proposed interim bridging outcome measures for currently planned trials in the ABSSSI indication.

These interim outcome measures are based on an evidence-based analysis of the historical literature that showed evidence for a treatment effect of antimicrobials in ABSSSI based on the control of lesion spread at 48-72 hours after the first dose of study drug and recently performed clinical trials. Although the proposed outcome measure is a biomarker, its use in this setting is based on the non-quantified observed relationship between erythema and how patients feel and function. The 48-72-hour time point shows a substantial treatment effect for antimicrobials, allowing assessment of non-inferiority of active agents at this time point. This large treatment effect (M1) provides a justification for selection of an M2 on the basis of clinical reasoning.

The Project Team agreed that the analyses evaluated by the group were relevant to the context of use in the studies evaluated. Sponsors may propose other outcome measures and timing of those measures in disease settings that differ substantially from the studies evaluated by the Project Team.

The Project Team agreed that later assessments are important to provide an evaluation of sustainability of initial response, as well as to provide other measures of how patients feel and function over a longer time period as well as to support proposed duration of dosing in the USPI. The Project Team agreed that these assessments should be included in studies of ABSSSI as in the recent ceftaroline FDA approval.

These interim outcome measures would allow studies to proceed while the Project Team plans future qualitative and quantitative research studies to evaluate the relationship between outcome measures in ABSSSI and the operational characteristics of various measurement methods in assessing outcomes in ABSSSI. These future studies are critical in addressing knowledge gaps that would aid in addressing unanswered questions related to designing trials in ABSSSI.

7 Appendix—Clinical Trial Data

7.1 Cerexa, Inc., Data

1. A brief overview of clinical trial data as analyzed by FDA was given. The analysis includes data from two identical, multinational, randomized, double-blinded non-inferiority Phase 3 studies of ceftaroline versus vancomycin plus aztreonam, using a non-inferiority margin of 10%, based on a pre-specified clinical response at TOC. Briefly, drug treatment was administered intravenously for 5–14 days with no oral step-down therapy. Daily "separate" lesion length times width and temperature measurements were recorded for the first 5 days of treatment.
2. Of the 1,396 subjects enrolled, approximately 57% were in the FDA-defined Modified Intent To Treat (MITT) population, with a lesion size at baseline ≥75 cm² originating from an infected wound, abscess (with > 5 cm surrounding cellulitis), or deep/extensive cellulitis or an ABSSSI in patients with diabetes mellitus or peripheral vascular disease. For the subgroup of patients with a lesion size greater than or equal to 75 cm², demographic data were presented on the percentage of patients with fever (41–47%) and increased WBC count (42–50%). The range of lesion size was 75–5,000 cm², with cellulitis as the major infection type. Clinical response for the exploratory analysis was defined as both "cessation of lesion spread" at Day 3 (comparing length times width measurements from baseline) and a temperature ≤37.6 °C.
3. Based on the primary Day 3 endpoint (cessation of spread and afebrile), the responder rates were in the upper 60% range for the vancomycin group and 74% for the ceftaroline group, with a treatment difference of approximately 6–9 percentage points in favor of ceftaroline. Evaluating percent reduction, patients considered a responder exhibited a 10–20% reduction in lesion size regardless of the treatment given, with a fairly constant treatment difference remaining throughout the percent reduction range.
4. In comparing patient populations with and without fever at baseline, "cessation of spread" occurred in 83–93%, regardless of baseline fever. The presence of fever at Day 3 did not affect the achievement of "cessation of spread." Also, it should be noted that about 5–10% of patients developed fever (or perhaps had documentation of pre-existing fever) after study enrollment.
5. Responder rates are in the 50–60% range with fever as a requirement at baseline, whereas without fever, the rate is in the 70% to upper 80% range.
6. In conclusion, for cessation of lesion spread (from 0% up to a 20% reduction), the absolute treatment difference remains relatively similar between groups; however, the responder rate decreases substantially; i.e., the power decreases by requiring increased percent reduction of lesion spread. From the FDA's perspective, it is most practical to use an absolute reduction compared to the baseline measure in lesion size as an endpoint as opposed to cessation of lesion spread compared to a post-randomization measurement obtained on the previous day. Cessation of lesion spread is unrelated to the presence of baseline fever; responder rate is influenced more by defervescence.

Clarification questions and responses:

1. Vancomycin fever, known to occur in approximately 5% of patients, might have confounded the data. Specifics about the frequency of this event are not readily available. Skin reactions (rash) to vancomycin generally appear on treatment Day 1, requiring treatment termination.
2. The circumstances surrounding patient study enrollment (e.g., hospital or emergency room [ER]) can greatly impact whether fever is documented. For example, in a hospital setting, the daily highest temperature is recorded, whereas in the ER, the temperature measurement is often a one-time event. The uncontrolled use of self-administered antipyretics is another problem that could influence temperature in the ER setting and was not controlled in these studies.
3. There were no differences between groups with respect to the length of therapy, which was 8.5 days throughout.
4. In analyzing the data, there is more signal when one looks at increasing the fraction of patients that have a substantial reduction of >50 or 75%. Also, when the measurement is performed later in the course of therapy, sensitivity is lost; if there is a treatment effect, there is greater sensitivity at Day 3. If there is a direct relationship between lesion size and pain, this is one example of capturing something of significance to patients.
5. We should be mindful about equating the Day 3 reduction in lesion size to how a patient functions or feels. It is tempting to think that a reduction in lesion size might make a difference to a patient; however, there is no definitive evidence thereof. Thus, one cannot conclude that it has any meaning to the patient, such as earlier discharge, reduced purulent complications, shorter treatment time, etc.
6. The fact that the response rates are equivalent with a substantial reduction in lesion size at EOT is what matters most to clinicians and patients.
7. Of note, an early measure of response is included in every clinical trial today; if improvement is not apparent at Day 3 or 4, the patient is withdrawn from the trial due to failure. One should be careful about saying the late TOC measure is flawed; rather, one should consider the total treatment course, incorporate early failure, and carry it forward. From the Snodgrass data and in CABP, there is a greater sense to look uniformly over time rather than an earlier time point.
8. There are two issues regarding clinical relevance. From the Snodgrass data, there is a concern (but not proof) that if one evaluates response later in treatment, success can be observed with an ineffective treatment. Thus, success rates that are similar "early" may reflect treatment effects, whereas success rates that are similar "late" may be consistent with a placebo effect.
9. The relationship of how patients feel, function, or survive to a reduction in lesion size from baseline of 30% or 50% is unknown, although clinicians believe patients have improved if the lesion has stopped spreading early in treatment. Snodgrass realized this by just looking at lesion spread. Stability versus baseline AND improvement versus baseline are being considered as the Day 3 endpoint because of the statistical concern regarding use of a post-baseline as baseline. Perhaps PRO data can provide more direct insight and guidance on how to define "x" in terms of clinical relevance. Also, how patients' signs relate to symptoms is a good example of what should be evaluated in the next phase of this project.
10. Clinical stability is an operational indicator used to make decisions about whether or not to escalate treatment or change to oral therapy. However, there are instances where different lesions, depending on anatomical sites, decrease in size more quickly than others. This complexity requires further investigation. Moreover, these observations only reflect what is

viewed on the surface and not the depth of the lesion that is present. We know from treating patients that there is a point where the lesion has stopped spreading, which is when the patient feels better.

11. Responder rates should only be high—roughly 90% for a drug like vancomycin—at an EOT or TOC that matters most to the patient. For a new drug, it is difficult to show superiority in a primary analysis; in a secondary analysis, it is very useful to support the non-inferiority finding.
12. Regarding erythema, the variations in measurement methods for assessments of lesion sizes have not been characterized.
13. With respect to EOT, if a patient fails on Day 3, then that is considered the EOT. Therefore, EOT is not a fixed point in time.

7.2 Durata Therapeutics Data

1. A reanalysis of the VER001-09 study in ABSSSI was performed with the goal of identifying novel endpoints to move the program forward. The data are from a randomized, double-blind, multicenter Phase 3 trial comparing intravenous linezolid with dalbavancin for treatment of complicated skin infections (one-third had cellulitis, one-third had abscess, and one-third had wound infection or "other" soft tissue infections, but many with "other" were reclassified as having cellulitis).
2. The purpose of the reanalysis was to examine changes in lesion size after the initiation of antimicrobial therapy and to better understand the performance of cessation of spread at Days 3 and 4. Thus, lesion size was measured at Days 3, 4, 8, and later. Patients received either dalbavancin on Day 1 and Day 8, or linezolid every 12 hours, but could switch to oral forms of the drug after Day 1 (most switched at about day 4).
3. This analysis was designed to assess response in the subgroup of patients with ≥ 75 cm^2 cellulitis.
4. The primary endpoint was clinical success in the dalbavancin group at Day 28. Inclusion/exclusion criteria were similar to the trials described previously, although MRSA patients were included in this study to better understand its effect on the drug. The median lesion size of these patients was 260 cm^2.
5. The response rates for dalbavancin and linezolid are significantly lower (69–78%) if patients with missing data are included as a failure. However, when the missing data are excluded, the cessation of spread success rate is greater for both groups (88–89%).
6. With a mean baseline lesion size of 260 cm^2 at Day 3, there was a 40% reduction (105 cm^2) at Day 4, and a further reduction in lesion size to 229 cm^2 by Day 8, indicating that about 50% of the lesion size diminution occurs at study Day 3 or 4. Patients considered a clinical failure experienced a 20% increase in lesion size above baseline.
7. Inclusion of fever resolution, cessation of spread, and missing data (regarded as failures) in the analysis produces a success rate of 78%, whereas exclusion of fever and missing data produces a response rate of 88–89%. Regardless, there were no treatment differences between the two drugs tested.
8. In summary, missing data complicates interpretation; better data are needed to correlate early responders to clinical outcome as defined at EOT. Inclusion of fever as a clinical marker is debatable.

Clarification questions and responses:

1. In assessing clinical failures, lesion size obtained early in the study should be correlated with lesion size at EOT or TOC.
2. Temperature does not appear to be an effect modifier and does not change the difference between these two drugs; therefore, inclusion of fever does not add value to the overall study conclusions.

7.3 Pfizer Data

1. The following analyses are from the Pfizer study "Linezolid versus Vancomycin in Treatment of Complicated Skin and Soft Tissue Infections" (Weigelt et al. 2005).
2. The study was a Phase 4, open-label, randomized non-inferiority trial including 1,200 patients enrolled throughout 16 countries. Hospitalized patients (n = 1,180) with a wound infection, cellulitis, abscess, infected ulcer, burn, or other soft tissue infection were randomized to receive linezolid via intravenous followed by change to oral administration (600 mg, q12h; n = 592) or vancomycin (1 g, q12h; n = 588) via intravenous administration followed by change to appropriate orally administered antibacterial drug for a total of 7–14 days of therapy. Other enrollment criteria included erythema ± induration (no size requirement), heat/localized warmth, and pain/tenderness or drainage/discharge plus evidence of one of the following systemic signs: fever, hypothermia, hypotension, and WBC count >10,000 mm^3 or > 15% immature neutrophils.
3. Scheduled patient visits were conducted at baseline, daily during the first 4 days of therapy, and at study Day 7. Patients who switched from intravenous to oral therapy received an additional visit at that time, plus visits at EOT and TOC (7 days after EOT).
4. Erythema was measured by both length and width; if applicable, induration length and width were also recorded on study visit case report forms. No information is available regarding measurement methodology. Separate length times width lesion measurements were analyzed as well as a "length and width" analyses of the ceftaroline data only; however, these data are unavailable until QC is performed.
5. The primary objectives of the analyses were (1) to define the time course for lesion measurements and (2) to fully characterize what constitutes a responder (success). Success is defined as a change from baseline of lesion size < x% at the 48- and 72-hour assessments compared to the previous lesion size measurement and no increase in lesion size at the 48- or 72-hour assessment. Accordingly, if x = 0%, the lesion size cannot have increased compared to baseline; if x = a negative percent (x < 0%), the lesion must have decreased compared to baseline; if x = a positive percent, the lesion size can be larger than baseline.

Results:

1. Demographics: Most patients were middle-aged men with cellulitis (5%) or major abscess (28%). Elevated WBC count (66%) was the most prevalent systemic sign in the study population, followed by fever (41%). The predominant local signs and symptoms were heat/localized warmth (93.4%) and pain/tenderness (92.2%). Fifty-five percent (55%) received antibiotics for the entire 14-day study course. Concomitant antibiotic use was allowed for treatment of gram-negative bacteria. In cases of methicillin-susceptible *S. aureus* (MSSA; 21.3%), vancomycin could be switched to oxacillin, nafcillin, flucloxacillin or

dicloxacillin. Other identifiable pathogens in patients included MRSA (29.6%) and *Streptococcus pyogenes* (4.2%).

2. The incidence of MRSA (29.65%) should not be interpreted as a less severe group of infections. It is a very aggressive pathogen, and patients can become very ill, even with small lesions. Also, there are distinctions between MRSA-induced abscesses and streptococcal cellulitis—lesions of cellulitis have larger surface areas. However, cellulitis is also commonly associated with abscess.
3. Regarding lesion size data, the mean erythematous area for all patients in the trial regardless of infection type was 400 cm^2, with a median area of 138 cm^2. Patients with cellulitis had the largest lesion area (592 cm^2), whereas abscesses were the smallest (184 cm^2). The lack of enrollment limitations with respect to lesion size contributed to large standard deviations. Induration measurements followed a similar pattern, with a mean area of 206.76 cm^2 among all patients, with the largest mean area in cellulitis (340.80 cm^2). In considering pathogen type, the highest mean area is found in *S. pyogenes* patients, followed by those infected with MSSA or MRSA.
4. In the box plot analyses, a definite downward trend in erythema area and induration over time is apparent, though not pronounced. In lesions < 75 cm^2 assessed by study day, the median erythema area progressively decreased, with outliers clustered in the 75% range. Induration lesion size > 75 cm^2 displays a similar downward trending pattern.
5. The evaluation of erythema area by study day also shows a downward trend for clinical successes. However, this trend is not apparent in the erythema area EOT analysis of clinical failures or indeterminate outcomes. Overall, the mean and median erythema and induration lesion size, as measured by area, length, or width decreases over the study periods examined. These result are consistent for lesion sizes greater or less than 75 cm^2; lesion size increased in a small patient subgroup (< 10%).

<u>Clarification questions and responses:</u>

1. EOT ranges from 7 to 14 days, but in patients considered a failure/indeterminate, the EOT is earlier. Lesion measurements from study Day 7 are not available.
2. Of concern, there are a large number of patients missing in the analysis of erythema/induration lesion measurements at later study days: the number changes substantially from baseline; e.g., 962 patients at Day 1 versus only 618 patients at Day 4. The loss of so many patients from analysis is an important consideration in designing future clinical trials. Some patients with missing data could be a failure, in which case the lesion size would have been larger. Failures that occurred on Day 3 or Day 4 were included in the analysis. However, if patients failed on Day 2, the lesion size measurement was removed from the analysis and not carried over to Day 3, subsequently lowering the overall numbers for Day 3 and Day 4. Other possible explanations include patients being switched to oral equivalents on an outpatient basis (hence not returning for subsequent study visits) or patients experiencing a successful outcome at Day 3.
3. The definition of clinical failure was not described in the presentation. (From the Weigelt paper, patients were "considered failed if they exhibited persistence or progression of the baseline clinical signs and symptoms of infection, development of new clinical findings consistent with active infection, or an inability to complete the study because of adverse events.") Also, it is not clear what proportion of patients had lesions that increased from Day 1 and decreased at Day 2. In general, the study was not designed specifically to examine

erythema measurements alone, although it is part of the study protocol. Clinical assessment at TOC is the primary objective for which measurements are available.

4. To better define what constitutes success for "control of lesion spread", the percent decrease in lesion spread "between" and "within" time points was examined. According to criteria established by the Project Team at the June 2, 2010 meeting, "success of control of lesion spread" includes: (1) change from baseline in lesion size of < x% at the 48- to 72-hour assessment and (2) compared to the previous lesion size measurement, no increase in lesion size at the 48- to 72-hour assessment. Increases in lesion size were easily identified using relative percent change in lesion size from study Day 3 versus Day 2, or study Day 4 versus Day 3. Assessments of decrease in lesion size were more problematic as evidenced by oscillating incremental changes. However, there were no increases in lesion size in patients responsive at EOT, though the analysis does not include missing data, which could confound interpretation of the data.
5. Sensitivity and specificity were examined by comparing no increase from baseline (as agreed to previously) as opposed to a baseline decrease where x = -5%. When the criterion is no increase in lesion size from baseline, the clinical response rate is 91.5%, whereas reducing the required change to -5% lowers the clinical success rate to 86.1%. Thus, increasing specificity modestly reduces the clinical response rate.
6. Regarding antibiotic use, patients without prior antibiotic exposure had a slightly better outcome (93%) with respect to achieving cessation of lesion spread than did those with previous antibiotic treatment (90%). Better outcomes were also observed in patients who had undergone incision and drainage than those who did not.
7. Based on erythema measurements, patient outcomes with regard to control of lesion spread were higher at study Day 4 than at study Day 3. Control of lesion spread, as determined by measuring the change in erythema from baseline to Day 3, produced higher outcomes than if one used measurement definitions established by the Project Team on June 2, 2010.
8. In maximizing sensitivity and specificity, one needs confidence in the endpoint of the clinical investigator's assessment, which is not clearly a gold standard, particularly since intra-operator reproducibility is unknown.
9. Whether measurement of induration is an accurate measurement of erythema is unclear, but induration does occur subsequent to erythema.
10. As summarized, success for control of erythema lesion spread is high (89%– 90%); however, there are drawbacks. The definition does not include missing patient data and does not consider patients with a prior history of antibiotic use as a treatment failure. Control of lesion spread (erythema) differs only slightly based on antibiotic use within the prior 14 days, and a greater proportion of patients who had undergone incision and drainage of cutaneous abscesses had control of lesion spread. When comparing day 3 assessments between patients who were considered failures versus successes at EOT, there was not a trend in differences between lesion size assessments at day 3.

8 Project Team Members

The conclusions described within this document represent the work of the FNIH Biomarkers Consortium Project "Developing Endpoints for Clinical Trials of Drugs for Treatment of Acute Bacterial Skin and Skin Structure Infections and Community-Acquired Bacterial Pneumonia (Phases 1 and 2)."

- Joseph Toerner, M.D., M.P.H. (Co-chair; FDA, non-voting)
- George H. Talbot, M.D. (Co-chair; IDSA)
- Jeff Alder, Ph.D. (Bayer)
- Paul G. Ambrose, Pharm.D. (Inst. for Clinical Pharmacodynamics; University of Oxford)
- Helen Boucher, M.D. (Tufts University)
- John Bradley, M.D. (University of California, San Diego)
- Laurie Burke, R.Ph., M.P.H. (FDA, non-voting)
- Aaron Dane, Ph.D. (AstraZeneca);
- Anita Das, Ph.D. (AxiStat)
- Dennis Dixon, Ph.D. (NIH/NIAID)
- Mike Dunne, M.D. (Durata Therapeutics)
- Barry Eisenstein, M.D. (Cubist)
- Thomas Fleming, Ph.D. (University of Washington)
- Dean Follmann, Ph.D. (NIH/NIAID)
- David Friedland, M.D. (Cerexa)
- Ian Friedland, M.D. (Cubist)
- Nickolas Kartsonis, M.D. (Merck)
- Achim Kaufhold, M.D., Ph.D. (Basilea)
- Scott Komo, Dr.P.H. (FDA; observer, non-voting)
- Mike Kurilla, M.D., Ph.D. (NIH/NIAID)
- Kim Lindfield, Ph.D. (Cubist)
- Lily Llorens, Ph.D. (Cerexa)
- Shawnmarie Mayrand-Chung, J.D., Ph.D. (FDA, non-voting)
- Susan Moriarty, M.D. (Cempra)
- Sumati Nambiar, M.D. (FDA; observer, non-voting)
- Gary Noel, M.D. (Paratek Pharmaceuticals)
- David Oldach, M.D. (Cempra)
- Elektra Papadopoulos, M.D. (FDA, non-voting)
- Roger Pomerantz, M.D. (Merck)
- John Powers, M.D. (SAIC in support of NIH/NIAID)
- Philippe Prokocimer, M.D. (Trius)
- John Quinn, M.D. (Pfizer)
- John Rex, M.D. (AstraZeneca)
- Jennifer Schranz, M.D. (Cempra)
- Judith A. Siuciak, Ph.D. (FNIH, non-voting)
- William Stubbings, Ph.D. (Basilea)
- David Wholley (FNIH, non-voting)

9 References

Boucher, H. W., G. H. Talbot, et al. (2009). "Bad Bugs, No Drugs: No ESKAPE! An Update from the Infectious Diseases Society of America." Clin Infect Dis **48**(1): 1-12.

Dieffenbach C.W., E.C.Tramont, S.F. Plaeger. (2010). Innate (General or Nonspecific) Host Defense Mechanisms, Chapter 4 from Mandell, Douglas, and Bennett's Principles and Practice of Infectious Diseases (7th Edition). Mandell G.L., J.E. Bennett, and R. Dolin, Eds. Churchill, Livingstone, Elsevier (Philadelphia, PA).

European Medicines Agency. (2011). Workshop on Antibacterials: Report of the Workshop. 17 March 2011. Retrieved from http://www.ema.europa.eu/docs/en_GB/document_library/Report/2011/04/WC500105473.pdf

Fleming, T.F. (2011). Addressing missing data in clinical trials. Ann Intern Med **154**: 113-117.

Food and Drug Administration, Center for Drug Evaluation. (2010, August). "Guidance for Industry: Acute Bacterial Skin and Skin Structure Infections: Developing Drugs for Treatment." Retrieved from http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm071185.pdf

Food and Drug Administration, Center for Drug Evaluation. (2010, October). "Guidance for Industry: Qualification Process for Drug Development Tools." Retrieved from http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM230597.pdf

Langemo, D.K., H. Melland, D. Hanson, B. Olson, S. Hunter, S.J. Henly. (1998). Two-dimensional wound measurement: comparison of 4 techniques. Adv Wound Care **11**(7): 337-43.

Micheel, C.M., & Ball J.R. (Eds.). (2010). Evaluation of Biomarkers and Surrogate Endpoints in Chronic Disease. Institute of Medicine: Committee on Qualifications of Biomarkers and Surrogate Endpoints in Chronic Disease. Available at http://www.nap.edu/catalog/12869.html

Moran, G.J., A. Krishnadasan, R.J. Gorwitz, et al. (2006). "Methicillin-resistant S. aureus among patients in the emergency department." N Engl J Med **355**(7): 666-74.

Nauseef W.M., R.A. Clark. (2010). Granulocytic phagocytes, Chapter 8 from Mandell, Douglas, and Bennett's Principles and Practice of Infectious Diseases (7th Edition). Mandell G.L., J.E. Bennett, and R. Dolin, Eds. Churchill, Livingstone, Elsevier (Philadelphia, PA).

Snodgrass, W.R., & T. Anderson. (1937). Prontosil in Erysipelas. Br Med J **2**(3993): 101-104.

Snodgrass, W.R., & T. Anderson. (1937). Sulphanilamide in the Treatment of Erysipelas. Br Med J **2**(4014): 1156-1159.

Spellberg, B. (2011). The tribulations of trials for antibacterial drugs. Clin Invest (in press).

Weigelt, J., K. Itani, D. Stevens, W. Lau, M. Dryden, C. Knirsch, the Linezolid CSSTI Study Group. (2005). Linezolid versus Vancomycin in Treatment of Complicated Skin and Soft Tissue Infections. Antimicrob Agents Chemother **49**(6); 2260-2266.