Talbot GH, Powers JH, Hoffmann SC. Developing Outcomes Assessments as Endpoints for Registrational Clinical Trials of Antibacterial Drugs: 2015 Update from the Biomarkers Consortium of the Foundation for the National Institutes of Health. Clin Infect Dis 2016.

Table 1: Exploratory Trials and Analyses of the Early Response Endpoint in the ABSSSI and CABP Indications

Indication	Methodology	Major Findings, Conclusions, and Recommendations	Date (Reference #) (found below)
ABSSSI	Double-blind, multicenter, phase 2 noninferiority study treated 161 patients for 7 to 14 days, testing the efficacy of JNJ-Q2 versus linezolid twice a day. *Post hoc* analyses were based on the 2010 FDA guidance, which defines treatment success as lack of lesion spread and afebrile status within 48	For the early response endpoint, JNJ-Q2 was statistically noninferior to linezolid (61.4% versus 57.7%, respectively; P = 0.024). Despite evidence of systemic disease, <5% of patients presented with fever, suggesting fever is not a compelling surrogate measure of systemic disease resolution for this indication.	2010 (1)

	to 72 h post-randomization.		
ABSSSI	Retrospective analysis to compare the outcomes of this Phase 2 study using criteria defined in prior and new FDA guidance. Patients were randomized 1:1:1 to 200 mg, 300 mg, or 400 mg oral tedizolid once daily for 5-- 7 days. Response to therapy was assessed examining cessation of spread of lesion and resolution of fever (<38degC) at the Day 3 visit (48 hours).	By Day 3, 172/181 (95%) patients showed no increase in lesion size from baseline. There was no difference in response by dose regimen. Patients with a baseline lesion surface area (length x width) ≥75 cm2 had a similar response at the Day 3 visit. Cessation of lesion spread and fever respond rapidly to therapy by Day 3.	2011 (2)
ABSSSI	Phase 3, randomized, double-blind study of 6-days tedizolid phosphate vs 10-days linezolid. Investigators	At Baseline, digital planimetry tends to provide area size measurements ~1/3 smaller than manual measurements. At the 48-72 hour visit the two methods yielded	2011 (3)

	were instructed to mark and outline the edge of erythema from severe abscesses, cellulitis and infected wounds using a surgical marker. Measurement of the erythema was to be in a head-to-toe orientation. Lesion surface area (cm2) was automatically calculated. Digital photographic images of the lesions were captured at the Screening, Day 3, and End-of-Therapy visits. If lesions were 3-dimensional in nature wrapping around the circumference of body parts (arms, legs, torso), then	comparable results in assessing cessation of spread of the lesion. High rates of response (as measured by cessation of spread) were observed at 48-72 hours, though higher for abscesses (~97%) and wound infections (~97%) than for cellulitis (~86%).

multiple images were to be captured perpendicular to the body surface. The digital images were uploaded by site personnel into the clinical trial database. Once uploaded, the digital images were organized and processed using Photoshop® image editing software. After the images were enhanced to provide the best possible image to analyze, PictZar®-CDM digital planimetry wound measurement software was used to assess lesion size. A comparison of the results of manual measurement at the bedside by site

	personnel versus an analysis of lesion size via digital planimetry is presented.		
ABSSSI	Retrospective analysis of the new FDA endpoint in a previously completed registrational trial (VER-009) to compare the FDA outcome to the protocol prespecified primary endpoint of clinical response at Day 28. The primary endpoint at Day 28 was originally defined in the clinically evaluable population which was then further analyzed in subgroups of	Dalbavancin noninferiority relative to linezolid as assessed by the prespecified primary analysis was reinforced with an early responder analysis performed at Day 3-4. The addition of resolution of fever to the early response definition did not further differentiate between treatment regimens beyond lesion measurement alone. Most patients who were an early responder were also a clinical success at End of Therapy. However, most patients who were an early nonresponder became a	2011 (4)

	patients meeting the newly defined FDA inclusion criteria (surface area of lesion >75cm2; one sign of either fever, elevated white blood cell count or bandemia) as well as the early response endpoint at Day 3/4 (lesion size the same or smaller relative to baseline and temperature <37.6°C).	clinical success at End of Therapy. The early response endpoint had limitations, specifically a low negative predictive value. A validated measure of response at End of Therapy and Test-of-Cure that is acceptable to FDA and other regulatory agencies would be more intuitive to clinicians and more relevant to patients.	
ABSSSI	Multicenter, randomized, double-blind, phase 2 US trial of adults with infection size ≥75cm2 of erythema or induration, lymph node enlargement, and one or more signs of systemic infection. Patients were	For the 256 patients randomized, the mean area of erythema of all infections was 292 cm2; 10% of patients had fever at baseline. Standard deviation for digital measurement was one third that of manual measurement. The three antibiotics were comparably efficacious. The	2012 (5)

	randomized 1:1:1 to receive delafloxacin, linezolid, or vancomycin for 5 – 14 days. The total area of erythema, and length of leading edge of major and minor axes of erythema, was measured digitally using acetate tracings and manually with a disposable rulers. A comparison in the treatment arms of the objective measures of response rate of either cessation of lesion spread or reduction of lesion spread and absence or resolution of fever was performed on the intent to treat population 48-72 hours after the first	time point of maximal benefit was 72 hours after initiation of therapy. Body temperature did not correlate with resolution or worsening of infection.

	dose of study drug by the Cochran-Mantel-Hanszel test.		
ABSSSI	Randomized, investigator-blind, multicenter phase 2 trial in which patients were randomized 1:1 to intravenous omadacycline or linezolid twice daily with an option to transition to oral therapy	In a retrospective analysis of the ITT patients for whom data were available to assess response during the first 72 h after starting therapy, including the cessation of an increase in maximal lesion dimension and the absence of fever (temperature < 38.0°C), 96.8% (30/31) of omadacycline-treated patients and 94.4% (34/36) of linezolid-treated patients met these two criteria.	2012 (6)
ABSSSI	20-center, randomized, double-blind, phase 2 study in which patients were randomized to receive either one of two different intravenous doses of BC-3781 or intravenous vancomycin q12 hour.	The early responder rate was lower when absence of fever was required for "success". Most early responders became a success at Test-of-Cure; the same was true for early non-responders. Only 4.8% of early responder successes became a failure at Test-of-Cure.	2012 (7)

	Early Response success or failure was determined using erythema length/width and temperature as recorded daily during the first 5 days of the study according to criteria from the FDA or the Foundation for the National Institutes of Health.		
ABSSSI	This was a multicenter, observational study in patients with an ABSSSI enrolled sequentially. All measurement/procedures were performed at the baseline visit. At least two, and not more than three, observers measured the erythema associated with the ABSSSI by	42 patients meeting criteria for an ABSSSI were enrolled; 39 were evaluated for efficacy. Overall, both the repeated (intra-) observer and between (inter-) observer ruler measurements showed almost perfect reliability with ICC's >0.9 for area, length and width. The percent differences in ruler measurements between different observers were slightly larger than between repeated measurements by the same individual (17.4%	2012 (8)

	ruler; correlation with other techniques such asdigital imaging, planimetric assessment of transparencies and thermal imaging was also performed. The primary endpoint was assessment of the intra-observer and inter-observer variability of the erythema as measured by ruler and determined by the intraclass correlation coefficient (ICC).	vs. 4.1%, respectively). Variability was higher in smaller lesions. There was almost perfect intra-observer reliability (ICC>0.9) comparing the longest and shortest lesion measurements by ruler compared to measurements derived programmatically from a computer-assisted planimetry assessment of lesion taken from the transparency grid. Lesions measured by digital camera and transparency were 35% and 31% smaller than the ruler, respectively. Measurement of the erythema associated with an ABSSSI can be reliably performed using a ruler.	
ABSSSI	Retrospective analysis using a day 3 clinical endpoint was conducted in	Day 3 integrated clinical response rates were 74.0% (296/400) for ceftaroline and 66.2% (263/397)	2012

	two Phase 3 trials of intravenous ceftaroline fosamil vs. vancomycin plus aztreonam q12h for 5 to 14 days. Clinical response at day 3, defined as cessation of infection spread and absence of fever, was analyzed in patients with a lesion size of >75 cm2 and either deep and/or extensive cellulitis, major abscess, or an infected wound.	for vancomycin plus aztreonam (difference, 7.8%; 95% confidence interval 1.3% to 14.0%). For ABSSSI due to MRSA, response rates were 81.7% and 77.4% in the ceftaroline and vancomycin plus aztreonam groups, respectively.	(9)
ABSSSI	These two Phase 3 trials were randomized, double-blinded studies in which patients with a skin infection lesion >75cm2 in area and either fever, WBC >12k cells/mm3	The majority (945/1046, 90.3%) of patients who responded favorably to treatment by 72 h were ultimately cured. Most (129/182, 70.9%) patients who were non-responders on Day 3 were also subsequently cured. 72/1046 patients (6.9%) who were	2013 (10)

	or bands >10% were randomized to receive either dalbavancin on Day 1 and Day 8 or vancomycin 1 gm every 12 h with an option to switch to linezolid po q12h after 3 days of therapy. Measurements obtained at 48–72 h recorded the cessation of spread of the lesion and absence of fever. Patients were not required to discontinue therapy based on these assessments. Clinical status was assessed programmatically based on the resolution of signs and	an early responder were a clinical failure at End of Treatment while 50/182 (27.5%) early non-responders were clinical failures at EOT. A measurement of early clinical response at 48–72 hours after initiation of treatment can help predict outcome and guide treatment of ABSSSI in addition to anchoring a non-inferiority trial design. All combinations of cessation of lesion spread and/or absence of fever had >90% sensitivity and could help identify patients who would ultimately be cured Cessation of lesion spread with an assessment of pain can

	symptoms of infection at End of Treatment on Day 14.	help predict those patients who might ultimately fail treatment	
ABSSSI	Prospective observational study of adult patients admitted to hospital with cellulitis in western Norway	Local response was recorded in 42% of cases at day 1. A decrease in leukocyte count ≥20% at day 1 was seen in 63% of these responders compared to 37% in cases with slower response (p=0.001). At day 2, 82% had local response, 59% of initially febrile patients had become afebrile, and 73% had ≥20% decline in CRP level. Logistic regression analysis identified symptom duration ≥2 days before admission as an independent predictor of local response at day 1 (Odds-ratio 2.60, p=0.031)."	2013 (11)
CABP	Two global, randomised, double-blind, multicentre trials assessed the efficacy and safety of ceftaroline fosamil vs ceftriaxone. The trials	A differential clinical benefit was observed at Day 4 in patients with CAP due to an atypical pathogen only, specifically *M. pneumoniae* and/or *C. pneumoniae*, and who received 1 day of clarithromycin (76.6% vs 57.6%).	2012 (12)

	were identical except for a 24-h course of clarithromycin on Day 1 in the first trial. This difference in design offered an opportunity to assess the effect of a macrolide on the outcome of CAP caused by >=1 atypical pathogens.	No difference was seen at the test-of-cure visit, perhaps due to the self-limited course of many atypical pneumonias. The addition of clarithromycin made no difference for typical pathogens, either at day 4 or test of cure. In patients with *L. pneumophila* only, a difference was not seen early, but a numerical difference favouring the clarithromycin group occurred at TOC. Although not a randomised comparison, our observations suggest that outcome assessment at an early time point may better identify differential effects of 2 treatments than a later evaluation and that empiric atypical coverage may impact early clinical response. Additional study is warranted.	
	Multicenter, double-blind, randomized phase 2 study in which	Patients were enrolled from 26 centers in the United States and 4 centers in Canada. A total of 132 patients	2013

	patients were randomized (1:1) to either 800 mg solithromycin orally on day 1, followed by 400 mg orally daily on days 2 to 5, or 750 mg levofloxacin orally daily on days 1 to 5. Early clinical response in the ITT population at day 3 was assessed programmatically. To be considered a success at day 3 using the FNIH criteria, patients had to report improvement in at least two cardinal symptoms (cough, chest pain, shortness of breath, and sputum production) without worsening in any of these four	received at least 1 dose of study drug. Early clinical response rates in the ITT population, from a *post hoc* analysis at day 3 by the FNIH criteria, were comparable in the solithromycin (72.3%) and levofloxacin (71.6%) treatment groups.	(13)

	symptoms.		
	A retrospective cohort study was conducted among adult patients who received ceftriaxone and azithromycin for CAP of Pneumonia Outcomes Research Team (PORT) risk class III and IV at an academic medical center. Clinical response was defined as clinical stability for 24 h with improvement in at least one pneumonia symptom and with no symptom worsening. A classification and regression tree	A total of 250 patients were included. In the CART analysis, adverse clinical outcomes were higher among day 5 nonresponders than those who responded by day 5 (22.4% versus 6.9%, **P**=0.001). The findings from this study indicate that time to clinical response, as defined by the recent FDA guidance, is a reasonable prognostic indictor of real-world effectiveness outcomes among hospitalized PORT risk class III and IV patients with CAP who received ceftriaxone and azithromycin.	2014 (14)

	(CART) was used to determine the delay in response time, measured in days, associated with the greatest risk of a prolonged hospital length of stay (LOS) and adverse outcomes (in-hospital mortality or 30-day CAP-related readmission).		
	The primary objective of this study was to assess health outcomes (length of stay [LOS] and hospital charges) between responders and nonresponders at day 4 of therapy. The Premier database was used to identify adult patients from 4 participating hospitals. Chart review	Of 666 patients who met study criteria, 277 (41.6%) achieved clinical response by day 4. The unadjusted mean (SD) LOS was 6.3 (2.8) days for responders and 7.4 (5.6) days for non-responders (*P* = 0.0009). Respective unadjusted total hospital charges were $22,827 (SD, $17,724) and $26,403 ($36,882) (*P* = 0.0031). Adjusted for demographics and clinical factors,	2014 (15)

	extracted data for demographics, clinical efficacy variables at day 4, LOS, and total hospital charges.	nonresponders compared with responders had an increased LOS of 0.9 days (8.4 vs 7.5 days; *P* = 0.0008), resulting in associated charges of approximately $2500 ($34,139 vs $36,629; *P* = 0.0768).	
CABP	Global, pivotal Phase 3 clinical trial of solithromycin vs moxifiloxacin, each administered orally for 5 days.	In the intent-to-treat population, solithromycin met the primary objective of statistical non-inferiority (10% non-inferiority margin) of the early clinical response at 72 (-12/+36) hours after initiation of therapy compared to moxifloxacin. The point estimates for the early clinical response were 78.2% for solithromycin and 77.9% for moxifloxacin (95% confidence interval for the treatment difference: -5.5% and 6.1%, respectively). Solithromycin also met the secondary objectives of non-inferiority in clinical success at the short term follow up visit, 5-10 days after the end of therapy.	2015 (16)

ABSSSI: Acute bacterial skin and skin structure infections

CABP: Community-acquired bacterial pneumonia

Citations

1. Covington P, Davenport JM, Andrae D, et al. Randomized, double-blind, phase II, multicenter study evaluating the safety/tolerability and efficacy of JNJ-Q2, a novel fluoroquinolone, compared with linezolid for treatment of acute bacterial skin and skin structure infection. Antimicrob Agents Chemother. **2011**; 55:5790-7. Epub September 26 2011.

2. DeAnda C, Das A, Fang E, Prokocimer P. Acute bacterial skin and skin structure infection (ABSSSI) dose-ranging phase 2 tedizolid phosphate (TR-701) study: Assessment of efficacy with new FDA guidance [Poster L1-1496]. In: Program and abstracts of 51st Interscience Conference on Antimicrobial Agents and Chemotherapy (Chicago). Washington, DC: American Society for Microbiology, 2011.

3. Bien P, DeAnda C, Fang E, Prokocimer P. Correlation between digital planimetry and a traditional manual method in the measurement of lesion size in a phase 3 acute bacterial skin and skin structure infection (ABSSSI) study [Poster L1-1494]. In: Program and abstracts of 51st Interscience Conference on Antimicrobial Agents and Chemotherapy (Chicago). Washington,

DC: American Society for Microbiology, 2011.

4. Dunne MW, Talbot GH, Das AF. Dalbavancin vs linezolid for treatment of acute bacterial infections of the skin: A comparison of early and standard outcome measures in Study Ver001-9 [Abstract P1530]. In: Program and abstracts of the 21st European Congress of Clinical Microbiology and Infectious Diseases / 27th International Congress on Chemotherapy (Milan)., 2011.

5. Longcor J, Lawrence L, Duffy E, Hopkins S. Objective measures of clinical efficacy in a phase 2b exploratory study of delafloxacin compared to vancomycin and linezolid in adults with acute bacterial skin and skin structure infections (ABSSSI) [Poster L1-1667c]. In: Program and abstracts of the 52nd Interscience Conference on Antimicrobial Agents and Chemotherapy (San Francisco), Washington, DC: American Society for Microbiology, 2012.

6. Noel GJ, Draper MP, Hait H, Tanaka SK, Arbeit RD. A randomized, evaluator-blind, phase 2 study comparing the safety and efficacy of omadacycline to those of linezolid for treatment of complicated skin and skin structure infections. Antimicrob Agents Chemother. **2012**; 56: 5650–5654.

7. Prince W, Obermayr F, Lell C, Das A, Talbot G. Assessment of two early-response outcome measures in a phase 2 clinical trial of the pleuromutilin BC-3781 in acute bacterial skin and skin structure infections [Poster P 693]. In: Program and abstracts of the 22nd European Congress of Clinical Microbiology and Infectious Diseases (London), 2012.

8. Dunne MD, Mehra P, Green S, et al. A study to assess skin lesion measurement techniques related to acute bacterial skin and skin structure infections. [Poster 1624]. In: Program and abstracts of IDWeek 2012 (San Diego), 2012.

9. Friedland HD, O'Neal T, Biek D, et al. CANVAS 1 and 2: Analysis of clinical response at day 3 in two phase 3 trials of ceftaroline fosamil versus vancomycin plus aztreonam in treatment of acute bacterial skin and skin structure infections. *Antimicrob Agents Chemother.* **2012**; 56: 2231–2236.

10. Dunne M, Puttagunta S, Wilcox M, Talbot G. Concordance of clinical response at 48-72 hours after initiation of therapy and end of treatment (EOT) in patients with acute bacterial skin and skin structure infection (abSSSI) in the DISCOVER studies. [Poster 1340]. In: Program and abstracts of IDWeek 2013 (San Francisco), 2013.

11. Bruun T, Oppegaard O, Skred S. Early treatment response in cellulitis. [Poster L-196]. In: Program and abstracts of the 53rd Interscience Conference on Antimicrobial Agents and Chemotherapy (Denver). Washington, DC: American Society for

Microbiology, 2013.

12. File T, Eckburg P, Low D, et al. Assessment of outcomes at an early time point may identify a differential effect of macrolide therapy on community-acquired pneumonia due to atypical pathogens. [Poster P721]. In: Program and abstracts of the 22nd European Congress of Clinical Microbiology and Infectious Diseases (London), 2012.

13. Oldach D, Clark, K, Schranz J, et al. Randomized, double-blind, multicenter phase 2 study comparing the efficacy and safety of oral solithromycin (CEM-101) to those of oral levofloxacin in the treatment of patients with community-acquired bacterial pneumonia. Antimicrob Agents Chemother. **2013**; 57:2526-34. Epub March 18, 2013.

14. Zasowski E, Butterfield JM, McNutt L-A, et al. Relationship between time to clinical response and outcomes among Pneumonia Outcomes Research Team (PORT) risk class III and IV hospitalized patients with community-acquired pneumonia who received ceftriaxone and azithromycin**.** Antimicrob Agents Chemother**. 2014**; 58:3804-13. Epub 2014 Apr 21.

15. Robinson SB, Ernst FR, Lipkin C, Huang, X. Patient outcomes on day 4 of intravenous antibiotic therapy in non–intensive care unit hospitalized adults with community-acquired bacterial pneumonia. Infect Dis Clin Pract. 2014; 22: 320–325.

16. Cempra Press Release Chapel Hill, NC (January 4, 2015). Cempra announces positive topline phase 3 clinical results for oral solithromycin in the treatment of community acquired bacterial pneumonia. http://investor.cempra.com/releasedetail.cfm?ReleaseID=889300. Accessed 1 February 2015.

Talbot GH, Powers JH, Hoffmann SC. Developing Outcomes Assessments as Endpoints for Registrational Clinical Trials of Antibacterial Drugs: 2015 Update from the Biomarkers Consortium of the Foundation for the National Institutes of Health. Clin Infect Dis 2016.

Table 2: FNIH Publications of Patient-reported Outcome Measure Development in the ABSSSI and CABP indications*

Indication	Objective	Methodology	Major Findings, Conclusions, and Recommendations	Date (Reference #) (found below)
ABSSSI	Literature review to identify clinical measurement concepts for other skin abnormalities that could be applied to ABSSSI.	A search was conducted in OVID. MEDLINE (1946-present) and EMBASE (1988-2012) were searched using terms related to ClinRO, measurement tools and devices, diagnostic tools, and skin diseases/abnormalities.	Of 428 abstracts. 381 were excluded based on pre-specified criteria. The remaining 47 full-text articles were scrutinized for eligibility, resulting in 30 that met the inclusion criteria. No ABSSSI-specific ClinROs or measurement tools were identified in the literature.	2013 (1)
ABSSSI	Literature review to investigate existing PRO measures used in	A search was conducted in OVID. MEDLINE (1946 to present) and EMBASE (1988 to 2012) were		2013 (2)

	studies of ABSSSI and to explore signs and symptoms of ABSSSI in order to inform the development of a disease model.	searched using terms for signs and symptoms and existing PRO measures, specifically related to skin infections.		
ABSSSI	Development and qualification of a new ABSSSI PRO instrument incorporating reliable, well-defined and relevant endpoints for patients in terms of how they feel and function	We adhered to the U.S. Food and Drug Administration (FDA) PRO Guidance for instrument development (2009) and the 2010 FDA qualification process for drug development tools (DDTs). Once qualified, drug developers can use DDTs for the qualified context in Investigational New Drug and	The initial phase of instrument development included a literature review and gap analysis and interviews with 9 clinical experts. The most commonly reported symptoms were pain and tenderness across all ABSSSI subtypes-cellulitis, wound infection, and abscess. A study protocol and	2013 (3)

	in clinical trials of antibacterial drugs for ABSSSI.	New Drug Application /Biological License Application submissions without FDA reconsideration of the DDTs' suitability.	interview guide were developed to elicit concepts from ABSSSI patients. A draft PRO will be evaluated by an expert panel and refined through cognitive debriefing interviews with patients.	
ABSSSI	Explore concepts of skin infection as reported by patients, and comprehensively capture these symptoms.	One-on-one telephone interviews were conducted with ABSSSI patients diagnosed within the past 4-7 days in the United States. Patients were asked to describe their skin infection and how it may have affected their life. The data were continually analyzed using	Thirty-four patients participated in concept elicitation interviews from four sites. Thirteen patients were diagnosed with a major abscess, twelve with wound infection, and nine with cellulitis. The main themes included signs (e.g., growth, color), symptoms (e.g., soreness), and	2013 (4)

		an iterative process to identify themes and concepts which were recorded in a saturation grid. Saturation was monitored according to the FDA PRO guidance.	impacts on functionality (e.g., social, physical) related to the skin infection. The most commonly reported symptoms included experiencing pain, swelling, and drainage or leakage at the site of the infection.	
ABSSSI	Develop a PRO instrument to assess ABSSSI symptoms in patients in clinical trials of antibacterial drugs, consistent with FDA PRO Guidance.	A comprehensive review of the literature and interviews with nine US and European clinical experts informed the development of a concept elicitation (CE) interview guide, and a hypothetical conceptual framework and disease model exploring patients'	Symptoms were common across all ABSSSI subtypes and supported the saturation of concepts. Items were generated for the PRO Instrument using patient terminology. Subsequent cognitive debriefing with patients demonstrated that the items	2015 (5)

	experience with symptoms of ABSSSI. CE was based on telephone interviews with 34 patients, after which saturation of emergent concepts was reached. Items and response options were generated based on the qualitative data and a draft instrument was prepared with input and review from an international project team of academic and industry antibacterial experts. Subsequently, cognitive debriefing interviews were conducted with 15 ABSSSI patients and 3 clinical	were understandable, relevant, and interpreted as intended. SKINFECT is a PRO instrument developed to evaluate ABSSSI patient symptoms and functioning in clinical studies with documented evidence of content validity. SKINFECT is now ready for psychometric reliability and validity testing.

		experts to assess item readability, relevance, comprehensiveness, and content validity. Items were edited based on feedback from the patients.		
CABP	Literature review to identify signs, symptoms, and measurement tools associated with patient experience of CABP.	A search was conducted using OVID. MEDLINE (1946-present) and EMBASE (1988-2012) using terms related to signs and symptoms of CABP and existing measurement and diagnostic tools.	Of 2158 abstracts, 940 were excluded based on pre-specified criteria. Of the remaining 1218, 39 met the inclusion criteria. Thirty-four articles focusing on CABP signs and symptoms were identified. The most commonly reported symptoms were cough, chest pain, dyspnea, sputum production, and fatigue.	2014 (6)

			Generic PRO instruments and an interviewer-administered measure including 10 CABP symptoms have been used in CABP studies. Four CABP-specific instruments that assess patient-reported symptoms revealed notable methodological limitations; these were developed prior to the FDA PRO Guidance.	
CABP	Development and qualification of a new CABP PRO instrument incorporating reliable, well-defined, and	We adhered to the FDA PRO Guidance for instrument development and the 2010 FDA qualification process for DDTs.	The initial phase of instrument development included a literature review, a gap analysis, and interviews with six clinical experts. The most commonly reported	2014 (7)

	relevant endpoints for patients in terms of how they feel and function in clinical trials of antibacterial drugs for CABP		symptoms identified were cough, chest pain, dyspnea, sputum production, and fatigue. These findings informed the development of a study protocol and interview guide to elicit concepts from CABP patients. A draft PRO will be evaluated by an expert panel and refined through cognitive debriefing interviews with patients.	
CABP	Explore CABP symptoms as reported by patients, and to develop a draft PRO	Concept elicitation was conducted by telephone interviews with patients within 10 days of CABP diagnosis. Data were analyzed	Twenty patients participated in concept elicitation interviews. The most common symptoms reported included a lack of energy or	2015 (8)

	instrument designed to comprehensively assess these symptoms.	using an iterative process to identify themes and concepts and was recorded in a saturation grid. Saturation was monitored according to the FDA PRO guidance. Using this qualitative data, a draft PRO instrument was prepared. Cognitive debriefing interviews were conducted to assess item readability, relevance, comprehensiveness, and content validity.	tiredness, cough, and shortness of breath. Nearly half the patients also reported fever, chest pain and general aches/pain as well as significant impacts on their social and physical functioning. Subsequent cognitive debriefing in 9 patients and 3 clinical experts demonstrated that the items were understandable, relevant, and interpreted as intended. These patient-reported CABP symptoms were shown to demonstrate content saturation and concept validity and provide unique

			information important for both comprehensive evaluation of individuals with CABP and evaluation of new antibacterial treatments.	
HABP	Literature review was to identify signs, symptoms, and measurement tools associated with patients' experience of HABP.	MEDLINE (1946 to 2014) and EMBASE (1988 to 2014) databases were searched individually and in combination using terms related to Hospital-Acquired Pneumonia (HAP), HABP, signs and symptoms, and patient-reported outcomes.	Of 1384 abstracts, 225 were excluded as duplicates or for missing content and a further 1145 based on pre-specified criteria. Six articles met the inclusion criteria. The most frequently cited signs and symptoms of HABP were fever, cough, purulent sputum, dyspnea, rales, chest pain, and elevated respiratory rate. No PRO measures for assessing HABP	2015 (9)

	signs and symptoms were identified in the literature. Current HABP clinical trials have not included end points that directly measure how a patient feels and functions.

* In collaboration with Oxford Outcomes Research of ICON Plc.

FNIH: Foundation for the National Institutes of Health

ABSSSI: Acute bacterial skin and skin structure infections

CABP: Community-acquired bacterial pneumonia

HABP: Hospital-acquired bacterial pneumonia

Citations

1. Cimms TA, DeBusk K, Howard K, et al. Clinical measurement concepts in acute skin and skin structure infections and other skin abnormalities: A comprehensive literature review [Abstract PRM104]. In: Program and abstracts of the International Society for Pharmacoeconomics and Outcomes Research 18th Annual International Meeting (New Orleans), 2013.

2. Cimms TA, DeBusk K, Howard K, et al. Acute bacterial skin and skin structure infections (ABSSSI) signs/ symptoms and PROs: A comprehensive literature review. [Abstract PRM110]. In: Program and abstracts of the International Society for Pharmacoeconomics and Outcomes Research 18th Annual International Meeting. (New Orleans), 2013.

3. Powers JH, Siuciak JA, Llorens L, et al. Development of a new patient reported outcome (PRO) measure for acute bacterial skin and skin structure infections (ABSSSI). [Abstract PRM181]. In: Program and abstracts of the International Society for Pharmacoeconomics and Outcomes Research 18th Annual International Meeting (New Orleans), 2013.

4. Powers JH, Portalupi S, Devine J, et al. Acute bacterial skin and skin structure infections (ABSSSI): Development of a new patient reported outcome (PRO). [Abstract PSS31]. In: Program and abstracts of the International Society for Pharmacoeconomics and Outcomes Research 18th Annual International Meeting (New Orleans), 2013.

5. Powers JH, Howard K, Saretsky T, et al. Development of a patient-reported outcome instrument (Skinfect-PRO) to standardize and qualify symptoms of acute bacterial skin and skin structure infection (ABSSSI) [Abstract PIN85]. In: Program and abstracts of the International Society for Pharmacoeconomics and Outcomes Research 20th Annual International Meeting (Philadelphia), 2015.

6. Cimms TA, Howard K, Portalupi S, et al. Signs, symptoms, and existing patient reported outcome (PRO) measures In community-acquired bacterial pneumonia (CABP): A comprehensive literature review. [Abstract PIN84]. In: Program and abstracts of the International Society for Pharmacoeconomics and Outcomes Research 19th Annual International Meeting (Montreal), 2014.

7. Howard K, Portalupi S, Hoffmann S, et al. Development of a new patient-reported outcome (PRO) measure for community-acquired bacterial pneumonia (CABP). [Abstract PIN85]. In: Program and abstracts of the International Society for Pharmacoeconomics and Outcomes Research 19th Annual International Meeting (Montreal), 2014.

8. Howard K, Clifford S, Powers JH, et al. Community-acquired bacterial pneumonia (CABP): Development of a new patient-reported outcome (PRO) measure [Abstract PIN84]. In: Program and abstracts of the International Society for Pharmacoeconomics and Outcomes Research 20th Annual International Meeting (Philadelphia), 2015.

9. Saretsky T, Clifford S, Hoffmann SC, et al. Signs, symptoms, and existing patient-reported outcome (PRO) measures in hospital-acquired bacterial pneumonia (HABP): A comprehensive literature review [Abstract PIN86]. In: Program and abstracts of the International Society for Pharmacoeconomics and Outcomes Research 20th Annual International Meeting (Philadelphia), 2015.

FDA Definition of a Patient-Reported Outcome

"A PRO is any report of the status of a patient's health condition that comes directly from the patient, without interpretation of the patient's response by a clinician or anyone else. The outcome can be measured in absolute terms (e.g., severity of a ***symptom***, ***sign***, or state of a disease) or as a change from a previous measure. In clinical trials, a PRO instrument can be used to measure the effect of a medical intervention on one or more ***concepts*** (i.e., the *thing* being measured, such as a symptom or group of symptoms, effects on a particular function or group of functions, or a group of symptoms or functions shown to measure the severity of a health condition)."

Citation: US Food and Drug Administration, Center for Drug Evaluation. (2009, December). Guidance for Industry and FDA Staff: Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims. http://www.fda.gov/downloads/Drugs/Guidances/UCM193282.pdf Accessed 29 March 2015.

(page 2)

Talbot GH, Powers JH, Hoffmann SC. Developing Outcomes Assessments as Endpoints for Registrational Clinical Trials of Antibacterial Drugs: 2015 Update from the Biomarkers Consortium of the Foundation for the National Institutes of Health. Clin Infect Dis 2016.

Table 3: FNIH Project Milestones and Deliverables

Acute Bacterial Skin and Skin Structure Infections (ABSSSI)	
Working Group Endpoint Initiation Meeting	May 5, 2010
Recommendations to the FDA for Interim Endpoints for Clinical Trials: Submission to Docket #FDA-2013-N-0556[1]	August 26, 2011
ABSSSI Endpoint and Patient-reported Outcome (PRO) Development Project Launch	July 27, 2012
DDT Qualification Letter of Intent: PRO Measure for Symptoms of ABSSSI	October, 4, 2012
Initial DDT Qualification Submission: ABSSSI PRO Interim Briefing Package (DDT COA 000018)	June 18, 2013
[Final Guidance for Industry Acute Bacterial Skin and Skin Structure Infections: Developing Drugs for Treatment][2]	October 2013
Draft ABSSSI PRO (Skinfect-PRO) Completed	November 11, 2015

Final DDT Qualification Submission: ABSSSI PRO Interim Briefing Package (DDT COA 000018)	January 23, 2015
Community-acquired Bacterial Pneumonia (CABP)	
Working Group Endpoint Initiation Meeting (with ABSSSI)	May 5, 2010
Recommendations to the FDA for Interim Endpoints for Clinical Trials: Submission to Docket #FDA-2009-D-0136[1]	August 26, 2011
CABP Endpoint and PRO Development Project Launch	January 29, 2013
DDT Qualification Letter of Intent: PRO Measure for Symptoms of CABP	April 2, 2013
Initial DDT Qualification Submission: CABP PRO Interim Briefing Package (DDT COA 000019)	July, 26, 2013
[Draft Guidance for Industry Community-Acquired Bacterial Pneumonia: Developing Drugs for Treatment][2]	January 2014
Draft CABP PRO (CAPBAC-PRO) Completed	January 19, 2014
Final DDT Qualification Submission: CABP PRO Interim Briefing	April 1, 2014

Package (DDT COA 000019)	
Hospital-acquired Bacterial Pneumonia (HABP) and Ventilator-associated Bacterial Pneumonia (VABP)	
Working Group Endpoint Initiation Meeting	November 1, 2012
Interim Considerations for Clinical Trial Design: Submission to Docket #FDA-2013-N-05562	July 15, 2013
HABP PRO White Paper (FDABAA-13-00119): Development of a Patient Reported Outcome Instrument in Hospital-Acquired Bacterial Pneumonia	August, 9 2013
[Draft Guidance for Industry Hospital-Acquired Bacterial Pneumonia and Ventilator-Associated Bacterial Pneumonia: Developing Drugs for Treatment][2]	May 2014
DDT Qualification Letter of Intent: PRO Measure for Symptoms of HABP	December 4, 2014
HABP/VABP Endpoint Development Project Launch	December 16, 2014

HABP PRO Development Project Launch	October 22, 2014

ABSSSI: Acute bacterial skin and skin structure infections

CABP: Community-acquired bacterial pneumonia

HABP: Hospital-acquired bacterial pneumonia

VABP: Ventilator-associated bacterial pneumonia

[1]Talbot et al., Clin Infect Dis. **2012** Oct;55(8):1114-21 and Toerner et al., Clin Infect Dis. **2012** Oct;55(8):1122-3.

[2]FDA-Issued Guidance Document

Talbot GH, Powers JH, Hoffmann SC. Developing Outcomes Assessments as Endpoints for Registrational Clinical Trials of Antibacterial Drugs: 2015 Update from the Biomarkers Consortium of the Foundation for the National Institutes of Health. Clin Infect Dis 2016.

Interim Considerations for Clinical Trial Design for the Study of Hospital-Acquired Bacterial Pneumonia and Ventilator-Associated Bacterial Pneumonia

Foundation for the National Institutes of Health

Biomarkers Consortium HABP/VABP Working Group

July 15, 2013

For submission to Docket #FDA-2013-N-0556

Executive Summary

- At FDA's request, this Working Group has been constituted to provide recommendations to support FDA's goal of articulating scientifically rigorous and clinically relevant hospital-acquired bacterial pneumonia (HABP)/ventilator-associated bacterial pneumonia (VABP) drug development based on a non-inferiority (NI) design.
- Despite the potential clinical trial implementation feasibility issues that have been raised with current FDA HABP/VABP Guidance, including an all-cause mortality (ACM) endpoint, most Working Group participants are comfortable with ACM as an endpoint, especially for VABP, if trial feasibility could be addressed by changing other parameters of study design.
- The outstanding questions for use of ACM relate to timing of its assessment, as well as to whether there are suitable intermediate clinical endpoints. One concern with ACM is its lower incidence in registrational trials versus "real life." It is hypothesized that making exclusion criteria less restrictive, and thereby increasing the severity of illness in the enrolled population, has the potential to facilitate enrollment. The past practice of excluding those more ill patients who could have a poor outcome exacts a cost to a study of limited enrollment, lower ACM, and decreased

generalizability. The practical consequence of this challenge is that when the mortality rate for enrolled subjects falls much below 15% to 20%, trial sizes rapidly enlarge based on a change to the odds ratio metric. It is anticipated that enrolling a population with increased severity of illness would make trials more broadly generalizable while decreasing sample size of trials based on a risk-difference metric.

- All-Cause Mortality in VABP could be evaluated at day 14, or at day 28, or sometime in between (e.g., day 21); rates of ACM would be expected to be 10% to 15% (e.g., for day 14 ACM) and 20% to 25% (e.g., for day 28 ACM).
- For VABP sample size estimation and analyses, when mortality is at least 15% on the active control regimen, a risk-difference metric with an NI margin of 10% could be used.
- For HABP, a clinically meaningful endpoint of symptom improvement plus survival for non-ventilated patients could be based on the historical data for community-acquired bacterial pneumonia, for which there is a large treatment effect to day 7 of antibacterial drug therapy.
- There was some concern in the Working Group that mortality and other differences between HABP and VABP suggest these are different diseases, meaning that combining both in a single trial could raise methodological issues.
- A number of candidate changes to other aspects of trial design (e.g., primary analysis set) were identified as promising potential approaches to improving feasibility, while maintaining scientific validity.

- The Working Group remains very interested in evaluating the potential application of alternative endpoints (e.g., improved oxygenation) for VABP and considering how they could be evaluated and qualified as endpoints.
- The next step for the Working Group is to examine the data from a single HABP/VABP trial, contributed in kind, to understand the data that are available (e.g., prevalence of respiratory symptoms at trial enrollment and their severity, mortality rate over time).
- Subsequently, a formal statistical analysis plan will be drafted; data from a number of HABP/VABP trials contributed in kind will be analyzed, with the results used to inform the Working Group's final recommendations to FDA.

Background

At FDA's request, this Working Group has been constituted to provide recommendations to support FDA's goal of articulating scientifically rigorous and clinically relevant HABP/VABP drug development guidance that is also feasible for sponsors to implement in terms of both financial cost and time. The HABP/VABP Working Group is building upon the work of the FNIH Biomarkers Consortium Community-Acquired Bacterial Pneumonia (CABP)/Acute Bacterial Skin and Skin Structure Infections (ABSSSI) Project Team. Forthcoming recommendations will be based on an evidenced-based, hypothesis-testing endeavor through analysis of HABP/VABP clinical trial data contributed in kind. In addition to FDA, other likely beneficiaries include clinicians, investigators, and patients.

Working Group Goal and Processes

The goal of the HABP/VABP Working Group's efforts is to identify potential changes to study design and analysis that could improve the feasibility—while retaining reliability, scientific validity, and meaningfulness for patients, caregivers, and clinicians—of HABP/VABP registrational clinical trials based on an NI design. HABP/VABP registrational trials based on a superiority design, as for narrow-spectrum antimicrobials intended to treat uncommon, multidrug-resistant pathogens, are not a focus of this group's initial deliberations.

This goal will be achieved via the following process:

- Evaluating the medical literature to determine those HABP/VABP signs, symptoms, and measures of function that are clinically relevant to treatment outcome, including mortality;
- Identifying feasibility constraints imposed by clinical trial requirements other than the choice of the primary endpoint (e.g., definition of statistical analysis populations);
- Determining whether it is possible to identify non-mortality endpoints, with specific regard to their potential use as the primary endpoint or as part of a composite primary endpoint. Using hypotheses generated based on the medical literature followed by examination of data from modern-day clinical trials, work on this question will focus specifically on defining the variables in such endpoints and quantifying a treatment effect on how patients feel and/or function; and
- Performing sensitivity analyses to understand how certain assumptions (e.g., day of endpoint assessment) impact these parameters.

Primary Endpoint: All-Cause Mortality vs. Non-Mortality vs. Composite (ACM Plus Non-Mortality)

- In a review of the literature for non-mortality clinical endpoints in HABP/VABP, FDA found 16 papers providing historical evidence for sensitivity to drug effects on the ACM endpoint. Only two papers described non-mortality outcome measures: Luna et al. [1] showed that serial Clinical Pulmonary Infection Score (CPIS) values did not improve among non-survivors and that PaO2/FiO2 did not improve or worsened among non-survivors, and Dennesen et al. [2] showed that PaO2/FiO2 correlated with clinical resolution. A discussion of non-mortality measures at the 2009 FDA co-sponsored workshop included PaO2/FiO2, time-on-ventilator (for VABP), and time-in-hospital. Subsequently, Esperatti et al. [3] found that an increase in the Sequential Organ Failure Assessment score from day 1 to day 5 of treatment and lack of improvement in PaO2/FiO2 were independently associated with increased 28-day mortality.
- In November 2011, the Anti-Infective Drugs Advisory Committee reviewed the HABP/VABP mortality endpoint based on four studies with mortality data at days 14, 21, and 28 [4]. In this set of data, a higher Acute Physiology and Chronic Health Evaluation II (APACHE II) score at baseline correlated with a higher mortality rate. The day 14 mortality rate is no more than 10% lower than 28-day mortality in a combined population. In HABP, mortality in treated patients ranged between 14% and 17% at day 14, and as expected, the VABP 14-day rate in treated patients was somewhat higher at 10% to 20%.

- However, data from trials recently conducted by one sponsor suggest a lower mortality rate can be observed (Rebecca Redman, written communication, June 2013). Evaluation of the data sets contributed in kind to the FNIH as part of this Working Group initiative should elucidate the mortality rate that has been observed in a broad range of modern-day trials using current enrollment criteria.
- The principal strengths of ACM are its simplicity of measurement (decreasing missing data) and unequivocal clinical relevance. ACM is, and will always remain, an acceptable choice as the primary endpoint in any nosocomial pneumonia trial. In addition, other endpoints cannot be evaluated independent of mortality since patients must be alive in order for measurements to be obtained (i.e., patients are not excluded from analysis due to death). Also, there is a clear and large treatment effect of antibiotics on ACM that justifies the NI trial design with an ACM endpoint.
- Concerns raised with ACM as an endpoint may include both competing risks (causes) of death and the timing of assessment. Potential concerns of some participants are that in some patients death may be caused a) by comorbidities that cannot be resolved by antimicrobial therapy and/or b) by withdrawal of care (especially for VABP). Some participants suggested that a day 28 assessment might enhance these potential concerns. Other participants countered that while these concerns would be relevant for superiority trials against active control antibiotics, they are not influential in the NI setting

because the historical evidence, despite its modest size, establishes that the treatment effects of antibiotics are very large even in the presence of competing risks.

- Concerns were also raised that a smaller clinical trial sample size could result in meaningful imbalances between treatment groups for variables that impact mortality such as chronic health conditions, acute comorbid diseases, and pharmacological interventions. Other participants countered that such concerns would apply to all endpoints, not only to ACM.
- Some participants noted that while it is always important to measure mortality—and certainly a drug should not increase mortality—alternative endpoints can also provide important indirect or direct insights about benefits for patients. Data from recent HABP/VABP registrational trials provide evidence of a non-mortality endpoint that is sensitive to treatment effect and is correlated with ACM [5] [6] [7], although substantive evidence currently is not available to determine whether an estimated treatment effect on that non-mortality endpoint would provide reliable insights about the true effect of treatment on direct measures of how a patient feels, functions, or survives [8].
- A composite endpoint that includes mortality could be clinically relevant; reflect how a patient feels, functions, or survives; and allow for more feasible clinical trials, although some noted that such an approach could have the undesirable effect of increasing clinical trial sample sizes and would still require justification of an NI margin.

- Global harmonization has suffered since, in general, the European Medicines Agency remains focused on non-mortality endpoints.

Statistical Considerations and Feasibility

- The decision to use either an odds-ratio or a risk-difference approach will have an impact on sample size requirements. If a fixed risk-difference margin is used, it will be important to indicate the lowest mortality rate at which the fixed risk difference would still be able to reliably demonstrate NI.
- The ACM endpoint is well defined, reliable, and clinically meaningful. Its strong level of clinical relevance justifies requiring only a single trial for registration, which in itself increases the feasibility of clinical trial conduct for this indication. Furthermore, when using absolute differences for an ACM endpoint, there is evidence-based justification for a 10% margin when mortality is at least 15% on the active control. Data reviewed by FDA from prior VABP trials confirm mortality should be in the range of 15% to 24% in the interval between day 14 to day 28. With the 10% margin and an active control survival of 15%, the total sample size would be approximately 544 patients in an intent-to-treat (ITT) analysis. Positive results would be obtained if the estimated ACM on the experimental regimen did not exceed that on the active control by at

least 3.5%. ACM also allows cost savings due to the simplicity of measuring that endpoint and enhances trial integrity by reducing the risk of missing data.

- However, if a microbiologically confirmed ITT (micro-ITT) primary analysis population is required, the required number of enrolled patients becomes much larger.
- The question of feasibility of various endpoints and trial design characteristics remains the subject of active debate in the Working Group. For example, whether a 500-patient sample size is achievable depends on the type of patient and other factors such as enrollment criteria and choice of analysis population; e.g., 500 PORT V CABP patients would be difficult to enroll. Other important impacts on feasibility include the ability to capture information and subsequent missing data. For example, it is likely not feasible to collect information on a biomarker every hour in an intensive care unit setting. One benefit of the ACM endpoint is that it is an easy endpoint to capture, so discussion of a composite endpoint should include consideration of the feasibility of assessing the chosen endpoint and risk of missing data for that endpoint. It is expected that analysis of actual clinical trial data will facilitate consensus-building on these important considerations.
- Other elements of NI trial design beyond choice of endpoint impact feasibility: the NI margin, the primary analysis population to determine treatment effect, and the inclusion/exclusion criteria.

- While the traditional outcome measure of clinical response is perceived by some experts as a subjective endpoint that is an indirect measure of patient benefit and currently cannot be used to set an NI margin, stakeholders agree that ACM is evaluated in any trial. ACM is a direct measure of patient benefit for which evidence exists to justify an NI margin. Further, ACM is a valid endpoint in that it is well defined and reliable and characterized by minimal ascertainment and measurement bias.
- Concerns about a lack of clinical trial feasibility based on the sample size required for an ACM endpoint may be minimized if this endpoint is analyzed using the ITT as the primary analysis set.
- Other endpoints may be considered, such as 14-day (or 21-day) ACM or a composite index, with the requisite requirements for validation and determination of the relationship between indirect and direct measures as well as justification of an NI margin. With regard to the NI margin, the 10% margin proposed by FDA is based on ACM in the ITT, not the micro-ITT, analysis population.
- Either a 14-day or a 28-day evaluation is likely an informative time point for ACM. At 14 days, the proportion of deaths from "other causes" (e.g., comorbid conditions) may be lower. However, the mortality rate will be lower at 14 days than at 28 days, and this may lead to the need for a larger study if the odds-ratio metric is used. Furthermore, for a marginally effective antibacterial drug, the time to death from inadequate treatment of infection

could be delayed beyond 14 days. If a 14-day ACM mortality assessment is chosen as the primary endpoint for efficacy, evaluation of 28-day ACM would still be a key secondary efficacy endpoint and an important safety assessment.

- Selection of the primary analysis set may well represent the greatest opportunity for decreasing sample size while maintaining scientific validity. The primary analysis set is defined as the population with the disease of interest in which a treatment effect will be evaluated; however, there are no "gold standard" criteria for the diagnosis of VABP. Diagnosis by lung histopathology and respiratory tract culture may lack sensitivity and specificity and, accordingly, may not add value to the diagnosis of VABP. Issues with microbiological confirmation include a large number of different sampling techniques, a high rate of false-positives, inability to distinguish between colonizing and pathogenic bacteria, and the lack of quantitative biotechnology sophistication at certain sites involved in a global trial. Thus, microbiological documentation is viewed by some as an imperfect clinical tool that justifies choice of an antibiotic against the possible causative pathogen isolated from the respiratory tract but not as a diagnostic tool for the disease of nosocomial pneumonia in a registrational trial. However, others expressed concern with this approach, citing a recent analysis showing that patients with microbiologically documented HABP/VABP have different baseline characteristics and different mortality rates than those without such documentation [9].

Specifically, following adjustment for potential confounders, patients with positive microbiology at baseline had higher hospital mortality and lower 90-day survival but, notably, a non-significantly lower 28-day survival. A proposed approach, balancing these varied concerns, suggests that with the present state of the art regarding microbiological diagnosis (e.g., availability of results with some meaningful delay), it makes sense to assess ACM primarily in the ITT analysis set but to place a substantial emphasis on results in the micro-ITT population. One option suggested is to require that a minimum percentage of the ITT population be microbiologically documented (e.g., 50% of ITT be micro-ITT), the results of which would be expected to be consistent with those in the primary ITT analysis.

- Accordingly, various values of the microbiological evaluability rate (the 50% rate noted above as well as alternative values) will be considered during the Working Group's data analysis, based upon previously observed data, and performance characteristics related to the expected sample size in a microbiologically confirmed population (i.e., uncertainty in the estimates received) will be used to provide a recommendation.
- Regardless, if employing ITT as the primary analysis population, it will be essential to ensure that the patients enrolled do not have an etiology other than HABP/VABP (e.g., pulmonary edema or venous thromboembolism).

- Overall, considering the ITT population as the primary efficacy set found broad support within the Working Group, with the micro-ITT population as a key secondary subset for sensitivity analyses.
- Using ITT as the primary efficacy set is most logical when the test agent is reasonably likely to have utility in the bulk of the enrolled population. For agents with a reasonably broad spectrum (e.g., an agent active against most gram-negative pathogens), this is a good assumption. For a narrow-spectrum agent (e.g., an agent active only against *Acinetobacter spp.*), population enrichment via enrollment based on rapid and sufficiently predictive microbiologic tools would seem necessary.
- The impact of prior antibiotic use has been an area of much discussion in HABP/VABP and other indications (especially CABP). Although allowing up to 24 to 48 hours of prior antibiotic treatment before study enrollment may substantially enhance trial feasibility in this indication, effective antibacterial drugs given promptly for the treatment of HABP/VABP may result in interpretability and integrity issues for an NI trial design. Some participants suggested that interpretability and integrity issues due to 24 hours of prior antibacterial drug therapy may be of less concern with a 28-day ACM endpoint. As a potential solution, trial sites should be encouraged to employ prompt or even “anticipatory” enrollment procedures so that for some patients the antibacterial drug therapy for

HABP/VABP can be initiated promptly within the context of the trial. Regardless, the Working Group expects that available data sets could be used to further evaluate this issue.

- Finally, beyond ACM, all other potential alternative outcome measures will have to be "well defined and reliable," define concepts of direct relevance to patients or have evidence that indirect benefits reflect direct benefits, allow justification of an NI margin, and prove useful for increasing trial feasibility. Study feasibility is defined not only based on the trial size but also the ability to be conducted globally, to obtain valid measurements with minimal missing data, and to reach completion within a reasonable timeframe while maintaining scientific validity. Furthermore, study costs should not preclude small companies from embarking on such trials. In short, the feasibility of the study will depend on considering all drivers of study design: outcome measure, enrollment criteria, NI margin, inclusion and exclusion criteria, primary analysis set, and meaningfulness of the design and results in providing benefits to patients. Lastly, as an extension to addressing issues related to the individual clinical trial, the Working Group will address feasibility and scientific validity considerations of conducting separate or combined trials for HABP and VABP as well as of conducting separate development programs in these two indications.

Endpoints for HABP vs. VABP Other Than ACM

The endpoint for HABP vs. VABP trials could well be different given the differences in patient populations, diagnostic modalities, and mortality. While the 2010 FDA Guidance advised separating the indications, comments posted to the docket discouraged this approach, and now the FDA in fact would consider proposed approaches to study both indications in a single trial, at least as a starting point. Nonetheless, there was concern among some members in the Working Group that mortality rate and other differences between the two infections may suggest these are different diseases, which means that for a single trial enrolling both populations, a sponsor may have to prespecify the hypotheses based on the proportion of patients who have VABP or HABP. Even within the VABP subset there may be a bimodal distribution of patients based on drug clearance, for example. Most notably, a small subset of VABP patients could alter conclusions from what is primarily a HABP study, or NI on HABP could mask differences in VABP patients.

Potential HABP Endpoints

An FDA review of historical papers for a CABP clinical endpoint included a cross-study comparison of patient recovery before the availability of antibiotic drugs versus after. These data demonstrate a large difference in clinical resolution of symptoms between treated and untreated patients that begins as soon as day 1 and extends to day 7. In this approach, historical "cross-study" comparisons represent the most appropriate data to justify a clinical recovery endpoint on symptoms for the NI trial

design in HABP/VABP, specifically a clinical recovery endpoint up to day 7. Collection of carefully defined patient symptoms would fulfill the requirement to assess how a patient feels, functions, and survives. Since HABP is more severe than CABP, the treatment effect should be large. Recent evidence from older hospitalized patients with CABP with comorbidities (patients similar to those with HABP) shows substantial symptom burden in these patients [10].

Potential VABP Endpoints

For VABP, a different approach may be necessary given the inability of most intubated patients to reliably report symptoms.

- *A priori*, various parameters were seen as biologically plausible elements of an early endpoint, but the possible choices vary in their strengths and weaknesses. It was agreed that the CPIS criteria are not adequate in terms of following patients over time given existing evidence on the lack of reliability of this measure. (See Schurink et al. [11] and Zilberberg et al. [12]; CPIS and other severity scores were discussed at the 2009 FDA co-sponsored workshop [13].)
- Improvement in oxygenation was another focal point of discussion. While the literature to date suggests that oxygenation status is prognostic of outcome, no data are available to show that it is sensitive to antibiotic treatment effect or that it has been rigorously evaluated as a surrogate endpoint for mortality. Specifically, while there are valid clinical uses for markers such as oxygenation and temperature and although data on some of these measures show a relationship to death,

correlation is not sufficient evidence for surrogacy. Comparator data are critical to establish the validity of a surrogate endpoint. On the other hand, this measure has significant face validity—it is not possible to survive if oxygenation does not ultimately improve. However, it is possible that patients could die despite improvements in oxygenation or that factors other than antibiotics might affect oxygenation. (See Guérin et al. [14]; improvement in oxygenation was discussed at the 2009 FDA co-sponsored workshop [13].)

- Similarly, although time-to-extubation or ventilator-free days could be endpoints of direct relevance to the VABP patient, the consensus was that variability in decisions to intubate and extubate could be problematic in developing these measurements into an endpoint, especially in a global trial with a multitude of investigator sites across which standards of critical care practice may vary. An additional requirement would be obtaining data on treatment effects for these measures to justify an NI hypothesis.
- For VABP, robust evidence defines a large treatment effect on the ACM endpoint, which thereby provides flexibility on the margin. Using an ITT analysis population, specifically for a broad-spectrum antibiotic trial, adds to the feasibility. On this basis, a single trial in VABP with a mortality endpoint at 21 to 28 days for a 10% margin on 20% mortality would require approximately 674 patients (337 per arm) at 90% power. A population with 15% mortality would require a sample size of approximately 544 patients. A sensitivity analysis in the micro-ITT population or any other relevant subgroups would not

require formal demonstration of NI; it should also include an analysis of prior antibiotic use as another sensitivity analysis. Whether this sample size reflects a feasible study was a matter of some debate.

Formulation of a Statistical Analysis Plan

Steps to guide initial development of a statistical analysis plan (SAP) include the following:

- Examining the definitions of various outcome measures in current trials;
- Examining the sensitivity to treatment effect of an earlier (e.g., 14-day or 21-day) ACM endpoint and its relationship to later time points like 28-day mortality;
- Exploring other potential alternative clinical endpoints such as symptoms (for which treatment effects are already known) for HABP and VABP by determining the frequency of the proposed parameter(s) at baseline and then over time during treatment;
- Evaluating potential indirect measures of treatment effects, including their definitions, timing, relationship to direct benefits, and effect sizes;

- Determining the impact on sample size of a potential new endpoint, including ACM at 14 days (or 21 days), an ITT primary analysis population, and altered inclusion/criteria such as prior antibiotic use (e.g., allowing 24 hours of prior therapy specifically for non-28-day ACM), increased severity of illness at baseline, and subgroups with and without receipt of prior therapy;
- Examining differing NI margin requirements;
- Examining performance characteristics related to the expected sample size in a microbiologically confirmed population;
- Standardizing the definition of pneumonia across the various databases for easier comparability (e.g., American Thoracic Society/Infectious Diseases Society of America criteria); and
- Identifying a set of prognostic factors (e.g., APACHE-II) based on the literature to estimate the association with mortality.

Significant discussion centered on how best to analyze the data to assess mortality, particularly in terms of understanding whether assessment can reasonably occur at an earlier time point, to avoid some of the possible attenuation of treatment effects due to non-infection-related deaths at day 28. Options included seeking evidence that allows a comparison of the

survival curves between active agents and either a non-specific therapy control or a control that has inferior effect on mortality. For these comparisons, from randomization to day 28, the goal would be to seek to identify where they diverge as well as examine instantaneous risk of mortality between groups at slices of time.

Determining attributable mortality was discussed, but it was generally agreed that attribution in a clinical trial setting is so challenging that it cannot be done in this context. Most participants considered that attribution is not necessary given the large treatment effects of antibiotics (M1) relative to the surrogate estimate of placebo effect via inadequate or delayed treatment [15]. However, others were concerned that rates of non-attributable mortality could be so high as to reduce assay sensitivity of ACM [16]; it was also noted that the historical evidence of treatment effect derives from a relatively small data set [15].

As noted above, one hypothesis can be based on the approach in ABSSSI and CABP. Given the biological and clinical similarities of HABP and CABP, assessing the quality of the data sets to support a CABP-like 4-point symptom measure for HABP seemed a reasonable starting point to most participants as an interim outcome measure. What is needed for HABP/VABP is to understand when deaths occurred and also, using information on baseline prevalence of symptoms, when these symptoms changed over time, and the distribution of outcomes/rates. The effect of changing the number of symptoms, the amount of improvement required over baseline, and/or the timing of assessment will be examined. With regard to VABP, it was noted

that the work done in ABSSSI relied on a historical data set to establish the marker of erythema. Similarly, the oxygenation ratio as a marker tracks with the mortality outcome and was argued by some to be on the causal pathway of disease. However, others cautioned against extrapolating the strategy of ABSSSI to establish oxygenation as a surrogate endpoint for VABP given that lesion size is a clinician-reported outcome with a demonstrated relationship to patient pain while oxygenation is a biomarker whose relationship to direct measures of benefit remains to be defined.

Next Steps

- The first step will be a high-level descriptive statistical analysis of the available data to inform development of the formal SAP.
- The group accepted that a preliminary hypothesis for HABP is that a symptom plus mortality-based endpoint built on the model of CABP could perform well. To that end, the first pass through the data can assess whether these symptom data exist and, if so, what was their severity and distribution at baseline and the frequency of measurement and rates over time.
- For VABP, a working hypothesis is that the ACM endpoint could be assessed at an earlier time point; thus, the first review of the data sets will be with an eye toward determining if the data can support that hypothesis. In addition, other clinical markers of interest in VABP will be explored (e.g., improvement in oxygenation).

- These exploratory descriptive analyses will be performed on a single HABP/VABP trial available to the Working Group.
- Thereafter, a formal SAP will be articulated, approved by the Working Group, and implemented using all the in-kind clinical trial data sets. The focus will be on understanding the impact of differing outcome definitions, outcome timing, analysis population assumptions, and enrollment criteria on the feasibility of HABP/VABP trial conduct.

Working Group Members

The conclusions described within this document represent the work of the FNIH Biomarkers Consortium HABP/VABP Working Group:

- Jeff Alder, Ph.D. (Bayer)
- Partha Bagchi, Ph.D. (Johnson and Johnson)
- Steve Barriere, Pharm.D. (Theravance)
- Helen W. Boucher, M.D., FACP (Tufts University, IDSA)
- Laurie Burke, R.Ph., M.P.H. (FDA)
- Becky Coleman, Pharm.D. (Theravance)
- Lynn Connolly, M.D. (Achaogen)
- Edward Cox, M.D. (FDA)
- Aaron Dane, M.Sc. (AstraZeneca)
- Anita Das, Ph.D. (InClin)
- Dennis M. Dixon, Ph.D. (NIH/NIAID)

- Michael Dudley, Pharm.D. (Rempex)
- Barry Eisenstein, M.D., FACP, FIDSA (Cubist)
- Thomas File, M.D. (Summa Health System)
- Tom Fleming, Ph.D. (University of Washington)
- Dean Follmann, Ph.D. (NIH/NIAID)
- David Friedland, M.D. (Cerexa)
- Ian Friedland, M.D. (Cubist)
- Kenneth Hillan, Ph.D. (Achaogen)
- Alan Hopkins, Ph.D. (Theravance)
- Nicholas Kartsonis, M.D. (Merck)
- Charles Knirsch, M.D., M.P.H. (Pfizer)
- Mark Kunkel, M.D. (Pfizer)
- Chin-Yu Lin, Ph.D. (Achaogen)
- Lily Llorens, Ph.D. (Cerexa)
- Daniel Meyer, Ph.D. (Pfizer)

- Roger Novak, Ph.D. (Sanofi)
- Elektra Papadopoulos, M.D. (FDA)
- John Powers, M.D. (NIH)
- Philippe Prokocimer, M.D. (Trius)
- Rebecca Redman, M.D. (Johnson and Johnson)
- John Rex, M.D., FACP (AstraZeneca)
- Dan Rubin, Ph.D. (FDA)
- Ashley Slagle, Ph.D. (FDA/CDER, CTR)
- Judy Siuciak, Ph.D. (FNIH)
- Will Stubbings, Ph.D. (now Achim Kaufhold, M.D.) (Basilea Pharmaceutica, Ltd.)
- Anthony Suffredini, M.D. (NIH/CC/CCMD)
- George H. Talbot, M.D., Co-Chair (Talbot Advisors, IDSA)
- Joe Toerner, M.D., Co-Chair (FDA)
- John Tomayko, M.D. (GSK)
- Antoni Torres, M.D. (Catedràtic de Medicina Hospital Clínic, Barcelona, Spain)

References

[1] Luna CM, Blanzaco D, Niederman MS, et al. Resolution of ventilator-associated pneumonia: prospective evaluation of the clinical pulmonary infection score as an early clinical predictor of outcome. Crit Care Med. **2003**;31(3):676–82.

[2] Dennesen PJ, van der Ven AJ, Kessels AG, Ramsay G, Bonten MJ. Resolution of infectious parameters after antimicrobial therapy in patients with ventilator-associated pneumonia. Am J Respir Crit Care Med. **2001**;163(6):1371–75.

[3] Esperatti M, Ferrer M, Giunta V, et al. Validation of predictors of adverse outcomes in hospital-acquired pneumonia in the ICU. Crit Care Med. **2013** Jun 11 [Epub ahead of print].

[4] Meeting materials, November 4, 2011, Meeting of the Anti-Infective Drugs Advisory Committee. Available at: http://www.fda.gov/AdvisoryCommittees/CommitteesMeetingMaterials/Drugs/Anti-InfectiveDrugsAdvisoryCommittee/ucm242307.htm

[5] Kollef MH, Chastre J, Clavel M, et al. A randomized trial of 7-day doripenem versus 10-day imipenem-cilastatin for ventilator-associated pneumonia. Crit Care. **2012**;16(6):R218.

[6] Ramirez J, Dartois N, Gandjini H, et al. A randomized phase 2 trial to evaluate the clinical efficacy of two high tigecycline dosage regimens versus imipenem/cilastatin in hospital-acquired pneumonia. Antimicrob Agents Chemother. **2013**;57(4):1756–62.

[7] Wunderink RG, Niederman MS, Kollef MH, et al. Linezolid in methicillin-resistant *Staphylococcus aureus* nosocomial pneumonia: a randomized, controlled study. Clin Infect Dis. **2012**;54(5):621–29.

[8] Fleming TR, Powers JH. Biomarkers and surrogate endpoints in clinical trials. Stat Med. **2012**;31(25):2973–84.

---

[9] Giunta V, Ferrer M, Esperatti M, et al. Intensive care unit acquired pneumonia with or without etiologic diagnosis: a comparison of outcomes. Crit Care Med. In press.

[10] Wyrwich KW, Yu H, Sato R, Strutton D, Powers JH. Community-acquired pneumonia: symptoms and burden of illness at diagnosis among US adults aged 50 years and older. Patient. **2013**;6(2):125–34.

[11] Schurink CA, Van Nieuwenhoven CA, Jacobs JA, et al. Clinical pulmonary infection score for ventilator-associated pneumonia: accuracy and inter-observer variability. Intensive Care Med. **2004**;30(2):217–24.

[12] Zilberberg MD, Shorr, AF. Ventilator-associated pneumonia: the clinical pulmonary infection score as a surrogate for diagnostics and outcome. Clin Infect Dis. **2010**;51 Suppl 1:S131–35.

[13] Meeting materials, "Clinical Trial Design for Hospital-Acquired Pneumonia and Ventilator-Associated Pneumonia," a public workshop co-sponsored by the FDA and professional societies. Available at: http://www.fda.gov/Drugs/NewsEvents/ucm169877.htm

[14] Guérin C, Reignier J, Richard J-C, et al. Prone positioning in severe acute respiratory distress syndrome. N Engl J Med. **2013**;368(23):2159–68.

[15] Sorbello A, Komo S, Valappil T, Nambiar S. Registration trials of antibacterial drugs for the treatment of nosocomial pneumonia. Clin Infect Dis. **2010**;51 Suppl 1:S36–41.

[16] Muscedere JG, Day A, Heyland DK. Mortality, attributable mortality, and clinical events as end points for clinical trials of ventilator-associated pneumonia and hospital-acquired pneumonia. Clin Infect Dis. **2010**;51 Suppl 1:S120–25.