

Noninferiority Trials in Community-
Acquired Bacterial Pneumonia
(CABP)

James Floyd, M.D.
University of Washington
FDA Anti-Infective Drugs Advisory Committee
December 9, 2009

The draft guidance under discussion clarifies several important issues regarding the appropriate design and conduct of noninferiority (NI) trials for community-acquired bacterial pneumonia (CABP). Overall, the recommendations in the guidance are a dramatic improvement over the current practice, and the FDA should be congratulated for their efforts. Because essentially all clinical trials of new antibiotics for CABP over the last two decades have been NI trials, I will limit my comments to this topic.

Opening Remarks

- New antibiotics are desirable, but only if they are effective
- First, must design trials that have capacity to show that treatments are effective and safe
- Then can consider issues of feasibility

Everyone here agrees that new antibiotics, particularly those active against resistant pathogens, are desirable. However, new antibiotics are desirable only if they are effective. Approving new antibiotics that are less effective than current treatments under the guise of similar efficacy puts patients at risk. First, trials must be designed so that they have the capacity to demonstrate efficacy. Then, issues of feasibility can be considered.

“Adequate and well-controlled”

- Amendment to FDCA in 1962 established “adequate and well-controlled studies” as the basis for determining efficacy of new drugs
- 21 CFR 314.126, on active control studies:
“Similarity of test drug and active control can mean either that both drugs were effective **or that neither was effective.**”
- Further detailed in ICH E-9 and E-10

An amendment to the federal Food, Drug, and Cosmetic Act in 1962 established “adequate and well-controlled studies” as the basis for determining the efficacy of new drugs, and this was codified in 1985. On active control studies, 21 CFR 314.126 states that “Similarity of test drug and active control can mean either that both drugs were effective or that neither was effective.” The features of trial design and conduct that allow one to distinguish between these two scenarios are detailed by the International Conference on Harmonization (ICH) in documents E-9 and E-10, and my comments are derived from the scientific issues presented in these documents.

Noninferiority Trials

- Reliable and reproducible evidence of efficacy of the active control
- Study design
 - Patient population, disease severity, outcome
- Study conduct
 - Minimize bias towards similarity in outcomes
- Margin (M2)
 - Must be clinically justified

NI trials do not show that two treatments are equivalent to each other. Rather, they exclude that one treatment is worse than another by a prespecified, clinically acceptable margin. For a NI comparison to be valid: (1) evidence of a treatment effect of the active control must be reliable and reproducible, (2) the trial design must reflect the patient and disease characteristics, concomitant treatments, and outcomes of previous controlled trials, (3) the trial must be conducted in a way that minimizes bias towards similar treatment effect, and (4) the margin selected must be justified clinically.

Effectiveness of Active Control

- Patients had pneumococcal pneumonia
 - No evidence for atypical pathogens
- Patients were at high risk of death
 - Effect size greatest in patients with bacteremia and age > 50
- Evidence of effect for mortality only
 - Observational studies with signs and symptoms as endpoint and Kingston et al. JAMA. 1961;176:118.
 - Decrease in biomarkers and symptoms not a well-defined or reliable measure of treatment effect

The FDA has provided compelling evidence from historical data and previous controlled trials of a substantial treatment effect for antibiotics in CABP, and estimates of this effect are presented in the draft guidance. A few points about these data are worth mentioning. First, nearly all patients had pneumococcal or lobar pneumonia. There is little reliable and reproducible evidence for a treatment effect with atypical pathogens from which a noninferiority margin can be calculated. Second, the effect size was greatest in patients at highest risk of death, those with bacteremia and over age 50. Third, an arguably reliable estimate for a treatment effect exists for the endpoint of mortality, but not for intermediate clinical outcomes or surrogate endpoints.

Today, the FDA presented data from observational studies that strongly suggest a treatment effect for signs and symptoms of pneumonia, and these studies are similar in design to the studies of mortality. This was an impressive review of the evidence, and I have little doubt that effective antibiotics improve symptoms. Then why is mortality a reasonable endpoint for NI trials while signs and symptoms are not? Signs and symptoms are subjective and were poorly defined in these early studies, rendering them irreproducible. Even contemporary studies suggest that clinicians have difficulty corroborating physical findings. On the other hand, death is an objective outcome, the most objective outcome, and has not varied from 1940 to 2009.

Additionally, the Infectious Diseases Society of America position paper on the design of clinical trials in CABP refers to a placebo-controlled RCT of tetracycline in young men with mild CAP, 46% of whom had evidence of *Mycoplasma pneumoniae* (Ref 1). This trial showed that tetracycline improved biomarkers, such as temperature and infiltrate on CXR, as well as symptoms, such as cough and fatigue (Ref 2). However, a valid endpoint cannot be constructed from these biomarkers and symptoms, and therefore an accurate estimate of effect size cannot be determined from this trial.

Age group, variable	Bacteremic subjects		Nonbacteremic subjects	
	No specific treatment	Antibiotic treatment	No specific treatment	Antibiotic treatment
12-29 years				
No. of deaths/no. of patients (% mortality)	81/126 (64.29)	5/21 (23.81)	90/1036 (8.69)	4/97 (4.12)
% Difference (95% CI)	40 (20-61)		5 (0-9)	
Proposed margin, %	10		None	
30-49 years				
No. of deaths/no. of patients (% mortality)	319/428 (74.53)	23/81 (28.40)	218/1222 (17.84)	14/171 (8.19)
% Difference (95% CI)	46 (35-57)		10 (5-14)	
Proposed margin, %	10		2.5	
≥50 years				
No. of deaths/no. of patients (% mortality)	368/395 (93.16)	40/84 (47.62)	353/715 (49.37)	23/143 (16.08)
% Difference (95% CI)	46 (35-57)		33 (26-40)	
Proposed margin, %	10		10	

Fleming TR, Powers JH. CID. 2008;47:S108-200.

This table comes from a review of the evidence for a treatment effect of antibiotics in CABP by Fleming and Powers, with mortality stratified by age and bacteremia status (Ref 3). The evidence for a treatment effect on mortality is much clearer in bacteremic patients and those over the age of 50, as is effect modification by these variables.

Study Design
<ul style="list-style-type: none"> • Patients must be at high risk of death <ul style="list-style-type: none"> – High proportion of patients with age > 50, bacteremia • Patients must have CABP with typical pathogens <ul style="list-style-type: none"> – Microbiological ITT should be primary analysis

For one to assume constancy of effect in contemporary trials, patients who are at high risk of death must be enrolled. Also, the primary efficacy analysis must include only those patients with microbiologic evidence of typical bacterial pathogens, called the MITT population in the draft guidance. Failing to do so may result in similar outcomes between treatment groups in the study population, while in patients

at higher risk of death and with typical pathogens the experimental treatment may be substantially worse than the active control. The draft guidance makes both of these points very clearly.

Study Design

- Patients who have already receive Abx should not be randomized
- Post-randomization exclusions must be avoided
 - Concomitant Abx should be avoided
 - Pts who receive concomitant Abx may not be excluded from analysis

Patients who receive any prior antibacterial drug therapy should not be enrolled in NI trials, as this minimizes any difference in treatment effect between two drugs and can lead to a false NI inference. There is evidence to support this, and by default we should assume that this is the case. Concomitant antibacterial drug therapy should not be allowed for similar reasons. Any patients who do receive concomitant antibacterial drug therapy should not be excluded from analysis, because such post-randomization exclusions are likely to introduce bias. The current draft guidance also makes these points very clearly.

An Example: Cethromycin

- Pts were not sick
 - Half of patients had a PORT score of 1
- Pts did not have CABP
 - Typical bacterial pathogens in < 25%
- Prior Abx in over 10% of patients
- 15% of patients excluded from “clinically evaluable” population after randomization

A recent example highlights the pitfalls of improperly designed and executed NI trials. In June 2009, this advisory committee evaluated two trials that compared cethromycin with clarithromycin for the outpatient treatment of mild to moderate CABP (Ref 4). In these trials, half of patients had a PORT score of 1 and less than 1% died. Typical bacterial pathogens were recovered in less than a quarter of patients. Prior antibiotics were given in over 10% of the study population and nearly 15% of patients were excluded from one of two primary efficacy analyses based on post-randomization events, including the use of concomitant antibiotics.

An Example: Cethromycin

Analysis Population	Cethromycin cure/N (%)	Clarithromycin cure/N (%)	Treatment Difference, (95% CI) *
ITT (n=510)	213/257 (82.9)	224/253 (88.5)	-5.7 (-12.1, 0.8)
CE (n=445)	205/224 (91.5)	212/221 (95.9)	-4.4 (-9.3, 0.5)
<ul style="list-style-type: none"> Excluding patients with PORT = 1, atypical pathogens, and prior Abx → treatment effect of cethromycin looked worse 			
ITT (n=510)	86/113 (76.1)	90/105 (85.7)	-9.6 (-20.9, 1.6)
CE (n=445)	85/98 (86.7)	86/92 (93.5)	-6.7 (-16.2, 2.7)

When patients with a PORT score of 1, patients with atypical pathogens, and patients who received prior antibiotics were excluded, less than half the initial study population remained and cure rates were much lower. In one of the two studies, excluding these patients made the treatment difference larger, with cethromycin appearing worse than clarithromycin. In other words, diluting the sample with patients who were not sick and did not have typical pathogens and giving prior antibiotics may have obscured the inferiority of cethromycin. The deficiencies in these trials are not unique – they plague nearly all contemporary trials in CABP and bring into question the claims of efficacy of some previously approved treatments.

Can We Extrapolate Outcomes?

- Rationale for extrapolating to “clinical response” from mortality
 - Mortality is a subset of clinical response
 - Any treatment effect on clinical failure would be larger than the effect on mortality
 - A treatment effect on mortality is unlikely to be observed in a contemporary NI trial because of advances in supportive care

One flaw in the draft guidance is the designation of “clinical response” as an appropriate primary endpoint in NI trials. This is a composite outcome that includes mortality and the resolution of signs and symptoms. In doing so, the following assumptions are made: (1) mortality is one of several types of clinical failure, (2) any treatment effect on clinical failure will be larger than the effect on mortality, and (3) a treatment effect on mortality is unlikely to be observed in contemporary trials because of rescue therapies and supportive care, such as mechanical ventilation. Let us explore the second and third assumptions.

Can We Extrapolate Outcomes?

- 2) Any treatment effect on clinical failure would be larger than the effect on mortality
 - No evidence that effect on resolution of signs and symptoms predicts effect on mortality
 - Doripenem for VAP: DORI-09

Clinical Response	Doripenem n/N (%)	Piperacillin/tazobactam n/N (%)	Difference % (95% CI)
Clinically evaluable	109/134 (81.3)	95/119 (79.8)	1.5 (-9.1, 12.1)
Clinical MITT	148/215 (69.5)	134/209 (64.1)	5.4 (-4.1, 14.8)

Mortality	Doripenem (N=225) n (%)	Piperacillin/tazobactam (N=221) n (%)	Relative Risk (Dori/Pip-tazo) (95% CI)
During iv therapy	21 (9.4)	9 (4.1)	2.3 (1.1, 4.9)

First, there is no evidence that a treatment effect on biomarkers of disease and resolution of symptoms reliably predicts an effect on mortality. In fact, the opposite can be seen. In DORI-09, a multicenter phase 3 trial that compared doripenem to piperacillin/tazobactam in patients with ventilator association pneumonia, there was no difference in clinical response rates (Ref 5). However, doripenem resulted in a significant increase in mortality during treatment.

Can We Extrapolate Outcomes?

- 3) A treatment effect on mortality is unlikely to be observed, because of advances in supportive care
 - Mortality from CABP remains high, 15% in a recent observational study
 - Low mortality in contemporary NI trials of CABP reflects selection of patients with low baseline risk

Second, the myth that patients no longer die of pneumonia because of rescue therapies and advances in supportive care is not supported by evidence. 55,477 Americans died of pneumonia in 2006, according to the National Center for Health Statistics (Ref 6), and current mortality from CABP in hospitalized patients was estimated at 15% in a recent observational study (Ref 7). The low mortality rate seen in contemporary trials of CABP reflects the sampling of patients who are at low risk of death. Furthermore, challenging the assumption of constancy of effect because of the much lower mortality rate observed in contemporary trials does not argue for changing the endpoint to an unvalidated surrogate measure. It argues against the use of NI trials to demonstrate efficacy and in favor conducting superiority trials.

The assumptions required to extrapolate from mortality to clinical response are not based on evidence, and they appear likely to be untrue. This is precisely the reason that any endpoint evaluated in a NI trial must be one for which there is reliable and reproducible evidence of a treatment effect for the active control.

NI Margin

- M1 based on estimate of treatment effect of active control
 - Observational studies: 15-22%
 - Controlled trials: 8.8%
 - Varies by patient subgroup
- M2 based on clinical judgment
 - Difficult to justify M2 > 10%

To derive the NI margin, both statistical reasoning and clinical judgment must be used. M1, the lower bound of the confidence interval for the estimated treatment effect of antibiotics in CABP is estimated at 9-15% by the FDA. Clinical judgment must be used to determine how much loss of efficacy is acceptable in setting M2. Although M1 may be large, especially in subgroups at higher risk of death, a large M2 cannot be justified for a life-saving therapy. Most clinicians would find it difficult to justify using a treatment that results in a 10% increase in mortality over existing effective therapies.

Consistent with the presentation by the FDA today, an odds ratio based approach to NI margins may offer more flexibility in enrolling some patients who are not in the highest risk categories, while maintaining the validity of the margin selected. Nonetheless, study populations will be smaller if patients at higher risk are enrolled, creating a clear incentive for this favorable feature of trial design.

Conclusion

- If principles of good study design and conduct are violated, NI trials will be unable to demonstrate efficacy
- A trial that is “feasible” but cannot provide evidence of efficacy is useless
- Antibiotics of questionable efficacy are not helpful to clinicians

The issues raised today are not arbitrary or merely statistical side notes. NI trials that violate principles of good study design and conduct are unable to demonstrate the efficacy of new treatments. If a trial has no capacity to demonstrate efficacy, it no longer matters whether or not it is feasible. As a clinician, I am not interested in a new antibacterial therapy if I have to worry about whether or not it is less effective than existing treatments because of poor study design and conduct.

References

- 1) Kingston JR, et al. JAMA. 1961;176:118-125.
- 2) Spellberg B, et al. Clin Inf Dis. 2008;47:S249-265.
- 3) Fleming TR, Powers JH. Clin Inf Dis. 2008;47:S108-200.
- 4) FDA Briefing document, Anti-Infective Drugs Advisory Committee, June 2, 2009.
<http://www.fda.gov/downloads/AdvisoryCommittees/CommitteesMeetingMaterials/Drugs/Anti-InfectiveDrugsAdvisoryCommittee/UCM161847.pdf>.
- 5) FDA Briefing document, Anti-Infective Drugs Advisory Committee, July 19, 2008.
<http://www.fda.gov/ohrms/dockets/ac/08/briefing/2008-4364b1-01-FDA.pdf>.
- 6) National Center for Health Statistics. Deaths: Final Data for 2006.
http://www.cdc.gov/NCHS/data/nvsr/nvsr57/nvsr57_14.pdf.
- 7) Ochoa-Gondar O, et al. BMC Public Health. 2008;8:222.