AAC Journals.ASM.org

COMMENTARY

Early Endpoints for Acute Bacterial Skin and Skin Structure Infections

George L. Drusano

Institute for Therapeutic Innovation, University of Florida School of Medicine, Albany, New York, USA

In this issue of the journal, Friedland et al. (4) present a retrospective analysis of the two CANVAS trials (3), employing the newer endpoints from the U.S. Food and Drug Administration (FDA) (7). First, they are to be congratulated for having the foresight to include sufficient measurements in their trial to allow the reanalysis even prior to the issuance of the new guidelines. The analysis employing the earlier endpoint was overall concordant with the analyses using the traditional test-of-cure endpoint. Both showed that when data from both trials were pooled, the results demonstrated ceftaroline to be significantly better than the combination of vancomycin plus aztreonam. Individually, one of the trials demonstrated significantly better effect, but the second trial did not, preventing an outright claim of superiority.

While the reanalysis was important to see, there are, I believe, other issues as important as or more important than the analysis *per se*. These have to do with the new FDA guidelines for acute bacterial skin and skin structure infections (ABSSSI).

(i) Clearly, the speed with which improvement takes place may be an important issue. However, equally important is whether the patient remains in the "cure" category at test of cure (TOC). To this end, it would have been nice to identify any discordances between endpoints for the subset of patients for whom both endpoints were determined. One could imagine that some early responders might fail at TOC. Likewise, given enough time, it is highly likely that patients failing early could still reach cure status at TOC.

(ii) Was what was measured the best endpoint? In the paper by Friedland et al. (4), the two early endpoints measured were cessation of infection spread and afebrility at day 3 in a subset of patients with lesion sizes of >75 cm^2 and deep and/or extensive cellulitis, major abscess, or an extensive wound. The first issue is with the choice of fever, as only a minority of patients were febrile at study entry. Impaired subjects, such as the frail elderly, may not be able to mount a febrile response yet may be very seriously ill. The second part of the endpoint was cessation of spread and not some more difficult-to-achieve endpoint such as reduction (by a percent) in involved area. Bhavnani et al. presented an abstract at the 51st Interscience Conference on Antimicrobial Agents and Chemotherapy in 2011 (2) describing the analysis of a phase II study of a new agent for ABSSSI. In this analysis, two groups were contrasted, patients with AUC/MIC ratios (ratios of the area under the concentration-time curve to the MIC) at or above a given breakpoint and those with ratios below that breakpoint. Two endpoints were examined: reduction in area of erythema by different percentages (10 to 70% reduction) at multiple times and this same sort of analysis for swelling (measurements were made daily to day 7). This analysis employed a Kaplan-Meier approach and showed for both endpoints that the greatest value for delta between treatment groups occurred on either day 3 or day 4. Also, the amount of effect showed maximal significance at the 10 to 30% reduction thresholds for both endpoints. This makes clinical sense, as higher degrees of decline in endpoint would occur less frequently. Nonetheless, with two groups with differing exposure to the same drug, it is clear that there is both a time of evaluation and a degree of effect that optimally separates the two groups. This sort of analysis needs to be considered for a contrast between an experimental agent and a control agent so that the most scientifically informative trials can be performed. The important issue (expanded below) is that the endpoints chosen ought to be derived not from historical data from the 1930s but from modern data identified with modern techniques (allowing percent reduction in involvement estimation) and employing modern mathematical approaches.

(iii) Perhaps the most important issue is the generation of endpoints from 1937 data sets, as published. The papers by Snodgrass and Anderson (5, 6) were groundbreaking for the time. There are so many issues with these data, however, that it is mind-boggling that they serve not as a springboard to design exploratory analyses with modern data but as the basis on which to generate a guidance document.

The FDA has (absolutely correctly) denied approval of the anti-MRSA cephalosporin ceftobiprole in recent times because of the inability to verify the source data in a sufficiently high number of patients from the database. Yet the data in the papers by Snodgrass and Anderson are completely nonverifiable. Further, it is highly likely that the endpoint is nondiscriminatory. When one reads the sulfanilamide paper, it is important to realize that 60% of patients had experienced a cessation of lesion spread on day 0 in the sulfanilamide group, while 39% experienced cessation on day 0 in the UV-light control group! Taking out 60% of the population on day 0 markedly diminishes the ability to discriminate the effectiveness of two drugs. Further, the endpoint is cease of infection spread, not regression of the involvement. As Bhavnani et al. demonstrated (2), the difference between groups depends on how much lesion regression is demanded and when the evaluation occurs. Should we not revisit the endpoint based on modern quantitative methodology to identify a required degree of effect and time of evaluation that provides optimal discriminatory power?

(iv) Another issue, not so much in the realm of endpoint, is the philosophy behind calculation of a noninferiority (NI) margin. While it is recognized that the estimation of drug effect on the NI margin should probably be somewhat conservative, the issue devolves into whether one wishes to be conservative or straightforwardly biased. The best (most likely to be correct) estimate of the drug effect is the actual difference between groups (drug versus placebo). There are obvious difficulties performing placebo-con-

Published ahead of print 6 February 2012

Address correspondence to George.Drusano@medicine.ufl.edu.

Copyright © 2012, American Society for Microbiology. All Rights Reserved.

doi:10.1128/AAC.00157-12

The views expressed in this Commentary do not necessarily reflect the views of the journal or of ASM.

trolled trials in seriously ill infected patients. Consequently, when trials are performed with an active control, the measure of difference is already a conservative estimate for the new agent. However, it may be wise to be somewhat more conservative. Consequently, taking the upper 95% confidence bound on the control effect may be an acceptable way to make certain that a mistake will not be made which will propagate through the system with other new agents. Currently, we do more than this: we take both upper and lower 95% confidence intervals for the two groups. This throws away 95% of the distribution for both groups. This now verges not on conservatism but on absolute bias in the estimation of effect of the new agent. This estimate is the so-called m1 estimate. As if this were not bad enough, yet another "conservative correction" is introduced. This is the totally empirical discounting of the m1 estimate by a percentage. In two different sets of guidelines, the discount was 30% to 50%. This estimate is referred to as the m2 estimate. There is absolutely no scientific basis for the m2 discount. Indeed, this takes an already highly biased (low) estimate of effect and decreases it further without any justification. One must call the whole process into question.

(v) Finally, the statistical approach employed is a frequentist one, in which only the data and nothing else is considered in the estimation of effect. Ambrose et al. (1) recently published a paper in which a Bayesian approach was employed in the calculation of effect of a drug in patients with hospital-acquired bacterial pneumonia and ventilator-acquired bacterial pneumonia. One should recognize that before a drug is introduced into clinical testing, there are series of data that are generated both *in vitro* and *in vivo* (animal model systems) where antibacterial activity is demonstrated. While no one would call for a cessation of clinical drug testing on the basis of these data, they have probative value regarding the question of whether the drug has activity and will have a positive impact on clinical outcome due to inhibition or killing of the infecting organism. Clearly, the issue of human toxicity is a separate one and is one of the reasons why clinical trials in infected patients will always be required. However, to throw out all prior information regarding drug activity would seem improvident. The approach used by Ambrose et al. demonstrated that the major impact of employing Bayesian estimation was on the confidence bounds around the point estimate of the effect.

In summary, Friedland et al. are to be commended for their paper. Ceftaroline is a valuable addition to the physician's armamentarium. The larger issue, however is the use of the new early endpoint and the appropriateness of its adoption from 1937 data that are not verifiable and where the constancy assumption is clearly violated (eggs and onions were interdicted, multiple enemas were employed, and paraffin was employed as a laxative, etc.). Further it is highly likely that the early endpoint as currently constituted is not optimally discriminatory between interventions. It is also important to see how early and late endpoints line up. Finally, the statistical approach in which we make the estimation of effect strongly biased low needs to be revisited.

ACKNOWLEDGMENT

I have served as a consultant for Cerexa and received payment for research investigation from them.

REFERENCES

1. **Ambrose PG, et al.** 2012. Frequentist and Bayesian pharmacometric-based approach to facilitate critically needed new antibiotic development: Overcoming lies, damn lies and statistics. Antimicrob. Agents Chemother. **56**: 1466–1470.
2. **Bhavnani SM, et al.** 2011. Pharmacokinetic-pharmacodynamic analysis for efficacy of BC 3781 using new clinical trial endpoints in patients with acute bacterial skin and skin-structure infections, abstr. A2-025. Abstr. 51st Intersci. Conf. Antimicrob. Agents Chemother. American Society for Microbiology, Washington, DC.
3. **Corey GR, et al.** 2010. Integrated analysis of CANVAS 1 and 2: phase 3, multi-center, randomized, double-blind studies to evaluate the safety and efficacy of ceftaroline versus vancomycin plus aztreonam in complicated skin and skin-structure infection. Clin. Infect. Dis. **51**:641–650.
4. **Friedland HD, et al.** 2012. 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. **56**:2231–2236.
5. **Snodgrass WR, Anderson T.** 1937. Sulphanilamide in the treatment of erysipelas. Br. Med. J. **2**:1156–1159.
6. **Snodgrass WR, Anderson T.** 1937. Prontosil in the treatment of erysipelas. Br. Med. J. **2**:101–104.
7. **U.S. Food and Drug Administration.** 2010. Guidance for industry. Acute bacterial skin and skin structure infections: developing drugs for treatment. U.S. Food and Drug Administration, Rockville, MD.