

Designing Trials and Deciding on Therapy in Complex Diseases: What Can We Do to Improve?

William A. Knaus, MD

The article by Dr. Johnston in this issue titled “Determinants of Mortality in Patients with Severe Sepsis”¹ reports a new analysis of results from a high-profile clinical trial of recombinant human activated protein C in patients with severe sepsis.² They demonstrate that when one performs a disease-specific statistical control for a patient’s level of severity, the total explanatory power for short-term outcomes (hospital mortality) improves compared to older risk assessment approaches. The article by Dr. Johnston is important because it points to specific ways to improve the analytic design component of clinical trials so that both the results and their interpretation can be made more useful for clinical medicine.

The background that underlies Dr. Johnston’s analysis first surfaced in the 1990s with a new class of drugs—monoclonal antibodies—that were aimed at modulating the inflammatory cascade in patients with serious hospital-acquired infections (sepsis). Then and now sepsis represents a major cause of morbidity and mortality in seriously ill hospitalized patients, with an estimated 500,000 episodes annually in the United States and a death rate between 30% and 50%. Monoclonal antibodies seemed well designed to interrupt the untoward action of early response cytokines such as tumor necrosis factor α , interleukin-1 β , and interleukin-8 released by tissue macrophages in response to infection.³

The septic inflammatory cascade, however, is complex. From an evolutionary viewpoint, inflammatory responses were designed to mobilize the body’s immune defense in an all-out battle for survival against the invading organism. It was only after the widespread availability of intensive care, with its life support capabilities that enabled individuals who would have previously died quickly to survive, that we began to think of these inflammatory mediators as harmful. We had at that time limited knowledge, however, of

which component is more important, how they interact, and precisely when their presence transitions from defense to damage. Despite these concerns, the initial experimental results using these agents were supportive and preliminary phase II trials, although small and severely underpowered, had tantalizingly positive trends. Fueled by huge amounts of investment capital, established pharmaceutical and startup biotech companies raced headlong into phase III clinical trials.

They discovered that, like the underlying inflammatory cascade, sepsis syndrome itself was tough to describe. It affected many different organ systems simultaneously; it occurred in patients with a variety of different primary diseases from trauma to cancer and was common in patients with preexisting chronic medical conditions. To address this challenge, we assembled many of these preexisting patient characteristics as components of a sepsis-specific risk model that we intended to be used as a prospective control for severity and preexisting risk in clinical trials.⁴ We believed at the time that any new treatment—no matter how successful in preliminary studies—would not work equally in all patients eligible for the phase III trials. There would be variations in efficacy based on the level of severity—patient characteristics, a classic signal to noise challenge. New drugs might have an important contribution to outcome—the signal—but there were so many other powerful preexisting patient characteristics—the noise—that we needed to control for them prospectively. Gail and others had already demonstrated that univariate analyses of dichotomous outcomes (like 28-d mortality—the main endpoint in most sepsis trials) produce treatment effect estimates

From the Department of Public Health Sciences, University of Virginia School of Medicine, Charlottesville, Virginia.

Address correspondence to William A. Knaus, MD, Department of Public Health Sciences, University of Virginia School of Medicine, Box 800717, Charlottesville, VA 22908-0717; e-mail: wknaus@virginia.edu.

DOI: 10.1177/0272989X05279246

that are biased toward no effect, compared to risk-adjusted analyses in the presence of patient heterogeneity.⁵ Heterogeneity in the risk of bad outcome is a hallmark of sepsis trials. Having a good risk model results in about 5% to 10% greater statistical power.

Introducing a comprehensive measurement of severity as a critical component of trial design was not rapidly accepted. Prespecifying risk adjustment was untraditional in primary analytic plans for clinical trials, and it would have added to the already substantial data collection burden. There was also concern on the part of the trial's sponsors that explicitly acknowledging that these new therapies' efficacies might be related to levels of severity could make marketing a successful drug more complex. The late Dr. Roger Bone organized a consensus development conference that discussed and debated these issues and produced useful uniform and subsequently widely used definitions for sepsis syndrome and severe sepsis. These were categorical definitions designed to create uniform entry criteria; they did not have the specificity needed for risk assessment.⁶

As a result, most of the clinical trials performed on sepsis populations relied on the measurement of APACHE II, a simple, easy-to-collect approach to primary risk assessment.⁷ We did have the opportunity after the failure of another high-profile phase III trial of an interleukin-II antagonist to retrospectively illustrate how more comprehensive control for severity made it easier to detect efficacy and identify the patients most likely to benefit.⁸ A few subsequent trials used our model⁴ as a secondary analysis, but even the Food and Drug Administration (FDA) was cautious of introducing new complexities into trial design, especially for therapies that had yet to prove their efficacy.

The FDA's concern proved to be well founded, as one large sepsis trial after another failed to return positive results. Then near the end of the decade, Eli Lilly entered the troubled waters with drotrecogin alfa (activated) or recombinant human activated protein C, a drug now also known by its proprietary name, Xigris. Unlike the early response cytokine antagonists, activated protein C specifically targeted the coagulation cascade in sepsis. We knew there were important molecular links between the procoagulant and inflammatory mechanisms in the pathogenesis of organ failure in patients with sepsis. There were data from large animal models that activated protein C reversed the procoagulant and inflammatory effects of sepsis and improved survival.³ Eli Lilly's path toward clinical approval seemed deliberate and well planned. It was with a deep sense of disappointment, therefore, when we learned that the design for the pivotal trial of activated

protein C again called for the use of APACHE II as the primary adjustment rather than the sepsis-specific model that had greater explanatory power.

The clinical trial named PROWESS for Protein C Worldwide Evaluation in Severe Sepsis had clearly positive results, with an overall reduction in 28-d mortality of 6% in 1690 patients treated in 11 countries.² There was finally a new therapeutic approach in sepsis. There was also evidence, however, that protein C did not have uniform efficacy across all patients and that those who were more severely ill when treated (defined as APACHE II scores >25) were the ones who benefited. These subgroup results were so striking that the FDA included calculation of an APACHE II score as a part of the drug's indications—an unprecedented action.⁹

But when the major article announcing the trial results was published in *The New England Journal of Medicine* in 2001, the interaction of severity and efficacy was missing.² An editorial in the same issue of the journal, drawing on the description of the phase III results, stated that, "The treatment was effective regardless of age, severity of illness. . . ."³ The discrepancy between the results that had been given to the FDA and those reported in *The New England Journal of Medicine* article was so striking that Dr. Jay Siegel, the FDA official who had been prominently involved in Xigris and many previous trials in sepsis, wrote a commentary to the journal that said, in part,

Analysis of the treatment effect according to the quartile of APACHE II score was prespecified in the PROWESS trial as an important analysis and the principal analysis of outcome according to measures of severity or risk. . . . Indeed, each of the three risk related components of the APACHE II score—acute physiologic changes, older age, and the assignment of chronic health points—correlated with a greater treatment effect. . . . the APACHE II score was the best predictor of survival benefit from activated protein C. All benefit was observed among the half of subjects who had APACHE II scores of 25 or more (a 13 percent reduction). . . . subjects with lower risk showed no benefit. . . . and patients in the lowest APACHE II quartile had somewhat higher mortality with activated protein C than with placebo.⁹

This variation in presentation of the trial's major findings created confusion in the clinical community that has persisted over the ensuing 4 years. Articles have been published without new results or sufficient new patients to reach significant conclusions that encouraged clinicians to rely on their "experience" or on

other subjective measures of sepsis that had not been demonstrated to be related to the drug's efficacy as guides toward prescribing Xigris.^{10,11} Follow-up trials¹² and those that contained new information, however, continued to link the drug's short- and long-term efficacy and its cost-effectiveness explicitly and exclusively to APACHE II scores with conclusions: "Our findings suggest that the use of drotrecogin alfa in patients with severe sepsis is associated with a favorable cost-effectiveness profile, especially if restricted to the FDA approved use."¹³

But maintaining the controversy were recent remarks in a *Wall Street Journal* interview that discussed potentially misleading medical journal publications and quoted an Eli Lilly official and PROWESS's principal investigator as stating that they used another unnamed severity adjustment in their primary analysis that demonstrated that Xigris worked across all levels of illness.¹⁴ Recently emphasized, there is now wide variation in the agreements that investigators complete with the pharmaceutical industry, and many prohibit independent disclosure and publication of all trial data.¹⁵

It is in the context of this history that I hope you will read Dr. Johnston's work. Dr. Johnston's model for severity adjustment uses cubic splines and other contemporary methods to build a new sepsis-specific risk-adjustment model based on the placebo patients in PROWESS. He compares it to the "Knaus" model, which is not the sepsis model we published in 1993.⁴ The Knaus model referred to by Dr. Johnston uses the APACHE II variables along with the addition of the timing of treatment.

Dr. Johnston's 4th "new specific-disease" model adds new physiologic variables, prothrombin time and platelet count, that are closely linked to the sepsis cascade and are important additions. Dr. Johnston's 4th model performs better than the Knaus model but not quite as well as our 1993 sepsis model. When we applied our 1993 sepsis-specific model to an independent database, the resulting Receiver Operating Characteristic area was 0.80 compared to Dr. Johnston's model performance of 0.79. I disagree when Dr. Johnston concludes that the APACHE II score used in PROWESS does not perform as well as a disease-specific severity model because it was designed for 1st-day application in the intensive care unit and many septic patients are treated later. Although the interval between onset of sepsis and therapy is an important variable, the more important fact is that APACHE II contains fewer variables, each weighted less accurately than in his or our work on APACHE since 1985.

This issue has important considerations beyond APACHE, the statistics of sepsis, and the prescribing of Xigris. Activated protein C is just one of the early entrants of what is going to be an increasingly common scenario in clinical research. As molecular and genetic research improves, more and more drugs like Xigris that attack very specific components of large complex pathologic pathways in patient populations with enormous heterogeneity will be developed and tested. To efficiently and accurately test and ultimately prescribe these compounds, I contend we need to reconsider how carefully we describe our patients within the clinical trials, how we design our primary analyses, and how we link them to subsequent indications for use. I believe we will have to be more specific in our description of patients and not assume that randomization and a statistically significant *P* value are sufficient to control for confounding.

We will also have to be more cautious in applying a one-size-fits-all approach to prescribing these agents. We will need to keep in mind that these new compounds may be dramatically life saving for a few select individuals, but they may also be ineffective or harmful for other patients—some with very similar but not identical clinical presentations.

At the time of approval, the FDA called for a 13,000-patient postmarketing trial aimed at evaluating Xigris's safety and efficacy for patients at low levels of severity, in children, and in combination with low-dose heparin.⁹ An 11,000-patient study in low-risk hospitalized adults with sepsis, called the ADDRESS (Administration of Drotrecogin Alfa [Activated] Early Stage Severe Sepsis) trial, began enrollment in September 2002 and was stopped at the recommendation of the data-monitoring committee in February 2004 when 2640 patients (1033 in the US) had been enrolled.¹⁶ The overall 28-d mortality for the Xigris group was 18.5% versus 17.0% for placebo (odds ratio [OR] 1.08, *P* = 0.34). There was also a significant increase in serious bleeding events in the first 7 d for the Xigris group (2.4% v. 1.2%, *P* = 0.02), although there was no significant difference in intracranial hemorrhage (0.3% v. 0.2%, *P* = 0.72). A phase III trial of Xigris in children was also recently halted because of a similar lack of benefit and an increased incidence of hemorrhage in the treatment arm.

These results make Dr. Johnston's study all the more relevant and raise the issue that, in this age of computer information, computer-based prediction rules may be necessary for the optimal use of molecules such as Xigris. The initial and ongoing controversy over prescribing indications may have led to the substantially

less-than-optimal use of Xigris for patients with severe sepsis. In addition to ensuring that all results from trials are published, we may also need to consider additional regulatory steps to strengthen the FDA to ensure that appropriate treatment guidelines are prominently and consistently described for the profession and the public as drugs are marketed.

In conclusion, I would encourage future evaluations of therapies of complex diseases to consider incorporating comprehensive risk modeling into trial design as part of the primary analysis. This is not a current recommended standard for trial design and reporting.¹⁷ Since our publication demonstrate improved power using a prespecified comprehensive risk model on a sepsis trial population, however, there are at least 2 other articles that also found risk adjustment increased effect size in experiments in other disease conditions.^{18,19} I would also urge that the model used be evaluated on a recently treated patient cohort. In the forthcoming release of APACHE IV, there are substantial improvements in outcomes for some of the major disease categories admitted to intensive care units in the decade between APACHE III and IV (Jack Zimmerman, personal communication, April 2005). Prognostic models need to be current with contemporary therapeutic capabilities even as they help us decide how we might improve on these outcomes.

REFERENCES

1. Johnston JA. Determinants of mortality in patients with severe sepsis. *Med Decis Making*. 2005;4:374–86.
2. Bernard GR, Vincent J-L, Laterre P-F, et al. Efficacy and safety of recombinant human activated protein C for severe sepsis. *N Engl J Med*. 2001;344:699–709.
3. Matthay MA. Severe sepsis—a new treatment with both anticoagulant and anti-inflammatory properties. *N Engl J Med*. 2001;344(10):759–62.
4. Knaus WA, Harrell FE, Fisher CJ, et al. The clinical evaluation of new drugs for sepsis. *JAMA*. 1993;270:376–93.
5. Gail MH, Wieand S, Piantadosi S. Biased estimates of treatment effect in randomized experiments with nonlinear regressions and omitted covariates. *Biometrika*. 1984;71:431–44.
6. American College of Chest Physicians/Society of Critical Care Medicine Consensus Conference. Definitions for sepsis and organ failure and guidelines for the use of innovative therapies in sepsis. *Crit Care Med*. 1992;20:864–74.
7. Knaus WA, Draper EA, Wagner DP, Zimmerman JE. APACHE II: a severity of disease classification system. *Crit Care Med*. 1985;13:818–29.
8. Knaus WA, Harrell FE, LaBrecque MS, et al. Use of predicted risk of mortality to evaluate the efficacy of anticytokine therapy in sepsis. *Crit Care Med*. 1996;24:46–56.
9. Siegel JP. Assessing the use of protein C in the treatment of severe sepsis. *N Engl J Med*. 2002;347:1030–4.
10. Rice TW, Bernard GR. Drotrecogin alfa (activated) for the treatment of severe sepsis and septic shock. *Am J Med Sci*. 2004;328(4):205–14.
11. Bernard GR, Margolis BD, Shanies HM, et al. Extended evaluation of recombinant human activated protein C United States trial (ENHANCE US)—a single-arm, phase 3B, multicenter study of drotrecogin alfa (activated) in severe sepsis. *Chest*. 2004;125:2206–16.
12. Angus DC, Laterre P-F, Hetterbrand J, et al. The effect of drotrecogin alfa (activated) on long-term survival after severe sepsis. *Crit Care Med*. 2004;32(11):2199–206.
13. Angus DC, Linde-Zwirble WT, Clermont G, et al. Cost-effectiveness of drotrecogin alfa (activated) in the treatment of severe sepsis. *Crit Care Med*. 2003;31(1):1–11.
14. Mathews AW. Worrisome ailment in medicine: misleading journal articles. *Wall Street Journal*. 10 May 2005;245(91):A1–9.
15. Steinbrook R. Gag clauses in clinical trial agreements. *N Engl J Med*. 2005;352(21):2160–2.
16. Deans KJ, Minneci PC, Banks SM, Natanson C, Eichacker PQ. Substantiating the concerns about recombinant human activated protein C use in sepsis. *Crit Care Med*. 2004;32:2542–3.
17. Altman DG, Schulz KF, Moher D, et al. The revised CONSORT statement for reporting randomized trials: explanation and elaboration. *Ann Intern Med*. 2001;134(8):663–94.
18. Johnson KC, Connors AF Jr, Wagner DP, Haley EC Jr. Risk adjustment effect on stroke clinical trials. *Stroke*. 2004;35:e43–5.
19. Steyerberg EW, Bossuyt PMM, Lee KL. Clinical trials in acute myocardial infarction: should we adjust for baseline characteristics? *Am Heart J*. 2000;139:745–51.