

COMMENTARY

Severe sepsis: are PROWESS and PROWESS-SHOCK trials comparable? A clinical and statistical heterogeneity analysis

Andre C Kalil* and Diana F Florescu

Abstract

Despite the same manufacturer, the same drotrecogin alfa activated dose, and the same placebo-controlled design, the negative result from the PROWESS-SHOCK trial contradicted the survival benefit observed in the PROWESS trial. We hypothesize that the different results were due to factors other than the experimental therapy and performed an analysis of the clinical heterogeneity (differences related to the trials' clinical aspects) and the statistical heterogeneity (differences related to the trials' statistical aspects) between these trials. Baseline characteristics and co-interventions were analyzed by chi-square testing and mortality was analyzed by random-effects modeling and I^2 . Our findings show that clinical variables presented significant heterogeneity, and that up to 90% of the mortality differences between both trials were not due to chance. These results demonstrate that PROWESS and PROWESS-SHOCK are not comparable trials due to the highly significant clinical and statistical heterogeneity. We propose a new and pragmatic solution.

Drotrecogin alfa activated (DAA) was approved for treatment of patients with severe sepsis in 2001 based on a large randomized double-blind clinical trial, PROWESS [1]. A second randomized clinical trial, PROWESS-SHOCK [2], was recently completed, but the survival benefits observed with the original trial were not reproduced in the second trial. The results from PROWESS-SHOCK culminated in the removal of this drug from the market in 2011. Both trials were multicenter, randomized, and double-blinded, and both used the same drug manufacturer. Which trial should we believe in? What

should be done while one-third of our patients with severe sepsis are still dying despite the best standard of care? Our paper aims to explore the reasons for this discrepancy and offer new solutions.

A total of 3,370 patients with severe sepsis were enrolled in both trials, for which we performed an analysis of the clinical heterogeneity (differences related to the trials' clinical aspects) and the statistical heterogeneity (differences related to the trials' statistical aspects) between these trials. Baseline characteristics, infection etiologies and sites, and co-interventions were compared by chi-square testing for the clinical analysis, while random-effects modeling and I^2 were performed for the statistical analysis. All results are shown in Tables 1 and 2. Our clinical findings demonstrate that infection sites, etiology, co-interventions, and geographic enrollment were all significantly different between the two trials. Moreover, the use of appropriate antibiotics, low-dose steroids, and heparin were all significantly different. Based on 28-day mortality, we also found a highly significant statistical heterogeneity: up to 90% of the mortality differences between the trials were not due to chance. This heterogeneity remained consistently high even when the analysis was done by shock status, number of organ failures, or APACHE II (Acute Physiology and Chronic Health Evaluation II) scores.

A recent study by Levy and colleagues [3] showed that the absolute mortality of severe sepsis is different between the US and EU; hence, the very different rates of geographic enrollment between PROWESS and PROWESS-SHOCK may explain, in part, the mortality differences. Infection site is known to be a major determinant of survival in patients with severe sepsis, so the significant differences we found regarding these sites provide more corroborating evidence of clinical heterogeneity. The diversity of microbiological etiologies between studies also provides evidence for the clinical differences; similarly, the rate of appropriate use of antibiotics was not comparable. Let alone the fact that co-interventions (for example, heparin and low-dose steroids) were significantly different between trials.

*Correspondence: akalil@unmc.edu
Infectious Diseases Division, University of Nebraska Medical Center, Omaha, NE 68198-5400, USA

Table 1. Clinical heterogeneity: PROWESS versus PROWESS-SHOCK - trial characteristics

Trial characteristic	Trial	Proportion with characteristic	Comparison P-value
Enrolled in North America	PROWESS	54.7%	<i>P</i> < 0.0001
	PROWESS-SHOCK	14.1%	
Chronic obstructive pulmonary disease	PROWESS	24.1%	<i>P</i> < 0.0001
	PROWESS-SHOCK	15.5%	
Lung as site of infection	PROWESS	53.6%	<i>P</i> < 0.0001
	PROWESS-SHOCK	43.9%	
Presence of shock	PROWESS	71%	<i>P</i> < 0.0001
	PROWESS-SHOCK	100%	
<i>S. pneumoniae</i> infections	PROWESS	11.9%	<i>P</i> < 0.0001
	PROWESS-SHOCK	4.5%	
<i>S. aureus</i> infections	PROWESS	14.3%	<i>P</i> < 0.0001
	PROWESS-SHOCK	3.7%	
<i>Escherichia coli</i> infections	PROWESS	16.5%	<i>P</i> < 0.0001
	PROWESS-SHOCK	10.3%	
<i>Klebsiella</i> spp. infections	PROWESS	6.2%	<i>P</i> < 0.0001
	PROWESS-SHOCK	2.2%	
Use of appropriate antibiotics	PROWESS	91.2%	<i>P</i> < 0.0001
	PROWESS-SHOCK	84.2%	
Use of steroids for septic shock	PROWESS	13.5%	<i>P</i> < 0.0001
	PROWESS-SHOCK	49.5%	
Use of heparin	PROWESS	75.2%	<i>P</i> < 0.0001
	PROWESS-SHOCK	44.7%	
Serious bleeding with drotrecogin	PROWESS	3.5%	<i>P</i> = 0.0017
	PROWESS-SHOCK	1.2%	

Table 2. Statistical heterogeneity: PROWESS versus PROWESS-SHOCK - 28-day mortality

Patient population	Trial	Risk ratio (95% CI)	Heterogeneity P-value	Heterogeneity I-square
All	PROWESS	0.80 (0.69-0.94)	0.008	85.8%
	PROWESS-SHOCK	1.09 (0.92-1.28)		
	Overall	0.93 (0.69-1.26)		
Shock Only	PROWESS	0.77 (0.65-0.92)	0.005	87.4%
	PROWESS-SHOCK	1.09 (0.92-1.28)		
	Overall	0.92 (0.65-1.29)		
APACHE II ≥ 25	PROWESS	0.71 (0.59-0.85)	0.002	89.9%
	PROWESS-SHOCK	1.11 (0.90-1.37)		
	Overall	0.88 (0.57-1.37)		
Organ failure ≥ 3	PROWESS	0.77 (0.63-0.94)	0.017	82.3%
	PROWESS-SHOCK	1.04 (0.90-1.20)		
	Overall	0.90 (0.68-1.21)		

CI, confidence interval.

The statistical heterogeneity analysis demonstrates that the vast majority (80 to 90%) of the detected heterogeneity regarding survival outcomes between these trials could not be explained by chance. This is quite remarkable because it points out that the reasons for this large heterogeneity derive from differences related to the trials themselves, in this case, patient population, baseline infection, and co-interventions. Moreover, even after we stratified the survival outcome analysis by disease severity, the elevated heterogeneity did not change. We conjecture that the different clinical characteristics and co-interventions were most likely the cause for this irreversible statistical heterogeneity.

Another complicating factor is that the PROWESS-SHOCK trial was substantially underpowered: 42% probability of false-negative results. Although a frequent question since the completion of PROWESS-SHOCK has been ‘which trial should we believe in?’, we propose that this is not the 64 million-dollar question; the one that is begging for an answer is ‘will we ever be able to replicate the design of the PROWESS trial?’ If we aim for a control mortality of 35%, the answer is ‘no’ because recent phase III trials [2,4] have shown that the mortality associated with severe sepsis now ranges from 24 to 28%. If we slightly modify the question to ‘can we perform another phase III trial on DAA with adequate statistical power?’

the answer is yes on two accounts: 1) a large sample size ($N = 2,500$ to $3,000$) would fulfill the frequentist (classical) statistical approach; and 2) a smaller sample size (500 to $1,000$) would fulfill the adaptive Bayesian statistical approach, as we explained in a previous manuscript [5]. What about financial support? The financial and logistic challenges would be enormous for the frequentist approach, but definitely more feasible for the Bayesian approach. Would it be ethical to perform a third trial? Yes, a study we published recently [6] demonstrated that, in real-life application outside phase III trials, DAA significantly reduced in-hospital mortality by 18% (95% confidence interval 13 to 22%) in patients with severe sepsis ($N = 41,401$ patients). How would this trial be designed? First, an individual-patient data meta-analysis to combine all randomized trials at the patient-level would provide the most accurate and statistically powerful way to reduce the current scientific uncertainty; second, the concomitant use of both frequentist and Bayesian methodologies [7] would maximize the opportunity to gather the most valuable scientific information on the efficacy of DAA; and third, the findings from this new analysis would provide the necessary tools to optimize the design of the next randomized trial. Thus, it is our responsibility to not stop our scientific investigation here, especially considering that the 3,370 patients who gave their consent to participate in these clinical trials were assured that their information would be fully utilized for fostering progress in medical science and for the betterment of future patients afflicted by severe sepsis.

Conclusion

PROWESS and PROWESS-SHOCK trials are not comparable based on both clinical and statistical heterogeneity. Hence, the true effect of DAA in patients with severe sepsis remains to be defined. Unless the totality of the available evidence is thoroughly evaluated through an individual-patient data meta-analysis, and an adaptive

Bayesian clinical trial is performed, we will continue treating our patients with the appalling sensation that we are not improving their survival due to our own inability to advance the quality of clinical research in the sepsis field.

Abbreviations

DAA, drotrecogin alfa activated.

Competing interests

ACK has received research grant support from Eisai, AstraZeneca, Agennix, and Asahi Kasei. DFF has no competing interests.

Published: 4 July 2013

References

1. Bernard GR, Vincent JL, Laterre PF, LaRosa SP, Dhainaut JF, Lopez-Rodriguez A, Steingrub JS, Garber GE, Helterbrand JD, Ely EW: **Efficacy and safety of recombinant human activated protein C for severe sepsis.** *N Engl J Med* 2001, **344**:699-709.
2. Ranieri VM, Thompson BT, Barie PS, Dhainaut JF, Douglas IS, Finfer S, Gardlund B, Marshall JC, Rhodes A, Artigas A, Payen D, Tenhunen J, Al-Khalidi HR, Thompson V, Janes J, Macias WL, Vangerow B, Williams MD, PROWESS-SHOCK Study Group: **Drotrecogin alfa (activated) in adults with septic shock.** *N Engl J Med* 2012, **366**:2055-2064.
3. Levy MM, Artigas A, Phillips GS, Rhodes A, Beale R, Osborn T, Vincent JL, Townsend S, Lemeshow S, Dellinger RP: **Outcomes of the Surviving Sepsis Campaign in intensive care units in the USA and Europe: a prospective cohort study.** *Lancet Infect Dis* 2012, **12**:919-924.
4. Opal SM, Laterre PF, Francois B, LaRosa SP, Angus DC, Mira JP, Wittebole X, Dugernier T, Perrotin D, Tidswell M, Jauregui L, Krell K, Pachi J, Takahashi T, Peckelsen C, Cordasco E, Chang CS, Oeyen S, Aikawa N, Maruyama T, Schein R, Kalil AC, Van Nuffelen M, Lynn M, Rossignol DP, Gogate J, Roberts MB, Wheeler JL, Vincent JL, ACCESS Study Group: **Effect of eritoran, an antagonist of MD2-TLR4, on mortality in patients with severe sepsis: the ACCESS randomized trial.** *JAMA* 2013, **309**:1154-1162.
5. Kalil AC, Sun J: **How many patients with severe sepsis are needed to confirm the efficacy of drotrecogin alfa activated? A Bayesian design.** *Intensive Care Med* 2008, **34**:1804-1811.
6. Kalil AC, LaRosa SP: **Effectiveness and safety of drotrecogin alfa (activated) for severe sepsis: a meta-analysis and metaregression.** *Lancet Infect Dis* 2012, **12**:678-686.
7. Kalil AC, Sun J: **Why are clinicians not embracing the results from pivotal clinical trials in severe sepsis? A bayesian analysis.** *PLoS ONE* 2008, **3**:e2291.

doi:10.1186/cc12752

Cite this article as: Kalil AC, Florescu DF: Severe sepsis: are PROWESS and PROWESS-SHOCK trials comparable? A clinical and statistical heterogeneity analysis. *Critical Care* 2013, **17**:167.