Published in final edited form as:

Crit Care Med. 2012 November ; 40(11): 3090–3092. doi:10.1097/CCM.0b013e31826536b7.

Resolving conflicting comparative effectiveness research in critical care

Christopher W. Seymour, MD MSc¹ and Jeremy M. Kahn, MD MSc^{1,2,3}

¹Clinical Research, Investigation and Systems Modeling of Acute Illness (CRISMA) Center, Department of Critical Care Medicine, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania, USA

²Department of Critical Care Medicine, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania, USA

³Department of Health Policy & Management, University of Pittsburgh Graduate School of Public Health, Pittsburgh, Pennsylvania, USA

Keywords

Randomized trials; observational studies; comparative effectiveness research; drotrecogin alfa; propensity score

Comparative effectiveness research (CER) has received keen interest from healthcare professionals, funding agencies, and researchers as a way to improve healthcare quality, efficiency, and costs. Broadly defined, CER includes both observational studies and randomized controlled trials (RCTs). Of these, RCTs are widely considered near the top of the hierarchy of scientific evidence. In critical care, however, RCTs are clearly not feasible for every research question. They are expensive, time consuming, restricted to narrow patient populations, and usually conducted in an academic community setting. All of these issues limit their practicality and generalizability.

As a result, observational research is an important and increasing complement to RCTs as we develop best practice in the intensive care unit (ICU).(1) In contrast to RCTs, observational research is often more efficient and generalizable.(2) Administrative databases and clinical registries offer opportunities to study large, diverse patient populations with relative ease, and emerging statistical techniques such as propensity scores and instrumental variables purport to overcome the indication biases that traditionally plague observational studies of treatment effects.(3) Yet even in these cases, the results of observational trials often conflict with RCTs, leading to erroneous conclusions and inappropriate treatment recommendations.(4) In these cases, it is left to the consumers of research—patients, clinicians and policy makers—to reconcile the results.

Contact: Christopher W. Seymour, M.D., M.Sc., Assistant Professor, Departments of Critical Care and Emergency Medicine, University of Pittsburgh School of Medicine, Clinical Research, Investigation, and Systems Modeling of Acute illness (CRISMA) Center, Pittsburgh, PA, 3550 Terrace St, Scaife Hall, #639, Pittsburgh PA 15261, Phone: 425-864-2993, Fax: (412) 647-8060, seymourcw@upmc.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Financial disclosures: The authors have no conflicts of interest to disclose.

In this issue of *Critical Care Medicine*, Rimmer and colleagues present us with that very challenge.(ref) They report a multi-national, multi-center observational study of recombinant human activated protein C (rhAPC, drotrecogin alfa) among 7392 subjects with septic shock enrolled from 1997 to 2007. They observed an overall low rate of rhAPC use among patients with septic shock (4.7%), and a significant association between rhAPC and 30-day mortality in a propensity matched analysis (hazard ratio = 0.72, 95% CI 0.52–1.00, p=0.05). Such findings in favor of rhAPC use are consistent an early randomized trial and multiple subsequent observational studies.(5, 6) And yet, Eli Lilly and Company conducted a confirmatory, randomized trial (PROWESS-SHOCK) to refine target population and risk-benefit ratios for rhAPC. This trial found no significant benefit for rhAPC on 28-day mortality, and the drug was withdrawn worldwide.(7)

So when studies conflict, what is the “right” answer? The first inclination is to rely on RCTs, and indeed there are many reasons why their conclusions are likely to be most reliable and valid. In contrast to other study designs, through the act of randomization RCTs protect against a common source of confounding termed “indication bias.” This bias arises from the notion that patients who receive a treatment are fundamentally different than those who don’t.(8) For example, ICU patients with severe sepsis who require vasopressors are more severely ill than those who don’t. If we were to study the outcomes of patients who received vasopressors against controls in a large cohort study, our results may typically show that vasopressors are associated with harm. Yet these results would likely be due to the fact that treated patients were sicker in ways that cannot be measured through traditional approaches.

To attempt to overcome indication bias in their observational study of rhAPC, Rimmer and colleagues applied propensity score methods. They used all the variables they could measure to predict the probability a patient would receive rhAPC (“the propensity score”). Then they matched subjects receiving rhAPC to control subjects who did not on the propensity score. The propensity score matching helps balance baseline characteristics – a sort of pseudo-randomization. Yet in the end, propensity score methods are only as good as the variables measured in the dataset. The cannot account for unmeasured covariates, especially the subtle “unmeasurables” that govern the decision to prescribe a drug like rhAPC. Thus they are far from a panacea as a method to remove bias.(3) Only randomization can fully overcome these biases, supporting the role of RCTs as the best available clinical evidence (5).

Yet there are also valid reasons why observational studies can also be “right”, and can indeed provide better data than even RCTs. Unlike RCTs, which typically are conducted in the academic setting under ideal conditions and with innumerable exclusion criteria, observational studies provide invaluable knowledge of practice patterns, treatment effects among “real” patients, and study phenotypes or sub-phenotypes not necessarily enrolled in pre-approval randomized trials. They also may study patient-centered outcomes not mandated by drug approval agencies, and could include more patients with greater power to detect small differences in treatment effects. Negative RCTs may be negative not because the drug doesn’t work, but because the study was not conducted in the right setting, in the right patients, or with the power to detect clinically meaningful effect sizes in small subgroups. Ignoring this reality downplays the important role of observational studies in CER, so long as that research is rigorously conducted. Recently, the Patient-Centered Outcomes Research Institute (PCORI) offered methodological standards for observational research that compliment standards for reporting.(9, 10) These provide a roadmap for the future of comparative effectiveness research in critical care, without always relying on a randomized design.

The correct answer, then, is that neither RCTs nor observational research is always “right”. Instead, we should approach each individual study with a critical eye weighing its strengths and limitations in the context of the previous literature and our prior beliefs. This is the idea of “Bayesian reasoning”—rather than providing irrefutable evidence, new clinical research should serve to inform and improve our decisions, not completely dictate our practice.⁽¹¹⁾ Indeed, up to one third of the most cited clinical research trials are refuted or found to overestimate treatment effects, including many in critical care.^(12–14) Rather than expect perfection out of research, both clinicians and researchers must sharpen their interpretations and be wary of the strengths and weaknesses of randomized and non-randomized study designs. We must move away from labeling studies “right” or “wrong”. As long as they are well-conducted, conflicting studies provide complimentary findings that build a patchwork of evidence with which to guide our decisions.

Of course, in the case of rhAPC this discussion is purely academic. Now off the market, the drug will not be given to patients in septic shock. Yet the conflicting evidence between early RCTs and post-approval observational studies is a warning for when we encounter the next rhAPC. The only “wrong” action is to be overly optimistic about a single RCT or to overly dismiss observational studies because patients are not randomized. Just like at the bedside in the ICU, our interpretation of comparative effectiveness research should consider the entire body of evidence.

References

1. Rubenfeld GD, Angus DC, Pinsky MR, et al. Outcomes research in critical care: results of the American Thoracic Society Critical Care Assembly Workshop on Outcomes Research. The Members of the Outcomes Research Workshop. *Am J Respir Crit Care Med*. 1999; 160(1):358–367. [PubMed: 10390426]
2. Concato J, Shah N, Horwitz RI. Randomized, controlled trials, observational studies, and the hierarchy of research designs. *N Engl J Med*. 2000; 342(25):1887–1892. [PubMed: 10861325]
3. Stukel TA, Fisher ES, Wennberg DE, et al. Analysis of observational studies in the presence of treatment selection bias: effects of invasive cardiac management on AMI survival using propensity score and instrumental variable methods. *JAMA*. 2007; 297(3):278–285. [PubMed: 17227979]
4. Rossouw JE, Anderson GL, Prentice RL, et al. Risks and benefits of estrogen plus progestin in healthy postmenopausal women: principal results From the Women's Health Initiative randomized controlled trial. *JAMA*. 2002; 288(3):321–333. [PubMed: 12117397]
5. Lindenauer PK, Rothberg MB, Nathanson BH, et al. Activated protein C and hospital mortality in septic shock: a propensity-matched analysis. *Crit Care Med*. 2010; 38(4):1101–1107. [PubMed: 20154607]
6. Martin G, Brunkhorst FM, Janes JM, et al. The international PROGRESS registry of patients with severe sepsis: drotrecogin alfa (activated) use and patient outcomes. *Crit Care*. 2009; 13(3):R103. [PubMed: 19566927]
7. Ranieri VM, Thompson BT, Barie PS, et al. Drotrecogin alfa (activated) in adults with septic shock. *N Engl J Med*. 2012; 366(22):2055–2064. [PubMed: 22616830]
8. Salas M, Hofman A, Stricker BH. Confounding by indication: an example of variation in the use of epidemiologic terminology. *Am J Epidemiol*. 1999; 149(11):981–983. [PubMed: 10355372]
9. Vandenbroucke JP, von Elm E, Altman DG, et al. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. *Ann Intern Med*. 2007; 147(8):W163–W194. [PubMed: 17938389]
10. Methodological standards and patient-centeredness in comparative effectiveness research: the PCORI perspective. *JAMA*. 2012; 307(15):1636–1640. [PubMed: 22511692]
11. Brophy JM, Joseph L. Placing trials in context using Bayesian analysis. GUSTO revisited by Reverend Bayes. *JAMA*. 1995; 273(11):871–875. [PubMed: 7869558]
12. Ioannidis JP. Contradicted and initially stronger effects in highly cited clinical research. *JAMA*. 2005; 294(2):218–228. [PubMed: 16014596]

13. Bernard GR, Vincent JL, Laterre PF, et al. Efficacy and safety of recombinant human activated protein C for severe sepsis. *N Engl J Med*. 2001; 344(10):699–709. [PubMed: 11236773]
14. van den Berghe G, Wouters P, Weekers F, et al. Intensive insulin therapy in critically ill patients. *N Engl J Med*. 2001; 345(19):1359–1367. [PubMed: 11794168]