

example, a propensity analysis performed on a US database created a lot of turmoil when it raised serious concerns about the use of the pulmonary artery catheter,⁵ but a similar analysis on the European Sepsis Occurrence in Acutely Ill Patients database using more meticulous adjustment did not yield similar findings.⁶

Thanks to the large database from the Sepsis Occurrence in Acutely Ill Patients study,⁷ the propensity analysis in our recent article² included a large number of variables, allowing extensive and reliable adjustment for confounders. Unfortunately, the accompanying editorial by Drs. Nuttall and Houle¹ was misleading when it suggested that the time to survival was considered as from intensive care unit admission: As stated in the article,² the time to survival was counted from the day on which patients received a blood transfusion, and patients who were not transfused were censored at the time of intensive care unit discharge; hence, the time factor was taken into account in our analysis. Drs. Nuttall and Houle correctly stated that performance of a propensity analysis is weak when there are seven or fewer events per confounding variable, but this limitation does not apply when the number of patients is large, as in our study.⁸ The editorial also described the statistical process as being opaque, simulating a “black box”; this could have been a relevant argument if covariate adjustment were performed using stratification according to propensity scores or the simple use of this score as a covariate in a multivariate analysis, where adjustment is not performed at the individual level and results cannot be examined for balance between treatment groups.⁹ However, we performed case matching, with an excellent match between patients (table 6 of our article²).

Therefore, the editorialists are correct in underlining the potential limitations of a propensity analysis, but we believe our statistical

methodology was strong enough to respond to most of these concerns.

Jean-Louis Vincent, M.D., Ph.D.,* Yasser Sakr, M.B., B.Ch., Ph.D. *Erasme Hospital, Université libre de Bruxelles, Brussels, Belgium. jlvincen@ulb.ac.be

References

1. Nuttall GA, Houle TT: Liars, damn liars, and propensity scores. *ANESTHESIOLOGY* 2008; 108:3–4
2. Vincent JL, Sakr Y, Sprung CL, Harboe S, Damas P: Are blood transfusions associated with greater mortality rates? *ANESTHESIOLOGY* 2008; 108:31–9
3. Deans KJ, Minneci PC, Suffredini AF, Danner RL, Hoffman WD, Ciu X, Klein HG, Schechter AN, Banks SM, Eichacker PQ, Natanson C: Randomization in clinical trials of titrated therapies: Unintended consequences of using fixed treatment protocols. *Crit Care Med* 2007; 35:1509–16
4. Hébert PC, Wells G, Blajchman MA, Marshall J, Martin C, Pagliarello G, Tweeddale M, Schweitzer I, Yetisir E: A multicenter, randomized, controlled clinical trial of transfusion requirements in critical care. *N Engl J Med* 1999; 340:409–17
5. Connors AF, Speroff T, Dawson NV, Thomas C, Harrell FE, Wagner D, Desbiens N, Goldman L, Wu AW, Califf RM, Fulkerson WJ, Vidaillet H, Broste S, Bellamy P, Lynn J, Knaus W, SUPPORT Investigators: The effectiveness of right heart catheterization in the initial care of critically ill patients. *JAMA* 1996;276:889–97
6. Sakr Y, Vincent JL, Reinhart K, Payen D, Wiedermann CJ, Zandstra DF, Sprung CL: Use of the pulmonary artery catheter is not associated with worse outcome in the intensive care unit. *Chest* 2005; 128:2722–31
7. Vincent JL, Sakr Y, Sprung CL, Ranieri VM, Reinhart K, Gerlach H, Moreno R, Carlet J, Le Gall JR, Payen D: Sepsis in European intensive care units: Results of the SOAP study. *Crit Care Med* 2006; 34:344–53
8. Cepeda MS, Boston R, Farrar JT, Strom BL: Comparison of logistic regression versus propensity score when the number of events is low and there are multiple confounders. *Am J Epidemiol* 2003; 158:280–7
9. Rosenbaum PR, Rubin DB: The central role of the propensity score on observational studies for causal effects. *Biometrika* 1983; 70:41–55

(Accepted for publication April 30, 2008.)

Anesthesiology 2008; 109:745–6

Copyright © 2008, the American Society of Anesthesiologists, Inc. Lippincott Williams & Wilkins, Inc.

The Name of the Game: No Transfusion (or Nontransfusion) by Cookbook

To the Editor:—The editorial by Nuttall and Houle¹ on the article by Vincent *et al.*² is long on method but short on biology. Nonstatisticians—the majority of readers—will be trying to get the article in clinical context. The editorial does not help them in this, and its pejorative title gets it off to a bad start. Nuttall and Houle¹ give a useful assessment of propensity scoring (in general) but barely mention the data (in this study), and so risk giving the reader the impression that the content should be given only limited credence. An editorial that gave more prominence to the biology would have collated the evidence and achieved broader perspective.

The article of Vincent *et al.*² is a hypothesis-generating study that questions the current consensus on erythrocyte transfusion therapy, in a similar manner to the findings of Connors *et al.*³ (with respect to pulmonary artery catheterization) and, more recently, Karkouti *et al.*⁴ and Mangano *et al.*⁵ (with respect to aprotinin in cardiac surgery). The conclusions of Vincent *et al.*² may be disturbing, but to summarize the study with the truism “interpret with caution”—on the grounds of methodology—is an incomplete response that serves nobody. The key question is: Do the article’s findings reflect flawed methods, or do they suggest a problem with generalizability (e.g., might previous data derived from randomized controlled trials (RCTs) be driving current practice inappro-

riately)? The editorialists omit the latter possibility altogether, which is unfortunate because it may be the most important lesson from the article.

In looking at two studies with disparate results, such as those of Vincent *et al.*² and the landmark Transfusion in Critical Care (TRICC) study,⁶ the most useful initial response is to try to understand how they can be reconciled, or how what was apparently true before might not be true now. Transfusion practice has changed as a result of the TRICC trial, and transfusion of leukodepleted erythrocytes is now widespread. If these changes are truly beneficial, we would expect the impact of transfusion decisions to change also, with a reduction in “harmful transfusion.” If the changes in practice had resulted in overly conservative decision making, we might observe an increase in harm from “harmful nontransfusion.” Successful RCTs that are followed by evidence of a “downside” are not novel; the Randomized Aldactone Evaluation Study,⁷ which showed improved survival in patients receiving spironolactone, was followed by observational data suggesting an increase in morbidity and mortality from hyperkalemia.⁸ Although the “harm” component in the TRICC trial seemed to stem from liberal transfusion in younger, healthier patients, later analysis suggested possible harm also from *not* transfusing in TRICC participants with known coronary artery disease.⁹

The editorialists are right to address study methodology, but they should not leave the reader with an indictment of propensity scores and, by extension, observational studies. When discussing methodologic issues, we should keep in mind the suggestion that recent high-quality observational studies and RCTs often arrive at similar conclusions,¹⁰ the fact that highly cited randomized trials may produce

The above letter was sent to the author of the referenced article by Vincent *et al.* The author did not feel that a response was required.—James C. Eisenach, M.D., Editor-in-Chief.

incorrect or exaggerated results,¹¹ and the suggestion that the durability of medical knowledge is unrelated to methodologic quality.¹²

Even the best observational study is limited by an inability to draw causal inferences and by the presence of confounders. RCT design takes causality as a given and puts its trust in an ability to minimize—of course it does not eliminate—confounders by randomization. But the problem of “unknown unknowns” remains, and the greater the number of unknown confounders that exist, the greater the likelihood of an imbalance. This problem is common to RCTs and observational studies alike and is probably most likely in small studies where our understanding of disease pathogenesis is limited. In a study with total $n \approx 1,600$, where five independent confounders exist, each with an incidence of 20%, the probability of an imbalance for at least one confounder is almost 25%.¹³ So studies A and B might disagree because A has greater balance of unknown confounders than B, and thus a better balance of confounders in a large observational study might “trump” randomization in a small RCT. This does not upgrade the status of observational studies, but it does explain why well-designed observational studies often arrive at similar conclusions relative to RCTs, and why some of the time they will correctly contradict previous RCT data. The controversial articles by Karkouti *et al.*⁴ and Mangano *et al.*⁵ may exemplify this—as suggested by the results of the recent Blood Conservation Using Antifibrinolytics in a Randomized Trial.¹⁴

The article of Vincent *et al.* discusses whether leukoreduction might account for the findings but provides no data²; the editorial does not mention it.¹ Neither the original article nor the editorial provides any convincing explanation (*i.e.*, biologic basis) for the reported effect. We wonder whether additional analysis of the data in the article of Vincent *et al.*² might shed light on whether leukoreduction may be responsible for the apparently altered impact of transfusion, as has been suggested previously.^{15,16}

The data of Vincent *et al.*² and the recent TRICC reanalysis by Deans *et al.*⁹ suggest that outcome is changing over time and that the interpretation of the TRICC trial is more complex than we thought. It will be some time before we get a clearer picture, but in the meantime, we should not treat propensity scoring as a straw man. Reading the article of Vincent *et al.*,² we experience the judgment under uncertainty that pervades clinical life. Decisions to transfuse—and not to transfuse—are not made lightly, so it is a truism that these data should be viewed with caution. The function of the article, however, is to make us view with caution things that we think we know.

John F. Boylan, M.B., F.R.C.P.C.,* Brian P. Kavanagh, M.B., F.R.C.P.C. *St. Vincent's University Hospital, Dublin, Ireland. jboylan@iol.ie

References

1. Nuttall GA, Houle TT: Liars, damn liars, and propensity scores. *ANESTHESIOLOGY* 2008; 108:3–4

Anesthesiology 2008; 109:746–7

Copyright © 2008, the American Society of Anesthesiologists, Inc. Lippincott Williams & Wilkins, Inc.

Propensity Scores Do Not Necessarily Lie!

To the Editor:—Recently, an Editorial View was published on propensity score methods.¹ The editorial describes strengths and weaknesses of propensity score methods in observational therapeutic studies. The authors apparently refer in their title to a quote said by the English Prime Minister Benjamin Disraeli (1804–1880) in the 19th century: “There are lies, damn lies and statistics.” In general, we appreciate links to statements from outside the clinical research world. However, the title of the Editorial View may be misinterpreted as a statement against

2. Vincent JL, Sakr Y, Sprung C, Harboe S, Damas P, on behalf of the Sepsis Occurrence in Acutely Ill Patients (SOAP) Investigators: Are blood transfusions associated with greater mortality rates? Results of the Sepsis Occurrence in Acutely Ill Patients Study. *ANESTHESIOLOGY* 2008; 108:31–9

3. Connors AF Jr, Speroff T, Dawson NV, Thomas C, Harrell FE Jr, Wagner D, Desbiens N, Goldman L, Wu AW, Califf RM, Fulkerson WJ Jr, Vidaillet H, Broste S, Bellamy P, Lynn J, Knaus WA: The effectiveness of right heart catheterization in the initial care of critically ill patients. SUPPORT Investigators. *JAMA* 1996; 276:889–97

4. Karkouti K, Beattie WS, Dattilo KM, McCluskey SA, Ghannam M, Hamdy A, Wijesundera DN, Fedorko L, Yau TM: A propensity score case-control comparison of aprotinin and tranexamic acid in high-transfusion-risk cardiac surgery. *Transfusion* 2006; 46:327–38

5. Mangano DT, Tudor IC, Dietzel C: The risk associated with aprotinin in cardiac surgery. *N Engl J Med* 2006; 354:353–65

6. Hébert PC, Wells G, Blajchman MA, Marshall J, Martin C, Pagliarello G, Tweeddale M, Schweitzer I, Yetisir E: A multicenter, randomized, controlled clinical trial of transfusion requirements in critical care. Transfusion Requirements in Critical Care Investigators, Canadian Critical Care Trials Group. *N Engl J Med* 1999; 340:409–17

7. Pitt B, Zannad F, Remme WJ, Cody R, Castaigne A, Perez A, Palensky J, Wittes J: The effect of spironolactone on morbidity and mortality in patients with severe heart failure. Randomized Aldactone Evaluation Study Investigators. *N Engl J Med* 1999; 341:709–17

8. Juurlink DN, Mamdani MM, Lee DS, Kopp A, Austin PC, Laupacis A, Redelmeier DA: Rates of hyperkalemia after publication of the Randomized Aldactone Evaluation Study. *N Engl J Med* 2004; 351:543–51

9. Deans KJ, Minneci PC, Suffredini AF, Danner RL, Hoffman WD, Ciu X, Klein HG, Schechter AN, Banks SM, Eichacker PQ, Natanson C: Randomization in clinical trials of titrated therapies: Unintended consequences of using fixed treatment protocols. *Crit Care Med* 2007; 35:1509–16

10. Benson K, Hartz AJ: A comparison of observational studies and randomized, controlled trials. *N Engl J Med* 2000; 342:1878–86

11. Ioannidis JP: Contradicted and initially stronger effects in highly cited clinical research. *JAMA* 2005; 294:218–28

12. Poynard T, Munteanu M, Ratzu V, Benhamou Y, Di Martino V, Taieb J, Opolon P: Truth survival in clinical research: An evidence-based requiem? *Ann Intern Med* 2002; 136:888–95

13. Shrier I, Platt RW, Steele RJ: Mega-trials *versus* meta-analysis: Precision *versus* heterogeneity? *Contemporary Clin Trials* 2007; 28:324–8

14. Fergusson DA, Hébert PC, Mazer CD, Fries S, MacAdams C, Murkin JM, Teoh K, Duke PC, Arellano R, Blajchman MA, Bussières JS, Côté D, Karski J, Martineau R, Robblee JA, Rodger M, Wells G, Clinch J, Pretorius R, for the BART Investigators: A comparison of aprotinin and lysine analogues in high-risk cardiac surgery. *N Engl J Med* 2008; 358:2319–31

15. Hébert PC, Fergusson D, Blajchman MA, Wells GA, Kmetz A, Coyle D, Heddle N, Germain M, Goldman M, Töy B, Schweitzer I, van Walraven C, Devine D, Sher GD: Clinical outcomes following institution of the Canadian universal leukoreduction program for red blood cell transfusions. *JAMA* 2003; 289:1941–9

16. Fergusson D, Hébert PC, Lee SK, Walker CR, Barrington KJ, Joseph L, Blajchman MA, Shapiro S: Clinical outcomes following institution of universal leukoreduction of blood transfusions for premature infants. *JAMA* 2003; 289:1950–6

(Accepted for publication April 30, 2008.)

propensity scores. Readers of *ANESTHESIOLOGY* in general are not professional statisticians and may be reluctant to use propensity scores, even in appropriate situations, because of such a title.

The editorial is of value because it reviews an important problem of observational therapeutic studies. In such studies, investigators do not have control over who is or is not receiving the index treatment, which potentially results in imbalance of prognostic factors across the treatment arms. In the absence of randomization, treatment indication and assignment are typically related to the prognosis of the patient. For example, more patients with advanced disease may be given the index treatment than patients with early disease stages. As a consequence,

Supported by grant No. ZON-MW 917.46.360 from The Netherlands Organization for Scientific Research, Den Haag, The Netherlands.