Anesthesiology 2008; 109:744

Propensity Analysis: A Tool to Complement Randomized Studies

To the Editor:—Studies that use propensity analysis, like the study of Vincent *et al.*,[1] should not be perceived to be inferior to the gold standard of prospective randomized studies. Rather, propensity analysis and prospective randomized studies should be interpreted as complementary methods for finding the truth. Despite Nuttall and Houle's assertion that randomized controlled studies, unlike propensity analyses, do not have "the limitation that remaining unmeasured confounding variables may still be present,"[2] both measured and unmeasured confounding variables may still be present. Randomized studies rely on the assumption (or hope) that these variables will be equally distributed between the groups. Who the anesthesiologist is or who harvests the saphenous vein may have a profound effect on outcome after cardiac surgery,[3,4] but random studies involving cardiac surgery rarely stratify by these factors or even measure them. Even small differences between groups in measured variables in randomized trials may lead to erroneous statistically significant outcomes.[5]

Prospective randomized studies may be limited by the inability to randomize for important variables. In evaluating an intervention, such as activated protein C on mortality of intensive care unit patients, it is necessary that nonrandom but important factors, such as which intensive care unit treats the patient, be controlled. Typically, this is done with severity scores such as the Acute Physiology and Chronic Health Evaluation and Mortality Probability Model. Although the word *propensity* is not used to describe the Acute Physiology and Chronic Health Evaluation or Mortality Probability Model, these scores are the likelihood or the propensity that a patient will die, and these scores are then included (the same as a propensity score determining the likelihood of receiving a transfusion would be included in a study of blood transfusion and sepsis[1]) in the analysis to partially control for some of the confounders in the randomized controlled trial.

Another limitation of randomized controlled trials is their lack of generalizability. In determining the benefits or harm of transfusion, Hébert *et al.*[6] evaluated 6,451 persons to randomize 838 subjects (13%); 5,613 patients were excluded from their study. Physician belief in equipoise, the patient's or family's beliefs, or excluding patients based on age or comorbidities may produce nonrepresentative populations in randomized trials and severely limit the generalizability of the results.[7–9] In addition, crossover of subjects from one arm to the other arm of the trial or subject withdrawal may make the results hard to interpret.

Observational studies are not necessarily inferior to randomized studies. Both have advantages and disadvantages. Observational studies should be encouraged as a complement to randomized studies. They include a greater variety of patients, many of whom would be excluded by randomized studies, and can be performed for a small fraction of the cost. Sophisticated and innovative statistical techniques, such as multivariable analysis, propensity, and instrumental variables[10] should be used to help separate gold from fool's gold.

Milo Engoren, M.D., University of Toledo College of Medicine, St. Vincent Mercy Medical Center, Toledo, Ohio. engoren@pol.net

References

1. Vincent J-L, Sakr Y, Sprung C, Harboe S, Damas P: Are blood transfusions associated with greater mortality rates? Results of the Sepsis Occurrence in Acutely Ill Patients study. ANESTHESIOLOGY 2008; 108:31–9
2. Nuttall GA, Houle TT: Liars, damn liars, and propensity scores. ANESTHESIOLOGY 2008; 108:3–4
3. Slogoff S, Keats AS: Does perioperative myocardial ischemia lead to postoperative myocardial infarction? ANESTHESIOLOGY 1985; 65:107–14
4. Poston RS, Gu J, Brown JM, Gammie JS, White C, Nie L, Pierson RN III, Griffith BP: Endothelial injury and acquired aspirin resistance as promoters of regional thrombin formation and early vein graft failure after coronary artery bypass grafting. J Thorac Cardiovasc Surg 2006; 131:122–30
5. Lee JL, McNeer JF, Starmer CF, Harris PJ, Rosarti RA: Clinical judgment and statistics: Lessons from a simulated randomized trial in coronary artery disease. Circulation 1980; 61:508–15
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. Sprung CL, Eidelman LA: Ethical issues of clinical trials for the pulmonary artery catheter. New Horiz 1997; 5:264–7
8. Sen Biswas M, Newby LK, Bastian LA, Peterson ED, Sugarman J: Who refuses enrollment in cardiac clinical trials? Clin Trials 2007; 4:258–63
9. Britton A, McKee M, Black N, McPherson K, Sanderson C, Bain C: Threats to applicability of randomised trials: Exclusions and selective participation. J Health Serv Res Policy 1999; 4:112–21
10. Newhouse JP, McClellan M: Econometrics in outcomes research: The use of instrumental variables. Annu Rev Public Health 1998; 19:17–34

(Accepted for publication April 30, 2008.)

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.

Anesthesiology 2008; 109:744–5

Our European Study on Blood Transfusions: Three Quarters Full or One Quarter Empty?

To the Editor:—We appreciate the editorial[1] accompanying our article[2] and agree with the need to stress the limitations of propensity scores. A prospective randomized controlled trial (RCT), where possible, is always preferable to an observational study. However, RCTs have their own limitations, and prospective studies on blood transfusions based on hemoglobin thresholds are no exception. The exclusion of various diseases groups, such as patients with coronary artery disease, and the choice of treatment modality in the control group may challenge the applicability of the results of RCTs in everyday practice.[3] Indeed, Deans *et al.*[3] recently highlighted the presence of coronary artery disease as a confounding factor in the RCT of Hébert *et al.*[4] More specifically, a liberal blood transfusion strategy seemed to result in a higher mortality rate in younger patients with lower severity scores, but a lower mortality rate in the subgroup of patients with coronary artery disease.

Meticulous analyses, performed on large, unselected cohorts of critically ill patients, may provide useful additional information that can generate hypotheses and set the stage for subsequent RCTs. For

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 article[2] included a large number of variables, allowing extensive and reliable adjustment for confounders. Unfortunately, the accompanying editorial by Drs. Nuttall and Houle[1] 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 article[2]).

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

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

To the Editor:—The editorial by Nuttall and Houle[1] 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 Houle[1] 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 inappropriately)? 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.