Observing Blood Management Programs through the Retrospectroscope

Richard J. Cook, B.Sc., M.Math, Ph.D., Richard B. Weiskopf, M.D.

In this issue of Anesthesiology, Gupta et al.\(^1\) report on the results of a retrospective cohort study comparing transfusion practices and clinical outcomes before and after the implementation of a blood management program in orthopedic surgery at a single center. The motivation stems in part from interest in reducing the number of transfusions by the use of more restrictive hemoglobin triggers for red blood cell transfusions in orthopedic surgery patients. The authors observe that both lower utilization and comparable or improved patient outcomes followed implementation of the blood management program and conclude that a “hemoglobin threshold of 7 g/dl appears to be safe for many orthopedic patients.” There is a clear need to understand the relationship between transfusion triggers and outcomes to ensure that limited resources are used judiciously, to minimize exposures of patients, and to optimize patient outcomes.

The study by Gupta et al.\(^1\) offers a good illustration of the kinds of retrospective analyses often conducted based on data from large registries or administrative databases. There have been a number of studies in transfusion medicine which involved retrospective database analyses yielding findings that, when tested in prospective randomized trials, were not validated. For example, in a large retrospective analysis of 4,470 intensive care unit patients, Hébert et al.\(^2\) observed an association between lower hemoglobin concentrations and death. However, when they tested this hypothesis in a subsequent randomized trial in 838 similar intensive care unit patients, there was no evidence of differences in mortality between the liberally and restrictively transfused groups that were transfused to hemoglobin concentrations of 10.7 and 8.5 g/dl.\(^3\) In a retrospective analysis directed at the effect of red blood cell storage duration, Koch et al.\(^4\) suggested that cardiac surgery patients transfused with red cells stored for longer periods of time experienced a higher mortality rate than did patients transfused with red cells having shorter storage durations; a subsequent larger retrospective analysis of all transfusions in Denmark and Sweden by Edgren et al.\(^5\) showed different results. Exposure of patients with cardiovascular disease to red cells of particularly long storage duration was also associated with increased in-hospital mortality in a prospective registry analysis by Eikelboom et al.\(^6\) but these findings were not validated in a subsequent analysis based on an expanded dataset.\(^7\) Seven prospective randomized trials addressing this question did not substantiate the findings of Koch et al., finding no difference between red cells storage duration and mortality, change in multiple organ dysfunction scores, composite morbidity, pulmonary and immune function, lactate clearance, and reversal of anemia-induced neurocognitive function in a wide range of populations: cardiac surgery, critically ill adults, children with severe anemia, low-weight premature infants, all hospitalized patients, and healthy volunteers.\(^8\)–\(^14\)

The current publication affords an opportunity to discuss some challenges arising in retrospective analyses, which are highlighted below. The themes include the post hoc definition exposure variables and the interpretation of their effects, the challenge of dealing completely and rigorously with the effect of confounding variables, incomplete data, and the use of composite outcomes. These, and other issues, are important to bear in mind when trying to explain conflicting findings between publications on different database analyses, and the results of randomized trials.

**Post Hoc Definition of Exposure**

Although a central theme of this work is examination of the effect of a new blood management program on red blood cell utilization and comparable or improved patient outcomes, it is important to note that this effect was not observed in all patient populations. For example, in a study comparing transfusion exposure variables and the interpretation of their effects, the challenge of dealing completely and rigorously with the effect of confounding variables, incomplete data, and the use of composite outcomes. These, and other issues, are important to bear in mind when trying to explain conflicting findings between publications on different database analyses, and the results of randomized trials.

Large databases are appealing to exploit for investigating focused scientific questions, but the data necessary for a rigorous analysis are often lacking ...
use and outcome, some statements made by the authors suggest a causal effect of hemoglobin threshold on clinical outcome. The *post hoc* definition of hemoglobin threshold used here is “the lowest (nadir) hemoglobin concentration during the hospital stay.” This crude summary of exposure over the course of a hospitalization may be reasonable for descriptive purposes, but there is danger in overinterpreting the relation between such an exposure summary and its association with the composite morbidity and mortality outcome. The principle reason is that this minimum hemoglobin concentration was simply an observed value over a period of time, with an unknown temporal relationship to any morbid event, rather than an actual predefined threshold as would be specified in a prospective randomized trial. The statement “[t]o our knowledge, ours is the first study in orthopedics to assess hemoglobin thresholds as low as 7 g/dl” and the concluding statement in the abstract are therefore inappropriate.

**Ecological Fallacy and Confounding**

Large databases are appealing to exploit for investigating focused scientific questions, but the data necessary for a rigorous analysis are often lacking; the lack of preoperative hemoglobin values is one such example in this article, but there is a myriad of factors influencing patient care, some of which are dynamic and responsive to early treatment. When a relatively small number of factors are adjusted for such as age, sex, hip fracture status, surgical procedure, and a case-mix index are used, concerns arise about whether the data are sufficient and the adjustments are adequate. A prospective observational study would have enabled collection of a more comprehensive set of variables enabling a more complete adjustment for potential confounders and findings more consistent with the prospective interpretation of real clinical interest. The rationale for many interventions is not generally available in databases; information on the reason for the transfusion would also be useful for causal analysis and would be easily collected in a prospective study. We also note that the propensity score used by the authors appears to include the same factors adjusted for in the multivariate regression analysis, so it does not represent a casual sensitivity analysis in the usual sense. Some principle advantages of propensity score analyses include the ability to adjust for a larger number of confounders while maintaining a relatively simply model for the outcome. This can be achieved by matching, stratification, or regression on the propensity score, or using inverse probability of exposure weights. The latter would yield estimates of the effect of the blood management program on the response more in line with what would be estimated in a randomized clinical trial.

The selection criteria for potentially confounding variables to adjust for raises challenging issues. Confounding variables have an association with the outcome and the exposure variable, and although this is a relatively simple concept when dealing with cross-sectional studies, when exposure variables change over time in complex feedback systems, identifying, selecting, and modeling the effect of confounding variables is a daunting challenge. Moreover, selection of variables to adjust for in causal analyses should be based on scientific context rather than statistical significance. We refer readers to a timely, sobering, and stimulating recent paper by Hernán, from which one can learn to calibrate expectations and interpretations from registry-based studies.

**Composite Outcomes**

There is often compelling practical rationale for use of composite outcomes but these also introduce substantial challenges in interpretation of findings. This challenge is particularly important when the components of a composite outcome are of unequal importance, when they represent quite different clinical outcomes, and when the relative weighting of the components is unclear in the final analysis. The magnitude of the effect reported on the composite outcome in Gupta *et al.* is striking, but for reasons stated earlier caution is warranted before attributing this large effect to a lower hemoglobin threshold. With such a large effect, however, it should be feasible to carry out a randomized clinical trial confirming this finding in the elderly population of orthopedic surgery patients. Some may consider that this had already been investigated in the prospective, randomized Transfusion Trigger Trial for Functional Outcomes in Cardiovascular Patients Undergoing Surgical Hip Fracture Repair (FOCUS) trial, in which patients age 65 yr or more with a history of, or risk factors for, cardiovascular disease and undergoing surgical repair of hip fracture were randomly allocated to a liberal or restrictive transfusion protocol. There were no differences in functional recovery or mortality found in this trial. Even such a randomized, clinical trial can have limitations, however; there was a highly significant increase in the use of “rescue” transfusion for cardiovascular symptoms (*i.e.*, red blood cell transfusion at a trigger greater than that specified by the protocol) in the restrictive group compared with the liberal group. Moreover, the population in the FOCUS trial was quite different than that analyzed by Gupta *et al.* We note that the numbers of patients experiencing even the composite outcome are quite low. To gain better insight into the nature of any effects, larger samples would be useful so that the effects could be explored in the component outcomes of the composite outcome in a meaningful way. Finally, we note that the FOCUS trial is yet another example of a randomized trial yielding different results from the database analyses from which it was spawned.

In many circumstances it is not possible to conduct randomized clinical trials, and other types of data are the best that can be obtained. Randomized trials are possible to test the hypothesis that less frequent red cell transfusion does not increase risk, and the results have been mixed, attesting to the challenges in conducting experimental research in complex settings. Comroe in his book, *Retrospectroscope*, described the origins of some great discoveries in medicine. He did not envision his imaginary instrument being used for retrospective database examinations; in such settings the instrument’s
lens can indeed be quite clouded. However, results generated from retrospective database analyses can be thought provoking, hypothesis generating, and help set the agenda for future investigations, as do the findings of Gupta et al.

Competing Interests
Dr. Cook consults for the following entities that have an interest in red cell transfusion: Canadian Blood Services (Ottawa, Canada), U.S. Food and Drug Administration (Silver Spring, Maryland), TerumoBCT (Lakewood, Colorado), HbO2 Therapeutics (Souderton, Pennsylvania). Dr. Weiskopf consults for the following entities that have an interest in red cell transfusion: U.S. Department of Defense (Arlington, Virginia), U.S. Food and Drug Administration, TerumoBCT, HbO2 Therapeutics.

Correspondence
Address correspondence to Dr. Weiskopf: rbw@theweiskopf-group.com

References
9. Heddle NM, Cook RJ: Composite outcomes in clinical trials: What are they and when should they be used? Transfusion 2011; 51:11–3
19. Heddle NM, Cook RJ: Composite outcomes in clinical trials: What are they and when should they be used? Transfusion 2011; 51:11–3