Evidence Based Reviews in Surgery
a joint program of
Canadian Association of General Surgeons
L’Association Canadienne des Chirurgiens Généraux
&
American College of Surgeons

Package #6

Articles for Review:


Please read the above articles and be prepared to discuss the following:

1. What is the clinical question being addressed?
2. What is the study design?
3. What is the source of data
4. Are the data accurate and valid?
5. Is the matching appropriate and are the 2 groups similar?
6. Identify potential biases that might account for differences in the conclusions?
7. What outcomes were assessed and are they clinically relevant and sensitive?
8. Are the differences clinically significant?
9. Are further studies required?
10. State the conclusion. Have the authors addressed the clinical question posed?
11. Do the data support the conclusions?
Operative Blood Loss, Blood Transfusion, and 30-Day Mortality in Older Patients After Major Noncardiac Surgery

Wen-Chih Wu, MD,*†‡ Tracy S. Smith, MS,§ William G. Henderson, PhD,¶ Charles B. Eaton, MD, MS,** Roy M. Poses, MD,† Georgette Uttley, RN,** Vincent Mor, PhD,‡ Satish C. Sharma, MD,†‡ Michael Vezieridis, MD,‡ Shukri F. Khuri, MD,‡‡ and Peter D. Friedman, MD, MPH*†

Objective: Anemia and operative blood loss are common in the elderly, but evidence is lacking on whether intraoperative blood transfusions can reduce the risk of postoperative death.

Methods: We analyzed retrospective data from 239,286 patients 65 years of age or older who underwent major noncardiac surgery in 1997 to 2004 at veteran hospitals nationwide. Propensity-score matching was used to adjust for differences between patients who received intraoperative blood transfusions (9.4%) and those who did not, and data were used to determine the association between intraoperative blood transfusion and 30-day postoperative mortality.

Results: After propensity-score matching, intraoperative blood transfusion was associated with mortality risk reductions in patients with preoperative hematocrit levels of <24% (odds ratio: 0.60, 95% CI: 0.41–0.87), and in patients with hematocrit of 30% or greater when there is substantial (500–999 mL) blood loss (odds ratio: 0.35, 95% CI: 0.22–0.56 for hematocrit levels between 30%–35.9% and 0.78, 95% CI: 0.62–0.97 for hematocrit levels of 36% or greater). When operative blood loss was <500 mL, transfusion was not associated with mortality reductions for patients with hematocrit levels of 30% or greater when there is substantial blood loss.

Conclusions: Intraoperative blood transfusion is associated with a lower 30-day postoperative mortality among elderly patients undergoing major noncardiac surgery if there is substantial operative blood loss or low preoperative hematocrit levels (<24%). Transfusion is associated with increased mortality risks for those with preoperative hematocrit levels between 30% and 35.9% and <500 mL of blood loss.

Both anemia and blood loss are common risk factors for mortality in patients undergoing surgery; yet, the benefits and the hematocrit trigger at which intraoperative blood transfusion should be given to treat anemia and blood loss are still unknown. While rigorous investigations have examined the optimal hematocrit range at which transfusion is beneficial in intensive care and coronary care unit settings, similar studies in surgery are lacking. Available reports are limited by small samples and/or single surgical specialties and have not documented efficacy of surgical blood transfusion. Indeed, various small sample studies reported intraoperative blood use to be either ineffective to improve postoperative outcomes or found it to be detrimental. Current recommendations on perioperative blood transfusion are mostly based on expert consensus, and not controlled studies.

Elderly patients undergoing surgery are particularly vulnerable to blood loss and anemia’s adverse effects, since they have limited physiological reserve and a high prevalence of underlying coronary disease. As a result, patients greater than 65 years of age receive nearly half of all blood transfusions, while evidence is lacking in its use in the operating room. Limited evidence complicates the transfusion decision, and is 1 source of the great variability in transfusion practices in the operating room with large impact in cost and quality of care. To this effect, the current study analyzes a nationwide cohort of war veteran patients, older than age 65 years, who underwent major noncardiac surgery to determine the effect of intraoperative blood transfusion on 30-day postoperative mortality.

METHODS

The institutional review boards at the Providence VA Medical Center, VA Boston Healthcare System and the University of Colorado Multiple (COMIRB) sites approved the study. We analyzed retrospective cohort data from the Department of Veterans Affairs (VA) National Surgical Quality Improvement Program (NSQIP), which has been described in detail. Briefly, NSQIP is a VA-wide initiative to improve surgical care through prospective collection of data on patient risk for major surgery and report of risk-adjusted postoperative outcomes. It includes the major procedures (defined as performed in the operating room that required general, spinal, or epidural anesthesia) in 8 surgical subspecialties (general surgery, vascular surgery, orthopedics, urology, noncardiac thoracic surgery, neurosurgery, otolaryngology, and plastic surgery) from 132 VA hospitals since 1991. Excluded were repeated operations on patients entered into the dataset in the previous 30 days, and surgical series.
procedures with very low morbidity and mortality, such as ophthalmology, auditory, or nasal procedures. Nurse reviewers in each participating center extracted the data, with periodic reabstraction of the data by external reviewers showing very good agreement. The NSQIP dataset can be merged with other VA datasets to obtain additional information. In this study, the NSQIP dataset was merged with the VA Patient Treatment Files to obtain records on the use of autologous transfusions and prior hospitalizations related to coronary artery disease, and to the Outpatient Care files to identify patients with cell saver procedure within 30 days prior to surgery and those with stable coronary artery disease treated in the ambulatory setting.

Analytic Sample

The analytic cohort was war veteran patients aged 65 years or older with a documented preoperative hematocrit value who underwent major noncardiac surgery during years 1997 to 2004 (310,311 cases). Random selection restricted those patients with multiple operations during the study period to 1 case per patient (70,506 cases excluded). We excluded 12 patients with missing transfusion information and 507 patients with preoperative hematocrit levels of 54% or greater, for a final analytic sample of 239,286 patients.

Preoperative Hematocrit Values

Preoperative hematocrit was defined as the last hematocrit measurement prior to the index operation. The median duration between the last hematocrit measurement and the index operation is 3 days with an interquartile range of 1 to 11 days. Based on previous work, 10 preoperative hematocrit values were stratified a priori into 4 categories: <24%, 24% to 29.9%, 30% to 35.9%, and 36% or greater, to detect the pretransfusion hematocrit threshold at which intraoperative blood use might be associated with postoperative outcomes.

Intraoperative Blood Transfusion

Patients were classified as having received intraoperative blood transfusions if they received whole blood or packed red blood cells during their procedure. Although this definition included blood transfused to the patient collected from a cell saver, only 113 patients had an intraoperative autologous transfusion (2 patients within the 24%–29.9% preoperative hematocrit category, 26 patients within the 30%–35.9% hematocrit category and 85 patients within the 36%–53.9% hematocrit category).

Estimated Intraoperative Blood Loss

Intraoperative blood loss was calculated based on a modification of the formula of allowable blood loss by Drs. Gross 22 and Hahn 23 adapted to a population of predominantly male, elderly veterans 65 years or older, with variable perioperative blood volume shifts (as opposed to a constant proposed by the equations), taking into account intraoperative blood transfusions. To derive an equation suitable to our study sample, we used blood loss data (per anesthesiologist and/or surgeon) that was collected from the surgical package records of 71,232 veteran patients 65 years or older who underwent similar major noncardiac surgeries during October 1991 to August 1995 (data no longer collected in the NSQIP thereafter). Based on this earlier blood loss data, we were able to derive an equation to estimate intraoperative blood loss based on preoperative and postoperative hematocrit levels, while accounting for packed red blood cells transfused during surgery:

\[
\text{Intraoperative blood loss (in mL)} = 31.265 \times \text{preoperative hematocrit} - 29.83 \\
\times \text{postoperative hematocrit} + 269.67 \\
\times \text{units of red blood cells transfused intraoperatively}
\]

This equation showed a very good prediction of the recorded blood loss (\(r^2 = 0.62\)).

Postoperative Outcome

Chart abstraction indicated death and postoperative complications within 30 days of the index surgery, with complete 30-day postoperative follow-up.

Statistical Analysis

We compared demographic and clinical characteristics of patients who received intraoperative blood transfusion and those who did not to identify potential confounders. \(x^2\) analyses for categorical variable and t-tests for continuous variables tested differences between groups.

Propensity analysis was used to match patients for comparison of 30-day postoperative mortality rates between transfused and those not transfused intraoperatively. 24 A nonparsimonious logistic regression model estimated a propensity score for the presence of intraoperative blood transfusions. The model included the patient’s preoperative demographic and clinical characteristics, hospital of the surgery, postgraduate year and surgical subspecialty of the surgeon, complexity of the surgery as determined by relative value units, and the 4 preoperative hematocrit categories. An indicator variable denoted whether a complication occurred on the day of surgery or first postoperative day to account for a complicated procedure. The resulting logistic regression computed each patient’s probability of intraoperative blood transfusion. Patients who received intraoperative blood transfusion were matched 1:1 to those who did not, based on the propensity for transfusion and the year of surgery. 26 We then examined the matched cohort for balance of baseline characteristics between the transfused and the nontransfused groups. Conditional logistic regression determined the effect of intraoperative blood transfusion on 30-day postoperative mortality risk, adjusting for residual differences. Since the effects of intraoperative blood transfusion might vary by the severity of anemia and bleeding, the association between blood transfusion and 30-day postoperative mortality was examined in the 4 preoperative hematocrit subgroups in the propensity-matched cohort, stratified by estimated intraoperative blood loss (less than 500 mL, 500–999 mL, and 1000 mL or greater). 3

RESULTS

The analytic sample included 239,286 patients; 9.4% received at least 1 unit of red blood cells during surgery. For both the transfused and the nontransfused patients, 98% were male, with an average age of 73 years (Table 1). The rate of intraoperative blood transfusion decreased as preoperative hematocrit values increased, from 54.5% among patients with a preoperative hematocrit level of <24% to 5.8% in patients with a preoperative hematocrit value of 36% or greater. Transfused patients tended to have lower preoperative hematocrit values, were less likely to be white, and had a higher prevalence of severe systemic disease (American Society of Anesthesiologists class 4 or 5), cardiac disease (eg, congestive heart failure, myocardial infarction), neurologic disorder (eg, cerebrovascular disease, coma), pulmonary disorder (eg, obstructive pulmonary disease, ventilator dependence), hematologic problems (eg, bleeding, postoperative blood use greater than 4 units), ascites and renal failure. Patients who received intraoperative blood transfusions were also more likely to have more complex and lengthy surgical procedures, to undergo emergency surgery, vascular surgery and general anesthesia, and have morbidity within the first postoperative day (Table 1; and Expanded Table 1, Supplemental Digital Content 1, online only, Available at: http://links.lww.com/SLA/A49). They were also more likely to have abnormal preoperative laboratory values, transferred from another hospital or nursing home and
TABLE 1. Characteristics of All Patients in the Study Who Received Intraoperative Blood Transfusions and Patients Who Did Not

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Operative Transfusion (n = 22,515)</th>
<th>Nontransfused (n = 216,771)</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (yr, ±SD)</td>
<td>73.8 ± 5.9</td>
<td>73.5 ± 5.8</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Male gender (%)</td>
<td>98.4</td>
<td>98.1</td>
<td>0.005</td>
</tr>
<tr>
<td>White race (%)</td>
<td>76.2</td>
<td>80.5</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Preoperative hematocrit &lt;24.0% (%)</td>
<td>3.3</td>
<td>0.3</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Preoperative hematocrit 24%–29.9% (%)</td>
<td>20.1</td>
<td>4.7</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Preoperative hematocrit 30%–35.9% (%)</td>
<td>30.6</td>
<td>17.9</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Preoperative hematocrit &gt;35.9% (%)</td>
<td>46.0</td>
<td>77.1</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>American society of anesthesiologist class 4 (%)</td>
<td>27.6</td>
<td>12.5</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>American society of anesthesiologist class 5 (%)</td>
<td>2.1</td>
<td>0.3</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Congestive heart failure within 30 d (%)</td>
<td>6.2</td>
<td>3.6</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Angina admission within 30 d (%)</td>
<td>28.1</td>
<td>18.9</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Myocardial Infarction within 6 mo (%)</td>
<td>2.4</td>
<td>1.0</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Diabetes treated with oral agents or insulin (%)</td>
<td>22.9</td>
<td>21.3</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Independent functional status (%)</td>
<td>77.2</td>
<td>85.1</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Coma (%)</td>
<td>0.3</td>
<td>0.1</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>CVA with neurological deficit (%)</td>
<td>8.8</td>
<td>7.8</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>History of transient ischemic attack</td>
<td>4.7</td>
<td>5.2</td>
<td>0.005</td>
</tr>
<tr>
<td>Dyspnea at rest or with moderate exertion (%)</td>
<td>25.7</td>
<td>19.4</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Current smoker (%)</td>
<td>27.9</td>
<td>22.9</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>History of COPD (%)</td>
<td>23.6</td>
<td>18.9</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Current pneumonia (%)</td>
<td>2.8</td>
<td>1.2</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Ventilator dependent greater than 48 h (%)</td>
<td>2.7</td>
<td>0.6</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Alcohol greater than 2 drinks per day (%)</td>
<td>8.7</td>
<td>7.2</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Asceptes (%)</td>
<td>1.7</td>
<td>0.6</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Acute renal failure (%)</td>
<td>2.1</td>
<td>0.8</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>On dialysis (%)</td>
<td>1.8</td>
<td>1.1</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Bleeding disorder (%)</td>
<td>6.1</td>
<td>2.4</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Disseminated cancer (%)</td>
<td>5.0</td>
<td>2.5</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Preoperative blood transfusion &gt;4 units (%)</td>
<td>5.1</td>
<td>0.6</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

*Emergency surgery is considered to exist when the anesthesiologist and the attending surgeon document a case to be “emergent” in the anesthesia record and the operative report. Surgical morbidity included cardiac arrest, myocardial infarction, stroke, coma, graft or prosthetic failure, sepsis, pneumonia, unplanned intubation or failure to wean from ventilator, pulmonary embolism or deep venous thrombosis, urinary tract infection, renal failure, wound infection, and bleeding.

COPD indicates chronic obstructive pulmonary disease; CVA, cerebrovascular accident.

Blood loss and transfusion in surgery

© 2010 Lippincott Williams & Wilkins

DISCUSSION

Although preoperative anemia and operative blood loss are both associated with increased morbidity and mortality in elderly patients undergoing major noncardiac surgery,1–3 there is limited evidence on treatment of perioperative anemia and blood loss with intraoperative blood transfusions. Our findings suggest that the association between intraoperative blood transfusion and 30-day operated on in hospitals with higher average surgical case loads per year (Expanded Table 1, online only).

In generating the propensity score for intraoperative blood transfusion, the logistic regression model had excellent predictive discrimination, with a c-index of 0.883. Based on the propensity score, 83% of transfused patients (n = 18,646) were matched with nontransfused controls (n = 18,646). Transfused patients received an average of 2.6 units of red blood cells intraoperatively. Examination of this matched cohort (Table 2; and Expanded Table 2, Supplemental Digital Content 2, online only, available at: http://links.lww.com/SLA/A50) showed successful matching, with balance of almost all the baseline characteristics between the transfused and nontransfused patients. Only 3 of 58 variables had clinically and statistically significant imbalance between the transfused and nontransfused matched-samples: mean operative time (3.8 vs. 2.8 hours), proportion of American Society of Anesthesiologist class 5 patients (1.5% vs. 1.0%), and rate of general anesthesia use (89.3% vs. 80%).

Analysis of the propensity-matched cohort showed 30-day postoperative mortality rate to be higher in transfused (10.5%) as compared with nontransfused patients (8.6%). After adjusting for mean operative time, ASA classification, and the rate of general anesthesia with conditional logistic regression, transfused patients were at increased risk for 30-day postoperative mortality compared with nontransfused counterparts (adjusted odds ratio: 1.37, 95% CI: 1.27–1.48). However, subgroup analyses of the matched cohort found that risk to vary according to preoperative hematocrit levels and the amount of operative blood loss (Fig. 1). Intraoperative blood transfusion conferred a reduced 30-day postoperative mortality risk in patients with a preoperative hematocrit value less than 24% (odds ratio: 0.60, 95% CI: 0.41–0.87) overall; and no significant change in mortality risks in patients with a preoperative hematocrit value between 24% and 29.9% (odds ratio: 1.04, 95% CI: 0.91–1.20), although all but 6 nontransfused patients in this hematocrit category had an operative blood loss of less than 500 mL. Among patients with preoperative hematocrit values of 30% and more, intraoperative blood transfusion was associated with reduced 30-day postoperative mortality risks when there was substantial blood loss of 500 to 999 mL (odds ratio: 0.35, 95% CI: 0.22–0.56 for hematocrit levels between 30%–35.9%; and 0.78, 95% CI: 0.62–0.97 for hematocrit levels of 36% or greater), with no change in risk or possibility of harm when blood loss is less than 500 mL (odds ratio: 1.29, 95% CI: 1.04–1.60 for hematocrit levels between 30% and 35.9%; and 1.40, 95% CI: 0.68–2.88 for hematocrit levels of 36% or greater). There were 6393 (18%) patients with an operative blood loss of 1000 mL or greater, of which only 6 patients were not transfused intraoperatively, and all of them had preoperative hematocrit levels of 30% and greater. Mortality rates were 10.2% for the transfused patients and 16.7% for the nontransfused (P = nonsignificant).

Analyses of late postoperative complications (after day 1) among the deceased patients showed an overall higher incidence of postoperative respiratory complications (failure to wean from ventilator, unplanned intubation) and sepsis, and to a smaller degree, postoperative renal failure, and bleeding among those who received intraoperative blood transfusions compared with those who were not transfused (Table 3).

© 2010 Lippincott Williams & Wilkins | www.annalsofsurgery.com | 13

Copyright © Lippincott Williams & Wilkins. Unauthorized reproduction of this article is prohibited.
postoperative mortality depends substantially on preoperative hematocrit levels and intraoperative blood loss. Intraoperative blood transfusion conferred a reduced 30-day postoperative mortality risk in patients with a preoperative hematocrit value less than 24%, and no significant change in mortality risks when preoperative hematocrit levels are between 24% and 29.9%. At higher preoperative hematocrit levels of 30% or greater, intraoperative blood transfusion conferred reduced 30-day postoperative mortality risks when there is substantial blood loss of 500 to 999 mL, and no change in risk or potential for harm when blood loss is less than 500 mL.

The relationship between perioperative anemia and postoperative adverse events has been described in several reports. Wu et al found a 1.6% increase in 30-day postoperative risk of death for every preoperative hematocrit point below 39% in a cohort of elderly veteran patients 65 years or older undergoing major noncardiac surgery.1 In patients undergoing vascular surgery or radical prostatectomy, postoperative hematocrit levels below 28% were found to be associated with intraoperative and postoperative cardiac morbidity and ischemic events.2,28 Carson et al found both preoperative and postoperative anemia to be related to postoperative death, especially in patients with cardiovascular diseases.2 In addition, perioperative anemia can be worsened by operative blood loss. Studies have shown an interaction between hemoglobin levels and surgical blood loss on postoperative outcomes.2 Both Carson and Spence and their respective coauthors found an increased hospital mortality for those patients whose operative blood loss exceeds 500 mL.3,29 These studies suggest that both perioperative anemia and operative blood loss are risk factors for adverse postoperative outcomes. Consistent with this rationale, the current study found that the use of intraoperative blood transfusion in the treatment of anemia was related to reduced mortality in older patients with preoperative hematocrit levels of less than 24% and those with higher hematocrit values who experienced substantial intraoperative blood losses (500–999 mL).

However, blood transfusion has its risks, including transmission of blood-borne pathogens,10 hemolytic reactions, acute lung injury, volume overload, and immunosuppression, among others.31 Thus, in patients with lesser degrees of anemia or blood loss, the potential adverse effects of transfusion might balance against, and in some cases, outweigh its benefits. The current study found no significant mortality benefits of intraoperative blood transfusion for patients with operative blood loss of <500 mL when their preoperative hematocrit levels were 24% or greater, and a potential for harm among patients with hematocrit levels of 30% to 35.9%. Comparable to our results, elderly patients undergoing hip surgery with hemoglobin values of 8 to 9 g/dL (approximately equal to hematocrit levels of 24%–27% in our study) also appear to experience no mortality benefit from blood transfusions.1,24 In patients undergoing vascular surgery or radical prostatectomy, postoperative hematocrit levels below 28% were found to be associated with intraoperative and postoperative cardiac morbidity and ischemic events.2,28 Carson et al found both preoperative and postoperative anemia to be related to postoperative death, especially in patients with cardiovascular diseases.2 In addition, perioperative anemia can be worsened by operative blood loss. Studies have shown an interaction between hemoglobin levels and surgical blood loss on postoperative outcomes.2 Both Carson and Spence and their respective coauthors found an increased hospital mortality for those patients whose operative blood loss exceeds 500 mL.3,29 These studies suggest that both perioperative anemia and operative blood loss are risk factors for adverse postoperative outcomes. Consistent with this rationale, the current study found that the use of intraoperative blood transfusion in the treatment of anemia was related to reduced mortality in older patients with preoperative hematocrit levels of less than 24% and those with higher hematocrit values who experienced substantial intraoperative blood losses (500–999 mL).

### Table 2. Characteristics of the Patients in the Propensity-Matched Cohort Who Received Operative Blood Transfusions and Those Who Did Not

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Operative Transfusion (n = 18,646)</th>
<th>Nontransfused (n = 18,646)</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (yr, ±SD)</td>
<td>73.8 ± 5.9</td>
<td>73.8 ± 5.9</td>
<td>0.889</td>
</tr>
<tr>
<td>Male gender (%)</td>
<td>98.3</td>
<td>98.4</td>
<td>0.422</td>
</tr>
<tr>
<td>White race (%)</td>
<td>76.7</td>
<td>77.0</td>
<td>0.818</td>
</tr>
<tr>
<td>Preoperative hematocrit &lt;24.0% (%)</td>
<td>2.1</td>
<td>1.8</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Preoperative hematocrit 24%–29.9% (%)</td>
<td>17.3</td>
<td>18.1</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Preoperative hematocrit 30%–35.9% (%)</td>
<td>30.8</td>
<td>32.3</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Preoperative hematocrit 36%–53.9% (%)</td>
<td>49.8</td>
<td>47.9</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>American society of anesthesiologist class 4 (%)</td>
<td>24.8</td>
<td>25.2</td>
<td>0.003</td>
</tr>
<tr>
<td>American society of anesthesiologist class 5 (%)</td>
<td>1.5</td>
<td>1.0</td>
<td>0.003</td>
</tr>
<tr>
<td>Congestive heart failure &lt;30 d (%)</td>
<td>5.9</td>
<td>6.2</td>
<td>0.216</td>
</tr>
<tr>
<td>Angina admission within 30 d (%)</td>
<td>27.3</td>
<td>28.2</td>
<td>0.051</td>
</tr>
<tr>
<td>Myocardial infarction within 6 mo (%)</td>
<td>2.2</td>
<td>1.8</td>
<td>0.009</td>
</tr>
<tr>
<td>Diabetes treated with oral agents or insulin (%)</td>
<td>23.3</td>
<td>23.4</td>
<td>0.509</td>
</tr>
<tr>
<td>Independent functional status (%)</td>
<td>77.8</td>
<td>77.1</td>
<td>0.257</td>
</tr>
<tr>
<td>Coma (%)</td>
<td>0.3</td>
<td>0.3</td>
<td>0.843</td>
</tr>
<tr>
<td>CVA with neurological deficit (%)</td>
<td>8.9</td>
<td>8.9</td>
<td>0.873</td>
</tr>
<tr>
<td>History of transient ischemic attack (%)</td>
<td>5.0</td>
<td>4.8</td>
<td>0.518</td>
</tr>
<tr>
<td>Dyspnea at rest or with moderate exertion (%)</td>
<td>24.7</td>
<td>25.0</td>
<td>0.729</td>
</tr>
<tr>
<td>Current smoker (%)</td>
<td>26.9</td>
<td>27.0</td>
<td>0.846</td>
</tr>
<tr>
<td>History of COPD (%)</td>
<td>22.7</td>
<td>23.0</td>
<td>0.488</td>
</tr>
<tr>
<td>Current pneumonia (%)</td>
<td>2.5</td>
<td>2.8</td>
<td>0.156</td>
</tr>
<tr>
<td>Ventilator dependent greater than 48 h (%)</td>
<td>2.3</td>
<td>2.2</td>
<td>0.528</td>
</tr>
<tr>
<td>Alcohol greater than 2 drinks per day (%)</td>
<td>8.4</td>
<td>8.8</td>
<td>0.162</td>
</tr>
<tr>
<td>Ascites (%)</td>
<td>1.5</td>
<td>1.4</td>
<td>0.785</td>
</tr>
<tr>
<td>Acute renal failure (%)</td>
<td>1.9</td>
<td>1.9</td>
<td>0.888</td>
</tr>
<tr>
<td>On dialysis (%)</td>
<td>1.8</td>
<td>1.9</td>
<td>0.358</td>
</tr>
<tr>
<td>Bleeding disorder (%)</td>
<td>5.2</td>
<td>5.2</td>
<td>0.817</td>
</tr>
<tr>
<td>Disseminated cancer (%)</td>
<td>4.9</td>
<td>5.0</td>
<td>0.537</td>
</tr>
<tr>
<td>Preoperative blood transfusion &gt;4 units (%)</td>
<td>3.3</td>
<td>3.2</td>
<td>0.501</td>
</tr>
<tr>
<td>Emergency surgery (%)</td>
<td>14.0</td>
<td>13.8</td>
<td>0.611</td>
</tr>
<tr>
<td>Mean operative time in h (±SD)</td>
<td>3.8 (±2.2)</td>
<td>2.8 (±1.9)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>General anesthesia (%)</td>
<td>89.3</td>
<td>80.0</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Units of red blood cells transfused in the operating room (±SD)</td>
<td>2.6 (±2.5)</td>
<td>0 (±0)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Surgical morbidity that occurred within postoperative d 1 (%)</td>
<td>5.6</td>
<td>5.6</td>
<td>0.857</td>
</tr>
</tbody>
</table>

*Emergency surgery is considered to exist when the anesthesiologist and the attending surgeon document a case to be “emergent” in the anesthesiologist record and the operative report. Surgical morbidity included cardiac arrest, myocardial infarction, stroke, coma, graft or prosthetic failure, sepsis, pneumonia, unplanned intubation or failure to wean from ventilator, pulmonary embolism or deep venous thrombosis, urinary tract infection, renal failure, wound infection, and bleeding.

COPD indicates Chronic obstructive pulmonary disease; CVA, Cerebrovascular accident.
Further studies are needed to validate this observation.

The NIH Consensus Statement on Perioperative Red Cell Transfusion and the American Society of Anesthesiologists practice guidelines suggest that red blood cell transfusion is usually indicated in perioperative patients with hemoglobin values of 6 to 7 g/dL or less and is rarely indicated in patients with hemoglobin values greater than 10 g/dL, while leaving an uncertain gap between hemoglobin values of 7 to 10 g/dL. Additionally, there is no information on how blood loss may change the need for transfusion. Our findings inform future recommendations by showing mortality risk reductions associated with transfusion in patients with preoperative hematocrit levels of 24% (approximately a hemoglobin of 8 g/dL), or in patients with mild to no preoperative anemia (hematocrit of 30% or greater) when there is substantial blood loss (500–999 mL). They also confirm that intraoperative transfusion is not helpful for patients with hematocrit levels of 24% or greater, when the estimated blood loss was 500 mL. This is important since red blood cell transfusions also represent a substantial financial burden to healthcare, with each unit of blood from acquisition to transfusion costing approximately $1600 to $2400 US dollars.

This study has several limitations and strengths. First, the study sample is mostly male, so results may not generalize to the older female

![FIGURE 1. Association between intraoperative blood transfusion and 30-day postoperative mortality in the propensity matched dataset stratified by preoperative hematocrit and intraoperative blood loss. There were 6393 patients with an operative blood loss of 1000 mL or greater, of which only 6 patients were not transfused intraoperatively; hence data was not included in the figure. All of them had preoperative hematocrit levels of 30% and greater. There were 1949 patients (5%) of the entire propensity matched cohort not included in the subanalysis based on intraoperative blood loss due to lack of postoperative hematocrit data. HCT indicates hematocrit.]

<table>
<thead>
<tr>
<th>Operative Transfusion</th>
<th>Nontransfused</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cardiac arrest</td>
<td>19.3%</td>
<td>18.5%</td>
</tr>
<tr>
<td>Myocardial infarction</td>
<td>4.2%</td>
<td>3.4%</td>
</tr>
<tr>
<td>Coma</td>
<td>2.5%</td>
<td>3.0%</td>
</tr>
<tr>
<td>Stroke</td>
<td>2.0%</td>
<td>1.3%</td>
</tr>
<tr>
<td>Failure to wean from ventilator</td>
<td>26.9%</td>
<td>20.7%</td>
</tr>
<tr>
<td>Renal failure</td>
<td>7.5%</td>
<td>5.7%</td>
</tr>
<tr>
<td>Bleeding</td>
<td>3.3%</td>
<td>1.6%</td>
</tr>
<tr>
<td>Deep venous thrombosis</td>
<td>1.0%</td>
<td>0.8%</td>
</tr>
<tr>
<td>Sepsis</td>
<td>14.2%</td>
<td>11.0%</td>
</tr>
<tr>
<td>Pneumonia</td>
<td>19.2%</td>
<td>17.9%</td>
</tr>
<tr>
<td>Pulmonary embolism</td>
<td>1.1%</td>
<td>1.6%</td>
</tr>
<tr>
<td>Unplanned intubation</td>
<td>20.4%</td>
<td>16.4%</td>
</tr>
<tr>
<td>Deep wound infection</td>
<td>2.9%</td>
<td>2.0%</td>
</tr>
</tbody>
</table>

**TABLE 3.** Late Postoperative Complications (After Postoperative Day 1) Among Deceased Patients
population undergoing major noncardiac surgery. This is especially important given gender differences in preoperative hematocrit levels, blood volume and pattern of blood transfusion, showing women in general to have lower preoperative hematocrit levels and smaller blood volume, and receive more intraoperative blood transfusions than men.\textsuperscript{34–36} Gender differences in blood transfusion persisted even after controlling for age, weight (accounting for blood volume), duration of surgery, and preoperative hematocrit levels.\textsuperscript{36} Future studies are needed to address the blood transfusion needs of women during surgery and whether gender differences in transfusion pattern may relate to postoperative outcomes. Second, we do not have information on the chronicity of anemia, the age of the transfused blood, or information on whether the blood was leukodepleted. Third, blood transfusion during surgery was not randomly assigned and therefore, the results may be confounded by indications for blood transfusion. Although state-of-the-art propensity methods matched patients on a broad spectrum of clinical characteristics, residual confounding may still exist. Fourth, the blood loss variable is not directly captured, but only an estimate based on blood transfused during surgery and the difference between the preoperative and nadir postoperative hematocrit levels. Since hematocrit is a volume-based measure, volume shifts in the perioperative period, which can range from dehydration to hypervolemia due to different clinical scenarios (eg, preoperative fasting, intravenous fluid replacements, postoperative diuresis), are likely to change the precision of the estimates. In addition, the timing of postoperative hematocrit measurement (not standardized) might also limit the precision of the estimated blood loss equation, since surgical blood loss would be underestimated if any postoperative blood transfusion should occur before the postoperative hematocrit is drawn. Therefore, although this equation showed a very good prediction against the recorded blood loss at the population level, it should be interpreted with caution when applied to individual patients, since it was not validated against other methods of blood loss estimation such as suction losses, sponge weights, or others. Lastly, our study would not be applicable to specific situations in which cell saver and autologous blood transfusions are commonly used. Nonetheless, this study’s strengths are its large, nationwide sample, comprehensive and reliable data extraction, and breadth of procedures and patients with diverse coexisting morbidities.

Among elderly patients who underwent major noncardiac surgery, intraoperative blood transfusion was associated with mortality risk reduction in patients with preoperative hematocrit levels of <24\% or patients with mild to no preoperative anemia (hematocrit of 30\% or greater) when there is substantial blood loss (500–999 mL). Intraoperative transfusion is not helpful for patients with hematocrit levels of 24\% or greater when the estimated blood loss was <500 mL, and it may be harmful if their preoperative hematocrit levels were between 30\% to 35.9\%. Ideally, randomized-controlled trials would confirm these findings, but such trials would be difficult to design, and providers and patients may be unwilling to randomize or be randomized. Until such trial can be mounted, this study’s findings provide some of the best evidence to date to guide blood use in the operative setting.

REFERENCES

27. Hogue CW, Goodnough LT, Monk TG. Perioperative myocardial ischemic episodes are related to hematocrit level in patients undergoing radical prostatectomy. Transfusion. 1998;38:924–931.


Estimating Causal Effects from Large Data Sets Using Propensity Scores

Donald B. Rubin, PhD

The aim of many analyses of large databases is to draw causal inferences about the effects of actions, treatments, or interventions. Examples include the effects of various options available to a physician for treating a particular patient, the relative efficacies of various health care providers, and the consequences of implementing a new national health care policy. A complication of using large databases to achieve such aims is that their data are almost always observational rather than experimental. That is, the data in most large data sets are not based on the results of carefully conducted randomized clinical trials, but rather represent data collected through the observation of systems as they operate in normal practice without any interventions implemented by randomized assignment rules. Such data are relatively inexpensive to obtain, however, and often do represent the spectrum of medical practice better than the settings of randomized experiments. Consequently, it is sensible to try to estimate the effects of treatments from such large data sets, even if only to help design a new randomized experiment or shed light on the generalizability of results from existing randomized experiments. However, standard methods of analysis using available statistical software (such as linear or logistic regression) can be deceptive for these objectives because they provide no warnings about their propriety. Propensity score methods are more reliable tools for addressing such objectives because the assumptions needed to make their answers appropriate are more assessable and transparent to the investigator.


From Harvard University, Cambridge, Massachusetts. For the current author address, see end of text.

Many observational studies based on large databases attempt to estimate the causal effects of some new treatment or exposure relative to a control condition, such as the effect of smoking on mortality. In most such studies, it is necessary to control for naturally occurring systematic differences in background characteristics between the treatment group and the control group, such as age or sex distributions, that would not occur in the context of a randomized experiment. Typically, many background characteristics need to be controlled.

Propensity score technology, introduced by Rosenbaum and Rubin (1), addresses this situation by reducing the entire collection of background characteristics to a single composite characteristic that appropriately summarizes the collection. This reduction from many characteristics to one composite characteristic allows the straightforward assessment of whether the treatment and control groups overlap enough with respect to background characteristics to allow a sensible estimation of treatment versus control effects from the data set. Moreover, when such overlap is present, the propensity score approach allows a straightforward estimation of treatment versus control effects that reflects adjustment for differences in all observed background characteristics.

Subclassification on One Confounding Variable

Before describing the use of propensity scores in the statistical analysis of observational studies with many confounding background characteristics, I begin with an example showing how subclassification adjusts for a single confounding covariate, such as age, in a study of smoking and mortality. I then show how propensity score methods generalize subclassification in the presence of many confounding covariates, such as age, region of the country, and sex.

The potential for a large database to suggest causal effects of treatments is indicated in Table 1, adapted from Cochran's work (2), which concerns mortality rates per 1000 person-years for nonsmokers, cigarette smokers, and cigar and pipe smokers drawn from three large databases in the United States, the United Kingdom, and Canada. The treatment factor here involves three levels of smoking. The unadjusted mortality rates in Table 1 make it seem that cigarette smoking is good for health, especially relative to cigar and pipe smoking; clearly, this result is contrary to current wisdom. A problem with this naive conclusion is exposed in Table 1, where the average ages of the subpopulations are given. Age correlates with both mortality rates and smoking behavior. In this example, age is a confounding covariate, and conclusions about the effects of smoking should be adjusted for its effects.

A straightforward way of adjusting for age is to 1) divide the population into age categories of approximately equal size (such as younger and older if two categories are appropriate; younger, middle-aged, and older if three are appropriate; and so on), 2) compare mortality rates within an age category (for example, compare mortality rates for the three treatment groups within the younger population and similarly for the older population), and 3) average the age-group-specific comparisons to obtain overall
estimates of the age-adjusted mortality rates per 1000 person-years for each of the three groups. Table 1 shows the results for different numbers of age categories where the subclass-age boundaries were defined to have equal numbers of nonsmokers in each subclass. These results align better than the unadjusted mortality rates with our current understanding of the effects of smoking, especially when 9 to 11 subclasses are used. Incidentally, having approximately equal numbers of nonsmokers within each subclass is not necessary, but if the nonsmokers are considered the baseline group, it is a convenient and efficient choice because then the overall estimated effect is the simple unweighted average of the subclass-specific results. That is, the mortality rates in all three groups are being standardized (3) to the age distribution of nonsmokers as defined by their subclass counts.

Cochran (2) calls this method subclassification and offers theoretical results showing that as long as the treatment and exposure groups overlap in their age distributions (that is, as long as a reasonable number of persons from each treatment group are in each subclass), comparisons using five or six subclasses will typically remove 90% or more of the bias present in the raw comparisons shown in Table 1. More than five subclasses were used for the adjusted mortality rates because the large size of the data sets made it possible to do so.

A particular statistical model, such as a linear regression (or a logistic regression model; or in other settings, a hazard model) could have been used to adjust for age, but subclassification has two distinct advantages over such models, at least for offering initial trustworthy comparisons that are easy to communicate. First, if the treatment or exposure groups do not adequately overlap on the confounding covariate age, the investigator will see it immediately and be warned. Thus, if members of one group have ages outside the range of another group’s ages, it will be obvious because one or more age-specific subclasses will consist almost solely of members exposed to one treatment. In contrast, nothing in the standard output of any regression modeling software will display this critical fact; the reason is that models predict an outcome (such as death) from regressors (such as age and treatment indicators), and standard regression diagnostics do not include careful analysis of the joint distribution of the regressors (such as a comparison of the distributions of age across treatment groups). When the overlap on age is too limited, the database, no matter how large, cannot support any causal conclusions about the differential effects of the treatments. For example, comparing 5-year survival rates among 70-year-old smokers and 40-year-old nonsmokers gives essentially no information about the effect of smoking or nonsmoking for either 70-year-old or 40-year-old persons.

The second reason for preferring subclassification to models concerns situations such as that found in Table 1, in which the groups overlap enough on the confounding covariate to make a comparison possible. Subclassification does not rely on any particular functional form, such as linearity, for the relation between the outcome (death) and the covariate (age) within each treatment group, whereas models do. If the groups have similar distributions of the covariate, such specific assumptions like linearity are usually harmless, but when the groups have different covariate distributions, model-based methods of adjustment are dependent on the specific form of the model (for example, linearity or log linearity) and their results are determined by untrustworthy extrapolations.

If standard models can be so dangerous, why are they commonly used for such adjustments when large databases are examined for estimates of causal effects? One reason is the ease with which automatic data analysis can be done using existing, pervasive software on plentiful, speedy hardware. A second reason is the seeming difficulty of using subclassification when many confounding covariates need adjustment, which is the common case. Standard modeling software can automatically handle many regressor variables and produce results, although they can be remarkably misleading. With many confounding covariates, however, the issues of

<table>
<thead>
<tr>
<th>Variable</th>
<th>Canadian Study</th>
<th>United Kingdom Study</th>
<th>United States Study</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Nonsmokers</td>
<td>Cigarette</td>
<td>Cigar and</td>
</tr>
<tr>
<td>Mortality rates per 1000</td>
<td></td>
<td>Smokers</td>
<td>Pipe Smokers</td>
</tr>
<tr>
<td>person-years, %</td>
<td>20.2</td>
<td>20.5</td>
<td>35.5</td>
</tr>
<tr>
<td>Average age, y</td>
<td>54.9</td>
<td>50.5</td>
<td>65.9</td>
</tr>
<tr>
<td>Adjusted mortality rates using subclasses, %</td>
<td>2 subclasses</td>
<td>20.2</td>
<td>26.4</td>
</tr>
<tr>
<td>3 subclasses</td>
<td>20.2</td>
<td>28.3</td>
<td>21.2</td>
</tr>
<tr>
<td>9–11 subclasses</td>
<td>20.2</td>
<td>29.5</td>
<td>19.8</td>
</tr>
</tbody>
</table>

* Adapted from Tables 1–3 in Cochran (2).
lack of adequate overlap and reliance on untrustworthy model-based extrapolations are even more serious than with only one confounding covariate. The reason is that small differences in many covariates can accumulate into a substantial overall difference. For example, if members of one treatment or exposure group are slightly older, have slightly higher cholesterol levels, and have slightly more familial history of cancer, that group may be substantially less healthy. Moreover, although standard comparisons of means between the groups like those in Table 1, or comparisons of histograms for each confounding covariate among groups are adequate with one covariate, they are inadequate with more than one. The groups may differ in a multivariate direction to an extent that cannot be discerned from separate analyses of each covariate. This multivariate direction is closely related to the statistical concept of the best linear discriminant and intuitively is the single combination of the covariates on which the treatment groups are farthest apart.

Subclassification techniques can be applied with many covariates with almost the same reliability as with only one covariate. The key idea is to use propensity score techniques, as developed by Rosenbaum and Rubin (1). These methods can be viewed as important extensions of discriminant matching techniques, which calculate the best linear discriminant between the treatment groups and match on it (4).

Since their introduction approximately 15 years ago, propensity score methods have been used in various applied problems in medical and other research disciplines (5-23) but not nearly as frequently as they should have been relative to model-based methods.

**Propensity Score Methods**

Propensity score methods must be applied to groups two at a time. Therefore, an example with three treatment or exposure conditions will generally yield three distinct propensity scores, one for each comparison (for the example in Table 1, nonsmokers compared with cigarette smokers, non-smokers compared with cigar and pipe smokers, and cigarette smokers compared with cigar and pipe smokers). To describe the way propensity scores work, I first assume two treatment conditions. Cases with more than two treatment groups are considered later.

The basic idea of propensity score methods is to replace the collection of confounding covariates in an observational study with one function of these covariates, called the propensity score (that is, the propensity to receive treatment 1 rather than treatment 2). This score is then used just as if it were the only confounding covariate. Thus, the collection of predictors is collapsed into a single predictor. The propensity score is found by predicting treatment group membership (that is, the indicator variable for being in treatment group 1 as opposed to treatment group 2) from the confounding covariates, for example, by a logistic regression or discriminant analysis. In this prediction of treatment group membership, it is critically important that the outcome variable (for example, death) play no role; the prediction of treatment group must involve only the covariates. Each person in the database then has an estimated propensity score, which is the estimated probability (as determined by that person's covariate values) of being exposed to treatment 1 rather than treatment 2. This propensity score is then the single summarized confounding covariate to be used for subclassification.

Subclassification into about five groups on the basis of the propensity score then has the rather remarkable property of adjusting for all of the covariates that went into its estimation, no matter how many there are. This is a large-sample claim that relies on certain conditions dealt with in technical statistical publications, but it is nevertheless an extremely useful guide for practice. The intuition behind the validity of this claim is fairly straightforward and proceeds as follows.

If two persons, one exposed to treatment 1 and the other exposed to treatment 2, had the same value of the propensity score, these two persons would then have the same predicted probability of being assigned to treatment 1 or treatment 2. Thus, as far as we can tell from the values of the confounding covariates, a coin was tossed to decide who received treatment 1 and who received treatment 2. Now suppose that we have a collection of persons receiving treatment 1 and a collection of persons receiving treatment 2 and that the distributions of the propensity scores are the same in both groups (as is approximately true within each propensity subclass). In subclass 1, the persons who received treatment 1 were essentially chosen randomly from the pool of all persons in subclass 1, and analogously for each subclass. As a result, within each subclass, the multivariate distribution of the covariates used to estimate the propensity score differs only randomly between the two treatment groups.

The formal proof of this result appears in Rosenbaum and Rubin (1). Research on how well this theoretical result is satisfied when using estimated rather than true propensity scores is the topic of technical statistical publications (24-28). Generally, the conclusion is that using estimated propensity scores in place of true propensity scores works very well.
Table 2. Estimated 5-Year Survival Rates for Node-Negative Patients in Six Randomized Experiments*

<table>
<thead>
<tr>
<th>Study</th>
<th>Treatment</th>
<th>Women</th>
<th>Estimated Survival Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>US-NCI†</td>
<td>Breast conservation</td>
<td>74</td>
<td>93.9</td>
</tr>
<tr>
<td></td>
<td>Mastectomy</td>
<td>67</td>
<td>94.7</td>
</tr>
<tr>
<td>Milanese†</td>
<td>Breast conservation</td>
<td>257</td>
<td>93.5</td>
</tr>
<tr>
<td></td>
<td>Mastectomy</td>
<td>263</td>
<td>93.0</td>
</tr>
<tr>
<td>Danish‡</td>
<td>Breast conservation</td>
<td>59</td>
<td>94.9</td>
</tr>
<tr>
<td></td>
<td>Mastectomy</td>
<td>62</td>
<td>95.2</td>
</tr>
<tr>
<td>EORTC§</td>
<td>Breast conservation</td>
<td>288</td>
<td>87.4</td>
</tr>
<tr>
<td></td>
<td>Mastectomy</td>
<td>288</td>
<td>85.9</td>
</tr>
<tr>
<td>US-NSABP‡</td>
<td>Breast conservation</td>
<td>330</td>
<td>89.0</td>
</tr>
<tr>
<td></td>
<td>Mastectomy</td>
<td>309</td>
<td>88.8</td>
</tr>
</tbody>
</table>

* Adapted from Table 2 in reference 3
† Single-center trial
‡ Multicenter trial

Propensity Subclassification

Several years ago, the U.S. Government Accounting Office (29) summarized results from randomized experiments comparing mastectomy (removal of the breast but not the pectoral muscle with nodal dissection but no radiation) and breast conservation therapy (lumpectomy, nodal dissection, and radiation) for the treatment of breast cancer in node-negative patients. The results, shown in Table 2 (29), provide no evidence of differential treatment effect, at least for the type of women who participated in these informed consent clinical trials and who received the kind of care dispensed at the centers participating in these trials. The question remained, however, how broadly these results could be generalized to other node-negative women and other medical facilities. The U.S. Government Accounting Office used the National Cancer Institute’s SEER (Surveillance, Epidemiology and End Results) observational database to address this question. Restrictions (including node-negative diagnosis, 70 years of age or younger, and tumor size ≤4 cm [29]) were applied to match criteria for the randomized experiments. These restrictions reduced the database to 1106 women who received breast conservation therapy and 4220 who received mastectomy, for a total of 5326 women.

The U.S. Government Accounting Office used propensity score methods on the SEER database to compare the two treatments for breast cancer. First, approximately 30 potential confounding covariates and interactions were identified: year of diagnosis (1983–1985), age category (4 levels), tumor size, geographical registry (9 levels), race (4 levels), marital status (4 levels), and interactions of year and registry. A logistic regression was then used to predict treatment (mastectomy compared with conservation therapy) from these confounding covariates on the basis of data from the 5326 women. Each woman was then assigned an estimated propensity score, which was her probability, on the basis of her covariate values, of receiving breast conservation therapy rather than mastectomy. The group was then divided into five subclasses of approximately equal size on the basis of the women’s individual propensity scores: 1064 in the most mastectomy-oriented subclass, 1070 in the next subclass, 1059 in the middle subclass, 1067 in the next subclass, and 1066 in the most breast conservation–oriented subclass.

Before examining any outcomes (5-year survival results), the subclasses were checked for balance with respect to the covariates. Propensity score theory claims that if the propensity scores are relatively constant within each subclass, then within each subclass, the distribution of all covariates should be approximately the same in both treatment groups. This balance was found to be satisfactory. If important within-subclass differences between treatment groups had been found on some covariates, then either the propensity score prediction model would need to be reformulated or it would have been concluded that the covariate distributions did not overlap sufficiently to allow subclassification to adjust for these covariates. This process of cycling between checking for balance on the covariates and reformulating the propensity score model is described by Rosenbaum and Rubin (18) in the context of a study investigating coronary bypass surgery. For example, when the variances of an important covariate were found to differ importantly between treatment and control groups, then the square of that covariate was included in the revised propensity score model. For another example, if the correlations between two important covariates differed between the groups, then the product of the covariates was added to the propensity score model.

The estimates of 5-year survival rates made on basis of the resulting propensity score subclassification are given in Table 3 (29). Total rates and rates excluding deaths unrelated to cancer are shown. Several features of Table 3 are particularly striking, especially when compared with the results of the randomized experiments shown in Table 2. First, the general conclusion of similar performance of both treatments is maintained. Second, although overall survival is similar across treatment groups, the results indicate that survival in general practice may be slightly lower than suggested by data from the population of women and types of clinics participating in the randomized clinical trials, especially in the single-clinic studies.

Third, results slightly indicate that, in general practice, women and their physicians may be mak-
Table 3. Estimated 5-Year Survival Rates for Node-Negative Patients in the SEER Database within Each of Five Propensity Score Subclasses*

<table>
<thead>
<tr>
<th>Propensity Score Subclass</th>
<th>Treatment</th>
<th>Women</th>
<th>Estimated Survival Rate for Women</th>
<th>Omitting Women Whose Deaths Were Unrelated to Cancer</th>
<th>Estimated Survival Rates Omitting Women Whose Deaths Were Unrelated to Cancer</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>n</td>
<td>%</td>
<td>n</td>
<td>%</td>
</tr>
<tr>
<td>1</td>
<td>Breast conservation</td>
<td>56</td>
<td>85.6</td>
<td>54</td>
<td>88.8</td>
</tr>
<tr>
<td></td>
<td>Mastectomy</td>
<td>1008</td>
<td>86.7</td>
<td>966</td>
<td>90.5</td>
</tr>
<tr>
<td>2</td>
<td>Breast conservation</td>
<td>106</td>
<td>82.8</td>
<td>102</td>
<td>86.0</td>
</tr>
<tr>
<td></td>
<td>Mastectomy</td>
<td>964</td>
<td>83.4</td>
<td>917</td>
<td>87.7</td>
</tr>
<tr>
<td>3</td>
<td>Breast conservation</td>
<td>193</td>
<td>85.2</td>
<td>184</td>
<td>89.4</td>
</tr>
<tr>
<td></td>
<td>Mastectomy</td>
<td>866</td>
<td>88.8</td>
<td>841</td>
<td>91.4</td>
</tr>
<tr>
<td>4</td>
<td>Breast conservation</td>
<td>289</td>
<td>88.7</td>
<td>279</td>
<td>91.0</td>
</tr>
<tr>
<td></td>
<td>Mastectomy</td>
<td>978</td>
<td>87.3</td>
<td>742</td>
<td>91.5</td>
</tr>
<tr>
<td>5</td>
<td>Breast conservation</td>
<td>462</td>
<td>89.0</td>
<td>453</td>
<td>90.7</td>
</tr>
<tr>
<td></td>
<td>Mastectomy</td>
<td>604</td>
<td>88.5</td>
<td>589</td>
<td>90.7</td>
</tr>
</tbody>
</table>

* Adapted from Tables 5 and 7 in reference 3. SEER = Surveillance, Epidemiology, and End Results.

ing beneficial choices. More precisely, women in propensity subclasses 1 to 3, composed of patients whose characteristics (including age, size of tumor, and region of country) make them relatively more likely to receive mastectomy than breast conservation therapy, seem to show better 5-year survival with mastectomy than with breast conservation therapy. In contrast, for women in propensity subclasses 4 and 5 (whose characteristics make them relatively more likely to receive breast conservation therapy than mastectomy), there seems to be no advantage to mastectomy and possibly a slight advantage to breast conservation therapy. Of course, this last interpretation is subject to two caveats. First, we only adjusted for the covariates that were used to estimate the propensity score; hence, other hidden covariates may alter this interpretation. In a randomized experiment, the effects of these hidden covariates are reflected in the SEs of the estimates, but in an observational study, these effects can create bias not reflected in the SEs. Second, the sampling variability (that is, SEs) of the results do not permit firm conclusions, even if the collection of confounding covariates was sufficient to remove bias in this observational study.

Although there is no randomized assignment in the SEER database, the propensity score analyses seem to provide useful suggestive results, especially when coupled with the results of the randomized experiments, with which they are consistent.

**More Than Two Treatment Conditions**

With more than two treatment conditions, the propensity score usually differs for each pair of treatment groups being compared (that is, with three treatment groups labelled A, B, and C, there are three propensity scores: A compared with B, A compared with C, and B compared with C). At first, this may seem to be a limitation of propensity score technology relative to a model-based analysis, but in fact it is an important strength and points to further weaknesses in a model-based approach. We show this by exploring a range of hypothetical modifications to Cochran's (2) smoking example.

First, consider what we could have learned if the nonsmokers and cigarette smokers had had adequately overlapping age distributions, but the cigar and pipe smokers had been substantially older than persons in either of the other groups, with essentially no overlap with the cigarette smokers or the nonsmokers. Even with only one covariate, with more than two groups, the groups in one two-group comparison (nonsmokers compared with cigarette smokers) may overlap adequately, whereas for all other comparisons (in this example, those involving cigar and pipe smokers), the overlap may be inadequate. A typical model-based analysis would use all the data to provide estimates for all three two-group comparisons, even using the data from the cigar and pipe smokers to influence the comparison between the nonsmokers and the cigarette smokers, with no warning of either the extreme extrapolations involved in two of the three two-group comparisons or the use of data on cigar and pipe smokers to help estimate the comparison of nonsmokers and cigarette smokers.

Let us again modify the Cochran (2) smoking example but now include an additional covariate: an index of socioeconomic status. We assume that nonsmokers and cigarette smokers have adequate overlap in their age distributions but not much overlap in their socioeconomic status distributions, with nonsmokers having higher socioeconomic status values. In contrast, we suppose that nonsmokers and cigar and pipe smokers have substantial overlap in their socioeconomic distributions but have essentially no overlap in their age distributions. This scenario illustrates that with more than two groups and more than one covariate, the comparison of one
pair of groups can be compromised by one covariate and the comparison of another pair of groups can be compromised by a different covariate. As discussed earlier, typical model-based analyses provide no warning that comparisons may be based on extreme extrapolations, nor do they show that the extrapolations include data from groups that are not in the pair of groups being compared.

Now suppose that the nonsmokers and cigarette smokers have the same age distributions and adequately overlapping socioeconomic status distributions. For this comparison, age needs no adjustment but socioeconomic status does need to be adjusted. The propensity score for the comparison would essentially equal socioeconomic status because it, and not age, would predict being a cigarette smoker as opposed to being a nonsmoker. Thus, for this comparison, adjusting for the propensity score would be the same as adjusting for socioeconomic status. Now also assume that the nonsmokers and cigar and pipe smokers have the same socioeconomic status distributions, so that socioeconomic status needs no adjustment, and have adequately overlapping age distributions that need adjustment. Then the propensity score for this comparison would equal age, and therefore, adjusting for the propensity score would be the same as adjusting for age. Thus, the propensity score for a comparison of one pair of groups generally needs to be different from that for a comparison of a different pair of groups. To complete the current scenario, assume that cigarette smokers and cigar and pipe smokers have adequate overlap in both age and socioeconomic status and that both need adjustment. The propensity score for this comparison would involve both age and socioeconomic status because both help to predict cigarette group membership, as opposed to cigar and pipe smoking group membership, and adjusting for this propensity score would adjust for both age and socioeconomic status. Clearly, different propensity score models are needed to adjust appropriately for different comparisons. Estimating all effects by using one model in our example with three groups and adequate overlap on all covariates can be even more deceptive than estimation in the two-group setting because the model being used to compare one pair of groups (for example, nonsmokers compared with cigarette smokers) is affected by the data from the third group (here cigar and pipe smokers), which probably has covariate values that differ from those in either one of the other two groups being compared.

**Limitations of Propensity Scores**

Despite the broad utility of propensity score methods, when addressing causal questions from nonrandomized studies, it is important to keep in mind that even propensity score methods can only adjust for observed confounding covariates and not for unobserved ones. This is always a limitation of nonrandomized studies compared with randomized studies, where the randomization tends to balance the distribution of all covariates, observed and unobserved.

In observational studies, confidence in causal conclusions must be built by seeing how consistent the obtained answers are with other evidence (such as results from related experiments) and how sensitive the conclusions are to reasonable deviations from assumptions, as illustrated by Connors and colleagues (20), who used techniques from Rosenbaum and Rubin's work (30). Such sensitivity analyses suppose that a relevant but unobserved covariate has been left out of the propensity score model. By explicating how this hypothetical unmeasured covariate is related to treatment assignment and outcome, we can obtain an estimate of the treatment effect that adjusts for it as well as for measured covariates and thereby investigate how answers might change if such a covariate were available for adjustment. Of course, medical knowledge is needed when assessing whether the posited relations involving the hypothetical unmeasured covariate are realistic or extreme. Clarifications of nomenclature and extended sensitivity analyses reported by Lin and colleagues (31) moderate the initial conclusions of Connors and colleagues (20).

Another limitation of propensity score methods is that they work better in larger samples for the following reason. The distributional balance of observed covariates created by subclassifying on the propensity score is an expected balance, just as the balance of all covariates in a randomized experiment is an expected balance. In a small randomized experiment, random imbalances of some covariates can be substantial despite randomization; analogously, in a small observational study, substantial imbalances of some covariates may be unavoidable despite subclassification using a sensibly estimated propensity score. The larger the study, the more minor are such imbalances.

A final possible limitation of propensity score methods is that a covariate related to treatment assignment but not to outcome is handled the same as a covariate with the same relation to treatment assignment but strongly related to outcome. This feature can be a limitation of propensity scores because inclusion of irrelevant covariates reduces the efficiency of the control on the relevant covariates. However, recent work (28) suggests that, at least in modest or large studies, the biasing effects of leaving out even a weakly predictive covariate dominate the efficiency gains from not using such a
covariate. Thus, in practice, this limitation may not be substantial if investigators use some judgment.

**Conclusion**

Large databases have tremendous potential for addressing (although not necessarily settling) important medical questions, including important causal questions involving issues of policy. Addressing these causal questions using standard statistical (or econometric, psychometric, or neural net) models can be fraught with pitfalls because of their possible reliance on unwarranted assumptions and extrapolations without any warning. Propensity score methods are more reliable; they generalize the straightforward technique of subclassification with one confounding covariate to allow simultaneous adjustment for many covariates. One critical advantage of propensity score methods is that they can warn the investigator that, because of inadequately overlapping covariate distributions, a particular database cannot address the causal question at hand without relying on untrustworthy model-dependent extrapolation or restricting attention to the type of person adequately represented in both treatment groups. Because of this advantage, any causal questions put to a large database should be first approached using propensity score methods to see whether the question can be legitimately addressed. If so, subclassification on a well-estimated propensity score can be used to provide reliable results, which are adjusted for the covariates used to estimate the propensity score and which can be clearly displayed. After that, modeling can play a useful role. For example, standard statistical models, such as least-squares regression, can be safely applied within propensity score subclasses to adjust for minor within-subclass differences in covariate distributions between treatment groups. This was done in the example of the study by the U.S. Government Accounting Office (29). Of course, it always must be remembered that propensity scores only adjust for the observed covariates that went into their estimation.

**Grant Support:** In part by a grant from the National Science Foundation (SES-9207456).

**Acknowledgments:** The author thanks Jennifer Hill and Frederick Mosteller for helpful editorial comments on an earlier draft of this article.

**Requests for Reprints:** Donald B. Rubin, PhD, Harvard University, Department of Statistics, Science Center, 6th Floor, 1 Oxford Street, Cambridge, MA 02138.

**References**

1. What is the clinical question being addressed?

“In elderly patients with variable preoperative anemia and volume of blood loss during major non-cardiac surgery, does intra-operative blood transfusion reduce the risk of perioperative death?”

2. What is the study design?

Retrospective cohort study. Two cohorts of patients who were and those who were not transfused were assembled retrospectively using the National Surgical Quality Improvement Program (NSQIP) database. Although the NSQIP data are collected prospectively, this study was done after the data were collected and therefore is considered a “retrospective” cohort study.

3. What is the source of data?

The NSQIP database, a Veterans Administration database is used to identify patients who underwent noncardiac surgery from 1997 to 2004 (310,311 patients). This is merged with VA patient treatment files and Outpatient Care files to enhance the available clinical data including data on the use of autologous transfusions, prior hospitalizations related to coronary artery disease, cell saver procedures within 30 days prior to surgery and treatment of coronary artery disease treated in the ambulatory setting.

4. Are the data accurate and valid?

Likely. Patients were identified from the NSQIP data base. Data in the NSQIP database are extracted from patient records by nurses and are periodically re-extracted by external reviewers to ensure good agreement.

For a propensity analysis to be valid, the variables used as covariates in the propensity score must be accurate. No information is provided about the validity or accuracy of the VA . NSQIP data but the authors have previously shown that the NSQIP data to be more complete than that extracted from medical records charts. It is likely that the patient treatment files and the outpatient care files are administrative data and may be less complete and accurate. Furthermore, an important variable in this equation is the intraoperative blood loss. Because intraoperative blood loss volume was no longer recorded in the NSQIP after August 1995, the volume of blood loss for the patients included in this study had to be estimated, and for this a formula was derived, which may make the data less accurate.
5. Is the matching appropriate and are the 2 groups similar?

In the unmatched cohorts, there are statistically significant differences between patients who did and did not receive blood transfusions with respect to most measured variables (Table 1). This is not surprising since patients who receive a blood transfusion are usually selected to receive one because of certain patient characteristics or operative factors. However, these confounders would invalidate any comparison to determine whether blood transfusion has detrimental effects and confirms the need for a propensity score technique for analysis. As stated by Rubin, propensity score technology reduces the entire collection of background characteristics to a single composite characteristic that appropriately summarizes the potential collection of confounders'.

In the propensity matched cohorts (Table 2), the above clinically and statistically significant differences between transfused and nontransfused groups are no longer present for all variables except the subclasses of preoperative hematocrit, mean operative time (3.8 versus 2.8 hours), proportion having a general anesthetic (89.3% versus 80%), proportion of ASA class 4 or 5 patients (24.8% versus 25.2%, and 1.5% versus 1.0% respectively), myocardial infarction within 6 months (2.2% versus 1.8%), mean operative time (3.8 h versus 2.8 hours) and as expected, number of units of blood cells transfused (2.6u versus 0 units).

Propensity scores allow groups to be matched according to those variables that can be matched although, unlike with randomization, groups may not be similar with respect to unmeasured variables.

6. Identify potential biases that might account for differences in the conclusions?

This study used retrospective data from a database not designed to answer this clinical question. Furthermore, patients were not randomized and as stated above, there may be unknown or unmeasurable variables which lead to differences in the two groups. Because blood transfusion during major non cardiac surgery is a situational variable, the decision to transfuse is best understood when studied in context and this limited dataset likely is not detailed enough to provide all the information that led to the decision to transfuse. Specifically, the paucity of data with respect to volume of blood loss intraoperatively, and the need to calculate estimates, could be a source of inaccuracy. In addition to that, a surgeon’s personal bias for transfusion might be partially controlled for by including his/her post graduate year and specialty into the propensity score, but the human nature of this decision making process might never be standardized as well as if the decision to transfuse was based on a standardized protocol. Certainly most would agree that clinicians are more likely to transfuse patients who are already anemic prior to surgery regardless of intraoperative events. Finally, the risk of blood loss in emergency surgery may vary depending on the specialty and this is not captured within the model, leading the reader to assume that urgent surgery for appendicitis and rupture abdominal aortic aneurysms are given equal weight – despite the different blood transfusion requirement re in each case.

Another confounder that is not controlled for in this model is the fact that only one procedure is considered for each patient. There is a statistically significant difference in the number of patient with post operative hemorrhage between groups transfused and not transfused. Knowledge of need for subsequent, secondary intervention might have a large impact on 30 day mortality.
7. What outcomes were assessed and are they clinically relevant and sensitive?

The primary outcome was the adjusted odds ratio of 30 day mortality. The authors subdivide this outcome based on preoperative hematocrit, and then further group outcomes based on the estimated blood loss in the index operation. These outcomes and subgroup analyses are clinically relevant and appropriate, with the caveat that the blood loss used is an estimated value. In addition, post operative complications were also compared.

8. Are the differences clinically significant?

Yes. The authors found that the unadjusted 30 day mortality for patients who had intraoperative transfusion was 10.5%, and for those who did not get transfusion was 8.6%. This corresponds to an adjusted odds ratio of 1.37 (95% confidence interval 1.27 – 1.48). Once sub-grouped based on preoperative hematocrit level, the results revealed that the risk varied depending on the preoperative hematocrit, where mortality risk was reduced for patients with a low initial hematocrit undergoing transfusion (OR 0.6, 95% CI 0.41 – 0.87), but this was only true for patients with higher preoperative hematocrit levels if the blood loss was 500 -999mL [(pre-op hematocrit 30 – 35%, OR 0.35 (95% CI 0.22 – 0.56) and pre-op hematocrit 36 – 53.9%, OR 0.78 (95% CI 0.62 – 0.97)].

9. Are further studies required?

Yes. Propensity score analysis is useful when analyzing observational data because it helps to control for the effect that some covariates have on the measured outcome. However, only measured variables can be included in the analysis. A randomized controlled trial would eliminate differences in unmeasured or unknown variables and the result would be more valid. As well, the population in this study does not contain a representative sample of women, so the results are not generalizable to women receiving intraoperative blood transfusion.

10. State the conclusion. Have the authors addressed the question posed?

The authors concluded: “Intraoperative blood transfusion is associated with a lower 30-day postoperative mortality among elderly patients undergoing major non cardiac surgery if there is substantial operative blood loss or low preoperative hematocrit levels (<24%). Transfusion is associated with increased mortality risks for those with preoperative hematocrit levels between 30% and 35.9% and <500mL blood loss.” The authors have addressed the original clinical question but only in the patient population studied which were predominantly elderly males,

11. Do the data support the conclusions?

Within the caveats of a retrospective analysis of an administrative dataset, the data do support the conclusions for male patients >65 years. There is need for further research into the effect of intraoperative transfusion in the female population.

Tara Mastracci and Members of the Evidence Based Reviews in Surgery Committee
References:


Anemia and blood loss are very common problems in surgery, and are effectively treated with red blood cell (RBC) transfusion. Transfusions however, can be either life-saving or harmful and there is a lot of uncertainty on which thresholds to use for this therapy. Even perioperative transfusion guidelines are mostly based on expert consensus and are not evidence-based\textsuperscript{1,2}. The present study is relevant as it addresses intraoperative transfusion threshold, a common and important surgical issue. Wu et al demonstrated that intraoperative RBC transfusion to elderly patients with low preoperative hematocrit (HT) or substantial operative blood loss was associated with lower mortality. On the other hand, transfusion was associated with increased mortality when given to patients with preoperative HT of 30 to 35% and without significant intraoperative blood loss.

The benefit of transfusing RBC is that it increases oxygen delivery, which depends on hemoglobin level as well as cardiac output, oxygen saturation and PaO\textsubscript{2}. The harm may be explained by the fact that despite increasing oxygen delivery, RBC transfusion may impair microcirculation flow, increase inflammation and not necessarily increase oxygen consumption by tissues. Every RBC transfusion increases blood viscosity, may result in the infusion of cytokines, activated leukocytes and hemolysed free hemoglobin and reduces NO availability leading to vasoconstriction. Stored RBC are less deformable reducing their ability to cross capillaries and the decreased\textsuperscript{2,3} Diphosphoglycerate (DPG) levels reduces oxygen release to tissues. Blood transfusion also has infectious and non-infectious complications (transfusion related acute lung injury (TRALI), transfusion-associated circulatory overload (TACO), hemolytic reaction, anaphylaxis, etc), is expensive and scarce. Thus the determination of intraoperative transfusion thresholds that improve survival or reduce harm and cost is tremendously important and relevant.

Many points in the present study cooperate to make the results generalizable. The authors chose the most vulnerable surgical population (elderly), thus the one more likely to benefit or be harmed by the intervention (transfusion). Elderly patients have more co-morbidities particularly ischemic heart disease (IHD); high APACHE scores; diseases such as cancer, sepsis and even trauma; high operative risk; undergo urgent surgery; have shock; etc. All these factors are commonly considered when deciding whether to transfuse a patient. In addition the study included a large number of patients (239,286) from 132 VA hospitals from across the United States and undergoing non-cardiac surgery by 8 different subspecialties. It also excluded patients with low morbidity and mortality and required only a simple and frequently done preoperative lab test (HT).
Generalizability suffers because the study included almost only elderly (≥65 years) war veteran male patients (98%) who were white (about 80%). The lack of women and other races limits the application of the results. Despite such limitations, I agree with the authors that this study is arguably “the best evidence to date to guide blood use in the operative setting”, thus generalizable for this population.

There is a general lack of studies on RBC transfusion thresholds. The best recent studies come from Critical Care with a strong Canadian participation. Arguably the most important study on the topic is the Transfusion Requirements in Critical Care (TRICC) study. This trial, which included surgical and trauma patients, proposed that RBC transfusion is not required in critically ill euvolemic patients until the Hb concentration drops below 70g/L\(^3\). The threshold for unstable patients or with ischemic heart disease (who may be less able to tolerate anemia) is mostly unknown but a trigger between 80-100g/L is commonly used.

Unfortunately even less is evidence exists on intraoperative RBC transfusion. A study on Jehovah’s Witness patients who refused blood transfusion showed postoperative mortality increases significantly if preoperative Hg levels were below 40g/L\(^4\). Another study by the same authors showed that among surgical patients with ischemic heart disease, the postoperative mortality increased significantly if preoperative Hg levels were below 90g/L\(^5\).

The results of the present study are similar to those above. Wu et al demonstrated an association between intraoperative transfusion and lower 30-day mortality in patients with preoperative HT levels <24% (approximately 80g/L). Interestingly they also demonstrated an increased mortality if patients with preoperative HT levels ≥30% were transfused, except when the transfusion was for significant intraoperative blood loss (500-999ml). Thus the results of the present study are similar to others suggesting restricting RBC transfused to either Hg levels below 70-80g/L (or HT levels below 24%), except when there is significant (500-999ml) blood loss.

The impact of this study will depend on the policies and guidelines of each institution and on how strong the results are believed to be. Many hospitals already practice restrictive policies that limit RBC transfusions to specific situations and patients. Decisions regarding transfusions are commonly made in collaboration with transfusion specialists and undergo periodical audits. In such institutions, transfusion triggers around HT <24% (or Hg levels <70-80g/L) are already in place and the present study would only reinforce their practice. However many institutions, including large academic centers, have more liberal practices. For these institutions, the evidence from this study and others could be used to change towards more restrictive indications.

Concerning how strong the results are and whether they should change practice, it is important to consider that they are similar to those from other studies (i.e. TRICC). Furthermore, the results of the present study is also in line with the prevalent concept that transfusions should be restricted to specific situations and prescribed (when possible) based on evidence.

Further research is needed without question. Transfusions in certain situations are life-saving while in others they cause harm. Determining optimal transfusion triggers may reduce mortality and morbidity as well as inappropriate use of blood, an expensive and scarce resource.
Future studies should consider including a broader population (women); better evaluation of anemia (HT has many limitations) and of intraoperative blood loss (the formula is of no clinical use). The ideal future study would be randomized and controlled and should also consider leukodepletion, blood age, duration of surgery and patient weight among other factors.

Most RBC transfusion are mostly done based on lab tests, so the implementation of a transfusion threshold (HT <24%) should be easy to implement. The challenge would be to have intraoperative transfusion being dictated by blood loss. The formula used in the study is not applicable, consequently other ways to estimate blood would be needed.

Sandro Rizoli MD PhD FRCSC FACS  
Associate Professor of Surgery and Critical Care Medicine  
University of Toronto

Bartolomeu Nascimento MD  
Clinical Associate Transfusion Medicine and Critical Care Medicine  
University of Toronto

References:

Decisions surrounding perioperative blood transfusions are commonplace in clinical medicine and have significant implications. It is estimated that approximately 40% of patients scheduled for noncardiac surgery have preoperative anemia and it is well established that these patients are at an increased risk for postoperative morbidity and mortality. Intraoperative blood loss can exacerbate anemia, and has also been shown to increase mortality. As the mainstay of treatment for acute anemia in the operative setting remains blood transfusions, the use of such products and associated costs are substantial. In the United States, 13.9 million RBC units were transfused to 4.9 million patients in 2001, and in 2001/2002, the Canadian Blood Services expenditures totalled 638.8 million dollars. However, as outlined by Wu et al, blood transfusions do not offer a simple solution to this clinical problem as they are independently associated with increased levels of morbidity and mortality. Mortality is most often attributable to acute lung injury, hemolytic reactions and circulatory overload.

The evidence around clinical parameters that may serve as transfusion “triggers” is equivocal and the decision making process remains largely a matter of clinical gestalt. Practice guidelines set out by the American Society of Anesthesiologist (ASA) recommend that in the event of intraoperative blood loss or ischemia, a hemoglobin value of less than 6g/dl warrants blood transfusion, whereas transfusions should be withheld for hemoglobin levels of 10g/dl or higher. Transfusion decisions for hemoglobin levels between these thresholds should be based on indications such as ongoing bleeding and ischemia, intravascular volume and underlying risk factors (low cardiopulmonary reserve, high oxygen consumption). Given the adverse implications associated with anemia, blood loss and transfusions, as well as the large window of decision-making subjectivity (Hb 6-10g/dl) for which research remains insufficient, the clinical and academic relevance of this study by Wu et al cannot be overstated.

With over 239,000 patients (over 37,000 in the propensity-matched analysis) from 132 Veteran Affairs hospitals, the results of this study can be extrapolated and applied to a broad patient population across North America. The most suitable patients would be those that reflect the study sample—elderly males undergoing noncardiac surgery. As mentioned by the study authors, it would be challenging to apply these findings to a female population given their higher likelihood of preoperative anemia and of receiving blood transfusions. The results should also be limited to patients within the general age group of the study participants (74±6). The postoperative mortality risks associated with anemia and blood loss most likely result from tissue injury to vital organs. Given that younger patients have a higher physiological reserve and are...
less likely to have underlying medical conditions (i.e. coronary disease), their capacity to tolerate blood loss and thresholds for transfusions may differ from the elderly patient population addressed in this study.

As outlined in the 2006 guidelines by the ASA *Task Force on Perioperative Blood Transfusions and Adjuvant Therapies*, “the information needed to define precisely when a blood transfusion should be given is not available in the literature”

The majority of studies investigating the outcomes of blood transfusions have correlated results solely to preoperative hemoglobin or hematocrit levels. The evidence base around blood transfusions remains controversial, as some studies suggest transfusions increase mortality whereas other studies show no differences in adverse outcomes. A recently published Cochrane systematic review of 17 randomized trials (8 surgical) comparing “restrictive” to “liberal” transfusion triggers, found that although a restrictive strategy was associated with an absolute risk reduction of 33% in receiving a transfusion, there was no difference in adverse events between the groups. The authors concluded that further high quality clinical trials are needed to delineate the effect varying transfusion thresholds have on adverse outcomes. The prevailing notion from the results of these various studies suggest that blood transfusions do not routinely improve outcomes in anemic patients and have a narrow therapeutic window of benefit.

The study by Wu *et al* provides a novel perspective by assessing outcomes in relation to both hematocrit levels and intraoperative blood loss. Accordingly, it uniquely attempts to challenge and add further specifications to the ASA guidelines which offer minimal decision-making guidance when patients’ hemoglobin levels are between 6-10g/dl. In particular, the study suggests that blood transfusions may provide a mortality benefit to those patients with a preoperative hematocrit of less than 24% (about 8g/dl) irrespective of blood loss, and those with a hematocrit of greater than 30% in the case of significant blood loss (greater than 500ml).

Although this study does not warrant a paradigm shift in management decisions regarding blood transfusion, it does provide insight into a common clinical problem plagued with uncertainty. The decision making model put forth by this study, which accounts for hematocrit levels in the context of intraoperative blood loss, offers a potential approach to minimizing the uncertainty of deciding when to transfuse surgical patients. However, to warrant implementation into routine clinical practice, a high-quality randomized controlled trial would be optimal to investigate the true effects of these hematocrit and blood loss thresholds as transfusion triggers.

*Raman Mundi MD, Mohit Bhandari MD, PhD, FRCSC
Division of Orthopaedic Surgery
McMaster University*

References:


Wu and colleagues have performed a retrospective cohort study utilizing National Surgical Quality Improvement Program (NSQIP) data to determine whether intraoperative blood transfusions reduce mortality in VA patients. The study population included 239,286 patients from 132 VAs who underwent major surgery between 1997 and 2004 by 8 surgical subspecialties. The investigators compared patients who received blood transfusions to those who did not using propensity-score matching to control for the intrinsic differences between the 2 populations.

The authors found that, after propensity matching, patients who received blood transfusions had a higher mortality than those who did not. Subgroup analyses of the matched cohort revealed that patients with a preoperative hematocrit < 24% had a better chance of surviving when they were transfused regardless of the amount of blood loss. Patients with a preoperative hematocrit > 30% with an estimated blood loss between 500 and 999 ml also had improved survival. Mortality was higher in transfused patients with a preoperative hematocrit between 30 and 35.9 whose estimated blood loss was less than 500 ml.

This study represents a critical addition to the literature because it helps to define the population of patients who benefit from blood transfusions. There is a growing body of data suggesting that blood transfusions are harmful. In a large randomized controlled trial, Hébert et al showed that there is no benefit and possible harm associated with a liberal transfusion strategy defined as a hemoglobin transfusion trigger of 10.0 grams/dl. Patients with active bleeding, chronic anemia and active angina were excluded from Hebert’s study. Others have shown that blood transfusions are independently associated with increased mortality, multiple organ failure and increased infection rates. The current study is unique in that it is very large, it evaluates peri-operative patients and extensive subgroup analyses have been performed.

The primary limitation of this study is that it is retrospective. Propensity matching is a powerful tool to control for potential biases but it does not overcome the fact that the 2 populations were fundamentally different. As shown in Table 2, the transfused population and the non-transfused population differed with respect to all of the evaluated characteristics. For the most part transfused patients had significantly more comorbidities and underwent more complex and longer surgeries. Even after propensity-matching, significant differences persisted in important
parameters like preoperative hematocrit, ASA classification, history of myocardial infarction, mean operative time and usage of general anesthesia.

Estimated blood loss is an important parameter in this study. The variables used to determine the estimated blood loss included pre-operative hematocrit, post-operative hematocrit and packed red blood cells transfused. This equation does not take into consideration the variable quantity of crystalloid and colloid given intraoperatively which produces hemodilution and affects the postoperative hematocrit. It is not clear that the authors used the best possible method to estimate operative blood loss.

Another limitation includes the failure to discuss post-operative morbidity in patients who survived. The incidence of pulmonary complications, multiple organ failure and sepsis in surviving patients would be important data points. Increased morbidity in some of the subgroups could affect the decision to transfuse during the perioperative period.

The results of this study can not be generalized to a broad population. The study group was greater than or equal to 65 years old and almost exclusively male. Despite all of the limitations of the study, the point that pre-operative hematocrit and estimated blood loss could potentially be used to identify patients who may benefit from transfusions is an important one and further study is certainly indicated. A randomized controlled trial performed in a variable aged population of mixed genders that randomized patients to transfusion or no transfusion based on pre-operative hematocrit and estimated blood loss would be feasible and would provide important information about the indications for peri-operative transfusion. Such a study should include perioperative morbidity and mortality as endpoints.

Martin A. Schreiber, MD FACS
Oregon Health Science University

References: