

Previous Presentation: This paper was presented at the American College of Surgeons 95th Clinical Congress; October 12, 2009; Chicago, Illinois; and is published after peer review and revision.

REFERENCES

1. MacKenzie EJ, Rivara FP, Jurkovich GJ, et al. A national evaluation of the effect of trauma-center care on mortality. *N Engl J Med*. 2006;354(4):366-378.
2. Haas B, Jurkovich GJ, Wang J, Rivara FP, Mackenzie EJ, Nathens AB. Survival advantage in trauma centers: expeditious intervention or experience? *J Am Coll Surg*. 2009;208(1):28-36.
3. D'Agostino RB Jr. Propensity scores in cardiovascular research. *Circulation*. 2007;115(17):2340-2343.
4. Rubin DB. Estimating causal effects from large data sets using propensity scores. *Ann Intern Med*. 1997;127(8, pt 2):757-763.
5. Kao LS, Lally KP, Thomas EJ, Tyson JE. Improving quality improvement: a methodologic framework for evaluating effectiveness of surgical quality improvement. *J Am Coll Surg*. 2009;208(4):621-626.
6. Adamina M, Guller U, Weber WP, Oertli D. Propensity scores and the surgeon. *Br J Surg*. 2006;93(4):389-394.
7. Austin PC, Grootendorst P, Anderson GM. A comparison of the ability of different propensity score models to balance measured variables between treated and untreated subjects: a Monte Carlo study. *Stat Med*. 2007;26(4):734-753.
8. Stukel TA, Fisher ES, Wennberg DE, Alter DA, Gottlieb DJ, Vermeulen MJ. Analysis of observational studies in the presence of treatment selection bias: effects of invasive cardiac management on AMI survival using propensity score and instrumental variable methods. *JAMA*. 2007;297(3):278-285.
9. Guller U, Hervey S, Purves H, et al. Laparoscopic versus open appendectomy: outcomes comparison based on a large administrative database. *Ann Surg*. 2004;239(1):43-52.
10. Faiz O, Clark J, Brown T, et al. Traditional and laparoscopic appendectomy in adults: outcomes in English NHS hospitals between 1996 and 2006. *Ann Surg*. 2008;248(5):800-806.
11. Dimick JB, Chen SL, Taheri PA, Henderson WG, Khuri SF, Campbell DA Jr. Hospital costs associated with surgical complications: a report from the private-sector National Surgical Quality Improvement Program. *J Am Coll Surg*. 2004;199(4):531-537.
12. Hemmila MR, Jakubis JL, Maggio PM, et al. Real money: complications and hospital costs in trauma patients. *Surgery*. 2008;144(2):307-316.
13. Sauerland S, Lefering R, Neugebauer EA. Laparoscopic versus open surgery for suspected appendicitis. *Cochrane Database Syst Rev*. 2004;4(4):CD001546.
14. D'Agostino RB Jr. Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. *Stat Med*. 1998;17(19):2265-2281.
15. Braitman LE, Rosenbaum PR. Rare outcomes, common treatments: analytic strategies using propensity scores. *Ann Intern Med*. 2002;137(8):693-695.
16. Cepeda MS, Boston R, Farrar JT, Strom BL. Comparison of logistic regression versus propensity score when the number of events is low and there are multiple confounders. *Am J Epidemiol*. 2003;158(3):280-287.

INVITED CRITIQUE

Are Surgeons Ready to Embrace a Paradigm Shift in Surgical Comparative Effectiveness Research?

Congratulations to Dr Hemmila and colleagues on another strong article in the arena of surgical outcomes research using novel, cutting-edge, advanced biostatistical methods. The authors are among a new breed of surgical scientists who are changing the face of surgical investigation and resident training. Cutting-edge surgical research has historically been based in basic science laboratories using bench assays, cells, and animal surgery as the standard fare. Only more recently has the world of surgical outcomes research begun to flourish. It would have been an absurd request for me to spend my research time during residency doing "outcomes research" and taking epidemiology or biostatistics classes 15 years ago. However, times are changing, and surgical educators see the benefit in training residents (and faculty) for the research they will eventually perform. I am not the only surgical faculty member who has not performed a "laboratory experiment" or written a basic science article since my research years during residency. We do clinical research; we should follow the lead of these authors and invest the time and energy needed to learn the important tools of the trade.

This project begins with an important clinical question: "Which is better: open or laparoscopic appendec-

tomy?" It then systematically shows the reader the steps taken to answer this question using a variety of biostatistical techniques. Along the way, propensity scoring is explained in enough detail to understand the basic concepts, but not in enough depth for casual readers to perform their own similar analyses. In the typical learning approach to surgery, we have now "seen one" but I doubt that many readers are ready to "do one" or "teach one" quite yet.

As with every statistical technique, if any key assumptions are violated, the results will be biased. I raise the following questions about these critical assumptions for this propensity score-based analysis. First, is there adequate overlap between the groups? What happened to data from patients who were unable to be matched? If their data were ignored, does the study truly reflect the entire population of patients undergoing appendectomy? Second, are there any remaining residual confounders? If so, propensity score matching may actually make the situation worse because "eliminating bias (or imbalance) due to one confounder may awaken and unleash bias due to dormant, unmeasured confounders."¹

The final questions I pose are more philosophical. What should be done when different (yet equally valid)

biostatistical approaches to the same underlying research question using the same data yield different results? What if your propensity scoring approach and “regular” multivariable methods gave diametrically opposed conclusions? Which would have been reported? Which would you want us to believe as “the truth”? If the results are the same (as is the case in the present study), is there any benefit to using this more complex technique as opposed to the standard multivariable regression that most surgeons have come to know and understand?

Elliott R. Haut, MD

Author Affiliation: Department of Surgery, The Johns Hopkins School of Medicine, Baltimore, Maryland.

Correspondence: Dr Haut, Department of Surgery, The Johns Hopkins School of Medicine, 600 N Wolfe St, Baltimore, MD 21287 (ehaut1@jhmi.edu).

Financial Disclosure: None reported.

Funding/Support: Dr Haut is supported by a Mentored Clinical Scientist Development Award (K08 1K08HS017952-01) from the Agency for Healthcare Research and Quality titled “Does Screening Variability Make DVT an Unreliable Quality Measure of Trauma Care?” and is a PhD candidate in the Graduate Training Program in Clinical Investigation at The Johns Hopkins University Bloomberg School of Public Health.

1. Pearl J. The foundations of causal inference. *Sociol Methodol*. In press. http://ftp.cs.ucla.edu/pub/stat_ser/r355.pdf. Accessed July 8, 2010.

Call for Papers

Health Care Reform and Comparative Effectiveness: Implications for Surgeons

In 2010, we are soliciting manuscripts on the topic Health Care Reform and Comparative Effectiveness: Implications for Surgeons. The editorial board has selected this topic as our year-long theme. We are interested in papers that compare different interventions and strategies to treat surgical conditions. These interventions could be randomized, preferably prospective, and have adequate follow-up so that patients and other clinicians can use the information to make decisions on care. Registries, health data sets, and meta-analyses would be excellent sources for articles. Accepted manuscripts will be given priority publication. Please indicate in your cover letter that your manuscript is to be considered for this theme.

We have seen a change in the manuscripts submitted to *Archives* over the past year that has led to our increased impact factor as more surgeons and physicians cite our studies in their articles. Now we would like to increase our impact factor in patient care by publishing articles that focus on comparative effectiveness.