VALIDITY ISSUES HARM STUDY FUNDED BY NURSE ANESTHETISTS

EXECUTIVE SUMMARY

In the 2010 article "No Harm Found When Nurse Anesthetists Work Without Supervision By Physicians" (the "Study"), the researchers defined treatment as the anesthesia staffing model – the type of provider(s) involved in the anesthesia care; and examined two outcomes related to health of patients after surgery: mortality and complications.[2] However, drawing valid conclusions based on a statistical analysis of treatment (e.g., anesthesia staffing model) and outcome (e.g., health of the patient after surgery) is more complicated than one might think. First, it requires reliable measures – that is, accurate and relevant data must be used as input to the analysis. Second, it requires a statistical model that uses correct and complete information to define the complex relationships among treatment, outcome and the many other factors that impact outcome.

The Study suffers from several critical deficiencies by using poor or unreliable measures and basing the analysis on a very weak statistical model of these complex relationships.

- First, the main treatment measure used to identify nurse anesthetist-only staffing is ambiguous, at best, and based on a flawed mechanism to identify the actual provider(s) of care.
- Second, the use of "all-cause mortality" (death occurring for any reason) as an outcome measure likely masked any relationship between staffing model and outcome because of the overwhelming influence of other factors on patient mortality not included in the statistical model used.
- Third, the statistical model lacked appropriate risk adjustment methods to appropriately capture the health status of patients.

These deficiencies are substantial enough to impair one's ability to draw conclusions from the Study. Specifically, the Study's inference that no harm is found when anesthesia nurses work unsupervised has essentially no validity.

In sum, this study should not be used as the basis for legislative or regulatory decision-making.

INTRODUCTION

The concept of validity in research refers to the approximate truth of an inference, interpretation, or conclusion. *Statistical conclusion* validity concerns the truth of inferences about the correlation, or association, between treatment and outcome.[1] In the 2010 article "No Harm Found When Nurse Anesthetists Work Without Supervision By Physicians" (the "Study"), the researchers defined treatment as the anesthesia staffing model and examined two outcomes related to health of the patients after surgery: mortality and complications.[2]

Threats to validity are specific reasons why inferences or conclusions can be partly or completely wrong. Two major threats to the validity of the conclusions drawn in this Study funded by the American Association of Nurse Anesthetists are the **unreliability of measures** and **weak statistical model**. The authors are experienced health services researchers; but without relevant and reliable measures and lacking a model that closely reflects the reality of the complex relationships between outcome and the myriad factors affecting that outcome, demonstrating a statistically significant relationship between treatment and outcome is nearly impossible.

Such unreliable or inaccurate data and the use of weak statistical models render a study's findings and conclusions essentially irrelevant. Reading beyond the "Study Data and Methods" section of an article's text becomes unnecessary. However, since policymakers have been directed to this Study as alleged evidence to support potential legislative changes, this current assessment of the validity of the Study's inferences and conclusions is warranted.

USE OF LIMITED AND UNRELIABLE MEASURES

Reliability is the confidence that a measurement or data element actually reflects what is being measured and that it is "stable" over time and across researchers who use the measure. Although Medicare claims data are often used in health services research, there are clear limitations to their use, including the low reliability of certain measures derived from the claims. These limitations are especially evident when using such administrative data to develop clinical outcome and risk adjustment measures.[3,4]

To identify the treatment (i.e., anesthesia staffing model), the Study uses a measure based on the Medicare Carrier (Part B) claims procedure modifier fields. The Study assumes the QZ modifier (CRNA without medical direction by a physician) represents nurse anesthetist-only staffing. Unfortunately, this assumption is flawed. Many practices use the QZ modifier when anesthesiologists are supervising the cases. As Byrd and colleagues note, "the QZ modifier has been twisted and contorted from its intended purpose and has become a 'catch-all' modifier for some practicing in the care team model."[5] Therefore, the procedure modifier field represents an unreliable measure; and any inference or conclusion regarding the relationship between the treatment and the outcome has extremely low validity.

All-cause mortality is one of only two outcome measures used in the Study. Based on the Institute of Medicine's estimates of the incidence of anesthesia-related deaths[6] and the mortality rates presented in the Study, surgical mortality is 5,000 to 7,000 times more likely to be related to something <u>other than anesthesia</u>. Distinguishing anesthesia-related mortality requires a well-defined statistical model and substantially more observations than provided in the Study. Simply put, claims-based, all-cause mortality is an unreliable measure of anesthesia quality. Any inference regarding the relationship between treatment (QZ modifier) and outcome (all-cause mortality) – both unreliable measures – has extremely low validity.

The authors recognize the higher complexity of cases directed by anesthesiologists and they adjusted mortality rates "by applying to the anesthesiologist solo group the nurse anesthetist case-mix for surgeries that the two providers had in common."[15] Yet, the nurse anesthetist-only staffing model is rarely, if ever, used for many complex procedures (e.g., organ transplants) so such an adjustment of "common" procedures would not be possible. Notably, the substantially higher incidence of anesthesiology-led care for complex cases is a recognition that "treatment" choice does matter.

In addition to the unreliability of the measures for both treatment and one of two outcome indicators, a major threat to the Study's *statistical conclusion validity* is the use of claims data for risk adjustment and measures of case complexity or clinical difficulty. The lack of clinical data is problematic and numerous studies attest to the unreliability of claims data for purposes of risk adjustment.[3,4,7-11] The Study's authors use anesthesia base units as the primary measure of procedure complexity and risk, stating "we believe that using additional measures of complexity would not qualitatively change our results."[12] Of course, whether the use of appropriate additional measures of complexity would impact the findings is an empirical question rather than simply a matter of belief. In addition, patients undergoing procedures with the same number of anesthesia base units can vary substantially in clinical risk. The use of age, race, and sex to adjust for differences in anesthesia-related risk is simply inadequate, even when limited to claims data.

In the published study *Anesthesiologist Direction and Patient Outcomes*, Silber and colleagues also used Medicare claims data.[13] However, these researchers used numerous variables to measure risk; and adjusted for severity of illness and case mix based on 64 patient characteristics and 42 DRG-principal procedure variables. They also adjusted for MedisGroups (MediQual Inc., Westborough, MA) severity scores and 11 hospital characteristics in additional analyses.[14] This rigorous approach enhanced the reliability of the measures and strengthened the validity of the inferences drawn, finding higher mortality and failure-to-rescue rates for surgical patients without medical direction by an anesthesiologist.[13]

USE OF A WEAK STATISTICAL MODEL

The Study presents the results of multivariate logistic regression models predicting mortality and complications. However, omitted from the Study are critical model fit statistics which assess how well these models actually predict the outcome measures.[16-18] A poor fitting model will likely have treatment effect estimates with substantial error (variance), *even if* the measures used to identify treatment and outcome were reliable, which they are not.

As with many analyses limited to claims data, there are likely numerous omitted and unobserved but important clinical confounders. Confounders are simply factors and variables other than the treatment (i.e., staffing model) that impact outcome; and some of these also may be related to the treatment. Examples include the characteristics of the surgical team; the specific equipment used; the processes and activities occurring to keep the patient safe; the number and types of patient comorbidities (other disease processes present); measures of obesity, frailty, and other health status indicators; heart, liver, and kidney function; and relevant clinical outcomes. Omitting these variables reduces the ability to detect differences resulting from differences in the anesthesia model of care, *even if* a reliable measure of anesthesia staffing were used, which it was not.

With unreliable measures and a weak statistical model, any inferences or conclusions regarding the association between treatment and outcome will have low or no validity. Even with reliable and complete data and a well-defined statistical model, we know from principles of hypothesis testing that *not* finding a significant correlation between treatment and outcome does not allow one to draw conclusions about the likelihood that the null hypothesis (i.e., the hypothesis that there is no difference in outcome based on treatment) is true. [19-20] This Study exhibits neither high reliability measures or a well-defined statistical model, and its predictive power (i.e., model fit statistic) is unreported; therefore, drawing any inferences is essentially pointless.

POLICY POINTS

Research efforts examining patient safety and quality and the models of care in anesthesia should include the development of relevant primary data and meaningful clinical data for risk adjustment. Published outcomes-based studies without appropriate clinical measures should recognize the various limitations of claims-based data. Analyses of claims data should employ current best practices in risk adjustment and statistical modeling to mitigate problems associated with unreliable or incomplete data. As recommended by Fleisher and colleagues, the use of claims data are appropriate for exploratory analysis (hypothesis generation), but should not be used for confirmatory analysis (hypothesis testing).[21]

It would be a considerable understatement simply to urge a cautious interpretation of the results of this Study which are based on unreliable measures from administrative data and a very weak statistical model. The Study's inferences have low or, perhaps, no statistical conclusion validity; and therefore no conclusions can be drawn from the misguided inference that there is "no correlation between treatment and outcome."

Simply stated, any policy decision based on this Study would be without merit.

CONTRIBUTING AUTHOR:
Thomas R. Miller, Ph.D., M.B.A.
ASA Director of Health Policy Research

REFERENCES

1. Shadish WR, Cook TD, Campbell: Experimental and quasi-experimental designs for generalized causal inference. Boston, Houghton Mifflin Company, 2002, 38

2. Dulisse B, Cromwell J: No harm found when nurse anesthetists work without supervision by physicians. Health Affairs 2010; 29:1469-75

3. Glance LG, Neuman M, Martinez EA, Pauker KY, Dutton RP: Performance measurement at a "Tipping Point". Anesthesia & Analgesia 2011; 112(4): 958-66

4. Iezzoni, LI: Risk adjustment for measuring health care outcomes, 4th edition. Chicago, Health Administration Press, 2013

5. Byrd JR, Merrick SK, Stead SW: Billing of anesthesia service and the QZ modifier: A lurking problem. ASA Newsletter 2011; 75(6):37

6. Committee on Quality of Healthcare in America, Institute of Medicine: To err is human, building a safer health system. Washington, National Academy Press, 1999, 241

7. Fry DE, Pine MB, Jordan HS, Hoaglin DC, Jones B, Meimban R: The hazards of using administrative data to measure surgical quality. The American Surgeon 2006; 72(11):1031-37

8. Hawker GA, Coyte PC, Wright JG, Paul JE, Bombardier C: Accuracy of administrative data for assessing outcomes after knee replacement surgery. J Clin Epidemiol 1997; 50(3):265-73

9. Jang WM, Park JH, Park JH, Oh JH, Kim Y: Improving the performance of risk-adjusted mortality modeling for colorectal cancer surgery by combining claims data and clinical data. J Prev Med Public Health 2013; 46:74-81

10. Nathan H, Pawlik TM: Limitations of claims and registry data in surgical oncology research. Annals of Surgical Oncology 2008; 15(2):415-423

11. Pine M, Jordan HS, Elixhauser A, Fry DE, Hoaglin DC, Jones B, Meimban R, Warner D, Gonzales J: Enhancement of claims data to improve risk adjustment of hospital mortality. JAMA 2007; 297(1):71-76

12. Dulisse B, Cromwell J: No harm found when nurse anesthetists work without supervision by physicians. Health Affairs 2010; 29:1474

13. Silber JH, Kennedy SK, Even-Shoshan O, Chen W, Koziol, LF, Showan AM, Longnecker DC: Anesthesiologist direction and patient outcomes. Anesthesiology 2000; 93(1):152-63

14. Silber JH, Kennedy SK, Even-Shoshan O, Chen W, Koziol, LF, Showan AM, Longnecker DC: Anesthesiologist direction and patient outcomes. Anesthesiology 2000; 93(1):157

15. Dulisse B, Cromwell J: No harm found when nurse anesthetists work without supervision by physicians. Health Affairs 2010; 29:1471

16. Austin PC, Steyeberg EW: Interpreting the concordance statistic of a logistic regression model: relation to the variance and odds ratio of a continuous explanatory variable. BMC Medical Research Methodology 2012; 12:82

17. Hanley JA, McNeil BJ: The meaning and use of the area under a Receiver Operating Characteristic (ROC) curve. Radiology 1982; 143(1):29-36

18. Hosmer DW, Lemeshow S: Applied logistic regression. New York, Wiley, 2000

19. Nickerson RS: Null hypothesis significance testing: A review of an old and continuing controversy. Psychological Methods 2000; 5:241-301

20. Verdam MGE, Oort, FJ, Sprangers MAG: Significance, truth and proof of p values: reminders about common misconceptions regarding null hypothesis significance testing. Quality of Life Research 2013; May 23

21. Fleisher L, Anderson G: Perioperative risk: How can we study the influence of provider characteristics? Anesthesiology 2002; 96(5):1039-41

POSTAGE