

Do Not Use Hierarchical Logistic Regression Models with Low-incidence Outcome Data to Compare Anesthesiologists in Your Department

Franklin Dexter, M.D., Ph.D., Bradley J. Hindman, M.D.

IN this issue of *ANESTHESIOLOGY*, Glance *et al.*¹ compare statistical methods for risk-adjusted comparisons among providers (*e.g.*, hospitals and anesthesiologists). They present their findings in the context of hospital *versus* “physician-based measures for Merit-Based Incentive Payment.”¹ There are multiple reasons to evaluate the performance of hospitals and their anesthesia departments as single teams.² Glance *et al.*¹ summarize the policy options well. In this editorial, we consider the implications of the article for evaluating individual anesthesiologists.

Individuals are hired, are credentialed by hospitals, and are promoted. Consequently, reasonably, there are multiple *requirements* from accreditation agencies (*e.g.*, The Joint Commission, Oak Brook, Illinois) and corporations (*e.g.*, universities) to evaluate *individual* anesthesiologists’ clinical performance.

When comparing low-incidence binary data (*e.g.*, patient mortality) among anesthesiologists, one must (1) know patient conditions (risk factors) upon admission, (2) adjust for those risks statistically, and (3) compare among anesthesiologists using hierarchical modeling.^{1,3,4} Unless risk-adjusted hierarchical modeling is used, the chance of falsely detecting anesthesiologists as having below-average performance can be greater than 50% (*i.e.*, worse than flipping a coin).¹

The results of the study by Glance *et al.*¹ are convincing because their findings are (reasonably) biased toward *underestimating* false discovery rates (*i.e.*, incorrectly reporting average anesthesiologists as low performers). First, their simulations assume that the risk adjustment model and the data collected are both perfect, which, of course, is untrue with real (clinical) data. Second, all providers are assumed to have performed the same numbers of cases, which, again, will be untrue. With imbalance in case numbers, the 95% CIs



“Unless [appropriate] modeling is used, the chance of falsely detecting anesthesiologists as [being below average] can be greater than 50%...”

calculated by the authors would be less accurate (*e.g.*, greater false discovery rates).⁵

Collecting patient risk factor data and performing hierarchical logistic regression modeling take substantial resources (*e.g.*, analysts).⁶ The expertise for this *versus* Student’s *t* test is analogous to comparing anesthesia expertise for cardiac surgery *versus* diagnostic colonoscopy. Yet, if your department reports low-incidence adverse events (*e.g.*, less than or equal to the 2.7% incidence simulated by Glance *et al.*¹) by an anesthesiologist, the results show that your department should use risk-adjusted hierarchical logistic regression modeling.^{1,7}

In our opinion, hiring analysts for this purpose is not worthwhile. Suppose your department accepts a false discovery rate (see Glance *et al.*¹) of approximately 5%. Then, even with unrealistically large $n = 1,000$ patients per anesthesiologist per evaluation period for an endpoint, Glance *et al.*¹ show that there is only a 14.2% sensitivity to detect anesthesiologists with 50% greater than average rates of adverse outcomes. Thus, even for the highest risk procedures (*e.g.*, cardiac surgery), typically a small proportion of the total anesthesia caseload, poorly performing anesthesiologists cannot reliably be identified.^{1,4,8} The reason is that serious adverse events are simply too infrequent for accurate comparisons of individual anesthesiologists. As Glance *et al.*¹ recommend, public reporting and merit-based payment should be by *hospital*.

Comparing individual anesthesiologists based on clinical performance measures that occur more frequently also has been fruitless.^{9–12} For example, pain upon arrival in the postanesthesia care unit needs to be risk adjusted for factors often not known accurately (*e.g.*, the specific postanesthesia care unit nurse obtaining the pain score and patient chronic opioid use).⁹ When the risk adjustments are made, differences

Image: John Ursino, ImagePower Productions.

Corresponding article on page 1092.

Accepted for publication May 27, 2016. From the Division of Management Consulting (F.D.), Department of Anesthesia (F.D., B.J.H.), University of Iowa, Iowa City, Iowa.

Copyright © 2016, the American Society of Anesthesiologists, Inc. Wolters Kluwer Health, Inc. All Rights Reserved. *Anesthesiology* 2016; 125:1083–4

among anesthesiologists are not detected.⁹ Patient satisfaction with the anesthesiologist lacks face (content) validity because amnesia is a fundamental part of anesthesia.¹⁰ After controlling for relevant covariates including patient waiting from surgical start times, there are not significant differences among individual anesthesiologists.¹¹ Finally, prolonged times to extubation differ substantively among patients but not among anesthesiologists.¹² Consequently, in our opinion, rely on the results of the study by Glance *et al*¹ and previous work.^{7,12} Do not use risk-adjusted hierarchical logistic regression models with low-incidence clinical outcomes and performance measures for comparing individual anesthesiologists.

Acknowledgment

The authors thank Ms. Jennifer Espy, B.A. (University of Iowa, Iowa City, Iowa), for assisting with the editing of this manuscript.

Research Support

Supported by funding from the Department of Anesthesia at the University of Iowa, Iowa City, Iowa.

Competing Interests

The authors are not supported by, nor maintain any financial interest in, any commercial activity that may be associated with the topic of this article.

Correspondence

Address correspondence to Dr. Dexter: franklin-dexter@uiowa.edu

References

1. Glance LG, Li Y, Dick AW: Quality of quality measurement: Impact of risk adjustment, hospital volume, and hospital performance. *ANESTHESIOLOGY* 2016; 125:1092–1102
2. Dexter F, Epstein RH: Associated roles of perioperative medical directors and anesthesia: Hospital agreements for operating room management. *Anesth Analg* 2015; 121:1469–78
3. Dalton JE, Glance LG, Mascha EJ, Ehrlinger J, Chamoun N, Sessler DI: Impact of present-on-admission indicators on risk-adjusted hospital mortality measurement. *ANESTHESIOLOGY* 2013; 118:1298–306
4. Glance LG, Hannan EL, Fleisher LA, Eaton MP, Dutton RP, Lustik SJ, Li Y, Dick AW: Feasibility of report cards for measuring anesthesiologist quality for cardiac surgery. *Anesth Analg* 2016; 122:1603–13
5. Gamage J, Mathew T, Weerahandi S: Generalized prediction intervals for BLUPs in mixed models. *J Multivar Anal* 2013; 120:226–33
6. Dexter F, Wachtel RE, Todd MM, Hindman BJ: The “Fourth Mission”: The time commitment of anesthesiology faculty for management is comparable to their time commitments to education, research, and indirect patient care. *A A Case Rep* 2015; 5:206–11
7. Bayman EO, Dexter F, Todd MM: Assessing and comparing anesthesiologists’ performance on mandated metrics using a Bayesian approach. *ANESTHESIOLOGY* 2015; 123:101–15
8. Hyder JA, Niconchuk J, Glance LG, Neuman MD, Cima RR, Dutton RP, Nguyen LL, Fleisher LA, Bader AM: What can the national quality forum tell us about performance measurement in anesthesiology? *Anesth Analg* 2015; 120:440–8
9. Wanderer JP, Shi Y, Schildcrout JS, Ehrenfeld JM, Epstein RH: Supervising anesthesiologists cannot be effectively compared according to their patients’ postanesthesia care unit admission pain scores. *Anesth Analg* 2015; 120:923–32
10. Chen Y, Cai A, Dexter F, Pryor KO, Jacobsohn EM, Glick DB, Willingham MD, Escallier K, Winter A, Avidan MS: Amnesia of the operating room in the B-Unaware and BAG-RECALL clinical trials. *Anesth Analg* 2016; 122:1158–68
11. Kynes JM, Schildcrout JS, Hickson GB, Pichert JW, Han X, Ehrenfeld JM, Westlake MW, Catron T, Jacques PS: An analysis of risk factors for patient complaints about ambulatory anesthesiology care. *Anesth Analg* 2013; 116:1325–32
12. Bayman EO, Dexter F, Todd MM: Prolonged operative time to extubation is not a useful metric for comparing the performance of individual anesthesia providers. *ANESTHESIOLOGY* 2016; 124:322–38