CrossMark

IN REPLY

In reply: Clinical supervision: what does it mean to be better?

Franklin Dexter, MD, PhD (iD) · **Johannes Ledolter, PhD** · **Bradley J. Hindman, MD**

To the Editor,

We would like to address Dr. Ramsay's comment on our recent report showing that mixed effects logistic regression can be used for biannual evaluation of anesthesiologists' supervision while adjusting for the leniency of the raters.[1] Dr. Ramsay's letter questions the validity of our analyses based on a hypothetical scenario of an anesthesiologist with substance abuse achieving consistently high scores punctuated by very low scores.[2]

As stated in Table 2.13,[1] we have indeed observed a few anesthesiologists who had very low scores in succession.[3] To detect such patterns quickly – as would be desired if there were impairment from whatever cause – we nightly perform Bernoulli Cumulative Sum (CUSUM) monitoring (the mathematics of this monitoring was described in our *Anesthesia and Analgesia* Statistical Grand Rounds article.[3] Low score detection occurs within 50 ±14 (median ± quartile deviation) days.[3] Thus, contrary to the letter, low scores are detected (using Bernoulli

CUSUM) far before biannual ongoing professional practice evaluation analyses are performed using mixed effects logistic regression.[1,4]

As stated in Appendix 3, there is significant positive correlation between mean scores and percentages of scores equal to the maximum (Kendall's τ_b +0.36, P <0.001).[1] This is evidence of concurrent validity and the opposite of Dr. Ramsay's hypothetical scenario.

As stated in Table 2.2,[1] questions' responses are appropriately highly correlated (Cronbach alpha 0.948 ± 0.001 standard error), which is shown also in Appendix 3 and the first paragraph of the results.[5] Individual question responses do not create many different and commonly observed scores (contrary to the letter).[5]

Thus, the scenario and examples considered in Dr. Ramsay's letter are unrealistic. In addition, the following addresses the implication that the comparisons made with our described analyses are inaccurate.

As shown in Figures 3-6 and Appendices 7 and 8, valid comparison to the mixed-effects logistic regression's inference is each anesthesiologist's mean score equally weighting each rater – not the mean pooled score, as used in the scenario.[1,6] We previously showed that the two are different because of the inequality of the variabilities of scores among raters ($P < 0.001$).[6]

As shown in Figures 3 and 4 and Appendix 5, the mean scores cannot reliably be compared among anesthesiologists, once the anesthesiologists have received feedback and learned how to provide better supervision, unless adjustment is made for the leniency of raters.[1] Thus, the letter's use of mean scores is unreliable in practice.[1] As shown in Appendix 2, models of mean scores adjusted for rater leniency violate the basic statistical assumptions of the inference (e.g., normal distributions).[1]

F. Dexter, MD, PhD (✉)
Division of Management Consulting, Department of Anesthesia, University of Iowa, Iowa City, IA, USA
e-mail: Franklin-Dexter@UIowa.edu
URL: http://www.FranklinDexter.net

J. Ledolter, PhD
Henry B. Tippie College of Business, University of Iowa, Iowa City, IA, USA

J. Ledolter, PhD
College of Liberal Arts and Sciences, University of Iowa, Iowa City, IA, USA

B. J. Hindman, MD
Department of Anesthesia, University of Iowa, Iowa City, IA, USA

Consequently, and in noted contrast to our approach, even though use of mean scores may be desired, Dr. Ramsay's letter does not suggest how they can be used reliably or with any validity.

Conflicts of interest None declared.

Editorial responsibility This submission was handled by Dr. Hilary P. Grocott, Editor-in-Chief, *Canadian Journal of Anesthesia.*

Funding Departmental.

References

1. *Dexter F*, *Ledolter J*, *Hindman BJ*. Measurement of faculty anesthesiologists' quality of clinical supervision has greater reliability when controlling for the leniency of the rating anesthesia resident: a retrospective cohort study. Can J Anesth 2017; 64: 643-55.
2. *Ramsay T*. Clinical supervision: what does it mean to be better? Can J Anesth 2017; 64: DOI:10.1007/s12630-017-0961-6.
3. *Dexter F*, *Ledolter J*, *Hindman BJ*. Bernoulli cumulative sum (CUSUM) control charts for monitoring of anesthesiologists' performance in supervising anesthesia residents and nurse anesthetists. Anesth Analg 2014; 119: 679-85.
4. *Dexter F*, *Szeluga D*, *Hindman BJ*. Content analysis of resident evaluations of faculty anesthesiologists: supervision encompasses some attributes of the professionalism core competency. Can J Anesth 2017; 64: 506-12.
5. *Dexter F*, *Szeluga D*, *Masursky D*, *Hindman BJ*. Written comments made by anesthesia residents when providing below average scores for the supervision provided by the faculty anesthesiologist. Anesth Analg 2016; 122: 2000-6.
6. *Dexter F*, *Ledolter J*, *Smith TC*, *Griffiths D*, *Hindman BJ*. Influence of provider type (nurse anesthetist or resident physician), staff assignments, and other covariates on daily evaluations of anesthesiologists' quality of supervision. Anesth Analg 2014; 119: 670-8.