
(Accepted for publication January 24, 2012.)

In Reply:
We thank Watkinson and Tarassenko for the interest they have taken in our review on patient surveillance.1

We are in agreement that in their cited work, they tested mandated five-channel physiologic monitoring versus standard care,2 and that extra monitoring with Biosign™ (OBS Medical Ltd., Abingdon, Oxon, United Kingdom) had no effect on adverse event rates or mortality, which we had summarized as patient outcome. As we pointed out in our review, the study of measuring the impact of patient surveillance or continuous monitoring is challenging.

In addition, we also agree, as described in our review, that deterioration detection must move beyond the use of just static alarm threshold to optimize the balance between “true” and nuisance alarms. In their mentioned recent study of their Biosign™ monitor in a step-down unit,3 use of the device decreased the authors’ cardiorespiratory instability criteria without changing the rate of medical emergency team (MET) activations. In our own work,4 we consider a reduction of MET calls a success, a sign of an intervention that happened at an earlier stage of deterioration and therefore prevented a physiologic derangement that necessitated a MET activation. But, if on the other hand, monitoring is used to trigger MET alerts, an increase in activations may be a desirable outcome.

At Dartmouth, all medical and surgical patients are now continuously monitored; we have a physiologic database containing more than 3 million hours of patient physiologic data and more than 20 trillion individual data points of inpatient oxygen saturations and heart rates. Like the Oxford research group, we have seen little variation across patient groups. What we have seen is a hospital-wide reduction of MET alerts and transfers to the intensive care unit: between 10–67% depending on the unit.

The research done at Oxford, Pittsburgh, and other locations is outstanding: more pieces are being added to solve the puzzle of patient surveillance, and our patients benefit.

Andreas H. Taenzer, M.D., M.S., F.A.A.P.,* Susan McGrath, Ph.D. *Dartmouth Medical School, Dartmouth-Hitchcock Medical Center, Lebanon, New Hampshire. andreas.taenzer@dartmouth.edu

References

(Accepted for publication January 24, 2012.)

Are Faculties Another Brick in the Wall?

To the Editor:
The investigation in education by Baker1 is an excellent source of information for practicing academic anesthesiologists involved in resident education. Despite efforts to normalize evaluations and potentially control for bias, concerns remain about the structure of this or any evaluation system because of unavoidable introduction of faculty bias. Baker’s results show that none of the residents received a score less than 3 and evaluations consistently increased with progression through CA-1 to CA-3 yr. Although acknowledging the finding, does Baker have any explanation for this, or whether this may be related to the faculty cohort here or for all faculty? In addition to unintentional penalty for being less experienced as a CA-1, the more senior the resident, the longer the time that a particular faculty has spent with a resident. A longer professional relationship will likely lead to greater confidence in assigning a higher evaluation grade, especially if residents with lower scores have been removed from the peer group by attrition.

Of greater significance, is the faculty really free not to be biased? When faculty are aware that their resident evaluations are accessible to the resident, acknowledging that their own teaching evaluations may subsequently be affected in a retaliatory fashion, a positive bias is expected.

As doubts remain whether faculty are assigning unbiased scores of resident evaluations, more information is required from academic educators, including how well residents and faculty are matched in terms of time and cases done together to allow accurate assessment of performance.

Davide Cattano, M.D., Ph.D., The University of Texas, Medical School at Houston, Houston, Texas. davide.cattano@uth.tmc.edu

Dr. Cattano is a paid consultant for Smith Medical (Dublin, Ohio), and receives grant support from Covidien (Mansfield, Massachusetts) and Karl Storz Endoskope (Tuttlingen, Germany), and research supplies from Haemonetics Corporation (Braintree, Massachusetts).
Likert or Not, We Are Biased

To the Editor:
I read with interest the recent article by Baker regarding the value of normalizing resident evaluation scores to eliminate individual faculty evaluator bias. Without unduly underestimating the importance of this study, I have concern about the statistical handling of Likert scores. Likert scores were used to create individual faculty member mean scores, faculty score standard deviations, and average resident scores when more than one core competency section was included. The central issue is that Likert scales involve ordinal data, or categories falling in a hierarchy. Because the numbers in a Likert scale represent verbal statements of rank order (eg, 5 = distinctly above peer level), summarizing such ordinal data with a mean value is inappropriate by strict statistical methodology. Moreover, the intervals between data points on a Likert scale are not necessarily equal or even certain. To put this in the context of the study, consider this example from the relative performance designation used in the study: a score of “4” is “somewhat above peer level” and a score of “5” is “distinctly above peer level”; however, an average score of “4.5” cannot be said to represent “somewhat-above-peer-level-and-a-half.” Similarly, on the absolute/anchored competency designation, the difference between a score of “5” (performed in a fully independent manner) and a score of “6” (able to serve as a consultant to other physicians) is not necessarily equivalent to the difference between a score of “2” (needed moderate assistance) and a score of “3” (needed only minimal assistance). It is difficult to determine what, if any, limitation was imposed on the study as a result of this violation of statistical propriety. Nevertheless, although a purist may pine for cleaner data and analysis, this distraction can be mitigated by considering what Stevens wrote in 1946: “for this ’illegal’ statisticizing there can be invoked a kind of pragmatic sanction: In numerous instances it leads to fruitful results.”

I look forward to future contributions from Baker. When I was a fellow his efforts sparked my interest in resident education and continue to do so now.

Nicholas C. Watson, M.D., UMass Memorial Medical Center, University of Massachusetts Medical School, Worcester, Massachusetts. nicholas.watson1@gmail.com

References

Errors in Assessment of Resident Performance

To the Editor:
In a recent innovative study, Baker used relative Z scores (Zrel) to correct for observer bias in the assessment of 108 anesthesiology residents. We have concerns about the statistical methodology used in this study and believe there is a need for caution before his approach is widely adopted.

Baker distinguishes three groups: those “reliably above average,” “reliably below average,” and “not reliably different from average.” His criterion for identifying a resident who is above average is that 1.96 times the SEM for that individual’s Z score (a 95% CI for the SEM) does not overlap with zero. A similar criterion is used to identify “below average” residents. This approach is problematic.

Although Baker identifies 30% of residents as “reliably below average,” with sufficient assessments, 50% would be “reliably below average” because the width of the CIs would decrease. It is trivially true that, as long as the distribution is symmetric, 50% of people are “below average,” but this does not imply that all “below average” residents require what Baker terms “performance interventions.” Baker’s Z scores could be applied to any group of residents, even a sample of entirely competent anesthesiologists, and would still identify a proportion as “below average.” Without a clinically relevant benchmark, Baker’s approach cannot be used to identify anesthetic competence.

In translating an overall assessment of “anesthetic competence” into a Z score, Baker makes certain assumptions. One of these is that the competence of anesthesiologists is an underlying, continuous variable that can be normalized. Although this assumption cannot be validated, it can be simulated using a Monte-Carlo approach. Figure 1 shows the results of a single run of such a simulation. The assumptions are: that each of 100 individuals has intraindividual variation in Zrel scores that is normally distributed, and that the mean score for each individual is offset by a value that is similarly, randomly sampled from a normal distribution (“interindividual variation”), with a known SD (SDadj). As both the generated SD

Supplemental digital content is available for this article. Direct URL citations appear in the printed text and are available in both the HTML and PDF versions of this article. Links to the digital files are provided in the HTML text of this article on the Journal’s Web site (www.anesthesiology.org).