Critical Review Form

Diagnostic Test

Gill M, et al, A Study of the Bispectral Index Monitor During Procedural Sedation and Analgesia in the Emergency Department, *Annals EM* **41:**2;234-241

Objective: "To define, through descriptive analyses if possible, any Bispectral Index Monitor thresholds that might accurately distinguish traditional states of sedation or general anesthesia." (p. 235)

Methods: A prospective, observational study consisting of a convenience sampling from a single ED (Loma Linda, California) of all patients receiving procedural sedation for any indication. "Neither the medications used nor the route (or dose) of administration was controlled, both being selected by the treating physicians in accordance with their standard practices" (p. 235). Exclusion criteria included patients unable to consent for enrollment, unable to tolerate the forehead BIS monitor leads, or those receiving ketamine sedation. The numeric BIS score ranges from 0 to 100 on an absolute unitless scale and was correlated with a modified Ramsay Sedation Scale which has scores ranging from 1 to 8 with the higher score representing more sedate patients.

Guide		Comments
I.	**Are the results valid?**	
A.	**Did clinicians face diagnostic uncertainty?**	Yes. There is no "gold standard" for depth of sedation.
B.	**Was there a blind comparison with an independent gold standard applied similarly to the treatment group and to the control group?**	The comparison was not blinded, nor were there a treatment and control group. Each patient served as their own control for repeated measure at multiple times. IMPORTANT: Each paired data point can't be considered an independent result. There was no accounting for the repeated measures effect (i.e. a patient with disparity between results at one point may be more likely to have a disparity at another point in time. Given that different patients had different numbers of paired data points collected, statistical accounting for this effect is necessary to minimize the effect of undue influence of the data.
C.	**Did the results of the test being evaluated influence the decision to perform the gold standard?**	No. Both tests were applied at each time.
II.	**What are the results?**	
A.	**What likelihood ratios were associated with the range of possible test results?**	Likelihood ratios of failing to be under general anesthesia at BIS>83 was 3.6. At BIS>69 was 18.8.

III.	How can I apply the results to patient care?	
A.	**Will the reproducibility of the test result and its interpretation be satisfactory in my clinical setting?**	No. The variability in distribution of scores compared with the modified Ramsey score makes use of BIS unreliable (assuming the Ramsey is the "gold standard")
B.	**Are the results applicable to the patients in my practice?**	Not well stated.
C.	**Will the results change my management strategy?**	Clinical gestalt and more specifically, the Ramsey score seem to be more reliable indicators of sedation depth.
D.	**Will patients be better off as a result of the test?**	No. The variability in BIS score implies a potential harm to patients. A high BIS in a patient with deep sedation could cause inappropriate over-sedation. A low BIS in an awake patient could result in inadequate procedural sedation.

Bottom Line

This paper was not well received at JC, but possibly unfairly. It was not attempting to be landmark, aggressive or cutting edge, but merely attempted to describe pilot data that showed a poor correlation between BIS and modified Ramsey score. While there are some statistical errors (no repeated measure effect, no calculation of a kappa), it accomplishes that. It's tough enough for an article to stand up to scrutiny, but it is impossible if we tear it apart for not doing what we wanted. My approach? To ask if it would be better not to have this data /study done. In this case, I think the data and study were worth a read, despite the negative result and the flaws.

The technology is conceptually intriguing, but probably not ready for prime time. From the discussion and from Dr. DeWitt's experience, we might be able to use it in paralyzed patients (since clinical exam is unreliable), but using for brief procedures, associated with little clinically significant respiratory depression, isn't indicated, and could potentially cause harm.