

Comparative Effectiveness Research It Is Not

To the Editor:

I take issue with Memtsoudis and Liu's editorial acceptance¹ of the study by Mashour *et al.*² being referred to as comparative effectiveness research (CER), or the actual study being labeled as a practical clinical trial. It is neither. Although the study did indeed enroll a large, "real-world" patient population undergoing various procedures, the hospital setting was specific, and even more importantly, the interventions being compared could hardly represent what the Institute of Medicine expects of alternative treatment methods used in typical daily practice to be evaluated with CER.³ Put simply, the study results are not generalizable.

CER is the direct comparison of existing (that is, commonly used) interventions, aiming to determine which treatment works best for whom and under what circumstances.³ An important ingredient of CER is pragmatic (or practical), effectiveness trials that reflect everyday clinical practice.³⁻⁶ The explicit purpose of pragmatic trials is to inform decision makers about study interventions commonly used in practice and to measure clinically important outcomes in typical populations.^{4,7-11} The design and interpretation of large pragmatic trials has been a major interest of mine for more than 15 yr.^{9,10,12} Pragmatic trials test commonly used interventions in a wide variety of healthcare settings—that is, they should represent "real-world" practice.

The study by Mashour *et al.*² occurred in a single university hospital system (albeit with three hospitals) in one city in the United States which used an electronic perioperative information system in all of its operating rooms. Furthermore, they used automated real-time analysis of bispectral index (BIS) values or minimum alveolar concentration every 5 min and provider-specific electronic alphanumeric paging alerts sent in under 1 min. This does not represent real-world practice in most of the western world (United States included). The most widely used approaches anesthesiologists use to avoid awareness include delivery of an appropriate concentration of hypnotic and other drugs, typically titrated to autonomic signs focusing on blood pressure and heart rate, and to look for patient movement—this is real-world monitoring of anesthetic depth. It is hard to see how the results of the study by Mashour *et al.*² can be applied outside of their specific setting. In contrast, the B-Unaware trial at least tested two broadly accessible, competing interventions.¹³

Memtsoudis and Liu go on to describe the B-Aware trial,¹⁴ for which I was the principal investigator, as "a traditional randomized trial" despite it clearly being a pragmatic, effectiveness trial. Unlike the study by Mashour *et al.*,² the B-Aware trial was conducted in a much broader range of hospital settings, involving 21 hospitals in 5 countries. The decision to use a high-risk population in the B-Aware trial

should not be confused with the bias inherent in the strict selection criteria seen in some trials, which are intended to exclude high-risk patients. Enrolling an at-risk cohort does not infer a lack of generalizability, for it has been repeatedly shown that study results (relative effect) are nearly always in the same direction and of similar magnitude in unselected settings^{8,15}; in fact, a high-risk study population may underestimate the treatment effect.^{16,17} The decision to use a high-risk group adds to study power because of the increased event rate and so it is more efficient; no more, no less. Finally, the view that we did not compare BIS with an alternative strategy to decrease the risk of intraoperative awareness is blatantly wrong. As stated above, the near-universal approach to avoiding intraoperative awareness—in other words, a relevant "competing intervention"³—is traditional monitoring of (predominantly) patient movement, hypnotic agent delivery, and autonomic signs. It is hardly a "no intervention" group. The B-Aware trial, therefore, was a pragmatic trial, and it remains highly relevant to inform CER.

Finally, CER is not only the generation of relevant evidence but also the synthesis of that evidence, with the latter best achieved using systematic review.^{3,6} I therefore provide results of broadly applicable CER evaluating the evidence for BIS monitoring in anesthesia, using an updated systematic review and meta-analysis of the pertinent randomized trials using Revman 5.1 software (Cochrane IMS team; fig. 1). This analysis includes input of the most robust data from the study by Mashour *et al.*,² based on their as-treated data, because the presumptions of an intention-to-treat analysis cannot be claimed because of the gross failure of the intended intervention in 36% of the BIS group.

There are six pertinent trials^{2,13,14,18-20} and one large observational study.²¹ The results of this CER approach depend on the local setting or scenario being considered:

1. The most clinically relevant question is whether the inclusion of BIS monitoring reduces the risk of awareness compared with traditional monitoring that usually includes end-tidal agent monitoring—for which the pooled data show that this significantly reduces the risk of awareness by 79% (95% CI: 21–95%), $P = 0.02$. The number needed to treat in a high-risk setting (incidence 1%) is about 130, and the number needed to treat in a low-risk setting (incidence 0.1%) is about 1,300. These conclusions are consistent with a large observational study in everyday practice²¹ and the *post hoc* findings from Mashour *et al.*,² adding further weight to this finding.
2. If considering the question of whether BIS monitoring provides additional benefit compared with any type of "traditional" monitoring that usually includes end-tidal volatile agent monitoring, and could include having the agent monitoring alarms activated, it is unclear whether BIS monitoring reduces the risk of awareness, pooled risk reduction 39% (95% CI: –16 to 68%), $P = 0.13$.

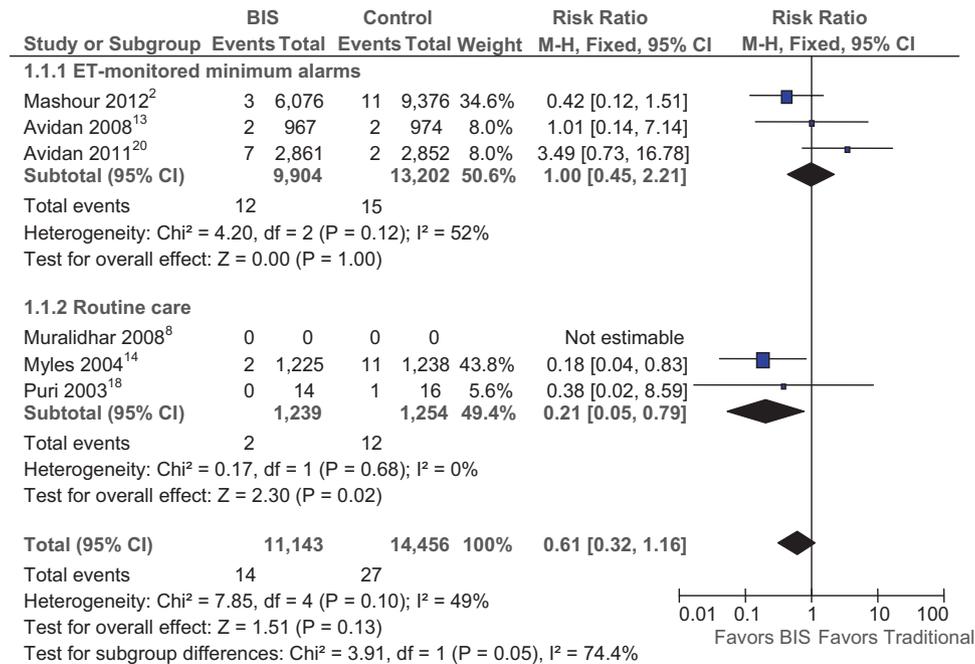


Fig 1. Risk of awareness with bispectral index (BIS) monitoring combined with traditional monitoring versus traditional monitoring alone (“Control”), in which an end-tidal (ET) minimum alveolar concentration can be selected, with or without alarms being set to avoid low levels of volatile agent delivery during anesthesia. The BIS group can include patients undergoing total intravenous anesthesia. Both groups could include an electronic record system and immediate automated paging to alert the anesthesiologist, as used by Mashour *et al.*² The “routine care” subgroup studies did not insist on the settings of ET alarms for a minimum alveolar concentration.

But, in my view, this is an ill-focused and unhelpful clinical question because the non-BIS options of monitoring (that is, the alternative treatment options) will be known to the decision-maker at the time. The inconsistency statistic, I^2 , highlights the heterogeneity when pooling disparate studies.

- When comparing end-tidal agent monitoring with the alarms activated to avoid delivery of less than 0.7 minimum alveolar concentration,^{13,20} it is unclear whether BIS monitoring modifies the risk of awareness, pooled risk increase 225% (95% CI: -31 to 731%), $P = 0.18$. The range of possible risk reduction afforded by BIS monitoring extends from a 31% risk reduction to a 731% risk increase. This nonsignificant finding might lead decision-makers to regard the two monitoring options as equivalent, but such a conclusion is fraught with error.²²
- In settings where an electronic record system incorporating automated alerts is used, it is unclear whether the addition of BIS monitoring reduces the risk of awareness compared with end-tidal agent monitoring, pooled risk reduction 58% (95% CI: -151 to 88%), $P = 0.18$. The range of possible risk reduction afforded by BIS monitoring extends from an 88% risk reduction to a 51% risk increase.

These up-to-date findings can be used by all anesthesia providers to determine whether or not to rely only on traditional monitoring, avoid total intravenous anesthesia because of the inability to use end-tidal volatile agent monitoring, with or

without activation of the volatile agent monitoring alarms, use BIS monitoring, or purchase/configure an electronic monitoring system with automated paging alerts. Scenario (a) would be most relevant to most anesthesia providers in the western world.

Another important component of CER is to consider efficiency: is the problem clinically important and is the intervention worthwhile and cost-effective?^{3,6} Regard for patient values and expectations, and consideration of benefits and harms, acquisition costs, and the number needed to treat should assist in such clinical and policy decision-making.

There have been dozens of high-quality large clinical trials guided by principles of CER in perioperative medicine over the past decade, including in the field of anesthesiology. The shame is that these are often overlooked or otherwise not incorporated in clinical practice more quickly. That is what our patients expect.

Paul S. Myles, M.B.B.S., M.P.H., M.D., F.C.A.R.C.S.I., F.A.N.Z.C.A., F.R.C.A., Alfred Hospital and Monash University, Melbourne, Victoria, Australia. p.myles@alfred.org.au

References

- Memtsoudis SG, Liu SS: Bispectral index *versus* minimum alveolar concentration for prevention of intraoperative awareness: Does a practical controlled trial provide CERTainty? *ANESTHESIOLOGY* 2012; 117:693–5
- Mashour GA, Shanks A, Tremper KK, Kheterpal S, Turner CR, Ramachandran SK, Picton P, Schueller C, Morris M, Vandervest

- JC, Lin N, Avidan MS: Prevention of intraoperative awareness with explicit recall in an unselected surgical population: A randomized comparative effectiveness trial. *ANESTHESIOLOGY* 2012; 117:717–25
3. Sox HC: Defining comparative effectiveness research: The importance of getting it right. *Med Care* 2010; 48:S7–8
 4. Tunis SR, Stryer DB, Clancy CM: Practical clinical trials: Increasing the value of clinical research for decision making in clinical and health policy. *JAMA* 2003; 290:1624–32
 5. Luce BR, Kramer JM, Goodman SN, Connor JT, Tunis S, Whicher D, Schwartz JS: Rethinking randomized clinical trials for comparative effectiveness research: The need for transformational change. *Ann Intern Med* 2009; 151:206–9
 6. Scott IA, Glasziou PP: Improving the effectiveness of clinical medicine: The need for better science. *Med J Aust* 2012; 196:304–8
 7. McPeck B: Inference, generalizability, and a major change in anesthetic practice. *ANESTHESIOLOGY* 1987; 66:723–4
 8. Yusuf S, Collins R, Peto R: Why do we need some large, simple randomized trials? *Stat Med* 1984; 3:409–22
 9. Myles PS: Why we need large randomized studies in anaesthesia. *Br J Anaesth* 1999; 83:833–4
 10. Myles PS: Why we need large trials in anaesthesia and analgesia, An Evidence Based Resource in Anaesthesia and Analgesia. 2nd edition. Edited by Tramer MR, London, BMJ Publishing Group, 2003, pp 12–21
 11. Collins R, MacMahon S: Reliable assessment of the effects of treatment on mortality and major morbidity, I: Clinical trials. *Lancet* 2001; 357:373–80
 12. Myles PS, Rigg JR: Anaesthesia and pain, Textbook of Clinical Trials, 2nd edition. Edited by Machin D, Day S, Green S. Chichester, United Kingdom, John Wiley & Sons Ltd., 2006, pp 519–42
 13. Avidan MS, Zhang L, Burnside BA, Finkel KJ, Searleman AC, Selvidge JA, Saager L, Turner MS, Rao S, Bottros M, Hantler C, Jacobsohn E, Evers AS: Anesthesia awareness and the bispectral index. *N Engl J Med* 2008; 358:1097–108
 14. Myles PS, Leslie K, McNeil J, Forbes A, Chan MT: Bispectral index monitoring to prevent awareness during anaesthesia: The B-Aware randomised controlled trial. *Lancet* 2004; 363:1757–63
 15. Proschan MA, Brittain EH, Fay MP: Does treatment effect depend on control event rate? Revisiting a meta-analysis of suicidality and antidepressant use in children. *Clin Trials* 2010; 7:109–17; discussion 118–20
 16. Weisberg HI, Hayden VC, Pontes VP: Selection criteria and generalizability within the counterfactual framework: Explaining the paradox of antidepressant-induced suicidality? *Clin Trials* 2009; 6:109–18
 17. Rothwell PM: Can overall results of clinical trials be applied to all patients? *Lancet* 1995; 345:1616–9
 18. Puri GD, Murthy SS: Bispectral index monitoring in patients undergoing cardiac surgery under cardiopulmonary bypass. *Eur J Anaesthesiol* 2003; 20:451–6
 19. Muralidhar K, Banakal S, Murthy K, Garg R, Rani GR, Dinesh R: Bispectral index-guided anaesthesia for off-pump coronary artery bypass grafting. *Ann Card Anaesth* 2008; 11:105–10
 20. Avidan MS, Jacobsohn E, Glick D, Burnside BA, Zhang L, Villafranca A, Karl L, Kamal S, Torres B, O'Connor M, Evers AS, Gradwohl S, Lin N, Palanca BJ, Mashour GA; BAG-RECALL Research Group: Prevention of intraoperative awareness in a high-risk surgical population. *N Engl J Med* 2011; 365:591–600
 21. Ekman A, Lindholm ML, Lennmarken C, Sandin R: Reduction in the incidence of awareness using BIS monitoring. *Acta Anaesthesiol Scand* 2004; 48:20–6
 22. Götzsche PC: Lessons from and cautions about noninferiority and equivalence randomized trials. *JAMA* 2006; 295:1172–4

(Accepted for publication January 17, 2013.)

Why Does Bispectral Index Monitoring Not Perform Better?

To the Editor:

We read with great interest the results of the largest Bispectral Index (BIS) monitoring study ever performed, which was published in the October 2012 issue of *ANESTHESIOLOGY*.¹ No significant difference in intraoperative awareness with explicit recall was detected between BIS and anesthetic concentration protocols (0.08 vs. 0.12%, $P = 0.48$) in an unselected surgical population of 21,601 patients. Initial multicenter studies suggested that BIS monitoring could reduce the incidence of explicit recall in high-risk surgical patients,² but later studies that compared BIS monitoring with carefully guided dosing schemes with audible alerts for low concentrations of the anesthetic failed to demonstrate such benefit.^{3,4} Now, this negative result was corroborated in a “normal” population (BIS < 60 vs. minimum alveolar concentration > 0.5). What went wrong? Why does BIS monitoring not perform better?

We believe that there are two main reasons. First, the suggested intraoperative “therapeutic window” (BIS 40–60) to guide anesthetic dosing is not optimal for preventing unintended awareness and is most probably dictated by manufacturer’s aspiration to not prolong awakening after anesthesia. The scientific evidence that BIS should be kept below 60 to prevent awareness is extremely weak if not totally nonexistent. We find it incomprehensible that this fundamental issue is not dealt with in the literature. Every anesthesiologist who has used BIS monitoring knows that BIS level 60 represents a labile “depth of anesthesia,” and even a small surgical or other irritation can lead to arousal and awakening. Deepening anesthesia induces characteristic electroencephalographic changes, and lowering the reference range would undoubtedly improve the sensitivity of BIS to prevent awareness despite the wide interpatient variability in its concentration–response curves and partially distinct electroencephalographic effects of different anesthetic agents. Because of the nonlinear behavior of BIS,⁵ keeping it close to 40 is actually relatively easy.

Our recent positron emission tomography imaging study with anesthetized healthy subjects suggests another reason for the poor performance of BIS. The emergence of consciousness after anesthetic-induced unconsciousness, as assessed with a motor response to a spoken command, was found to be associated with activation of deep, primitive brain structures rather than the evolutionary younger neocortex.⁶ Unexpectedly, activation of these central core structures was enough for the arousal and behavioral expression of subjective awareness. Because BIS is based on cortical electroencephalographic measurement (*i.e.*, measuring electrical signals on the surface of the scalp that arise from the brain’s cortical surface), these results help to understand why BIS fails in differentiating the conscious and unconscious states in the subtle transition phase during emergence⁷ and why patient awareness during general anesthesia may not always be detected.