

## Pros and Cons of Composite Endpoints in Anesthesia Trials

IT is widely accepted that the randomized controlled trial is the optimal method to evaluate the efficacy of an intervention.<sup>1</sup> Clinical research aiming to inform and improve patient care should evaluate outcomes rated as important by our patients.<sup>2–4</sup> Unfortunately, many trials focus on surrogate outcomes that are of questionable significance.<sup>5</sup> Given that 5 to 10% of trial patients will suffer a serious postoperative complication, and that perioperative treatments are likely to have only a moderate effect on outcome, large numbers of patients are required to have sufficient trial power to detect a modest but clinically important effect.<sup>1,2</sup>

A systematic review of head injury trials published before December 1998 identified 203 studies in which 16,613 patients were enrolled.<sup>6</sup> Thus, the average number of patients in each trial was 82. Only 4% of these trials were large enough to detect a clinically important difference. We believe similar problems with underpowered studies exist in the anesthesia, perioperative, and pain medicine literature.<sup>7–9</sup>

Large trials can provide definitive evidence to guide clinical practice.<sup>1–3</sup> They provide insights in a broad range of clinical settings and offer an opportunity to identify patient, clinician, and institutional factors that may influence outcome.<sup>2</sup> They should have sufficient power to address clinically important questions, balance a range of known and unknown confounding factors, and provide precise estimates of effect—the latter being indicated by narrow confidence intervals. However, large multicenter trials take a lot of time and effort to establish and can be expensive to run.<sup>10</sup>

The required sample size needed to reliably demonstrate a clinically important effect should be based on the nominated primary endpoint of the study<sup>11</sup>; this is sometimes lacking in anesthesia trials.<sup>8,9</sup> There may be several secondary study outcomes, and each of these typically relates to mechanisms of action, specific organ effects, and potential adverse effects of therapy. If the primary endpoint of the trial is a serious complication (e.g., death), it may only occur in 1 to 2% of patients, which greatly increases the sample size required to detect a clinically useful effect.<sup>12</sup> For this reason, some investigators combine several outcomes into a single composite, or pooled, endpoint.<sup>13–15</sup> For example, antithrombotic treatments may reduce nonfatal myocardial infarction (MI), nonfatal stroke, and death. Similar approaches have or are being

used in perioperative anesthesia trials.<sup>16–18</sup> A summary measure of each of these effects can be defined. The composite endpoint will have a higher incidence than each of the individual outcomes, and as long as there is a reasonably similar effect across the individual components of the composite, this will reduce the sample size requirement for the trial. Because of the higher event rate, composite endpoints can provide increased statistical precision and efficiency. Clinical trials can be smaller, less costly, completed earlier, and the results made available sooner.

However composite endpoints have an inherent assumption in that each component of the endpoint has a similar burden on health.<sup>13–15,19–21</sup> This is often not the case. Endpoints of least importance to patients, such as a single episode of angina, as opposed to MI-induced heart failure, stroke, or death, typically contribute most to trial events.<sup>15</sup> A recent systematic review found that in approximately half the trials reviewed, there were large or moderate gradients in both importance to patients and magnitude of effect across components.<sup>15</sup> The most serious and highly rated events typically occurred least often, so they provided the least information (in terms of numerical value) to the composite endpoint. This may lead to misleading impressions of the true clinical value of the treatment.<sup>22</sup>

Should endpoints be weighted, and if so, how? We believe this should be guided by the perceived clinical importance of each individual endpoint, and emphasis should be placed on patient values. Objective justification can assist in this regard, using patient and clinician surveys, disability scores, and eventual health care utilization. There may be competing risks between endpoints; a positive composite endpoint may conceal a negative individual (more serious) outcome such as death. The most extreme example of this is when early death prevents the possibility of other less serious outcomes, such as subclinical MI or prolonged hospital stay. Clinicians should be wary of *post hoc* definition of composite endpoints as secondary measures when the primary outcome fails to show a treatment effect. This is one of several reasons why findings from the ENIGMA trial<sup>23</sup> are being followed up with a larger, more specifically designed trial focusing on cardiovascular outcomes.<sup>18</sup>

Cardiovascular trials commonly use major adverse cardiac events as a composite primary endpoint.<sup>14,21</sup> But there is no standard definition for major adverse cardiac events. Most include MI, stroke, and death; others include revasculariza-

Accepted for publication June 2, 2010. 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.

tions, and some include cardiac arrest, heart failure, or bleeding complications. For a technical term to have value, it must have consistency in interpretation and usage. Major adverse cardiac events fail in this regard and should be abandoned because it risks erroneous assumptions. This highlights the need to clearly define each component and justify their inclusion in a composite endpoint. Sometimes a composite endpoint plans to combine efficacy and safety components, and sometimes it is assumed that a similar effect across outcomes will exist, but it does not. Perhaps the best and most recent example is the Perioperative Ischemic Evaluation Study trial,<sup>16</sup> which identified a significant reduction in perioperative MI with  $\beta$ -blocker therapy but at a cost of excess death and stroke. There was also more hypotension and bradycardia in the  $\beta$ -blocker group.

It is unlikely that a common definition of perioperative composite endpoints can be achieved by independent investigators, but we believe that some consensus is needed so that different studies can be compared or, when undertaking meta-analysis, their results pooled. Outcome after surgery and anesthesia is sometimes defined by pain score, functional status, or satisfaction, using scales or instruments as endpoints. Some of these have never been validated in the perioperative setting, and their relationship with true morbid events, such as dementia, stroke, and long-term disability, is unclear.

The decision to stop a trial early because of apparent benefit or harm needs to consider a range of issues, including the statistical uncertainty inherent in a reduced sample size.<sup>24</sup> This is more so if the decision is based on the monitoring of a composite endpoint, particularly if this is driven by a relatively minor yet frequent outcome dominating the composite endpoint.<sup>22</sup> Such an approach may lead to an overestimation of benefit and underestimation of risk. Critical review of the frequency of each of the individual components of the composite endpoint, and whether the treatment effect seems to be consistent for each of these, is strongly recommended.<sup>13–15,19,20</sup> If large variations exist between components, then the value of the composite endpoint is diminished. The same process should be undertaken at the end of the trial when considering the final results.

Composite endpoints are useful in that they provide an overall summary of effect, which may be readily appreciated by clinicians and their patients. When correctly constructed, they enhance comprehension, study power, and precision, and these should lead to earlier identification of real improvements in care. But poorly constructed composites, or insufficient consideration of the rates of each of the individual endpoints making up the composite, can be misleading. Better appreciation of these issues will lead to improved clinical research, its interpretation, and implementation of evidence-based care.

**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.,\* Philip J. Devereaux, M.D., Ph.D.†** \*Department of Anaesthesia and Perioperative Medicine, Alfred Hospital, Academic Board of Anaesthe-

sia and Perioperative Medicine, Monash University, Melbourne, Victoria, Australia. p.myles@alfred.org.au. †Departments of Medicine and Clinical Epidemiology & Biostatistics, McMaster University, Hamilton, Ontario, Canada.

## References

- Collins R, MacMahon S: Reliable assessment of the effects of treatment on mortality and major morbidity, I: Clinical trials. *Lancet* 2001; 357:373–80
- 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
- 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
- Guyatt GH, Montori V, Devereaux PJ, Schunemann H, Bhandari M: Patients at the center: In our practice, and in our use of language (editorial). *ACP J Club* 2004; 140: A11–2
- Fisher DM: Surrogate outcomes: Meaningful not! *ANESTHESIOLOGY* 1999; 90:355–6
- Dickinson K, Bunn F, Wentz R, Edwards P, Roberts I: Size and quality of randomised controlled trials in head injury: Review of published studies. *BMJ* 2000; 320:1308–11
- Kelly MJ, Wadsworth J: What price inconclusive clinical trials? *Ann R Coll Surg Engl* 1993; 75:145–6
- Pua HL, Lerman J, Crawford MW, Wright JG: An evaluation of the quality of clinical trials in anesthesia. *ANESTHESIOLOGY* 2001; 95:1068–73
- Greenfield ML, Mhyre JM, Mashour GA, Blum JM, Yen EC, Rosenberg AL: Improvement in the quality of randomized controlled trials among general anesthesiology journals 2000 to 2006: A 6-year follow-up. *Anesth Analg* 2009; 108:1916–21
- Snowdon C, Elbourne DR, Garcia J, Campbell MK, Entwistle VA, Francis D, Grant AM, Knight RC, McDonald AM, Roberts I: Financial considerations in the conduct of multi-centre randomised controlled trials: Evidence from a qualitative study. *Trials* 2006; 7:34–52
- Moher D, Schulz KF, Altman DG: The CONSORT statement: Revised recommendations for improving the quality of reports of parallel-group randomised trials. *Lancet* 2001; 357:1191–4
- Freiman JA, Chalmers TC, Smith H Jr, Kuebler RR: The importance of beta, the type II error and sample size in the design and interpretation of the randomized controlled trial. Survey of 71 “negative” trials. *N Engl J Med* 1978; 299:690–4
- Freemantle N, Calvert M: Composite and surrogate outcomes in randomised controlled trials. *BMJ* 2007; 334: 756–7
- Kip KE, Hollabaugh K, Marroquin OC, Williams DO: The problem with composite end points in cardiovascular studies: The story of major adverse cardiac events and percutaneous coronary intervention. *J Am Coll Cardiol* 2008; 51:701–7
- Ferreira-González I, Busse JW, Heels-Ansdell D, Montori VM, Akl EA, Bryant DM, Alonso-Coello P, Alonso J, Worster A, Upadhye S, Jaeschke R, Schünemann HJ, Permyer-Miralda G, Pacheco-Huergo V, Domingo-Salvany A, Wu P, Mills EJ, Guyatt GH: Problems with use of composite end points in cardiovascular trials: Systematic review of randomised controlled trials. *BMJ* 2007; 334:786–8
- POISE Study Group, Devereaux PJ, Yang H, Yusuf S, Guyatt G, Leslie K, Villar JC, Xavier D, Chrolavicius S, Greenspan L, Pogue J, Pais P, Liu L, Xu S, Málaga G, Avezum A, Chan M, Montori VM, Jacka M, Choi P: Effects of extended-

- release metoprolol succinate in patients undergoing non-cardiac surgery (POISE trial): A randomised controlled trial. *Lancet* 2008; 371:1839-47
17. Myles PS, Smith J, Knight J, Cooper DJ, Silbert B, McNeil J, Esmore DS, Buxton B, Krum H, Forbes A, Tonkin A, ATACAS Trial Group: Aspirin and Tranexamic Acid for Coronary Artery Surgery (ATACAS) Trial: Rationale and design. *Am Heart J* 2008; 155:224-30
  18. Myles PS, Leslie K, Peyton P, Paech M, Forbes A, Chan MT, Sessler D, Devereaux PJ, Silbert BS, Jamrozik K, Beattie S, Badner N, Tomlinson J, Wallace S, ANZCA Trials Group: Nitrous oxide and perioperative cardiac morbidity (ENIGMA-ID) Trial: Rationale and design. *Am Heart J* 2009; 157:488-94.e1
  19. Montori VM, Permyer-Miralda G, Ferreira-González I, Busse JW, Pacheco-Huergo V, Bryant D, Alonso J, Akl EA, Domingo-Salvany A, Mills E, Wu P, Schünemann HJ, Jaeschke R, Guyatt GH: Validity of composite end points in clinical trials. *BMJ* 2005; 330:594-6
  20. Neaton JD, Gray G, Zuckerman BD, Konstam MA: Key issues in end point selection for heart failure trials: Composite end points. *J Card Fail* 2005; 11:567-75
  21. Myles PS: What's new in trial design: Propensity scores, equivalence, and non-inferiority. *J Extra Corpor Technol* 2009; 41:P6-10
  22. Montori VM, Devereaux PJ, Adhikari NK, Burns KE, Eggert CH, Briel M, Lacchetti C, Leung TW, Darling E, Bryant DM, Bucher HC, Schünemann HJ, Meade MO, Cook DJ, Erwin PJ, Sood A, Sood R, Lo B, Thompson CA, Zhou Q, Mills E, Guyatt GH: Randomized trials stopped early for benefit: A systematic review. *JAMA* 2005; 294:2203-9
  23. Myles PS, Leslie K, Chan MT, Forbes A, Paech MJ, Peyton P, Silbert BS, Pascoe E, ENIGMA Trial Group: Avoidance of nitrous oxide for patients undergoing major surgery: A randomized controlled trial. *ANESTHESIOLOGY* 2007; 107:221-31
  24. Pocock SJ: When (not) to stop a clinical trial for benefit. *JAMA* 2005; 294:2228-30

## ANESTHESIOLOGY REFLECTIONS

### The 1847 Murphy Chloroform Inhaler



After relocating from Dublin to London in 1840, Edward William Murphy, M.D. (1802-1877), served as University College's Professor of Midwifery (1842-1865). Misdated by many historians, his Murphy Chloroform Inhaler (see the later model above, courtesy of the Wood Library-Museum) was actually first used clinically on "the 17th December, 1847" and was even depicted in his 1848 booklet *Chloroform in the Practice of Midwifery* as a chloroform-dampened sponge surrounded by a circular box with a "tin plate . . . interposed, leaving a small fissure" for vapor to pass *via* mouth-piece to the briefly nose-pinched patient. Expirations "passed through an opening in the upper part of the mouth-piece." The professor's interests in obstetric analgesia may have been reinforced by the eight childbirths endured by his wife, a fellow Dubliner whose maiden name, ironically, was "English." (Copyright © the American Society of Anesthesiologists, Inc. This image appears in color in the *Anesthesiology Reflections* online collection available at [www.anesthesiology.org](http://www.anesthesiology.org).)

George S. Bause, M.D., M.P.H., Honorary Curator, ASA's Wood Library-Museum of Anesthesiology, Park Ridge, Illinois, and Clinical Associate Professor, Case Western Reserve University, Cleveland, Ohio. [UJYC@aol.com](mailto:UJYC@aol.com).