



# The role of randomized cluster crossover trials for comparative effectiveness testing in anesthesia: design of the Benzodiazepine-Free Cardiac Anesthesia for Reduction in Postoperative Delirium (B-Free) trial

## La place des essais randomisés avec permutation de grappes pour les tests d'efficacité comparée en anesthésie: plan de l'essai d'anesthésie cardiaque sans benzodiazépine pour la réduction du délirium postopératoire (étude B-Free)

Jessica Spence, MD · Emilie Belley-Côté, MD, FRCPC · Shun Fu Lee, PhD · Shrikant Bangdiwala, PhD · Richard Whitlock, MD, PhD, FRCSC · Yannick LeManach, MD, PhD · Summer Syed, MD, MSc, FRCPC · Andre Lamy, MD, MHSc, FRCSC · Eric Jacobsohn, MBChB, MPHE, FRCPC · Sarah MacIsaac, HBSM, MSc · P. J. Devereaux, MD, PhD, FRCPC · Stuart Connolly, MD, FRCPC

Received: 9 November 2017 / Revised: 28 February 2018 / Accepted: 3 March 2018 / Published online: 18 April 2018  
© Canadian Anesthesiologists' Society 2018

**Abstract** Increasingly, clinicians and researchers recognize that studies of interventions need to evaluate not

only their therapeutic efficacy (i.e., the effect on an outcome in ideal, controlled settings) but also their real-world effectiveness in broad, unselected patient groups. Effectiveness trials inform clinical practice by comparing variations in therapeutic approaches that fall within the standard of care. In this article, we discuss the need for studies of comparative effectiveness in anesthesia and the limitations of individual patient randomized-controlled

This article is accompanied by an editorial. Please see Can J Anesth 2018; 65: this issue.

**Electronic supplementary material** The online version of this article (<https://doi.org/10.1007/s12630-018-1130-2>) contains supplementary material, which is available to authorized users.

J. Spence, MD · Y. LeManach, MD, PhD  
Departments of Anesthesia and Critical Care and Health Research Methods, Evidence, and Impact, Population Health Research Institute (PHRI), McMaster University, 2V9 – 1280 Main St. West, Hamilton, ON L8S 4K1, Canada  
e-mail: jessicaspence13@gmail.com

E. Belley-Côté, MD, FRCPC  
Departments of Anesthesia and Critical Care, Medicine (Cardiology), and Health Research Methods, Evidence, and Impact, Population Health Research Institute (PHRI), McMaster University, Hamilton, ON, Canada

S. F. Lee, PhD · S. Bangdiwala, PhD · S. MacIsaac, HBSM, MSc  
Population Health Research Institute (PHRI), Hamilton, ON, Canada

R. Whitlock, MD, PhD, FRCSC · A. Lamy, MD, MHSc, FRCSC  
Departments of Surgery (Cardiac Surgery) and Health Research Methods, Evidence, and Impact, Population Health Research Institute (PHRI), McMaster University, Hamilton, ON, Canada

S. Syed, MD, MSc, FRCPC  
Department of Anesthesia and Critical Care, McMaster University, Hamilton, ON, Canada

E. Jacobsohn, MBChB, MPHE, FRCPC  
Departments of Anesthesia and Internal Medicine, University of Manitoba, Winnipeg, MB, Canada

P. J. Devereaux, MD, PhD, FRCPC  
Departments of Medicine (Cardiology) and Health Research Methods, Evidence, and Impact, Population Health Research Institute (PHRI), McMaster University, Hamilton, ON, Canada

S. Connolly, MD, FRCPC  
Department of Medicine (Cardiology), Population Health Research Institute (PHRI), McMaster University, Hamilton, ON, Canada

trials in determining comparative effectiveness. We introduce the concept of randomized cluster crossover trials as a means of answering questions of comparative effectiveness in anesthesia, using the design of the Benzodiazepine-Free Cardiac Anesthesia for Reduction in Postoperative Delirium (B-Free) trial (Clinicaltrials.gov identifier NCT03053869).

**Résumé** De plus en plus, les cliniciens et chercheurs reconnaissent que les études sur des interventions doivent évaluer non seulement leur efficacité thérapeutique (c'est-à-dire l'effet sur un critère d'évaluation dans un cadre contrôlé, idéal), mais aussi son efficacité dans la vraie vie sur de larges groupes de patients non sélectionnés. Les études sur l'efficacité informent la pratique clinique en comparant des variations — entrant dans les normes de soins — dans les démarches thérapeutiques. Dans cet article, nous discutons de la nécessité d'avoir en anesthésie des études comparatives de l'efficacité et des limites des essais contrôlés avec des patients randomisés individuellement pour la détermination de l'efficacité comparative. Nous présentons le concept d'essais randomisés avec permutation (« crossover ») de grappes (« cluster ») comme moyen de répondre aux questions d'efficacité comparative en anesthésie en utilisant le plan de l'essai d'anesthésie cardiaque sans benzodiazépine pour la réduction du délirium postopératoire (étude B-Free).

### The need for comparative effectiveness trials in clinical research

Individual patient randomized-controlled trials (RCTs) have historically represented the gold standard methodologic approach used to establish the clinical efficacy of an intervention; however, individual patient RCTs are not the preferred design to answer certain types of questions. This is because their external validity (generalizability) may be limited by their selected and specific populations and their conduct in ideal clinical conditions. Though variations in approaches to improve the generalizability of individual patient RCTs have been described,<sup>1</sup> in general, individual patient RCTs are better suited to provide evidence of efficacy (i.e., effect under ideal circumstances), but less suited to answering questions of effectiveness (i.e., true benefit to all patients in routine practice).<sup>2,3</sup> This is, in part, because effectiveness research examines the effect of an intervention after the influences of patient-, provider-, and system-level factors that may moderate the effect of an intervention are taken into account. Introducing this additional noise often requires a larger sample size to show an intervention's effectiveness, even if it has already demonstrated efficacy. Though very

large ( $n > 10,000$ ) individual patient RCTs can evaluate questions of effectiveness, the time and expense associated with individual patient recruitment and randomization make them onerous and costly.

Beyond the limitations of their applicability, efficacy RCTs impose processes that are different from usual care, making their conduct costly and their implementation complicated.<sup>4</sup> Because of their complexity and cost, the optimal therapeutic approach has not been evaluated using RCT methodology in many clinical contexts.<sup>4</sup> As a result, many clinical decisions are only indirectly informed by the results of individual patient RCTs. While observational studies represent a more cost-effective and simple means of addressing comparative effectiveness, randomized comparisons are necessary to ensure balance of prognosis between treatment groups. To establish evidence that will inform day-to-day clinical decisions, researchers need economical and efficient methods to evaluate the relative effectiveness of therapeutic approaches. While individual patient RCTs generate important knowledge about therapeutic efficacy, they are costly and time-consuming, particularly when one considers the large sample size required to evaluate patient-important outcomes.<sup>5</sup> In general, cluster RCTs are the preferred method to evaluate questions of effectiveness,<sup>6</sup> though this is one of several possible approaches.

### The need for comparative effectiveness trials in anesthesia research

To date, clinical trials in anesthesia have focused on physiologic and pharmacologic phenomena occurring within the intra- and immediate postoperative period.<sup>7</sup> Surrogate outcomes—measures of treatment effect that correlate with, but have no guaranteed relationship with a patient-important outcome (e.g., biomarkers)—have been common in anesthesia research.<sup>8,9</sup> During the last two decades, this focus has broadened to include intermediate and long-term effects of anesthesia-related interventions on patient-important outcomes such as major morbidity and mortality.<sup>7,9</sup> Because of the low postoperative incidence of these events and their multiple causal pathways, anesthesia interventions are unlikely to have more than a modest effect and, thus, many thousand participants are required for a study to have adequate power to quantify the effect of an intervention. For example, in the case of cardiac surgery—which has among the highest perioperative mortality rates at 3%<sup>10</sup>—more than 20,000 participants are required for a trial to show a reduction in mortality from 3.0 to 2.4% (a relative risk reduction [RRR] of 20%).

In addition, anesthesia is a specialty in which there is wide variability in approach across individual practitioners,

even within the same institution.<sup>11-13</sup> This may, in part, be related to the limited evidence base for best practice, with clinical practice guidelines frequently based on “expert opinion” rather than a summary of RCT-derived evidence.<sup>14</sup> In addition, it has been repeatedly shown that the patients who consent to participate in research are fundamentally different from those who do not, a phenomenon referred to as volunteer bias.<sup>15-17</sup> Even when large, multicentre, individual-patient RCTs are conducted, the volunteer bias that results from the need for consent to participate limits the generalizability of study findings to the “average” patient, thus contributing to a lack of consensus as to what constitutes best practice in many clinical situations. Thus, anesthesia trials that address the optimal approaches to care for the average patient in routine practice (i.e., clinical effectiveness trials) are required. Although in principle this could be evaluated through an individual patient RCT, this is probably not the optimal approach to address broad questions of policy.

### Randomized cluster crossover trials as a method of evaluating relative effectiveness

In some situations, individual-patient randomization is not only impractical but inappropriate.<sup>18</sup> When the intervention being evaluated is a policy or approach to care, applied at the level of an institution, region, or healthcare system, evaluating the outcome at the level of the individual does not answer the question about the impact on the broader group. Historically, researchers have randomized at the level of a group of patients (i.e., the cluster) to evaluate the impact of standardizing an approach to care.<sup>18</sup> In most clinical trials, the cluster is the healthcare facility. Nevertheless, cluster trials that compare outcomes based on variations in hospital standard operating procedures (SOPs) have methodologic features that limit their statistical power. Individuals being studied within a cluster are usually more similar than individuals across clusters, a phenomenon described statistically by the intra-cluster correlation coefficient (ICC). For example, individuals who live within the same geographic region are more likely to have similar health outcomes because of shared cultural and socioeconomic characteristics. Thus, patients within a hospital may have more similar outcomes than those between hospitals, leading to overly precise estimates of effect and results that may have spuriously low *P* values if clustering is ignored.<sup>19</sup>

The randomized cluster crossover design, a variation of the cluster-randomized design, overcomes this issue and regains some of the statistical power lost by using the cluster design. In this approach, clusters are randomized to receive each of the evaluated interventions at least once

during separate crossover periods.<sup>19</sup> Clusters alternate, or cross over, from one intervention arm to the other one or more times. By exposing each cluster to both treatment arms, each cluster acts as its own control group, which mitigates the effect of imbalance between sites in patient and provider characteristics.<sup>19</sup> Nevertheless, because of the cluster effect, the application of standard sample size approaches leads to an underpowered study (type II error) and the use of standard analytic methods tends to bias *P* values, possibly resulting in spurious statistical significance (type I error). Thus, a correction factor is applied to sample size calculations, resulting in the need for a larger sample size requirement compared with that of an individual patient RCT. During the analysis stage, a random effects model may be used to account for the clustering effect.

### The Benzodiazepine-Free Cardiac Anesthesia for Reduction in Postoperative Delirium (B-Free) trial

Delirium is a serious problem that affects 15-30% of patients after cardiac surgery.<sup>20,21</sup> It is an acute confusional state associated with prolonged hospital length of stay,<sup>22</sup> institutional discharge,<sup>22</sup> functional decline,<sup>22</sup> cognitive decline,<sup>23</sup> and death.<sup>24</sup> Evidence derived from individual patient RCTs conducted in the setting of the intensive care unit has suggested that benzodiazepines are associated with an increased risk of delirium, such that practice guidelines issued by the Society for Critical Care Medicine<sup>25</sup> and American Geriatric Society<sup>26</sup> have recommended that they be avoided in these populations. Nevertheless, despite benzodiazepine administration after cardiac surgery having been reduced, their intraoperative administration remains common.

As perioperative cardiovascular surgery clinicians, we were aware of significant variability in the use of benzodiazepines as part of anesthetic regimens for patients undergoing cardiac surgery. Some clinicians do not use them, others use large doses, and another subset gives them in moderate doses. Historical concerns about the need for benzodiazepines to ensure hemodynamic stability and prevent intraoperative awareness during cardiac surgery are balanced against a hypothetical association between benzodiazepines and postoperative delirium. Faced with this clinical equipoise, we identified the need to establish whether the routine or restricted use of benzodiazepines in cardiac anesthesia affected the incidence of postoperative delirium.

To evaluate the comparative effectiveness of two SOPs for cardiac anesthesia—both of which are used in clinical practice—we designed a cluster crossover trial with randomization at the hospital level. Within this trial, we seek to evaluate the impact of two different approaches to

cardiac anesthesia, one where nearly all patients receive intraoperative benzodiazepine unless there are contraindications (routine benzodiazepine arm) and the other where nearly all patients receive no intraoperative benzodiazepines unless there are contraindications (benzodiazepine-restricted arm). We will be evaluating the impact of these two strategies during 12 four-week crossover periods. This is a situation in which an individually randomized clinical trial is impractical.

### Data collection within randomized cluster crossover trials

To answer questions of clinical effectiveness, randomized cluster crossover trials must be integrated into the standard processes of clinical care. As such, data points and study outcomes must be routinely evaluated and documented by frontline healthcare providers. To minimize the time and costs associated with data collection, most or all data points should ideally be stored in electronic medical records (EMRs) suitable for exportation into the trial data set. In addition, study endpoints need to be objectively defined, measured in a similar way across clusters, and of central importance to patients. For the B-Free trial, we respected these principles by only collecting patient and surgical characteristics already collected for administrative purposes and choosing a primary outcome that was both important to patients and routinely assessed and documented by nursing staff providing postoperative care. Delirium after cardiac surgery is routinely measured as it is considered a quality of care metric in most centres. Two validated tools are recommended to assess delirium in the ICU:<sup>25</sup> the Confusion Assessment Method<sup>27</sup> and Intensive Care Delirium Screening Checklist.<sup>28</sup> These validated tools both generate a binary assessment of whether delirium is present or absent and have been shown to be both sensitive and specific in its detection.<sup>25</sup> All data collected within B-Free are either directly entered or scanned into the electronic medical record. Most data are available for download into the study database, with only a few data points manually extracted from patient charts by study personnel.

### Ethical considerations for randomized cluster crossover trials

The interventions evaluated within randomized cluster crossover trials are applied at the cluster level. Consent to participate is therefore obtained at the cluster rather than the patient level, thereby ensuring representative sampling and enrolment of all eligible patients. Though

randomization at the cluster level does not automatically abrogate the need for individual consent, for a randomized cluster crossover trial to evaluate the impact of policy change broadly applied at the level of a cluster, a research ethics board waiver of the need for individual consent is key. If this waiver cannot be obtained, the efficiency gained by cluster randomization is lost.

The Tri-Council Policy Statement (TCPS2) requires a study to fulfill several conditions for individual consent to be waived, most importantly:<sup>29</sup>

- (1) The alteration to consent requirements is necessary to address the research question.
- (2) The lack of *a priori* consent will not have an adverse impact on participant welfare.
- (3) The benefits of the research, whether direct, indirect, or societal, justify any risks associated with the absence of *a priori* consent.

In a cluster-randomized trial, the unit of randomization is the cluster rather than the individual. Although individual patients may participate in the research, it may not be possible for a patient to provide or withhold consent to an intervention as it is implemented at the level of the health system rather than the patient. Centres studying SOPs using this methodologic approach typically notify patients that a study of hospital policy is taking place and that their anonymous data are being collected. Though they are not able to opt out of the SOP being studied, patients may withdraw the data being collected that pertain to them. Consistent with the requirements of the TCPS2, this involves some form of notification, usually in the form of a letter of information, provided to patients prior to receiving the intervention being studied or having their data collected.

The TCPS2 allows for deviations from the traditional approach to individual informed consent only in specific circumstances. As discussed in a recent article by Murdoch and Caulfield,<sup>30</sup> the increased financial cost and lost efficiency associated with obtaining individual informed consent are not ethically sufficient to render a study impractical to conduct without a waiver of individual consent. Nevertheless, conducting the B-Free trial would have been impractical without a waiver of individual consent. The unbiased evaluation of alternate cardiac anesthesia policies requires their almost universal application. Both policies are associated with minimal risk and are currently used routinely in clinical practice. In preparation for the trial, we conducted a survey to formally measure variations in reported intraoperative use of benzodiazepine by Canadian cardiac anesthesiologists. This survey showed a wide range of benzodiazepine administration and established clinical equipoise.<sup>31</sup> Finally, establishing the optimal approach to intraoperative

benzodiazepine use is important to both guide anesthesia practice and possibly benefit patients and society by reducing delirium and its associated morbidity in patients undergoing cardiac surgery, thus satisfying the final requirement for waived individual consent.

For the B-free pilot trial, we obtained approval from the research ethics boards of both participating pilot sites; the need to obtain individual consent was waived. We notified patients by letter that a study evaluating intraoperative cardiac anesthesia practice was being undertaken and that they could choose to opt out, removing their data from the study database. Nevertheless, they would nonetheless be treated as per the policy in place at the time; SOPs being coordinated at the anesthesia-department level are not optional and represent a standardized approach to care at the level of the institution. As an example, consider a hospital evaluating the impact of two different types of surgical suture on postoperative infection rates. Although patients may be given the opportunity to opt out of having their data included in study analyses, they would nonetheless have their incisions stitched with the suture being evaluated at the time that they underwent surgery.

### Statistical considerations for randomized cluster crossover trials

#### Sample size

Cluster-randomized trials are statistically less efficient than individual-randomized trials because of the lack of independence between individuals in the same cluster. Randomized cluster crossover trials regain some of this lost statistical power by comparing the effects of an intervention within clusters. Nonetheless, the statistical power of a randomized-cluster crossover trial remains lower than that of an individual patient RCT. The unique statistical considerations of randomized-cluster crossover trials must be recognized when determining the sample size required for a randomized cluster crossover trial. These considerations are complex, and it is crucial that a biostatistician with expertise in randomized-cluster crossover trials be a member of the study team.

Given that the unit of randomization is the centre, sample size and statistical power depend on the ICC. The ICC describes how strongly patients within the same cluster are correlated with each other and how they differ from patients within other clusters.<sup>32</sup> Compared with a randomly selected sample from the same population, members of clusters are more likely to have similar outcomes in response to an intervention. The larger the ICC, the greater the dependence among observations within a cluster and the greater the number of clusters

needed to achieve target power. Obtaining estimates of the ICC can be difficult; previous studies have obtained estimates using administrative data.<sup>4</sup> The use of pilot trials to derive the ICC has been described,<sup>33</sup> though this is generally not recommended because the sample sizes from pilot studies are usually too small to allow for reliable estimates.<sup>34</sup> Given that the ICC has been found to vary with prevalence, the use of the prevalence of an outcome to estimate the ICC has been described.<sup>35</sup>

The other consideration for cluster crossover trials is the inter-period correlation (IPC), which describes how strongly different patients within the same cluster during different crossover periods resemble each other. It is often much more difficult to obtain estimates of the IPC,<sup>36</sup> and thus studies often use an assumed value that is some proportion of the ICC as, generally,  $IPC \leq ICC$ .<sup>4,32,35</sup> In any healthcare setting, the clinical correlates that the IPC represents are familiar to most practicing clinicians. For example, there are frequent anecdotal accounts that patients are more likely to present with severe illness after the winter holiday season or that patient outcomes are worse in July, which represents the start of the academic year when new trainees are commencing their clinical rotations. These period effects may be overcome by alternating between shorter crossover periods. Nevertheless, the benefit in IPC gained doing so must be balanced against the clinical impracticality of very frequent changes in SOP.

In the B-Free study, we used a formula derived by extending the approaches developed by Donner *et al.*,<sup>37</sup> Giraudeau *et al.*,<sup>32</sup> Forbes *et al.*,<sup>38</sup> and Taljaard *et al.*<sup>39</sup> for a cluster crossover of two periods to multiple periods to calculate a preliminary sample size. The formula itself, which was locally derived and validated, can be found in e-supplement 1.

In B-Free, the ICC will depend on the incidence of delirium in each participating cluster. The ICC has previously been shown to be related to outcome prevalence in clustered, binary data. Thus, prior to completing site recruitment for the full B-Free trial, we used our local delirium rates to estimate a conservative ICC of 0.02, based on values determined by Gulliford *et al.* using several large administrative data sets.<sup>35</sup> We undertook a preliminary and conservative sample size calculation using an estimated prevalence of delirium of 15% (derived from local administrative data) and predicted an ICC of 0.02<sup>35,36</sup> and an  $IPC = 0.5 * ICC$ .<sup>32</sup> The table illustrates the total number of patients (N) and number of hospitals ( $N/\bar{m}$ ) required to be randomized for each of two intervention groups to assure a sufficient power of 80% and a type I error probability of 5% (two sided) with an anticipated control event rate ( $p_1$ ) of 15% for different combinations of arithmetic mean total cluster size



over all periods ( $\bar{m}$ ), ICC, IPC, and RRR. Based on the assumption of an ICC = 0.02, an IPC = 0.5·ICC, an RR reduction of 15%, and an average total cluster size of 1,000, we will require a sample of 15,886 patients studied within 16 participating hospitals.

#### Data analysis

The data obtained from a randomized cluster crossover trial can be analyzed at the individual level in the sense that individuals within a cluster each provide a data point that is used within the overall analysis (e.g., the binary outcome of delirium). Nevertheless, to avoid having spurious statistical significance, a modelling approach that accounts for the correlation within centre across period (IPC) and within centre in a period (ICC) must be used.<sup>19</sup>

#### Economic considerations for randomized cluster crossover trials

A major advantage of randomized cluster crossover studies is—when they qualify for waiver of individual consent—they are of relatively low cost and rapid patient enrolment. Waived individual consent significantly reduces the cost and time associated with patient recruitment and individual randomization. Because the interventions being evaluated are integrated into patient care and, similarly, so are study processes, minimal research personnel are required for data collection. For example, if B-Free includes 16 clusters, with an average total cluster size of 1,000 patients (i.e., annual cardiac surgery case volume), the two SOPs will be evaluated in 16,000 patients over a single year. We estimate that the total expenditure (expressed in CAD) for each site will be 20,000 CAD, which will cover the costs of data extraction from EMRs and the salary of a 0.2

full-time-equivalent research assistant to extract data not stored electronically as well as coordinate communication to anesthesia staff. Considering an additional 150,000.00 CAD for central coordination costs, this translates to a total of 470,000.00 CAD to conduct a study of more than 15,000 patients. To put this economy of time and resources into perspective, we calculated a sample budget for B-Free as though it had been designed as an individual patient RCT. Based on the recently published ENIGMA-II trial of 7,112 patients, conducted over five years, we approximate five years for recruitment and completion of follow-up in our required sample size for an individual patient RCT of 7,409 patients (see Table). Accounting for five years of central coordination costs, site fees of 750 CAD per patient, and operating expenses, we estimate that conducting B-Free as an individual patient RCT would require more than 10 million CAD.

#### General considerations for randomized cluster crossover trials in anesthesia

The historically individual nature of anesthesia practice presents both a unique challenge to and an opportunity for the conduct of randomized cluster crossover trials. The degree of individual practice variability within the standard of care means that there are many clinical questions that would be best answered using randomized cluster crossover methodology. For example, liberal vs conservative oxygenation, routine use of pulmonary artery catheters in cardiac surgery vs only in certain circumstances, and intraoperative mean arterial blood pressure targets are all topics extremely well-suited to study using this methodology. On the other hand, the individual nature of anesthesia practice may present a challenge with respect to studying SOPs. As opposed to

**Table** Sample size considerations for a potential future cluster crossover RCT

RRR	Overall cluster size ( $\bar{m}$ )	No. periods	ICC	Total N (inflated)	Sample size for an individual patient RCT	No. hospitals ( $N/\bar{m}$ )
15%	1,000	12	0.01	11,648	7,409	12
		12	0.02	15,886	7,409	16
	800	12	0.01	10,785	7,409	14
		12	0.02	14,161	7,409	18
20%	1,000	12	0.01	6,402	4,072	7
		12	0.02	8,731	4,072	9
	800	12	0.01	5,928	4,072	8
		12	0.02	7,783	4,072	10

Total patients (N) and number of hospitals ( $N/\bar{m}$ ) required to be randomized for each of two intervention groups to assure a sufficient power of 80% and a type I error rate of 5% (two sided) with 12 periods, a coefficient of variation of 0.63 for an assumed unequal cluster size, an anticipated control event rate ( $p_1$ ) of 15% for different combinations of average cluster size per cluster ( $\bar{m}$ ), intra-cluster correlation (ICC), and relative risk reduction (RRR)

RCT = randomized controlled trials

other areas of medicine, few academic anesthesia departments employ SOPs, which may, in part, stem from the lack of evidence supporting one approach over another. As such, practitioners may be uncomfortable deviating from their usual practice patterns, even if the SOP under evaluation falls within the standard of care for the profession. This possible non-adherence to the SOP under evaluation poses a potential threat to the internal validity of the study and is an important issue that must be addressed within the design of a trial.

Within the B-Free trial, we recognized that, though there is community clinical equipoise with respect to intraoperative benzodiazepine administration to cardiac surgery patients, many practitioners have strong and dichotomous opinions about using or not using these medications. As such, before undertaking a large, multicentre trial, we conducted a vanguard study that included two separate centres: one where most practitioners routinely used at least moderate doses of benzodiazepine and the other where practitioners used low-dose or no benzodiazepines. In doing so, we sought to evaluate whether we could feasibly achieve adherence to each of the SOPs (i.e., administration of benzodiazepine to most patients *vs* no administration of benzodiazepine to most patients) under evaluation in at least 80% of patients managed during each intervention arm (Clinicaltrials.gov identifier NCT03053869).

### Current status of the B-Free trial

In both pilot sites we were able to get Research Ethics Board approval of a waived consent model and to engage cardiac anesthesiologists to participate in the trial. We are currently completing the pilot trial. We will refine the protocol for the full trial based on detailed analysis of pilot trial results and input from participating clusters. We expect to begin the trial in the summer of 2018.

### Discussion

Inter-practitioner variation in anesthesia practice often stems from a lack of knowledge about the relative effectiveness of different approaches to patient care. This variation reveals the knowledge gap, but also highlights the opportunity to improve care by comparatively testing the effectiveness of routine practices. The randomized cluster crossover design is the appropriate way to do so. Though individual RCTs represent the gold standard means of assessing the efficacy of treatments, they are difficult to implement and costly, and uncertainty may exist about the

generalizability of trial results. By incorporating the rigour of randomization into day-to-day clinical care, cluster randomization reduces the complexity of trial implementation and, when a waiver of individual patient consent is obtained, improves the generalizability of results. Nevertheless, this approach should only be used when the interventions being compared have minimal risk and there is clear clinical equipoise.

There are some limitations of cluster crossover trials for answering questions within anesthesia practice. First, the interventions being compared must be of minimal risk and clinical equipoise must be demonstrable. Often, documented proof of minimal risk—though frequently intuitive—is not available, even if both interventions are routinely being used in clinical practice. As such, researchers who are considering undertaking a trial that employs this design may need to quantify clinical equipoise and minimal risk, through either conduct of a survey of practitioners or review of administrative data. In addition, randomized cluster crossover trials are not suited to answering questions of clinical efficacy. This is because the noise introduced by the broad inclusion of patients who cannot benefit from the intervention will diminish the signal of effect that would otherwise be seen under the strictly controlled population and conditions of an individual patient RCT. Though strict inclusion criteria and study protocols could be applied to the randomized cluster crossover design, the benefits of diminished cost and increased efficiency would be abnegated.

To our knowledge, no previous anesthesia trials have been conducted using a randomized cluster crossover design. Its use has, however, become common in the setting of the intensive care unit,<sup>40,41</sup> where—similarly to anesthesia practice—interventions are likely to have small effect sizes on patient-important outcomes because of the multiple causal pathways that result in patient morbidity and mortality. Given the resultant need for an extremely large sample size, the costs and time required to evaluate an intervention using an individual patient RCT are often so prohibitive that simple questions of clinical effectiveness remain unresolved. The use of the cluster-randomized crossover design presents a feasible solution to this issue. Its widespread application will require increasing multicentre and multinational collaboration among anesthesia departments.

**Financial disclosure and conflict of interest** None.

**Editorial responsibility** This submission was handled by Dr. Gregory L. Bryson, Deputy Editor-in-Chief, *Canadian Journal of Anesthesia*.

**Author contributions** *Jessica Spence, Emilie Belley-Côté, P.J. Devereaux, and Stuart Connolly* contributed substantially to all

aspects of this manuscript, including conception and design of the B-Free trial and conception, design, and drafting of the article. *Shun Fu Lee, Shrikant Bangdiwala, and Yannick LeManach* contributed substantially to statistical planning for the B-Free trial and drafting of the article. *Richard Whitlock, Summer Syed, Andre Lamy, Eric Jacobsohn, and Sarah MacIsaac* contributed substantially to the conception and design of the B-Free trial and drafting of the article.

## References

- Karanicolas PJ, Montori VM, Devereaux PJ, Schunemann H, Guyatt GH. A new ‘mechanistic-practical’ framework for designing and interpreting randomized trials. *J Clin Epidemiol* 2009; 62: 479-84.
- Booth CM, Tannock IF. Randomised controlled trials and population-based observational research: partners in the evolution of medical evidence. *Br J Cancer* 2014; 110: 551-5.
- Platt R, Takvorian SU, Septimus E, et al. Cluster randomized trials in comparative effectiveness research: randomizing hospitals to test methods for prevention of healthcare-associated infections. *Med Care* 2010; 48: S52-7.
- Connolly SJ, Philippon F, Longtin Y, et al. Randomized cluster crossover trials for reliable, efficient, comparative effectiveness testing: design of the Prevention of Arrhythmia Device Infection Trial (PADIT). *Can J Cardiol* 2013; 29: 652-8.
- Yusuf S, Collins R, Peto R. Why do we need some large, simple randomized trials? *Stat Med* 1984; 3: 409-22.
- Fiore LD, Lavori PW. Integrating Randomized comparative effectiveness research with patient care. *N Engl J Med* 2016; 374: 2152-8.
- Sessler DI. Long-term consequences of anesthetic management. *Anesthesiology* 2009; 111: 1-4.
- Myles PS. Why we need large randomized studies in anaesthesia. *Br J Anaesth* 1999; 83: 833-4.
- Lurati Buse GA, Devereaux PJ. Anesthesia needs large international clinical trials. *HSR Proc Intensive Care Cardiovasc Anesth* 2010; 2: 153-5.
- Peterson ED, Coombs LP, DeLong ER, Haan CK, Ferguson TB. Procedural volume as a marker of quality for CABG surgery. *JAMA* 2004; 291: 195-201.
- Lilot M, Ehrenfeld JM, Lee C, Harrington B, Cannessa M, Rinehart J. Variability in practice and factors predictive of total crystalloid administration during abdominal surgery: retrospective two-centre analysis. *Br J Anaesth* 2015; 114: 767-76.
- Duggan A, Koff E, Marshall V. Clinical variation: why it matters. *Med J Aust* 2016; 205: S3-4.
- Maccario A, Chung A, Weinger MB. Variation in practice patterns of anesthesiologists in California for prophylaxis of postoperative nausea and vomiting. *J Clin Anesth* 2001; 13: 353-60.
- Nickinovich DG, Connis RT, Caplan RA, Arens JF, Apfelbaum JL. Evidence-based practice parameters: the American Society of Anesthesiologists. In: Fleischer LA, editor. *Evidence-Based Practice of Anesthesiology*. 3rd ed. Philadelphia, PA: Elsevier; 2013. p. 2-7.
- Dominioni L, Poli A, Mantovani W, Rotolo N, Imperatori A. Volunteer effect and compromised randomization in the Mayo Project of screening for lung cancer. *Eur J Epidemiol* 2011; 26: 79-80.
- Songstad NT, Roberts CT, Manley BJ, Owen LS, Davis PG; HIPSTER Trial Investigators. Retrospective consent in a neonatal randomized controlled trial. *Pediatrics* 2018. DOI: <https://doi.org/10.1542/peds.2017-2092>.
- Jordan S, Watkins A, Storey M, et al. Volunteer bias in recruitment, retention, and blood sample donation in a randomised controlled trial involving mothers and their children at six months and two years: a longitudinal analysis. *PLoS One* 2013; 8: e67912.
- Sim J, Dawson A. Informed consent and cluster-randomized trials. *Am J Public Health* 2012; 102: 480-5.
- Arnup SJ, Forbes AB, Kahan BC, Morgan KE, McKenzie JE. Appropriate statistical methods were infrequently used in cluster-randomized crossover trials. *J Clin Epidemiol* 2016; 74: 40-50.
- Arenson BG, MacDonald LA, Grocott HP, Hiebert BM, Arora RC. Effect of intensive care unit environment on in-hospital delirium after cardiac surgery. *J Thorac Cardiovasc Surg* 2013; 146: 172-8.
- Pauley E, Lishmanov A, Schumann S, Gala GJ, van Diepen S, Katz JN. Delirium is a robust predictor of morbidity and mortality among critically ill patients treated in the cardiac intensive care unit. *Am Heart J* 2015; 170: 79.e1-86.e1.
- Gleason LJ, Schmitt EM, Kosar CM, et al. Effect of delirium and other major complications on outcomes after elective surgery in older adults. *JAMA Surg* 2015; 150: 1134-40.
- Saczynski JS, Marcantonio ER, Quach L, et al. Cognitive trajectories after postoperative delirium. *N Engl J Med* 2012; 367: 30-9.
- Inouye SK, Westendorp RG, Saczynski JS. Delirium in elderly people. *Lancet* 2014; 383: 911-22.
- Barr J, Fraser GL, Puntillo K, et al. Clinical practice guidelines for the management of pain, agitation, and delirium in adult patients in the intensive care unit. *Crit Care Med* 2013; 41: 263-306.
- American Geriatrics Society Expert Panel on Postoperative Delirium in Older Adults. Postoperative delirium in older adults: best practice statement from the American Geriatrics Society. *J Am Coll Surg* 2015; 220: 136.e1-148.e1.
- Inouye SK, van Dyck CH, Alessi CA, Balkin S, Siegel AP, Horwitz RI. Clarifying confusion: the confusion assessment method. A new method for detection of delirium. *Ann Intern Med* 1990; 113: 941-8.
- Roberts B, Rickard CM, Rajbhandari D, et al. Multicentre study of delirium in ICU patients using a simple screening tool. *Aust Crit Care* 2005; 18: 6-16.
- Canadian Institutes of Health Research; Natural Sciences and Engineering Research Council of Canada; Social Sciences and Humanities Research Council of Canada. Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans, December 2014. Available from URL: <http://www.pre.ethics.gc.ca/eng/policy-politique/initiatives/tcps2-eptc2/Default/> (accessed March 2018).
- Murdoch B, Caulfield T. Pragmatic clinical trials and the consent process. *Research Ethics* 2017. DOI: <https://doi.org/10.1177/1747016117733506>.
- Spence J, Belley-Cote E, Devereaux PJ, et al. Benzodiazepine administration during adult cardiac surgery: a survey of current practice among Canadian anesthesiologists working in academic centres. *Can J Anesth* 2018; 65: 263-71.
- Giraudeau B, Ravaud P, Donner A. Sample size calculation for cluster randomized cross-over trials. *Stat Med* 2008; 27: 5578-85.
- McCluskey A, Ada L, Middleton S, et al. Improving quality of life by increasing outings after stroke: study protocol for the Out-and-About trial. *Int J Stroke* 2013; 8: 54-8.
- Eldridge SM, Ashby D, Kerry S. Sample size for cluster randomized trials: effect of coefficient of variation of cluster size and analysis method. *Int J Epidemiol* 2006; 35: 1292-300.



35. Gulliford MC, Adams G, Ukoumunne OC, Latinovic R, Chinn S, Campbell MJ. Intraclass correlation coefficient and outcome prevalence are associated in clustered binary data. *J Clin Epidemiol* 2005; 58: 246-51.
36. Thompson SG, Pyke SD, Hardy RJ. The design and analysis of paired cluster randomized trials: an application of meta-analysis techniques. *Stat Med* 1997; 16: 2063-79.
37. Donner A, Klar N, Zou G. Methods for the statistical analysis of binary data in split-cluster designs. *Biometrics* 2004; 60: 919-25.
38. Forbes AB, Akram M, Pilcher D, Cooper J, Bellomo R. Cluster randomised crossover trials with binary data and unbalanced cluster sizes: application to studies of near-universal interventions in intensive care. *Clin Trials* 2015; 12: 34-44.
39. Taljaard M, Teerenstra S, Ivers NM, Fergusson DA. Substantial risks associated with few clusters in cluster randomized and stepped wedge designs. *Clin Trials* 2016; 13: 459-63.
40. Young P, Bailey M, Beasley R, et al. Effect of a buffered crystalloid solution vs saline on acute kidney injury among patients in the intensive care unit: the SPLIT randomized clinical trial. *JAMA* 2015; 314: 1701-10.
41. Jaillette E, Girault C, Brunin G, et al. Impact of tapered-cuff tracheal tube on microaspiration of gastric contents in intubated critically ill patients: a multicenter cluster-randomized cross-over controlled trial. *Intensive Care Med* 2017; 43: 1562-71.