Critical Care Perspective

Positive Clinical Trials
Understand the Control Group before Implementing the Result

Christopher S. Parshuram and Brian P. Kavanagh

Department of Critical Care Medicine, and Lung Biology and Population Health Sciences Programs, Hospital for Sick Children; and Departments of Anesthesia and Pediatrics and Interdepartmental Division of Critical Care Medicine, University of Toronto, Toronto, Ontario, Canada

The reporting and analysis of control groups have caused considerable controversy in the interpretation of important clinical trials (1–5), causing interruption of ongoing studies and, in the views of some (6), mandating a fresh look at the ethical and regulatory issues governing clinical trials. This perspective explores novel approaches to the selection of control groups and analysis of study data that may assist in the interpretation of controversial findings and help to reduce future conflict.

The purpose of clinical trials is to inform clinical decision making by correctly identifying beneficial, nonbeneficial, and harmful therapies. In a clinical trial testing a new drug, the patients in the control group receive either no drug or a different drug; understanding the control group and the study result is often straightforward. In contrast, studies of singular therapies that are administered with different intensity or magnitude (e.g., different milliliters per kilogram in the case of tidal volume) can result in control groups that are difficult to interpret. For example, in studies of tidal volume (1), blood transfusion (7), and neonatal oxygenation (8), the standard of care in the control group has been debated (9–11). Nonetheless, each of these types of study involved what the authors considered “conventional” treatment in the control group.

We discuss here how the origins of between-group differences may be evaluated. Although classic principles of trial evaluation focus on comparison of baseline data, this perspective addresses the evaluation of between-group differences in clinical studies in which single treatments are applied with different intensity. We describe novel approaches for “anchoring” the control group to gain a sense of how conventional it really is (Table 1). This perspective is directed at the methodologic issues in this debate; ethical issues have been discussed elsewhere (11, 12). Finally, we suggest how researchers, as well as data safety monitoring boards, reviewers, and readers, may implement these principles into the design, conduct, appraisal, and interpretation of clinical studies.

“ANCHORING” THE CONTROL GROUP

To understand correctly the relevance of a between-group difference, the reader needs to understand the relationship between the treatment applied to the “control” group in the study and that applied in other patients not enrolled in the study. To address this relationship, three elements of comparison need to be considered: the patients (i.e., “who”), the measurements (i.e., “what”), and the method of comparison (i.e., “how”).

Comparison against Other Patients

Comparison against other populations, using tidal volume as an example, is considered below. The merits of including a “standard care” group, whether delivered as part of a protocol or as unrestricted care, has been reviewed in depth (11, 12). In the absence of such an approach, a comparison of the control group against concurrent study-eligible patients who for various reasons did not participate in the study may be an excellent means to evaluate the application of the intervention. Although description of these patients is necessary for the CONSORT criteria (13), their outcome might not be reported. These data may be instructive because such “eligible, nonrandomized” patients are contemporaneous to the study and are treated in the same institutions and by the same clinicians caring for all patients in the study. Such patients have met eligibility criteria and could have been in the trial and thus are likely to be very similar to those patients who did enter the trial. Ideally, data from “eligible, nonrandomized” patients can be collected prospectively; retrospective collection, although inferior, may nonetheless be useful, and simple data (e.g., survived versus died) should be easily obtainable.

Another comparison group is the enrolled patients themselves. In situations in which the treatment being studied involves modification of a pre-existing parameter (e.g., tidal volume), the postrandomization values can be compared against the prerandomization values. Such comparisons are available in several studies (4, 7, 14, 15). The treatment that the patients were receiving before randomization constitutes the actual “standard of care” for these patients, and discrepancies between prerandomization and postrandomization values in the control group may indicate important changes in treatment in the control patients as a result of entry into the trial.

Measurements for Comparison

Comparisons among groups or populations can be complex. We suggest that there are three approaches that may be help to discern if the treatment has been similarly applied in the study control group compared with other populations outside the study.

Direct quantification of therapy. The comparison of the dose or intensity of the treatment under study is intuitive and should be readily available. For example, after randomization, the control group in Acute Respiratory Distress Syndrome (ARDS) Network study had a mean ± SD tidal volume of 11.8 ± 0.8 compared with mean values of 12.4, 13.7, and 11.3 ml/kg, quoted for patients randomized to control groups in trials of ARDS therapy between 1989 and 1994 (16). From 1996, mean tidal volumes were 10.3 ml/kg before randomization in the ARDS Network study (1) and were 10.5 ml/kg in an international cross-sectional survey (17). Such comparisons provide the reader with a sense of where the control group lies with respect to other
“known” patients. The time of tidal volume measurement, or use of averaging, needs to be specified, and averaged values can conceal occasional extremes that might have significant impact on outcome.

**Surrogate markers.** Understanding the physiologic response to alterations in treatment forms the basis for monitoring, and titrating therapy, in critical care medicine. In the case of tidal volume, surrogate “cues” are available to the clinician. Just as “permissive hypercapnia” might represent beneficial limitation of tidal volume, low $P_{aCO_2}$, although the net result of several factors affecting production and elimination, probably indicates excessive ventilation in patients with ARDS (18) and as such may suggest an unreasonable standard of care.

**Outcomes.** It is often assumed that participation in clinical trials confers benefit to patients. This may be related to superior clinical care associated with academic “centers of excellence,” outstanding scientific merit of the research, or high standards required by institutional research ethics boards. However, the appearance of study-associated outcome benefit might be because of the need to select participants who have a reasonable (i.e., higher) chance of survival. Indeed, a recent meta-analysis suggests that participation in clinical trials may worsen, not enhance, chances of a better outcome (19). Either way, these conflicting concepts indicate that using control group mortality from other studies may be a poor means for “validating” the control group in this study.

Nonetheless, comparisons will always be made between the outcome in the current control group and the control groups of historical studies and may offer useful information. Of greatest concern are situations in which mortality in the control group is greater than expected (20). This may reflect the selection of sicker patients, although these tend not to be recruited into clinical trials. Unbalanced randomization of sicker patients into the control group should be apparent from inspection of baseline data. When historical comparison does not help, further evaluation is required to ascertain whether randomization to the control group resulted in worse than expected outcomes and thereby “accounted for” (or inflated) the between-group difference reported by the clinical trial. Such a finding does not imply that the between-group difference was due to unethical or unfair design (11, 12), rather that the result may reflect harm in the control group, instead of benefit in the intervention group. Conversely, comparable mortality rates may provide insufficient assurance that the mortality in the current control group, and by implication the treatment used, is appropriate.

Although suboptimal in disease-specific studies, a final approach is to estimate the predicted outcome using a mortality prediction model (21) and compare this result with the actual mortality in the control group. Such an approach has been employed in the study of permissive hypercapnia in ARDS (18). Whereas standard statistical tests are generally used to demonstrate whether a difference exists between groups, modifications can be used to demonstrate similarity between populations (e.g., $\chi^2$ goodness of fit test) in addition to visual inspection of graphical information and numerical expression of centrality and spread.

### DOSE–RESPONSE CHARACTERISTICS

Common sense dictates that in most situations the mortality rate will be 100% when any important physiologic parameter (e.g., tidal volume, blood pressure, $P_{aCO_2}$) registers at a level that is either close to zero or is extremely high. Thus, most dose–response relationships are nonlinear (“U-shaped” or parabolic), at least at extreme levels, and for a given population an optimal level of treatment may be identifiable. Furthermore, in situations in which the treatment in the control group cannot be easily anchored to known populations, evaluation of dose–response relationships might be used to determine whether the “better” of two treatment levels (e.g., milliliters per kilogram of tidal volume) is actually the “best.”

### Post Hoc Analysis of Two-Group Studies

Where appreciable within-group treatment variation occurs, the dose–response can be assessed either within each treatment group or for all patients. This approach acknowledges that clinical trials are themselves imperfect, and variations in the treatment delivered, including frank protocol violations, are a reality obscured by “intention to treat” analysis. An example of within-group variability yielding analyzable data comes from the ARDS Network study (1), where differences in dynamic compliance within each group were examined for mortality rates, as tidal volume could have been. In addition to revealing “hidden” differences in outcome, such analysis could confirm or refute the presence of a parabolic relationship between tidal volume and outcome. This was one of the elements of the controversial ARDS metaanalysis (10), subsequently countered by additional
analysis (22). Indeed, plateau pressure, reflecting the interaction of tidal volume and compliance, appeared to be directly related to mortality (23).

Finally, the impact of randomization might be most sensitively detected if outcomes were compared in those who were ventilated with the lowest tidal volumes before (but the highest tidal volumes after) randomization with those ventilated with the highest tidal volumes before (but the lowest tidal volumes after) randomization.

APPLICATION IN PRACTICE

It is apparent that to be reasonably termed a “control group,” demographics and treatment should be comparable to that of similar patients who were not enrolled in the study, and the outcome comparable or better. This, as discussed previously here, may not always be the case or may not be discernible. We suggest that there are three stages during which a clinical trial may be examined to test the “utility” (validity, reasonableness) of the control group: at the design phase, during the study, and after its completion.

During the design of the study, the therapy in the control group therapy would ideally be identical to a commonly accepted “standard care.” However, clinicians recognize that for continuous parameters (e.g., tidal volume, hemoglobin), a singular standard seldom exists. This can be overcome by assigning patients to a “control” group that involves “usual care” (i.e., no change in therapy associated with entry into the trial), the ethical implications of which have been discussed elsewhere (11, 12) and have proven to be a feasible option in important studies of mechanical ventilation (24–26). Alternatively, for patients not recruited into a study, the effects of no study entry, by following “eligible, nonrandomized” patients, are discussed above.

During the conduct of the study, the Data Safety Monitoring Board conventionally reviews mortality or other outcome trends, as well as individual reported adverse events (27). Aside from this conventional approach, there are two additional approaches that could be used to ensure that the question asked by the study is being answered by its conduct. First, the board could examine the intensity of therapy before and after randomization. If changes between prerandomization and postrandomization are apparent, mortality trends could be related to the magnitude and direction of changes in therapy. Discrepancies between prerandomization and postrandomization treatment in the control group would be a clear marker that entry into the study altered current care. Second, if mortality differences are associated with changes in treatment in the control group, this would provide additional evidence that the control group is not representative of “similar” patients outside the study. Thus, detection of major changes in the therapy of patients after randomization may indicate that the control group is no longer receiving “usual” care and that the answer resulting from the study will not address the hypothesis originally posed.

What’s the point of knowing whether entry into a study caused harm? If it was clear that the change in transfusion threshold, tidal volume, or oxygenation target in a control group accounted for the between-group difference in outcome, then the positive study results would have been caused not by the beneficial effects of the intervention strategy, but rather by the adverse effects of the control strategy. This does not suggest for a moment that any study was planned with that interpretation in mind, but it would be the inevitable post hoc conclusion. Because patients should not be disadvantaged by entering into a clinical trial, we believe that reviewers and readers of an article should be reluctant to adopt a “protective” or “superior” regimen, if the only reason for the positive result was a greater mortality in the control group as a result of changes in treatment after study entry.

We believe that despite the potential limitations of post hoc analysis that evaluation of dose–response relationships may reduce the exposure of further patients to less beneficial intensities of therapy (e.g., milliliters per kilogram of tidal volume), promote the efficient use of research resources, and usefully inform future clinical studies.

CONCLUSIONS

We believe that all clinicians wish to provide the “best practice,” an appealing but difficult concept, for their patients. However, it is an oversimplification to define the “best practice” on the basis that it has been shown to be “better” than another application of a therapy that could be administered at many different intensity levels. Every clinical trial provides data, and those that are of the highest quality and provide the most discussion about all of the associated patients and co-interventions are likely to provide the most reliable lessons. This perspective illustrates approaches that might maximize the information available from clinical studies.

Conflict of Interest Statement: C.S.P. does not have a financial relationship with a commercial entity that has an interest in the subject of this manuscript; B.P.K. does not have a financial relationship with a commercial entity that has an interest in the subject of this manuscript.

Acknowledgment: The authors are grateful for Drs. Paul Hebert, Niall Ferguson, and Donald Redelmeier for their constructive comments.

References


