Clinical Research Manuscripts in Anesthesiology

MORE than half of the original studies published in ANESTHESIOLOGY are Clinical Investigations. Most involve comparing the effects of various treatments or interventions on variables of interest to anesthesiologists, intensivists, and pain medicine specialists. They thus fall under the general heading of "clinical trials." Investigators often assume that clinical trials refer only to large, complex, multicenter outcome studies—and that the "rules" for clinical trials only apply to such work. This is clearly incorrect. Friedman *et al.*[1] state that "a clinical trial is defined as a prospective study comparing the effect and value of intervention(s) against a control in human beings." In fact, a clinical trial may involve anywhere from 10 to 10,000 subjects studied in 1–100 centers, and it may be performed on volunteers instead of patients. It may involve either simple or complex interventions, and it may involve testing a drug, a technique, a new piece of equipment, or a monitoring modality. In most cases, such trials involve randomized treatment assignment ("randomized clinical trial"; RCT), usually with some degree of blinding. Regardless of these specifics, current standards developed on the basis of decades of experience (and uncountable errors) indicate that such studies should adhere to certain key design features. In fact, many of these "rules" apply equally well to nonrandomized or single-group experiments and to many of the other nonrandomized epidemiologic, descriptive, or mechanistic studies that we publish.

Two articles that appear in this issue of ANESTHESIOLOGY prompted this editorial. The first is an article by Pua *et al.*[2] examining the reporting of *a priori* sample size calculations or power analysis and the presence of other related "errors" in articles published in four major anesthesia journals. Sample size calculations are important to any study for two reasons. First, they minimize the chances of type I and type II statistical errors. A type I error is a conclusion that there is an intergroup difference when, in fact, significance was achieved simply by chance. A type II error is the failure to detect a real intergroup difference because of inadequate statistical power. Second, a sample size calculation cannot be performed until the authors carefully define a clear and quantifiable hypothesis based on detecting a clinically or biologically meaningful intergroup difference in one or two "primary outcome variables." Pua *et al.*[2] note that there has been a clear improvement in the fraction of articles that include these key components, but they also note that a disturbing fraction of articles still fail to provide this information.

The second article is a report by Norris *et al.*[3] This study is a traditional RCT examining the effects of intraoperative thoracic epidural anesthesia combined with light general anesthesia *versus* general anesthesia alone, followed by postoperative patient-controlled anesthesia *versus* patient-controlled epidural analgesia (four groups). The authors conclude, "In patients undergoing surgery of the abdominal aorta, thoracic epidural anesthesia combined with a light general anesthesia and followed by either intravenous or epidural patient-controlled analgesia offers no major advantage or disadvantage when compared with general anesthesia alone followed by either intravenous or epidural patient-controlled analgesia."[3] This is clearly an important finding for many practitioners. However, this study is also an excellent example of how a clinical trial should be designed and reported.

Seven years ago, Warren Browner[4] wrote an article for this journal entitled "Clinical Research: A Simple Recipe for Doing It Well." Two years later, an international group of trialists, statisticians, epidemiologists, and biomedical editors published the Consolidated Standards of Reporting Clinical Trials (CONSORT) statement in the *Journal of the American Medical Association*.[5] This document has recently been updated.[6] Although the CONSORT process has been criticized,[7] these articles discuss a number of key issues that should generally be addressed in the design and description of a clinical trial. I would also like to comment on several important items that are often overlooked in articles submitted to (and occasionally even published in) ANESTHESIOLOGY.

A Clearly Defined and Unambiguous *A Priori* Hypothesis

Too many investigators undertake a study with no clear idea of what they are trying to prove. A common (incorrect) design is to assign patients randomly to one of two groups, administer a treatment in an appropriately blinded fashion, and measure changes in many different variables. Then, when data collection is complete, multiple statistical comparisons looking for inter-

group differences are performed. The conclusion is then often based on one or more "differences" that emerge from that analysis. This approach is unquestionably flawed (although on occasion, it might be acceptable for pilot studies when too little is known to permit formulation of a clear hypothesis). The problem, of course, is that if one performs enough statistical comparisons, one or more differences may achieve P values less than 0.05 or even less than 0.01 by simple chance. The correct approach is to define a hypothesis *in advance* (not after the study has been completed), and then design a study to test that hypothesis. This hypothesis must be clear and quantifiable. Statements to the effect of "we conducted this study to examine the effects of drug X" or "we hypothesized that patients receiving treatment A would do better than those given treatment B" are incorrect simply because "effects of" and "doing better" are not quantitative. A useful hypothesis should clearly define the specific variable being examined (the primary outcome variable) and the magnitude of the treatment effects that would be sufficient to conclude that a treatment effect was present. For example, in the article by Norris *et al.*,[3] the authors state, "To separate the influence of time period and technique, remove physician bias, and provide comparable perioperative care, we conducted a double-masked randomized clinical trial comparing alternate combinations of intraoperative anesthesia and postoperative analgesia with respect to LOS [length of stay] in patients undergoing surgery of the abdominal aorta."[3]

Sample Size Calculation and Power Analysis

As noted above, each study should clearly define a primary outcome to be studied, base the hypothesis to be tested on that outcome, and calculate a sample size that has adequate power to detect a difference of meaningful magnitude while simultaneously minimizing the changes of detecting a difference by chance. The study by Norris *et al.*[3] contains the following statement: "The study population size for this trial was 204 patients. Based on a review of 234 consecutive patients undergoing abdominal aortic reconstruction at the JHH [Johns Hopkins Hospital], we found a mean LOS of 12.7 days (SD = 4.5). We considered a 2.5-day reduction (20%) in LOS to be both clinically and economically important. Based on the formula for normal theory and assuming a two-sided type I error protection of 0.05 and a power of 0.80, 51 patients in each of the four groups were required to reveal a reduction in mean LOS of 2.5 days in any group."[3]

Defined Secondary Outcomes

A study is designed and powered to prove or disprove a hypothesis based on one or two (rarely three) primary outcome variables. Obviously, many other results may also be of interest to the investigator or reader and may provide important support for the primary hypothesis. Again, however, it is most appropriate to define these other variables clearly. In addition, it is critical that the investigators restrict their primary conclusions to differences in the primary variables. It is rarely permissible to conclude that a treatment has some effects when there were no statistically significant differences between groups in the primary outcome, even when one or more of the secondary outcomes differ. If such a secondary outcome difference is deemed of particular importance but does not support the primary hypothesis, the best approach may be to undertake a new trial that focuses on this alternative outcome.

Patient Inclusion and Exclusion

The specific criteria by which patients are deemed eligible and ineligible for enrollment in a trial should be stated in advance. Once a patient is enrolled, there are very few legitimate reasons for them to be subsequently excluded from the analysis, even if there is a protocol violation. To quote Dr. Browner[4]:

"Sometimes a patient is assigned to receive a therapy but does not, for example, because a contraindication develops after the randomization process. RCTs should follow the rule 'once randomized, always analyzed': a subject is always considered a member of the original randomization group. This rule means that RCTs are comparing assignment to a particular therapy, rather than to the therapy itself. The alternative—to analyze just those subjects who actually receive the treatment or the control—introduces a potential bias, because subjects who are lost to follow-up or who refuse treatment are likely to be different in important ways from the other subjects. Because of this rule, before a subject is randomized, the investigator should be absolutely certain that the patient is eligible, has given informed consent, and can be followed for the length of the study."

This approach— once randomized, always analyzed—is another way of saying that data should most commonly be analyzed according to "intention-to-treat" rules. Analysis of "protocol-compliant" patients may be acceptable but usually only as a supplement to intention-to-treat, not instead of it.

Adverse Events and Safety

A recent article in the *Journal of the American Medical Association*[8] clearly indicated that many articles give short shrift to the reporting of adverse events. They note that "An evaluation of safety reporting in randomized trials across seven different medical areas proves that safety reporting is often inadequate and neglected . . . with one exception, safety reporting takes less than a half page in the average trial report; at least as much space is taken by the listing of the names and affiliations of the trial contributors and authors."[8]

In many cases, adverse events may be uncommon or of minor medical importance. In other cases, underreporting may be unintentional— or in the worst case, may represent active concealment of problems that are not thought to be "important" but might compromise acceptance of the treatment. The recent experience of the anesthesia community with rapacuronium—which was found only after release to be associated with an unacceptably high incidence of sometimes fatal bronchospasm[9–12]—reinforces the importance of clear and dispassionate reporting even of "minor adverse events."

Authors should recognize that small clinical studies might easily miss rare but nevertheless catastrophic problems. Even if a study showed no complications, a conclusion that the treatment is "safe" may be unwarranted if the study is not adequately powered. If clinicians would be unwilling to accept a major complication rate of 1 in 100 cases, then a conclusion that a treatment is safe would need to show no complications in more than 300 subjects to convincingly rule out an incidence of 1%. A conclusion that a treatment is "safe" should be made cautiously, and only when it can be statistically supported.

As the Editor-in-Chief of ANESTHESIOLOGY, I believe that much more careful attention needs to be paid by our authors and reviewers—and our readers—to these and other aspects of study design and reporting. Interested individuals (including most authors) are encouraged to read any of several texts devoted to clinical trial design and conduct; a particularly useful, short (and quite readable) book is *Fundamentals of Clinical Trials* by Friedman, Furberg, and DeMets.[1] We have also posted a copy of Dr. Browner's article and have provided a link to the CONSORT documents and checklist in our Guide for Authors on the Journal's Web site (www.anesthesiology.org). In addition, the National Institutes of Health Clinical Center has developed an on-line clinical research training program that is available to any interested investigator (http://www.cc.nih.gov/ccc/cr/training.html).

I should note that the CONSORT statement contains a checklist and flow diagram. While authors are encouraged to use this checklist, there is no intent to enforce rigid adherence to it; the complexity of clinical research may make certain items inapplicable. The purpose of providing these materials is to encourage investigators to consider them when *designing* their studies and preparing their manuscripts, to aid reviewers (and editors) in ensuring that certain key features have not been overlooked and to help readers better determine the quality of a study. The greater goal is to ensure that all of the articles we publish meet the high standards that our readers have come to expect of ANESTHESIOLOGY.

Michael M. Todd, M.D., Editor-in-Chief, ANESTHESIOLOGY, and Professor of Anesthesia, University of Iowa College of Medicine, Iowa City, Iowa. anesthesiology@uiowa.edu

References

1. Friedman LM, Furberg, CD, DeMets DL: Fundamentals of Clinical Trials, 3rd edition. New York, Springer-Verlag, 1998

2. Pua HL, Lerman J, Crawford MW, Wright JG: An evaluation of the quality of clinical trials in anesthesia. ANESTHESIOLOGY 2001; 95:1068–73

3. Norris EJ, Beattie C, Perler BA, Martinez EA, Grass JA, Meinert CL, Andersen GF, Sakima NT, Gorman R, Achuff SC, Minken S, Williams M, Traystman RJ: A double-masked randomized trial comparing alternative combinations of intraoperative anesthesia and postoperative analgesia in abdominal aortic surgery. ANESTHESIOLOGY 2001; 95:1054–67

4. Browner WS: Clinical research: A simple recipe for doing it well. ANESTHESIOLOGY 1994; 80:923–8

5. Begg CB, Cho MK, Eastwood S, Horton, R, Moher D, Orkin I, Pitkin R, Rennie D, Schulz K, Simel D, Stroup D: Improving the quality of reporting of randomized controlled trials: The CONSORT statement. JAMA 1996; 276:637–9

6. Moher D, Schulz KF, Altman D for the CONSORT Group: The CONSORT statement: Revised recommendations for improving the quality of reports of parallel-group randomized trials. JAMA 2001; 285:1987–91

7. Meinert C: Beyond CONSORT: The need to improve reporting standards for clinical trials. JAMA 1998; 279:1487–9

8. Ioannidis JPA, Lau J: Completeness of safety reporting in randomized trials: An evaluation of 7 medical areas. JAMA 2001; 285:437–43

9. Goudsouzian NG: Rapacuronium and bronchospasm. ANESTHESIOLOGY 2001; 94:727–8

10. Kron S: Severe bronchospasm and desaturation in a child associated with rapacuronium. ANESTHESIOLOGY 2001; 94:923–4

11. Naguib M: How serious is the bronchospasm induced by rapacuronium? ANESTHESIOLOGY 2001; 94:924–5

12. Meakin GH, Pronske EH, Lerman J, Orr R, Joffe D, Savarese AM, Lynn AM: Bronchospasm after rapacuronium in infants and children. ANESTHESIOLOGY 2001; 94:926–7