

Anesthesiology
2000; 92:928-30
© 2000 American Society of Anesthesiologists, Inc.
Lippincott Williams & Wilkins, Inc.

Randomized and Nonrandomized Clinical Studies

Statistical Considerations

THE article in this issue of ANESTHESIOLOGY by O'Hara *et al.*¹ provides a good opportunity to review different clinical study designs and statistical issues associated with analyzing data from randomized and nonrandomized comparative studies. Statistics are necessary to analyze clinical data because the response to intervention usually varies widely among patients.² Most biostatisticians favor the use of randomized trials to compare interventions. To understand why, consider the question posed by O'Hara *et al.*:¹ whether either of two interventions—regional or general anesthesia—led to greater mortality in hip fracture patients. Ideally one would compare mortality after all patients received regional, and then, after turning back the clock, after the same patients all received general. Such a study design would eliminate the possibility that different outcomes in regional or general resulted from differences inherent to the patients. Of course, this study design is impossible to implement, but it sets a standard for evaluation.

Randomly assigning subjects to regional or general anesthesia comes closest to the ideal situation. Instead of comparing regional and general in the same subjects, a randomized trial compares regional and general in subjects with the same distribution of observed and unobserved risk factors. In other words, in a randomized trial, observed or unobserved risk factors would have the same chance of occurrence in subjects who receive regional anesthesia as in subjects who receive general. In

practice, the observed risk factors may not be allocated exactly the same in subjects who receive regional as in subjects who receive general. The *P* values and confidence intervals take into account the possibility of different allocations of observed and unobserved risk factors.²

In some situations, a satisfactory randomized trial is not feasible. Expanding on Byar,³ some reasons include (1) enrollment of sufficient numbers of patients is too time consuming, (2) the cost or necessary effort is excessive, (3) the time until the endpoint is reached is too long, and (4) investigators would need to confront various ethical issues. O'Hara *et al.*¹ justify an observational study by stating that a large number of subjects would be necessary for a randomized trial. However, an unbiased nonrandomized study would necessitate approximately the same number of subjects to detect the same effect. The underlying reason for not performing a randomized trial of regional anesthesia *versus* general is that enrollment of sufficient numbers of patients is too time consuming or the cost or necessary effort is excessive.

For some clinically important questions for which a single, large, randomized trial is difficult to implement, results from various small randomized trials have been published. If small, randomized trials are performed instead of one large trial, one can increase power relative to a single small trial by using a meta-analysis, which is a weighted average of the estimates from each trial. A large, carefully conducted randomized trial is generally preferable to a meta-analysis because a meta-analysis can give misleading results if some trials are conducted poorly or if the interventions are very different. However, when the interventions are reasonably similar, a good meta-analysis can provide useful information. We performed a meta-analysis⁴ of regional *versus* general anesthesia using the nine studies analyzed by Parker *et al.*,⁵ along with three other studies.⁶⁻⁸ The endpoint was a 1-month mortality, if reported; otherwise it was a 1-week or in-hospital mortality. The estimated difference in the probability of short-term mortality between general and regional anesthesia was 1.5%, with a 95% confidence interval of -0.6%, 5.4%. To put this result into perspective, apply-

This Editorial View accompanies the following article: O'Hara DA, Duff A, Berlin JA, Poses RM, Lawrence VA, Huber EC, Noveck H, Strom BL, Carson JL: The effect of anesthetic technique on postoperative outcomes in hip fracture repair. ANESTHESIOLOGY 2000; 92:947-57.

Accepted for publication November 9, 1999.

Key words: Anesthesia; general; observational studies; paired availability design; regional.

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.

ing the adjusted odds ratio results of the study by O'Hara *et al.*¹ to a baseline mortality rate of 4.8%, the estimated difference in the probability of 1-month mortality rate between general and regional was 0.4%, with a 95% confidence interval of -0.8%, 1.8%. Thus, both approaches give the same conclusion of no effect of regional *versus* general anesthesia on short-term mortality, although the meta-analysis suggests that the short-term mortality rate may be slightly higher with general.

Because of the study by O'Hara *et al.*¹ is observational, how confident can we be in the results? The difficulty interpreting data from a study without random allocation to regional or general anesthesia is that the type of patient who receives regional may differ from the type of patient who receives general. Instead of evaluating the effect of regional *versus* general anesthesia, one is evaluating regional in one type of patient *versus* general in another type of patient. Another way of looking at the problem is that there is a risk factor for mortality that could occur more frequently than by chance in subjects who receive regional anesthesia than in subjects who receive general. In this case, comparing regional and general could give an incorrect result.

To compare the effect of regional and general anesthesia on the mortality in an observational study, O'Hara *et al.*¹ used a logistic regression model to adjust for many baseline risk factors related to intervention and mortality. By including these risk factors in the logistic regression, one can avoid a bias when the risk factors occur more frequently in subjects who receive regional than in subjects who receive general anesthesia. O'Hara *et al.*¹ did well to include demographic variables, laboratory results, cointerventions, and types of surgery. The authors were also very careful to exclude variables, such as blood pressure, that occurred during or after the initiation of anesthesia. Because blood pressure is affected by anesthesia and may predict mortality, its inclusion would increase bias and not eliminate it. A limitation of logistic regression analysis is the assumption of a particular mathematical relation between risk factors and mortality. As a check, by using propensity scores, which do not necessitate this assumption,⁹ O'Hara *et al.*¹ obtained a similar result. However, even the best multivariate adjustment can be biased if it misses an important risk factor related to why a subject receives one intervention and not the another. In the most extreme case, an omitted covariate could lead one to conclude the opposite of the truth in what is known as Simpson's paradox.¹⁰ In a classic article, The Coronary Drug Project Research Group used logistic regression to compare mortality be-

tween poor and good adherers to clofibrate and found a statistically significant difference ($P = 0.0001$), even though the randomized trial showed no significant effect.¹¹ As another example, using propensity scores, Lieberman *et al.*¹² found a significant effect of labor epidural analgesia on the probability of cesarean section, but a meta-analysis of randomized trials indicated no significant effect.¹³ Sometimes a multivariate adjustment can give the same result as a randomized trial.^{9,14,15} Thus, the level of confidence in multivariate adjustments depends on how strongly one believes that all baseline risk factors related to intervention and mortality have been included.

Another type of nonrandomized clinical study involves historical controls. The traditional method of using historical controls compares outcome in a previous group that received treatment A with outcome in a current group that is receiving treatment B. The major problem is that the criteria for selecting patients to receive treatment A may differ from the criteria for selecting patients to receive treatment B.¹⁶ To reduce this selection bias with historical controls, Baker and Lindeman¹⁷ proposed the paired availability design in the context of estimating the effect of epidural analgesia on the rate of cesarean section (C/S). In hospitals with a sudden change in the availability of epidural analgesia, one compares the rate of cesarean section before and after the increased availability of epidural analgesia among all eligible subjects, not just among those who received epidural analgesia after the change *versus* no epidural analgesia before the change. Applying the method to data from 11 hospitals with a change in the availability of epidural analgesia, Baker¹⁸ obtained a point estimate similar to that from randomized trials.

In summary, the comparison of interventions using logistic regression from a large database is typically much more difficult and less definitive than the analysis of data from a randomized trial because of the need to identify all important risk factors. For nonrandomized studies, the paired availability design for historical controls represents a new approach that may have less bias. For further nontechnical reading, see the References section and the book *Nonrandomized Comparative Clinical Studies*, edited by U. Abel and A. Koch, which is available at <http://www.symposium.com/nrccs>.

Stuart G. Baker, Sc.D.

Mathematical Statistician,

Biometry Research Group,

Division of Cancer Prevention, National Cancer Institute

Bethesda, Maryland 20892-7354
sb16i@nih.gov

Karen S. Lindeman, M.D.

Associate Professor
Department of Anesthesiology/Critical Care Medicine
The Johns Hopkins Medical Institutions,
Baltimore, Maryland
klindema@jhmi.edu

References

1. O'Hara DA, Duff A, Berlin JA, Poses RM, Lawrence VA, Huber EC, Noveck H, Strom BL, Carson JL: The effect of anesthetic technique on postoperative outcomes in hip fracture repair. *ANESTHESIOLOGY* 2000; 92:947-57
2. Green SB: Patient heterogeneity and the need for randomized clinical trials. *Control Clin Trials* 1982; 3:189-98
3. Byar D: Why data bases should not replace randomized clinical trials. *Biometrics* 1980; 36:337-42
4. Follman DA, Proshan MA: Valid inference in random effects meta-analysis. *Biometrics*, 1999; 732-7
5. Parker MJ, Urwin SC, Handoll HHG, Griffiths R: General versus spinal/epidural analgesia for hip fractures in adults, Issue 4 (Cochrane review). Oxford, The Cochrane Library, 1993. Update Software
6. Bode RH, Lewis KP, Zarich SW, Pierce ET, Roberts M, Kowalchuk GJ, Satwicz PR, Gibbons GW, Hunger JA, Espanola CC, Nesto RW: Cardiac outcome after peripheral vascular surgery: Comparison of general and regional anesthesia. *ANESTHESIOLOGY* 1996; 84:3-13
7. Cook PT, Davies MJ, Cronin KD, Moran P: A prospective randomised trial comparing spinal anaesthesia using hyperbaric cinchocaine with general anaesthesia for lower limb vascular surgery. *Anaesth Intensive Care* 1986; 14:373-80
8. Christopherson R, Beattie C, Frank SM, Morris EJ, Meinert L, Gottlieb SO, Yates H, Rock P, Parker SD, Perler BA: Perioperative morbidity in patients randomized to epidural or general anesthesia for lower extremity vascular surgery. Perioperative Ischemia Randomized Anesthesia Trial Study Group. *ANESTHESIOLOGY* 1993; 79(3):422-34
9. Rubin DB: Estimating causal effects from large data sets using propensity scores. *Ann Intern Med* 1997; 127:757-63
10. Green SG, Byar DB: Using observational data from registries to compare treatments: The fallacy of omnimetrics. *Stat Med* 1984; 3:361-70
11. The Coronary Drug Project Research Group: Influence of adherence to treatment and response of cholesterol on mortality in the coronary drug project. *N Engl J Med* 1980; 303:1038-41
12. Lieberman E, Lang J, Cohen A, D'Agostino R, Datta S, Frigoletto F: Association of epidural analgesia with cesarean delivery in nulliparas. *Obstet Gynecol* 1996; 88:993-1000
13. Halpern SH, Leighton BL, Ohlsson A, Barrett JFR, Rice A: Effect of epidural vs parenteral opioid analgesia on the progress of labor. A meta-analysis. *JAMA* 1998; 280:2105-10
14. Horwitz RJ, Viscoli CM, Clemens JD, Sadock RT: Developing improved observational methods for evaluating therapeutic effectiveness. *Am J Med* 1990; 89:630-8
15. Abel U, Koch A: The role of randomization in clinical studies: Myths and beliefs. *J Clin Epidemiol* 1999; 52:487-97
16. Doll R, Peto R: Randomized controlled trials and retrospective controls (letter). *BMJ* 1980; i:44
17. Baker SG, Lindeman KS: The paired availability design: A proposal for evaluating epidural analgesia during labor. *Stat Med* 1994; 13:2269-78
18. Baker SG: The paired availability design for strengthening inference from historical controls: Generalization and validation. Paper presented at: Joint Statistical Meetings; August 8-12, 1999; Baltimore, MD