

Anesthesiology
66:723-724, 1987

Inference, Generalizability, and a Major Change in Anesthetic Practice

PROPERLY CONDUCTED CLINICAL trials provide the only reliable basis for evaluating the comparative efficacy of treatments. A clinical trial is a planned, scientific experiment, using a limited sample of patients to make inferences about how treatment should be given to a general population of future patients. Some other, simpler strategies can provide information about what happens after a treatment is administered, but all lack the strength of the controlled trial because their designs are inherently weaker. They lack strengthening features often found in trials, such as specific eligibility criteria, random allocation to treatment groups, defined therapies, careful followup, appropriate statistical analyses, and adequate sample size. Weaker substitutes for trials are almost always easier, less expensive, and "more practical" than a carefully executed clinical trial. Their inherently weaker designs never offer the strength of inference provided by a well-controlled trial.

In a trial, we seek evidence for a causal link between treatments and observed outcomes. Because the controlled trial depends on an argument based on exclusion (*i.e.*, no other causes or differences affected the experimental groups), we strengthen its inference by taking steps to exclude any such differences. The strength of a trial lies principally in the strength of its control.

In this issue of ANESTHESIOLOGY, Yeager *et al.*¹ report the results of a randomized, controlled clinical trial to evaluate the effect of epidural anesthesia and postoperative epidural analgesia in a group of high-risk surgical patients compared with a control group who received their institution's standard general anesthetic and analgesic techniques, without the use of epidural drugs.

The authors report statistically significant reductions

in the epidural group's postoperative complication rate, frequency of postoperative cardiovascular failure, and frequency of major postoperative infectious complications. Four of 25 patients in the standard treatment group died postoperatively. None of those receiving epidural anesthesia died.

These results are striking. Even though the sample sizes are modest, the findings suggest that significant improvement in outcomes might be made by using epidural anesthesia and analgesia for high-risk surgical patients. Such massive effects are unusual when therapies are carefully tested by randomized controlled trials. There is evidence that, at least for survival, gains of 10% or more occur in less than one-tenth of published trials.²

The primary concern in interpreting the results of this study is that this trial is a randomized study with a relatively small number of patients in each group. There remains the clear possibility that the control of potentially confounding factors normally achieved with randomization of a sufficiently large number of subjects has not been reached in this instance. The beauty of randomization as a technique for allocation to treatment is that, with a large enough number of patients in a trial, a truly random assignment assures that confounding factors (both identified and unidentified) are evenly distributed among the resulting groups. We would not raise this question if, rather than 28 patients and 25 patients per group, the study reported the results of an analysis of 128 or 125 patients per group. We wonder whether the luck of randomization has assigned a set of sicker patients to the control group, and whether this assignment has added to any outcome difference between this set of patients and those who received epidural anesthesia. Do the observed differences between the epidural group and the standard anesthesia group result from a true treatment difference, or some bizarre combination of confounding factors? Because the authors found an overall complication rate and complication intensity strikingly higher in the standard treatment patients, and believed this to be due to a true treatment effect, they felt it necessary to terminate their study.

Accepted for publication February 9, 1987.

Address reprint requests to Dr. McPeck: Department of Anaesthesia, Massachusetts General Hospital, Boston, Massachusetts 02114.

Key words: Analgesics. Analgesics: epidural. Anesthetic techniques: epidural. Complications. Surgery: postoperative morbidity.

Small trials, (*i.e.*, trials with relatively small numbers of subjects in each treatment group) cause difficulty in interpretation. Usually, this difficulty stems from a small trial's relatively weak power to detect differences between treatment groups. In this instance, the authors have detected a difference. Here, our concern is not with the trial's power, not that it may have missed an important true difference, but, rather, that it may have detected a difference where, in fact, no difference exists. This error, sometimes called a Type I or alpha error, would most likely be caused, in a small randomized trial, by initial inequality of the two groups. We see no convincing evidence to suggest such an error has occurred in this trial, but, with sample sizes of 25 and 28, our concerns remain.

If one assumes that the authors are correct and a real treatment difference exists, we, as readers, must take a further step. We must decide whether these results are generalizable to the patients we each see in our own practices.

Many who write about clinical trials draw a distinction between a trial designed to establish efficacy and one designed to establish effectiveness in clinical practice. These writers define a treatment's efficacy as the outcome observed in a specific population of patients who receive a narrowly defined treatment. Effectiveness is defined as the outcomes likely to be achieved when a treatment is introduced into wider clinical practice. If one narrowly defines the subjects admitted to the trial and the treatments applied, one sharpens the inference and increases the possibility that a real treatment effect will be uncovered. Yet, by limiting the patients involved and the treatments offered too narrowly, investigators impede generalizability. This tension between desirable homogeneity with resulting control and narrowness of inference is pitted against the desirability of heterogeneity of population and treatment yielding greater generalizability. In general, Yeager *et al.*'s design leans toward ease of generalization.

Nevertheless, generalizability remains a problem here. It is a problem each of us must consider as we decide how to use the authors' conclusions in our own work. Do the

patients the authors describe sound like patients from our own practices? Does the general anesthesia given the standard treatment group in New Hampshire sound like the general anesthetic technique we use? Do the postoperative outcomes, the length of postoperative intubation and ventilation, and postoperative complications sound like those we observe in our own practice with high-risk surgical patients? These are difficult questions, and we must each answer them for ourselves. Would the institution of epidural anesthesia and analgesia as used by the authors be feasible in our practices? Could we actually deliver the treatment? Physicians will best determine appropriate anesthetic choice by considering, not only the authors' findings, but also the local experience of the anesthetist and surgeon and their facilities for caring for critically ill patients.

Some physicians will doubtless want to move ahead to adopt, in their own practices, the epidural anesthesia and analgesia regimen used by the present authors. Others will be more cautious and will wait until these conclusions have been confirmed in another trial; not because they distrust the authors' work, but simply because of the small size of the trial, they want a broader base for making such a major change in anesthetic practice.

BUCKNAM MCPEEK, M.D.
Associate Professor of Anaesthesia
Harvard University
Anesthetist
Massachusetts General Hospital
Boston, Massachusetts 02114

References

1. Yeager MP, Glass DD, Neff RK, Brinck-Johnsen T: Epidural anesthesia and analgesia in high risk surgical patients. *Anesthesiology* 66:729-736, 1987
2. Gilbert JP, McPeck B, Mosteller F: Progress in surgery and anesthesia. Benefits and risks of innovative therapy, Costs, Risks and Benefits of Surgery. Edited by Bunker JP, Barnes BA, Mosteller F. New York, Oxford University Press, 1977, pp 124-169

Anesthesiology
66:724-726, 1987

Neurotoxicology of Spinal Agents

DIRECT DELIVERY OF PHARMACOLOGIC agents into the central nervous system promises to achieve enhanced

Accepted for publication January 27, 1987.

Address reprint requests to Dr. Coombs: Department of Surgery (Anes), Dartmouth-Hitchcock Medical Center, Hanover, New Hampshire 02755.

Key words: Analgesia. Anesthetics, local: bupivacaine. Injections, intrathecal. Pain, intractable.

neuropharmacologic activity through avoidance of the blood-brain barrier. Worldwide recognition of this fact has led to clinical introduction of prolonged spinal delivery of a variety of analgesics, including narcotics, alpha 2 agonists, and peptides.¹⁻⁴ In fact, chronic intraventricular and spinal drug infusion has rapidly burgeoned to now include non-cancer pain patients,⁵ neurologic patients with spasticity,⁶ amyotrophic lateral sclerosis,⁷ and Alzheimer's