CORRESPONDENCE

Anesthesiology
79:1155, 1993

Negative Inferences about Rare Events Require Large Samples

To the Editor:—Todd *et al.* compared three partially overlapping anesthetic combinations in a series of neurosurgery patients ($n < 40$ per group) and concluded that "the specific choice of anesthetic agent(s) may not be the most crucial aspect of successful neuroanesthetic practice."[1] It would be difficult to find an anesthesiologist who would have disagreed with this tentative conclusion before considering Todd and coauthors' results.

A more engaging conclusion is implied by the authors' penultimate summary statement: "Our data also support the belief that, despite the presumably different effects of these three anesthetic techniques on cerebrovascular physiology, only small differences in ICP were observed, and these differences were of minimal relevance in these elective patients." The inference here is that a statistically significant fourfold increase in the incidence of patients with intracranial pressures (ICP) greater than 24 mmHg, accompanied by "a greater degree of swelling at time of craniotomy" and a statistically significant twofold increase in recovery time, in the isoflurane/N_2O group, "is of minimal relevance."

When statistically significant findings of substantive magnitude regarding clinically relevant variables are diminished, if not dismissed, readers need to be especially wary about inferences drawn from results that were not statistically significant but would be of clinical concern if validated—especially when the events measured are known to be infrequent and sample sizes are too small to provide statistical power (probability of a false negative <0.20). For example, if Todd *et al.* had obtained the same proportionate results with sample sizes doubled, even their single-measurement *average* ICP differences would have been statistically significant (as distinct from differences in incidence of high ICP, which are of more clinical concern, and which *were* statistically significant). If sample sizes were tripled, their observed differences in frequency of mannitol use would have been statistically significant. And if sample sizes had been five times larger (still on the small side for making negative inferences about rare events), their observed twofold increase in "brain severely swollen, specific therapy required" patients (propofol/fentanyl and isoflurane/N_2O groups compared to the fentanyl/N_2O group) would have been statistically significant.

If only differences in incidence of "new or worsened neurologic deficit" warrant inferences about anesthetic selection for neurosurgery, any negative result derived from fewer than 200 patients per group should not be considered for publication, and a study that contains fewer than 100 patients per group should be considered not to have found statistically significant results by design. Todd *et al.* reported their results on neurologic deficits by stating, "Thirty-five of 121 patients (29%) had some new or worsened neurologic deficit when examined 24 h postoperatively, but there were no differences among anesthetics." Does this mean that no *statistically significant* differences were found, or does it mean no differences were *observed*? Was it a 12–12–11 split among the three groups? If not, was it a counterintuitive result in which the opioid/N_2O group—the group that had lowest average ICP, fewest ICPs >24, and fewest severely swollen brains—had the most neurologic deficits?

John Hartung, Ph.D.
Research Associate Professor

James E. Cottrell, M.D.
Professor and Chairman
Senior Associate Dean for Clinical Practice

Department of Anesthesiology
State University of New York Health Science Center
450 Clarkson Avenue
Brooklyn, New York 11203

Reference

1. Todd MM, Warner DS, Sokoll MD, Maktabi MA, Hindman BJ, Scamman FL, Kirschner J: A prospective, comparative trial of three anesthetics for elective supratentorial craniotomy. ANESTHESIOLOGY 78:1005–1020, 1993

(Accepted for publication August 18, 1993.)

Anesthesiology
79:1155–1156, 1993

In Reply:—In 1979, Eger noted in an editorial for the Journal that "long ago my father warned me that I could not disprove the existence of dragons."[1] There are many dragons in medicine, and some of these are believed to threaten the castle we know as neuroanesthesia. One concerns drug selection for elective intracranial surgery, and despite a lack of any substantive clinical evidence suggesting that one anesthetic(s) is better or worse than another, some continue to argue that we should or should not use particular agents during elective neurosurgery. Some might accuse us of violating the above caveat when we state that we could not find important intergroup differences, and Hartung and Cottrell suggest that our results fail to provide conclusive proof that differences do not exist. We cannot disagree with this premise, but believe that more important issues are involved.

Their letter focuses on the fact that "power analysis" suggests that

larger groups *might* have yielded significant intergroup differences that were not evident in our study. For example, the differences in average intracranial pressure might have been significant with groups of 80 patients each; or differences in mannitol use might have appeared with groups of 120 patients each (which is highly unlikely, given the noted pattern of mannitol administration); or groups of 200 patients each might have revealed differences in the incidence of severe swelling. They go so far as to argue that "any negative result derived from fewer than 200 patients per group should not be considered for publication." Unfortunately, these authors overlook the possibility that a larger trial might also *fail* to show any differences, and they ignore the enormous financial and human resources that would be needed. It is a trivial matter to state that someone needs to do a bigger trial; it is not a trivial matter to undertake one. This raises the issue of whether it is worth engaging in the multimillion dollar, multicenter trial (funded by whom?) to answer their challenges. In one way, our study can be viewed as a pilot. At the conclusion of any pilot trial, the investigators must decide whether it would be worthwhile to continue with a larger study. We had many discussion on this issue, and we firmly decided that the data so far collected did not warrant the effort and cost that would be needed. If Hartung and Cottrell would like to undertake this large study, we would gladly participate and provide them with any insights we might have gained. In the meantime, we have chosen to pursue questions that we believe are far more important.

We carried out what we believe to be the largest prospective comparative trial yet dealing with the issue of anesthetic selection in neurosurgery. We presented the results in an exhaustive fashion (perhaps excessively so) to allow readers to draw their own conclusions, which may be different than our own. Because we realize the limitations of such a trial (including those engendered by its relatively small size), we were extremely cautious about interpretation. In the absence of any other evidence, we believe that our goal should be to disprove the null hypothesis. We could not do this in most cases and therefore concluded that there were no important differences among the three anesthetics. We also realize that our clinical impressions, as well as our data, probably played some role in these conclusions, but we stand by those conclusions. More importantly, we believe that the burden of proof now resides with those who suggest that one of the commonly used anesthetic agents is indeed "better or worse" for the brain than another.

Michael M. Todd, M.D.
David S. Warner, M.D.
Martin D. Sokoll, M.D.
Bradley Hindman, M.D.
Mazen Maktabi, M.D.
Franklin L. Scamman, M.D.
Jerry Kirchner, B.S.
Department of Anesthesia
University of Iowa, College of Medicine
Iowa City, Iowa 52242

Reference

1. Eger EI II: Dragons and other scientific hazards. ANESTHESIOLOGY 50:1, 1979

(Accepted for publication August 18, 1993.)

Anesthesiology
79:1156–1157, 1993

Anterior Spinal Artery Syndrome?

To the Editor:—Parris and Kirshner recently reported an unfortunate case of long-lasting motor paralysis of the lower extremities after lumbar sympathetic block.[1] In this case, lumbar sympathetic block at L2 was abandoned because of arterial blood flow from the spinal needle. It was reattempted at L3, but motor and sensory disturbances of both lower extremities occurred, preceded by lower back pain radiating down both legs immediately after the injection of local anesthetic. The authors speculated that arteriospasm resulting from inadvertent puncture of the anterior spinal artery or the artery of Adamkiewicz may have produced progressive ischemia to the anterior segment of the spinal cord, presenting as anterior spinal artery syndrome.

However, there are a few problems with that speculation. First, motor and sensory disturbances followed a series of events after the injection of local anesthetic, and not the puncture of the artery. It is more reasonable to speculate that motor and sensory disturbances resulted from the injection of local anesthetic. Second, in most reported cases of anterior spinal artery syndrome, pain in the region fed by these arteries did not precede the motor and sensory disturbances, as in this case.[2–5]

Parris and Kirshner did not describe how the position of the needle tip was ascertained. If radiologic aid had not been used, the needle tip might have been malpositioned. We think that it is necessary to consider another possibility: local anesthetic may have been injected into the spinal cord directly or through a peripheral nerve, which would have caused severe pain and spinal cord damage,[5,6] resulting in the intraspinal lesion observed in this patient.

Takumi Nagaro, M.D.
Associate Professor of Anesthesiology and Resuscitology

Tatsuru Arai, M.D.
Professor and Chairman of Anesthesiology and Resuscitology

Ehime University School of Medicine
Shigenobu-cho, Onsen-gun, Ehime-ken
791-02, Japan