

CORRESPONDENCE

Anesthesiology
79:1155, 1993
© 1993 American Society of Anesthesiologists, Inc.
J. B. Lippincott Company, Philadelphia

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 neuro-anesthetic 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/N₂O 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/N₂O groups compared to the fentanyl/N₂O group) would have been statistically significant.

Anesthesiology
79:1155–1156, 1993
© 1993 American Society of Anesthesiologists, Inc.
J. B. Lippincott Company, Philadelphia

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

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/N₂O 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.)

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