Anesthesia and Outcome after Neonatal Surgery

The Role for Randomized Trials

Editor's Note: This is the second in a three-part series of Editorial Views regarding design of clinical trials to address the effect of anesthesia on the developing brain. Animal studies have suggested that anesthetic exposure could affect neurocognitive development, and there is an urgent need for clinical trials to determine whether this effect occurs in humans. This series presents the opinions of three world thought leaders in the possible designs of such clinical trials.

James C. Eisenach, M.D., Editor-in-Chief

SEVERAL studies in young rodents and primates have identified anesthetic-induced neuronal cell death.[1] There is also evidence that surgery in some human infant populations is associated with poor neurodevelopment outcome.[2,3] The relevance of these findings to clinical anesthetic practice is an important and controversial question—a question unlikely to be answered with a single study. To address this issue, various studies with various designs have been proposed. In this editorial, we describe a randomized trial that we and colleagues are conducting. Randomized trials provide strong clinical evidence but are not without limitations and challenges. These include the ethics of randomization, choice of comparison group, choice of relevant dose in the treatment arm, definition of suitable outcome measure, and protocol adherence. By describing this trial, we illustrate not only the advantages and rationale for randomized trials, but also some of the problems in general for such trials in infant populations.

The concern for neurotoxicity of anesthesia has arisen in an unusual way. Rather than clinical findings prompting animal studies, the question arose from almost incidental findings in the laboratory, after many decades of seemingly unremarkable and routine anesthesia in neonates. How could clinically important neurotoxicity have been missed? The first relevant point is that many anesthetics have not been adequately studied in neonates and are often not licensed for such use. The responsibility to correct this sad state of affairs must be shared by pharmaceutical companies, regulatory authorities, and indeed our own profession. Second, it may be difficult to identify an association when there is a long lag time between exposure and measurable outcome, particularly in the presence of important confounding factors.

Although further laboratory work remains to be done to elucidate the agents, doses, and growth periods that may pose greatest risk, as well as a greater understanding of the mechanisms of toxicity and possible methods to reduce such toxicity, clinical studies must also be done to determine the clinical relevance and reliability of any laboratory findings. There are numerous examples of laboratory studies failing to translate into clinical practice.[4]

Clinical studies could use observational cohort or randomized controlled trial designs. The former could involve examining cohorts of children previously anesthetized at particular developmental stages and comparing their neurocognitive development with that of a control group or an appropriate standard. This approach has recently been reviewed by Sun *et al.*[5] in ANESTHESIOLOGY. One attraction of this approach is that an answer can be obtained relatively quickly, but it has the weakness that however well the comparison group is chosen, it is impossible to be certain that there are not recognized or unrecognized confounding variables that are more common in one group than the other, resulting in group differences that are not solely dependent on the variable of interest, in this case anesthesia. Unfortunately, there are several likely confounding variables favoring worse outcome in the children exposed to anesthesia. Children have anesthesia for the treatment or diagnosis of a particular medical or surgical condition. The particular condition, other conditions or syndromes associated with the condition, and other treatments, as well as the direct effects of the surgery or investigation itself, may all influence developmental outcome. It may be possible to reduce the potential bias in the association between exposure to anesthesia and outcome, by matching the control group or adjusting statistically for some of these factors, but matching and adjustment are always imperfect, and it is impossible to control for unknown confounding factors.[6–8] Nevertheless, cohort studies are still a useful first step. If no association is found, that is some evidence against clinically relevant toxicity. However, if some associations are found, the exact causation cannot be determined and stronger levels of evidence are desirable.[7] Large randomized controlled trials are the best

---

941

way to reduce the influence of known and unknown confounding variables, thus providing the strongest evidence for causation, linking exposure to outcome.[7] The larger the sample size is in a randomized trial, the less likely it is that confounding can produce major bias.[7,8] A randomized trial should in theory avoid many of the limitations of a cohort study.

Randomized trials involve the random allocation of patients to different treatment groups. This raises particular ethical issues, especially in infants. It has been suggested that randomly assigning a subject to one treatment or another is only ethical if there is equipoise, where the balance of evidence suggests that no difference exists between treatment groups.[9] However, this is contentious. When planning a trial, we often hope or expect that one treatment is better, and rarely do we expect there to be no difference. This possible challenge to equipoise may be ethically balanced if, for humanitarian reasons and with informed consent, a subject chooses to participate in a randomized trial and accept the perceived risk of receiving only the standard treatment (or placebo), or the subject places equal weight to the comfort of a known standard treatment *versus* what is potentially better but also unknown and potentially ineffective or risky. Infants cannot give such informed consent, and in many countries, laws or ethical guidelines specifically limit the capacity of others to make these decisions for them. In infants, we have to be more certain that equipoise is present. As an interesting aside, this poses the ethical contradiction that clinical trials may be hindered by the ethical requirement for equipoise, while it is also unethical to disadvantage this particular population by discouraging clinical trials.

For a trial, there must be a comparison. The comparison group could be a placebo or another form of treatment. To answer the specific question of whether general anesthetic is neurotoxic, a placebo would be the best comparator. Because surgical stress may have an influence on outcome and because anesthesia may alter the impact of surgical stress on outcome, conducting the trial without surgery would provide the clearest answer to the question of anesthesia neurotoxicity. However, such a design is not possible. It would be unethical to expose infants to the risk of anesthesia for no reason, and it would be unethical to expose children to surgery with no anesthesia. Therefore, the comparison group must be another form of anesthetic in the setting of surgery. We are using sevoflurane general anesthesia as the treatment group and awake spinal anesthesia as the control. Awake spinal anesthesia uses none of the agents implicated in neurotoxicity and is an established and commonly used technique for inguinal hernia repair in neonates. Spinal anesthesia has often been preferred because of the respiratory complications of general anesthesia. With newer general anesthetics, these benefits of spinal anesthesia are less clear, and because the

evidence that general anesthesia actually causes harm is tenuous, we believe that there is a strong case for equipoise.

Using spinal anesthetic as a comparison rather than placebo and performing the trial in the setting of surgery has important limitations when trying to determine the actual effect of the treatment of interest (general anesthesia).[10] These limitations occur when there is reason to suspect that the control (spinal anesthesia) may itself directly influence outcome. This would be so if spinal anesthesia were to cause neurologic injury, or conversely if spinal anesthesia were to be a more (or less) effective way to reduce some other form of injury due to the surgery. Therefore, with respect to the issue of neurotoxicity, finding no difference between spinal and general anesthesia could mean either no toxicity from either technique or equal toxicity of both techniques; and finding a difference could mean no toxicity from one technique and toxicity from the other, or toxicity from both with different degrees of toxicity, or different degrees in protection from the effect of surgery. If evidence emerges to suspect local anesthetic toxicity, the trial results should be interpreted accordingly. From the above discussion, it can be seen that the randomized trial design has potential limitations in its ability to directly answer whether general anesthetics are neurotoxic in humans, but from a clinical perspective, the trial is still far from useless. Regardless of the issue of direct effect of local anesthetic, the trial can be important clinically as a comparison of two techniques. Given that there is an association between surgery and poor outcome in some infant populations, the trial may still indicate which anesthetic gives better outcome or whether there is no difference between them. This will guide clinical practice independently to the issues of neurotoxicity.

A randomized trial can use a superiority or equivalence design.[9] A superiority design tests whether one treatment is clinically better than another. In an equivalence design, two treatments are compared and analysis is performed to exclude a predefined difference between the two treatment groups. Such trials cannot actually demonstrate equivalence; they allow investigators to exclude a predefined difference between the treatment groups with a predefined level of confidence. The choice of design is driven by the aim of the study. We consider that the aim of the study is to determine whether general anesthesia is as safe as spinal anesthesia. The broader implication being that exposure to general anesthesia produces no clinically relevant difference in neurodevelopmental outcome compared with spinal, and assuming that the risk from spinal is small, general anesthesia may be used in populations without having to consider potentially riskier options that involve avoiding general anesthesia. The proposal is therefore for an equivalence design. The study will enroll at least 598 infants in total so that if the two methods of anesthesia

really are close to equivalent and we assume an expected difference of only 1 IQ point, there is a 90% chance that a 95% confidence interval will exclude a difference of more than 5 IQ points.

In many respects, the differences between a superiority design and an equivalence design are subtle. A superiority trial can still provide evidence for equivalence, and an equivalence trial can provide evidence for superiority. One practical difference is that an equivalence trial follows a per-protocol analysis, and the trial is weakened if substantial numbers do not follow the protocol. For this reason, the trial only includes infants where spinal anesthesia is highly likely to be effective and the family is highly likely to return for follow-up. This limits the numbers of suitable subjects at each site, and hence the trial is run at multiple sites (each of which has a track record of effective spinal anesthesia). Criteria for inclusion may vary subtly between sites, so to reduce potential random bias, the randomization is stratified by site. Multicenter designs have the added advantage of providing a broad population and hence the opportunity to identify other patient, clinician, and institutional factors that may influence outcome.[8]

It is difficult to infer from animal data the dose of or duration of exposure to general anesthetic needed to produce neuronal injury in humans. In the trial, we chose a relatively brief exposure that matches the duration of spinal anesthesia. There are limitations wherever the dose is set. If equivalence is demonstrated at a particular exposure, we can make no assumptions about the toxicity that might appear with a longer exposure, and if we find a difference, we can make no assumptions that toxicity will persist with shorter exposure.

A crucial component of any trial, or indeed any research question, is the outcome measure. As Sun *et al.* have indicated, it is difficult to know which neurodevelopmental outcome is most likely to be altered by any anesthesia toxicity. Laboratory work can guide this choice, but the link between animal work and human neurodevelopmental tests remain tenuous. Choosing a single primary outcome is important. If it is unclear which single outcome is most likely to be altered, a logical alternative is to choose a primary outcome that is easily tested and regarded as a clinically important outcome. For the trial, IQ at 5 yr of age is the primary outcome. Although defining a primary outcome *a priori* is important, it is still valid to analyze multiple secondary outcomes. In our trial, we will also test a broad range of neurodevelopmental outcomes at 2 and 5 yr of age. At 2 yr of age, we will test each of the three subscales of the Bayley III (cognitive, motor, and language), and tests of adaptive behavior. At 5 yr, we will apply each of the three subscales of the Wechsler Preschool and Primary Scale of Intelligence, third edition; selected NEPSY II subtests; and the Behavioral Rating of Executive Func-

tion, looking at attention, working memory, information processing, and executive function.

Neurodevelopment outcomes after exposure in infancy pose a particular problem. As a child becomes older, the measures of neurodevelopment become more robust and predictive of future development. It may be many years before the best outcome can be measured. This wait may seem extraordinary from an anesthesia perspective, and indeed does pose particular issues if funding agencies prefer results sooner rather than later, but it is not so extraordinary in the context of most scientific endeavors. It may seem a long time for a simple experiment, but not so long compared with the usual programs required to solve major scientific questions. Certainly, waiting 2–5 yr for accurate and strong evidence is far wiser and more cost-effective than compromising design or using misleading and inaccurate measures earlier.

Although neurotoxicity of general anesthesia is currently of great interest, there are many other areas where anesthesiologists may have a role in improving neurodevelopmental outcome after surgery in infants. These include managing fluid, respiratory, cardiovascular, glucose, and pain, as well as the inflammatory and humoral responses to surgery. These may be far more important than neurotoxicity. Few clinical trials have been performed to determine the optimal management of these factors.

A large multicenter randomized trial in neonates is a daunting task. There is, however, a strong precedent for performing randomized trials in neonates; it is not so long ago that Anand *et al.*[11] performed groundbreaking anesthesia randomized trials in neonatal populations, and neonatologists regularly perform large multicenter trials.[12,13] Such trials are expensive, require particular skills and collaborations, and have their limitations, but given the complexity of neurotoxicity and the many other important questions surrounding our management of infants and neurodevelopmental outcome, randomized trials such as this one will provide the best evidence to guide actual clinical practice.

Andrew J. Davidson, M.B., B.S., M.D., F.A.N.Z.C.A.,* Mary Ellen McCann, M.D., M.P.H.,† Neil S. Morton, M.B., Ch.B., F.R.C.A.,‡ Paul S. Myles, M.D., M.P.H., F.A.N.Z.C.A., F.C.A.R.S.C.I.§ *Department of Anaesthesia, Royal Children's Hospital, Melbourne, Australia. andrew.davidson@rch.org.au. †Department of Anesthesia, Children's Hospital, Boston, Massachusetts. ‡Department of Anaesthesia, Royal Hospital for Sick Children, Glasgow, United Kingdom. §Department of Anaesthesia and Perioperative Medicine, Alfred Hospital, Melbourne, Australia.

References

1. Mellon RD, Simone AF, Rappaport BA: Use of anesthetic agents in neonates and young children. Anesth Analg 2007; 104:509–20

2. Surgery and the tiny baby: Sensorineural outcome at 5 years of age. The Victorian Infant Collaborative Study Group. J Paediatr Child Health 1996; 32: 167–72

3. Loepke AW, Soriano SG: An assessment of the effects of general anesthetics on developing brain structure and neurocognitive function. Anesth Analg 2008; 106:1681–707

4. van der Worp HB, de Haan P, Morrema E, Kalkman CJ: Methodological quality of animal studies on neuroprotection in focal cerebral ischaemia. J Neurol 2005; 252:1108–14

5. Sun LS, Li G, DiMaggio C, Byrne M, Rauh V, Brooks-Gunn J, Kakavouli A, Wood A, Co-Investigators of the PANDA Research Network: Anesthesia and neurodevelopment in children: Time for an answer? ANESTHESIOLOGY 2008; 109: 757–61

6. Sackett DL: Bias in analytic research. J Chronic Dis 1979; 32:51–63

7. Collins R, MacMahon S: Reliable assessment of the effects of treatment on mortality and major morbidity, I: Clinical trials. Lancet 2001; 357:373–80

8. Myles P, Rigg J: Anaesthesia and pain, Textbook of Clinical Trials, 2nd edition. Edited by Machin D, Day S, Green S. Chichester, UK, John Wiley & Sons, 2006, pp 519–42

9. Machin D: General statistical issues, Textbook of Clinical Trials, 2nd edition. Edited by Machin D, Day S, Green S. Chichester, UK, John Wiley & Sons, 2006, pp 13–38

10. Temple R, Ellenberg SS: Placebo-controlled trials and active-control trials in the evaluation of new treatments, part 1: Ethical and scientific issues. Ann Intern Med 2000; 133:455–63

11. Anand KJ, Sippell WG, Aynsley-Green A: Randomised trial of fentanyl anaesthesia in preterm babies undergoing surgery: Effects on the stress response. Lancet 1987; 1:62–6

12. Anand KJ, Hall RW, Desai N, Shephard B, Bergqvist LL, Young TE, Boyle EM, Carbajal R, Bhutani VK, Moore MB, Kronsberg SS, Barton BA: Effects of morphine analgesia in ventilated preterm neonates: Primary outcomes from the NEOPAIN randomised trial. Lancet 2004; 363:1673–82

13. Schmidt B, Davis P, Moddemann D, Ohlsson A, Roberts RS, Saigal S, Solimano A, Vincer M, Wright LL: Long-term effects of indomethacin prophylaxis in extremely-low-birth-weight infants. N Engl J Med 2001; 344:1966–72