

Anesthetic-related Neurotoxicity in the Young and Outcome Measures

The Devil Is in the Details

Randall P. Flick, M.D., Michael E. Nemergut, M.D., Ph.D., Kaare Christensen, M.D., Ph.D., D.M.Sc., Tom G. Hansen, M.D., Ph.D.

THERE are few issues facing the field that are more concerning and contentious than the possible neurotoxic effects of anesthetics on children. Although laboratory studies report that virtually all commonly used anesthetics invariably induce neurodegeneration in the developing animal brain, observational studies are less conclusive with some reporting an association between exposure to anesthesia/surgery and adverse neurobehavioral outcome, whereas others do not.¹ Among the many methodologic problems associated with human studies are the outcome measures available to the investigators.^{1,2} As virtually all these studies are retrospective, the outcome is not chosen by the investigator and therefore may not provide the most meaningful measure of the cognitive or behavioral effect. In addition, the various neurocognitive outcomes may or may not be comparable as few studies have reported more than a single end point. In this issue of *ANESTHESIOLOGY*, Ing *et al.*³ have attempted to provide a structured comparison of outcome measures representative of those found in most studies of this type. Similar to their previous publication,⁴ data from the Raine Study, a cohort of 2,868 children born from 1989 to 1992 in Western Australia, were examined for an association between exposure to anesthesia/surgery in children before the age of 3 yr and three different but closely related outcomes including direct neuropsychological



“The lack of an obvious human phenotype for anesthetic neurotoxicity represents a major obstacle to study design and interpretation.”

studies with negative results use broad measures of academic performance such as group tests of achievement (California Achievement Test and Danish standardized test of achievement) and teacher–parent rating scales very similar to that used in this study.^{6–9} Studies using individual tests of cognitive performance have been uniformly positive, commonly in areas of speech and language. The larger studies performed in Europe utilizing group tests (or similar) tend to be negative, whereas smaller studies using individual neurobehavioral tests more frequently are positive.

testing, International Classification of Diseases, 9th Revision (ICD-9)–coded clinical disorders, and a group test of academic achievement. Of the 781 children included, 112 had been exposed to anesthesia/surgery, and among those exposed, the risk of deficits in individual language assessments and ICD-9 codes for language or cognitive disorders was increased. In contrast, exposed and unexposed children did not differ with regard to academic achievement. The authors conclude that these data explain some of the variation in the literature and underscore the importance of the outcome measure when interpreting studies of cognitive function. Similar findings have previously been noted in other studies using more than a single measure of neurodevelopment.⁵

A cursory review of the literature suggests that the majority of

Image: ©Thinkstock.

Corresponding article on page 1319.

Accepted for publication February 28, 2014. From the Department of Anesthesiology, College of Medicine, Mayo Clinic, Rochester, Minnesota (R.P.F., M.E.N.); Institute of Public Health, Epidemiology, University of Southern Denmark, and Department of Clinical Biochemistry and Pharmacology and Department of Clinical Genetics, Odense University Hospital, Odense, Denmark (K.C.); and Department of Anesthesiology and Intensive Care, Odense University Hospital, and Clinical Institute, Anesthesiology, University of Southern Denmark, Odense, Denmark (T.G.H.).

Copyright © 2014, the American Society of Anesthesiologists, Inc. Lippincott Williams & Wilkins. *Anesthesiology* 2014; 120:1303-5

Utilization of ICD-9 codes in epidemiologic research is common as administrative data are widely available and often represent the only source of information related to an outcome of interest. Unfortunately, errors in coding are exceedingly common and represent a source of significant bias.¹⁰ Attention deficit hyperactivity disorder (ADHD) provides an instructive example alluded to by the authors. Ing *et al.* utilized ICD-9 codes as a means of identifying relevant behavioral or cognitive outcomes including ADHD, the diagnosis of which is clearly delineated within the Diagnostic and Statistical Manual of Mental Disorders, 4th Edition. However, in studies of ADHD diagnostic accuracy, only one third of children diagnosed with ADHD have been subject to the Diagnostic and Statistical Manual of Mental Disorders, 4th Edition, criteria and as many as two thirds of children with ADHD have a diagnosed learning disability that may or may not be identified with a specific ICD-9 code.¹¹ It is therefore difficult to be certain whether a child has the outcome of interest (ADHD) or has a similar outcome that may confound the relationship (learning disability). In the case of the study by Ing *et al.*, the problem of mis-coding was magnified by assigning codes from parental reports of childhood illness, rather than medical records, an additional source of potential bias. Ing *et al.* somewhat inaccurately compares ADHD as an outcome in this study with that in the study by Sprung *et al.*¹² The comparison provides an instructive example of how apparently identical outcome measures may differ in profound ways. In the study Sprung *et al.*, ADHD was diagnosed by strict Diagnostic and Statistical Manual of Mental Disorders, 4th Edition, criteria using a robust medical record and unique access to school records—information unavailable to Ing *et al.* In addition, Sprung, but not Ing, was able to separate those children with ADHD alone from those with a learning disability and ADHD to examine the effects of these overlapping cognitive disorders separately. Consequently, the methodology in the study by Ing *et al.* almost certainly overestimates the frequency of ADHD, cannot determine whether the observed differences are truly driven by ADHD, or is the result of confounding between ADHD and learning disability. As such these data should be compared with that in the study by Sprung with great caution, if at all.

The lack of an obvious human phenotype for anesthetic neurotoxicity represents a major obstacle to study design and interpretation. The study by Ing *et al.* is intended in part to identify a robust end point for evaluating existing work as well as designing future studies that may be more informative. The unique feature of the data reported by Ing is the extensive neurodevelopmental testing that was performed repeatedly for each of the studied subjects. No other study to date contains as much cognitive outcome data as this and their previous publication using the same data. In addition to studies from the Mayo Clinic, those by Ing *et al.* are the only extant studies that contain data from individually administered tests of cognition. It is striking that

these studies are both positive and report disproportionate effects on speech and language. Nonetheless, as mentioned above, caution should also be used when interpreting these data as many of the outcomes are interrelated and the use of multiple tests increases the risk of a type 1 statistical error. Noteworthy is the observation that 25% of the exposed comprised children undergoing myringotomies—a population notoriously known to suffer from later language and learning problems.¹³

Ing *et al.* suggest that group tests may lack sufficient sensitivity to detect small differences in performance that may exist between those exposed and those not exposed, but that these minor differences may not be clinically or academically meaningful. They also suggest that studies using large cohorts but insensitive outcomes are likely to be negative and should be interpreted with caution; studies using individually administered tests of cognition may be more likely to be positive and can provide insight into phenotype (*i.e.*, abnormalities in speech and language). However, the value of ICD-9 or other administrative data in this setting as an end point is unclear and awaits the results of studies that examine the correlation between such codes and direct testing depending on location and time. Moreover, studies using comprehensive cognitive testing are laborious and expensive; therefore, the sample size in these studies will invariably be small. If this approach is used more widely in the future, a possible consequence is the accumulation of limited powered studies that might overestimate the effects we are looking for (type I error) or fail to detect a difference (type II error) based on limited sample size. Indeed, similar concerns have been raised regarding studies on postoperative cognitive dysfunction (POCD) in the elderly.^{14,15} POCD researchers still have no tools available that can reliably assess the presence of POCD, and increasing the number of tests used to classify POCD increases the sensitivity to change not only in postoperative patients but also in the controls.¹⁴

Ing *et al.* should be congratulated for their contribution to the understanding of the growing concerns related to the effects of exposure to anesthetic agents on young children. However, not all outcome measures are created equally—the devil is truly in the details with regard to not only outcome but also many other aspects of study design and conduct not discussed here. However, the problems with the POCD studies suggest that one must ascertain under what circumstances individual cognitive testings are also meaningful human outcome measures. Indeed, exactly how different are individually administered tests of speech and language and school tests—certainly, good school test scores require adequate speech and learning skills?

Acknowledgments

Dr. Flick is supported by the Eunice Kennedy Shriver National Institute of Child Health and Human Development (Bethesda, Maryland) to study this topic (grant no. R01 HD 071907-01).

Competing Interests

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.

Correspondence

Address correspondence to Dr. Hansen: tomghansen@dadlnet.dk or tom.g.hansen@rsyd.dk

References

1. Vutskits L, Davis PJ, Hansen TG: Anesthetics and the developing brain: Time for a change in practice? A pro/con debate. *Paediatr Anaesth* 2012; 22:973–80
2. Hansen TG, Flick R; Danish Registry Study Group; Mayo Clinic Pediatric Anesthesia and Learning Disabilities Study Group: Anesthetic effects on the developing brain: Insights from epidemiology. *ANESTHESIOLOGY* 2009; 110:1–3
3. Ing CH, DiMaggio CJ, Malacova E, Whitehouse AJ, Hegarty MK, Feng T, Brady JE, von Ungern-Sternberg BS, Davidson AJ, Davidson AJ, Wall MM, Wood AJJ, Li G, Sun LS: Comparative analysis of outcome measures used in examining neurodevelopmental effects of early childhood anesthesia exposure. *ANESTHESIOLOGY* 2014; 120:1319–32
4. Ing C, DiMaggio C, Whitehouse A, Hegarty MK, Brady J, von Ungern-Sternberg BS, Davidson A, Wood AJ, Li G, Sun LS: Long-term differences in language and cognitive function after childhood exposure to anesthesia. *Pediatrics* 2012; 130:e476–85
5. Flick RP, Katusic SK, Colligan RC, Wilder RT, Voigt RG, Olson MD, Sprung J, Weaver AL, Schroeder DR, Warner DO: Cognitive and behavioral outcomes after early exposure to anesthesia and surgery. *Pediatrics* 2011; 128:e1053–61
6. Hansen TG, Pedersen JK, Henneberg SW, Pedersen DA, Murray JC, Morton NS, Christensen K: Academic performance in adolescence after inguinal hernia repair in infancy: A nationwide cohort study. *ANESTHESIOLOGY* 2011; 114:1076–85
7. Kalkman CJ, Peelen L, Moons KG, Veenhuizen M, Bruens M, Sinnema G, de Jong TP: Behavior and development in children and age at the time of first anesthetic exposure. *ANESTHESIOLOGY* 2009; 110:805–12
8. DiMaggio C, Sun LS, Li G: Early childhood exposure to anesthesia and risk of developmental and behavioral disorders in a sibling birth cohort. *Anesth Analg* 2011; 113:1143–51
9. Bartels M, Althoff RR, Boomsma DI: Anesthesia and cognitive performance in children: No evidence for a causal relationship. *Twin Res Hum Genet* 2009; 12:246–53
10. O'Malley KJ, Cook KF, Price MD, Wildes KR, Hurdle JF, Ashton CM: Measuring diagnoses: ICD code accuracy. *Health Serv Res* 2005; 40(5 Pt 2):1620–39
11. Rowland AS, Lesesne CA, Abramowitz AJ: The epidemiology of attention-deficit/hyperactivity disorder (ADHD): A public health view. *Ment Retard Dev Disabil Res Rev* 2002; 8:162–70
12. Sprung J, Flick RP, Katusic SK, Colligan RC, Barbaresi WJ, Bojanić K, Welch TL, Olson MD, Hanson AC, Schroeder DR, Wilder RT, Warner DO: Attention-deficit/hyperactivity disorder after early exposure to procedures requiring general anesthesia. *Mayo Clin Proc* 2012; 87:120–9
13. Browning GG, Rovers MM, Williamson I, Lous J, Burton MJ: Grommets (ventilation tube) for hearing loss associated with otitis media with effusion in children (Review). *Cochrane Database Syst Rev* 2010: CD001801
14. Lewis MS, Maruff P, Silbert BS, Evered LA, Scott DA: Detection of postoperative cognitive decline after coronary artery bypass graft surgery is affected by the number of neuropsychological tests in the assessment battery. *Ann Thorac Surg* 2006; 81:2097–104
15. Selnes OA, Gottesman RF, Grega MA, Baumgartner WA, Zeger SL, McKhann GM: Cognitive and neurologic outcomes after coronary-artery bypass surgery. *N Engl J Med* 2012; 366:250–7