Published in final edited form as:

Anesth Analg. 2017 December ; 125(6): 1845–1848. doi:10.1213/ANE.0000000000002567.

Association of Surgery and Anesthesia with Mental Disorder Diagnoses: What Would Sir Austin Bradford Hill Say?

Olubukola O. Nafiu, MD, FRCA, MS¹ and Peter J. Davis, MD²

¹Department of Anesthesiology, Section of Pediatric Anesthesiology, University of Michigan, Ann Arbor, Michigan, USA

²Professor of Anesthesiology and Pediatrics; Dr. Joseph H. Marcy Endowed Chair in Pediatric Anesthesiology; Professor of Anesthesia and Pediatrics, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania; Anesthesiologist-in Chief, Children's Hospital of Pittsburgh of UPMC

Keywords

General anesthesia; neurotoxicity; children; mental disorder; learning disability

In the current issue of the journal, Ing and colleagues further contribute to the increasing research on whether exposure to anesthesia at an early age is detrimental to the developing brain (1).

The potential for anesthetic agents to increase the risk of neurotoxicity has become both a paramount concern and a source of intense debate among perioperative caregivers over the last two decades (2–5). The fundamental question is whether children who undergo surgery and general anesthesia (GA) in the first few years of life are at risk for long-term or permanent neuro-behavioral abnormalities. Given that millions of children undergo procedures requiring GA every year in the United States and around the world, this is an important clinical and epidemiological question.

One of the earliest studies on this topic demonstrated increased rates of neuronal death (neuro-apoptosis) following the administration of NMDA receptor blockers to neonatal rats (6). These early investigators subtly concluded that “these findings may have implications for pediatric anesthesia and that the mechanism of drug-induced neuro-apoptosis could contribute to a variety of neuropsychiatric disorders” (6). Since these early laboratory studies, an ever-increasing number of observational studies on human infants and preschool children have examined the association of multiple or single GA exposure and surgery with subsequent development of neurobehavioral problems in childhood. (5,7–15) These human studies showed a variety of primary outcomes on factors ranging from learning disabilities, school performance evaluations, intelligence testing (IQ), and parental behavior assessments.

Olubukola O Nafiu: This author helped with idea concept, literature review and manuscript preparation

Peter J. Davis: This author helped with manuscript preparation and editing.

Conflict of interest: none for any of the authors have any conflict of interest to declare

Not surprisingly, the findings of these studies have been inconclusive and have generated more questions than answers (8).

What complicates the question of surgery and anesthesia-related neurotoxicity is that there are no clear biomarkers or phenotypes for outcomes. Unlike other clinical problems, anesthesia-related neurotoxicity is an issue of “reverse association”; that is, laboratory studies in animals have preceded clinical observations. Animal studies are important in helping to delineate such issues, but one must remember that an animal model does not always equate to human conditions. One such example is studies on rodents that suggest that Sevoflurane causes renal toxicity (16,17). Now, millions of anesthetics later, the animal model of renal toxicity does not appear to reflect the human experience. Given that studying anesthesia-related neurotoxicity is not amenable to classic randomized controlled trials (hard to find volunteers for surgery without anesthesia), most published works on this subject are by default, non-experimental (2–5).

What is not clear is which factors, if any, are associated with neurocognitive or behavioral impairment. Is there a threshold value for anesthetic toxicity? Does patient age at exposure, drug dosage, duration of anesthetic exposure, number of anesthetic exposures, or the area under the anesthetic concentration-time curve of anesthetic exposure influence neurocognitive outcome?

Human studies, whether retrospective (6,7) or prospective (9) are conducted in inherently “noisy” environments. For example, standardizing the exposure variable (GA) is impossible. GA is a “state,” and no two GA concentrations produce the same anesthetic state. Also, is the patient inherently predisposed to the outcome, or did surgery/anesthesia simply unmask this inherent defect; i.e. are surgery and anesthesia simply innocent bystanders?

In this current issue of *A&A*, Ing *et al.* (1) retrospectively studied a large cohort of children (N=38,493) from the New York and Texas Medicaid databases who had undergone either pyloric stenosis, inguinal hernia, circumcision, or tonsillectomy and/or adenoidectomy before five years of age. The primary objective was to determine whether the age at GA exposure affected the risk of developing a mental disorder for up to 12 years after GA exposure. Mental disorder diagnosis was identified using ICD-9 coding and included: schizophrenia, bipolar disorder, depressive disorder, anxiety disorder, conduct disorders, autism, attention deficit hyperactivity disorder (ADHD), intellectual disability, developmental delay, delusional disorders, dissociative and somatosensory disorders, personality disorders, and adjustment disorders. The authors noted that exposed patients had an increased hazard ratio (HR) of 1.26 (95% CI 1.20–1.32) for developmental delay and 1.31 (95% CI 1.25–1.37) for ADHD, and that there did not appear to be an effect with respect to the age of the patient at the time of the anesthetic exposure.

Unfortunately, the study design (retrospective database analyses) makes it difficult to draw any conclusions other than an association. Propensity matching involving 50 variables was used for comparative purposes to identify a group of unexposed patients. However, no mention was made as to whether children were raised in a single parent or two-parent household, the number of household members, the level of maternal education, any history

of parental substance abuse, or a family history of depression or mental illness. In addition, there was no way to ascertain who made the ICD-9 diagnosis of mental disorder, let alone the accuracy of the diagnosis.

Although retrospective database studies make drawing any conclusions about causation difficult, the authors use of two different states' databases makes for some interesting observations. It is notable that unexposed children in Texas had the same incidence of mental disorders (3.5–5.4 diagnoses per 100 person years) as the exposed children in New York (3.0–5.3 diagnoses per 100 person years). Analyzing a claims database for clinical outcomes is always going to be difficult given the inherent “noise” associated with these observations. Possible limitations noted with propensity matching coupled with possible misclassification errors when combining distinct databases further limits drawing any conclusions about causality. Another concern about the study was the inclusion of children who underwent adenotonsillectomy (T&A) in their analyses. Given that adeno-tonsillar hypertrophy and sleep-disordered breathing are often associated with the outcome variables reported in their study (ADHD and learning disability) (18), this creates significant issues with confounding, which propensity score matching may not fix.

Although association does not mean causation, nevertheless, investigators and the lay public often attempt to determine causation from non-experimental studies. Given the limitations of determining causation factors with patients exposed to anesthesia, the epidemiological approach involving the Bradford Hill causality criteria (19) can be utilized to examine the hypothesis that GA may cause neurobehavioral anomalies or mental disorder diagnoses. Thus, we ask the question: what would Sir Austin Bradford Hill do?

In 1965, Sir Austin Bradford Hill proposed a set of nine criteria to determine causality in observational or epidemiological conditions (19). Establishing an argument for causation is a critical research activity because it affects the delivery of safe medical care. Given that anesthesia/surgery-related neurotoxicity can only be subjected to observational cohort studies (surgery without anesthesia is unethical), it represents the ideal scenario to subject to the Bradford Hill's criteria for causality. The criteria are as follows: (1) strength of association, (2) consistency, (3) specificity, (4) temporality, (5) biological gradient, (6) plausibility, (7) coherence, (8) experiment, and (9) analogy.

- 1 Strength of the observed association:** This was listed as the first criterion in Bradford Hill's paper (19). The premise is that if condition A causes an outcome B, then A&B can be demonstrably associated with each other. The association should be strong enough to be judged as clinically significant. However, Sir Bradford Hill cautioned that we should not be too ready to discard slight associations (19). A small association does not mean causal association is absent, although the larger the association, the more likely that it is causal.

Additionally, this criterion ignores the fact that not every component cause will have a strong association with the outcome produced. Although the overall HR of mental disorder diagnosis in the study by Ing *et al.* was only 1.26 (1.22–1.30), not a strong effect, other anesthesia and neurocognitive studies have found similarly small to moderate measures of association (3,4). Generally, an odds

ratio (OR) of 1.5 or less is considered small, OR = 2 moderate and OR ≥ 3 large effect size (20).

- 2 **Consistency of the observed association (reproducibility):** Bradford Hill suggests that reproducible findings observed by different persons in different settings strengthens the likelihood of causation. Although the data by Ing *et al.* will appear to score highly on this criterion, given that there are now many studies that have examined the association between GA and a wide variety of neurodevelopmental outcomes (2–8), certain caveats are necessary. The results from a review of these studies have been mixed largely due to inconsistency in the definition of the outcome variable (8). Diagnostic criteria for anesthesia-related neurobehavioral changes have varied widely and many of the studies have used a variety of diagnostic instruments, making comparison between studies extremely difficult. Another important caveat with the consistency criterion is that it is highly subject to publication bias because of the higher likelihood that positive findings get published. On the other hand, non-reproducible or inconsistent findings may be due to differences in research methodology. Furthermore, causal agents might require the presence of other factors. For example, does anesthesia/surgery-related neurotoxicity require hypoxia, hypercarbia, or other factors that present day limited knowledge has not identified? Finally, as Bradford Hill cautioned, shared flaws in study design would tend to replicate the same wrong conclusions (19).
- 3 **Specificity:** This is probably the weakest of all the Hill criteria to establish, given that epidemiological diseases are rarely caused by a single agent. There is no specific “causal agent” to be found in the report by Ing *et al.* Consider the outcome variable of ADHD. Was it the GA? The surgery? The prevailing environmental conditions in the states studied? Without knowing a threshold value for anesthetic toxicity, vulnerable age range of exposure, drug dosage, duration of anesthetic exposure, number of anesthetic exposures, or the area under the anesthetic concentration-time curve makes specificity a difficult component to achieve. However, the lack of specificity should in no way detract from the argument for causation (19).
- 4 **Temporality:** The exposure must precede the disease or the outcome it is supposed to cause. The premise is that if A causes B, then A should occur before B, or as Sir Bradford Hill reflected, (19) “which is the cart, and which is the horse?” The best way to determine temporality in health research is to conduct prospective trials where subjects are examined for the presence of the outcome prior to the exposure variable. Subjects are then monitored for the outcome after a sufficient lag period.

In the case of anesthesia-related neurotoxicity, GA administration must predate the disorder it is purported to cause. This is ostensibly the easiest of the Hill criteria to fulfill but, in reality, the most difficult. GA is a clear categorical variable. A patient either received GA or they didn't. Unfortunately, this clear variable can be difficult to establish from retrospective database queries. In their report, Ing *et al.* utilized surgical procedure as a proxy for GA,

arguing that it is impossible to perform the selected procedures without anesthesia. While this may be true for T&A and pyloromyotomy, some of the circumcision and herniorrhaphy cases could have been performed without GA, which would introduce recall or attribution bias to the outcome of the study. Furthermore, since mental disorders were not prospectively evaluated before the anesthetic exposure, it is impossible to determine whether the patients had pre-procedural behavioral disturbances from a simple retrospective database study. This limits the assessment of temporality.

- 5 **Biological gradient:** Sir Bradford Hill suggested that evidence of a dose-response relationship (or biological gradient) was indicative of a causal relationship (19). Greater exposure should generally lead to greater incidence of the effect. While this criterion may be required for an outcome due to exposure to pathogens or toxins, it may not be appropriate or easy to establish in anesthesia-related neurotoxicity. Though studies have shown patients with repeated exposures have fared worse, it is unclear whether the repeat exposure or the patient's underlying disease process that is responsible. As previously noted, it becomes essential to discern a threshold value for anesthetic toxicity, vulnerable age, drug dosage, duration of anesthetic exposure, number of anesthetic exposures, or the area under the anesthetic concentration-time curve of anesthetic exposure. Furthermore, what is the best measure of the putative brain insult resulting from GA? Is it language impairment, behavioral abnormalities, or something else? Most outcomes such as language delay and attention deficit disorder are indirect measures of brain injury because they may be influenced by non-cerebral factors, such as parental education, income, home stressors such as parental abuse, relocation, job losses, etc. None of these issues appeared to be addressed in a propensity matching.
- 6 **Plausibility:** Does it make biologic sense that the exposure is causing the outcome? Sir Bradford Hill was quick to point out that this is a feature we cannot demand in order to establish causality (19). The principal reason is that biologic plausibility is often determined by the prevailing knowledge and belief system, because "truth" changes with our belief system. For example, it was once "true" that blood-letting was the standard of care for the prevention and treatment of illnesses. Also, it was generally accepted until the late 1980s that neonates did not feel pain and major surgeries were performed on babies without anesthesia or analgesia (21). These lines of thinking, popular in those days, makes no sense in contemporary medicine. Therefore, biological plausibility depends on the prevailing knowledge of the day. Consequently, we must be willing to accept, regardless of what we believe about anesthesia-related neurobehavioral disorders, that an argument that makes no sense to us does not mean it isn't true.

The remaining three Bradford Hill criteria (coherence, experiment, and analogy) can be succinctly summarized. *Coherence* between known information, clinical and laboratory, increases the likelihood of causality. Unfortunately, there is no consensus on the role of GA with neurobehavioral or mental disorder diagnoses. There is presently insufficient data to support the *experimental evidence* criterion because randomizing children to surgery without

anesthesia is unethical. Finally, regarding *analogous* evidence, if neurobehavioral changes are observed in other situations comparable to GA, this could provide analogous evidence. For example, children sedated in the intensive care unit may provide some analogous evidence. However, eliminating confounders in this setting would be almost impossible.

Summary

Although each of the Bradford Hill criteria are not a *sine qua non* for causality, the criteria do provide a framework against which anesthetic exposure and outcome can be examined. The question underlying the hypothesis that GA causes neurobehavioral or mental disorder is far from being answered. To this end, further research on this subject is needed. However, future studies should address the methodological flaws of existing studies and re-examine the value of retrospective database studies, with their inherent biases and presumed matching criteria.

Finally, as with many other studies in this area, “what we don’t know, we don’t know” may be more of a contributor to the findings than what we think we know. Studies that can identify a biomarker or develop a phenotype will become important elements in determining causation, as well as defining threshold values for toxicity with respect to anesthetic agents, doses, and duration and age of exposure.

Acknowledgments

Financial statement: National Institute of General Medical Sciences (NIGMS) grant number K23 GM104354 supported Dr. Nafiu’s work.

The National Institute of General Medical Sciences (NIGMS) supported Dr. Nafiu’s work (grant number K23 GM104354).

References

1. Ing C, Sun M, Olfson M, DiMaggio CJ, Sun LS, Wall MM, Li G. Age at Exposure to Surgery and Anesthesia in Children and Association with Mental Disorder Diagnosis. *Anesth Analg*. 2017 xx: xx–xx.
2. Wilder RT, Flick RP, Sprung J, Katusic SK, Barbaresi WJ, Mickelson C, Gleich SJ, Schroeder DR, Weaver AL, Warner DO. Early exposure to anesthesia and learning disabilities in a population-based birth cohort. *Anesthesiology*. 2009; 110:796–804. [PubMed: 19293700]
3. Sun LS, Li G, Miller TL, Salorio C, Byrne MW, Bellinger DC, Ing C, Park R, Radcliffe J, Hays SR, DiMaggio CJ, Cooper TJ, Rauh V, Maxwell LG, Youn A, McGowan FX. Association Between a Single General Anesthesia Exposure Before Age 36 Months and Neurocognitive Outcomes in Later Childhood. *JAMA*. 2016; 315(21):2312–20. [PubMed: 27272582]
4. DiMaggio C, Sun LS, Kakavouli A, Byrne MW, Li G. A retrospective cohort study of the association of anesthesia and hernia repair surgery with behavioral and developmental disorders in young children. *J Neurosurg Anesthesiol*. 2009; 21(4):286–91. [PubMed: 19955889]
5. Sprung J, Flick RP, Katusic SK, Colligan RC, Barbaresi WJ, Bojanic K, Welch TL, Olson MD, Hanson AC, Schroeder RD, Wilder RT, Warner DO. Attention-deficit/hyperactivity disorder after early exposure to procedures requiring general anesthesia. *Mayo Clin Proc*. 2012; 87:120–9. [PubMed: 22305025]
6. Ikonomidou C, Bosch F, Miksa M, Bittigau P, Vöckler J, Dikranian K, Tenkova TI, Stefovskva V, Turski L, Olney JW. Blockade of NMDA receptors and apoptotic neurodegeneration in the developing brain. *Science*. 1999; 283(5398):70–4. [PubMed: 9872743]

7. 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; 113:1143–1151. [PubMed: 21415431]
8. Vutskits L, Davidson A. Update on developmental anesthesia neurotoxicity. *Curr Opin Anaesthesiol*. 2017; 30:337–42. [PubMed: 28277380]
9. Bartels M, Althoff RR, et al. Anesthesia and cognitive performance in children: no evidence for a causal relationship. *Twin Res Hum Genet*. 2009; 12(3):246–53. [PubMed: 19456216]
10. Flick RP, Katusic SK, et al. Cognitive and behavioral outcomes after early exposure to anesthesia and surgery. *Pediatrics*. 2011; 128(5):e1053–61. [PubMed: 21969289]
11. Glatz P, Sandin RH, et al. Association of Anesthesia and Surgery During Childhood With Long-term Academic Performance. *JAMA Pediatr*. 2017; 171(1):e163470. [PubMed: 27820621]
12. Hansen TG, Pedersen JK, et al. Academic performance in adolescence after inguinal hernia repair in infancy: a nationwide cohort study. *Anesthesiology*. 2011; 114(5):1076–85. [PubMed: 21368654]
13. Ing C, DiMaggio C, et al. Long-term differences in language and cognitive function after childhood exposure to anesthesia. *Pediatrics*. 2012; 130(3):e476–85. [PubMed: 22908104]
14. O’Leary JD, Janus M, et al. A Population-based Study Evaluating the Association between Surgery in Early Life and Child Development at Primary School Entry. *Anesthesiology*. 2016; 125(2):272–9. [PubMed: 27433745]
15. Davidson AJ, Disma N, et al. Neurodevelopmental outcome at 2 years of age after general anaesthesia and awake-regional anaesthesia in infancy (GAS): an international multicentre, randomised controlled trial. *Lancet*. 2016; 387(10015):239–50. [PubMed: 26507180]
16. Martin JL, Laster MJ, et al. Metabolism of compound A by renal cysteine-S-conjugate beta-lyase is not the mechanism of compound A-induced renal injury in the rat. *Anesth Analg*. 1996; 82(4):770–4. [PubMed: 8615496]
17. Gonsowski CT1, Laster MJ, et al. Toxicity of compound A in rats. Effect of a 3-hour administration. *Anesthesiology*. 1994; 80(3):556–65. [PubMed: 8141452]
18. Sedky K, Bennett DS, Carvalho KS. Attention deficit hyperactivity disorder and sleep disordered breathing in pediatric populations: a meta-analysis. *Sleep Med Rev*. 2014; 18(4):349–56. [PubMed: 24581717]
19. Hill AB. The environment and disease: Association or causation? *Proceedings of the Royal Medical Society*. 1965; 58:295–300.
20. Sullivan GM, Feinn R. Using Effect Size-or Why the P Value Is Not Enough. *J Grad Med Educ*. 2012; 4(3):279–82. [PubMed: 23997866]
21. Owens M, Todt E. Pain in infancy: neonatal reaction to a heel lance. *Pain*. 1984; 20(1):77–86. [PubMed: 6493791]