

Clinical Evidence for Any Effect of Anesthesia on the Developing Brain

Andrew J. Davidson, M.B.B.S., M.D., F.A.N.Z.C.A., Lena S. Sun, M.D., F.A.A.P., D.A.B.A.

ABSTRACT

A recent U.S. Food and Drug Administration warning advised that prolonged or repeated exposure to general anesthetics may affect neurodevelopment in children. This warning is based on a wealth of preclinical animal studies and relatively few human studies. The human studies include a variety of different populations with several different outcome measures. Interpreting the results requires consideration of the outcome used, the power of the study, the length of exposure and the efforts to reduce the confounding effects of comorbidity and surgery. Most, but not all, of the large population-based studies find evidence for associations between surgery in early childhood and slightly worse subsequent academic achievement or increased risk for later diagnosis of a behavioral disability. In several studies, the amount of added risk is very small; however, there is some evidence for a greater association with multiple exposures. These results may be consistent with the preclinical data, but the possibility of confounding means the positive associations can only be regarded as weak evidence for causation. Finally, there is strong evidence that brief exposure is not associated with any long term risk in humans. (**ANESTHESIOLOGY 2018; 128:840-53**)

THE US Food and Drug Administration (FDA) recently issued a change in labeling regarding the safe use of anesthetic and sedative agents (<http://www.fda.gov/Drugs/DrugSafety/ucm532356.htm>; accessed November 10, 2017).

The opening sentence states that “repeated or lengthy use of general anesthetic and sedation drugs during surgeries or procedures in children younger than 3 years or in pregnant women during their third trimester may affect the development of children’s brains.” This warning then suggests that brief exposure is probably safe and continues to a summary of available preclinical and human evidence. It also encourages healthcare professionals to consider the risks of delaying surgery. The issue of how, or even if, anesthetic agents affect the developing brain has been the subject of a great deal of research over the last couple of decades. There is now overwhelming preclinical evidence that most anesthetic agents can, in some situations, modulate various aspects of brain development, but there is great uncertainty over how these findings in animal models translate to clinically relevant human scenarios.^{1–3} Evaluating the strengths and limitations of human evidence is essential when determining potential change in practice. Several important human studies have been published recently, and with the recent FDA warning, it is an opportune time for this review which aims to

critically appraise the human evidence. This review will focus on studies that provide the strongest evidence, and will provide a summary of the current human evidence.

What Are We Looking for in Human Studies?

The question of how anesthesia exposure in childhood influences neurodevelopmental outcome arose after findings made in the laboratory. Before these findings, a link between major surgery in the neonate and increased risk of neurodevelopmental problems was suspected, however, anesthesia toxicity was not considered as the cause for this association.^{4–6} No link between surgery and neurodevelopmental outcome had been noticed outside the neonatal period; there was no obvious clinical problem. However, as the preclinical data has grown, there has become an increasingly pressing need to determine if anesthesia exposure does indeed cause clinically relevant changes in neurodevelopment of children. Answering this question is not straightforward. There are several facets or domains of neurodevelopmental function that could be affected. The effect may be dependent on time or duration of anesthesia administration, or the effect may be only apparent in subpopulations. In the absence of an

This article is featured in “This Month in Anesthesiology,” page 1A. Corresponding articles on pages 693 and 700.

Submitted for publication February 14, 2017. Accepted for publication October 5, 2017. From the Department of Anesthesia and Pain Management, Royal Children’s Hospital, the Melbourne Children’s Trials Centre, Murdoch Children’s Research Institute, Melbourne, Australia (A.J.D.); and the College of Physicians and Surgeons, Columbia University, New York, New York (L.S.S.).

Copyright © 2018, the American Society of Anesthesiologists, Inc. Wolters Kluwer Health, Inc. All Rights Reserved. Anesthesiology 2018; 128:840–53

obvious clinical problem when designing human studies, the preclinical data should drive not only the choice of population to be studied in terms of age and duration at exposure, but also the choice of domain of neurodevelopmental function to be assessed. Unfortunately, translating preclinical data in humans is an inherently imprecise science. Also, even if the preclinical data could be easily translated, the preclinical data show a range of effects over a range of ages and duration of exposures. A recent review of 440 preclinical studies found a greater proportion of studies reporting abnormalities with progressively longer exposures, but no clear exposure duration threshold below which no abnormalities were seen.² Many of the early preclinical studies suggested that only newborn animals were affected; however, the same review of the entire range of literature found that while the vast majority of preclinical studies were in very young animals (equivalent to the fetus or neonate), there was no clear evidence that abnormalities were absent in older animals.² The neurodevelopmental functional domain that is likely to be affected is also uncertain. The diffuse nature of the injury seen in preclinical studies would suggest that domains such as higher executive function are most likely affected in humans. Animal studies have demonstrated functional defects in memory and learning, while recent non-human primate studies have shown functional defects in behavior. Thus, in terms of population and neurodevelopmental outcome, it is difficult to know where to look in human studies.

Given the uncertainty of translating preclinical data, it is impossible to design a single definitive human study that could confirm, or rule out, an effect of anesthesia on neurodevelopment, and it is appropriate for a range of studies to examine a range of outcomes, anesthetic durations and ages at exposure. In this review, we will group the human studies in terms of outcomes assessed. When considering the evidence provided by each study, it is important to consider the size or power of the study, and the likelihood of confounding factors.

Outcomes Considered

The clinical studies can be broadly divided into three groups in terms of outcomes: group or population administered academic performance or school readiness tests; a diagnosis of a neurodevelopmental or behavioral disorder or learning disability; and abnormalities in neurocognitive function or behavior based on validated neuropsychologic or behavior assessment tools.⁷ One study has also looked at structural changes on magnetic resonance imaging.⁸ It is important to consider all domains, as brain insult early in life may result in a deficit in one of these outcome groups without a deficit in the others. For example, an event may increase the risk of behavioral disorder without having an impact on cognition, and *vice versa*.

Academic Performance and School Readiness as Outcome Measures

Measures of academic performance or school grades are administered in a standard way in large populations of

children, and school readiness tests are administered to preschool children to determine if they are ready for school. Neurodevelopmental outcome data for very large numbers of children can be relatively easily extracted from existing databases. School grades are important outcomes from the child's perspective but they are imperfect ways to assess possible brain injury and may be insensitive to defects in some domains of neurodevelopment. Many other factors can affect academic performance. School grades are also only available for those children that have sufficient capacity to be in the school system. Table 1 summarizes the studies which have used academic performance or school readiness as an outcome measure. School grades are important, and easy to obtain in large numbers, but imperfect measures of neurodevelopment.

Using the Young Netherlands Twin Registry, Bartels *et al.*, examined two different outcomes: education achievement (using standardized achievement scores) and cognitive problems (identified using the Conners' Teacher Short Form).⁹ Twin studies are particularly useful as twins are very similar in terms of genetics and environment. Parents were questioned with regard to whether their children had received anesthesia both before the age of 3 yr and by the age of 12 yr. The twins in the registry were then divided into three groups: twins who were concordant for exposure to anesthesia at age 3 or by age 12, twins who were concordant for no anesthesia exposure before age 12, and twins who were discordant for anesthesia exposure before age 3 or by age 12. The results of this analysis found there were no differences detected in outcomes of twins who were discordant for exposure; however, the small numbers of study participants limit the power of this study. Interestingly, the concordant, unexposed twins had more favorable outcomes compared to both of the other two groups. The authors suggested that this may be interpreted as the need for anesthesia exposure being a genetically mediated link to vulnerability and that exposure was not the cause for any adverse neurodevelopmental outcome, *per se*.

Hansen *et al.* conducted a birth cohort study using the Danish National Patient Register (1986 to 1990) in infants (under 1 yr) exposed to anesthesia during inguinal hernia surgery,¹⁰ or in those under 3 months of age who were exposed to anesthesia during pyloromyotomy¹¹ and compared them to a random age-matched 5% sample of the entire population using ninth grade test scores and teachers' scores as the primary outcome. Their results found no evidence for differences between anesthetized and control children in primary outcomes, but children exposed to anesthesia were more likely to be categorized by "nonattainment of score." Nonattainment is defined as the child being unable to sit for the test, for whatever reason. This includes children that are unable to sit for tests due to significant neurodevelopmental or behavioral problems. In another population-based Danish study, Clausen *et al.* compared academic achievement in 509 adolescents with cleft lip and/or palate that had received anesthesia compared to a 5% sample of the population

Table 1. Published Studies with Academic Performance/School Readiness as Main Outcome Measures

Author, Yr (Setting)	Design	Data Source	Study Subjects		Age of Exposure	Main Outcomes
			Exposed	Unexposed		
Bartels, 2009 (The Netherlands)	Retrospective	Netherlands Birth Registry (monozygotic twins only) (1986–1995)	One twin had anesthesia	One twin no anesthesia	0–3 yr 0–12 yr	Age 12 yr No difference in national educational achievement test or teacher rating of cognitive problems
Hansen, 2011 (Denmark)	Retrospective	Danish Patient Register, a population-based cohort (1986–1990)	Inguinal hernia repair (n = 2,689)	Randomly selected 5% age-matched population sample (n = 14,575)	≤ 12 months	Age 15 or 16 yr No difference in average ninth grade test scores and teacher rating Higher rate of nonattainment in exposed
Flick, 2011 (Mayo Clinic, Minnesota, USA)	Retrospective	Olmsted County birth cohort (1976–1982)	Children who had anesthesia for all procedures (n = 350)	No anesthesia, matched controls (n = 700)	< 2 yr	Age ≥ 5 yr No difference in group administered achievement tests
Block, 2012 (Iowa, USA)	Retrospective	Hospital records and available standardized composite IOWA scores (1990–2001)	General anesthesia for circumcision, pyloromyotomy, or inguinal hernia repair (n = 133)	Population norm scores	≤ 12 months	Age 7–17.9 yr No differences in average scores. An over-representation of those scoring at bottom 5%
Hansen, 2013 (Denmark)	Retrospective	Danish Patient Register, a population-based cohort (1986–1990)	Pyloric stenosis repair before age 3 months (n = 779)	Randomly selected 5% age-matched population sample (n = 14,465)	≤ 3 months	Age 15 or 16 yr No difference in average ninth grade test scores and teacher rating Higher rate of nonattainment in exposed
Bong, 2013 (Singapore)	Retrospective	Hospital birth registry (1998–1999)	ASA Physical Status I and II Anesthesia for minor surgery (n = 100)	Age-matched with no anesthesia (n = 100)	≤ 12 months	Age 12 yr No difference in Singapore standardized test scores
Ing, 2014 (Western Australia)	Retrospective	Raine birth cohort (1989–1992)	Anesthesia for any procedure (n = 112)	No anesthesia (n = 669)	≤ 3 yr	Age 10 yr No difference in Western Australian standardized test scores
Williams, 2014 (Vermont, USA)	Retrospective	Vermont Spinal Infant Registry (1989–2003) and Vermont education records (2000–2011)	Spinal anesthesia for circumcision, pyloromyotomy, or inguinal hernia repair (n = 260)	Population norm scores	≤ 12 months	Grades 3 to 5 No difference in standardized test scores or need for individualized educational program
Graham, 2016 (Manitoba, Canada)	Retrospective	Population-based Manitoba Provincial Health administrative database and EDI (2006, 2007, 2009, 2011)	Child with surgery/anesthesia (n = 4,470)	No surgery/anesthesia (n = 13,586)	0–4 yr	Kindergarten age For entire cohort – small decrease in total EDI score. Deficits in EDI domains: communication/general knowledge, language/cognition Strong interaction with age (effect greater in 2–4 age group)

(Continued)

Table 1. (Continued)

Author, Yr (Setting)	Design	Data Source	Study Subjects		Age of Exposure	Main Outcomes
			Exposed	Unexposed		
O'Leary, 2016 (Ontario, Canada)	Retrospective	Population-based Ontario Provincial Health administrative database and EDI (2004–2012)	Child with surgery/anesthesia (n = 28,366)	No surgery/anesthesia (n = 55,910)	0–4 yr	Kindergarten age increased developmental vulnerability (scoring in lowest 10%) EDI domain deficits: Physical health, emotional health, communication/general knowledge Effect only seen in those in 2–4 yr age band.
Clausen, 2016 (Denmark)	Retrospective	Danish Patient Register, a population-based cohort (1986–1990)	Oral cleft surgery (total n = 508; cleft lip = 171, cleft palate = 195, cleft lip and palate = 222)	Randomly selected 5% age-matched population sample (n = 14,677)	Age 15 or 16 yr	No difference in average ninth grade test scores and teacher rating in children with cleft lip or cleft lip and palate; lower scores in children with cleft palate
Glatz, 2017 (Sweden)	Retrospective	Population-based Swedish Patient Register birth cohort (1973–1993)	Anesthesia exposures (n = 33,514)	No anesthesia (n = 159,619)	0–4 yr	Age 16 yr Mean school grades difference 0.41% Age 18 yr IQ test scores (National Military Con- scription Review Process) Mean difference 0.97%

ASA = American Society of Anesthesiologists; EDI = Early Development Instrument; IOWA = Iowa Tests of Basic Skills and Educational Development.

(14,677 adolescents).¹² The analysis was adjusted for sex, birth weight, parental age, and parental level of education. Compared to controls, there was evidence that children with cleft palate alone had lower test scores, while there was no evidence of differences compared to controls for either cleft lip alone or combined cleft lip and palate. The cleft palate alone group was also more likely to have nonattainment of scores.

A similar but smaller study from the University of Iowa¹³ examined infants (under 1 yr) who had one of three surgeries: inguinal hernia, with or without orchiopexy; pyloromyotomy; and circumcision. The outcome measure used was the composite score on the Iowa Tests of Basic Skills and Educational Development (IOWA test scores). The authors examined both the mean scores and the proportion of the group that had scores that represented the lowest fifth percentile of the population norm. From the initial total of 519 subjects whose relevant medical records were retrievable, 287 had available IOWA test scores (group 1). Of these, 133 agreed to have their medical records examined (group 2). The authors then applied a list of 18 prespecified central nervous system-related problems or risk factors, and 58 of the 133 were found to have no risk factors (group 3). In both groups 1 and 2, the mean scores were significantly lower than the population mean, but not in group 3. However, in all three groups, there was a disproportionately larger percentage of the group that scored in the lowest fifth percentile for the IOWA tests (12% in group 1, 11% in group 2, and 14% in group 3). The authors concluded that anesthesia and surgery during infancy was associated with over-representation of very low IOWA test scores. The Iowa study is consistent with the Danish studies where mean scores were similar between those with and without exposure to surgery/anesthesia; however, there was a slightly increased risk of being more likely to have very poor performance scores or nonattainment of the standardized tests.

Williams *et al.* examined academic performances of individuals who received spinal anesthesia as infants in Vermont for the same three surgical procedures that the Iowa study evaluated, and found no evidence of any differences between infants who had surgery/spinal anesthesia and the general population.¹⁴ Bong *et al.*¹⁵ matched 100 healthy children in Singapore who had minor surgery (inguinal hernias, circumcisions, cystoscopies, and pyloromyotomies) under general anesthesia before age one with 106 age-matched children without anesthesia or sedation exposure. Two different outcome measures were measured at age 12 yr: academic performance using the standardized aggregate Primary School Leaving Examination scores; and diagnosis of learning disability. They found no evidence for a difference in examination scores but children having surgery had a greater risk for having a formal diagnosis of a learning disability.

The Bartels, Block, Williams, and Bong studies all had relatively small samples. Small sample sizes will inevitably provide imprecise results with wide 95% CI that may overlap

with clinically relevant added risk. Thus, when small studies find no evidence for an association or added risk it should be noted that they may not be sufficiently large enough to rule out a clinically relevant risk.

A recent large Swedish study compared 33,514 children that had received one anesthetic before age four with 159,619 matched controls.¹⁶ They also included a subgroup of 3,640 children who have been exposed to multiple anesthetics. The primary outcome was school grades at age 16 yr. In a subgroup, they also compared the IQ in boys that were tested as part of their national military service. The analysis was adjusted for sex, month of birth, gestational age, Apgar score, parental education, household income, cohabiting parents, and number of siblings. They found strong evidence of a small difference. One anesthetic was associated with a 0.41% (95% CI, 0.12 to 0.70%) lower score in school grades. In those with IQ scores, one anesthetic was associated with a 0.97% (95% CI, 0.15 to 1.78%) lower IQ score. The impact was greatest with ear, nose, and throat surgery, and interestingly, the impact was greater in children exposed at an older age. There was also some evidence that the impact was greater with multiple anesthetic exposures, as compared to single anesthetic exposure. Exposure to two anesthetics was associated with a 1.41% (95% CI, 0.50 to 2.31%) lower score in school grades and having three or more was associated with a 1.82% (95% CI, 0.15 to 3.49%) lower score. While this study did find evidence of an association, it is important to note that the added risk was very small – a difference of 1% or less in school grades. This is compared to a difference of 10% in school grades that is associated with sex or maternal education.

Two very similar Canadian studies examined the association between surgery in early childhood and the Early Development Index (EDI). The EDI is a test of readiness for school and is administered at around 5 yr of age. It has five domains (physical health and well being, social knowledge and competence, emotional health and maturity, language and cognitive development, and communication skills and general knowledge).

The first study, from Ontario, matched 28,366 children that had surgery before the EDI, with 55,910 controls.¹⁷ They excluded children with physical disability, health-related causes of impaired development, and any diagnosis of a behavioral or learning problem. Children were matched on gestational age, maternal age, rurality, sex, and year and quartile of birth. The primary outcome was any EDI domain score below the tenth percentile. The analysis was adjusted for aboriginal status, age and household income. They found weak evidence for a small difference in the percentage of children with one or more domain scores below the tenth percentile. There were 25.6% with such low scores in the surgical group compared with 25.0% in the controls. The difference was largest in the physical health and well-being, and social knowledge and competence domains. In sub-analyses, there was evidence for a difference in scores in

children aged 2 to 4 yr at time of surgery, but insufficient power to conclude whether or not there was any difference in scores in the 0 to 2 yr old group. They found no evidence that number of surgeries had an impact. The other Canadian study from Manitoba compared 4,470 children that had surgery before the age of four with 13,586 matched controls, excluding children with any diagnosed developmental disability.¹⁸ They matched children according to gestational age, maternal age, rurality, income quartile, sex, and year of birth; their analysis adjusted for welfare status, being small for age at birth, maternal age, child's age, and John Hopkins Resource Utilization Band. They found strong evidence of a very small difference, with surgical children doing slightly worse. The effect was greatest in communication skills and general knowledge, and language and cognitive development domains. Of note, the developmental areas affected were different compared to the other Canadian study. The Manitoba study also found no evidence of a difference between single and multiple exposure, but did find strong evidence of an interaction between age of exposure and outcome, with the risk greater in older children. Studies using databases with large numbers of patients, such as the Canadian studies and the Swedish study, can detect differences that are very small in magnitude, but are statistically significant. The clinical significance of the very small added risk reported in these studies remains uncertain.

In summary, most, but not all, large population-based studies looking at school performance have found evidence of small differences in performance in children that had surgery in early childhood. The increased risk is considerably smaller than other factors that have an impact on performance. The studies do not indicate that exposure at 0 to 2 yr is worse than 2 to 4 yr (indeed, some show the opposite). They are mostly insufficiently powered to determine if multiple exposures pose any greater risk than single exposures. Interestingly, some studies have found that surgery increased the risk of nonattainment without a difference in mean scores. This may be consistent with exposure being associated with a small increase in risk for more severe learning disabilities. This is addressed further in the studies in the next section.

As mentioned above, school assessment may be insensitive. Using the Raine birth cohort, Ing *et al.* examined academic performance, clinical diagnoses, and direct neuropsychologic testing results.¹⁹ They demonstrated that academic performance measures using standardized test scores were the least sensitive, and those from direct assessment using validated neuropsychologic instruments were most sensitive in detecting a difference between exposure to surgery and no exposure.

Clinical Outcomes or Diagnoses as Outcome Measures

Clinical outcomes and diagnoses of particular behavioral or learning disabilities are also important outcomes for the child. Like school grades, they too can be accessed

using data linkage in large population-based studies. Table 2 summarizes the studies that used these outcome measures. Attention Deficit Hyperactivity Disorder (ADHD) and autism both have genetic risk factors; however, a variety of early life insults have also been found to be associated with an increased risk of these disorders.²⁰ It is therefore biologically plausible to explore an association with surgery and anesthesia. It is important to note the limitation of these studies. Definitions of behavioral disorders and learning disability change over time and are inconsistently applied across populations. This makes comparisons between studies difficult and limits the generalization of results.

A learning disability in reading, verbal language, or mathematics, or a clinical outcome/diagnosis were outcome measures used in several studies that examined the effects of early childhood exposure using the Olmsted County birth cohort from 1976 to 1982.²¹ The first study had a total of 539 children who underwent a total of 875 procedures that required general anesthesia before the age of four.²¹ Anesthetics used were predominantly halothane and nitrous oxide (88% received halothane, 91% received nitrous oxide), and ketamine was used in most of the remaining cases (9% received ketamine). The authors reported an increased risk for learning disability associated with multiple, but not single, anesthetic exposure. To further determine whether the frequency of exposure may be a consequence of the health status of the child, the authors used the same cohort in a second study but conducted a matched cohort design in the subgroup that had general anesthesia before age two.²² They used two methods to adjust health status: the American Society of Anesthesiologists Physical Status assignment and the Hopkins Adjusted Clinical Groups case-mix system. Their results were the same as in the original study, showing that multiple, but not single, anesthetic exposure was associated with increased risk of learning disability. In addition, there was also demonstrable increased risk for the need of language Individualized Education Programs. The same authors also used the Olmsted County birth cohort from 1976 to 1982 in another study, but the clinical condition of ADHD was the outcome measure.²³ They found multiple, but not single, anesthetic exposures before age two increased the risk for ADHD. The same group recently performed another study looking at children from Olmsted County exposed to surgery between 1996 and 2000 (when they would have had more modern anesthetic care). In this study they also found that multiple, but not single, exposures increased the risk of a diagnosis of a learning disability and ADHD as compared to children that had no exposure to surgery and anesthesia. Multiple and single exposures were also associated with decreases in some aspects of academic achievement.²⁴ The Olmsted County cohort studies have consistently found that multiple, but not single, exposures to anesthesia in early childhood increased the risk of some learning disabilities and subsequent diagnoses of ADHD.

Using the International Classification of Diseases, Ninth Revision (ICD-9) coded diagnoses for developmental delay, mental retardation, autism spectrum disorders, speech/language problems, and behavior problems as the outcome variable, DiMaggio *et al.* created a birth cohort from all children enrolled in the New York State Medicaid program from 1999 to 2001. The ICD-9 code for inguinal hernia surgery before age 36 months was used as the exposure variable.²⁵ For comparison, the authors used a group derived from random sampling, matched for age that did not have the procedure code for inguinal hernia surgery. In their analyses, they adjusted for low birth weight, perinatal hypoxia, perinatal infections, and central nervous system anomalies. Their conclusion was that inguinal hernia surgery and anesthesia were associated with increased risk of developmental and behavioral disorders. A second study by DiMaggio *et al.* used a birth cohort of twins from the New York State Medicaid dataset (1999 to 2005).²⁶ Any child was considered to have been exposed to anesthesia if there was an ICD-9 procedure code for any type of surgery before 36 months of age, and was included in the analysis if there was no history of developmental disorder at the time of surgery. ICD-9 codes for developmental delay, mental retardation, autism spectrum disorders, speech/language problems, and behavior problems were used as the outcome variable. The analysis was adjusted for low birth weight, perinatal hypoxia, perinatal infections, and neurologic anomalies. The study found that anesthesia exposure and surgery before age three were associated with an increased risk of subsequent diagnosis of developmental or behavior disorders.

A Taiwanese study by Ko *et al.* examined a birth cohort of 114,435 children, amongst whom 5,197 received general anesthesia before 2 yr of age.²⁷ These were matched with 20,788 unexposed children. They found no evidence of increased risk of autism in the exposed group, after single or multiple exposures. In another manuscript, Ko *et al.* also examined the association with ADHD.²⁸ In a cohort of 16,465 children, 3,293 children exposed to anesthesia before the age of three were matched to unexposed children. Unlike the Olmsted County studies, the Taiwanese study found no evidence for increased incidence of ADHD in the exposed group, after multiple or single exposures.

In summary, several, but not all studies, have found evidence for an association between surgery and anesthesia in early life, and increased risk of behavioral disorder or learning disability diagnoses. The association is greater with multiple exposures.

Neuropsychologic Testing Results as Outcome Measures

Table 3 summarizes the studies that have used neuropsychologic tests as outcomes. There are many different neuropsychologic tests available, and they are divided into domains: intelligence, language, learning and memory, visual-spatial skills, attention and executive functions, and motor and psychomotor abilities.⁷ Broadly speaking, there are apical tests, such as the intelligence quotient (IQ), which measure

Table 2. Published Studies with Clinical Diagnosis as Main Outcome Measures

Author, Yr (Setting)	Design	Data Source	Study Subjects		Age of Exposure	Main Outcomes
			Exposed	Unexposed		
Wilder, 2009 (Mayo Clinic, Minnesota, USA)	Retrospective	Population-based birth cohort in Olmsted County (1976–1982)	Children had anesthetic at Mayo Clinic (n = 593)	Children had no anesthesia at Mayo Clinic (n = 4,764)	< 4 yr	Age ≥ 5 yr ≥ 2 exposures Increased risk for learning disability in reading OR written language OR math 1 exposure No difference in learning disability Increased risk for diagnosis of behavioral or developmental disorder
DiMaggio, 2009 (New York, USA)	Retrospective	Birth cohort from New York State Medicaid (1999–2002)	Diagnostic code for inguinal hernia (n = 383)	No inguinal hernia (n = 5,050)	< 3 yr	Increased risk for diagnosis of behavioral or developmental disorder
DiMaggio, 2011 (New York, USA)	Retrospective	Sibling birth cohort from New York State Medicaid (1999–2005)	Diagnostic code for any procedure (n = 304)	No codes for any procedure (n = 10,146)	< 3 yr	Increased risk for diagnosis of behavioral or developmental disorder
Flick, 2011 (Mayo Clinic, Minnesota, USA)	Retrospective	Olmsted County birth cohort (1976–1982)	Children who had anesthesia for all procedures (n = 350)	No anesthesia, matched for age (n = 700)	< 2 yr	Age ≥ 5 yr No difference in need for individualized education program for emotional/behavior disorder ≥ 2 exposures Increased risk for learning disability in reading OR written language OR math 1 exposure No difference in learning disability
Sprung, 2012 (Mayo Clinic, Minnesota, USA)	Retrospective	Olmsted County birth cohort (1976–1982)	Children who had anesthesia for all procedures (n = 350)	No anesthesia (n = 5,007)	< 2 yr	Age ≥ 5 yr ≥ 2 exposures Increased risk for identified as having ADHD 1 exposure No difference in ADHD
Bong, 2013 (Singapore)	Retrospective	Hospital birth registry (1998–1999)	ASA Physical Status I and II Anesthesia for minor surgery (n = 100)	Age-matched with no anesthesia (n = 100)	≤ 1 yr	Age 12 yr Increased risk for formally receiving a diagnosis of learning disability
Ing, 2014 (Western Australia)	Retrospective	Raine birth cohort (1989–1992)	Anesthesia for any procedure (n = 112)	No anesthesia (n = 669)	≤ 3 yr	Age 10 yr Increased risk for diagnosis of behavioral or language disorders
Ko, 2014 (Taiwan)	Retrospective	Birth cohort from National Health Insurance Research Database of Taiwan	GA exposure and without diagnosis of ADHD (n = 3293)	No anesthesia exposure (n = 13,172)	< 3 yr	Database from 2005 and 2010 No increase in risk for diagnosis of ADHD
Ko, 2015 (Taiwan)	Retrospective	Birth cohort from National Health Insurance Research Database of Taiwan (2001–2007)	GA exposure and without diagnosis of autistic disorders (n = 5197)	No anesthesia exposure (n = 109,238)	< 2 yr	Database from 2005 and 2010 No increase in risk for diagnosis of autistic disorder
Hu, 2017 (Mayo Clinic, Minnesota, USA)	Retrospective	Olmsted County birth cohort (1996–2000)	Children who had anesthesia for all procedures (n = 573)	No anesthesia (n = 463)	< 3 yr	Up to age 18 yr ≥ 2 exposures Increased risk for ADHD and LD Deficits in achievement test scores 1 exposure No differences in ADHD and LD Deficits in reading achievement test

ADHD = Attention Deficit Hyperactivity Disorder; ASA = American Society of Anesthesiologists; GA = glutaraldehyde; LD = learning disability.

Table 3. Published Studies with Neuropsychologic Tests as Main Outcome Measure

Author, Yr (Setting)	Design	Data Source	Study Subjects		Age of Exposure	Main Outcomes
			Exposed	Unexposed		
Stratmann, 2014 (San Francisco, USA)	Ambidirectional	UCSF clinical database	Procedure with anesthesia (n = 28)	Age and sex-matched (n = 28)	< 1 yr	Age 6–11 yr No difference in IQ or behavior, or familiarity memory Impaired recollection memory
Ing, 2012 (Western Australia)	Retrospective	Raine birth cohort (1989–1992)	Exposure to anesthesia/surgery (n = 321)	No exposure to anesthesia/surgery (n = 2,287)	< 3 yr	Age 10 yr Expressive and receptive language scores and abstract reasoning by Colored Progressive Matrix
Ing, 2014 (Western Australia)	Retrospective	Raine birth cohort (1989–1992)	Exposure to anesthesia/surgery Age 3–5 yr (n = 220) Age 5–10 yr (n = 375)	No exposure (n = 1,952)	3–10 yr	Age 10 yr No difference in cognitive ability or behavior Difference in motor function by MAND
Backeljauw, 2015 (Cincinnati, USA [Cincinnati Children's])	Retrospective	Language development MRI study at Cincinnati Children's	Anesthesia exposure (n = 53)	No anesthesia (n = 53)	< 4 yr	Age 5–18 yr Lower scores in comprehension language Lower performance IQ
Kermary, 2016 (Iran)	Retrospective	Medical and anesthesia records with phonemic and semantic verbal fluency, digit span test results	6–15 yr old glaucoma patients with history of anesthesia (n = 22 single anesthetic, n = 46 ≥ 2 anesthetics)	6–15 yr old glaucoma patients with no previous anesthesia (n = 47)	< 3 yr	≥ 6 months after last anesthetic No difference in cognitive ability
De Heer, 2016 (The Netherlands)	Retrospective	Generation R cohort study; Parental response on questionnaire regarding anesthesia exposure and in children with available IQ scores	Received anesthesia before age 5 yr according to parental response on questionnaire and had available IQ scores (n = 415)	No anesthesia by parental report and had available IQ scores (n = 3,026)	< 5 yr	Age 6 yr IQ scores were 2.7 (1.2–4.2, 95% CI) lower in exposed compared to unexposed Only 35.2% data were included in analysis, data were missing for 64.8%
Sun, 2016 (Multi-site PANDA study; USA)	Ambidirectional	Enrolled sibling-matched cohort in New York, Boston, Nashville, and Philadelphia	ASA Physical Status I or II GA ≥ 36 weeks inguinal hernia Single episode (n = 105)	Sibling of exposed (n = 105)	< 3 yr	Age 8–15 yr No difference in scores: IQ, memory, learning language, attention, executive function, motor/processing speed, visuospatial function, behavior
Davidson, 2016 (International; GAS trial)	Randomized trial	Enrolled infants for hernia surgery, 27 tertiary hospitals	Sevoflurane anesthesia (n = 238)	Awake regional anesthesia (n = 294)	≤ 60 weeks postmenstrual age	Age 2 yr No difference in all Bayley-III measures

ASA = American Society of Anesthesiologists; GA = glutaraldehyde; MAND = McCarron Assessment of Neuromuscular Development; MRI = magnetic resonance imaging; PANDA = Pediatric Anesthesia NeuroDevelopment Assessment; UCSF = University of California, San Francisco.

and collate function over a range of domains, and other tests that focus on particular subdomains or functions. Apical tests have better psychometrics and can more accurately predict a child's overall future functioning. Tests focusing on particular subdomains may indicate a deficit that is not detectable with apical tests; however, the relevance of that deficit on future function is less certain. Usually, a battery of tests is applied. These will be a mix of apical tests that have clearer implications for the child's future, and other tests that focus on particular sub-domains of interest. The choice of domains that are examined in more detail should be driven by the findings from preclinical studies. Preclinical studies have suggested that cognition, learning, memory, and executive function should be of particular interest.

Testing at an older age is generally preferable as tests in older children have greater capacity to predict future function. Some domains, such as executive function and some aspects of memory, are not fully developed until the child is older. Some children will also "grow into" a defect; wherein the defect is only apparent as the period between exposure and test lengthens. In other cases, a child may recover over time and the defect becomes undetectable. Using neuropsychologic tests as outcomes is logistically more difficult and considerably more expensive than using data linkage to identify school grades or diagnoses.

Several studies have made use of data already collected in various birth cohorts. The Western Australian Pregnancy Cohort (Raine) study enrolled 2,900 pregnant women in their early pregnancy.²⁹ The Raine birth cohort includes 2,868 children born to these mothers. The health and other related data in these children have been reviewed in detail at ages 1, 2, 3, 5, 8, 10, 14, 17, 18, 20, and 23 yr. Direct neuropsychologic testing and parental interviews occurred at these different ages. The largest number of tests were performed at age 10, and included testing for cognitive function (using Symbol Digit Modality Test and Raven's Colored Progressive Matrices), language (using Clinical Evaluation of Language Fundamentals Peabody Picture Vocabulary), and motor function (using McCarron Assessment of Neuromuscular Development), as well as reports of behavior (using Child Behavior Checklist). A total of 2,608 children, age 10, were included in the analysis. Ing *et al.* divided the children into an exposed cohort: those who had received surgery/anesthesia before age 3 yr ($n = 321$) and those who had not ($n = 2,287$);²⁹ in the analysis adjusted for family income, maternal education, and birth weight. The exposed cohort had significantly lower scores in language, receptive, expressive, and total, as well as in abstract reasoning, but not in any other neuropsychologic domains or in behavior. This was found with either single or multiple episodes of anesthesia exposure. In another analysis of this cohort, Ing *et al.* examined outcomes in children exposed between the ages of 3 and 5 yr, and 5 and 8 yr, as compared to unexposed children.³⁰ In both of these relatively older age groups, there was no evidence of worse cognitive and language outcomes

in the exposed groups; however, both exposed age groups did have decreased motor function.

De Heer *et al.*³¹ examined data from the Dutch birth cohort, "Generation R." The cohort consisted of 9,901 children born between 2002 and 2006. The outcome of interest was nonverbal IQ measured at age six. IQ data was available for 3,441 children. Of these, 415 had been exposed to anesthesia before the age of five. The authors reported an association between anesthesia and IQ, after adjusting for sex, prematurity, maternal education, IQ, smoking history, and alcohol use.

Stratmann *et al.*³² studied 28 children aged 6 to 11 yr old who had general anesthesia exposure before age one. They compared recollection and familiarity memory, IQ scores, and behavior with 28 age- and sex-matched control children. They found impaired recollection memory, but otherwise no differences were found between groups.³² Backeljauw *et al.*, from Cincinnati Children's Hospital, used a cohort of healthy participants, ages 5 to 18, in a language development magnetic resonance imaging study that examined the effects of anesthesia and surgery before age four.⁸ The comparison group was matched for age, sex, handedness and family income. The exposed group ($n = 53$) scored lower than the control group ($n = 53$) in performance IQ scores, and in listening comprehension. This was for both single and multiple episodes of anesthesia exposure. Another small study compared 68 children who had surgery before 3 yr of age for glaucoma, with 47 children that had not.³³ They also examined a subgroup of children that had received multiple anesthetics. They found no evidence of differences when the children were tested for verbal fluency or digit span tests when tested at 5 to 16 yr of age.

The cognitive outcomes following a single and relatively brief anesthesia exposure were examined in the Pediatric Anesthesia NeuroDevelopment Assessment (PANDA) study in a sibling-matched cohort of healthy children.³⁴ The PANDA study is an ambidirectional cohort study of American Society of Anesthesiologists Physical Status I or II children who had a single episode of general anesthesia for inguinal hernia surgery at 36 months or before, as compared with their siblings who received no anesthesia or sedation before age 36 months. Use of siblings as controls is a powerful technique to reduce many of the confounding influences. The median duration of exposure was 80 min. The assessment was performed in both siblings using a comprehensive neuropsychologic battery between 8 to 15 yr of age. One-hundred-five sibling pairs were recruited (mean age: 17.3 months at surgery/anesthesia); with 95 males/10 females in the surgical group and 59 males/46 females in the unexposed siblings group. Full scale IQ was the primary outcome. The mean IQ was similar in both groups: 111 in the exposed group and 111 in the unexposed group, with a difference between groups (exposed – unexposed) of only 0.2 points (95% CI, -2.6 to 2.9). There was also no evidence for any significant difference in IQ sub-domains with a difference of 0.5 (95% CI, -2.7 to 3.7) in performance IQ and -0.5 (95% CI, -3.2 to 2.2) in the verbal IQ. There was

no evidence of any significant difference in the other tests of memory/learning, motor/processing speed, visuospatial function, attention, executive function, and language. In the unadjusted analysis, differences were seen in some aspects of behavior and verbal fluency, but these differences were no longer apparent once adjustments were made for sex. In a secondary analysis, proportions of children scoring below clinically relevant cut offs for behavior were also compared. Twenty-one percent of exposed and 10% of unexposed siblings did have abnormal Child Behavior Check List internalizing scores (greater than 60), which was statistically significant, after adjusting for sex. The significance of this result should be regarded with caution, as this was a secondary analysis of a secondary outcome. Subanalyses looking at those with longer exposure found no evidence of differences in IQ in those exposed to up to 120 min of anesthesia, and age of exposure had no impact on the outcome.

In the PANDA study, IQ scores had mean differences of 0.2 to 0.5 points and the 95% CI around the differences are within ± 4 points. This is because the PANDA study was designed to detect differences that have been reported to be significant in developmental neurotoxicology studies. Thus, it was adequately designed to provide evidence to rule out that a single episode of anesthetic exposure is likely to have any adverse effects on neurodevelopment.

Both single and multiple episodes of anesthesia exposures are being examined in the Mayo Anesthesia Safety in Kids study, which also uses an ambidirectional study design. The study's population included children who received general anesthesia before age three, either as a single episode of exposure or with multiple episodes of exposure.³⁵ The control group is a propensity age-matched group. Testing of the study subjects occurred at two later ages (age 8 to 12 yr or 15 to 19 yr). The neuropsychologic battery included the Operant Test Battery, a test that was used in an earlier nonprimate study. The results of the Mayo Anesthesia Safety in Kids study are still pending.

In summary, there is mixed evidence for an association between anesthesia exposure and deficits in neuropsychologic testing. Some well-powered and carefully conducted studies such as PANDA found no evidence of association with any deficit. Other studies found deficits in IQ, language, abstract reasoning, or some aspects of memory.

Surgery in the Neonatal Period

There have been several cohort studies that specifically examined the impact of surgery in the neonatal period.^{36,37} In one study, more than half of children with tracheoesophageal fistula repair had neurodevelopmental delays requiring referral to early intervention services.³⁸ Similarly, children who have had congenital diaphragmatic hernia repair had a high rate of poor neurodevelopmental outcomes,³⁹ and extremely premature neonates who underwent laparotomy had poorer neurodevelopmental outcomes compared with matched controls.⁵ Another study looked at surgery in very preterm infants, and found that when compared to a nonsurgical

group, the surgical group had lower mental development index scores, lower brain volumes, smaller deep nuclear gray matter volumes, and more white matter injury.⁴⁰ However, there was no evidence of differences in mental development scores when adjusted for potential confounding influences. In another matched cohort study, infants who underwent major surgery had poorer school grades compared to a matched control group of infants who had major nonsurgical medical conditions.⁶ In children weighing less than 1,000 g at birth, neurologic impairment was present in more children who had undergone surgery for patent ductus ligation compared with those that had received medical therapy.⁴¹ In another study of extremely preterm infants, the IQ of those who had undergone surgery was lower at 5 yr of age and they exhibited more sensorineural disability than those who had not undergone surgery.⁴ A recent systematic review that examined 23 studies found developmental delays in 23% of children that had neonatal surgery for non-cardiac conditions.³⁷ These studies used the Bayley Scales of Infant Development as a measure of neurodevelopment and when compared to population normative data, the meta-analysis of these studies found evidence of delays in both motor and cognitive subscales. While the majority of these studies show good evidence for significantly increased risk of poorer neurodevelopmental outcomes in neonates that had major surgery, nearly all these studies involve neonates that have many other risk factors for poor neurodevelopment, including presence of syndromes or premature birth.

Cardiac Surgery

Cardiac surgery in infants often requires long surgery with substantial exposure to anesthetic agents; however, children having surgery for congenital heart disease are at risk for adverse neurodevelopmental outcomes for many reasons. In neonatal cardiac surgery, Andropoulos *et al.* found an association between the dose of volatile anesthetic and neurodevelopmental outcome measured at 12 months of age.⁴² In contrast, Guerra *et al.* found no evidence of an association between cumulative dose of sedative/analgesia used before, during, or after cardiac surgery in infants younger than 6 months of age, and adverse neurodevelopmental outcome measured at 18 to 24 months of age.⁴³ However, when assessed at 4 yr of age they did find evidence of an association between cumulative dose of chloral hydrate and lower performance IQ, and between cumulative dose of benzodiazepines and lower visual motor integration scores.⁴⁴ In both studies it should be noted that children who required longer surgery or longer sedation were more likely to be the children that were sicker and had more major surgeries. Cohort studies cannot determine the impact of anesthesia exposure in children who have cardiac surgery.

Limitations of Cohort Studies: Confounding Factors

All previously reported studies are cohort studies. Cohort studies are inherently limited due to the possibility of

confounding factors. Confounding occurs when another factor directly influences both the likelihood of needing anesthesia and neurodevelopmental outcome, resulting in the false assumption that anesthesia causes the neurodevelopmental outcome. It is difficult to eliminate known or probable confounding factors in studies examining possible effects of anesthetics on neurodevelopment. Children receive anesthetics for a procedure or surgery. The procedure or surgery is often being done because the child has a condition that may directly affect neurodevelopment or is associated with conditions that affect neurodevelopment (*e.g.*, prematurity, syndromes, chromosomal abnormalities, or cerebral palsy). Other examples of potential confounding influences are that children with poor hearing are more likely to have a myringotomy to optimize their hearing, or children from low socioeconomic circumstances are more likely to have poor dentition. Similarly, uncooperative children with a yet to be diagnosed behavioral problem may be more likely to need anesthesia for procedures that healthy children may tolerate awake. The surgery itself may also have an impact on neurodevelopment. Major surgery is associated with a significant inflammatory response which may have an impact on the developing brain. There are other perioperative factors such as pain, low cerebral perfusion, hypoxia, and electrolyte disturbances that could all impact brain development. In summary, there are a great many possible sources of confounding factors in the association between anesthesia and neurodevelopmental outcome.

Careful patient selection, matching, and adjusted analyses can reduce some of the effects of known confounding; however, the matching and adjustments are not perfect and no adjustment can reduce the possible impact of unknown confounding factors. Most importantly, the studies cannot adjust for the confounding effect of the surgery itself. Thus, a positive finding of association in a cohort study can never be assumed to be due to the effect of the anesthetic. Sibling or twin matching is perhaps the best way to reduce the confounding influences of genetics and environment. It is important to note that the PANDA study, which used sibling matches, and the Bartels twin study found no evidence for an association.

Trials

The most effective way to reduce non-random confounding influences is with a randomized controlled trial. To reduce the confounding effect of surgery, the trial needs to randomly allocate surgical patients to anesthesia and no anesthesia. This is clearly not feasible, but a spinal anesthesia is a feasible alternative in some circumstances. It is unlikely that spinal anesthesia has any direct toxic effects on the brain.⁴⁵ This was the rationale behind the General Anesthesia compared to Spinal anesthesia trial (GAS).⁴⁶ Seven-hundred-twenty-two infants under 60 weeks postmenstrual age were randomized to awake-regional (predominantly spinal) or sevoflurane-based general anesthesia for inguinal

hernia repair. The median duration of anesthesia in the general anesthesia group was 54 min. The primary outcome was full scale IQ using the Wechsler Preschool and Primary Scale of Intelligence, Third Edition to be assessed at 5 yr of age. These results will not be available until early 2018. Neurodevelopment assessed by the Bayley-III at 2 yr of age was a predefined secondary outcome. The Bayley-III has five domains: cognitive, language, motor, social emotional, and adaptive behavior. Each domain has a normalized mean of 100 and a SD of 15 points. In the GAS trial, the difference in the cognitive composite score was 0.17 points (95% CI, -2.30 to 2.64). This was within the predefined equivalence margin of 5 points, which provides strong evidence for no difference between groups, and is well within any margin that would be regarded as clinically relevant. There was also strong evidence of equivalence in all the other four domains of the Bayley-III. There was no difference in results comparing intention-to-treat and as-per-protocol analyses, implying that the 19% failed spinal cases did not bias the outcome, and no difference between multiple imputation and complete case analyses, implying the 15% loss to follow-up did not bias the results either. It must be stressed, however, that neurodevelopmental testing at 2 yr of age is inherently limited for higher executive function and some aspects of memory.⁴⁷ Thus, while the preliminary results of the GAS trial found strong evidence of no added risk associated with an hour of general anesthesia as compared to spinal anesthesia, the results are not definitive.

Collating the Human Data

It is not possible to make a single definitive conclusion as to whether or not the human evidence supports or refutes the possibility that anesthetic exposure in children causes adverse effects on neurodevelopment. As Ted Eger pointed out previously, it is essentially impossible to prove absolutely that a technique is always safe, or a drug completely non-toxic.⁴⁸ No number of negative studies will ever prove that anesthetics have *no* impact on neurodevelopment; however, human studies can provide some idea of the likelihood for an underlying causative association, which populations and domains might be most at risk, and which strategies reduce the risk.

Overall human studies have found mixed evidence for an association between anesthesia exposure in early life and neurodevelopmental outcome. This is not inconsistent with the underlying effect, given the range of outcomes and populations studied. Some, but not all, large population-based studies have found evidence for a small difference in tests of academic achievement and school readiness, while some studies have found an increased risk of not being able to be tested. The difference in school grades is small and unlikely to have a measurable impact on a child's wellbeing. The added risk is far less than other factors such as sex or maternal education. Similarly, some but not all studies, have found evidence of an association between surgery and

anesthesia in early life, and increased risk of a diagnosis of a behavioral disorder or learning disability. The added risk is small, but given the implications of such a diagnosis, the overall impact on society is potentially worrisome. Lastly, there is mixed evidence of an association between anesthesia exposure and poorer outcome in some domains of neuropsychological testing.

Confounding factors are the greatest limitation for all the human cohort studies. The GAS trial is the only trial where randomization would minimize confounding factors, and the PANDA and Bartels studies are perhaps the most carefully matched cohort studies. Both the GAS and PANDA studies found no evidence of any difference in neurodevelopmental outcome in children having less than 2 h of anesthesia in infancy.

FDA Warning

The FDA warns that anesthesia exposure in children younger than 3 yr of age having long duration or multiple exposures may have an impact on neurodevelopment. The age limit is presumably derived from both preclinical and human data. There is very little human data to support or refute using 3 yr of age as the limit. Most human studies have focused on younger children so there is very limited data on children over 3 yr of age. A few studies have examined different age subgroups, and those have not consistently found evidence for greater risk in younger children. The impact of multiple exposures compared to single exposure is also unclear. Several, but not all, studies have found a greater impact associated with multiple exposures, and most studies do not have sufficient power to differentiate between single and multiple exposures. The greater impact with multiple exposures may also be explained by the greater influence of confounding factors—children that have more comorbidities require more surgeries. Human studies also shed little light on what duration is “safe.” Most anesthetics in children are under 2 h in duration. Thus, the large population-based studies would have a considerable number of these relatively “short” exposures. If the effect were duration dependent, this could explain the modest effect sizes seen in these studies. Some studies have looked at the impact of duration of exposure; PANDA found no evidence of a difference in outcome when comparing less than 1 h and 1 to 2 h exposure. There are very few human data about exposures greater than 3 or 4 h. Relatively few children have long procedures; longer procedures, like multiple procedures, may be associated with a greater influence of confounding factors. Many of the cases described in the neonatal and cardiac studies would be of long duration; however, these are also the groups with potentially the greatest likelihood of confounding influences.

In summary, there is only extremely weak human evidence to support the FDA warning that repeated or lengthy use of anesthetic drugs may affect the development of children's brains. The warning should be regarded as being largely based on the extensive and much more robust evidence

from preclinical studies. There is, however, more substantial human evidence to support the FDA statement that single, relatively short exposures are not associated with increased risk. There is very little, if any, human evidence to support a recommendation that a particular age is safe or unsafe.

Other Implications of the Human Studies

If the results from the preclinical studies were to be completely ignored, then what could be concluded from the results of the human studies alone? The majority of the studies found some association between surgery in early childhood and increased risk of adverse neurodevelopmental outcomes. The association may be due to pathology, or indication for surgery, but it may also be due to other perioperative factors which are under the control of the anesthesiologist. The PANDA study results could indicate that these perioperative factors are not likely to be a problem for healthy infants having short procedures, but they may still be important in other populations. When a detailed examination of the association between surgery and neurodevelopmental outcome is performed, considerations should be given to all potentially reversible causative factors, and not just neurotoxicity.

Future Directions

To further define whether or when anesthetics have a direct impact on neurodevelopment requires more high quality human studies, especially in children having prolonged or repeated exposure to anesthesia. Ideally, these would be randomized trials in healthy infants comparing anesthesia regimens that do and do not produce the changes seen in preclinical studies. These would not be easy trials to perform. Particular problems are the paucity of healthy infants having long procedures or repeated procedures, the uncertainty over which “nontoxic” anesthetic regimens would be clinically feasible, and the long delay between randomization and the ideal outcome measure. Another useful line of research would be to do larger and more detailed cohort studies to identify and further characterize those at greatest risk and determine what psychometric domains are most affected. Due to confounding factors, these more detailed cohort studies will always have limited capacity to provide evidence about whether or not anesthetics directly impact on neurodevelopment, but they may give a clue as to the other possible causes of the poor outcome. They will also provide valuable information about how to design future clinical trials comparing different strategies to reduce any impact on neurodevelopment.

Conclusions

The human studies provide mixed evidence of an association between anesthesia exposure in early childhood and later deficits in a range of neurodevelopmental outcomes. When added risk has been observed, it is very small. The variations in examined outcomes and generally small differences seen

in human studies are not inconsistent with the preclinical data given the predominantly short exposures and ranges of populations and outcomes assessed. However, the strong likelihood of confounding influences in these studies, which are predominantly cohort studies, means that the human evidence for any association can only be regarded as very weak evidence that anesthesia *actually causes* these poorer outcomes. Thus, any recommendations for changing practice, including the FDA warning, continue to be driven largely by the preclinical evidence. In contrast, there is stronger human evidence that a single brief exposure in a healthy infant is not associated with poorer neurodevelopmental outcome.

Research Support

Support was provided solely from institutional and/or departmental sources.

Competing Interests

The authors declare no competing interests.

Correspondence

Address correspondence to Dr. Sun: College of Physicians and Surgeons, Columbia University, CH 4-440 North, 622 West 168th Street, New York, New York 10032. lss4@cumc.columbia.edu. Information on purchasing reprints may be found at www.anesthesiology.org or on the masthead page at the beginning of this issue. ANESTHESIOLOGY's articles are made freely accessible to all readers, for personal use only, 6 months from the cover date of the issue.

References

- Vutskits L, Xie Z: Lasting impact of general anaesthesia on the brain: Mechanisms and relevance. *Nat Rev Neurosci* 2016; 17:705-17
- Lin EP, Lee JR, Lee CS, Deng M, Loepke AW: Do anesthetics harm the developing human brain? An integrative analysis of animal and human studies. *Neurotoxicol Teratol* 2017; 60:117-28
- Montana MC, Evers AS: Anesthetic neurotoxicity: New findings and future directions. *J Pediatr* 2017; 181:279-85
- Surgery and the tiny baby: sensorineural outcome at 5 years of age. The Victorian Infant Collaborative Study Group. *Journal of paediatrics and child health* 1996; 32:167-72
- Chacko J, Ford WD, Haslam R: Growth and neurodevelopmental outcome in extremely-low-birth-weight infants after laparotomy. *Pediatr Surg Int* 1999; 15:496-9
- Ludman L, Spitz L, Wade A: Educational attainments in early adolescence of infants who required major neonatal surgery. *J Pediatr Surg* 2001; 36:858-62
- Beers SR, Rofey DL, McIntyre KA: Neurodevelopmental assessment after anesthesia in childhood: Review of the literature and recommendations. *Anesth Analg* 2014; 119:661-9
- Backeljauw B, Holland SK, Altaye M, Loepke AW: Cognition and brain structure following early childhood surgery with anesthesia. *Pediatrics* 2015; 136:e1-12
- 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
- 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
- Hansen TG, Pedersen JK, Henneberg SW, Morton NS, Christensen K: Educational outcome in adolescence following pyloric stenosis repair before 3 months of age: A nationwide cohort study. *Paediatr Anaesth* 2013; 23:883-90
- Clausen NG, Pedersen DA, Pedersen JK, Møller SE, Grosen D, Wehby GL, Christensen K, Hansen TG: Oral clefts and academic performance in adolescence: The impact of anesthesia-related neurotoxicity, timing of surgery, and type of oral clefts. *Cleft Palate Craniofac J* 2017; 54:371-80
- Block RI, Thomas JJ, Bayman EO, Choi JY, Kimble KK, Todd MM: Are anesthesia and surgery during infancy associated with altered academic performance during childhood? *ANESTHESIOLOGY* 2012; 117:494-503
- Williams RK, Black IH, Howard DB, Adams DC, Mathews DM, Friend AF, Meyers HW: Cognitive outcome after spinal anesthesia and surgery during infancy. *Anesth Analg* 2014; 119:651-60
- Bong CL, Allen JC, Kim JT: The effects of exposure to general anesthesia in infancy on academic performance at age 12. *Anesth Analg* 2013; 117:1419-28
- Glatz P, Sandin RH, Pedersen NL, Bonamy AK, Eriksson LI, Granath F: Association of anesthesia and surgery during childhood with long-term academic performance. *JAMA Pediatr* 2017; 171:e163470
- O'Leary JD, Janus M, Duku E, Wijeyesundera DN, To T, Li P, Maynes JT, Crawford MW: A population-based study evaluating the association between surgery in early life and child development at primary school entry. *ANESTHESIOLOGY* 2016; 125:272-9
- Graham MR, Brownell M, Chateau DG, Dragan RD, Burchill C, Fransoo RR: Neurodevelopmental assessment in kindergarten in children exposed to general anesthesia before the age of 4 years: A retrospective matched cohort study. *ANESTHESIOLOGY* 2016; 125:667-77
- Ing CH, DiMaggio CJ, Malacova E, Whitehouse AJ, Hegarty MK, Feng T, Brady JE, von Ungern-Sternberg BS, Davidson AJ, Wall MM, Wood AJ, Li G, Sun LS: Comparative analysis of outcome measures used in examining neurodevelopmental effects of early childhood anesthesia exposure. *ANESTHESIOLOGY* 2014; 120:1319-32
- Efron D, Vutskits L, Davidson AJ: Can we really suggest that anesthesia might cause attention-deficit/hyperactivity disorder? *ANESTHESIOLOGY* 2017; 127:209-11
- 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
- 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
- Sprung J, Flick RP, Katusic SK, Colligan RC, Barbaresi WJ, Bojanic K, Welch TL, Olson MD, Hanson A, Schroeder DR, Wilder RT, Warner DO: Attention-deficit/hyperactivity disorder after early exposure to procedures requiring general anesthesia. *Mayo Clinic Proc* 2012; 87:120-9
- Hu D, Flick RP, Zaccariello MJ, Colligan RC, Katusic SK, Schroeder DR, Hanson AC, Buenvenida SL, Gleich SJ, Wilder RT, Sprung J, Warner DO: Association between exposure of young children to procedures requiring general anesthesia and learning and behavioral outcomes in a population-based birth cohort. *ANESTHESIOLOGY* 2017; 127:227-40
- 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; 21:286-91
- 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

27. Ko WR, Huang JY, Chiang YC, Nfor ON, Ko PC, Jan SR, Lung CC, Chang HC, Lin LY, Liaw YP: Risk of autistic disorder after exposure to general anaesthesia and surgery: A nationwide, retrospective matched cohort study. *Eur J Anaesthesiol* 2015; 32:303–10
28. Ko WR, Liaw YP, Huang JY, Zhao DH, Chang HC, Ko PC, Jan SR, Nfor ON, Chiang YC, Lin LY: Exposure to general anesthesia in early life and the risk of attention deficit/hyperactivity disorder development: A nationwide, retrospective matched-cohort study. *Paediatr Anaesth* 2014; 24:741–8
29. 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
30. Ing CH, DiMaggio CJ, Whitehouse AJ, Hegarty MK, Sun M, von Ungern-Sternberg BS, Davidson AJ, Wall MM, Li G, Sun LS: Neurodevelopmental outcomes after initial childhood anesthetic exposure between ages 3 and 10 years. *J Neurosurg Anesthesiol* 2014; 26:377–86
31. de Heer IJ, Tiemeier H, Hoeks SE, Weber F: Intelligence quotient scores at the age of 6 years in children anaesthetised before the age of 5 years. *Anaesthesia* 2017; 72:57–62
32. Stratmann G, Lee J, Sall JW, Lee BH, Alvi RS, Shih J, Rowe AM, Ramage TM, Chang FL, Alexander TG, Lempert DK, Lin N, Siu KH, Elphick SA, Wong A, Schnair CI, Vu AF, Chan JT, Zai H, Wong MK, Anthony AM, Barbour KC, Ben-Tzur D, Kazarian NE, Lee JY, Shen JR, Liu E, Behniwal GS, Lammers CR, Quinones Z, Aggarwal A, Cedars E, Yonelinas AP, Ghetti S: Effect of general anesthesia in infancy on long-term recognition memory in humans and rats. *Neuropsychopharmacology* 2014; 39:2275–87
33. Poor Zamany Nejat Kermany M, Roodneshin F, Ahmadi Dizgah N, Gerami E, Riahi E: Early childhood exposure to short periods of sevoflurane is not associated with later, lasting cognitive deficits. *Paediatr Anaesth* 2016; 26:1018–25
34. 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; 315:2312–20
35. Gleich SJ, Flick R, Hu D, Zaccariello MJ, Colligan RC, Katusic SK, Schroeder DR, Hanson A, Buenvenida S, Wilder RT, Sprung J, Voigt RG, Paule MG, Chelonis JJ, Warner DO: Neurodevelopment of children exposed to anesthesia: Design of the Mayo Anesthesia Safety in Kids (MASK) study. *Contemp Clin Trials* 2015; 41:45–54
36. Walker K, Holland AJ, Winlaw D, Sherwood M, Badawi N: Neurodevelopmental outcomes and surgery in neonates. *J Paediatr Child Health* 2006; 42:749–51
37. Stolwijk IJ, Lemmers PM, Harmsen M, Groenendaal F, de Vries LS, van der Zee DC, Benders MJ, van Herwaarden-Lindeboom MY: Neurodevelopmental Outcomes After Neonatal Surgery for Major Noncardiac Anomalies. *Pediatrics* 2016; 137:e20151728
38. Newton LE, Abdessalam SF, Raynor SC, Lyden ER, Rush ET, Needelman H, Cusick RA: Neurodevelopmental outcomes of tracheoesophageal fistulas. *J Pediatr Surg* 2016; 51:743–7
39. Hedrick HL, Crombleholme TM, Flake AW, Nance ML, von Allmen D, Howell LJ, Johnson MP, Wilson RD, Adzick NS: Right congenital diaphragmatic hernia: Prenatal assessment and outcome. *J Pediatr Surg* 2004; 39:319–23
40. Filan PM, Hunt RW, Anderson PJ, Doyle LW, Inder TE: Neurologic outcomes in very preterm infants undergoing surgery. *J Pediatr* 2012; 160:409–14
41. Kabra NS, Schmidt B, Roberts RS, Doyle LW, Papile L, Fanaroff A: Neurosensory impairment after surgical closure of patent ductus arteriosus in extremely low birth weight infants: Results from the Trial of Indomethacin Prophylaxis in Preterms. *J Pediatrics* 2007; 150: 229–34, 234.e1
42. Andropoulos DB, Ahmad HB, Haq T, Brady K, Stayer SA, Meador MR, Hunter JV, Rivera C, Voigt RG, Turcich M, He CQ, Shekerdemian LS, Dickerson HA, Fraser CD, Dean McKenzie E, Heinle JS, Blaine Easley R: The association between brain injury, perioperative anesthetic exposure, and 12-month neurodevelopmental outcomes after neonatal cardiac surgery: A retrospective cohort study. *Paediatr Anaesth* 2014; 24:266–74
43. Guerra GG, Robertson CM, Alton GY, Joffe AR, Cave DA, Dinu IA, Creighton DE, Ross DB, Rebeyka IM; Western Canadian Complex Pediatric Therapies Follow-up Group: Neurodevelopmental outcome following exposure to sedative and analgesic drugs for complex cardiac surgery in infancy. *Paediatr Anaesth* 2011; 21:932–41
44. Garcia Guerra G, Robertson CM, Alton GY, Joffe AR, Cave DA, Yasmin F, Dinu IA, Creighton DE, Ross DB, Rebeyka IM; Western Canadian Complex Pediatric Therapies Follow-up Group: Neurotoxicity of sedative and analgesia drugs in young infants with congenital heart disease: 4-year follow-up. *Paediatr Anaesth* 2014; 24:257–65
45. Hamurtekin E, Fitzsimmons BL, Shubayev VI, Grafe MR, Deumens R, Yaksh TL, Walker SM: Evaluation of spinal toxicity and long-term spinal reflex function after intrathecal levobupivacaine in the neonatal rat. *ANESTHESIOLOGY* 2013; 119:142–55
46. Davidson AJ, Disma N, de Graaff JC, Withington DE, Dorris L, Bell G, Stargatt R, Bellinger DC, Schuster T, Arnup SJ, Hardy P, Hunt RW, Takagi MJ, Giribaldi G, Hartmann PL, Salvo I, Morton NS, von Ungern Sternberg BS, Locatelli BG, Wilton N, Lynn A, Thomas JJ, Polaner D, Bagshaw O, Szmuk P, Absalom AR, Frawley G, Berde C, Ormond GD, Marmor J, McCann ME; GAS consortium: 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; 387:239–50
47. Warner DO, Flick RP: Anaesthetics, infants, and neurodevelopment: Case closed? *Lancet* 2016; 387:202–4
48. Eger EI II: Dragons and other scientific hazards. *ANESTHESIOLOGY* 1979; 50:1