
Publisher's PDF, also known as Version of record

License (if available):
CC BY

Link to published version (if available):
10.1213/ANE.0000000000004052

Link to publication record in Explore Bristol Research
PDF-document

This is the final published version of the article (version of record). It first appeared online via Wolters Kluwer at https://doi.org/10.1213/ANE.0000000000004052. Please refer to any applicable terms of use of the publisher.

University of Bristol - Explore Bristol Research
General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available:
http://www.bristol.ac.uk/pure/about/ebr-terms
Assessing Long-term Neurodevelopmental Outcome Following General Anesthesia in Early Childhood: Challenges and Opportunities

Graham J. Walkden, MBChB,*† Anthony E. Pickering, FRCA, PhD,*† and Hannah Gill, FRCA, PhD*†‡

Neurodegeneration has been reported in young animals after exposure to all commonly used general anesthetic agents. The brain may be particularly vulnerable to anesthetic toxicity during peak synaptogenesis (in gestation and infancy). Human studies of long-term neurodevelopmental outcome following general anesthesia in early childhood report contradictory findings. This review assesses the strengths and deficiencies in human research methodologies to inform future studies. We identified 76 studies, published between 1990 and 2017, of long-term neurodevelopmental outcome following early childhood or in utero general anesthesia exposure: 49 retrospective, 9 ambidirectional, 17 prospective cohort studies, and 1 randomized controlled trial. Forty-nine studies were explicitly concerned with anesthetic-induced neurotoxicity. Full texts were appraised for methodological challenges and possible solutions. Major challenges identified included delineating effects of anesthesia from surgery, defining the timing and duration of exposure, selection of a surgical cohort and intervention, addressing multiple confounding life course factors, detecting modest neurotoxic effects with small sample sizes (median, 131 children; interquartile range, 50–372), selection of sensitive neurodevelopmental outcomes at appropriate ages for different developmental domains, insufficient length of follow-up (median age, 6 years; interquartile range, 2–12 years), and sample attrition. We discuss potential solutions to these challenges. Further adequately powered, multicenter, prospective randomized controlled trials of anesthetic-induced neurotoxicity in children are required. However, we believe that the inherent methodological challenges of studying anesthetic-induced neurotoxicity necessitate the parallel use of well-designed observational cohort studies. (Anesth Analg 2019;128:681–94)
repeated general anesthesia or sedation in the third trimester and in children younger than 3 years old. Guidance from the United Kingdom and Ireland and a statement from European bodies advocate avoiding unnecessary general anesthesia but recommend no changes to clinical practice.

There has been much discussion of the limitations of the existing human evidence base for anesthetic-induced neurotoxicity. Therefore, to inform the design of future clinical studies, we identified and reviewed the 76 clinical studies of long-term neurodevelopmental outcome following early childhood or in utero general anesthesia exposure that were published between 1990 and April 2018 (Supplemental Digital Content, Appendix 1, http://links.lww.com/AA/C723) to identify particular challenges encountered in performing these types of studies, as well as feasible pragmatic methodological solutions. We sought methods used to isolate the effects of general anesthesia from surgery/disease, characterize anesthetic exposure and surgical intervention, address confounding, detect marginal neurotoxic effects, and define what the implications of the research are for clinical practice. These are summarized in Table 1.

DELINEATING THE NEUROTOXIC EFFECT OF ANESTHESIA

Perhaps the greatest challenge to studying anesthetic-induced neurotoxicity is in separating direct toxic effects of general anesthesia on the brain from indirect effects of anesthesia (disturbance of normal physiology, eg, hypoxia, hyperoxia, hypotension, and hypothermia), surgery (stress response and systemic inflammation), and the perioperative course (complications, pain, artificial or inadequate nutrition). We illustrate this concept in Figure 2A.

All but 2 studies make comparisons between general anesthesia and surgery groups, with or without control and therefore cannot distinguish anesthesia-induced effects from surgery-induced effects. Although methodologically ideal, a 2 × 2 factorial design (anesthesia yes/no × surgery yes/no) to determine the effect of anesthesia on neurodevelopment would be logistically and ethically challenging in children or animals and arguably not possible.

A pragmatic nonrandomized study might compare (a) general anesthesia without surgery, eg, undergoing imaging, endoscopic, or interventional procedures; (b) general anesthesia with surgery, and (c) no general anesthesia or surgery. Careful choice of the category (a) children would be required. For example, children undergoing neuroimaging may have comorbidities that are independent risk factors for poor neurodevelopmental outcome. Category (c) controls could be nonhospitalized siblings/classmates or hospitalized nonsurgical children. It is important that children who undergo additional surgeries in later childhood are not excluded from either the intervention or control groups to avoid selection biases.

Spinal anesthesia in immature rats has been shown to not accelerate neuronal apoptosis or cause neurobehavioral abnormality. An ideal randomized study, therefore, could compare (a) general anesthesia for surgery, (b) awake-regional anesthesia for surgery, and (c) no anesthesia or surgery controls. The General Anesthesia compared to Spinal trial adopted a similar strategy in children undergoing general anesthesia/surgery or intended to undergo awake-spinal anesthesia for inguinal herniorrhaphy. In reality, this approach restricts the sample to children undergoing infraumbilical procedures for which awake-neuraxial anesthesia is a feasible alternative to general anesthesia.
Table 1. Challenges and Potential Solutions in Human Studies of Anesthetic-Induced Neurotoxicity

<table>
<thead>
<tr>
<th>Challenges</th>
<th>Solutions</th>
</tr>
</thead>
</table>
| Measuring direct neurotoxic effects of general anesthesia | • Nonrandomized observational design: compare (a) general anesthesia only, (b) general anesthesia + surgery, (c) no general anesthesia or surgery
• Randomized controlled trial: compare (a) general anesthesia + surgery, (b) awake-regional anesthesia + surgery, (c) no general anesthesia or surgery |
| Defining the general anesthesia exposure | • Accurate ascertainment of general anesthesia exposure including duration, drugs, age-adjusted MAC-hours for inhalational agents, or cumulative mg/kg for IV agents
• Use of electronic anesthetic record keeping systems makes this feasible |
| Selection of surgical cohort and procedure | • Study otherwise healthy elective surgical cohorts that have no independent risk factors for poor neurodevelopmental outcome
• Select common, relatively minor surgery, eg, inguinal herniorrhaphy or surgery for solitary urogenital problems |
| Addressing multiple confounding factors | • Careful selection and thorough measurement of potential confounders
• Randomized controlled trial: evenly distribute known and unknown confounders across groups through randomization
• Nonrandomized observational design: (a) control for differences in known confounders using regression or matching techniques; (b) address potential impacts of unmeasured confounding through statistical simulation
• Undertake a Mendelian randomization study |
| Detecting modest neurotoxic effects | • Maximize statistical power by (a) studying large samples in the order of 10^3 to 10^5 children; (b) comparing exposed to unexposed children in 1:4 ratio
• Study longer duration general anesthesia, ie, ≥1 h
• Study general anesthesia during the “vulnerable time window” of brain development, 0-8-11, i.e., <3 years of age |
| Measurement of neurodevelopmental outcome | • Assessment in multiple domains of development using a battery of sensitive, validated outcomes
• Parallel and sequential analyses of noninferiority or equivalence trials where possible
• Randomized controlled trial: compare (a) general anesthesia + surgery, (b) awake-regional anesthesia + surgery, (c) no general anesthesia or surgery
• Prospective evaluation by trained, blinded assessors |
| Length of follow up and sample attrition | • Follow-up until at least school age to allow deficits to manifest in domains of development, which become amenable to thorough neuropsychological testing in school age children
• Ascertain and report reasons for loss to follow up
• Address missing data depending on the mechanism of data loss: missing at random—multiple imputation; missing not at random—sensitivity experiments |

Abbreviations: IV, intravenous; MAC, minimum alveolar concentration.

and may therefore limit external generalizability to other patient groups. Careful control or adjustment for differential incidence of deranged physiology between general anesthesia and awake-regional anesthesia groups (eg, significant hypotension more common in the former) is required to avoid biasing results. Furthermore, children with inadequate blocks or who do not tolerate awake-regional anesthesia may require sedation or conversion to general anesthesia (18% in the General Anesthesia compared to Spinal trial but may be up to 80%,35,63) which may defeat the purpose of the study design. However, per protocol analyses of noninferiority or equivalence trials where there is a crossover of patients between exposure categories would still test whether general anesthesia was harmful to child neurodevelopment.

THE TOXIC EXPOSURE TO GENERAL ANESTHESIA

Although brain structure and function develop throughout childhood, a period of peak synaptogenesis in early childhood has strong implications for later cognition, language, and social behavior. Exposure during this “vulnerable time window” of brain development ought to be the focus of anesthetic research. Although its timing is well defined in animal species, with the overwhelming majority of studies performed on postnatal day 7 in rats, human anesthetic-induced neurotoxicity studies have quoted a heterogeneous range of definitions, eg, “third trimester to 2 years,” “third trimester to 6 weeks,” “0–36 months,” “early gestation through to infancy,” or “birth to 2–3 years.” The concept of a single vulnerable time window may be an oversimplification since there are significant regional differences in the timing and pace of peak synaptogenesis, which are reflected in discordant results for different domains of neurodevelopment. Furthermore, the age of the neuron as opposed to the age of the child can determine vulnerability to anesthetics. At present, it seems pragmatic to investigate general anesthesia exposures up to 3 years of age.

Since most of the studies (n = 49; 64.5%) employ retrospective observational designs and many were not designed to investigate anesthetic-induced neurotoxicity per se (n=27, 35.5%), data concerning anesthetic exposure are often limited. Some investigators make assumptions that, if incorrect, could undermine their studies, eg, babies are presumed to undergo general anesthesia for minor procedures that may have been conducted under regional anesthesia. Whether randomized or nonrandomized prospective or retrospective designs, anesthetic-induced neurotoxicity studies need to strive to accurately ascertain the exposure of each child to avoid underestimating the true effect of general anesthesia (false-negative results).

A dose-response relationship has been detected with increasing numbers of coadministered anesthetic agents and been sought by comparing single versus multiple anesthetic exposures. However, because dose and duration of general anesthesia vary widely between procedures, these are poor surrogates for cumulative dose of anesthetic drug...
exposure. Furthermore, inaccurate reporting of composite procedures, e.g., adeno-nectomy/tonsillectomy/myringotomy, may lead to misclassification of children to the multiple-exposure group. Children requiring repeated procedures may have confounding reasons for poor neurodevelopmental outcome, which may not be captured in the study data set. Ideally, dose–response analyses ought to use a prospectively determined duration of anesthesia in minutes for specified drugs or dose in age-adjusted minimum alveolar concentration-hours for inhalational agents or cumulative milligram per kilogram for IV anesthesia. This level of detail may be more achievable with electronic anesthetic record-keeping systems.

**CHOICE OF INTERVENTION**

In observational studies, selection of participants in terms of their diagnosis/disease and surgical procedure ought to minimize “confounding by indication”—a scenario in which the disease or the surgery itself is an independent risk factor for poor neurodevelopmental outcome. Studies of neurosurgical and cardiothoracic surgical cohorts, as well as children operated on with major congenital or chromosomal abnormalities are classically affected. However, studies of general anesthesia for neuroimaging, some otorhinolaryngology procedures (e.g., adenotonsillectomy for obstructive sleep apnea associated with learning difficulty or myringotomy and grommet insertion associated with speech/language delay), pyloromyotomy associated with significant hyperbilirubinemia or nutritional inadequacy, gastroschisis, craniosynostosis, and cancer surgery may be similarly compromised.

When selecting study participants, a balance ought to be struck between the risk of confounding by indication and being as inclusive as possible to maximize external validity. A healthy, elective surgical cohort undergoing relatively minor surgery would be ideal. Inguinal herniorrphathy or surgery for solitary urogenital problems (e.g., circumcision or hypospadias repair) is common and has no known independent association with poor neurodevelopmental outcome. Particular care should be exercised if it is necessary to pool multiple surgical procedures to increase statistical power.

Anesthetic agents readily cross the placental barrier, which has previously permitted studies in children born to occupationally exposed mothers and children born by cesarean delivery under general anesthesia. These studies may not demonstrate anesthetic-induced neurotoxicity because of the poorly defined, chronic low-dose occupational exposure or the relatively brief exposure at cesarean delivery. Studying anesthetic-induced neurotoxicity in the context of (a) general anesthesia cesarean delivery versus (b) neuraxial anesthetic cesarean delivery and (c) spontaneous vertex delivery is also fraught with difficulty. Results may be confounded by opioids used for labor analgesia, which may cause neonatal respiratory depression; or the use of labor epidural analgesia, which may reduce stress response in the control group. The indication for cesarean delivery intervention, as well as an increased frequency of prematurity, complications of pregnancy, and perinatal insults in the intervention groups may also confound results. Studying intrauterine surgery to correct fetal abnormalities would offer a longer well-defined general anesthetic drug exposure, but no such work has been published.
Table 2. Potential Confounders of the Association Between Anesthesia and Neurodevelopment, Which Have Been Measured in Human Anesthetic-Induced Neurotoxicity Studies

<table>
<thead>
<tr>
<th>Child Demographics</th>
<th>Language 14, 59</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age/year of birth: 23, 33, 34–42, 71, 78, 81–85</td>
<td>Month/quarter of birth (accounts for school entry cohorts) 39, 47</td>
</tr>
<tr>
<td>Sex: 14, 21, 23, 29, 35, 32, 34–39–42, 46, 47, 53, 60, 64, 68, 73, 74, 78, 80, 83–86</td>
<td>Year of birth cohort (accounts for changes in assessment tool) 10, 39</td>
</tr>
<tr>
<td>Race/ethnicity: 14, 21, 30, 31, 34–42, 46, 64, 88</td>
<td>Received income support 9</td>
</tr>
</tbody>
</table>

Socioeconomic Characteristics

| Socioeconomic status: 23, 30, 40, 64, 74, 82, 85 | Involved in child welfare system 9 |
| Housing class: 10, 26 | Insurance system: eligibility status, provider 40, 80 |
| Household/maternal income: 9, 34, 46, 47, 66, 88 | Geographical location, eg, zone improvement plan code/postal code 38, 40, 48, 87 |
| Years/level of education: 10, 26, 29, 32, 41, 47, 53, 60, 64, 68, 73, 74, 78, 83, 86, 88 | Urbanity/rurality of residence: 9, 39, 80 |

Other Parent/Guardian Factors

| Parental intelligence quotient: 24, 73, 90 | Number of siblings: 17 |
| Age at birth of (first) child: 5, 27, 32, 36, 39, 53, 80, 87 | Birth order: 26, 87 |
| Maternal parity: 9, 10 | Sibships: 92 |
| Medical history: 26 | Race/nationality: 80 |
| Parental death: 89 | Language spoken at home: 22, 64, 67, 68, 83, 91 |
| Maternal alcohol consumption: never, prepregnancy, or in pregnancy: 73 | Maternal smoking: never, prepregnancy, or in pregnancy: 73 |
| Developmental delay, mental disability, psychiatric disorder, autism, or attention-deficit/hyperactivity disorder: 26 | Maternal alcohol consumption: never, prepregnancy, or in pregnancy: 73, 92 |
| Problems at school: 29 | Developmental delay, mental disability, psychiatric disorder, autism, or attention-deficit/hyperactivity disorder: 26 |
| Childhood trauma: 23 | Urgency of assisted or operative delivery |

Pregnancy and Peripartum

| Intrauterine growth retardation: 38, 52 | Labor analgesia: epidural, spinal, opioid, nitrous oxide |
| Prematurity: gestational age at birth: 14, 21, 24, 26, 27, 30, 34, 35, 41, 46, 47, 52, 53, 57, 60, 64, 66, 73, 75, 80, 83, 86, 90, 94, 95 | Mode of delivery: normal vaginal, assisted vaginal, Cesarean: 53, 84 |
| Birth weight (centile)/small or large for gestational age: 5–20, 21, 22, 29, 27, 25, 32, 34–36, 41, 46, 47, 53, 52, 60, 83, 86 | Fetal and neonatal hemorrhage, hemolytic disease of the newborn, or other hematological conditions: 38 |
| Induction of labor: 53 | Perinatal infection: 38, 46 |
| Prolonged labor | Disorders of digestive system: 38 |
| Maternal complications of pregnancy, labor, or delivery: 38, 42, 53, 80 | Preoperative neurodevelopmental scores: 32 |

Perinatal Fetal Morbidity

| Respiratory distress syndrome or other neonatal respiratory disorder: 26, 48, 80 | Hand dominance: 23 |
| Endocrine and metabolic disturbances: 38, 48 | Abnormal neurological examination: 23, 88 |
| Perinatal jaundice: 38, 80 | Mental/psychiatric disorder or disability: 24, 41, 89, 92 |

Past or Perioperative Neurological Status

| Microcephaly: 25, 95, head circumference: 38, 64 | Neurosurgery: 91 |
| Magnetic resonance imaging brain maturity score: 24 | Head trauma and/or loss of consciousness: 23 |
| Magnetic resonance imaging intracranial volume: 34 | Followed by cardiologist or cardiovascular service: 19 |
| Number of sedated magnetic resonance imaging scans: 95 | Age at general anesthesia: 21, 25, 30, 35, 57, 64–66, 88, 88 |
| Meningitis: 10 | Weight at general anesthesia: 25, 30, 64, 80, 85 |
| Seizures: 10, 97 | Bispectral index at the end of surgery: 21 |
| Traumatic or ischemic CNS injury: 14, 19, 90, 92, 95, 97 | Hemodynamic and respiratory instability during anesthetic: 20 |

Secondary Care Interaction

| Number/duration of admissions: 22, 24, 35, 46, 61, 63, 64, 65, 66, 68, 93 | Complications of surgery: 27 |
| Number of outpatient clinic attendances: 46 | Followed by cardiologist or cardiovascular service: 19 |

Anesthetic and Surgical Factors

| Indication for general anesthesia: type of surgery, imaging, or examination: 10, 48, 57 | Age at general anesthesia: 21, 25, 30, 35, 57, 64–66, 88, 88 |
| Diagnosis, eg, anatomical defect in cardiac surgery, site of craniosynostosis suture, and type of cleft lip and palate: Surgical center: 0, 40 | Weight at general anesthesia: 25, 30, 64, 80, 85 |
| Surgical approach: open, minimally invasive: 88 | Bispectral index at the end of surgery: 21 |

(Continued)
Table 2. Continued

Other Child Medical History

Genotype: genetic syndrome/genetic polymorphisms/chromosomal abnormality/predisposition to neurodevelopmental disorders
Phenotype: (major or multiple) congenital anomalies/birth defects/dysmorphic syndrome
Other comorbidities likely to influence neurodevelopment: cardiac disease, Hirschsprung disease, retinoblastoma, Sturge-Weber syndrome, severe renal disorders, endocrine disease, jaundice, chronic respiratory disease, bronchopulmonary dysplasia, strabismus surgery, cardiac surgery, or other significant health conditions requiring surgery

Past and Perioperative Critical Care Admissions

Number/duration of admissions
Composite scores, eg, Parmelee Postnatal Complication Scale
Invasive or noninvasive ventilation: use of, duration
Oxygen requirement, eg, preoperative Fo2
Lowest PaO2
Prenatal/postnatal steroids or respiratory distress
Hypotension
Blood loss and coagulopathy
Volume of blood components given

Duration/Cumulative Doses of Other Drugs
Opioids
Benzodiazepines
Ketamine

Cardiac Surgery
Prenatal diagnosis
Palliative or corrective surgery
Congenital lesion: cyanotic or acyanotic, 1- or 2-ventricle circulation, anatomical defect, eg, tetralogy of Fallot, transposition of the great arteries, ventricular septal defect
Composite scores, eg, Society of Thoracic Surgeons mortality category, Aristotle score, Risk Adjustment for Congenital Heart Surgery
Somatic and cerebral regional
Inotropes: duration, mean score
Duration of open sternum
Postoperative catheterization or reoperation
Erythropoetin or aprotinin administration
Anticoagulant or antiplatelet drug at discharge

Other Biochemical Measurements
Hypothalamic-pituitary-adrenal axis, eg, baseline adrenocorticotropic hormone
Adrenaline level

Composite morbidity scores, eg, Johns Hopkins Resource Utilization Band, John Hopkins Adjusted Diagnosis Group, and ASA physical status
History of fetal intervention (radiotherapy, brachytherapy, pharmacotherapy, chemotherapy)
Physical disabilities

Cardiopulmonary arrest
Nutrition: oral, artificial supplement, entirely artificial
Total parenteral nutrition: use of, prolonged use
Necrotizing enterocolitis
Gut perforation
Infections: number, complicated by sepsis
Lowest temperature
Other perioperative complications: thromboembolic, hemorrhagic, hypoxia, acidosis, hypocalcaemia, seizure

Chloral hydrate
Antiepileptic drugs
Substance abuse
Diameter ascending aorta
Indexed shunt diameter
Patent ductus arteriosus (distinguish surgically closed)
Prostaglandin dependent
Extracorporeal membranous oxygenation: use, duration
CPB: use, duration
Deep hypothermic circulatory arrest: use, duration
Aortic cross-clamp: use, duration
Selective cerebral perfusion time
Afterload reduction time
Hematocrit: intraoperatively/postoperatively lowest on CPB, at the end of bypass, after hemodilution
Worst lactate or base excess intraoperatively and postoperatively

Inflammation, eg, interleukin-6 at the end of surgery

Abbreviations: ASA, American Society of Anesthesiologists; CNS, central nervous system; CPB, cardiopulmonary bypass; Fo2, fractional concentration of inspired oxygen.

ADDRESSING CONFOUNDING

The association between general anesthesia and neurodevelopmental outcome is heavily confounded by factors throughout the life course (Figure 2B; Table 2). Properly conducted randomized controlled trials should ideally distribute known/measured and unknown/unmeasured confounders across groups at randomization, thereby overcoming confounder bias.

Observational studies of anesthetic-induced neurotoxicity must control (via restriction, stratification, or regression adjustment) for differences in known/measured confounders between groups to avoid extensive bias. However, data concerning pregnancy/peripartum factors (eg, prematurity, fetal acidosis, birth asphyxia) and perioperative factors (eg, temperature, hypoxia/hyperoxia, hemodynamics, adverse events) are often unknown, especially in retrospective studies. Some factors that ought to be adjusted for, eg, American Society of Anesthesiologists physical status, are not routinely recorded for nonexposed children, and smaller studies may make no attempt to adjust for confounders at all.

By definition, unknown/unmeasured confounders cannot be controlled for, but their potential impacts on the results of observational studies can be simulated statistically.

Adjustment for multiple potential confounders in observational studies is performed with the intention of reducing confounder bias. However, care must be exercised to avoid
“overadjustment” — whereby this very process decreases precision or paradoxically increases net bias though several mechanisms. First, attempting to control for increasing numbers of variables reduces the precision of the neurotoxic effect estimates generated by statistical models. Wide (imprecise) CIs around the effect estimates may mask any evidence of anesthetic-induced neurotoxicity, leading to false-negative conclusions. The second mechanism concerns “intermediate variables,” which are distinguished from confounders by lying on the causal pathway between exposure and outcome. For example, we might speculate that anesthetic-induced neurotoxicity is mediated via hypotension (Figure 2A). In the case of multiple causal pathways between exposure and outcome, then mistakenly controlling for hypotension (or some descending proxy thereof such as volume of crystalloids or amount of vasoactive drug administered) would produce a null-biased result, ie, falsely reducing the apparent strength of any neurotoxic effect estimate. Worse still, if the only causal path between general anesthesia exposure and impaired neurodevelopment were mediated through hypotension, then mistakenly controlling for this intermediate variable (or its proxies) ought to entirely nullify any neurotoxic effect estimate, again producing falsely reassuring conclusions. The third mechanism involves “collider variables,” which are defined as a common effect of the exposure and outcome (Figure 2C). Mistaken control for this common effect induces a spurious (noncausal) association between general anesthesia and neurodevelopmental outcome through which confounding can flow, paradoxically inducing bias (termed “collider-stratification bias”) into the neurotoxic effect estimate where none previously existed. An illustrative example comes from studies of prenatal pollutant exposure and long-term child neurodevelopment in which the pollutants also cause fetal loss.106 Since outcome can only be determined in live-born children, if investigators condition on live birth status (in this case by restriction to live-born children as is typical in pediatric cohort studies), bias arising from common causes of fetal death and long-term neurodevelopmental outcome (ie, confounders of the association between fetal death and neurodevelopment) is induced.

Collectively, the pitfalls of multivariable analysis necessitate thoughtful selection of potential confounders, which may be assisted by drawing a “directed acyclic graph” — a visual representation of the assumed associations among exposure, outcome, and other measured/unmeasured variables using unidirectional arrows to represent the direction of causality (and temporality). These graphs distill the causal model underlying the epidemiological problem, informing the choice of confounding, intermediate and collider variables, which would be required to build a statistical model to test for an unbiased relationship between general anesthesia and neurodevelopmental outcome. The aforementioned pitfalls of multiple confounder adjustment also necessitate cautious “stepwise” modeling whereby potential confounders are sequentially added to the developing statistical model and its output scrutinized at each step for paradoxical effects. A sudden reversal of the effect estimate following the stepwise incorporation of the latest potential confounder, for example, may prompt a reevaluation of the causal assumptions regarding that variable and whether it may operate as a collider as opposed to a confounder in the causal model. It would be dangerous to simply attempt to simultaneously adjust for all measured child characteristics in a nonrandomized anesthetic-induced neurotoxicity study.

Conventional techniques for confounder adjustment include various regression models (eg, linear, logistic, Poisson or Cox proportional hazards modeling) and matching techniques. Group/frequency matching ensures that the proportions of subjects with given characteristics are the same in each group.41,42 Individual/pair matching ensures that pairs of children, 1 from each group, share similar characteristics.39,40 Results from matched pairs are less confounded but require larger sample sizes to achieve the same precision.

More innovative approaches may help uncover associations. Propensity score analysis is a pragmatic choice of method to reduce the complexity and computational burden of statistical models, which attempt to control for a multitude of potential confounding variables in a nonrandomized study. It reduces the dimensionality of the data set from a large collection of variables to a single propensity score, which is generated by a regression model from those variables that are thought to influence membership to the general anesthesia group in the study. The propensity score assigned to each child would take a value between 0 and 1 and represent the estimated probability of general anesthesia group membership, conditional on the values of those variables thought to influence general anesthesia versus nongeneral anesthesia group membership. The propensity score can then be adjusted for as an independent variable in a regression model (as opposed to entering the collection of known/measured confounders). Alternatively, one can match individual children between general anesthesia and nongeneral anesthesia groups who have similar likelihoods of general anesthesia group membership (ie, similar propensity scores), such that known/measured confounders are balanced across the 2 groups.10,23,27,38,46,74 These “propensity-adjusted” or “propensity-matched” estimates of neurotoxic effect on neurodevelopment ought to be unbiased by known/measured confounders.

Mendelian randomization is an advance in observational epidemiology, which overcomes confounding by both known/measured and unknown/unmeasured factors. It can provide unbiased evidence for causal relations between a modifiable exposure and patient outcome.44,45 Instead of the traditional exposure variable (ie, general anesthesia/surgery), it considers “instrumental variables” (Figure 2D). These are either one or a combination of multiple genetic variants (ie, alleles or single nucleotide polymorphisms) that are randomly allocated to children at meiosis in human reproduction and are selected on the basis that they robustly predict general anesthesia exposure without directly influencing neurodevelopmental outcome (except via the general anesthesia exposure itself). Candidate genetic variants are typically identified from large genome-wide association studies but could conceivably be associated with certain disease states (increasing the propensity for general anesthesia to facilitate procedures, medical imaging, or surgery) or with suxamethonium apnea or malignant hyperpyrexia (reducing the propensity for general anesthesia where there is an established child or family history). Random natural assortment of genetic material ensures that instrumental variable status is independent of factors that confound the association between the traditional exposure variable (general anesthesia/surgery) and the neurodevelopmental outcome.
Once child outcomes are compared based on the instrumental variable (rather than general anesthesia exposure), then intergroup differences in general anesthesia exposure and neurodevelopment ought to reflect true, unconfounded causal relationships between general anesthesia/surgery and neurodevelopmental outcome (Figure 3). We believe that the Mendelian randomization approach to detecting anesthetic-induced neurotoxicity may be especially feasible using a “2-sample” Mendelian randomization in which data linking the chosen genetic variants to general anesthesia exposure need not come from the same sample as data that link general anesthesia exposure to neurodevelopment. No observational studies of anesthetic-induced neurotoxicity published to date have used Mendelian randomization. However, it offers the potential to elucidate an unconfounded link between anesthesia and neurodevelopment using what is an efficient natural analogy to a randomized controlled trial.

As an illustrative example, the effect of prenatal alcohol exposure on child academic achievement has been studied recently using the Mendelian randomization approach. Here, researchers have exploited genetic variation in the alcohol dehydrogenase gene as an instrument for in utero alcohol exposure. Mothers with the rare allele metabolize alcohol faster, resulting in more rapid production of ethanol metabolites that cause unpleasant symptoms. These mothers are shown to consume less alcohol. Investigators demonstrate that the instrumental variable, unlike alcohol consumption, is unrelated to potential confounders of the association between prenatal alcohol exposure and academic achievement such as socioeconomic status. While traditional regression analyses based on the alcohol consumption exposure variable have returned ambiguous results, presumably due to residual confounding (eg, maternal wine consumption being protective for child educational attainment), the instrumental variable analyses demonstrate robust positive effects on child educational achievement in children whose mothers were induced by their genotype to abstinence or lower alcohol consumption in pregnancy.

Twin or sibling studies attempt to eliminate confounding by genetic and environmental factors, eg, uterine environment, parental education, parenting style, home/family environment, neighborhood, educational, and socioeconomic factors. In a monozygotic concordant-discordant design, participants in each group share the same genetics and family-level environmental factors. Differences in neurodevelopmental outcome across groups would then reflect the toxic effect of general anesthesia/surgery.

Longitudinal study designs, where neurodevelopment is repeatedly assessed over time, allow children to serve as their own controls. This approach mitigates confounding by static confounders, eg, genetics and socioeconomic status.

Finally, other approaches may dispense with control groups altogether. One could focus on the interaction between general anesthesia and age at exposure, ie, compare children who undergo early versus late surgery. Associations would not be confounded by diagnosis and surgery/anesthetic factors since all subjects could be similarly exposed. However, this approach mandates that surgery can be postponed, which is not always feasible.

DETECTING MODEST NEUROTOXIC EFFECTS

In utero or early childhood exposure to a range of neurotoxicants (eg, metals, organic solvents, pesticides)

Figure 3. Contrasting the conduct of (A) randomized controlled trials and (B) Mendelian randomization studies. See text for full explanation.
can adversely affect neurobehavioral development.\textsuperscript{110,111} Ethanol, like anesthetic agents, acts at $\gamma$-aminobutyric acid and N-methyl-$\delta$-aspartic acid receptors and causes neuronal apoptosis in the developing brain.\textsuperscript{92} Robust detrimental associations between heavy and binge prenatal alcohol exposure and adverse child neurodevelopment are established.\textsuperscript{112,113} However, studies of light-to-moderate prenatal alcohol exposure have suffered from residual confounding and have reported inconsistent conclusions even with sample sizes in the order of 10,000 children. We can presume that large samples will similarly be required to reliably detect any long-term neurotoxic effects following childhood general anesthesia—an effect that may also be comparable or small relative to the effects of confounding factors.\textsuperscript{9,25,39,47,73,82} Large samples are also required to permit adjustment or matching techniques to account for confounding. Existing anesthetic-induced neurotoxicity studies vary in size between 15 and 125,000 subjects with a median 131 children (interquartile range, 50–372), so are often likely to be underpowered and potentially falsely reassuring.

Besides pursuing larger sample sizes, comparing exposed children with unexposed children in a 1:4 ratio to maximize statistical power,\textsuperscript{38,47,48} avoiding short-duration interventions (eg, maternal general anesthesia for cesarean delivery or myringotomy and grommet insertion), studying exposure during the “vulnerable time window” of brain development, and using sensitive outcome measures are strategies that may increase the likelihood of detecting the neurotoxic effects of general anesthesia.

**NEURODEVELOPMENTAL OUTCOME**

The neurodevelopmental outcome measures reported in the literature vary and encompass (a) intelligence/cognition, (b) academic achievement, (c) development/behavior, and (d) neuropsychiatric diagnoses, ie, attention-deficit/hyperactivity disorder, autism spectrum disorder, and learning disability.\textsuperscript{49} Prospective evaluation in multiple domains of development using a battery of sensitive, validated outcomes and trained, blinded assessors is the gold standard. However, the risk of detecting spurious associations increases with multiple outcomes. Therefore, it is wise to caution against the overinterpretation of solitary detrimental associations in the context of a panel of otherwise reassuring results.

Measures of intelligence/cognition are thought to remain stable throughout the life course unless disrupted by severe disease.\textsuperscript{49,114} However, assessment is not feasible until basic cognitive skills are achieved by 4–6 years of age.\textsuperscript{90,50} Age-normalized intelligence scores permit comparisons of outcome at different ages and enable referencing to population scores.\textsuperscript{104}

Academic achievement in standardized national tests reflects intelligence/cognition,\textsuperscript{115} but is muddied by multiple external factors, eg, self-esteem and lifestyle factors.\textsuperscript{49} School grade performance in children with dyslexia or dyspraxia may be boosted by extra help in school, mitigating any negative effect on academic achievement.\textsuperscript{28} Although standardized national tests are administered at population level, which makes them a feasible outcome for large population studies, not all children participate, eg, private schools or nonentry due to learning difficulty.\textsuperscript{8} However, investigating academic achievement does confer the pragmatic advantage that parents/guardians are likely to be highly invested in their child’s school performance.\textsuperscript{27}

Child development evolves in surges and plateaus, referenced to well-defined developmental milestones expected at certain ages, which permits outcome assessment even at the youngest ages.\textsuperscript{89} The reliability of subjective developmental/behavioral data collected through parental survey is questionable: developmental delay in language/speech, mathematics, and reading domains may not be noticed until challenged in school; behavioral problems may not manifest until children communicate and interact with their peers in school.\textsuperscript{34,38,45,68} An ideal anesthetic-induced neurotoxicity study should use trained, blinded assessors (eg, pediatric neuropsychologists) to measure outcome using a comprehensive battery of developmental assessments. Scores generated by this method of outcome assessment are objective and highly sensitive to subtle neurotoxic effects that may be difficult to detect clinically.\textsuperscript{24} The use of such comprehensive neurodevelopmental assessments is most feasible in smaller studies, which prospectively assess outcome,\textsuperscript{27} but it is also available in some retrospective data sets.\textsuperscript{73} The Bayley Scales of Infant Development\textsuperscript{28,34} is the most extensively used example,\textsuperscript{116} but the latest third version may overestimate development in certain groups,\textsuperscript{117,118} and caution is required if comparisons are made with scores from previous iterations.\textsuperscript{119}

Neuropsychiatric diagnoses for developmental/behavioral disorders are multifactorial in origin (including genetic predisposition), with a heterogeneous and changing clinical presentation over time.\textsuperscript{49} Children may spontaneously “catch-up”\textsuperscript{46} or benefit from supportive interventions in childhood.\textsuperscript{90,51} Neuropsychiatric diagnoses are almost exclusively parameterized as binary outcomes (eg, from *International Classification of Diseases, Ninth Revision* diagnosis codes, school or health care records) as opposed to “risk scores.” These binary outcomes are likely to be too crude/insensitive to detect any subtle effects of anesthetic exposure.\textsuperscript{23} Nondiagnosis (especially before the group communication/interaction and higher cognitive demands placed on schoolchildren\textsuperscript{14,48}), underreporting, and incorrect diagnosis coding in databases is likely to introduce misclassification bias. Studying learning disability confers particular advantages though: a high incidence (5%–10%) and recording in large educational databases.\textsuperscript{49}

**POSTOPERATIVE FOLLOW-UP AND SAMPLE ATTRITION**

The time interval between anesthesia and first neurodevelopmental assessment must be sufficiently long to distinguish long-term neurotoxic effects from short-term postoperative cognitive-behavioral changes (ie, 26 months\textsuperscript{96}). It must also allow sufficient latency for marginal neurodevelopmental deficits to manifest in domains of development, which emerge, differentiate, and are amenable to thorough neuropsychological testing at older ages, eg, cognitive skills such as language/speech/reading, mathematics, memory, and executive functioning from late childhood.\textsuperscript{14,24,30} Furthermore, neurodevelopmental evaluation in school children is known
to be more robust and predictive for adulthood than when measured in preschool children because of the variability in young children’s developmental trajectories. There has been concern that multiple life course factors may dilute any differences in outcome between exposed and unexposed children after such long follow-up. However, subtle associations between starting school in January versus December and educational achievement and intelligence quotient scores have been detected in large cohorts as late as 18 years old. Existing studies of anesthetic-induced neurotoxicity follow up children until a median age of 6 years (interquartile range, 2–12 years).

Prolonged follow-up makes retrospective or ambidirectional (meaning retrospective ascertainement of exposure but prospective measurement of outcome) studies inefficient compared to prospective randomized and nonrandomized designs. But it also makes sample attrition (e.g., due to withdrawal, death, migration, moving schools, or health care provider) a significant problem, eg, 50% of initially enrolled children completing assessment at 2 years in 1 study. Most observational studies report a “complete case analysis,” in which any children with missing data are disregarded. The amount of missing data and reasons for this are frequently omitted. As well as suffering a reduction in precision, their results may be biased when neurodevelopmental outcome data are missing nonrandomly. For example, if general anesthesia slowed child neurodevelopment, then exposed children may be lost to follow-up if they were unable or reluctant to engage in intelligence testing. Effect estimates would then underestimate the true effect of general anesthesia in the complete case analysis.

Even research funded to intensively follow up children in prospective randomized or nonrandomized studies will have missing data. Statistical methods can be used to permit unbiased analyses without excluding affected cases. Choice of method depends on the probable mechanism of data loss. Multiple imputation is a popular technique used when data are believed to be missing at random. Missing data are inferred from a rich observed data set to construct multiple plausible data sets, which are pooled to produce a result that reflects the uncertainty in the imputed data. Data that are missing not at random can only be addressed through experiments that test the sensitivity of results to different mechanisms of data loss.

INTERPRETING RESULTS IN CLINICAL PRACTICE

Despite considerable interest and anxiety, there is at present no conclusive evidence or consensus that general anesthesia harms the developing brain. Childhood general anesthesia typically comprises single short exposures and is likely to carry low risk. However, if general anesthesia is thought to pose long-term neurodevelopmental risks, then the impacts on clinical practice could be far reaching.

In considering the current clinical implications, it should be noted that the evidence base is comprised mainly of retrospective observational studies, whose subjects were anesthetized in the 1970s–1990s, since when there have been widespread changes in practice. Pediatric anesthesia may have become safer as isoflurane/sevoflurane and IV anesthesia have replaced the “Liverpool technique” (muscle relaxation and nitrous oxide for neonatal procedures), halothane, enflurane, and methoxyflurane, and our profession became more conscious of optimal fluid management, adopted obligatory multiparameter monitoring incorporating pulse oximetry and capnography; and there have been changes in who is delivering anesthetic care to children.

Nonetheless, if the evidence base becomes stronger, then surgeons, physicians, and general practitioners will require a new appreciation of the neurotoxic risks of anesthesia to inform clinical decision making and the consent process. Important topics for discussion with children, parents, or guardians would include which elective procedures could be deferred, the associated risks of delay, alternative anesthetic management (eg, alternative anesthetic agents or regional techniques), and possible mitigating or protective strategies.

Withholding general anesthetic drugs during neonatal surgery (e.g., the “Liverpool technique”) may not be an option today and is certainly unethical in later childhood. Painful stimulation and the associated strong stress response are also thought to impair neurodevelopment. Modifiable factors certainly include optimizing perioperative physiology, good perioperative analgesia, psychosocial support, and avoidance of unpleasant experiences or prolonged hospitalization. Determining which general anesthetic drugs and techniques might carry the lowest risk will require researchers to accurately quantify the duration, cumulative dose, and interactions of specific agents. Whether time to allow remodeling/recovery between sequential general anesthesia can mitigate neurotoxic damage could be investigated. Neuroprotection afforded by strict maintenance of physiological parameters, pharmacotherapies, preconditioning, and novel neurogenesis techniques are being researched. Maintaining cerebral glucose and oxygen delivery by minimizing cardiopulmonary bypass and deep hypothermic circulatory arrest times may play a role in pediatric cardiac surgery.

Most general anesthesia is provided for healthy elective cases. Here, the physical or psychosocial harms of deferring or cancelling surgery or procedures would need careful weighing against the risk and impact of potential neurodevelopmental impairment on the individual, especially for repeated or prolonged anesthesia. For example, impaired wound healing and cosmesis, concerns about impaired speech/language development, and social stigma may preclude deferral of surgery in cleft lip and palate. The current level of concern about neurotoxicity would not preclude the provision of general anesthesia for emergency surgery or cesarean delivery.

High-risk groups for poor developmental outcome (e.g., multiple prolonged general anesthesia) may require follow-up neurodevelopmental screening with the option of referral for early school intervention programs to attempt to mitigate any harms and improve developmental acquisition and school performance.

CONCLUSIONS

Despite growing international concern that general anesthesia in childhood leads to long-term neurodevelopmental impairment, delineating general anesthesia–induced effects from those of surgery remains a significant challenge in the
study of anesthetic-induced neurotoxicity. Deficiencies of existing research also include inconsistent exposure definitions, selection of cohorts with independent risk factors for impaired neurodevelopment, extensive confounding, the need to detect subtle neurotoxic effects, blunt neurodevelopmental assessment tools, and sample attrition over the long-term follow-up required.

Randomized controlled trials represent the gold standard tool in the present climate of clinical equipoise. However, randomly assigning children to general anesthesia-surgery versus regional anesthesia-surgery versus no anesthesia-no surgery poses significant ethical and logistical challenges, particularly if prolonged or repeated general anesthesia is to be studied. This coupled with the large sample sizes and prolonged follow-up required to detect neurotoxic effects necessitates the design of more efficient, sophisticated observational studies and has driven calls for the adoption of non-inferiority studies and has driven calls for the adoption of surrogate indices such as neuroimaging and biomarker techniques to evaluate neuronal inflammation and apoptosis.

Large observational studies can produce more precise, more timely results that are not constrained to studying single short general anesthesia exposures. We advocate prospective or ambidirectional cohort studies that accurately ascertain general anesthesia exposure, rigorously control for confounders, and prospectively follow up neurodevelopment into adolescence. They will also permit researchers to elucidate the role of potential mediators and effect modifiers of any neurotoxic effect to inform strategies to mitigate the potential neurotoxic risks of general anesthesia in early childhood.

In parallel, there is a need for ongoing animal work to characterize the mechanisms of anesthetic-induced neurotoxicity, the relative neurotoxic potentials of different anesthetic agents at different stages of development, and modifiable factors to reduce anesthetic-induced neurotoxicity. These animal studies will need to more carefully control physiological parameters and anesthetic dosing and more closely mimic the surgical insult if their findings are to be generalizable to human pediatric anesthesia.

Given the inherent challenges of studying anesthetic-induced neurotoxicity, we must acknowledge that it may never be possible to demonstrate anesthetic-induced neurotoxicity in conventional clinical trials. Ultimately, multiple complementary approaches are required to accumulate sufficient evidence to inform a consensus opinion on the neurotoxic potential of general anesthesia—currently, the single greatest issue in modern pediatric anesthetic practice.

DISCLOSURES

Name: Graham J. Walkden, MBChB.
Contribution: This author helped plan the study, the literature search, review process, and prepare the manuscript.
Name: Anthony E. Pickering, FRCA, PhD.
Contribution: This author helped plan the study and prepare the manuscript.
Name: Hannah Gill, FRCA, PhD.
Contribution: This author helped plan the study and prepare the manuscript.
This manuscript was handled by: Gregory J. Crosby, MD.

REFERENCES


86. Liew Z, Olsen J, Cui X, Ritz B, Arah OA. Bias from conditioning on live birth in pregnancy cohorts: an illustration based on...


