Published in final edited form as:

Anesthesiology. 2018 July ; 129(1): 89–105. doi:10.1097/ALN.0000000000002232.

Neuropsychological and behavioral outcomes after exposure of young children to procedures requiring general anesthesia: The MASK study

David O. Warner, M.D.[1], **Michael J. Zaccariello, Ph.D., L.P.**[2], **Slavica K. Katusic, M.D.**[3], **Darrell R. Schroeder, M.S.**[3], **Andrew C. Hanson, B.S.**[3], **Phillip J. Schulte, Ph.D.**[3], **Shonie L. Buenvenida, R.N.**[1], **Stephen J. Gleich, M.D.**[1], **Robert T. Wilder, M.D.**[1], **Juraj Sprung, M.D.**[1], **Danqing Hu, M.D.**[1], **Robert G. Voigt, M.D.**[5], **Merle G. Paule, Ph.D.**[6], **John J. Chelonis, Ph.D.**[6], and **Randall P. Flick, M.D., M.P.H.**[1]

[1]Department of Anesthesiology and Perioperative Medicine, Mayo Clinic, 200 1st ST SW, Rochester, Minnesota, 55905

[2]Department of Psychology, Mayo Clinic, 200 1st ST SW, Rochester, Minnesota, 55905

[3]Department of Health Sciences Research, Mayo Clinic, 200 1st ST SW, Rochester, Minnesota, 55905

[5]Department of Pediatric Medicine, Baylor College of Medicine, One Baylor Place, Houston, Texas 77030

[6]National Center for Toxicological Research, Food and Drug Administration, 3900 NCTR Road, Jefferson, Arkansas, 72079

Abstract

Background—Few studies of how exposure of children to anesthesia may affect neurodevelopment employ comprehensive neuropsychological assessments. This study tested the hypothesis that exposure to multiple, but not single, procedures requiring anesthesia prior to age 3 years is associated with adverse neurodevelopmental outcomes.

Methods—Unexposed, singly-exposed, and multiply-exposed children born in Olmsted County, MN from 1994-2007 were sampled using a propensity-guided approach and underwent neuropsychological testing at ages 8-12 or 15-20 years. The primary outcome was the Full-Scale intelligence quotient (IQ) standard score of the Wechsler Abbreviated Scale of Intelligence. Secondary outcomes included individual domains from a comprehensive neuropsychological assessment and parent reports.

Corresponding Author: David O. Warner, M.D., Mayo Clinic, Department of Anesthesiology and Perioperative Medicine, 200 1st St SW, Rochester, MN 55905, Ph: 507-255-4288, warner.david@mayo.edu.

Prior Presentations: None

Conflicts of interest: The authors declare no competing interests.

Results—997 children completed testing (411, 380, and 206 un-, singly-, and multiply-exposed, respectively). The primary outcome of IQ did not differ significantly according to exposure status; multiply-exposed and singly-exposed children scoring 1.3 points (95% CI -3.8 to 1.2, p=0.32) and 0.5 points (CI -2.8 to 1.9, p=0.70) lower than unexposed children, respectively. For secondary outcomes, processing speed and fine motor abilities were decreased in multiply- but not singly-exposed children; other domains did not differ. The parents of multiply-exposed children reported increased problems related to executive function, behavior, and reading.

Conclusions—Anesthesia exposure prior to age 3 years was not associated with deficits in the primary outcome of general intelligence. Although secondary outcomes must be interpreted cautiously, they suggest the hypothesis that multiple, but not single, exposures are associated with a pattern of changes in specific neuropsychological domains that is associated with behavioral and learning difficulties.

Introduction

Drugs producing general anesthesia can cause neurodegeneration and long-term deficits in learning and behavior in young animals (including non-human primates).[1-3] Numerous studies have sought evidence for similar effects in children. Most observational studies find that multiple exposures to procedures requiring general anesthesia are associated with deficits in learning and behavior, albeit with small effect sizes in some studies.[4-10] Some, but not all, human studies also find an association between single exposures and a variety of outcomes.[4,8,9,11-18] These studies employed a wide range of designs and outcomes. Only two studies reported a comprehensive assessment of neuropsychological function: an unmatched cohort study,[14] and another that carefully matched subjects who were and were not exposed to anesthesia, but included only children undergoing herniorraphy, who were predominantly male.[18] Thus, any specific pattern of neuropsychological changes associated with the exposure of a general population of children to procedures requiring anesthesia, if present, is still poorly defined.

The aim of the Mayo Anesthesia Safety in Kids (MASK) study was to test the hypothesis that exposure to multiple, but not single, procedures requiring general anesthesia prior to a child's third birthday is associated with adverse neurodevelopmental outcomes. Using a matched cohort design, this hypothesis was evaluated by prospective neuropsychological testing of a propensity-guided sample of children born in Olmsted County, MN from 1994 to 2007. The primary outcome for analysis was the Full Scale intelligence quotient (IQ) score of the Wechsler Abbreviated Scale of Intelligence (WASI). This score was chosen as the primary outcome based on comparability with other studies[10,12,13,18] and the availability at the time of study design of school achievement test data that permitted power calculations, with the assumption that IQ is related to achievement test performance.[5] Secondary outcomes included the results of a comprehensive battery of neuropsychological assessments and parent reports of behavior and learning difficulties.[19]

Materials and Methods

This study was approved by the Mayo Clinic and Olmsted Medical Center Institutional Review Boards, and written informed consent/assent was obtained. Study methods have

been previously published[19] and are here summarized. In addition to the neuropsychological testing reported here, children were tested on the National Center for Toxicological Research Operant Test Battery; results of this testing will be presented in a future analysis.

Subject recruitment

Children born from January 1, 1994 to December 31, 2007 in Olmsted County, MN, who resided within Olmsted County until their third birthday and who resided within 25 miles of Rochester, MN according to available records at study onset were identified using the resources of the Rochester Epidemiology Project, a medical records linkage system that provides access to the complete medical records of all Olmsted County residents, and birth certificate information obtained from the Division of Vital Statistics, Minnesota Department of Health. This date range was chosen as approximately coinciding with the more widespread use of sevoflurane into clinical practice, and as providing a sufficient number of children who could be tested at age 8 or greater during the study period. Birth certificate information was used to establish that children were born in Olmsted County, an approach that minimized the potential for referral bias and facilitated recruitment for testing.

Subjects were eligible for testing if enrolled between the ages 8 and 12 years or 15 and 19 years to allow evaluation of any evolution of anesthesia-associated changes. These age ranges were chosen to represent two developmental stages (pre-adolescence and adolescence), and based on preliminary estimates of the number of children who would be available for testing. Those who enrolled at age 19 years were tested even if they turned 20 years before testing could be scheduled.

Through medical records review each eligible child was classified as unexposed, singly-exposed or multiply-exposed to anesthesia prior to their third birthday. Target recruitment goals were initially determined based on considerations of statistical power and feasibility, goals that were adjusted at the approximate midpoint of subject recruitment to account for actual recruitment patterns while maintaining statistical power (Supplemental Digital Content 1 and 2, showing enrollment goals and statistical power).

It was initially estimated that it would be necessary to contact all eligible multiply-exposed children to meet recruitment goals. To minimize the potential for confounding, we sought to recruit singly- and non-exposed children who were best matched with multiply-exposed children on a variety of characteristics potentially affecting the outcomes of interest. Singly- and non-exposed children were selected for recruitment using a frequency-matched approach, with strata defined based on their propensity for receiving single and multiple exposures general anesthesia. Propensity scores were calculated using multinomial logistic regression including data available from the birth certificate (sex, gestational age at birth, birth weight, APGAR scores at 1 and 5 minutes, mother's and father's age and level of education), and health status from data available in the medical record as estimated using the Johns Hopkins ACG Case Mix System, which calculates 32 binary indicator variables representing co-morbidity clusters (aggregated diagnostic groups, ADGs). Based on quintiles of the observed distribution of propensity scores for single and multiple exposures, 50 sex-specific propensity-matched strata (25 each for males and females) were defined and used to select those singly- and un-exposed children to be randomly sampled within each

stratum that included at least one multiply-exposed child. Children were excluded if conditions that would preclude testing were present, including severe intellectual disability, limited English language proficiency, autism, and spastic cerebral palsy.

Assessments

Each subject was assessed by a trained psychometrist who tested domains typically measured in clinical practice (Supplemental Digital Content 3, which lists all tests).[19] The study psychometrists were chosen from a pool of eighteen who staff our clinical Psychometric Assessment Lab. Initially, psychometrists undergo at least 4 months of full-time training and are not deemed independent for roughly 12-months after they begin training. Periodically, thereafter they are observed to assure test administration fidelity. After each testing session, as a quality control measure, all neuropsychometric assessment data was reviewed by another psychometrist for accuracy. Parent/guardian questionnaires assessed perceived behavior and learning difficulties. A summary of how tests would be interpreted in terms of underlying domains measured was formulated prior to analysis (Table 1), including study-specific composite scores to increase the ability to detect effects when more than one instrument assessed a particular domain. This *a priori* approach provided an overall roadmap for how any observed differences would be interpreted.

Analysis

Details of the analysis plan were made available via a web-based repository prior to commencing analysis (https://osf.io/k93nb/).

Weighting procedure—The sampling strategy planned to invite all available multiply-exposed children to participate, and sample propensity-matched single- and un-exposed children in fixed ratios to the multiply-exposed children for each stratum. With ideal sampling, the characteristics of these singly- and un-exposed children would be distributed similarly to the multiply-exposed children. However, because not all who were invited participated, some sampling strata were missing subjects with a given exposure status. Thus, the primary analysis used inverse probability of treatment weighting (IPTW) to account for imbalances across exposure categories among children actually tested.[20] The approach weighted the observed sample of singly- and un-exposed children to mimic the originally-planned fixed-ratio sampling and thus balance potential confounders across exposure groups. The need for this procedure was identified early during the accrual period and reflected in the analysis plan developed before completing subject accrual and posted prior to analysis.

Propensity scores were estimated from a multinomial model. Using Z to denote exposure and X to denote the vector of explanatory covariates in the propensity model, three probabilities were estimated for each individual: $P(Z = 2|X)$, $P(Z = 1|X)$, and $P(Z = 0|X)$, using $Z = 2$ as shorthand for 2 or more exposures. Then, for a given individual with observed exposure group z (0=none, 1=single, 2=multiple) and explanatory covariates X, the weight for the individual is given by

$$w_{ATT} = 1 * I(Z = 2) + \frac{P(Z = 2|X) * I(Z = 1)}{P(Z = 1|X)} + \frac{P(Z = 2|X) * I(Z = 0)}{P(Z = 0|X)}$$

where $P(Z = z|X)$ indicates the probability of being in group z given covariates X, and $I(Z = z)$ is an indicator function taking values 1 if the individual was in group z and 0 if not. The weighted sample is expected to be balanced with respect to the distribution of baseline covariates across the three exposure groups, with a distribution similar to the population of multiply exposed individuals. Standardized differences in explanatory covariates between singly- vs un-exposed and multiply- vs un-exposed were compared before and after weighting to evaluate balance.

Prior literature suggests explanatory variables in propensity score models should include factors associated with the outcome, and which occur temporally prior to the exposure.[21] This would include potentially confounding variables. For the IPTW propensity score, most explanatory variables were retrieved from birth certificate data. Only ADGs occurring prior to age 3 were included in the model to best reflect subject characteristics over the period they were at risk for anesthesia exposure. Since some ADGs are sparsely represented or omnipresent, all of them could not be included in the propensity model without overfitting the model. Further, some may be highly correlated which may lead to collinearity in the propensity score model. We thus identified the subset of ADGs associated with the primary outcome of IQ using multivariable linear regression. Explanatory variables including sex, gestational age at birth, birth weight, APGAR scores at 1 and 5 minutes, mother's and father's age and level of education, socioeconomic status as measured by the HOUSES index,[22] and all ADGs were considered in a linear regression of the WASI IQ score outcome. Backwards selection was used to assess what ADGs were associated with this outcome, while keeping other explanatory variables in the model. A $p<0.1$ was used for stay criteria to conservatively include ADGs associated with outcome. Using this procedure, the final model for the IPTW propensity score thus included sex, gestational age at birth, birth weight, APGAR scores at 1 and 5 minutes, mother's and father's age and level of education, socioeconomic status, Dermatologic ADG, Psychosocial ADG, Minor infection ADG, Asthma ADG, and Major Infection ADG. Sex by characteristic interactions were also included. Data for birth characteristics and ADGs were complete for all individuals, whereas parental characteristics and socioeconomic status were complete for at least 97% of the study sample (socioeconomic status incomplete for 3%, father's age missing for 3%; ≤2% missing for other data). Multiple imputation (n=50 imputations) was performed to obtain complete datasets of characteristics necessary for the calculation of the IPTW propensity scores.

Primary analysis—Each endpoint was analyzed as a continuous variable, with transformations used as necessary to satisfy the distributional assumptions (normally distributed errors) implicit in the analysis model. Linear regression including IPTW weights evaluated the relationship between exposure status and each outcome, using generalized estimating equations and a robust variance. For both the primary and secondary endpoints (Table 1, Supplemental Digital Content 3) a two-tailed p-value of <0.05 was considered statistically significant for the overall 2 degrees of freedom tests across exposure categories.

Pairwise comparisons of single and multiple exposures versus no exposure were performed using $p<0.025$ (Bonferroni adjustment) to denote statistical significance. For all comparisons, findings were summarized using point-estimates and corresponding 95% confidence intervals, reflecting the combined analysis of multiple imputations. Age at testing was evaluated as a potential moderator of the effect of exposure on outcomes, with the interaction between exposure group and age at testing group (8-12 vs. 15-20 years) assessed for all outcomes.

Some of the variables measured have established cut-offs for defining clinically meaningful deficits, including the WASI Full Scale IQ score (<85) and the Child Behavior Checklist (>60). These variables were dichotomized accordingly and analyzed using logistic regression in additional analyses.

For endpoints found to be significantly associated with exposure, four potential moderators (sex, gestational age, birth weight, and socioeconomic status) were examined.[23] For each, regression analyses were performed which included explanatory variables for exposure category, the potential moderator variable, and the moderator-by-exposure interaction effect.

Sensitivity analyses—In the IPTW analysis, some combinations of covariates in the propensity score model may lead to small or large weights such that individuals have small or large amounts of influence on the exposure comparisons, which can lead to large variation of effect estimates. Weight truncation was performed to evaluate a possible bias-variance tradeoff and sensitivity to the original weights.[21,24] Truncations were performed on the distribution of weights using the 1st and 99th percentiles, 5th and 95th percentiles, and 10th and 90th percentiles. As another method to explore the potential effect of extreme weights, for each sex, subjects were stratified by quintiles of the propensity score distribution (10 strata total) among the multiply exposed. For this procedure, the propensity to be multiply exposed was estimated as in the primary analysis. Among multiply exposed, separately for each sex, quintiles were obtained and participants stratified according to those quintile and sex combinations. Because quintiles reflect the distribution of multiply exposed, this analysis also targets the average treatment effect among the multiply exposed. Results of the stratified analyses reflect the combined estimate across the strata.

In additional sensitivity analyses, multivariable regression models were used rather than IPTW to adjust for potential confounders. These models used those variables previously identified and used in the propensity score model, provided an estimate of an effect of exposure on treatment, conditional on baseline covariates. In another analysis, crude estimates of the differences between exposure groups were also performed without adjustment or weighting; this approach does not account for the sampling framework or any imbalance of covariates among the exposure groups. Finally, a *post hoc* sensitivity analysis using the primary IPTW approach excluded 18 children with cardiopulmonary bypass or intracranial procedures.

Statistical power—At the time of study design no estimate of effect size for the primary endpoint of IQ was available.[19] Prior work found that mean group academic achievement test scores in the multiply-exposed were lower than for those not exposed by approximately

0.4 standard deviation units and that the scores of those with single exposures were similar to those with no exposure.[5] Power calculations were based on these data. The originally targeted sample sizes provided statistical power (two-tailed, alpha=0.025) of 80% to detect a difference of 0.37 and 0.32 standard deviation units respectively within each age group for pairwise comparisons of multiply- and singly- exposure children versus those not exposed, respectively. Power calculations based on the actual numbers tested are provided in Supplemental Digital Content 2.

All data were analyzed using SAS 9.4 TS1M3 (SAS institute, INC, Cary, NC).

Results

Subjects were tested from November 2012 to November 2016. From the 19,296 children initially screened as potentially eligible for recruitment, 3,106 were invited to participate and 998 (32%) enrolled, with highest enrollment rates in the multiply-exposed (26%, 35%, and 43% of un-, singly-, and multiply-exposed children, respectively) (Figure 1). One subject refused all testing subsequent to enrollment. Those who enrolled had parents who were older, better educated, more likely to be married, and more likely to be white, but child characteristics were not different except that enrolled children were more likely to be the product of multiple births and small differences were present in the frequency of some individual ADG co-morbidity clusters (Table 2).

The median cumulative duration of anesthesia was 45 and 187 min in singly- and multiply-exposed children tested, respectively, with two-thirds of multiply-exposed children receiving more than 2 h of anesthesia (Table 3). The most common procedure type was otorhinolaryngologic (42% of all procedures, Table 4); cardiovascular and neurological surgeries comprised 4% of procedures (Supplemental Digital Content 4 presents details of all procedures for the multiply-exposed). The most common anesthetic agents utilized included sevoflurane and nitrous oxide (79% and 90% of procedures, respectively, Supplemental Digital Content 5, which lists agents utilized). Approximately half of children received at least one anesthetic after their third birthday (Supplemental Digital Content 6, which presents this subsequent exposure history). Parent and child characteristics were similar among exposure categories; exceptions included small differences in the education of the father, delivery method, and some individual ADG categories (Table 5). The standardized differences between the factors used in the propensity scoring for the primary analysis were small after IPTW adjustment (Figure 2).

Interactions between exposure and age at testing were not significant for any outcome ($p>0.05$); i.e., any effects of exposure did not depend on the age at testing. Age at testing (8-12 vs. 15-20 years) was still included in all models to account for any differences in outcomes that would depend on age.

The primary outcome of WASI Full Scale IQ did not differ significantly according to exposure status, with multiply-exposed children scoring 1.3 points (95% CI -3.8 to 1.2, p=0.32) lower and singly-exposed children scoring 0.5 points (CI -2.8 to 1.9, p=0.70)) lower than unexposed children on average (Table 6). For the other psychometrist-assessed

neuropsychological testing scores of *a* priori primary interest as secondary outcomes (Table 1), only processing speed/automaticity associated with reading skills (as determined by the Rapid Naming Composite of the Comprehensive Test of Phonological Processing [CTOPP]) and the fine motor study composite differed significantly between multiply- and un-exposed children (differences of -3.5 [-6.3 to -0.7] and -5.5[-8.4 to -2.6] respectively, both standard scores) (Table 6). These scores did not differ significantly between singly- and un-exposed children. There were no significant differences in measures of attention, memory, executive function, expressive language, visual-motor abilities, or visual-spatial abilities between unexposed and either exposure category (Table 6).

When all psychometrist-assessed scores were considered (Supplemental Digital Content 7 and 8, which present statistical comparisons and estimates, respectively, for all psychometrist-assessed scores and parent reports), of the eight scores that were both dependent on motor ability and had a timed component,[1] seven were significantly lower in multiply-exposed children. Thus, the results suggest a consistent impairment of fine motor function and processing speed associated with multiple exposures. No score was significantly different in singly-exposed children.

For the secondary outcome of parent reports, the Behavior Rating Inventory of Executive Function (BRIEF) and the Colorado Leaning Difficulties Questionnaire (CLDQ) Reading (but not Math) scales were significantly greater (indicating more problems) in both singly- and multiply-exposed children (Table 6) compared with unexposed children. All scales of the Child Behavior Checklist (CBCL) were significantly greater (indicating more problems) in multiply- but not singly-exposed children compared with unexposed children. The proportion of children with clinically-abnormal parent-reported scores was significantly greater for the CBCL Externalizing Problems scale in both singly- and multiply-exposed children (Supplemental Digital Content 7 and 8).

In moderator analysis examining those scores significantly different in multiply-exposed children, sex, gestational age, birth weight, and socioeconomic status did not moderate the association between exposures and any score, with the exception of the interaction term for socioeconomic status and CTOPP (Table 7). Given the multiple interaction terms sought across multiple outcomes, the significance of this isolated interaction term is unclear.

In sensitivity analyses (Supplemental Digital Content 9, which provides the results of these analyses), the crude analysis (i.e., no adjustments via weighting or other methods) produced the largest effect sizes for most scores, with trends observed in the primary analysis now statistically significant for several scores. For most scores, absolute effect sizes also were larger for covariate-adjusted analysis, and increased as the degree of IPTW truncation increased. Stratification of IPTW scores and imputation of missing outcome values had little effect. For all adjusted analyses, there was still little evidence in any analysis for exposure effects on any measure of attention, executive function, memory, expressive language, visual-motor abilities, or visual-spatial abilities. For the parent reports, several of the

[1]Grooved Pegboard dominant and other hand, Beery Motor Coordination, and the five Delis-Kaplan Executive Function System [D-KEFS] Trail-making tasks).

sensitivity analyses now demonstrated significant differences in all measures reported for singly-exposed children. These results suggest despite the propensity-guided recruitment strategy employed, there were still imbalances in the baseline characteristics among children who actually enrolled that affected some interpretations of exposure effects in terms of statistical significance for secondary outcomes (Figure 2).

Discussion

Regarding the primary outcome, exposure to procedures requiring general anesthesia before the age of 3 years was not associated with significant differences in general cognitive ability as quantified by Full Scale IQ score, relative to unexposed children. Regarding secondary outcomes, multiple, but not single, exposures were associated with decreases in a processing speed task related to retrieval of verbal codes associated with reading and fine motor coordination, but not other psychometrist-assessed domains. The parents of multiply-exposed children reported more problems related to executive function, behavior, and reading (but not math); the parents of singly-exposed children reported more problems related to executive function and reading. These findings did not depend on age at testing.

The absence of association between exposure and Full Scale IQ is consistent with several smaller prior studies.[12,13,18] A large population-based study[10] found small effects of exposures prior to age 4 years (decreases of 0.97 points [95% CI -1.78 to -0.15] for a single exposure and 1.02 points [95% CI -3.43 to 1.39] for two exposures) of similar magnitude to the present study, which was not powered to detect this small effect size. Thus, the present results add to the evidence that exposure is associated with no or a small effect on general intelligence.

Analysis of secondary outcomes revealed a specific pattern of changes. This analysis is of importance (albeit with appropriate cautions in interpretation since they are secondary outcomes) given that there was little understanding of a likely phenotype at the time of study design. Two prior studies have utilized comparable comprehensive neuropsychological assessments. A sibling-matched cohort study of 105 children singly-exposed prior to age 36 months (the PANDA study) found no significant differences in a battery of tests similar to ours, with the exception of more exposed children having abnormal CBCL internalizing scores.[18] Unlike our results, they found no differences in the BRIEF parent assessment of executive function. In a study of approximately 200 children who were singly or multiply (20% of children) exposed to anesthesia prior to age 3,[14] exposure was associated with significant deficits in performance IQ and language abilities as well as tendencies for decreases in combined fine and gross motor performance and increased CBCL problems. In an additional analysis, among those exposed between ages 3 to 5 years, motor performance but not other domains were significantly affected.[25] Two other small studies evaluated a more limited range of domains. Single exposures were associated with decrements in listening comprehension and performance IQ.[13] Children exposed prior to age 1 year assessed with an object recognition test had lower recollection memory scores, but no differences in the CBCL or familiarity scores.[12] Due to differences in study design and assessments, it is difficult to directly compare all of these results with ours. Broad areas of consistency include some evidence for exposure being associated with differences in

performance IQ, motor skills, and parent ratings of behavior. In contrast, we failed to identify associations with measures of language processing or memory, although we utilized different assessments that may have measured different constructs. Thus, our study is unique in finding decreases in scores reflecting fine motor skills and processing speed, in the absence of changes in scores assessing other cognitive domains, in children receiving multiple exposures. These decreases are modest (effect sizes of less than 0.5 SD) and occur in the context of relatively normal performance in unexposed children (see estimates in Supplemental Digital Content 8).

If confirmed in further analyses, would modest differences in these two domains be potentially relevant to children and their families? It is not possible to make definitive conclusions, but the parent-reported outcomes and results of prior studies may provide insights. Several reviews summarize studies examining the association between exposure and patient-relevant outcomes such as behavioral problems, learning difficulties, and academic achievement.[26,27] Multiple, but not single, exposures to anesthesia are associated with an increased risk of ADHD, learning disabilities and decreased performance in group-administered assessments of ability and achievement,[5-7,23] outcomes of potential relevance to children and families. The association between multiple exposures and parent reports of ADHD problems on the CBCL is consistent with this prior work, also performed in children born in Olmsted County, MN. The association of single exposures with reduced scores on reading, but not math, achievement tests[11] is also consistent with the current CLDQ results. The prior work also found an association of multiple exposures with decreases in both reading and math achievement tests;[5] in the current study the CLDQ differed significantly only for reading. However, most of the neuropsychological test results in the current study did not depend on exposure status, including some that may reflect problems with behavior or learning. For example, children with learning disabilities often exhibit impairment in domains such as attention, memory, and executive function,[28,29] yet these domains were not affected. Many children with ADHD exhibit deficits in executive function and attention tests,[30-32] but others do not, especially when evaluated in a focused laboratory setting as contrasted with their natural environment.[33] Also in the current study children being treated for ADHD were not instructed to discontinue medications for testing, which also could have affected results.

Nonetheless, motor deficits and decreases in processing speed are common in children with ADHD or reading disabilities.[34-36] For example, motor deficits are characteristic of Developmental Coordination Disorder, which is associated with ADHD and learning difficulties.[35,37] The fine motor composite was significantly correlated with both CBCL:ADHD Problems (Spearman's rho [r_s] = -0.22) and reading difficulties (r_s = -0.27), and the CTOPP was significantly correlated with both CBCL:ADHD Problems (r_s = -0.14) and reading difficulties (r_s = -0.31)(all $p < 0.0001$), suggesting that these changes may be related to these behavioral and learning difficulties. The finding of a correlation between the CTOPP and parent report of reading difficulties could also be explained by weaknesses with other fundamental skills needed for successful reading (e.g., phonological awareness, sight word vocabulary, and/or phonics), which in part also determine performance on the CTOPP. However, these were not formally assessed due to time constraints with the testing session. In addition, ADHD and learning disabilities frequently co-occur, and other studies suggest

that defects in processing speed may be an underlying explanatory cognitive risk factor for both conditions.[28,38] Our prior work found a high rate of concordance between ADHD and learning disabilities in children multiply-exposed to anesthesia.[23] Consistent with this observation, CBCL:ADHD Problems and CLDQ Reading Scales were correlated ($r_s = 0.42$, $p<0.0001$).

Although diagnoses such as ADHD and learning disabilities are clinically useful, their causes are multifactorial and similar phenotypes may result from different neuropsychological deficits.[33,39,40] If further evidence supports the hypothesis that anesthesia exposure causes a phenotype diagnosed as ADHD or learning disabilities, the underlying mechanism may be unique, and the pattern of neuropsychological abnormalities may differ from other children diagnosed with ADHD or learning disabilities who are not exposed to aesthesia.

These observations also provide context to interpret preclinical studies, although correlations between measures in humans and animals must be made cautiously. Most rodent studies consistently find sustained impairments in learning and memory,[1,12,41] as do the limited primate studies,[42,43] but we find no evidence for an association between exposure and these domains in humans. Most rodent studies find little effect of exposure on measures of attention or locomotor activity, or behavioral tests,[44-46] although some recent studies in mice suggest effects on social behavior.[47] In contrast, primate studies find effects on anxiety-related behaviors[48,49] and motor reflex deficits.[48] Deficits in response rates to operant test battery tasks dependent on motor skills and processing speed are also consistently observed in ketamine-exposed macaques.[43]

Limitations

As with all observational studies, unmeasured confounders may affect outcomes.[6,50,51] A propensity-guided strategy attempted to recruit children who were comparable for health status and other factors potentially relevant to neurodevelopment within a population-based sample to reduce the potential for referral bias, with IPTW used to account for residual imbalances between exposure groups. Still, children who need procedures differ from those who do not, and it is not possible to fully account for such differences.[6,23,52] This raises the potential for confounding by indication if the procedural indication affects neurodevelopmental outcomes,[53] such as may be the case with cardiopulmonary bypass and intracranial procedures.[54,55] However, a *post hoc* sensitivity analysis excluding children who received at least one of these procedures had little effect on the results (Supplemental Digital Content 9). The finding of a specific pattern of changes in secondary outcomes also argues against confounding by indication, as it is not immediately apparent what common underlying condition across all children receiving procedures would produce such a specific pattern. Finally, it is also possible that elements of procedural experience other than anesthesia exposure, such as a stress response to surgery and pain, may affect neurodevelopment. Thus, these findings cannot directly demonstrate causality, but should be interpreted in the context of other animal and human data.[52]

Selection bias is possible given that not all who were invited accepted, and some characteristics of parents who accepted differed from those who did not. Parents who

accepted may have been more concerned about their child's development than those who did not, which could bias parent reports of behavior and learning if such concerns differ according to exposure status. However, the alignment of the current results with our prior population-based studies of ADHD and learning disabilities based on records review[5-7,23] (in which potential selection bias is not an issue) argues against significant bias.

The need to adjust for residual imbalances even after propensity-guided sampling raises the potential for statistical artifacts. Sensitivity analyses revealed that although effect size estimates depended on the adjustment method, the overall pattern of results was little affected. Testing multiple secondary endpoints also has the potential to detect spurious associations (Type 1 error), prompting an *a priori* analysis plan that specified how results would be interpreted (Table 1) and the creation of study composites that reduced the number of comparisons. Although we did note a specific pattern of effects on these secondary endpoints, these results must be interpreted cautiously because they are secondary endpoints and multiple comparisons were made.

Although this study represents the largest in the field to employ detailed neuropsychological assessments, there may still be limitations in the ability to detect small differences according to exposure category. For example, in singly-exposed children, the observed effect sizes for some scores, including the CTOPP and fine motor composite, were intermediate between the effect sizes for un-exposed and multiply-exposed children, but their confidence intervals included 0 (i.e., they were not statistically significant). It is thus possible that even single exposures were associated with subtle changes in some scores, but that our study lacked sufficient power to detect these differences.

Other potential limitations include that 1) although most characteristics of Olmsted County residents resemble those of other Minnesotans, some differ from the US population as a whole;[56] 2) neuropsychological tests were selected to assess important domains across a wide range of ages within a feasible testing period, but there are strengths and weaknesses for all tests, and some relevant domains may have been missed, 3) approximately half of subjects had exposure after the age of 3 years, and these exposures could bias against finding differences if they too affect outcomes,[8-10] and; 4) the analysis examined mean effects over all children tested; this approach may not be sufficiently sensitive to detect significant effects if only some children are affected.

Conclusion

Exposure of children to procedures requiring general anesthesia prior to the age of 3 years is not associated with lower Full Scale IQ in later life (assessed as a primary outcome). In addition, single exposures are not associated with deficits in other neuropsychological domains (assessed as secondary outcomes). These findings should be reassuring to clinicians and families. However, multiple exposures are associated with modest decreases in processing speed and fine motor coordination, but not changes in other neuropsychological domains. Parents report that multiply-exposed children have more difficulties with behavior and reading. These secondary outcomes must be interpreted cautiously, but suggest the hypothesis, which will need to be evaluated in future work, that exposure to multiple

procedures requiring general anesthesia is associated with a subtle, specific pattern of injury that may have consequences for subsequent learning and behavior.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors would like to thank the staff at the Psychological Assessment Lab for their dedicated work in subject testing, Dr. Bradley Peterson (Children's Hospital of Los Angeles and the Keck School of Medicine at the University of Southern California, Los Angeles, CA) for his helpful comments, and Dr. Young Juhn (Mayo Clinic, Rochester, MN) for assistance with determining socioeconomic status as measured by the HOUSES index. The authors would also like to recognize the extraordinary contributions of Dr. Robert Colligan, who was crucial in the design of the study and died during its conduct; he is sorely missed.

Funding Statement: This work was supported by R01 HD071907 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development of the National Institutes of Health, and also utilized the resources of the Rochester Epidemiology Project, supported by R01 AG034676 from the National Institute on Aging of the National Institutes of Health.

References

1. Jevtovic-Todorovic V, Hartman RE, Izumi Y, Benshoff ND, Dikranian K, Zorumski CF, Olney JW, Wozniak DF. Early exposure to common anesthetic agents causes widespread neurodegeneration in the developing rat brain and persistent learning deficits. J Neurosci. 2003; 23:876–82. [PubMed: 12574416]
2. Brambrink AM, Evers AS, Avidan MS, Farber NB, Smith DJ, Zhang X, Dissen GA, Creeley CE, Olney JW. Isoflurane-induced neuroapoptosis in the neonatal rhesus macaque brain. Anesthesiology. 2010; 112:834–41. [PubMed: 20234312]
3. Disma N, Mondardini MC, Terrando N, Absalom AR, Bilotta F. A systematic review of methodology applied during preclinical anesthetic neurotoxicity studies: important issues and lessons relevant to the design of future clinical research. Paediatr Anaesth. 2016; 26:6–36. [PubMed: 26530523]
4. 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. [PubMed: 21415431]
5. Flick RP, Katusic SK, Colligan RC, Wilder RT, Voigt RG, Olson MD, Sprung J, Weaver AL, Schroeder DR, Warner DO. Cognitive and behavioral outcomes after early exposure to anesthesia and surgery. Pediatrics. 2012; 128:e1053–61.
6. Wilder RT, Flick RP, Sprung J, Katusic SK, Barbaresi WJ, Mickelson C, Gleich SJ, Schroeder DR, Weaver AL, Warner DO. Early exposure to anesthesia and learning disabilities in a population-based birth cohort. Anesthesiology. 2009; 110:796–804. [PubMed: 19293700]
7. Sprung J, Flick RP, Katusic SK, Colligan RC, Barbaresi WJ, Bojanic K, Welch TL, Olson MD, Hanson AC, Schroeder DR, Wilder RT, Warner DO. Attention-deficit/hyperactivity disorder after early exposure to procedures requiring general anesthesia. Mayo Clinic Proc. 2012; 87:120–9.
8. O'Leary JD, Janus M, Duku E, Wijeysundera 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. [PubMed: 27433745]
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; 125:667–677. [PubMed: 27655179]
10. 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. 2016:e163470.

11. 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. [PubMed: 19955889]
12. 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, Michelle KW, 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. Neuropsychopharmacol. 2014; 39:2275–2287.
13. Backeljauw B, Holland SK, Altaye M, Loepke AW. Cognition and brain structure following early childhood surgery with anesthesia. Pediatrics. 2015; 136:e1–12. [PubMed: 26055844]
14. 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. [PubMed: 22908104]
15. 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. [PubMed: 22801049]
16. 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. [PubMed: 21368654]
17. 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. Neurodevelopmental outcome at 2 years of age after general anaesthesia and awake-regional anaesthesia in infancy (GAS): an international multicentre, randomised controlled trial. Lancet. 2015; 387:239–250. [PubMed: 26507180]
18. 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. [PubMed: 27272582]
19. 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. 2014; 41:45–54. [PubMed: 25555440]
20. McCaffrey DF, Griffin BA, Almirall D, Slaughter ME, Ramchand R, Burgette LF. A tutorial on propensity score estimation for multiple treatments using generalized boosted models. Stat Med. 2013; 32:3388–414. [PubMed: 23508673]
21. Austin PC, Stuart EA. Moving towards best practice when using inverse probability of treatment weighting (IPTW) using the propensity score to estimate causal treatment effects in observational studies. Stat Med. 2015; 34:3661–79. [PubMed: 26238958]
22. Juhn YJ, Beebe TJ, Finnie DM, Sloan J, Wheeler PH, Yawn B, Williams AR. Development and initial testing of a new socioeconomic status measure based on housing data. J Urban Health. 2011; 88:933–44. [PubMed: 21499815]
23. 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; 127:227–240. [PubMed: 28609302]
24. Cole SR, Hernan MA. Constructing inverse probability weights for marginal structural models. Am J Epidemiol. 2008; 168:656–64. [PubMed: 18682488]
25. 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. 26:377–86. 014.

26. Zhang H, Du L, Du Z, Jiang H, Han D, Li Q. Association between childhood exposure to single general anesthesia and neurodevelopment: a systematic review and meta-analysis of cohort study. J Anesth. 2015; 29:749–57. [PubMed: 26002228]
27. Beers SR, Rofey DL, McIntyre KA. Neurodevelopmental assessment after anesthesia in childhood: review of the literature and recommendations. Anesth Analg. 2014; 119:661–9. [PubMed: 25137001]
28. Willcutt EG, Pennington BF, Olson RK, Chhabildas N, Hulslander J. Neuropsychological analyses of comorbidity between reading disability and attention deficit hyperactivity disorder: in search of the common deficit. Dev Neuropsychol. 2005; 27:35–78. [PubMed: 15737942]
29. Rigoli D, Piek JP, Kane R, Oosterlaan J. Motor coordination, working memory, and academic achievement in a normative adolescent sample: testing a mediation model. Arch Clin Neuropsychol. 2012; 27:766–80. [PubMed: 22777140]
30. Pasini A, Paloscia C, Alessandrelli R, Porfirio MC, Curatolo P. Attention and executive functions profile in drug naive ADHD subtypes. Brain Dev. 2007; 29:400–8. [PubMed: 17207595]
31. Nigg JT, Blaskey LG, Huang-Pollock CL, Rappley MD. Neuropsychological executive functions and DSM-IV ADHD subtypes. J Am Acad Child Adolesc Psychiatry. 2002; 41:59–66. [PubMed: 11800208]
32. Frazier TW, Demaree HA, Youngstrom EA. Meta-analysis of intellectual and neuropsychological test performance in attention-deficit/hyperactivity disorder. Neuropsychology. 2004; 18:543–55. [PubMed: 15291732]
33. Nigg JT, Willcutt EG, Doyle AE, Sonuga-Barke EJ. Causal heterogeneity in attention-deficit/hyperactivity disorder: do we need neuropsychologically impaired subtypes? Biol Psychiatry. 2005; 57:1224–30. [PubMed: 15949992]
34. Piek JP, Dyck MJ, Nieman A, Anderson M, Hay D, Smith LM, McCoy M, Hallmayer J. The relationship between motor coordination, executive functioning and attention in school aged children. Arch Clin Neuropsychol. 2004; 19:1063–76. [PubMed: 15533697]
35. Goulardins JB, Rigoli D, Licari M, Piek JP, Hasue RH, Oosterlaan J, Oliveira JA. Attention deficit hyperactivity disorder and developmental coordination disorder: Two separate disorders or do they share a common etiology. Behav Brain Res. 2015; 292:484–92. [PubMed: 26168770]
36. Kaiser ML, Schoemaker MM, Albaret JM, Geuze RH. What is the evidence of impaired motor skills and motor control among children with attention deficit hyperactivity disorder (ADHD)? Systematic review of the literature. Res Dev Disabil. 2014; 36C:338–357. [PubMed: 25462494]
37. Wocadlo C, Rieger I. Motor impairment and low achievement in very preterm children at eight years of age. Early Hum Dev. 2008; 84:769–76. [PubMed: 18639396]
38. Shanahan MA, Pennington BF, Yerys BE, Scott A, Boada R, Willcutt EG, Olson RK, DeFries JC. Processing speed deficits in attention deficit/hyperactivity disorder and reading disability. J Abnorm Child Psychol. 2006; 34:585–602. [PubMed: 16850284]
39. Willcutt EG, Doyle AE, Nigg JT, Faraone SV, Pennington BF. Validity of the executive function theory of attention-deficit/hyperactivity disorder: a meta-analytic review. Biol Psychiatry. 2005; 57:1336–46. [PubMed: 15950006]
40. Makris N, Biederman J, Monuteaux MC, Seidman LJ. Towards conceptualizing a neural systems-based anatomy of attention-deficit/hyperactivity disorder. Dev Neurosci. 2009; 31:36–49. [PubMed: 19372685]
41. Vutskits L, Xie Z. Lasting impact of general anaesthesia on the brain: mechanisms and relevance. Nat Rev Neurosci. 2016; 17:705–717. [PubMed: 27752068]
42. Alvarado MC, Murphy KL, Baxter MG. Visual recognition memory is impaired in rhesus monkeys repeatedly exposed to sevoflurane in infancy. Br J Anaesth. 2017
43. Paule MG, Li M, Allen RR, Liu F, Zou X, Hotchkiss C, Hanig JP, Patterson TA, Slikker W Jr, Wang C. Ketamine anesthesia during the first week of life can cause long-lasting cognitive deficits in rhesus monkeys. Neurotoxicol Teratol. 2011; 33:220–30. [PubMed: 21241795]
44. Murphy KL, McGaughy J, Croxson PL, Baxter MG. Exposure to sevoflurane anesthesia during development does not impair aspects of attention during adulthood in rats. Neurotoxicol Teratol. 2016

45. Xie H, She GM, Wang C, Zhang LY, Liu CF. The gender difference in effect of sevoflurane exposure on cognitive function and hippocampus neuronal apoptosis in rats. Eur Rev Medical Pharmacol Sci. 2015; 19:647–57.

46. Chung W, Ryu MJ, Heo JY, Lee S, Yoon S, Park H, Park S, Kim Y, Kim YH, Yoon SH, Shin YS, Lee WH, Ju X, Kweon GR, Ko Y. Sevoflurane exposure during the critical period affects synaptic transmission and mitochondrial respiration but not long-term behavior in mice. Anesthesiology. 2017; 126:288–299. [PubMed: 27922840]

47. Lin D, Liu J, Kramberg L, Ruggiero A, Cottrell J, Kass IS. Early-life single-episode sevoflurane exposure impairs social behavior and cognition later in life. Brain Behav. 2016; 6:e00514. [PubMed: 27688943]

48. Coleman K, Robertson ND, Dissen GA, Neuringer MD, Martin LD, Cuzon Carlson VC, Kroenke C, Fair D, Brambrink AM. Isoflurane anesthesia has long-term consequences on motor and behavioral development in infant rhesus macaques. Anesthesiology. 2016; 127:74–94.

49. Raper J, Alvarado MC, Murphy KL, Baxter MG. Multiple anesthetic exposures in infant monkeys alters emotional reactivity to an acute stressor. Anesthesiology. 2015; 123:1084–92. [PubMed: 26313293]

50. Hansen TG. Anesthesia-related neurotoxicity and the developing animal brain is not a significant problem in children. Paediatr Anaesth. 2015; 25:65–72. [PubMed: 25266176]

51. Davidson AJ, Becke K, de Graaff J, Giribaldi G, Habre W, Hansen T, Hunt RW, Ing C, Loepke A, McCann ME, Ormond GD, Pini Prato A, Salvo I, Sun L, Vutskits L, Walker S, Disma N. Anesthesia and the developing brain: a way forward for clinical research. Paediatr Anaesth. 2015; 25:447–52. [PubMed: 25818094]

52. Flick RP, Warner DO. A users' guide to interpreting observational studies of pediatric anesthetic neurotoxicity: The lessons of Sir Bradford Hill. Anesthesiology. 2012; 117:459–62. [PubMed: 22797284]

53. Salas M, Hofman A, Stricker BH. Confounding by indication: an example of variation in the use of epidemiologic terminology. Am J Epidemiol. 1999; 149:981–3. [PubMed: 10355372]

54. Bellinger DC, Jonas RA, Rappaport LA, Wypij D, Wernovsky G, Kuban KC, Barnes PD, Holmes GL, Hickey PR, Strand RD, Walsh AZ, Helmers SL, Constantinou JE, Carrazana EJ, Mayer JE, Hanley FL, Castaneda AR, Ware JH, Newburger JW. Developmental and neurologic status of children after heart surgery with hypothermic circulatory arrest or low-flow cardiopulmonary bypass. N Engl J Med. 1995; 332:549–55. [PubMed: 7838188]

55. Hansen TG, Pedersen JK, Henneberg SW, Morton NS, Christensen K. Neurosurgical conditions and procedures in infancy are associated with mortality and academic performances in adolescence: a nationwide cohort study. Paediatr Anaesth. 2014; 25:186–192. [PubMed: 25251293]

56. St Sauver JL, Grossardt BR, Yawn BP, Melton LJ 3rd, Pankratz JJ, Brue SM, Rocca WA. Data resource profile: the Rochester Epidemiology Project (REP) medical records-linkage system. Int J Epidemiol. 2012; 41:1614–24. [PubMed: 23159830]

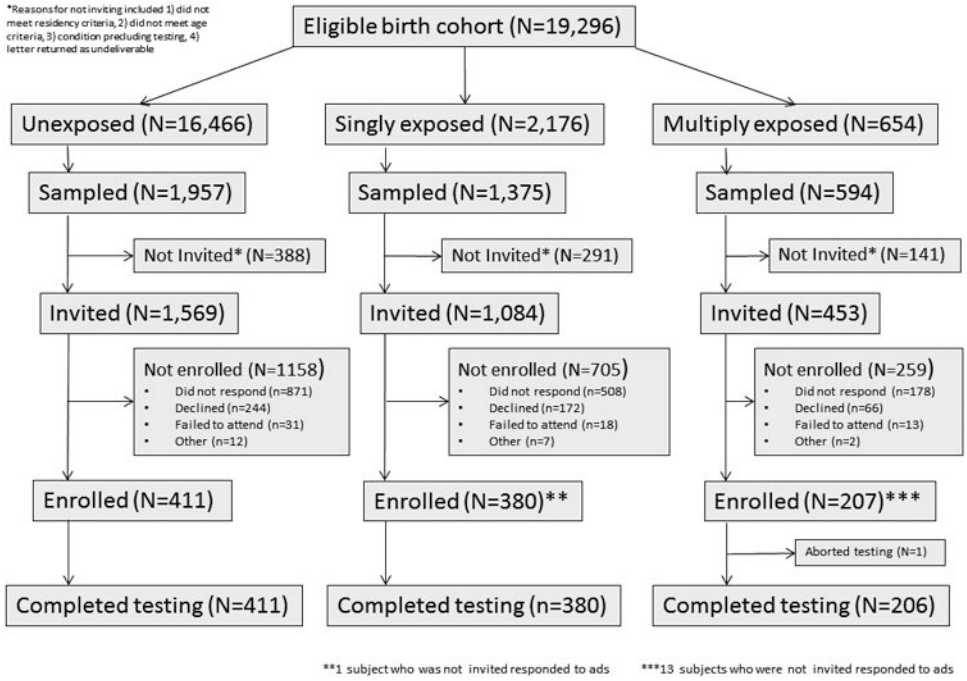

Figure 1.

Flow diagram of subject recruitment. Of the 27,213 children identified by the Minnesota Department of Health as being born in to mothers residing in Olmsted County from 1994-2007, the following were excluded from being potentially eligible for recruitment: 34 who could not be matched to a Rochester Epidemiology Project record, 3,455 who did not provide authorization to use their medical records for research, 2,784 who were not resident in Olmsted County until age 3, 161 who were deceased at the start of recruitment, and 1,483 who lived more than 25 miles from Rochester, MN at the start of recruitment. From the remaining 19,296, a total of 3,926 were sampled using propensity scores as described in the methods for possible recruitment. The original intention was to sample all multiply-exposed children, but recruitment goals were met before all were invited. Of those sampled, 820 were not invited for the reasons indicated. Regarding conditions precluding testing, among all exposure categories, 10 children (0.2% of those sampled) were excluded because of severe intellectual disability, 37 (0.9%) because of autism, 7 (0.2%) because of spastic cerebral palsy, and 59 (1.5%) because of limited English proficiency. Fourteen children responded directly to television advertisements directed towards the multiply-exposed without receiving invitations[19] (one child who was singly-exposed was originally thought to be multiply-exposed but actually had a procedure without general anesthesia). A total of 998 children were enrolled rather than the planned 1000 due to a clerical error that duplicated two children in the recruitment log.

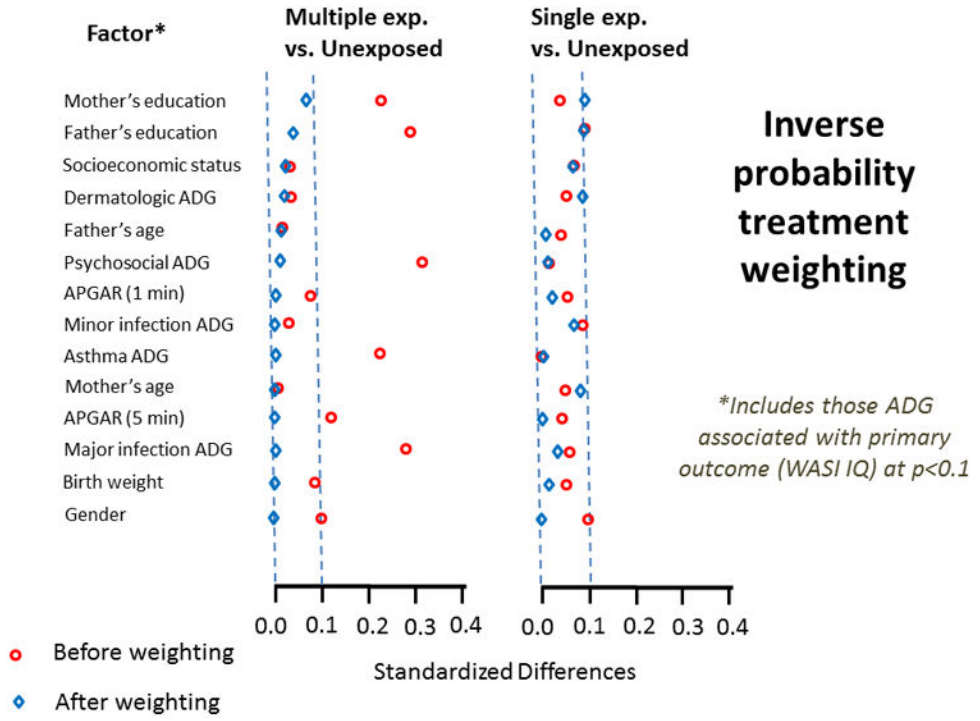

Figure 2.

Standardized differences between the factors used in the propensity scoring for the primary analysis before (red circles) and after (blue diamonds) inverse probability treatment weighting. WASI, Wechsler Abbreviated Scale of Intelligence; IQ, intelligence quotient; ADG, aggregated diagnostic groups; exp., exposed. For both mother's and father's education (the two factors with the largest differences after weighting in multiply-exposed subjects), the residual differences from the weighted estimates are in the direction of the unexposed group having parents who are slightly less educated than the exposed groups. Therefore, if lower parental education is associated with worse outcomes, any residual confounding should bias results toward the null.

Table 1

Summary of how the results of neuropsychological assessment battery instruments are interpreted in terms of domains measured

Psychometrist Assessments						
Instrument	**Domain Measured**	**Type of Score**	**Range**	**Mean**	**Standard Deviation**	**Cut-Off Score**
Wechsler Abbreviated Scale of Intelligence (WASI): Full Scale IQ Score		Composite Standard Score	40-160	100	15	<85
Wide Range Assessment of Memory and Learning-Second Edition (WRAML-2): Attention/Concentration index		Composite Standard Score	55-145	100	15	
Continuous Performance Test II (CPT-II): Hit Reaction Time		T-score	30-90	50	10	>60
WRAML-2: Verbal Memory Index (story memory and verbal learning subtests)		Composite Standard Score	55-145	100	15	
WRAML-2: “Delayed Verbal Recall” composite (Story Memory Delay Recall and Verbal Learning Delay Recall subtests)		Composite Standard Score *	55-145	100	15	
WRAML-2: Verbal Recognition Composite (Story Memory Recognition and Verbal Learning Recognition subtests)		Composite Standard Score	55-145	100	15	
WRAML-2: Design Memory subtest		Scaled Score	1-19	10	3	
WRAML-2: Design Recognition subtest		Scaled Score	1-19	10	3	
Delis-Kaplan Executive Function System (D-KEFs): Tower test, Total Achievement score		Scaled Score	1-19	10	3	
D-KEFs: Trail Making Test: Condition 4		Scaled score	1-19	10	3	
Wisconsin Card Sort: Perseverative Responses		T-score	<20->80	50	10	
Comprehensive Test of Phonological Processing (CTOPP): Rapid Naming Composite		Composite Standard Score	46-130	100	15	
Boston Naming, D-KEFS Category Fluency		Composite Standard Score *	55-145	100	15	
Beery Motor Coordination, Grooved Pegboard Dominant Hand	Fine Motor Skills	Composite Standard Score *	55-145	100	15	
Berry Visual Motor Integration (VMI)	Visual/Motor Integration Abilities	Composite Standard Score	45-130	100	15	
Beery Visual Perception, Judgement of Line Orientation		Composite Standard Score *	55-145	100	15	
Parental reports						
Instrument	**Domain Measured**	**Type of Score**	**Range**	**Mean**	**Standard Deviation**	**Cut-Off Score**

Psychometrist Assessments						
Instrument	**Domain Measured**	**Type of Score**	**Range**	**Mean**	**Standard Deviation**	**Cut-Off Score**
Colorado Learning Difficulties Questionnaire (CLDQ): Math Scale	Learning Difficulties: Math	Z-score	-0.86-3.59	0	1	
Colorado Learning Difficulties Questionnaire (CLDQ): Reading Scale	Learning Difficulties: Reading	Z-score	-0.81-3.63	0	1	
Behavior Rating Inventory of Executive Function (BRIEF): Global Executive Composite	Executive Function	T-score	30-100	50	10	>60
Child Behavior Checklist (CBCL): Internalizing Problems	Behavior	T-score	20-100	50	10	>60
Child Behavior Checklist (CBCL): Externalizing Problems	Behavior	T-score	20-100	50	10	>60
Child Behavior Checklist (CBCL): Total Problems	Behavior	T-score	20-100	50	10	>60
Child Behavior Checklist (CBCL): ADHD Problems	Behavior	T-score	20-100	50	10	>60

Tests include both in-person tests administered by psychometrists, and parental questionnaires. These domains mimic those typically assessed during a clinical exam. This table provides an overall roadmap for how any observed differences according to exposure status are interpreted in terms of effects on overall neuropsychological domains.

* In addition to scores derived from the test developers, 4 study-specific composite scores were defined a priori from individual tests to increase the ability to detect effects when more than one instrument assessed a particular domain and to simplify the consideration of a large number of individual assessments. Individual test scores were first converted to standard scores using available normative data from either the test developers or available scientific literature. The individual standard scores were then averaged to create an overall composite domain score (mean = 100, standard deviation = 15).

D-KEFS, Delis-Kaplan Executive Function System; WASI, Wechsler Abbreviated Scale of Intelligence; WRAML-2, Wide Range Assessment of Memory and Learning- Second Edition; CPT II, Connor's Continuous Performance Test II; CTOPP, Comprehensive Test of Phonological Processing; VMI, visual motor integration; ADHD, attention deficit hyperactivity disorder; CBCL, Child Behavior Checklist; CLDQ, Colorado Learning Difficulties Questionnaire.

Table 2
Characteristics of enrolled subjects compared to subjects invited but not enrolled

Variable	Not enrolled (N=2122)	Enrolled (N=998)	P
Mother's age (y)	29 (5)	30 (5)	<.001[1]
Father's age (y; n=2980)	31 (6)	32 (5)	0.001[1]
Mother's education (y; n=3083)			<.001[2]
<12	142 (7%)	22 (2%)	
12	464 (22%)	120 (12%)	
13-15	641 (31%)	264 (27%)	
16	533 (26%)	342 (34%)	
>16	308 (15%)	247 (25%)	
Father's education (y; n=2892)			<.001[2]
<12	88 (5%)	29 (3%)	
12	543 (28%)	174 (18%)	
13-15	466 (24%)	244 (25%)	
16	484 (25%)	314 (32%)	
>16	333 (17%)	217 (22%)	
Mother's ethnicity, Hispanic (n=3116)	51 (2%)	9 (1%)	0.004 [2]
Father's ethnicity, Hispanic (n=2926)	44 (2%)	18 (2%)	0.43[2]
Mother's race, white (n=3116)	1918 (90%)	971 (97%)	<.001[2]
Father's race, white (n=2928)	1751 (90%)	940 (95%)	<.001[2]
Number of prenatal care visits (n=3036)	11.7 (3.3)	11.9 (3.3)	0.19[1]
Marital status (n=3106)			<.001[2]
Married	1666 (79%)	881 (90%)	
Not married	456 (21%)	103 (10%)	
Sex			0.61[2]
Male	1245 (59%)	576 (58%)	
Female	877 (41%)	422 (42%)	
Estimated gestational age (w; n=3119)	38.5 (2.5)	38.5 (2.5)	0.60[1]
Apgar score, 1 minute (n=3118)	8 (2)	8 (2)	0.13[1]
Apgar score, 5 minutes (n=3117)	9 (1)	9 (1)	0.77[1]
Birth weight (g)	3324 (682)	3345 (700)	0.42[1]
Multiple births	116 (5%)	77 (8%)	0.015[2]
Delivery method (n=3114)			0.24[2]
Vaginal	1608 (76%)	739 (74%)	
C-section	508 (24%)	259 (26%)	
Johns Hopkins ACG Case Mix comorbidity clusters			
Time Limited: Minor	1880 (89%)	900 (90%)	0.19[2]
Time Limited: Minor-Primary Infections	2068 (97%)	961 (96%)	0.07[2]

Variable	Not enrolled (N=2122)	Enrolled (N=998)	P
Time Limited: Major	836 (39%)	450 (45%)	0.003[2]
Time Limited: Major-Primary Infections	782 (37%)	382 (38%)	0.44[2]
Allergies	376 (18%)	180 (18%)	0.83 [2]
Asthma	331 (16%)	165 (17%)	0.51[2]
Likely to Recur: Discrete	1214 (57%)	571 (57%)	1.00[2]
Likely to Recur: Discrete-Infection	1996 (94%)	922 (92%)	0.08[2]
Likely to Recur: Progressive	79 (4%)	48 (5%)	0.15[2]
Chronic Medical: Stable	638 (30%)	322 (32%)	0.21[2]
Chronic Medical: Unstable	970 (46%)	476 (48%)	0.30[2]
Chronic Specialty: Stable-Orthopedic	51 (2%)	35 (4%)	0.08[2]
Chronic Specialty: Stable-Ear, Nose, Throat	592 (28%)	322 (32%)	0.012[2]
Chronic Specialty: Stable-Eye	455 (21%)	215 (22%)	0.95[2]
Chronic Specialty: Unstable-Orthopedic	1 (0%)	4 (0%)	0.021[2]
Chronic Specialty: Unstable-Ear, Nose, Throat	201 (9%)	141 (14%)	<.001[2]
Chronic Specialty: Unstable-Eye	148 (7%)	75 (8%)	0.58[2]
Dermatologic	552 (26%)	269 (27%)	0.58[2]
Injuries/Adverse Effects: Minor	957 (45%)	412 (41%)	0.045[2]
Injuries/Adverse Effects: Major	865 (41%)	411 (41%)	0.82[2]
Psychosocial: Time Limited, Minor	510 (24%)	262 (26%)	0.18[2]
Psychosocial: Recurrent or Persistent, Stable	298 (14%)	170 (17%)	0.029[2]
Psychosocial: Recurrent or Persistent, Unstable	8 (0%)	4 (0%)	0.92[2]
Signs/Symptoms: Minor	1782 (84%)	837 (84%)	0.94[2]
Signs/Symptoms: Uncertain	1988 (94%)	935 (94%)	1.00[2]
Signs/Symptoms: Major	1446 (68%)	707 (71%)	0.13[2]
Discretionary	1599 (75%)	804 (81%)	0.001[2]
See and Reassure	1108 (52%)	552 (55%)	0.11[2]
Prevention/Administrative	2122 (100%)	998 (100%)	1.00[2]
Malignancy	16 (1%)	9 (1%)	0.67[2]
Dental	103 (5%)	45 (5%)	0.67[2]

[1] One-way ANOVA

[2] Pearson Chi-squared

Data are summarized as n (%) for categorical variables and mean (SD) for continuous variables. Values in parentheses refer to the number of individuals for whom data was available

Table 3
Anesthesia exposure characteristics for children tested

	Single exposure (N=380)	Multiple exposures (N=206)
ASA physical status		
1	281 (73.9%)	71 (34.5%)
2	90 (23.7%)	96 (46.6%)
3	9 (2.4%)	34 (16.5%)
4	0 (0.0%)	5 (2.4%)
Age at first exposure (years)		
0-0.9	138 (36.3%)	123 (59.7%)
1-1.9	150 (39.5%)	73 (35.4%)
2-2.9	92 (24.2%)	10 (4.9%)
Duration of anesthesia (minutes)		
mean (SD)	61 (51)	295 (354)
median (Q1, Q3)	45 (25, 81)	187 (99, 326)
1-30	138 (36.3%)	1 (0.5%)
31-60	91 (23.9%)	12 (5.8%)
61-90	79 (20.8%)	33 (16.0%)
91-120	38 (10.0%)	24 (11.7%)
121-180	19 (5.0%)	29 (14.1%)
181-240	10 (2.6%)	31 (15.0%)
≥241	5 (1.3%)	76 (36.9%)
Number of exposures for the multiply-exposed		
2	-	122 (59.2%)
3	-	36 (17.5%)
4-5	-	29 (14.1%)
6-7	-	6 (2.9%)
8-9	-	9 (4.4%)
≥10	-	4 (1.9%)

Q1, first quartile; Q3, third quartile; SD, standard deviation.
586 children underwent 1052 anesthetics. For the 206 children who underwent multiple (range, 2-29) anesthetics, the highest ASA physical status and the total cumulative duration of anesthesia are presented. The median (Q1, Q3) duration per anesthetic for all anesthetics was 58 (30, 100) minutes. For the four patients who had ≥10 anesthetics, exposure counts were 11, 15, 26, and 29.

Table 4

Types of procedures by exposure status

	Overall (N=1052)	Single exposure (N=380)	Multiple exposures (N=672)
Procedure type, n (%)			
General surgery	149 (14%)	41 (11%)	108 (16%)
Otorhinolaryngologic	441 (42%)	197 (52%)	244 (36%)
Neurologic surgery	11 (1%)	3 (1%)	8 (1%)
Urologic surgery	81 (8%)	36 (9%)	45 (7%)
Orthopedic surgery	38 (4%)	10 (3%)	28 (4%)
Plastic surgery	57 (5%)	16 (4%)	41 (6%)
Cardiovascular surgery	33 (3%)	3 (1%)	30 (4%)
Other*	242 (23%)	74 (19%)	168 (25%)

For those who were multiply exposed, data is presented for the 206 children who completed at least one neuropsychological test.

*"Other" procedures include oral surgeries, ophthalmology surgeries, diagnostic procedures, and examination during anesthesia

Table 5
Parent and child characteristics for those enrolled by exposure category

	No exposures (N=411)	Single exposure (N=380)	Multiple exposures (N=206)	P Value
Sex				0.06[1]
Male	219 (53%)	232 (61%)	124 (60%)	
Female	192 (47%)	148 (39%)	82 (40%)	
Age at testing (y)				0.06[1]
8-12	224 (55%)	200 (53%)	129 (63%)	
15-20	187 (45%)	180 (47%)	77 (37%)	
Mother's ethnicity, Hispanic (n=995)	4 (1%)	5 (1%)	0 (0%)	0.27[1]
Father's ethnicity, Hispanic (n=986)	5 (1%)	9 (2%)	4 (2%)	0.47[1]
Mother's race, white (n=995)	400 (97%)	374 (99%)	196 (96%)	0.07[1]
Father's race, white (n=986)	387 (95%)	362 (96%)	190 (94%)	0.56[1]
Mother's education (y; n=994)				0.32[1]
<12	11 (3%)	8 (2%)	3 (1%)	
12	43 (10%)	45 (12%)	32 (16%)	
13-15	106 (26%)	96 (25%)	62 (30%)	
16	145 (35%)	127 (34%)	69 (34%)	
>16	105 (26%)	103 (27%)	39 (19%)	
Father's education (y; n=977)				0.009[1]
<12	11 (3%)	12 (3%)	6 (3%)	
12	59 (15%)	59 (16%)	55 (28%)	
13-15	95 (24%)	107 (28%)	42 (21%)	
16	142 (36%)	115 (31%)	57 (29%)	
>16	93 (23%)	84 (22%)	40 (20%)	
Mother's age (y)	30.2 (5.0)	30.3 (4.8)	30.0 (4.9)	0.86[2]
Father's age (y; n=968)	32.3 (5.2)	31.9 (5.1)	31.9 (5.4)	0.50[2]
Marital status (n=983)				0.41[1]
Married	366 (89%)	345 (91%)	169 (88%)	
Not married	45 (11%)	34 (9%)	24 (12%)	
HOUSES index (quartile; n=967)*				0.88[1]
1	58 (14%)	55 (15%)	27 (14%)	
2	93 (23%)	102 (27%)	48 (25%)	
3	106 (26%)	98 (26%)	51 (27%)	
4	145 (36%)	119 (32%)	65 (34%)	
Estimated gestational age (w)	38.7 (2.4)	38.6 (2.4)	38.2 (2.8)	0.09[2]
Birth weight (g)	3341 (664)	3400 (697)	3267 (745)	0.08[2]
Multiple births	26 (6%)	37 (10%)	13 (6%)	0.14[1]
Apgar score, 1 minute	8 (1)	8 (2)	9 (2)	0.60[2]

	No exposures (N=411)	Single exposure (N=380)	Multiple exposures (N=206)	P Value
Apgar score, 5 minutes	9 (1)	9 (1)	9 (1)	0.25[2]
Delivery method				0.005[1]
Vaginal	321 (78%)	282 (74%)	136 (66%)	
C-section	90 (22%)	98 (26%)	70 (34%)	
Johns Hopkins ACG Case Mix comorbidity clusters				
Time Limited: Minor	356 (87%)	350 (92%)	193 (94%)	0.006[1]
Time Limited: Minor-Primary Infections	390 (95%)	371 (98%)	199 (97%)	0.12[1]
Time Limited: Major	161 (39%)	156 (41%)	132 (64%)	<.001[1]
Time Limited: Major-Primary Infections	139 (34%)	144 (38%)	98 (48%)	0.004[1]
Allergies	73 (18%)	68 (18%)	39 (19%)	0.93[1]
Asthma	60 (15%)	56 (15%)	49 (24%)	0.007[1]
Likely to Recur: Discrete	193 (47%)	225 (59%)	152 (74%)	<.001[1]
Likely to Recur: Discrete-Infection	365 (89%)	361 (95%)	195 (95%)	0.002[1]
Likely to Recur: Progressive	14 (3%)	16 (4%)	17 (8%)	0.023[1]
Chronic Medical: Stable	100 (24%)	115 (30%)	106 (51%)	<.001[1]
Chronic Medical: Unstable	184 (45%)	174 (46%)	117 (57%)	0.012[1]
Chronic Specialty: Stable-Orthopedic	13 (3%)	8 (2%)	14 (7%)	0.012[1]
Chronic Specialty: Stable-Ear, Nose, Throat	80 (19%)	149 (39%)	92 (45%)	<.001[1]
Chronic Specialty: Stable-Eye	76 (18%)	77 (20%)	61 (30%)	0.005[1]
Chronic Specialty: Unstable-Orthopedic	0 (0%)	2 (1%)	2 (1%)	0.18[1]
Chronic Specialty: Unstable-Ear, Nose, Throat	11 (3%)	82 (22%)	47 (23%)	<.001[1]
Chronic Specialty: Unstable-Eye	24 (6%)	22 (6%)	28 (14%)	<.001[1]
Dermatologic	114 (28%)	91 (24%)	64 (31%)	0.16[1]
Injuries/Adverse Effects: Minor	153 (37%)	169 (44%)	90 (44%)	0.09[1]
Injuries/Adverse Effects: Major	138 (34%)	153 (40%)	119 (58%)	<.001[1]
Psychosocial: Time Limited, Minor	98 (24%)	103 (27%)	60 (29%)	0.32[1]
Psychosocial: Recurrent or Persistent, Stable	57 (14%)	52 (14%)	60 (29%)	<.001[1]
Psychosocial: Recurrent or Persistent, Unstable	1 (0%)	0 (0%)	3 (1%)	0.023[1]
Signs/Symptoms: Minor	326 (79%)	324 (85%)	186 (90%)	0.001[1]
Signs/Symptoms: Uncertain	374 (91%)	359 (94%)	201 (98%)	0.005[1]
Signs/Symptoms: Major	274 (67%)	261 (69%)	171 (83%)	<.001[1]
Discretionary	305 (74%)	315 (83%)	183 (89%)	<.001[1]
See and Reassure	222 (54%)	201 (53%)	128 (62%)	0.08[1]
Malignancy	1 (0%)	0 (0%)	8 (4%)	<.001[1]

	No exposures (N=411)	Single exposure (N=380)	Multiple exposures (N=206)	P Value
Dental	18 (4%)	9 (2%)	18 (9%)	0.002[1]

[1] Pearson Chi-squared

[2] One-way ANOVA

One enrolled patient was excluded from the summary because they did not complete any testing.

[*] The HOUSES index is a measure of socioeconomic status.[22]

Table 6

Primary analysis for scores chosen *a priori* as of primary interest

Test	N	P*	Single exposure, estimate (95% CI) of difference from unexposed	P**	Multiple exposures, estimate (95% CI) of difference from unexposed	P**	Unexposed, estimate of mean (95% CI)†
Psychometrist-administered							
WASI: Full Scale IQ Score	993	0.599	-0.46 (-2.78, 1.87)	0.701	-1.27 (-3.75, 1.21)	0.315	106.0 (104.2, 107.8)
WRAML-2: Attention/Concentration Index (StdS)	926	0.404	0.59 (-2.34, 3.53)	0.692	-1.58 (-4.57, 1.40)	0.299	104.0 (102.1, 105.8)
CPT-II: Hit Reaction Time (TS)	981	0.504	0.20 (-1.89, 2.29)	0.853	1.09 (-0.84, 3.03)	0.268	51.6 (50.2, 53.0)
WRAML-2: Verbal Memory Index (StdS)	993	0.269	1.27 (-1.08, 3.63)	0.289	-0.82 (-3.23, 1.60)	0.507	105.7 (104.1, 107.2)
WRAML-2: Delayed Verbal Recall composite (StdS)††	994	0.340	1.13 (-0.84, 3.11)	0.261	-0.40 (-2.40, 1.60)	0.697	104.5 (103.2, 105.8)
WRAML-2: Verbal Recognition Composite (StdS)	992	0.343	0.78 (-1.46, 3.02)	0.495	-1.02 (-3.27, 1.24)	0.378	106.5 (105.0, 107.9)
WRAML-2: Design Memory subtest (SS)‡	995	**0.037**	0.11 (-0.37, 0.60)	0.648	-0.54 (-1.03, -0.05)	0.030	8.9 (8.6, 9.2)
WRAML-2: Design Recognition subtest (SS)	995	0.828	0.17 (-0.38, 0.71)	0.550	0.12 (-0.44, 0.68)	0.678	9.9 (9.5, 10.2)
D-KEFS Trail Making Test: Condition 4 (SS)	994	0.194	0.22 (-0.37, 0.80)	0.469	-0.33 (-0.99, 0.33)	0.323	10.0 (9.5, 10.5)
D-KEFS Tower Test: Total Achievement Score (SS)	996	0.350	0.02 (-0.42, 0.46)	0.939	-0.29 (-0.76, 0.17)	0.216	10.1 (9.8, 10.4)
Wisconsin Card Sort: Perseverative Responses (TS)	958	0.104	1.46 (-0.92, 3.85)	0.229	-1.40 (-3.97, 1.17)	0.286	55.4 (53.8, 57.0)
CTOPP: Rapid Naming Composite Score (StdS)	995	**0.047**	-2.03 (-4.78, 0.72)	0.147	-3.51 (-6.32, -0.70)	**0.014**	97.8 (95.9, 99.8)
Expressive language composite (StdS)††	882	0.202	1.31 (-1.21, 3.83)	0.309	-1.13 (-3.75, 1.49)	0.398	103.6 (101.9, 105.3)
Fine motor composite (StdS)††	988	**<.001**	-1.34 (-3.91, 1.23)	0.306	-5.53 (-8.42, -2.64)	**<.001**	93.7 (92.0, 95.4)
Beery: Visual-Motor Integration (StdS)	990	0.148	-2.48 (-5.00, 0.04)	0.054	-1.40 (-3.77, 0.97)	0.246	89.4 (87.6, 91.1)
Visual-spatial abilities composite (StdS)††,‡‡	529	0.490	-1.73 (-4.61, 1.15)	0.239	-0.95 (-3.73, 1.84)	0.506	99.4 (97.5, 101.2)
Parent report							
BRIEF: Global Executive Composite (TS)	840	**<.001**	4.64 (2.36, 6.92)	**<.001**	3.23 (0.89, 5.57)	**0.007**	46.3 (44.9, 47.8)
CBCL: Internalizing Problems (TS)	829	**0.038**	1.16 (-1.25, 3.57)	0.347	2.94 (0.68, 5.20)	**0.011**	47.5 (46.0, 49.1)
CBCL: Externalizing Problems (TS)	829	**<.001**	2.03 (-0.24, 4.30)	0.080	3.91 (1.90, 5.92)	**<.001**	43.2 (41.9, 44.6)
CBCL: Total Problems (TS)	828	**0.001**	2.10 (-0.68, 4.88)	0.139	4.75 (2.23, 7.27)	**<.001**	43.8 (41.9, 45.6)
CBCL: ADHD Problems (TS)	828	**0.026**	1.41 (-0.05, 2.86)	0.058	1.97 (0.51, 3.44)	**0.008**	52.6 (51.5, 53.7)
CLDQ: Math Scale (ZS)	894	0.210	0.17 (-0.09, 0.43)	0.204	0.23 (-0.04, 0.50)	0.089	0.04 (-0.15, 0.23)
CLDQ: Reading Scale (ZS)	895	**<.001**	0.24 (0.07, 0.42)	**0.005**	0.44 (0.24, 0.64)	**<.001**	-0.29 (-0.39, -0.19)

This primary analysis used inverse probability of treatment weighting (IPTW) to account for imbalances across exposure categories among children actually tested.

[*] 2 degrees of freedom test for overall difference, significance indicated in bold for $p < 0.05$

[**] Pairwise comparisons of single or multiple exposures versus no exposure, significance indicated in bold for $p<0.025$ (Bonferroni adjustment).

[†] Estimate for unexposed children from regression analysis (95% CI).

[††] In addition to scores derived from the test developers, 4 study- specific composite scores were defined a priori from individual tests to increase the ability to detect effects when more than one instrument assessed a particular domain (see Supplemental Digital Content 3 for components of individual composites) and to simplify the consideration of a large number of individual assessments. Individual test scores were first converted to standard- scores using available normative data from either the test developers or available scientific literature. The individual standard-scores were then averaged to create an overall composite domain score (mean = 100, standard deviation = 15).

[‡] The WRAML-2 Design Memory task requires drawing figures by pencil, thus motor coordination.

[‡‡] Because normative data was not available for the Judgement of Line Orientation for older subjects (ages 15-20), these older subjects were not included in the analysis of this study composite.

N indicated number of children completing a given test. TS, T score (mean = 50, standard deviation = 10); SS, Scaled score (mean = 10, standard deviation = 3); StdS, Standard score (mean = 100, standard deviation = 15); ZS, Z score (mean =0, standard deviation = 1). D-KEFS, Delis-Kaplan Executive Function System; WASI, Wechsler Abbreviated Scale of Intelligence; WRAML-2, Wide Range Assessment of Memory and Learning- Second Edition; CPT II, Connor's Continuous Performance Test II; CTOPP, Comprehensive Test of Phonological Processing; ADHD, attention deficit hyperactivity disorder; BRIEF, Behavior Rating Inventory of Executive Function; CLDQ, Colorado Learning Difficulties Questionnaire; CBCL, Child Behavior Checklist.

Table 7

Analysis of variables that potentially moderate those scores that differed significantly among groups

	Moderator variable				
	Male gender	**Gestational age**	**Birth weight**	**HOUSES index***	**Age group**
BRIEF: Global Executive Composite (TS)					
2df exposure × moderator	0.291	0.677	0.614	0.447	0.923
Single exposure × moderator	0.122	0.483	0.328	0.312	0.978
Multiple exposures × moderator	0.703	0.422	0.723	0.814	0.727
CBCL: ADHD Problems (TS)					
2df exposure × moderator	0.154	0.482	0.390	0.616	0.712
Single exposure × moderator	0.054	0.362	0.172	0.361	0.572
Multiple exposures × moderator	0.230	0.235	0.244	0.825	0.425
CBCL: Total Problems (TS)					
2df exposure × moderator	0.850	0.666	0.866	0.428	0.862
Single exposure × moderator	0.711	0.890	0.663	0.214	0.621
Multiple exposures × moderator	0.574	0.508	0.596	0.951	0.659
CLDQ: Reading Scale (ZS)					
2df exposure × moderator	0.839	0.299	0.357	0.760	0.805
Single exposure × moderator	0.595	0.417	0.751	0.801	0.875
Multiple exposures × moderator	0.702	0.145	0.199	0.420	0.512
CTOPP: Rapid Naming Composite Score (StdS)					
2df exposure × moderator	0.984	0.883	0.210	**0.005**	0.424
Single exposure × moderator	0.863	0.812	0.646	0.944	0.909
Multiple exposures × moderator	0.961	0.618	0.229	**0.004**	0.228
Fine motor composite (StdS)					
2df exposure × moderator	0.849	0.883	0.101	0.816	0.148
Single exposure × moderator	0.776	0.957	0.545	0.590	0.138
Multiple exposures × moderator	0.719	0.646	0.111	0.804	0.079

*The HOUSES index is a measure of socioeconomic status.[22]

Separate analyses were performed for each potential moderator variable for both the overall 2 degrees of freedom (df) test and each individual comparison of single or multiple exposures. Gestational age and birth weight were analyzed as continuous variables. Gender, HOUSES index, and age group were analyzed as categorical variables. The values presented in the table correspond to the p-value for the given moderator-by-anesthesia exposure interaction term.

TS, T score (mean = 50, standard deviation = 10); SS, Scaled score (mean = 10, standard deviation = 3); StdS, Standard score (mean = 100, standard deviation = 15); ZS, Z score (mean = 0, standard deviation = 1). CTOPP, Comprehensive Test of Phonological Processing; ADHD, attention deficit hyperactivity disorder; CLDQ, Colorado Learning Difficulties Questionnaire; CBCL, Child Behavior Checklist.