Anesthetic Exposure and Academic Achievement: Missing the Forest for the Trees

To the Editor:
I read with interest the recent report from Block et al., examining the relation between anesthetic exposure in the first year of life and academic achievement, but have concerns that their analysis inappropriately emphasizes the post hoc observation of an excess of subjects with extremely low achievement scores rather than overall academic achievements. The authors do not explicitly state their Null (Alternative) Hypotheses; however, based on the title of their report and its Introduction, I suspect that they were of the form "the distribution of achievement test scores in children exposed to anesthesia before age 1 yr does not (does) differ from that seen in children not so exposed." The analysis presented by the authors would be appropriate only to the far more specific Null (Alternative) Hypotheses that "children exposed to anesthesia before 1 yr of age are not (are) more likely to score in the lowest fifth percentile on tests of academic achievement." The authors' analysis unnecessarily discards almost all the information available from the distribution of scores.

When considering distributions of outcomes, a more appropriate analysis would be the chi-square goodness-of-fit test. Based on a minimum cell expected value of 5, we can divide the authors' data from figure 1 into 10 deciles, each with an expected value of 5.8 (table 1).

For this distribution, chi-square test is equal to 8.55 with 9 degrees of freedom ($df = 0.48$). This should be interpreted as meaning that the probability of randomly drawing a sample of 58 subjects from a uniformly distributed population, at least this extreme, is 48%. In other words, the distribution of the achievement scores for the authors' subjects is entirely consistent with a randomly selected sample from the population of all children taking the tests. A similar analysis of subjects' percentile ranks converted to z-scores yields the even less significant result of chi-square test of 2.16 for $df = 5 (P = 0.83)$.

An analogy may be helpful. We are given an icosahedral (12-sided) die and wish to determine if it is fair. The die is rolled 60 times so that the expected value for each possible outcome is five; the results are recorded in a table similar to that in the preceding paragraph (table 2). It is tempting to interpret the excess number of 4's rolled as evidence that the die is unfair; the result is certainly statistically "significant" for the binomial outcome "4" versus "not 4" (the 95% CI for the observed frequency for each result based on 60 rolls of a fair 12-sided die is 0–9, which does not include the observed value of 10). This conclusion would be incorrect. There was no a priori reason to suspect that the die would roll more 4's than any other number. Because we had neither specified nor predicted the outcome of interest, we must be prepared to analyze every possible outcome, including an appropriate adjustment to the test-wise level of $\alpha$ to maintain an experiment-wise level of $\alpha$ of 0.05. Because this adjustment can result in an extremely conservative test-wise $\alpha$, we instead look at the overall distribution of rolls; based on this (chi-square test = 9.60; $df = 11; P = 0.57$), we conclude that there is no evidence based on this experiment that the die is unfair. Of course, we may wish to repeat the experiment with the hypothesis that the die favors 4's, especially given the relative deficit of rolls on the side opposite the 4 (the 9).

In a similar way, it seems obvious that the excess number of subjects observed in the lowest fifth percentile must mean that the academic achievement was lower in the group exposed to anesthesia in the first year of life. However, there are other deciles whose deviation from the expected is more extreme (50–60 and 60–70%); we are not free to pick and choose which extreme deviations to treat as random chance and which to analyze as meaningful unless we have included a specific deviation in the Null and Alternative Hypotheses. There are many ways that exposure to anesthetics might impact academic achievement. Exposure might decrease each subject's achievement by a fairly consistent amount relative to that which would have been seen absent exposure, in effect shifting the entire normal curve to the left. Although this would indeed result in an excess of subjects in the lowest fifth percentile, it would also result in a relative deficit in the higher percentiles. This is not seen in the authors' results; indeed, two subjects scored at the 100th percentile, far above the expected 0.58 subjects. Alternatively, anesthetic exposure might result in disparate impact on the subjects with the highest potential achievement, resulting in a dearth of subjects with high scores, but no major alteration in the distribution of low scores; again, this is not seen. Examination of the authors' figure 1 shows that their results are almost entirely dependent on the presence of an unexpectedly large number of subjects (five) scoring at the second percentile; indeed, if we remove one of these subjects by excluding the four subjects who had additional surgical procedures beyond those in the inclusion criteria (authors' figs. 2 and 3 and related discussion; see next paragraph), the resulting distribution falls nicely along the line of identity with the expected uniform distribution (chi-square test = 8.96; $df = 9; P = 0.44$) and the mean percentile score of the remaining 54 subjects is 49.8. As per the authors' discussion, their results seem most consistent with a mechanism that leads to minimal or no impact in the vast majority of subjects and severe, pervasive neurocognitive impairment in a very small subset of subjects.

Finally, some observations about their figures 2 and 3: it seems that the significance of the linear relations reported is entirely dependent on the presence of subjects who either
had multiple exposures to anesthesia (fig. 2) or are extreme outliers with respect to duration of surgery (fig. 3). It seems unlikely that this subject, who scored at the second percentile and whose anesthetic lasted 4 h, had only a circumcision, herniorrhaphy, or pyloromyotomy; exclusion of this subject from figure 3 leaves a simple vertical cluster without any obvious trend. It would be interesting to see whether the relation between academic achievement and duration of surgery or anesthesia remains significant if the analysis is restricted to the 45 subjects who had only circumcision, herniorrhaphy, or pyloromyotomy. If nothing else, listing of the additional surgical procedures would permit the reader to decide whether or not the nine subjects who had additional surgical procedures were representative of the entire sample.

I salute the authors for attempting to address this complex and timely issue, but ask that they consider reanalyzing their data in light of my concerns. In particular, I would welcome publication of the details of the distribution of scores in their larger data sets, including analysis of goodness-of-fit as outlined above. If the larger data sets also include excess numbers of subjects with the very lowest scores, but are otherwise indistinguishable from the expected uniform distribution, then further, in-depth review of the medical and school records of the most impaired subjects might reveal common attributes that either increased their joint probability of poor academic achievement and the need for surgical intervention in the first year of life or their susceptibility to any neurocognitive impairment related to anesthetic or surgical exposure. Given the potentially incendiary impact of research on this topic, it is critical that results be carefully analyzed and thoughtfully presented. I urge the authors to consider repeating their experiment with the more specific Null and Alternative Hypotheses presented above to see whether their results are reproducible. Until then, their results only support the conclusion that the distribution of academic achievement scores in otherwise neurologically normal children with a single exposure to anesthesia in the first year of life for minor, peripheral surgery is completely consistent with that seen in the population at large. I find this conclusion reassuring.

Joel B. Gunter, M.D., Children's Hospital Medical Center, University of Cincinnati, Cincinnati, Ohio. joel.gunter@cchmc.org

Table 1. Frequency of Academic Achievement Score by Decile

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Observed frequency</td>
<td>9</td>
<td>5</td>
<td>5</td>
<td>6</td>
<td>5</td>
<td>2</td>
<td>10</td>
<td>5</td>
<td>7</td>
<td>4</td>
</tr>
</tbody>
</table>

Table 2. Frequency of Outcomes for 60 Rolls of a 12-sided Die

<table>
<thead>
<tr>
<th>Number rolled</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observed frequency</td>
<td>4</td>
<td>7</td>
<td>5</td>
<td>10</td>
<td>3</td>
<td>4</td>
<td>6</td>
<td>4</td>
<td>2</td>
<td>6</td>
<td>5</td>
<td>4</td>
</tr>
</tbody>
</table>

In Reply:

We would like to thank Dr. Gunter for his interest in our article1 and for his thoughts regarding alternative analytical approaches to the data. In fact, we presented our data in a manner that would permit such a reanalysis and we are pleased to see him taking the effort. We do not disagree with what we think is Gunter’s overarching theme in his letter; our results do not provide general support for the belief that anesthesia and surgery during infancy produce detrimental effects on academic performance during childhood. However, we have concerns regarding the possibility that there may be a small group of infants in whom there may be an association between anesthesia and surgery and subsequent performance—although, as we clearly stated, it is impossible to draw any cause-and-effect conclusions. This was the reason for one of our two predefined questions: do a disproportionate number of children who had anesthesia and surgery during infancy subsequently have very low achievement test scores (for instance, because anesthesia and surgery might be associated with large adverse effects on a small percentage of patients)? Our second question was closer to that proposed by Gunter: do children who had anesthesia and surgery during infancy subsequently have lower mean test scores (for instance, because anesthesia and surgery might be associated with mild adverse effects on a large percentage of patients)? Again, both questions were formulated a priori, and we have no idea why Gunter suspects that the first question was formulated on a post hoc basis. We cannot find anything in our article that would give this impression. Neither question was explicitly stated as a hypothesis in the Introduction, but we dealt with the two questions in parallel in the Abstract, Statistical Analysis subsection, and Results section.

As we discussed at some length in a recent review article concerning assessment of cognition after anesthesia and

Reference


(Accepted for publication March 11, 2013.)