IQ at Age 4 in Relation to Maternal Alcohol Use and Smoking During Pregnancy

Ann Pytkowicz Streissguth  
Department of Psychiatry and Behavioral Sciences,  
School of Medicine  
Child Development and Mental Retardation Center  
University of Washington

Helen M. Barr  
Department of Psychiatry and Behavioral Sciences,  
School of Medicine  
University of Washington

Paul D. Sampson  
Department of Statistics  
University of Washington

Donald C. Martin  
Department of Biostatistics, School of Public Health  
University of Washington

Examined the relationship of prenatal alcohol exposure to the IQ of children at age 4 in a longitudinal prospective, population-based study. Multiple-regression analyses on data from 421 children indicated that use of more than 1.5 oz (44 ml, or approximately 3 drinks) of alcohol per day during pregnancy was significantly related to an average IQ decrement of almost 5 IQ points (% of a standard deviation; p = .008), even after adjustment for maternal and paternal education, race, prenatal nutrition, aspirin and antibiotics, child's sex and birth order, mother-child interaction, and preschool attendance. We caution against using these data to describe "safe" drinking levels, because other outcomes, more sensitive than IQ tests, show significant alcohol effects at lower drinking levels.

In this study, we examine the early childhood consequences of prenatal exposure to alcohol and tobacco. Theoretically, this kind of research is conceptualized as human behavioral teratology. Teratogens are substances capable of producing death, malformations, growth deficiency, or behavioral aberrations as a result of prenatal exposure (Wilson, 1977a), and a behavioral teratogen is a substance that produces behavioral deviations as a result of prenatal exposure (Hutchings, 1980; Riley & Vorhees, 1987; Vorhees & Butcher, 1982). In this article, we examine IQ scores obtained on 4-year-old children as the outcome of interest and relate these scores to other markers of prenatal insult, including birth weight and morphology.

Behavioral teratology is a new field (Hutchings, 1980; Riley & Vorhees, 1987; Vorhees & Butcher, 1982; Wilson, 1977b), most frequently explored through research on laboratory animals, where tight experimental controls can be used in manipulation of the genetic background, administration of the agent, rearing of the young, and assessment of behavior. Although studies of behavioral teratogens in human beings are ultimately of more practical interest, the many complex factors that influence human behavior can never be completely controlled in the study design. Instead, these factors must be measured carefully and adjusted for statistically as much as possible.

The teratogenic properties of alcohol have been well documented through clinical studies of children with Fetal Alcohol Syndrome (FAS; e.g., Clarren & Smith, 1978; Streissguth, Landesman-Dwyer, Martin, & Smith, 1980) as well as over a thousand studies on laboratory animals (e.g., Abel, Randall, & Riley, 1983; West, 1986). Prenatally, alcohol crosses the placenta and can alter brain development throughout gestation. Alcohol can interfere with cell proliferation in the embryo, disorganize cell migration and development, and interfere with neurotransmitter production in the developing central nervous system (CNS; Porter & O'Connor, 1984; West, 1986). Although children with FAS show varying levels of mental retardation, their mothers are usually alcoholic or severely abusing alcohol. Questions about the childhood consequences of more moderate prenatal alcohol exposure are of considerable public health importance, and few controlled studies exist (see Streissguth, 1986a, for a review).
The offspring effects of maternal smoking have been examined in many epidemiological studies (for reviews, see Landesman-Dwyer & Emanuel, 1979; Streissguth, 1986b). Although several studies have shown long-term childhood effects of prenatal smoking on child IQ and achievement (Butler & Goldstein, 1973; Dunn, McBurney, Ingram, & Hunter, 1977; Fogelman, 1980; Goldstein, 1977), all of them were begun prior to the recognition of the teratogenicity of alcohol. Some of these studies controlled for socioeconomic variables, but none measured maternal drinking—although the two behaviors are often correlated.

In the present study, in which we obtained careful prenatal assessments of both drinking and smoking, it is possible to consider the relative contribution of each in its effect on child IQ at 4 years of age. Furthermore, it is also possible to consider the relationship of early markers of an adverse intrauterine environment (namely, low birth weight and dysmorphology) in terms of their predictive validity for later CNS effects as assessed by IQ.

Method

In the study design (Streissguth, Martin, Martin, & Barr, 1981), 1,529 pregnant women were interviewed in their homes during their fifth month of pregnancy in order to select a small follow-up cohort (approximately 500 children, stratified for alcohol and smoking), which has been examined at several ages.

Sample Selection and Characteristics

All women receiving prenatal care by the fifth month of pregnancy at two study hospitals over a 1-year period in 1974-1975 were invited to participate in the "Pregnancy and Health Study." The participation rate was 85%, and 1,529 pregnant women were given a structured confidential interview in their homes by highly trained female interviewers who were not part of the hospital staff. The interview covered pregnancy history, and social/demographic/life-style characteristics as well as beverage consumption, smoking, drug use, and diet (Streissguth et al., 1981).

The mothers were primarily White (88%), married (87%), and middle class. Their average age was 26 and they had an average of 13.5 years of education. The characteristics of the follow-up cohort (Table 1) closely resemble those of the large sample from which it was drawn and of the Seattle, Washington community. Two-thirds of the mothers received care at a health maintenance organization.

The follow-up cohort was selected at the time of birth using a computerized formula that prioritized infants of the interviewed women in order to oversample for heavier drinkers and smokers (Streissguth et al., 1981). Status of the infant at delivery was not a consideration in the selection of the follow-up cohort; which consisted of approximately 500 families in which maternal alcohol use during pregnancy ranged from none to heavy. Unlike some lower-class samples, heavier drinking mothers in this study had more education, fewer children, and were slightly older than light drinkers and abstainers.

Independent Predictor Variables

Maternal alcohol use was determined using self-report from a standard quantity-frequency-variability interview on wine, beer, and liquor consumption (Jessor, Gravens, Hanson, & Jessor, 1968). Average ounces of absolute alcohol consumed per day (AA), calculated according to Jessor et al., ranged from 0 to 25.8, with a mean of 0.63 (i.e., a little over one drink per day) and a median of 0.17, for the period before the recognition of pregnancy, and from 0 to 8.55, with a mean of 0.27 (i.e., one half of a drink per day) and a median of 0.06, during the fifth month of pregnancy. Approximately 28% of the follow-up sample were non-drinkers at the start of pregnancy; 11% reported consuming more than 1.5 oz (44 ml) per day (i.e., about three drinks, or more). Three mothers reporting major alcohol problems and one mother with a very discrepant AA score (AA > 10) were considered separately.

Other prenatal exposures were also obtained prenatally using self-report. Nicotine use was calculated using the number of cigarettes that were reported as smoked per day multiplied by the milligrams of nicotine per brand smoked. Nicotine scores ranged from 0 to 77, with the 131 smokers averaging 17.3 mg (i.e., over three-fourths of a pack of moderate nicotine cigarettes per day). Sixty-nine percent of the sample were nonsmokers, and 65% of the heavier drinkers did not smoke. Caffeine use was determined using self-reported use of coffee, tea, cola beverages, and chocolate, which were converted to a mg per day measurement (Barr, Streissguth, Martin, & Horst, 1981). Maternal caffeine use ranged from 0 to 2.053 mg per day, with a mean of 257 mg (i.e., the equivalent of about 3.5 cups of coffee). Only three mothers abstained from caffeine. Aspirin and acetaminophen (the most frequently reported drugs after alcohol) as well as other drugs (such as antibiotic use, which was studied to control for possible illness and infection during pregnancy) were coded as times used per month. Nine of the mothers reported use of "street drugs" (heroin, methadone, and others), 72 reported marijuana use, and 7 had used drugs considered to be "possibly teratogenic" (hydantoins, hydrochlorothiazide, chloridiazepoxide hydrochloride [Librium], and meprobamate). Maternal nutrition was quantified by summarizing the number of basic food groups for which a mother's dietary intake, according to her 24-hour recall during pregnancy, was adequate according to recommended allowances.

Genetic and demographic characteristics included race, age, education, and parity of the mother, education of the father, and sex and age of the child. Postnatal events evaluated included illnesses, accidents, hospitalizations, medical problems, minor illness on the day of the exam, preschool attendance, and major life changes in household (at 8 months, 18 months, and 4 years). Mother-child interaction was assessed in nine global assessments at 8 and 18 months of age, each rated on a 5-point scale (Streissguth, Barr, Darby, & Ervin, 1987).

The 4-Year Examination

Sample maintenance was achieved through an elaborate outreach system that kept the cohort intact over the 5-year interval. Only 13.7% of the candidates for the 4-year follow up could not be seen at 4 years. There was no difference in the rates of subject loss according to maternal alcohol use.

Testing procedures involved an examination at the laboratory when each child was 4 years and 3 months old (92% of the children were examined within 2 weeks of this age). A 2-hour test battery was administered that began with the Wechsler Preschool and Primary Scales of Intelligence (WPPSI), a standardized, individually administered, age-appropriate IQ scale. Pretesting revealed that the WPPSI at this age was free of floor and ceiling effects. (A companion article now in preparation will describe data from the fine and gross-motor battery.) Dysmorphology examinations were conducted on subsets of this cohort at birth (Hanson, Streissguth, & Smith, 1978) and at 4 years (Graham, Hanson, Darby, Barr, & Streissguth, 1988).

Exam conditions were highly standardized. Testing rooms, time of day for testing, and alcohol exposure of children were all counterbalanced between two female psychometrists who had been trained to a high level of reliability, which was maintained through monthly reliability examinations. The psychometrists were "blind" both before and after the examinations, with respect to alcohol exposure, family history, and children's previous test scores. In order to reduce bias, all scheduling and contacts with parents were handled by an outreach worker who knew the families, provided transportation as needed, and reimbursed families for their expenses. A parental questionnaire provided information.
Demographic Characteristics of 421 Mothers in the 4-Year Follow-Up Cohort

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>n</th>
<th>%</th>
<th>Characteristic</th>
<th>n</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Race</td>
<td></td>
<td></td>
<td>Education</td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>372</td>
<td>88</td>
<td>Graduate school</td>
<td>31</td>
<td>7</td>
</tr>
<tr>
<td>Black</td>
<td>22</td>
<td>5</td>
<td>College graduate</td>
<td>80</td>
<td>19</td>
</tr>
<tr>
<td>American Indian</td>
<td>3</td>
<td>1</td>
<td>Some college</td>
<td>130</td>
<td>31</td>
</tr>
<tr>
<td>Other</td>
<td>24</td>
<td>6</td>
<td>High school graduate</td>
<td>129</td>
<td>31</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Some high school</td>
<td>36</td>
<td>9</td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td></td>
<td>Junior high school</td>
<td>15</td>
<td>4</td>
</tr>
<tr>
<td>14-19</td>
<td>37</td>
<td>9</td>
<td>Less than 7th grade</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>20-24</td>
<td>108</td>
<td>26</td>
<td>Socioeconomic status</td>
<td></td>
<td></td>
</tr>
<tr>
<td>25-29</td>
<td>187</td>
<td>44</td>
<td>Upper</td>
<td>37</td>
<td>9</td>
</tr>
<tr>
<td>30-34</td>
<td>12</td>
<td>3</td>
<td>Upper middle</td>
<td>76</td>
<td>18</td>
</tr>
<tr>
<td>35-39</td>
<td>1</td>
<td>&lt;1</td>
<td>Middle</td>
<td>111</td>
<td>26</td>
</tr>
<tr>
<td>40 and over</td>
<td>179</td>
<td>43</td>
<td>Lower middle</td>
<td>155</td>
<td>37</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Lower</td>
<td>42</td>
<td>10</td>
</tr>
<tr>
<td>Child's birth order</td>
<td>242</td>
<td>57</td>
<td>Marital status</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Firstborn</td>
<td></td>
<td></td>
<td>Single</td>
<td>34</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Married</td>
<td>368</td>
<td>87</td>
</tr>
<tr>
<td>Other</td>
<td></td>
<td></td>
<td>Separated</td>
<td>10</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Divorced</td>
<td>8</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Widowed</td>
<td>1</td>
<td>&lt;1</td>
</tr>
</tbody>
</table>

Note. Age, education, and marital status pertain to the fifth month of pregnancy. Socioeconomic status was calculated by using the education of the mother and the occupation of the head of household, according to a modification of Hollingshead’s two-factor system.

regarding the family environment (e.g., divorces, separations, births, deaths, child care, preschool experience) and the child’s health history (illnesses, accidents, hospitalizations, fevers, medications, sensory problems, general development, etc.). Valid IQ scores were obtained for 452 of the 457 children examined. (Five subjects were deleted after “blind” examination of the records by Ann Streissguth or Betty Darby: 1 deaf child, 3 whose primary language was not English, and 1 who refused to cooperate.) The analyses reported here involved 421 children, that is the total number of those tested at 4 years for whom mother–child interaction ratings were also available.

Data Analysis

The primary outcome variable in this report was the WRAT IQ score, which ranged from 69 to 151, with a mean of 110.5 and a standard deviation of 14.4 for this cohort. These scores are congruent with the higher educational level of the mothers, other IQ studies in the Seattle, Washington area (Bec et al., 1982), and earlier scores on mental development tests from the same sample (Streissguth, Barr, Martin, & Herman, 1980). Fourteen children (3.3%) had IQ scores below 85.

Our general strategy for data analysis (Streissguth, Sampson, Barr, Clarke, & Martin, 1986) involved four phases. Phase 1 consisted of thorough descriptive analysis of three classes of variables: (a) outcome variables (here, IQ scores); (b) exposure variables (alcohol, nicotine, caffeine, etc.); and (c) covariates or potentially confounding variables. Appropriate transformations—such as the log for highly skewed distributions of exposure variables (Breslow & Day, 1980)—are applied so that findings are not unduly influenced by a few extreme scores. Table 2 presents the simple correlations among alcohol, nicotine, caffeine, and IQ and other potential covariates.

In Phase 2 we developed a “base model,” beginning with high priority covariates derived from literature reviews and past experience. Exposure variables (other than alcohol) were introduced to the model at this stage. The nominal significance of the contribution of each of the independent variables in the model was tested using partial t tests (Weisberg, 1985), which permitted adjustment for all the other variables in the regression model. Variables that were not significantly related to IQ according to the t test were deleted from the base model in the interest of building a parsimonious model for prediction of IQ.

Missing data are a problem often encountered in longitudinal studies. Although the present study had no missing data on any of the primary predictor variables (alcohol and nicotine) or the primary outcome variable (child IQ at 4 years), we did have missing data on two important covariates that were assessed at ages other than the 4-year follow up. The biological father’s education (missing on 36 subjects) was well predicted by mother’s education, age, and other variables, so the missing scores were predicted from the regression equation to avoid reducing the sample size. (For further discussion of the missing data problem in regression see Weisberg, 1985, Chap. 10.) Mother–child interaction scores (missing on 31 children) could not be successfully predicted, however, so the sample size for the final model was reduced to 421 to accommodate this important predictor variable.

In Phase 3 we introduced alcohol variables to the base model and considered a variety of alcohol scores and parameterizations. We used current diagnostic methods in the regression analysis literature to help determine whether nonlinear transformations of the alcohol scores seemed warranted. In particular, partial residual plots and the “ACE” algorithm for estimating optimal nonlinear transformations in regression analysis (Breiman & Friedman, 1985) suggested that mean IQ decrements begin to become manifest in the multiple regression models near AA scores of 1.5 oz (44 ml) per day. In fact, two alternative models indicated a nominal significant alcohol relationship with IQ. One describes a linear effect on mean IQ of log AA scores over 1.5. The other is a step function specifying a discrete mean IQ decrement for alcohol scores over 1.5 oz per day. We report only the latter model, because its estimate of the alcohol effect was the most stable when examined under the various conditions described in Table 4 and discussed below.

Phase 4 involved testing for the effects of other possible confounding variables and using diagnostic techniques (Weisberg, 1985) to examine the influence of extreme cases. Our analyses considered approximately 150 potential covariates or confounding variables (Streissguth et al., 1986). Table 3 presents the final multiple regression model including the 12 predictor variables for AQ. Table 4 lists 19 additional variables that were considered as potential confounding variables. We examined
Significance values are two-tailed.

Full-scale IQ: \( R = 0.62 \);
\( SE \) = 11.45; \( r(12, 408) = 21.43, p = .000 \). Partial \( t \) tests adjust each variable for the other 11 variables. \( SE \) denotes standard error of the coefficient.

\( * \) Significance values are two-tailed.

### Table 2

**Simple Product-Moment Correlations Between Selected Predictor Variables and Alcohol, Nicotine, Caffeine, and IQ**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Alcohol</th>
<th>Nicotine</th>
<th>Caffeine</th>
<th>IQ</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alcohol, prenatal</td>
<td>1.00</td>
<td>.12</td>
<td>.22</td>
<td>-0.07</td>
</tr>
<tr>
<td>Nicotine, prenatal</td>
<td>.12</td>
<td>1.00</td>
<td>.30</td>
<td>-0.18</td>
</tr>
<tr>
<td>Caffeine, prenatal</td>
<td>.22</td>
<td>.30</td>
<td>1.00</td>
<td>-0.04</td>
</tr>
<tr>
<td>Aspirin, prenatal</td>
<td>-.02</td>
<td>.06</td>
<td>.01</td>
<td>-0.18</td>
</tr>
<tr>
<td>Antibiotics, prenatal</td>
<td>-.02</td>
<td>.05</td>
<td>.01</td>
<td>-0.11</td>
</tr>
<tr>
<td>Nutrition, prenatal</td>
<td>.01</td>
<td>.03</td>
<td>.01</td>
<td>.22</td>
</tr>
<tr>
<td>Maternal education (years)</td>
<td>.08</td>
<td>-.27</td>
<td>-.06</td>
<td>.46</td>
</tr>
<tr>
<td>Paternal education (years)</td>
<td>.04</td>
<td>-.28</td>
<td>-.04</td>
<td>.43</td>
</tr>
<tr>
<td>Maternal age (years)</td>
<td>.08</td>
<td>-.18</td>
<td>.07</td>
<td>.24</td>
</tr>
<tr>
<td>Race (Black/other)</td>
<td>.12</td>
<td>.06</td>
<td>.06</td>
<td>-.27</td>
</tr>
<tr>
<td>Birth order (first/other)</td>
<td>.09</td>
<td>-.02</td>
<td>.03</td>
<td>.13</td>
</tr>
<tr>
<td>Child's sex (female/male)</td>
<td>.06</td>
<td>-.05</td>
<td>-.01</td>
<td>-.10</td>
</tr>
<tr>
<td>Mother–child interaction</td>
<td>.01</td>
<td>.06</td>
<td>-.01</td>
<td>-.21</td>
</tr>
<tr>
<td>Preschool</td>
<td>.10</td>
<td>-.17</td>
<td>.10</td>
<td>.23</td>
</tr>
</tbody>
</table>

Note. \( N = 421 \). The threshold for the two-tailed 5% level of significance is .095. For the 1% level of significance the threshold is .125. Alcohol use is a binary indicator of average use of more than 1.5 oz of absolute alcohol per day \( (n = 47; 11\% \text{ of } 421) \); Nicotine and Caffeine are \( \ln (\text{mg} + 1) \); Aspirin use is a binary indicator of use several times per week or more \( (n = 54; 1\% \text{ of } 421) \); Antibiotic use is binary for any use during pregnancy \( (n = 52; 12\% \text{ of } 421) \); Nutrition is a count from 0 to 4 of the number of food groups with adequate intake; Mother–child interaction is a scale from 0 to 13 (higher scores indicate deviant interaction); Preschool: \( \ln (\text{months} + 1) \).

### Results

The final regression model (Table 3) indicates that maternal alcohol was significantly related to child IQ at 4 years, even after adjustment for a wide variety of other potentially confounding variables, including the 11 other predictor variables in Table 3 and the 19 other potential covariates listed in Table 4. The partial \( t \) test for alcohol had a significance value of 0.008. The magnitude of the alcohol effect on child IQ at 4 years was estimated as 4.80 IQ points \( (SE = 1.81) \) in the binary threshold model, with a cut off at 1.5 oz of absolute alcohol per day \( (AA) \) prior to pregnancy recognition (Table 3).

AA scores from both pregnancy periods were significantly related to child IQ at 4 years, but the relationships from the “prior to pregnancy recognition” scores were stronger and more stable across different models. This phenomenon was first reported (Little, 1977) in a study of prenatal alcohol effects on infant birth weight and has been replicated in many outcomes assessed in the present series of studies (Streissguth et al., 1981).

Nicotine use was not significantly associated with child IQ in the Table 3 regression model after adjustment was made for alcohol use and the other variables. (Thus, nicotine appears in Table 4 rather than in Table 3.) Findings for number of cigarettes smoked per day were analogous to those for nicotine. Other exposure variables that did not significantly affect the al-

### Table 3

**Multiple Regression Analysis of 4-Year IQ on Maternal Alcohol Use and 11 Other Significant Predictors**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>SE</th>
<th>Standardized coefficient</th>
<th>( t(408) )</th>
<th>( p^* )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alcohol, prenatal</td>
<td>-4.80</td>
<td>1.81</td>
<td>-.11</td>
<td>2.65</td>
<td>.008</td>
</tr>
<tr>
<td>Aspirin, prenatal</td>
<td>-5.27</td>
<td>1.68</td>
<td>-.12</td>
<td>3.13</td>
<td>.002</td>
</tr>
<tr>
<td>Antibiotics, prenatal</td>
<td>-4.74</td>
<td>1.72</td>
<td>-.11</td>
<td>2.76</td>
<td>.006</td>
</tr>
<tr>
<td>Nutrition, prenatal</td>
<td>1.15</td>
<td>.53</td>
<td>.09</td>
<td>2.19</td>
<td>.029</td>
</tr>
<tr>
<td>Maternal education (years)</td>
<td>1.56</td>
<td>.32</td>
<td>.26</td>
<td>4.98</td>
<td>.000</td>
</tr>
<tr>
<td>Paternal education (years)</td>
<td>.82</td>
<td>.28</td>
<td>.16</td>
<td>2.96</td>
<td>.003</td>
</tr>
<tr>
<td>Mother–child interaction</td>
<td>-1.45</td>
<td>.33</td>
<td>-.18</td>
<td>4.46</td>
<td>.000</td>
</tr>
<tr>
<td>Child’s sex</td>
<td>-2.87</td>
<td>1.13</td>
<td>-.10</td>
<td>2.25</td>
<td>.011</td>
</tr>
<tr>
<td>Sex X Aspirin</td>
<td>8.83</td>
<td>3.37</td>
<td>.10</td>
<td>2.62</td>
<td>.009</td>
</tr>
<tr>
<td>Firstborn</td>
<td>3.90</td>
<td>1.15</td>
<td>.13</td>
<td>3.38</td>
<td>.001</td>
</tr>
<tr>
<td>Preschool</td>
<td>.96</td>
<td>.46</td>
<td>.09</td>
<td>2.09</td>
<td>.038</td>
</tr>
<tr>
<td>Race</td>
<td>-9.90</td>
<td>2.64</td>
<td>-.15</td>
<td>3.76</td>
<td>.000</td>
</tr>
</tbody>
</table>

Note. \( N = 421 \). Full-scale IQ: \( R = 0.62; \) \( SE \) regression = 11.45; \( F(12, 408) = 21.43, p = .000 \). Partial \( t \) tests adjust each variable for the other 11 variables. \( SE \) denotes standard error of the coefficient.

* Significance values are two-tailed.
PRENATAL ALCOHOL, SMOKING, AND AGE 4 IQ

Table 4
Other Possible Confounding Variables Having Essentially No Effect on the Magnitude or Significance of the Alcohol Findings in the Multiple Regression Analysis

<table>
<thead>
<tr>
<th>Variable</th>
<th>Effect of variable on child IQ (Standardized Coefficient)</th>
<th>Effect of alcohol use on child IQ</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coefficient</td>
<td>SE</td>
</tr>
<tr>
<td>Model from Table 3</td>
<td>---</td>
<td></td>
</tr>
<tr>
<td>Prenatal exposure</td>
<td>-0.01</td>
<td></td>
</tr>
<tr>
<td>Nicotine (131)</td>
<td>-0.06</td>
<td></td>
</tr>
<tr>
<td>Nicotine × Sex</td>
<td>-0.03</td>
<td></td>
</tr>
<tr>
<td>Caffeine (418)</td>
<td>0.03</td>
<td></td>
</tr>
<tr>
<td>Acetaminophen (183)</td>
<td>0.04</td>
<td></td>
</tr>
<tr>
<td>Diazepam (Valium) (31)</td>
<td>-0.01</td>
<td></td>
</tr>
<tr>
<td>Barbiturates (14)</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>Marijuana (72)</td>
<td>-0.01</td>
<td></td>
</tr>
<tr>
<td>Illicit drugs (9)</td>
<td>---</td>
<td></td>
</tr>
<tr>
<td>Questionable teratogenic drugs (7)</td>
<td>---</td>
<td></td>
</tr>
<tr>
<td>Major alcohol problem (3)</td>
<td>---</td>
<td></td>
</tr>
<tr>
<td>Alcohol &gt; 10 oz/day (1)</td>
<td>---</td>
<td></td>
</tr>
<tr>
<td>Postnatal environment:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hospitalizations (child) (84)</td>
<td>0.02</td>
<td></td>
</tr>
<tr>
<td>Accidents for which doctor called (160)</td>
<td>0.03</td>
<td></td>
</tr>
<tr>
<td>Illnesses, excluding colds (241)</td>
<td>0.05</td>
<td></td>
</tr>
<tr>
<td>Major life changes in household (354)</td>
<td>0.06</td>
<td></td>
</tr>
<tr>
<td>Medical problems being followed (99)</td>
<td>-0.07</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mother's age</td>
<td>0.02</td>
<td></td>
</tr>
<tr>
<td>Child's age at testing</td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td>Possibly ill on exam day (35)</td>
<td>0.00</td>
<td></td>
</tr>
</tbody>
</table>

Note. The numbers in parentheses refer to the number of subjects with positive scores for the category (from a total sample of 421). Variables were entered one at a time into the regression model described in Table 3 in order to avoid problems with multicollinearity.

* These were variables with such a small sample size that their effects were examined by running the regression model on Table 3, omitting subjects with positive scores for the given variable.

Cohab-by-IQ relationship (also listed in Table 4) included marijuana, other illicit drugs, other questionable teratogenic drugs, barbiturates, diazepam, and acetaminophen. As Table 3 indicates, all 12 of the variables included in the model were significantly related to child IQ even after adjustment for the other 11. Prenatal aspirin (Table 3) as a predictor of child IQ was consistent with the animal literature on aspirin as a behavioral teratogen (Streissguth, Treder, et al., 1987).

As the standardized coefficients from Table 3 indicate, the strongest predictors of child IQ at age 4 were maternal education, mother–infant interaction, paternal education, race, and birth order. The partial correlation coefficient for alcohol was -0.130, indicating that alcohol was not as strong a predictor of IQ as maternal education (with a partial correlation coefficient of 0.239) but about the same as parity (with a partial correlation coefficient of 0.165). Race was examined using a binary term for Blacks, but the relatively small number of Blacks in this sample did not permit us to model them separately or to test effectively for interactions between relevant covariates and the racial indicator. Preschool experience and sex of the child were also significant predictors of child IQ, but they were relatively less important in this sample after adjustment for demographic and significant exposure variables. Birth order was examined using a binary term for firstborns, as suggested by regression diagnostic tools (partial residual plots and the ACE algorithm).

Using a model similar to that described in Tables 3 and 4, gestational alcohol exposure (a binary threshold model with a threshold at 1.5 oz) was significantly and negatively related to Performance IQ (p = 0.002; coefficient = -5.44). Verbal IQ was also negatively related to alcohol, but the relationship was not as strong (p = 0.086; coefficient = -3.46). Neither Verbal IQ nor Performance IQ was significantly related to smoking. Those WPPSI subtest scores that were most related to the binary alcohol indicator (with a nominal significance value of less than 0.05) were Animal House, Mazes, Similarities, Vocabulary, and Picture Completion. Certain other subtests were more strongly related to other alcohol scores.

To avoid the problem of repeated tests with many different alcohol scores, we also applied a method of latent variable analysis (using partial least squares methods; Sampson, Streissguth, Barr, & Bookstein, in press) to incorporate simultaneously many measures of alcohol exposure (including the AA score) in the analysis of all the WPPSI subtest scores. This analysis confirmed the strength of the alcohol effect on the Performance Scale. Alcohol and IQ latent variables were computed using the cross-correlations between multiple alcohol measures and the IQ subtests (after adjusting for the covariates in Table 3). The subtests most highly correlated with alcohol exposure were from the Performance scale (Block Design, Picture Completion, Mazes) and Mathematics, from the Verbal scale.

In order to evaluate the clinical significance of the IQ findings, Figure 1 is presented to show the estimated effect of a 5-point decrement in IQ on a child of “average background” in our population. After adjusting for all other predictors of IQ in the Table 3 model, maternal consumption of over 3 drinks per day (1.5 oz) in early pregnancy was estimated to triple...
that were 10.5 points lower at 4 years than the 23 other neonates

dicators of later IQ was studied using data from a subset of 87
birth weight, even at the low end of the distribution, failed to
the lowest 10% of our birth-weight distribution) revealed that
action terms was significant. Similar analyses involving just
examined by adding terms for the multiplicative interaction of
weight and whether low-birth-weight offspring of smokers or
whether the alcohol effects on child IQ were mediated by birth
results. These latter effects were
mately 2% of our population, as well as a binary indicator for
The 10 infants classified as FAE at birth had IQ scores
son et al. examined who had exposure histories above 1.5 oz
model, we examined the IQ scores of the 33 children that Han-
ted 30 important independent variables
developmental toxicology studies share the same problem of mul-
tiple predictors. The 30 important independent variables
exposed at this level who had normal morphology and growth
Discussion
This study shows that maternal alcohol use during early preg-
nancy is significantly related to offspring IQ scores at 4 years,
even when a large number of other potentially confounding
variables are adjusted for in the statistical analyses. The step
function model at 1.5 oz per day provides a more robust test
than the linear or other nonlinear models (i.e., it is less sensitive
to the influence of extreme alcohol cases). This finding should
not, however, be interpreted as indicating a biological threshold
with respect to the effects of alcohol on the fetus, because other
parameterizations of alcohol also showed significant relations-
ships. We also note that other behavioral outcomes from this
study, measured on precise laboratory tasks, support linear
rather than "threshold" relationships to prenatal alcohol expo-
sure (Streissguth, Barr, & Martin, 1984).

We now address three questions of critical importance to be-
havioral teratology studies using human samples: (a) How ade-
quate is the measurement of "dose"? (b) How well do the find-
ings hold up when the possible influences of other covariates
are considered? and (c) What is the practical significance of the
findings and of the interpretation.

No human studies can measure "exposure" per se. The short
half-life of alcohol and the infeasibility of obtaining consecutive
blood samples from pregnant women make self-report the
method of choice for assessing exposure. Therefore, we must
examine what additional information is available regarding the
veracity of the alcohol reports. The test–retest reliability corre-
lation of the AA score was .90 over a 1-week interval (Streiss-
guth, Martin, & Buffington, 1976), the same as the reliability of
the caffeine score (Barr et al., 1981). Furthermore, the alcohol
effects held up in the present study, even after the deletion of the
three women from the present sample who reported alcohol-
related problems with health, job, the law, or marriage. The
findings also held up when one woman, who reported alcohol
use far in excess of the other mothers, was deleted from the
sample. In addition, the women were all interviewed prospectively
in midpregnancy; no woman who was not in prenatal care by
the fifth month of pregnancy was included in the interview sam-
ple. We went to great length to ensure a confidential interview
atmosphere (Streissguth et al., 1981). The women were inter-
viewed during 1974–1975, at a time when no information was
available about moderate drinking and pregnancy outcome. We
believe that these interview factors have resulted in as accurate
self-reports as possible and that the statistical procedures have
provided as conservative a test as possible against the possible
influence of outliers.

Several points are relevant with respect to the question of ade-
quate control of potentially confounding variables in the pre-
diction of IQ. First, all observational behavioral teratology/de-
velopmental toxicology studies share the same problem of mul-
tiple predictors. The 30 important independent variables
described in Tables 3 and 4 (or the approximately 150 consid-
ered in this study) cannot be meaningfully examined in a single
regression model. As with other researchers coping with this
problem (i.e., Bellinger, Needleman, Leviton, & Waternaux,
1984; Schroeder & Hawk, 1987), we have demonstrated a sys-
tematic procedure for examining multiple predictors of a given outcome such as IQ (Streissguth et al., 1986). When the research questions involve multiple outcomes (e.g., a comparison of all the WPPSI subtests) as well as multiple predictors (e.g., the contrast between many types of alcohol predictors), we believe that other procedures that are based on latent variable methods are preferable (Sampson et al., in press).

In the IQ analyses presented here, we have taken a conservative approach in that the statistical tests were two-tailed throughout. On the other hand, we cannot emphasize literal interpretation of significance values that do not take into account all the models considered en route to a “final” model. In order to ensure that the results we report as alcohol effects were not attributable to other causes, we have carefully examined 30 potentially confounding variables described in Tables 3 and 4 from an initial listing of approximately 150 in the database. Some variables, such as nicotine, were not significantly associated with IQ once other covariates were considered, nor did they affect the alcohol/IQ prediction—these were removed from the Table 3 final model. This study confirmed that IQ scores have many significant covariates (c.f., Bee et al., 1982; Broman, Nichols, & Kennedy, 1975). The strength of these other predictors was, however, of marginal interest to the goals of this study, and they primarily demonstrated that the important covariates have been taken into account before claiming a teratogenic effect. Human behavioral teratology studies, particularly those examining low-dose effects as manifest in nonclinical populations, do not hypothesize that the teratogen under examination is the best predictor of the outcome, only that the teratogen continues to predict the outcome after statistical adjustment for the other significant predictors.

One question that must be addressed when using IQ scores as teratogenic outcomes is whether the familial component of IQ has been adequately adjusted for. Although some researchers have argued that parental IQ should be examined directly (e.g., Smith, Delves, Lansdown, Clayton, & Ghahm, 1983), we note that the simple correlations reported between parental IQ and child IQ (.38 to .46; Bouchard & McGue, 1981; Smith et al., 1983) are comparable with the correlations that we present in Table 2 between parental education and child IQ (.46 and .43, for mothers and fathers, respectively). Considering the disadvantages of using parental IQ tests in complex longitudinal studies (e.g., cost, risk of subject loss because of refusals, poor validity of the abbreviated or specialized IQ tests often used, etc.), we conclude that our parental education scores are satisfactory measures of that aspect of child IQ that might otherwise be predicted by parent IQ. Note also that we have measured mother–child interactions using direct observation in this study and that an earlier examination of a subset of these families using Caldwell’s HOME scale did not reveal any significant alcohol-related differences in the level of stimulation at home or in the mother–child interaction (Ragozin, Landesman-Dwyer, & Streissguth, 1978). The basically well-educated, middle-class mothers who were drinking during pregnancy when the present study was carried out were not a deviant group, because this study predated warnings against drinking during pregnancy.

The absence of smoking effects on 4-year IQ (once alcohol and other covariates were adjusted for) is of considerable interest in light of earlier studies (Butler & Goldstein, 1973; Dunn, McBurney, Ingram, & Hunter, 1977; Fogelman, 1980; Goldstein, 1977) that reported smoking effects on cognitive outcomes. Although not all early studies found positive associations between smoking and later cognitive outcomes (see Streissguth, 1986b, for a review), these early studies usually failed to examine the potentially confounding effects of alcohol that are often highly correlated with cigarette use. An observational study of childhood attention (Landesman-Dwyer, Ragozin, & Little, 1981), which did not covary for both alcohol and smoking, found alcohol but not smoking effects. The present population was particularly well suited to sorting out alcohol and smoking effects, because both drinking and smoking were stratified in the cohort selection. Thus, 65% of our heaviest drinkers did not smoke, and 28% of our heaviest smokers did not drink or were very infrequent drinkers. Table 2 indicates, smokers and drinkers were quite different demographically. The smokers were primarily young and poorly educated (accounting for the simple correlation of -.18 between nicotine and child IQ presented in Table 2). The older, well-educated mothers had apparently stopped smoking during pregnancy or had heeded the warnings about not smoking at all.

Alcohol use, on the other hand, was reported by 80% of the mothers and was widely distributed across social class and educational levels. In fact, the educational group with the highest number of heavier drinkers was the most highly educated group of professional mothers. This factor contributes to the low negative simple correlation (see Table 2) between prenatal alcohol and child IQ (−.07). In conclusion, we believe that at least two factors may explain the divergence of the present results from past smoking/IQ studies: (a) It is quite possible that at the time that most of the smoking studies were carried out (in the 1960s), smoking during pregnancy was more widely distributed across demographic categories than is now the case and (b) the early studies all failed to examine the possible confounding effects of alcohol and may have attributed to cigarettes those IQ decrements actually arising from prenatal alcohol exposure.

The present negative findings with respect to smoking and IQ do not contradict the ample epidemiologic literature (Landesman-Dwyer & Emanuel, 1979) on the adverse effects of maternal smoking on the health of newborns. Without question, smokers’ offspring are at higher risk for miscarriage, stillbirth, prematurity, and low birth weight. Even in the present study, which failed to find a relationship between smoking and IQ, smoking had a much stronger effect on birth size than did drinking (Barr; Streissguth, Martin, & Herman, 1984). However, birth weight alone did not predict 4-year IQ, nor did the interaction of birth weight with either alcohol or nicotine. Thus, low birth weight was not a marker for those offspring of drinking or smoking mothers who would later show IQ decrements. Neonatal dysmorphology in combination with growth decrements (called Fetal Alcohol Effects) did predict those offspring of heavily drinking mothers who would have the most severe IQ decrements at 4 years of age.

The presence of earlier alcohol-related effects strengthens the argument that the IQ decrements observed at 4 years are related biologically to gestational alcohol exposure and are not just artifacts of differences in the postnatal environment. Previous reports from this study have shown prenatal alcohol effects on heart rate, Apgar scores, and respiratory distress at birth (Streissguth, Barr, & Martin, 1982) as well as a variety of neonatal manifestations of early CNS dysfunction including the fol-
lowing: poorer habituation, more low-level arousal, and increased lability of states (Streissguth, Barr, & Martin, 1983); increased tremors, head turns to the left, hand-to-mouth behavior, and low levels of arousal (Landesman-Dwyer, Keller, & Streissguth, 1978); and decreased sucking pressure (Stock, Streissguth, & Martin, 1985). Small decrements in mental and motor development at the age of 8 months were also significantly related to gestational alcohol exposure (Streissguth, Barr, et al., 1980) as well as attentional deficits and longer reaction time as assessed using a laboratory vigilance test at 4 years of age (Streissguth, Martin, et al., 1984). Smoking, on the other hand, has not shown a consistent association with postnatal behavioral outcomes despite the strong relationship between smoking and birth size.

This study indicates that prenatal alcohol exposure is significantly related to child IQ at 4 years of age, in a relatively healthy, generally middle-class sample. Self-reported consumption of over three drinks a day on the average was associated with an average IQ decrement of almost 5 IQ points, after adjustment for a wide variety of other factors that also predict child IQ. This decrement represents an estimated tripling of the risk of subnormal intelligence (i.e., IQ < 85) for a child of "average background" in our sample. For these 4-year-olds it is the performance rather than verbal aspects of intelligence that are most strongly correlated with prenatal alcohol exposure.

Epidemiologic, population-based studies such as this one, which have the capability of adjusting for a wide variety of confounding variables, are of importance in understanding the general relationships between exposures (in this case, prenatal alcohol) and child outcomes (in this case, general level of intellectual functioning). Several cautions are in order, however:

1. These data should not be used to predict intellectual deficits in the individual case where many other factors (both genetic and environmental) are associated with child IQ scores.

2. The statistical models referred to here as "threshold" models should not be regarded as "biological" thresholds, because other outcomes from this study have shown strong linear effects of prenatal alcohol exposure.

3. Correlational studies cannot be used in isolation to infer causation; therefore, the findings from this study should be considered in relation to the extensive literature on the teratogenic effects of alcohol in laboratory animals (e.g., Abel et al., 1983; Streissguth, Landesman-Dwyer, et al., 1980) and other studies examining alcohol effects on offspring in different populations (Harlap & Shiono, 1980; Kline, ShROUT, Stein, Susser, & Warner, 1980; Oullette, Rossett, Rosman, & Weiner, 1977; Sokol, Miller, & Reed, 1980; Streissguth, Landesman-Dwyer, et al., 1980).

4. The alcohol-use scores and thresholds described in this study do not directly measure actual fetal exposure. Research with inbred mouse strains (Chernoff, 1980) suggests that maternal and fetal metabolism of alcohol (not assessed in this study) have a further role in ameliorating or exacerbating the effects of alcohol in the individual case.

Timing of exposure to alcohol and variation in the duration and peaks of alcohol exposure may also be important case by case. Because different outcomes are associated with different levels and patterns of prenatal alcohol exposure, and because the children in this study were only 4 years old at the time of examination, we urge that caution be used in characterizing any of the exposure levels described in this paper as "safe.”

References


Received December 29, 1987
Revision received March 24, 1988
Accepted May 24, 1988.