Benchmark Dose Analysis of Developmental Toxicity in Rats Exposed to Boric Acid

BRUCE C. ALLEN,* PHILIP L. STRONG,† CATHERINE J. PRICE,‡ SUSAN A. HUBBARD,§ AND GEORGE P. DASTON


Received April 24, 1995; accepted March 27, 1996


Developmental toxicity risk assessment has typically relied on the estimation of reference doses or reference concentrations based on the use of no-observed-adverse-effect levels (NOAELs) divided by uncertainty factors. The benchmark dose (BMD) approach has been proposed as an alternative basis for reference value calculations. In this analysis of the developmental toxicity observed in rats exposed to boric acid in their diet, BMD analyses have been conducted using two existing studies. By considering various end points (rib XIII effects, variations of the first lumbar rib, and fetal weight changes) and various modeling approaches for those end points, the best approach for incorporating all of the information available from those studies could be determined. Particular emphasis has been placed on methods for combining data across studies and for combining potentially related effects (on rib XIII and on the first lumbar rib). The issues of study and end point selection are ones that will arise frequently in the process of estimating reference values. This example of boric acid suggests that the BMD approach provides a reasonable basis for appropriately comparing and combining study data, as opposed to ad hoc combinations of study results. Moreover, it is shown that the BMD approach can be used with combinations of end points considered to differ in severity. In this case, the preferred approach involved combining the data from the two studies, which were similarly designed and were conducted in the same laboratory, to calculate BMDs that were more accurate and more precise than those that could be derived from either study alone. It was determined that decreased fetal body weight provided the best basis for BMD calculations; BMDs calculated for fetal body weight changes were less than those for all other relevant end points. The appropriate BMD to use as the basis for boric acid reference dose calculation appears to be 59 mg/kg/day, which is very similar to the NOAEL observed in the second of the two studies (55 mg/kg/day). Although the first study failed to establish a NOAEL, the BMD approach could have been applied to that study, thereby avoiding the need for a repeat study. Similar BMD results were obtained in both studies.

For developmental toxicity, as well as many other types of noncancer end points, regulatory agencies have typically calculated reference doses (RfDs) or reference concentrations (RfCs) using the no-observed-adverse-effect level (NOAEL) approach. A reference dose or reference concentration is intended to be a level of exposure for the human population that is likely to be without appreciable risk of causing adverse effects, in this case in the developing organism. The NOAEL approach identifies the greatest experimental dose not associated with an increase above background in the most sensitive adverse effect, and then applies uncertainty factors and modifying factors to derive the reference dose

RfD_{DT} = NOAEL/(UF \times MF),

where the DT subscript refers to developmental toxicity, UF is the product of all the uncertainty factors, and MF is the modifying factor accounting for data base adequacy issues (USEPA, 1991).

An alternative approach known as the benchmark dose (BMD) approach has been proposed (Crum, 1984). This approach uses the entire dose–response data base to define a BMD, the statistical lower bound on a dose corresponding to a relatively low level of response, which is used as an alternative to the NOAEL. This approach was proposed in response to potential problems that have been identified with NOAELs: they do not consider all of the dose–response data, they are inappropriately related to sample size, and they may be inconsistent from study to study (and sometimes controversial) at least in part because a NOAEL is constrained to be one of the experimental doses. In contrast, the BMD approach incorporates all of the dose–response information, is sensitive to sample size, can be applied consistently from one study to another, and the BMD need not be one of the experimental test doses. A series of studies (Faustman et al., 1994; Allen et al., 1994a,b; Kavlock et al., 1995) has shown that the BMD approach is generally applicable to developmental toxicity studies. The purpose of this paper is to evaluate and use the BMD approach in the context of boric acid developmental toxicity.

Boron is a naturally occurring element found primarily as
oxide compounds such as boric acid and inorganic borates. The inorganic borates, though ubiquitous in the environment, are found in commercially important concentrations at relatively few locations throughout the world. Boric acid and borates are used in the glass and ceramic industries and in detergent bleach, fertilizers, flame retardants, pesticides, and wood preservatives (Rainer, 1994; Woods, 1994). Boron is an essential plant micronutrient which occurs naturally in fruits, vegetables, nuts, and other foods. The primary source of human exposure to boron compounds is through the diet (Rainer, 1994; Woods, 1994). Preliminary clinical research by the USDA and other researchers indicates that boron is nutritionally beneficial for animals and humans (Newnham, 1991, 1994; Nielsen, 1991, 1993, 1994). Thus, exposure to low levels of boron appears to be both inevitable and potentially beneficial.

However, boric acid, at high levels in the diet, has been shown to be a developmental toxicant in laboratory mammals (see Heindel et al., 1992, 1994). The National Toxicology Program (NTP) tested boric acid in mice, rats, and rabbits in standard segment II developmental toxicity studies to investigate the effects of in utero exposure. Boric acid exposure during gestation was associated with an increased incidence of malformations in multiple organ systems. For example, these studies found that incidences of enlarged lateral ventricles (in rats), cardiovascular interventricular septal defects (in rats and rabbits), and short or missing rib XIII (in mice and rats) were increased among the offspring in a dose-related manner. In addition, a common skeletal variation (the presence of full or rudimentary first lumbar rib) tended to be less prevalent at the higher doses in mice, rats, and rabbits than in concurrent controls. Fetal weights decreased as a function of dose level in rats and mice (Heindel et al., 1992, 1994). Rats appeared to be the most sensitive species, with decreased fetal body weights being observed even at the lowest dose (0.1% in feed, equivalent to 78 mg/kg/day, on gestation Days 0–20).

A follow-up rat study was conducted by the same investigators in order to establish a clear NOAEL for developmental toxicity in rats (Price et al., 1994, 1995). The protocol for the second study was nearly identical to that of the first study, except that generally lower dose levels were employed (although two dietary boric acid concentrations—0.1 and 0.2% on gestation Days 0–20—were common to the two studies).

The two rat studies (Heindel et al., 1992; Price et al., 1994, 1995) were the subject of the present investigation. The BMDs associated with selected end points have been calculated. Novel dose–response modeling approaches have been explored in relation to the combination and analysis of effects that may be mechanistically related but which have been historically reported as separate morphological entities of different severity. Moreover, the existence of the two rat studies provided an excellent opportunity to examine the BMD with respect to its predictions across studies and with respect to the combination of study results. Recommendations are provided for the use of the BMD results in boric acid reference dose estimation.

METHODS

Experimental Material

Time-mated Sprague–Dawley rats were exposed to boric acid in the diet, on gestational Days 0 through 20, in two separate experiments. These two studies were conducted in the same laboratory using a similar experimental design, as summarized in Table 1. The dose group given 0.8% boric acid in the feed in the Heindel et al. (1992) study, referred to as Study A in this paper, has not been included in the subsequent analyses because the timing of exposure was different from all other groups. The Price et al. (1994, 1995) study, referred to as Study B in this paper, included a postnatal recovery phase that has not been considered in the analyses discussed here.

In Study A, all live fetuses were examined for gross (external) defects. Then half the fetuses were decapitated and examined for visceral defects (including brain defects) and skeletal defects (other than those of the skull, since examination of the brain tissue precluded staining and observation of the skull bones). The fetuses that were not decapitated also had visceral examinations (excluding the brain) as well as skeletal examinations (including the skull).

In Study B, the protocol differed in that only the fetuses that were decapitated (about 50% of live fetuses) had visceral examinations, including the brain. Otherwise, the fetal examinations were identical to those of the previous study.

Analysis of End Points

The end points selected for modeling and BMD estimation were those determined to be significantly related to dose in Study A. Similar effects were seen in Study B. Significance was determined by applying Mantel–Haenszel trend tests to the dose–response data (Haesman, 1984). The end points that were significantly related to dose in Study A included incidence of total malformations, enlarged lateral ventricles in the brain, agenesis or shortening of rib XIII, and variations of the first lumbar rib, as well as fetal weight decrements. Total malformations and enlarged lateral ventricles, although used for BMD calculations, are not discussed here; those end points appeared to be high-dose effects (Heindel et al., 1992) and their BMDs were much greater than those of the remaining end points discussed here. The rib XIII effects and lumbar rib variations were considered together and separately (see below), in light of the possibility that they may be two manifestations of a common mechanism of action affecting adjacent regions of the vertebral axis, although there are, to date, no data to support such a hypothesis.

Analysis of Rib Effects

Because treatment-related rib effects were observed at both rib XIII and the first lumbar rib, the relationship between boric acid dose and such effects was analyzed in three ways. First, the dose-related increase in fetuses with rib XIII effects was modeled, and, separately, the dose-related decrease in the incidence of fetuses with lumbar ribs was considered by modeling the

1 The term “total malformations” includes enlarged lateral ventricles and agenesis or shortening of rib XIII according to the classifications used in Studies A and B. The laboratory in which these studies were conducted now classifies enlarged lateral ventricles and short rib XIII as variations (see Price et al., 1994, 1995, for further discussion).
complementary dose-related increase in the incidence of offspring without lumbar ribs. The focus on increased incidence of fetuses without the lumbar rib variant (which would otherwise be considered normal) is consistent with the possible mechanism of action discussed above whereby boric acid inhibits rib growth in adjacent regions of the vertebral axis. The two additional approaches for analyzing the rib effects explicitly combine the rib XIII and first lumbar rib effects; the approaches were based on either determining weighted proportions of affected fetuses within litters or by calculating rib counts for each fetus. The weighted proportions and rib counts are described below.

Weighted proportions. Because the presence of lumbar ribs is considered to be a variation, while shortening or agenesis of rib XIII was classified by the investigators as a malformation, an approach that considered the relative severity of the response was desired. This was accomplished by defining weighted proportions. Thus, shortening or agenesis of rib XIII was given full weight in the counting of affected fetuses; the absence of a full or rudimentary lumbar rib was given partial weight in the counting. Three weighting schemes (referred to as rib effects analyses 1a, 1b, and 1c) were examined. They differed with respect to the partial weight assigned to the absence of a lumbar rib; the first scheme (1a) assigned a weight of $\frac{1}{2}$; the second scheme (1b), a weight of $\frac{1}{4}$; the third scheme (1c), a weight of $\frac{1}{8}$. These values were chosen to represent the possible interpretations that a missing lumbar rib is trivially different from "normal," is intermediate in severity between normal and frankly abnormal, or is tantamount to a malformation, respectively.

Within each litter, the weighted proportion of affected fetuses was calculated: a fetus with lumbar ribs and normal rib XIII was considered to be unaffected (to show no signs of rib growth inhibition) and contributed 0 to the calculation of a weighted proportion. Fetuses lacking lumbar ribs but with normal rib XIII contributed either $\frac{1}{2}$, $\frac{1}{4}$, or $\frac{1}{8}$ to the calculation; those fetuses that lacked lumbar ribs and also had rib XIII shortening or agenesis contributed 1 to the calculation.$^2$ The contributions of each fetus were added together within each litter and the average for each litter was calculated. The minimum weighted proportion for a litter would be 0 (if all fetuses in that litter had lumbar ribs) and the maximum weighted proportion would be 1 (if all fetuses in that litter had shortening or agenesis of rib XIII).

Rib counts. A rib count was computed for each fetus according to the following algorithm. The base number of rib pairs was set at 13. To that base, the following modifiers were added to determine a rib count:

- Rib XIII agenesis (left or right) $-0.50$
- Rib XIII shortening (left or right)$^3$ $-0.25$
- Rudimentary left or right lumbar rib $+0.25$
- Rudimentary bilateral lumbar rib $+0.50$
- Full right or full left lumbar rib $+0.50$
- Full bilateral lumbar rib $+1.00$

Given the rib count in each fetus, an average count was computed for each litter. The litter averages were considered to be continuous variables. This approach to analyzing total rib effects was referred to as "rib effects analysis 2."

Analysis of Fetal Weight

Changes in fetal weight were analyzed in two ways. First, average live fetal weight was determined for each litter with live fetuses, and the litter averages were considered to represent variations in a continuous variable. Second, a definition of adversely low birth weight fetuses was applied; counts of the numbers of fetuses within each litter with adversely low weights were analyzed. Adversely low birth weight was considered to be any weight less than the 5th percentile of the fetal weights in the concurrent control group (Kavlock et al., 1995). These two major methodologies for analyzing fetal weight are referred to as "fetal weight analyses 1 and 2," respectively.

Modeling

The BMD approach requires that dose–response models be applied to the experimental data. Such modeling was applied only to those end points for which there was a significant dose effect. In the case of developmental toxicity studies as analyzed here, Allen et al. (1994a,b) and Kavlock et al. (1995) have investigated models that are capable of describing the dose–

$^2$ One fetus in Study B had agenesis of rib XIII and a rudimentary lumbar rib on the left side. This fetus was given full weight for calculations of weighted proportions. All other fetuses that had shortening or agenesis of rib XIII also lacked lumbar ribs.

$^3$ In the absence of more specific information, an observation of rib XIII agenesis or shortening has been considered to be unilateral for this analysis.
response data for malformations and fetal weight changes. Those models are described here. For analysis of those end points measured on a continuous scale (litter averages for fetal weights, weighted proportions of fetuses with rib effects, and rib counts), a continuous power model was used. The continuous power model was expressed as

\[ m(d) = \alpha - \beta \times d^r, \]  

(1)

where \( m(d) \) is the average litter mean at dose \( d \) (expressed in terms of mg/kg/day) and \( \alpha, \beta, \) and \( r \) were the three parameters estimated by maximum likelihood methods. This model assumed normal variation around the mean \( m(d) \). Separate variances were estimated for each dose group.

For analysis of rib XIII shortening or agenesis and lumbar rib variations (when they were considered separately from one another), a log-logistic model was applied. The log-logistic model was expressed as

\[ P(d,s) = \alpha + \theta s + [1 - \alpha - \theta s][1 + \exp(\beta + \theta s - \gamma \log(d))], \]  

(2)

where \( P(d,s) \) is the probability of response for dose \( d \) and litter size \( s \), and parameters \( \alpha, \beta, \gamma, \theta, \) and \( \theta \) were estimated for each end point by methods of maximum likelihood. A beta-binomial probability model was assumed when the log-logistic model was applied; i.e., the responses within a given litter were assumed to be described by a binomial distribution where the underlying probability of response varied from litter to litter according to a beta distribution. The mean of the beta distribution was given by Eq. (2). The other parameter of the beta-binomial distribution, the intralitter correlation coefficient, was estimated separately for each dose group.

The log-logistic model was also applied to the analysis of fetal weight when that end point was expressed in terms of the number of fetuses with abnormally low weights (fetal weight analysis 2).

A BMD and its corresponding maximum likelihood estimate (MLE) are defined in terms of a prespecified level of effect, referred to as the benchmark effect (BME) level (Kavlock et al., 1995). For mean fetal weight (fetal weight analysis 1), the BMD was defined either as the 95% lower bound on the dose corresponding to a 5% decrease in the mean (BMD was 5% decrease) or as the 95% lower bound on dose corresponding to a decrease in the mean equal to \( s_0(0)/2 \), where \( s_0(0) \) was the observed standard deviation of the control group mean weights (BMD was decrease by \( s_0(0)/2 \); Kavlock et al., 1995). These two variations of fetal weight analysis 1, selected because Kavlock et al. (1995) found such BMEs to yield BMDs close to corresponding NOAELs, on average, were referred to as fetal weight analyses 1a and 1b, respectively. For the rib count end point (rib effect analysis 2), the BMD was the 95% lower bound on dose corresponding to a rib count \( s_0(0)/2 \) units below the mean control count, where \( s_0(0) \) was the standard deviation of the control group mean rib counts (BMD was decrease by \( s_0(0)/2 \)). BMDs for all end points modeled using Eq. (2) were defined as the 95% lower bound on dose corresponding to an increase in the probability of response of 5% (i.e., the BME was 5% additional risk). A BME of 5% additional risk was also used for the weighted proportions of fetuses with rib effects (rib effect analyses 1a, 1b, and 1c). All confidence limits were calculated using the likelihood approach (Crump and Howe, 1985; Cox and Lindley, 1974). The definitions of the BMD utilized here were based on the findings of Alten et al. (1994a,b) and Kavlock et al. (1995) with respect to BMD estimation for developmental toxicity experiments in general. MLEs represent the best estimates of the designated doses, when maximum likelihood techniques are used, without representing the statistical uncertainties involved.

The models were examined for fit to the data. In the case of the log-logistic model, \( \chi^2 \) approaches for comparing observed and expected numbers of affected fetuses were employed. For the continuous power model, \( F \) tests that compared the lack of model fit to an estimate of pure error were employed.

### Combination of Study Results

For each end point under investigation, the results of the two studies were compared. In particular, the dose–response patterns were examined to determine if a single function (either Eq. (1) or Eq. (2)) could adequately describe the responses in both studies. This determination was based on a likelihood ratio test. The maximum log-likelihoods from the models fit to the two studies considered separately were added together; from this was subtracted the maximum log-likelihood for the model fit to the combined results. Twice that difference is distributed approximately as a \( \chi^2 \) random variable (Cox and Lindley, 1974). The degrees of freedom for that \( \chi^2 \) variable are equal to the number of parameters in the model plus 1; the additional degree of freedom was available because in the combined results, the two control groups were treated as one group, which eliminates the need to estimate one of the intralitter correlation coefficients (for beta-binomial random variables) or variances (for normal random variables) that was estimated when the studies were treated separately. The critical values from the appropriate \( \chi^2 \) distributions (associated with a \( p \) value of 0.01) were compared to the calculated values. When the calculated value was less than the corresponding critical value, the combined results were used to estimate BMDs; this result indicated that the responses from the two studies were consistent with a single dose–response function.

### RESULTS

The dose–response data used in the analyses are summarized in Table 2. As discussed under Methods, two types of data summaries were considered depending on the nature of the end point. First, for the continuous end points (e.g., average fetal weight, average rib count, or weighted proportion of fetuses with rib effects), means and standard deviations were analyzed. Second, for the quantal end points (e.g., malformations, variations, or low-weight fetuses), recorded as counts of affected fetuses, Table 2 presents the number of affected fetuses over the number of fetuses examined. The analysis of the latter end points actually considered the distribution of affected fetuses within individual litters.

For fetal weight analysis 1, utilizing litter averages which were treated as continuous variables, the two studies displayed good consistency. The dose–response patterns were very similar (Fig. 1) and consistent with a single dose–response relationship (Table 3; fetal weight analyses 1a and 1b). The two approaches to defining a BMD (by finding the dose corresponding to a 5% decrease in mean weight [1a] or by finding the dose giving a mean response half a standard deviation below the control mean [1b]) resulted in similar BMDs. Of particular interest is the fact that the analysis of the combined studies yielded a BMD at or above the BMD from the first study, even though the maximum likelihood estimates of doses for the combined studies were less than those from Study A. This is an illustration of the advantage of larger data sets for BMD estimation (see Discussion below).

Although the continuous treatment of fetal weight suggested that combination of the studies was appropriate, the analysis of counts of low-weight fetuses (fetal weight analysis 2) did not indicate that the end point should be combined
decreased average fetal weight is defined in such a way that change from control (and therefore lower dose associated with the background variability is not directly involved, i.e., by comparing mean to its corresponding mean. Consequently, for a given change in the mean than the 5th percentile from Study A was to its corresponding 5th percentile from Study B was closer to its corresponding 

The same phenomenon, less variability in the control group across studies. The fit of the log-logistic model to the results from Study B was relatively poor (goodness-of-fit p value equal to 0.01); this was due, at least in part, to the nonmonotonicity of the dose–response for Study B (Table 2 and Fig. 2). Figure 2 graphically highlights the differences in the best-fitting dose–response models for the two studies; it is not surprising that this end point did not appear to be compatible with a single dose–response function for the two studies. Because the variability in the Study B control group was less than that in the Study A control group (Table 2), the 5th percentile from Study B was closer to its corresponding mean than the 5th percentile from Study A was to its corresponding mean. Consequently, for a given change in the mean fetal weight as a function of dose, there will tend to be more low-weight fetuses in Study B. This is in fact observed, because the variability in the Study B control group was less than that in the Study A control group (Table 2), the 5th percentile from Study B was closer to its corresponding mean than the 5th percentile from Study A was to its corresponding mean.

The same phenomenon, less variability in the control group of Study B, results in smaller standard deviations (shorter error bars) for the control litter average weights, compared to Study A (Fig. 1). Thus, for fetal weight analysis lb, the BME level bars) for the control litter average weights, compared to Study A up to about 150 mg/kg/day (Fig. 2). Shortening or agenesis of rib XIII (see Discussion below). For that reason, the total rib effect has been summarized via weighted proportions of affected fetuses and average counts of ribs.

The weighted proportions of affected fetuses with rib effects as indicated.

### Table 2: Data from Two Studies of Fetal Effects Following Exposure to Boric Acid

<table>
<thead>
<tr>
<th>End point</th>
<th>Study</th>
<th>0</th>
<th>19</th>
<th>36</th>
<th>55</th>
<th>76</th>
<th>78</th>
<th>143</th>
<th>163</th>
<th>330</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fetal weight</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>analysis 1; litter averages</td>
<td>A</td>
<td>3.70</td>
<td>(0.32)</td>
<td>3.61</td>
<td>(0.24)</td>
<td>3.56</td>
<td>(0.23)</td>
<td>3.53</td>
<td>(0.28)</td>
<td>3.50</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fetal weight</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>analysis 2; low-weight fetuses*</td>
<td>A</td>
<td>2.14</td>
<td>(4.3 )</td>
<td>2.14</td>
<td>(5.0)</td>
<td>2.44</td>
<td>(5.2)</td>
<td>4.24</td>
<td>(9.6)</td>
<td>4.04</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shortening or agenesis of rib</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>XIII</td>
<td>A</td>
<td>2.43</td>
<td>(0.5 )</td>
<td>2.44</td>
<td>(0.2)</td>
<td>1.43</td>
<td>(0.2)</td>
<td>3.43</td>
<td>(0.7)</td>
<td>8.47</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>1.44</td>
<td>(0.2)</td>
<td>0.46</td>
<td>(0.0)</td>
<td>1.43</td>
<td>(0.2)</td>
<td>3.43</td>
<td>(0.7)</td>
<td>8.47</td>
</tr>
<tr>
<td>Missing lumber ribs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>3.69</td>
<td>(85.6)</td>
<td>4.04</td>
<td>(96.9)</td>
<td>4.55</td>
<td>(98.7)</td>
<td>4.25</td>
<td>(97.3)</td>
<td>4.33</td>
<td>(99.1)</td>
</tr>
<tr>
<td>B</td>
<td>4.04</td>
<td>(96.9)</td>
<td>4.55</td>
<td>(98.7)</td>
<td>4.25</td>
<td>(97.3)</td>
<td>4.33</td>
<td>(99.1)</td>
<td>4.64</td>
<td>(98.5)</td>
</tr>
<tr>
<td>Rib effects analysis</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1st 5/6th</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>weighting*</td>
<td>A</td>
<td>0.42</td>
<td>(0.08)</td>
<td>0.16</td>
<td>(0.01)</td>
<td>0.16</td>
<td>(0.02)</td>
<td>0.17</td>
<td>(0.02)</td>
<td>0.17</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>0.48</td>
<td>(0.03)</td>
<td>0.49</td>
<td>(0.02)</td>
<td>0.48</td>
<td>(0.03)</td>
<td>0.49</td>
<td>(0.02)</td>
<td>0.50</td>
</tr>
<tr>
<td>Rib effects analysis</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1st 5/6th</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>weighting*</td>
<td>A</td>
<td>0.71</td>
<td>(0.13)</td>
<td>0.82</td>
<td>(0.04)</td>
<td>0.82</td>
<td>(0.04)</td>
<td>0.84</td>
<td>(0.02)</td>
<td>0.90</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>0.87</td>
<td>(0.05)</td>
<td>0.82</td>
<td>(0.03)</td>
<td>0.80</td>
<td>(0.03)</td>
<td>0.84</td>
<td>(0.03)</td>
<td>0.82</td>
</tr>
<tr>
<td>Rib effects analysis</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2nd rib count</td>
<td>A</td>
<td>3.05</td>
<td>(0.07)</td>
<td>1.03</td>
<td>(0.01)</td>
<td>1.03</td>
<td>(0.04)</td>
<td>1.03</td>
<td>(0.02)</td>
<td>1.03</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>3.01</td>
<td>(0.07)</td>
<td>1.03</td>
<td>(0.01)</td>
<td>1.03</td>
<td>(0.04)</td>
<td>1.03</td>
<td>(0.02)</td>
<td>1.03</td>
</tr>
</tbody>
</table>

Note. Quantal (count) end point data are presented as number affected/total number of fetuses examined (with percentage response in parentheses). Continuous end point data are summarized by the mean (and standard deviation) across litters.

* Study A is from 1988 (Heindel et al., 1992); Study B is from 1994 (Price et al., 1994, 1995).

* Low-weight fetuses are those weighing less than the fifth percentile of the concurrent control fetal weights.

* Proportion of fetuses with effects on rib XIII or missing lumbar ribs; missing lumbar ribs are weighted in relation to rib XIII effects as indicated.
FIG. 1. Observed and predicted average fetal weights. Model predictions (solid and dotted curves) compare favorably to the observed means (± one standard deviation) for each study. The data from the two studies are, in fact, consistent with a common dose–response relationship.

In light of the results presented above, this discussion will highlight the following issues:

- repeated experimentation in the context of BMD analyses
- combination of data from different studies
- combination of related end points, if they differ in presumed severity
- practical consideration of different procedures for BMD definition for continuous variables
- a BMD of 59 mg/kg/day recommended as the basis for subsequent reference dose estimation.

Using current NOAEL-based risk assessment procedures, an RfD derived from the original rat study would have required application of an additional uncertainty factor (UF) to the LOAEL, the default value of that UF being a factor of 10. A 10-fold correction for the LOAEL to NOAEL conversion is clearly excessive for that study (Study A) considering that the 6% reduction in fetal weight at the LOAEL was close to the limit of statistical detection (the corresponding p value for testing dose–group differences was just less than 0.05). Study B was conducted to empirically establish a NOAEL for fetal weight. However, as demonstrated in this paper, the information from Study A can be used to determine a BMD, functionally substituting for a NOAEL, at considerably less expense and difficulty and avoiding altogether the issue of LOAEL-to-NOAEL uncertainty factors. As the discussions below will suggest, had Study A been the only data set available, the BMD estimate recommended for subsequent reference dose estimation would have been 56 mg/kg/day, almost identical to the NOAEL of 55 mg/kg/day identified in Study B. When use of the BMD approach is more universally accepted, it will no longer be necessary...
to define a NOAEL experimentally, and repeat testing, as in this case, may not be required.

Nevertheless, the availability of the two similarly conducted rat studies has afforded us the opportunity to investigate the use of data from different studies in the BMD approach. The combination of data is recommended only when studies are in reasonable agreement with respect to dose-response relationships, and statistical tests have been employed to formally test for such agreement. Toxicological issues of similarity of studies should also be considered, but in this case the studies shared many common features, not least of which was the common testing laboratory. Dose-response patterns for fetal weight, measured on a continuous scale (fetal weight analyses 1a and 1b), rib XIII effects, and lumbar rib variations were all comparable across the two studies.

In every case where combination of the two studies was appropriate, the BMDs estimated for the combined data were "closer" to the corresponding MLEs than for either of the studies alone. That is, the confidence intervals around the best estimates of dose corresponding to the selected response level were narrower. This increase in precision is expected when more animals are tested; in general, the more data points that are available, the more confidence one has in the estimates based on those data.

It is interesting to note, however, that merely increasing sample size does not always increase the precision of the
FIG. 2. Observed and predicted probabilities of low-weight fetuses. Mean probabilities of low-weight fetuses as a function of dose (large circles, with individual observations for each litter designated by smaller circles) were well-predicted by log-logistic model for Study A (solid curve; lack-of-fit p value was 0.44). Observed and predicted means for Study B (large squares and dotted curve, respectively) showed less agreement (lack-of-fit p value = 0.01). Study-specific dose–response curves were significantly better than a single common dose–response curve.

FIG. 3. Observed and predicted probabilities of missing lumbar ribs. The steep dose–response curves (solid and dotted curves) for both studies entail relatively low MLE and BMD estimates. The predicted background rates (for the average control litter size) differed across the two studies: for Study A the predicted background rate was 0.86, whereas for Study B that rate was 0.90, accounting for the differences in the curves at low doses.
estimates. Also affecting BMD estimation is the location of the experimental doses relative to the MLE and BMD estimates. When the doses corresponding to a selected effect level are estimated to be within the range of doses used in an experiment, then no extrapolation is required. In that case, much more confidence can be placed on the accuracy of the estimates; all models that fit the data will give approximately the same estimates for doses in this range. Moreover, the precision of the estimates is adversely affected by the need for extrapolation. An example of this is provided by shortening or agenesis of rib XIII. For Study B, the MLE dose for the BME of 5% additional risk was actually above the highest dose in that study; this is an example of extrapolation not typically considered in risk assessment, i.e., extrapolation above the dose range. Nevertheless, the value of the BMD relative to the MLE (BMD/MLE) was less for Study B than for Study A for this end point. The MLE from Study A was within the range of doses used in that study and no extrapolation was required. The BMD/MLE ratio for the combined studies was substantially greater than the ratio for either study considered alone, reflecting both the increased precision due to increased sample size and the fact that, with the combined data, no extrapolation was required.

Another consideration with respect to the accuracy of the estimates is the degree to which the dose–response relationship is well estimated. The BMD approach is designed to be sensitive to the dose–response pattern, so increased accuracy is attained when one can be more confident about the nature of that pattern. The best way to increase confidence is to increase the number of dose groups. Thus, although there is no measure of improved accuracy, the combination of studies that have the same dose–response pattern for a particular end point will increase confidence that the dose–response pattern has been estimated satisfactorily, especially when the experiments have employed different doses.

The effect with respect to characterization of the dose–response relationship attributable to increasing the number of dose groups can be illustrated by considering the changes in the MLEs when such combinations are accomplished. For fetal weight analyses 1a and 1b the MLEs for the combination of the studies fell within the estimates from the two studies considered separately. For shortening or agenesis of rib XIII, the combined MLE was less than either of the single-study MLE estimates. For missing lumbar ribs, the MLE was greater than either of the MLEs from the studies considered separately. These changes in the MLEs reflect changes in the characterization of the dose–response relationship.

The discussion above is based on a small set of examples from the analysis of boric acid. The relationships between BMD estimates and sample size, number of dose groups, and spacing of dose groups warrants further investigation. Both case studies and simulations can be used to extend the base of knowledge about BMD estimation for developmental toxicity and for other types of noncancer end points. Recent presentations (Kavlock et al., 1994; Weller et al., 1994) have considered some of these issues.

For boric acid in particular, the question that remains to be addressed is the choice of the BMD results to use for reference dose estimation. This issue is considered important, despite the fact that the postnatal component of Study B (Price et al., 1994, 1995) demonstrated that the effects in question were largely reversible or that there had been substantial “catching-up” for the developmental delays evident at the time of birth. It is far from clear that the postnatal results would completely eliminate concern about the prenatal observations, especially in regulatory contexts.

Among the end points considered in this analysis, fetal weight changes and rib effects, the BMDs for lumbar rib variations are the smallest that have been derived. The following discussion suggests, however, that those BMDs are not satisfactory for reference dose estimation.

The manner in which the lumbar rib variation has been treated is unusual. What has been modeled is the apparent dose-related loss of that variation. As noted above, that effect is not adverse. If the loss of the variations had been the only rib effect observed, then it would not have been used to assess the risk associated with boric acid exposure. However, it is possible that the dose-related decrease in numbers of fetuses with lumbar ribs is a less severe manifestation of the same mechanism causing shortening or agenesis of rib XIII. The effect of boric acid on the lumbar ribs may be analogous to slight changes in a clinical chemistry parameter, changes that remain within normal limits; such effects are not adverse but may be indicative of an adverse effect that may become apparent at higher doses. Thus, it is only in conjunction with the rib XIII abnormalities that the loss of variations may be hypothesized to be consistent with an adverse effect of boric acid; one should not consider the lumbar rib changes by themselves as indicative of an adverse effect. Thus, the BMDs for missing lumbar ribs are less appropriate for the interpretation of boric acid risks than are the BMDs for total rib effects that consider the variation in combination with the rib XIII effects.

The weighted proportions analyses and the rib count analysis were designed to consider the total rib effect. Because it is not entirely clear how one should relate the severity of rib XIII shortening or agenesis to the severity of missing lumbar ribs, three weighting schemes were investigated. The weights used in those schemes (assigning a weight of either 1/6, 1/2, or 5/6 to missing lumbar ribs relative to the rib XIII effects) cover much of the possible range (from 0 to 1). The weight of 1/6 represents an evaluation that missing lumbar ribs are considerably less serious than short or missing rib XIII; it may be most consistent with the fact that the “normal” state is without lumbar ribs (Charles River, 1993;
Chernoff et al., 1991). On the other hand, a weight of 5/6 implies that it is nearly as bad to be missing a lumbar rib as it is to have shortening or agenesis of rib XIII. The 5/6 weighting may be appropriate if it is believed that there is a strong relationship between boric acid-induced rib XIII effects and missing lumbar ribs. Clearly, this weighting suggests that at least 85% of control fetuses had effects tantamount to a malformation. This may be considered an exaggeration of the situation, but it represents a conservative approach to modeling severity.

Because any weighting scheme introduces a subjective component, it was considered desirable to try to obtain a more objective measure of the total rib effect. The one possibility included in this investigation of the boric acid studies was to determine the loss of ribs, starting in the lumbar region and moving up to the thoracic region. The rib count analysis (rib effects analysis 2) attempted to represent this approach. The values of the rib count end point for each fetus reflect the best estimate of the number of pairs of ribs that the fetus has. The primary difficulty with this measure of response is that the data available for getting counts (estimating the number of pairs of ribs) are not optimal. The values assigned (see Methods) were selected to estimate counts as consistently with the codings of the rib effects as possible, but this may have introduced some subjectivity, especially with respect to fetuses with combinations of rib XIII and lumbar rib effects. As an example, a fetus with agenesis of rib XIII was considered to have 12.5 pairs of ribs: it was assumed that the agenesis was only on one side. This assumption was made because some fetuses were recorded as having both agenesis and shortening of rib XIII. It should be noted that the difference between full ribs and shortened ribs, or between shortening of ribs and agenesis of ribs, may be very subtle, so the counts used in this analysis may exaggerate differences that are relatively minor. An alternative approach could be based on actual rib length measurements, which were not collected in these studies.

No matter whether weighted proportions or rib counts were considered, the BMDs for rib effects from either study were greater than those calculated for fetal weight changes. The analyses of rib effects have included several possibilities to reflect uncertainty with respect to the correct relationship, especially with respect to combinations of rib and lumbar rib effects. As an example, a fetus with agenesis of rib XIII was considered to have 12.5 pairs of ribs: it was assumed that the agenesis was only on one side. This assumption was made because some fetuses were recorded as having both agenesis and shortening of rib XIII. It should be noted that the difference between full ribs and shortened ribs, or between shortening of ribs and agenesis of ribs, may be very subtle, so the counts used in this analysis may exaggerate differences that are relatively minor. An alternative approach could be based on actual rib length measurements, which were not collected in these studies.

The preceding discussions lead to the conclusion that a BMD based on fetal weight changes should determine the reference dose. BMDs calculated for fetal body weight changes were less than those from all other relevant endpoints. Of the fetal weight BMD results presented in Table 3, one must decide on the appropriate choice.

Better results were available using fetal weight analysis 1. The model used in that analysis fit all of the data. Moreover, the combined data were consistent with a single dose–response function. Slight differences in the curve shapes for the studies considered individually (Fig. 1) led to differences in the study-specific BMDs. However, the fact that the two studies were consistent with respect to the dose–response of average fetal weight entails that the benefits of combining the studies, as discussed above, improve both the accuracy and the precision of the BMD procedure. Analysis of the combined results was in this case preferable to selection of one study over the other.

For fetal weight analysis 1, the two versions differ with respect to the BMD used to calculate the BMD. As discussed above in relation to fetal weight analysis 2, the BME that was defined directly in terms of control group standard deviation (s_d(0)/2 for fetal weight analysis 1b) may be overly sensitive to differences in estimated background variability. Even if everything else was the same for the two studies, the difference in control group variability (Table 2 and Fig. 1) would make the BMD from Study B less than that from Study A. Given that the two studies were so similar in other respects (especially dose–response) it appears more appropriate to rely on fetal weight analysis 1a that is not as sensitive to estimates of background variability.

The issues addressed in the case of boric acid will be similar to those faced with other compounds for which refer-
ence doses are desired. In particular, choices from different end points and different BMDs for any particular end point must be made. If, as in the case of boric acid, more than one study exists, then the relative merits of the studies must be considered. The BMD approach has an advantage over the NOAEL in that instance, over and above those already noted (Crump, 1984). That is, by appropriately combining the dose–response data, rather than the NOAELs, of two studies, increases in accuracy and precision can be attained and one can avoid ad hoc, and perhaps overly conservative, choices from among the study-specific outputs.

For the specific case of boric acid, in light of the results of both studies and all of the end points considered above, the fetal weight end point, analyzed using fetal weight analysis Ia (examining average weights within litters and using a BMD defined in relation to a 5% decrease in those averages) appears to be the best basis for subsequent reference dose estimation for boric acid developmental toxicity risk assessment. The specific BMD resulting from those considerations is 59 mg/kg/day. This BMD is based on the combined results of the two studies that were similarly designed and were conducted in the same laboratory. The selected value is not much less than the lowest dose tested in Study A (78 mg/kg/day, which was considered to be a LOAEL) and is very close to the NOAEL determined in Study B, 55 mg/kg/day.

REFERENCES


