Statistical Analysis of Nonmonotonic Dose-Response Relationships: Research Design and Analysis of Nasal Cell Proliferation in Rats Exposed to Formaldehyde

David W. Gaylor,*† Werner K. Lutz, ‡ and Rory B. Conolly‡

*Gaylor and Associates, Eureka Springs, Arkansas 72631; †Department of Toxicology, University of Würzburg, 97078 Würzburg, Germany; and ‡CIIT Centers for Health Research, Research Triangle Park, North Carolina 27709

Received May 13, 2003; accepted September 28, 2003

Statistical analyses of nonmonotonic dose-response curves are proposed, experimental designs to detect low-dose effects of J-shaped curves are suggested, and sample sizes are provided. For quantal data such as cancer incidence rates, much larger numbers of animals are required than for continuous data such as biomarker measurements. For example, 155 animals per dose group are required to have at least an 80% chance of detecting a decrease from a 20% incidence in controls to an incidence of 10% at a low dose. For a continuous measurement, only 14 animals per group are required to have at least an 80% chance of detecting a change of the mean by one standard deviation of the control group. Experimental designs based on three dose groups plus controls are discussed to detect nonmonotonicity or to estimate the zero equivalent dose (ZED), i.e., the dose that produces a response equal to the average response in the controls. Cell proliferation data in the nasal respiratory epithelium of rats exposed to formaldehyde by inhalation are used to illustrate the statistical procedures. Statistically significant departures from a monotonic dose response were obtained for time-weighted average labeling indices with an estimated ZED at a formaldehyde dose of 5.4 ppm, with a lower 95% confidence limit of 2.7 ppm. It is concluded that demonstration of a statistically significant bi-phasic dose-response curve, together with estimation of the resulting ZED, could serve as a point-of departure in establishing a reference dose for low-dose risk assessment.

Key Words: dose response; hormesis; risk assessment; cell proliferation; formaldehyde; statistical procedures.

In the foregoing paper, plausible mechanistic explanations for nonmonotonic (biphasic) dose-response relationships were provided (Conolly and Lutz, 2004). The hypothesis that this may be a generalizable and unifying concept (Calabrese and Baldwin, 2001) has now to be investigated more directly, i.e., with specifically designed experiments. Data have to be generated to validate the existence of a J-shaped dose response for the incidence or strength of an adverse effect. Also, for regulatory purposes, it may be useful to estimate the zero-equivalent dose (ZED) where the response is equal to the background response in unexposed control animals (Gaylor, 1994). In this setting, a lower confidence limit (LZED) to account for experimental variation in the estimation of the ZED could serve as a point of departure for setting a reference dose (RfD) associated with a negligible risk. That is,

\[ \text{RfD} = \frac{\text{LZED}}{U_A \times U_H \times U_S} \]  

where \( U_A, U_H, \) and \( U_S \) are uncertainty factors to account for extrapolation from animals to humans, for interindividual variation in sensitivity to a chemical exposure, and, if necessary, for extrapolation from short-term to chronic exposures, respectively. Presumably, an RfD need not be set below this value. This proposal for setting an RfD is more conservative than the current use of a lower bound on a benchmark dose corresponding to some level of increased risk, whereas the LZED is a lower boundary on the dose, estimated to truly produce no increased risk.

Principles of the statistical design of experiments will be utilized in order to determine the selection of experimental doses and the allocation of the numbers of animals to those doses, in order to maximize the power of statistical tests to detect J-shaped dose responses and to attain adequate precision of the estimate of the ZED. Obviously, it is necessary to demonstrate the existence of a statistically significant J-shaped dose-response curve in order to provide strong evidence for discontinuing the use of a conservative linear extrapolation to zero for low-dose risk assessment. The impact of such a demonstration would be to allow higher concentrations of such chemicals in the environment, workplace, or consumer products without increasing risks above background incidence rates. For the analysis of low-dose endocrine disruptor data, statistical issues have been discussed (Crump, 2001; Haseman et al., 2001), but guidelines for research design are scarce and limited (Bailer, 2001; Sielken and Stevenson, 1998).

Generally, an adequately powerful statistical test for com-
paring tumor incidence rates requires a large number of animals as the incidence rates depart from zero. A surrogate (precursor) measure for cancer that can be quantified on a continuous scale, e.g., a relevant DNA adduct concentration or cell division rate in the target tissue, may provide more precise information than quanital tumor-incidence data. Statistical tests will be investigated for detecting J-shaped dose-response curves for incidence rates or continuous data. The selection of doses and the numbers of animals assigned to doses will be investigated to achieve adequate power for detecting effects and to provide adequate precision for the estimation of the ZED in the setting of RfDs.

MATERIALS AND METHODS AND RESULTS

Experimental Design to Test for a J-shaped Curve

For purposes of discussion, the experimental response will represent an adverse health effect or surrogate for an adverse health effect. With a J-shaped dose response, there will be a decline in the level of the adverse health effect from the background level to a minimum response at a dose designated as \( d_m \) (Fig. 1). Above this dose, the response increases until the background level is again obtained at the ZED. The adverse response will continue to increase at doses above the ZED. A potential candidate for a reference dose (RfD) could be a lower confidence limit (LZED) divided by uncertainty factors.

It is clearly straightforward to demonstrate increases in adverse effects at high doses. It is assumed here that adverse effects have or can be demonstrated. The difficulty lies in demonstrating the presence of a J-shaped dose response. This requires showing a decrease in the adverse effect at low doses. Such decreases are likely to be relatively small compared to the increases observed for adverse effects at high doses.

The most powerful statistical test for a J-shaped dose response is obtained by comparison of the response for the control animals with the response at the \( d_m \). This is the dose that provides the maximum difference in response from the background level. A statistical test generally will be in the form of a \( t \)-test,

\[
t = \frac{(Y_0 - Y_m)\sqrt{(s_0^2/n_0) + (s_m^2/n_m)}}{s_0/s_m}^{1/2}
\]

where \( Y_0 \) and \( Y_m \) represent mean responses, \( s_0 \) and \( s_m \) are the standard deviations, and \( n_0 \) and \( n_m \) are the number of animals for the controls and \( d_m \) level, respectively, and \( N = (n_0 + n_m) \) is the total number of animals. As the value of \( t \) increases, the level of statistical significance increases. The value of \( t \) is maximized by selecting \( n_0 \) and \( n_m = (N - n_0) \), so that the denominator of \( t \) is minimized. Setting the first derivative of the denominator with respect to \( n_0 \) equal to zero gives:

\[
n_0/n_m = s_0/s_m
\]

This condition for relative sample sizes maximizes the power of the statistical test for detecting a decrease in an adverse health effect at low doses, indicating a J-shaped dose response. Generally, \( d_m \) will be relatively small. Thus, it is expected that \( s_0 \) and \( s_m \) will be similar, if not equal. Under this condition from Equation 3, the maximum power of the statistical test for continuous data is achieved by setting \( n_0 = n_m \), i.e., placing half of the animals in the control group and half at the \( d_m \). However, in general, the exact value of \( d_m \) will be unknown. It may be possible from the development of the biologically based model to make a preliminary estimate of \( d_m \) (Conolly and Lutz, 2004). However, it probably would be unwise to assign half of the animals to this dose. Due to the uncertainty in the initial prediction of the \( d_m \), it is suggested that dose levels below and above the suspected \( d_m \) should also be used in order to hopefully have two doses that span the true \( d_m \). This span may need to be quite wide if there is considerable uncertainty in the value of \( d_m \), which may require additional doses.

Let \( D_m \) represent an initial estimate of the value of \( d_m \). It is suggested that the experimental design for testing for a J-shaped dose response should consist of control animals (\( D = 0 \)) and animals at \( (D_m/k) \), \( (D_m) \), and \( (k \times D_m) \), where \( k \) is some factor greater than one. For example, with \( k = 2 \), animals would be administered doses of 0, \( D_m/2 \), \( D_m \), and \( 2D_m \).

Thus, the recommended experimental design to detect a J-shaped curve consists of \( n \) animals at each of four doses (total number of animals = \( 4n \)), consisting of controls, the dose estimated to minimize the adverse health effect, and one dose below and one above this estimated minimum effect dose. The number of animals per dose (\( n \)) required in order to achieve adequate power for detecting a J-shaped dose response will be discussed later.

For quantal data with a proportion \( (p) \) of exhibiting a biological effect, the standard deviation is \( s = \sqrt{p(1 - p)} \). From Equation 2, the optimum allocation of the numbers of animals is \( n_0/n_m = \frac{p_0(1 - p_0)}{p_m(1 - p_m)}^{1/2} \). That is, the optimum allocation of animals would assign more to the control group.

\[
FIG. 1. Schematic representation of a J-shaped dose-response curve, as one possibility of a nonmonotonic dose-response relationship; \( d_m \), dose showing minimum response; \( d_m \), dose where control level is reached after decrease at low dose (=ZED), zero equivalent dose. Bars represent standard deviations that apply for continuous measurements.
\]
For example, if \( p_0 = 0.4 \) and \( p_n = 0.2 \), the optimum allocation would assign \( n_o = (1.22 \times n_m) \).

**Statistical Tests for J-Shaped Curves**

For quantal data, each animal is categorized as normal or as demonstrating an adverse effect, e.g., diagnosed with a particular type of tumor. The proportion of animals with the tumor in a dosed group is compared with the proportion of tumor-bearing animals in the control group. In order to demonstrate the presence of a J-shaped dose response, it will be necessary that at least a moderate incidence of animals with the adverse effect is present in the control animals in order for a reduction in the incidence to be demonstrated in dosed animals. Fisher’s exact test can be employed to test for a statistically significant reduction in the proportion of dosed animals with the adverse health effect.

For continuous data ordinary \( t \)-tests generally can be conducted. Since biological measurements often are restricted to be positive and a small proportion of animals may exhibit high values, i.e., a positively skewed distribution, a transformation of the data may be necessary. Often, biological measurements are approximately log-normally distributed among animals. That is, the logarithm of the measurement will be approximately normally distributed so that a \( t \)-test can be used and standard confidence limits can be calculated for a mean value. In some extreme cases it may be necessary to employ nonparametric statistical procedures.

**Multiple Comparisons**

For the experimental design discussed above with three dose groups and controls, results at the three doses can be compared to the controls. This provides an increased opportunity for a false-positive result. Requiring a higher level of statistical significance for some tests can control the false-positive rate. If it is desired to maintain the false positive rate at \( \alpha \), then the comparison among the three dosed groups showing the minimum \( P \)-value would be required to obtain a statistical significance level of \( P \leq \alpha/3 \). Generally, \( \alpha \) is set equal to 0.05. If this difference is not statistically significant at the \( \alpha/3 \) level, testing stops. If this test achieves statistical significance, the test for the second largest difference is conducted at the \( P \leq \alpha/2 \) level. If this second test is not statistically significant, again testing stops. If the second test is also significant, the third (final) test with the smallest difference from the controls is tested at the \( P \leq \alpha \) level. Testing in this manner maintains the overall false-positive rate at \( P \leq \alpha \) (Holm, 1979).

**Sample Size for Tests of Proportions (Quantal Data)**

Proportions arise from data that are variously referred to as quantal, binary, or dichotomous data. The sample size \( (n) \) required per group to have a specified probability (power) to detect a difference between two proportions depends on the proportion in the controls and the size of the difference in proportions being tested. Recall that there are three dosed groups plus controls. Thus, the total number of animals required is \( N = 4n \). Sample sizes required to have an 80% chance of detecting various differences in proportions are displayed in Table 1. Even with these sample sizes, there is a 20% chance that an experiment would fail to attain a statistically significant difference, although a true difference exists. In order to decrease this probability of a false-negative result requires even larger sample sizes. The true differences are expressed as percent reductions from the background rate for the controls. For example, the first entry shows that \( n = 2525 \) animals would be required in order in order to have an 80% chance of detecting a 20% reduction in a background rate of 10%. That is, if the true background tumor incidence is 10%, a reduction of this rate by 20% \((0.2 \times 10\% = 2\%)\) to 8% in dosed animals would require \( n = 2525 \) animals in each group. Clearly, such an experiment probably would not be conducted under these conditions. At the other extreme, the last entry in Table 1 shows that for a tumor type that has a high incidence of 50% in the controls, that a 50% reduction in this rate to a tumor incidence of 25% in the dosed group would require only 45 animals. This certainly is a reasonable size for an experiment, but recall that the probability of a false negative is still 20%. Also, it is quite unlikely that a small dose of a carcinogen will reduce the tumor incidence rate by half.

Examination of Table 1 reveals that in order to conduct an experiment to show, e.g., a reduction in tumor incidence at low doses with less than 100 animals per group will require a tumor type with a background incidence of about 30% or more. Further, it will not be possible to detect subtle reductions in tumor incidence rates at low doses. Sizable reductions from the background incidence greater than 30% are required in order to be detected. Thus, the demonstration of the reduction of tumor incidence at low doses is restricted to the unlikely event of sizable reductions in tumor incidence.

**Table 1**

<table>
<thead>
<tr>
<th>True percentage reduction in the incidence in controls</th>
<th>20</th>
<th>30</th>
<th>40</th>
<th>50</th>
</tr>
</thead>
<tbody>
<tr>
<td>True incidence</td>
<td>10%</td>
<td>20%</td>
<td>30%</td>
<td>40%</td>
</tr>
<tr>
<td>10%</td>
<td>2525</td>
<td>1137</td>
<td>675</td>
<td>444</td>
</tr>
<tr>
<td>20%</td>
<td>1061</td>
<td>482</td>
<td>288</td>
<td>192</td>
</tr>
<tr>
<td>30%</td>
<td>547</td>
<td>256</td>
<td>154</td>
<td>104</td>
</tr>
<tr>
<td>40%</td>
<td>330</td>
<td>155</td>
<td>94</td>
<td>63</td>
</tr>
<tr>
<td>50%</td>
<td>150</td>
<td>94</td>
<td>63</td>
<td>45</td>
</tr>
</tbody>
</table>

Note. Testing was only for determining decreases in incidence; \( P \), total \( p \), one-sided; \( n \), sample size.
Sample Size for Tests with Continuous Data

Development of biologically based models that describe the modes of action for processes producing J-shaped dose responses may lead to measures of effects on continuous scales. This provides more precise information that can be used to detect more subtle differences at low doses than is possible by direct observation of incidence rates. For example, if a critical process that affects the carcinogenic process at low doses can be identified, this dose response curve can be used to demonstrate a J-shaped curve and to predict the ZED. Sample sizes \( n \) are displayed in Table 2, which are required in order to have an 80% chance to detect changes in continuous measures. The changes are expressed in terms of a shift of the mean response from the 50th percentile to some specified percentile of the distribution of the measurements among control animals. It is assumed that the measurements or a transformation of the measurements, e.g., logarithm of the measurement, are normally distributed. The shift in the mean response at low doses then can be expressed in terms relative to the size of the standard deviation or coefficient of variation of the response among animals.

For example, suppose the distribution of responses among control animals is normal, with a mean response (50th percentile) of 100 units and a coefficient of variation of 20%, i.e., a standard deviation (\( \sigma \)) of 20 units. The 40th percentile of this distribution is approximately 0.25 \( \sigma \) below the mean \( [100 - (0.25)(20)] = 95 \) units. If a low dose of a chemical shifts the true mean response from 100 to 95 units, with 200 animals in the control group and 200 animals in the dosed group, there is an 80% chance of obtaining a statistically significant \( P < 0.05 \) with a one-sided \( t \)-test. If a low dose of a chemical shifts the mean response to a lower level corresponding to the 31st percentile of the control animal distribution (0.5 \( \sigma \) below the mean), i.e., to \( [100 - (0.50)(20)] = 90 \) units, then 51 animals per group would be required to have a relatively high probability (80%) of detecting this difference.

<table>
<thead>
<tr>
<th>Change of mean response expressed as a factor of the SD (( \sigma ))</th>
<th>( n )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.25</td>
<td>200</td>
</tr>
<tr>
<td>0.50</td>
<td>51</td>
</tr>
<tr>
<td>0.75</td>
<td>23</td>
</tr>
<tr>
<td>1.00</td>
<td>14</td>
</tr>
<tr>
<td>1.25</td>
<td>9</td>
</tr>
<tr>
<td>1.50</td>
<td>7</td>
</tr>
</tbody>
</table>

Note. Data indicates direction of change (increase or decrease) observed in the mean response; SD, standard deviation; \( \sigma \), change of mean; \( n \), sample size postulated by the biologically based model.

In general, experiments to detect J-shaped dose-response curves will involve smaller effects at low doses than are typically observed in toxicological studies conducted at higher doses. Thus, much larger numbers of animals per dose group will be required to investigate J-shaped dose response curves than are typically employed. It is more likely that a J-shaped dose response curve can be detected with a moderate number of animals for a surrogate (precursor) for cancer that can be quantified on a continuous scale rather than sole use of quantal tumor-incidence data.

Alternative Test for a J-Shaped Curve

The mathematical function, \( Y = f(d) \), relating a response (\( Y \)) to the dose (\( d \)) of a chemical is likely to be a complex function containing several parameters. An experimental design to fit and test this function by a goodness-of-fit test likely would require several dose groups. Since a segment of any smooth mathematical function can be approximated by a polynomial function, a simple approximation in the dose range from zero to around the ZED may be provided by a linear-quadratic function:

\[
Y = b_0 + b_1d + b_2d^2
\]  

If high doses are associated with an adverse effect, the sign of \( b_2 \) is positive. Then, a necessary and sufficient condition for Equation 4 to be a J-shaped curve is that the sign of \( b_1 \) is negative. Hence, a test showing that \( b_1 \) is statistically different from zero and negative, cautiously suggests a J-shaped curve.

Unless the nonmonotonic effect is pronounced, the approach based on a nonlinear model fit may possibly be the better method to detect nonmonotonic dose-response relationships, due to the generally large sample sizes required to detect effects from pair-wise comparisons. Fitting nonlinear models is not trivial and will require the use of statistical software. Unrestricted multistage and logistic models could be considered.

Experimental Design for Estimation of the ZED

One advantage of establishing a J-shaped dose response would be to estimate the no-effect level (ZED). Its lower confidence limit (LZED) could possibly serve as a point-of-departure for setting an RfD associated with a negligible risk. The RfD is calculated as described in Equation 1. This raises the question of the experimental design to use to estimate the ZED and obtain the LZED.

Presumably, an initial estimate (\( D_{0.5} \)) of the zero-equivalent dose can be obtained from prior information about the nature of the J-shaped, biologically based model. Specifically, let the model be represented by \( Y = f(d) \), where \( Y \) is the expected (average) response and \( f(d) \) represents a mathematical function of the dose (\( d \)). If \( Y_0 \) is the average response observed in unexposed control animals, an estimate of the ZED is obtained.
by setting \( Y_0 = f(d) \), and calculating the value of \( d \) that satisfies this equation, giving \( d = D_z \). Due to the uncertainty in this initial estimate of the ZED, it is suggested that an experimental design for the estimation of the ZED should at a minimum consist of a dose below, at, and above \( D_z \) at doses of \( D_z/k \), \( D_z \), and \( kD_z \), where \( k \) is a factor greater than one. For example, if \( k = 2 \), animals would be administered doses of \( D_z/2 \), \( D_z \), and \( 2D_z \). Hopefully, ZED will be within the range of \( D_z/k \) to \( kD_z \). Failure to span the ZED in this range does not necessarily preclude obtaining an adequate estimate of the ZED, but it would require an extrapolation. The choice of \( k \) under different conditions could be a topic for further research.

The function \( Y = f(d) \) is likely to be a complex function containing several parameters. An experimental design to fit and test this function by a goodness-of-fit test likely would require several dose groups. Three data points probably will not be adequate for this task, but may be adequate to estimate the ZED. Since a segment of any smooth mathematical function often can be approximated adequately by a polynomial function, a simple estimate of the ZED can be obtained from fitting a linear-quadratic function, as described in Equation 4, to these three data points and determining the dose where this function predicts a response equal to the average response \( (Y_0) \) observed in unexposed control animals. Fitting this function can be accomplished readily using the EPA benchmark dose software (BMDS), available online at www.epa.gov/ncea/bmds.html. In using this software program, select the “continuous” data category under the model type and, under the choice of a model, select “polynomial”. For each of the three doses, enter the dose level, average response, number of animals, and standard deviation. For the benchmark-response (BMR) type, choose the “point” option and set this equal to the average response for the controls \( (Y_0) \). BMDS provides an estimate of the benchmark dose (BMD) (in this case BMD = ZED) and its lower 95% confidence limit (any other confidence level can be specified). If the goodness-of-fit is improved by including the control animals, they could be added.

**ZED Example**

For example, suppose a preliminary estimate of the ZED \( (D_z) \) was 4.0 mg/kg body weight per day of a chemical and an experiment was conducted with 5 animals each at doses of 2, 4, and 8 mg/kg per day for the purpose of estimating the ZED. Further, suppose the average responses were 4.5, 5.0, and 7.0 units with standard deviations of 2.2, 2.9, and 2.4 units at these doses, respectively, and the average response for the controls was 5.9 units. Fitting a polynomial (linear-quadratic) function to these three data points and estimating the dose with the same average response of 5.9 units as the controls gave ZED = 6.1 mg/kg per day with a lower 95% confidence limit of LZED = 2.5 mg/kg per day. That is, there is a 5% chance that the true ZED is less than 2.5 mg/kg per day. This is a fairly well-behaved example with the ZED = 6.1 mg/kg per day reasonably close to the initial estimate of 4 mg/kg per day, resulting in the experimental doses spanning the ZED and a confidence limit within a factor of 2.5 of the estimate.

The issue of sample size for the estimation of ZED is much more difficult than determining sample sizes for testing hypotheses considered previously, where relatively simple formulas exist for that purpose. The precision of the estimate of ZED depends not only on the standard deviations, but also on the selection of dose levels in relation to the ZED, shape of the dose-response curve, and the average response of the controls. The simplest approach to studying the effect of sample size is to perform a sensitivity analysis for any given set of data by keeping all of the results the same while varying the sample sizes in the BMDS runs. In the above example, if the sample size were changed to 10 animals per group, while keeping all of the other data entries the same, the ZED = 6.1 mg/kg per day of course remains the same, but now a tighter confidence interval is obtained with the LZED = 4.4 mg/kg per day. Thus, raising the LZED from 2.5 mg/kg per day based on \( n = 5 \) to a value of 4.4 mg/kg per day based on \( n = 10 \) would give a higher point of departure for setting an RfD. It would have to be decided if this were worth the additional costs. That is, in this case a doubling of the sample size would result in increasing the estimate of a “safe” or “acceptable” dose by a factor of \((4.4 / 2.5) = 1.8\).

**Experimental Design for Both Testing and Estimation**

If it is desirable to both test for the existence of a J-shaped dose response and estimate the ZED, this can be accomplished by combining the two experimental designs discussed above. From a preliminary estimate \( (D_m) \) of the dose associated with the minimum response, it is suggested that \( n_1 \) animals be tested at doses of \( 0, D_m/k_1, D_m, \) and \( k_1D_m \), where \( k_1 \) is greater than one. As noted before, the value of \( n_1 \) required for testing for subtle effects at low doses will generally need to be considerably larger than typically used in toxicology studies. Also, from a preliminary estimate of the dose for the ZED \( (D_z) \), it is suggested that \( n_2 \) animals be tested at doses of \( D_z/k_2, D_z, \) and \( k_2D_z \), where \( k_2 \) is greater than one and may be equal to \( k_1 \). The value of \( n_2 \) likely will be smaller than \( n_1 \) and on the order of sample sizes typically used in toxicology studies.

If there is no duplication of doses, combining these two designs results in six dose groups plus controls and a total of \( N = (4n_1 + 3n_2) \) animals. This design provides data for testing for reductions from the controls to provide experimental evidence for a J-Shaped dose response and additional data for estimating the ZED that can be used for risk assessment and regulation. Further, with seven dose groups it may be possible to fit the biologically based model and perform a goodness-of-fit test. Deviations from the model may indicate modifications to the underlying processes that generated the model. As in other areas of science, it is unlikely that adequate answers will be forthcoming from early experiments and an iterative process of postulation and experimentation will evolve.
Depending on the shape of the dose-response curve and the initial estimates of the minimum and the ZED, the doses included to span the minimum and the doses included to estimate the ZED may overlap. Thus, it may not be necessary to include animals at the D_0/k_2 or perhaps even at D_0 if the doses of k_1D_m and perhaps D_m, that use more animals per dose, are in the same dosage range. Presumably, effects have been demonstrated previously at high doses and it is not necessary to include high doses in experiments to test for J-shaped curves or to estimate the ZED.

**Example: Formaldehyde**

Extensive studies have been conducted by CIIT Centers for Health Research on formaldehyde (CIIT, 1999). Fisher 344 rats were exposed to formaldehyde by inhalation at concentrations of 0, 0.7, 2.0, 6.0, 10, and 15 ppm for 6 h per day, 5 days per week, for up to two years. In this section, rates of cell proliferation will be examined. Increased values of the labeling index (ULLI) in the nasal respiratory epithelium were observed at 6.0, 10, and 15 ppm, but not at 0.7 or 2.0 ppm (Conolly et al., 2002).

At the time of sacrifice of an animal, measurement of the labeling index was obtained for several sites of the nasal respiratory epithelium. Measurements of the labeling index are analyzed here for the two lowest doses of formaldehyde exposures (0.7 and 2.0 ppm) and the controls at five sites: anterior lateral meatus (ALM), posterior lateral meatus (PLM), anterior medial septom (AMS), posterior medial septum (PMS), and medial maxilloturbinate (MMT). Measurements were made at 8 time points ranging from 1 day to 78 weeks after the initiation of formaldehyde exposure. A two-way crossed analysis of variance (three doses × eight time periods) was conducted for each site. The interaction mean square with (2/H11005 for the ith time period (i = 1, 2, ..., 8). Let w_i represent the weight given to the measurement for the ith time period, i.e., the proportion of the 2-year study covered by the ith time period. The length of a time period was taken to be the time-weighted average of the controls: 

\[ \text{time-weighted average of the controls: } w_i = \frac{\text{time (in weeks)}}{78} \]

For a given dose group, let L_i represent the labeling index for the ith time period (i = 1, 2, ..., 8). Let w_i represent the weight given to the measurement for the ith time period, i.e., the proportion of the 2-year study covered by the ith time period. The length of a time period was taken to be the time-weighted average of the controls: 

\[ \text{time-weighted average of the controls: } w_i = \frac{\text{time (in weeks)}}{78} \]

All of the time-weighted averages of ULLI in the 0.7- and 2.0-ppm exposure groups were consistently lower than the controls at each of the five tissue sites. With ULLI exceeding the controls at 6, 10, and 15 ppm of formaldehyde. No statistically significant reductions in the labeling index were detected in the ALM, AMS, or MMT sites. The time-weighted ULLI averaged across the five tissue sites were 10.7, 8.2, 6.8, 12.5, 40.3, and 67.6 at formaldehyde concentrations of 0, 0.7, 2.0, 6.0, 10.0, and 15.0 ppm, respectively.

For PLM, the time-weighted averages for the ULLI were 8.97, 6.64, and 7.05 for the controls, and the 0.7- and 2.0-ppm groups, respectively. The estimate of the variance with 14 degrees of freedom from the two-way analysis of variance was 

\[ s^2 = \frac{10.45 - 3.025d^2}{2} \] 

The t-test for the comparison of the 0.7-ppm group with the controls is: 

\[ t = \frac{[8.97 - 6.64] - 2.7819}{0.2514} = 1.97 \]

Comparing this value with the one-sided t-table values with 14 degrees of freedom yields a false positive probability of \( P = 0.037 \). When comparing two dose groups with controls, the procedure suggested by Holm (Holm, 1979) indicates that the largest difference should be tested at the 0.05/2 = 0.025 level of statistical significance in order to preserve an overall false positive rate of \( P = 0.05 \) for the PLM site. Under this restriction, the reduction in the ULLI at 0.7 ppm is not statistically significant, nor is the reduction at 2.0 ppm at the PLM site.

Large reductions in ULLI were measured at the PMS site. The time-weighted averages were 22.86, 15.60, and 10.16 for the controls, 0.7, and 2.0 exposure groups, respectively, with \( s^2 = 26.8699 \). The largest reduction from the controls occurred for the 2.0 ppm group with \( t = 3.45 \) which is highly significant with \( P = 0.003 \). Statistical tests at two doses and five tissue sites—10 possible comparisons—increase the chance of a false-positive response. According to the procedure by Holm (Holm, 1979), in order to preserve an overall false positive rate of \( P \leq 0.05 \), the most significant test would require \( P \leq (0.05/10) = 0.005 \), which is achieved here. Hence, the 2.0-ppm group demonstrated a statistically significant reduction in ULLI from the controls at the PMS site.

In order to illustrate the curve-fitting procedure, results for the average time-weighted ULLI across the five tissue sites were utilized. The average ULLI were: 10.70, 8.18, 6.85, and 12.51 at formaldehyde concentrations of 0, 0.7, 2.0, and 7.0 ppm, respectively. A polynomial model fit to these results, using the computer software S-Plus, gave a J-shaped curve:

\[ ULLI = (10.45 - 3.025d + 0.56171d^2) \]

where \( d \) represents the formaldehyde concentration. The estimate of the coefficient of the linear term was negative (−3.025), with a standard error of 1.487. The t-test was \( t = (-3.025 / 1.487) = -2.03 \), with 12 degrees of freedom, gave a one-sided \( P < 0.05 \). Since a single test was involved with this approach, no correction for multiple comparisons was required. The statistically significant negative linear coefficient suggests a J-shaped curve. Also, the ULLI averaged across the five sites for a formaldehyde concentration of 2.0 ppm showed a statistically significant \( (P < 0.025) \) reduction from the control average. The estimate of the concentration of formaldehyde where the average ULLI across the five tissue sites was equal to the controls, i.e., the ZED occurred at a formaldehyde concentration of 5.4 ppm, with a lower 95% confidence limit of 2.7 ppm.

For a cancer-risk assessment for formaldehyde, these data support the hypothesis that the threshold-type dose response
for nasal tumor incidence is the result of a minor genotoxicity at low dose that is superimposed by a J-shaped dose response for tumor promotion by cell proliferation at high cytotoxic dose levels (Lutz, 1998). At low dose, the incremental DNA damage may be cancelled out by a reduction in cell proliferation.

**DISCUSSION**

Identifying a nonmonotonic dose-response curve in the low-dose range can lead to an estimate of a no-effect level, ZED. A lower confidence limit for the ZED can be used as a point of departure for setting a reference dose (acceptable daily intake). However, establishing a nonmonotonic dose-response curve in the low-dose range can be problematic. Toxicology studies often are conducted only at high doses in order to screen for potential toxic effects requiring relatively small numbers of animals. Nonmonotonicity at low doses may only reflect relatively small effects. Designing experiments to investigate low-dose effects, while possible, will require theoretical calculations from biologically based models and/or experimental data in order to identify the general range of doses where nonmonotonicity occurs.

The precision of quantal data, such as tumor incidence, generally will be inadequate to detect subtle effects without using prohibitively large numbers of animals. Biomarkers for cancer utilizing continuous measurement data, such as the labeling index for cell proliferation or DNA adduct levels, may be adequate to establish nonmonotonic dose-response curves. However, such studies generally would require lower doses and more animals per dose than typically are used in screening studies for toxicity. When nonmonotonicity can be demonstrated with statistical tests, then an estimate can be justified of the no-effect level equivalent to the background response in controls. This zero-equivalent dose could then be utilized for improved risk assessments. In setting an RfD, presumably a smaller uncertainty factor could be used with the LZED compared to a point-of-departure based on a benchmark dose associated with some level of risk. If linear extrapolation is deemed necessary, presumably extrapolation to the LZED could replace extrapolation to zero.

**ACKNOWLEDGMENTS**

This work was funded by the United States Air Force through subcontract Number 740889.3000-00 with Parsons Engineering Science, Inc., and CIIT Centers for Health Research.

**REFERENCES**


