Regulatory Toxicology and Pharmacology **33,** 37–59 (2001)
doi:10.1006/rtph.2000.1453, available online at http://www.idealibrary.com on IDEAL®

Evaluation of Subchronic Toxicity Data Using the Benchmark Dose Approach

L. A. Gephart,*,[1] W. F. Salminen,* M. J. Nicolich,† and M. Pelekis*

*Toxicology and Environmental Sciences Division and †Occupational Health and Epidemiology Division, ExxonMobil Biomedical Sciences, Inc., P.O. Box 971, 1545 Route 22 East, Annandale, New Jersey 08801-0971

Received June 1, 2000

We used the benchmark dose (BMD) methodology devised by Crump (*Fundam. Appl. Toxicol.* 4, 854–871, 1984) to estimate BMDs for 90-day toxicological data and several fabricated data sets. From a toxicological perspective, dose–response modeling offers certain advantages over using a point estimate, such as the currently used no-observable-adverse-effect level (NOAEL) approach. However, there are many variables associated with the BMD that could be set to produce unreasonable BMD estimates. Some of these variables and decisions are examined in this study. BMDs were calculated for discrete and continuous endpoints using a variety of different variables (e.g., maximum likelihood estimates [MLEs], lower-confidence limits [LCLs], and different risk levels). In addition, the fabricated data sets were manipulated (i.e., dose groups eliminated) and the BMDs recalculated. This process tested how the BMD estimates varied using different forms of the data. For the 90-day toxicological studies, the BMDs were typically within an order of magnitude of the NOAEL for discrete endpoints. For the discrete endpoints, the MLEs were typically greater than the NOAEL and the LCLs were typically less than the NOAEL. The BMD was insensitive to changes in the data points one to two dose groups beyond the NOAEL/LOAEL. With the continuous data, the ratios of MLEs and LCLs to the NOAEL were highly variable, and no general trend could be determined. The BMD methodology offers potential improvements in the risk assessment process since dose–response characteristics are used to calculate the BMD. Depending upon how the BMD is defined, i.e., the form of the dose–response model, and how the BMD is used in the risk assessment process, BMD estimates may produce reference doses/concentrations that are more or less conservative than the NOAEL approach. Active involvement in discussions with regulatory agencies is needed to ensure that inappropriate models and unreasonable BMDs are not used. In addition, further discussions on how BMDs should be used in the risk assessment process are needed.

Key Words: **dose–response modeling; BMD; subchronic studies; discrete and continuous endpoints.**

INTRODUCTION

The United States Environmental Protection Agency (EPA) uses reference doses/concentrations (RfD/Cs) in the risk assessment of noncancer chemicals. The RfD/C is an estimate of the daily exposure that a person, including sensitive individuals, can be exposed to without experiencing adverse effects. Traditionally, the RfD/C is calculated from the NOAEL for a critical effect. The no-observable-adverse-effect level (NOAEL) is the highest experimental dose that failed to cause an adverse effect. Despite the tradition of using the NOAEL, there are many arguments in favor of using dose–response modeling in the risk assessment of noncancer chemicals.

Comments favoring the use of the NOAEL include the following: (i) it is easy to understand, (ii) it has intuitive appeal, (iii) it is not dependent on the assumed dose–response model, and (iv) it can be used for continuous or discrete responses. Typical arguments against its use include the following: (i) it is limited to the doses tested in the experiment, (ii) it rewards poor experimental design, (iii) it does not provide an indication of the magnitude of the risk, and (iv) it does not provide a slope of the dose–response curve. These flaws have spurred the development of more detailed methods, such as the benchmark dose (BMD), to estimate the risk associated with a given exposure. However, few studies have been used in the validation of these methods.

DOSE–RESPONSE MODELING

Dose–response modeling is a commonly accepted method of estimating the response associated with a given exposure. Crump (1984) suggested that dose–response

[1] To whom correspondence should be addressed. Fax: 908-730-1198. E-mail: lagepha@erenj.com.

0273-2300/01 $35.00

modeling should be applied to noncarcinogens and proposed the benchmark dose methodology. Since then, many others have advocated this approach including the EPA (1995). Dose–response modeling for noncarcinogens is similar to that used for carcinogens, but extrapolation to extremely low risk or dose levels is not performed. Typically, extrapolation is restricted to doses associated with a 1–10% response rate. When dose extrapolation is restricted to responses in the "observable range," all dose–response models are supposed to yield similar results. A few studies have tested this hypothesis and concluded that the BMD methodology is robust. Unfortunately, these studies used a narrow range of endpoints from developmental toxicity studies, a limited number of dose–response models, and models with similar mathematical forms.

There are many desirable features that are believed to be associated with dose–response modeling of noncarcinogens. These include the following: (i) the dose associated with any response rate can be calculated, (ii) the slope of the dose–response curve can be calculated, and (iii) well-designed experiments are rewarded since they generate higher BMDs. However, the behavior of dose–response models with various nondevelopmental data has not been investigated.

In this study we tested the limits and behavior of the BMD methodology using data from a variety of 90-day subchronic toxicity studies which are routinely used for the regulation/registration of chemicals. Therefore, examining the behavior of the BMD methodology with these data sets represents a real-world application. The behavior of the methodology was also tested using fabricated discrete data sets, which were manipulated (i.e., certain dose groups removed) in order to determine how the BMD estimates change with alterations of the data.

COMPARISON OF THE NOAEL AND BENCHMARK DOSE METHODOLOGIES

1. Calculation of a NOAEL

The NOAEL is defined as the highest dose that fails to cause an adverse effect. The term adverse has a great deal of uncertainty associated with it. Expert judgement is used to determine whether an effect is considered adverse (i.e., biologically significant). Often, biological parameters will be altered, but those alterations will be considered side effects of the treatment. For example, animals fed a large quantity of nonnutritive chemical in the diet often lose weight. The weight loss is due to the decrease in nutritive food consumption and not to the chemical itself. In this case, the weight loss is not considered adverse, since it is not a direct effect of the chemical. Less clear-cut associations between dose and effect are often found. In these situations, the NOAEL may depend upon the expert interpreting the data. For these reasons, a great deal of uncertainty is often associated with the NOAEL. Statistical approaches have been used to estimate NOAELs (Faustman *et al.*, 1994), but a statistically significant finding does not necessarily mean that the finding is biologically relevant. Therefore, expert judgement remains an integral part of calculating a NOAEL.

TABLE 1
Ninety-Day Inhalation Study of Ethylene Glycol Monohexyl Ether

Vapor concentration (ppm)	Serum ALT activity (I.U./L)	Mean liver/body weight (%)	Mean kidney/body weight (%)
0 (control)	55	3.08	0.69
20	61	3.23	0.71
41	50	3.35*	0.72
71	33*	3.57*	0.76*

Note. * Denotes values statistically different from control.

a. Example of the role of scientific judgement in selecting a NOAEL. Data from a 90-day inhalation toxicity study are shown in Table 1. Alanine aminotransferase (ALT) activity is an indicator of liver damage, with higher values representing greater liver damage. Values with an asterisk are statistically different from the control value. The role of expert judgement in calculating a NOAEL for these data follows. The NOAEL for this study was determined to be 41 ppm. This is not obvious from the statistical analysis, since mean liver/body weight was increased at 41 ppm. A change in relative liver weight was not considered strong evidence of an adverse effect. Instead, the weight of the evidence was used to determine the NOAEL. At 71 ppm, relative kidney and liver weight were increased; therefore, this vapor concentration was considered to cause an adverse effect since both organs were affected. To confuse matters further, ALT activity actually decreased at 71 ppm, indicating that higher vapor concentrations prevented liver damage. However, ALT levels were given less weight than increases in organ weight, so the conclusion was that 41 ppm was the NOAEL and 71 ppm was the lowest-observed-adverse-effect level (LOAEL). As shown in the previous example, the NOAEL is not always a clear-cut dose that is without an effect. Expert judgement plays a pivotal role in determining what effects are considered adverse. Although the BMD methodology can model the dose-response, expert judgement is still essential for determining what effects are considered adverse.

2. Calculation of a BMD

In contrast to the NOAEL, the BMD is not limited to the experimental doses. The BMD is the dose producing

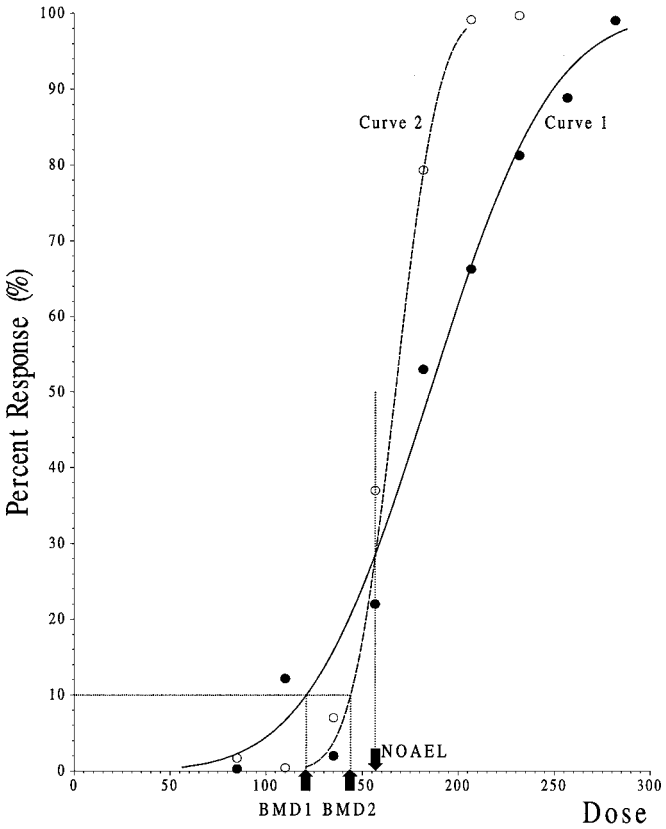


FIG. 1. BMDs for two dose–response curves that have the same NOAELs.

a predetermined change in response (i.e., the benchmark response [BMR]) and is calculated by statistically fitting a dose–response model to experimental data. The response rate associated with any dose can be estimated from the dose–response curve.

a. Choice of effects to model. As with the NOAEL, the effects to be modeled must be chosen. Therefore, expert judgement plays a critical role in calculating a BMD. Since the BMD is dependent upon the slope of the dose response, effects that have the same NOAEL are likely to have different BMDs. For effects that have the same NOAEL, the effect with the steeper dose-response will have a smaller BMD. This is illustrated in Fig. 1. Since the slope is used in the BMD estimation, it is possible that endpoints that do not dictate the NOAEL may dictate the BMD.

b. Data format. BMD models exist for modeling discrete and continuous data. Discrete data are typically binary—effect or no effect. For example, if an animal develops a tumor it is considered affected; if it does not, it is considered not affected. Discrete data do not give a measure of the magnitude of effect. Discrete data can consist of more than two categories; however, all discrete data must be transformed into binary format before they can be modeled with currently used quantal BMD models. Continuous data are not restricted to groupings and the magnitude of effect is represented in the data. An example of continuous data is serum ALT activity and liver/body weight (see Table 1). Continuous data can be transformed into quantal data using thresholds; however, this process introduces uncertainty into the final results.

c. Model selection. A variety of dose–response models are available for calculating a BMD. The models can be separated into those that model discrete data and those that model continuous data. The models are further separated based upon the mathematical equation they use. A few examples are given,

Quantal data BMD models

Quantal linear regression	$P(d) = c + (1 - c) \times \{1 - \exp[-q_1(d - d_0)]\}$
Quantal Weibull	$P(d) = c + (1 - c) \times \{1 - \exp[q_1(d - d_0)^k]\}$
Log-normal	$P(d) = c + (1 - c)N \times (a + b \log[d - d_0])$

Continuous data BMD models

Continuous linear regression	$m(d) = c + q_1(d - d_0)$
Continuous polynomial regression	$m(d) = c + q_1 d + \cdots + q_k(d - d_0)^k$
Continuous power	$m(d) = c + q_1(d - d_0)^k,$

where $P(d)$ is the probability of a response at the dose d; $m(d)$ is the mean response at the dose d, c and $q_1, \ldots, q_k$ are parameters estimated from the data, $N(x)$ denotes the cumulative normal distribution function, and d_0 is the threshold dose.

A highly touted strength of the BMD methodology is that different models should calculate similar BMDs for the same data set. This is because dose–response estimates are restricted to response rates in the observable range. For extremely low response rates outside the observable range, the BMDs become highly model dependent. However, the behavior of different BMD models has not been thoroughly tested. A summary of available models from K. S. Crump Associates is given in Appendix A.

d. Model fit and error estimation. Once selected, the dose–response model is fit to the data. The model is typically fit to the data using maximum likelihood methods. The maximum likelihood estimate (MLE) of a dose is the point on the estimated dose–response curve corresponding to a given response. The fit of the model to the data can be determined through a variety of statistical tests of fit, but not a single test is universally accepted. The degree of fit is an indication of how well the model reflects the data. It is possible to eliminate data points or use different fitting models to improve model

fit. However, just because a model fits the data does not mean that it accurately models the dose-response in the region of interest, usually the 1–10% response area. For example, a model may closely fit the data in the high-dose/high-exposure region of the experiment, but provide a poor fit at the low-response region where the BMD is usually estimated. Most summary estimates of model fit consider an average fit, rather than the fit in a specific region of the model. For this reason, we recommend always plotting the data and the estimated model to assess the adequacy of the fit in the region of the estimated BMD.

The adequacy of the fit of the model to the data is usually measured by the magnitude of the log-likelihood function. The log-likelihood function is a number which reflects how close the fitted model is to the observed data. The closeness is measured by the likelihood, which the estimation seeks to maximize (hence, maximum likelihood estimation). To simplify calculations, the method considers the log of the likelihood, which, because the likelihood is less than 1, is a negative number. The less negative the log of the likelihood is the greater the likelihood, so to maximize the likelihood one chooses an estimated model that maximizes the log of the likelihood function. When comparing two estimated models, the one with the larger log-likelihood function is preferred. It is also possible to statistically estimate the improvement of one model over another or the improvement of a model over no change in response over dose. These statistics are based on the score statistic (Breslow and Day, 1980). The uncertainty associated with the dose–response curve is represented by confidence intervals or confidence limits (CLs) placed around MLEs of dose or risk. The EPA has endorsed the use of CLs for deriving the BMD (EPA, 1995). The most common method of estimating CLs is based on the asymptotic methods using the likelihood function (Crump and Howe, 1985); however, other techniques such as bootstrapping can be used. The 95% confidence interval of an estimated dose for a particular risk represents a range of values that bracket the true dose in 95% of the experiments using similar techniques. The more certain the probability of inclusion (the higher the confidence level), the wider the interval is. The limits usually range from 90% (less certain) to 99% (more certain), with 95% being the most common. The 95% CL is typically used with the LMS for cancer dose–response modeling (Beliles and Shulz, 1994).

e. Definition of risk and level of response. Two additional variables must be accounted for before a BMD can be estimated. First, the definition of risk must be defined. Two common ways of expressing risk are additional risk and extra risk (Gaylor and Slikker, 1990). Additional risk is defined as

$$\mathrm{AR}(d) = P(d) - P(0)$$

and extra risk is defined as

$$\mathrm{ER}(d) = [P(d) - P(0)]/[1 - P(0)],$$

where $P(d)$ is the probability of response at dose d, and $P(0)$ is the probability of response in the absence of exposure or the background response ($d = 0$).

The benchmark response rate (i.e., the BMR) is the other value to be specified, and it represents the response rate for which the corresponding dose will be estimated. The BMD approach avoids the pitfall of extrapolation to low response rates by restricting the BMR to responses in the observable range. A BMR of 1–10% is typically within the observable range of most experiments (EPA, 1995; Foster and Auton, 1995). In an experiment using 10 animals per group, a 10% response rate could be detected (i.e., 1 in 10 animals would show a response). To detect a 1% response, the group size would have to be increased to 100. The EPA has discouraged extrapolating below the 1–10% response range. However, they stated that it is possible to use the BMD methodology to extrapolate to lower levels of risk as long as the inherent instability of the approach is recognized (EPA, 1995). This approach seems to be at odds with the major benefit of the BMD, staying within the observable range.

f. BMD calculation. Once the BMR is chosen, the dose associated with the BMR is determined using the dose–response curve. The dose corresponding to a given BMR is considered the BMD. The BMD is then used in place of the NOAEL to derive an RfD/C. However, there are several ways of estimating the BMD from the dose–response curve. The EPA uses a lower confidence limit (LCL) on the dose to determine the BMD. The BMD based on the LCL is also referred to as the lower confidence limit of the effective dose (LED). The effective dose (ED) is simply the MLE point-estimate at a given response rate. The EPA prefers the LED over the ED since the LED accounts for uncertainty associated with calculation of the ED. The use of the LCL and MLE dose–response curves to derive an ED_{10} and an LED_{10} (i.e., the ED or LED associated with a 10% response rate) is shown in Fig. 2. Murrell *et al.* (1998) suggested that LCLs should not be used to determine the BMD. They felt that the size of the study should not bias a characteristic dose statistic in either direction. Instead, they suggested that the MLE should be used to determine the BMD.

ISSUES SURROUNDING THE NOAEL AND BMD APPROACHES

The NOAEL is often criticized as being archaic. It lacks most of the information associated with dose–response curves (e.g., number of animals affected and dose–response slope). Therefore, the magnitude of the effect and margin of safety cannot be assessed using

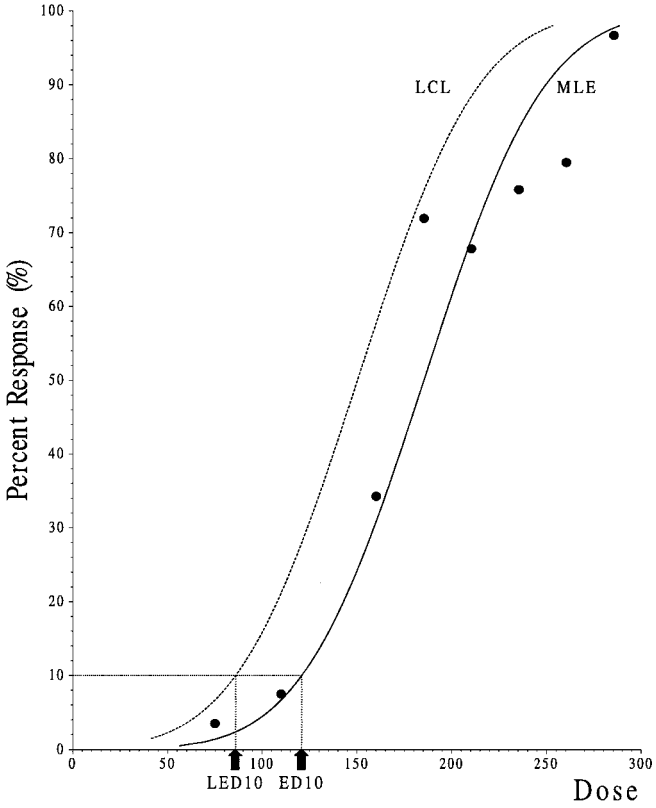


FIG. 2. Estimation of an ED_{10} and a LED_{10}.

a NOAEL. To overcome these inadequacies, dose–response modeling has been advocated for estimating RfD/Cs. The most popular of these techniques is the BMD methodology, which is endorsed by the EPA (1995). The following list highlights some of the benefits that are believed to be associated with the BMD versus the NOAEL.

1. Theoretical Benefits of the BMD vs NOAEL

- The BMD is unrestricted, which means it can be calculated for any response rate. In contrast, the NOAEL is restricted to the experimental doses.
- The BMD is less sensitive to dose level selection.
- The BMD provides a slope of the dose–response curve and the uncertainty associated with the BMD. These values can be used to improve the risk assessment process.
- The BMD encourages better experimentation, since larger group sizes should result in higher BMDs when based on the LCL. This is because larger groups decrease the uncertainty associated with the dose–response curve. In contrast, larger group sizes typically decrease the NOAEL since lower response rates can be detected.
- A BMD can be calculated from an experiment lacking a NOAEL. Therefore, repeat of experiments lacking a NOAEL may not be necessary.

Theoretically, the BMD is superior to the NOAEL. However, even if rigorous testing shows that the BMD methodology is robust, there are many issues associated with the BMD methodology that greatly influence BMD estimates. The following list highlights some of these issues.

2. Model and Data Selection

- What model(s) should be used to calculate the BMD?
- Should a threshold parameter be incorporated into the model? Noncarcinogens often have thresholds below which no effect occurs.
- Should a background parameter be incorporated into the model? Many effects have background response rates independent of exposure.
- How should risk be defined? Extra risk. Additional risk.
- What should the BMR be set at? One to 10% is believed to be reasonable.
- Should continuous data be transformed into quantal data prior to modeling?
- How should the risk associated with continuous responses be defined? Percentage of change relative to control value. Percentage of change relative to the maximum response expected.

3. Uncertainty Calculation and Model Fit

- How should the uncertainty associated with the dose–response curve be calculated? Asymptotic methods of Crump and Howe (1985). Bootstrapping techniques.
- How important is model fit? The EPA (1995) suggested that the model with the best fit should be used.
- Should data points be eliminated to provide a better fit to the model? Saturation of a response at high doses often results in poor model fit.
- Can correlated effects be accurately modeled?
- What is the minimum number of responding groups required for calculation of a BMD?

0.0.1. 4. Risk Assessment and the BMD

- Should the slope of the dose–response curve be used in risk assessment?
- Should the BMD be defined as the MLE or the LCL?
- Should the BMD be compared against the NOAEL?
- Should multiple effects be modeled or only the effect with the lowest NOAEL? An effect with a low NOAEL may have a higher BMD than an effect with a higher NOAEL and a steeper dose–response slope.
- Should extrapolation to extremely low BMRs be done?

• What uncertainty factors should be applied to the BMD to derive an RfD or RfC?

BMD CALCULATIONS

A. 90-Day Toxicity Studies

1. General Approach

We derived BMDs for typical 90-day toxicological studies that were available in the published literature. These studies are reviewed in Appendix B. We chose these types of studies since they are often used for the regulation and hazard evaluation of chemicals. If the BMD methodology is accepted, it will probably be routinely applied to these types of data. The selected studies used inhalation or oral exposures and a NOAEL could be established for all of them. In addition, all of the studies had at least one dose level above the NOAEL. No restriction was placed upon the type of endpoint that dictated the NOAEL; however, organ weights and/or histopathology were typically the critical endpoints.

We used two different discrete (a polynomial and a Weibul with threshold parameters known as THRESH and THRESHW, respectively) and two different continuous BMD models (a multistage and a Weibul with threshold parameters known as THC and THWC, respectively). These four models and others are reviewed in Appendix A. The NOAEL was determined for each study, and the critical effect(s) was used to determine the BMD. We did not calculate BMDs for endpoints that did not dictate the NOAEL. In addition, the data format for some of the critical endpoints was not adequate for calculating a BMD (e.g., standard deviations were not available). In these cases, the BMDs were calculated using the remaining critical endpoint(s). The MLE and 90, 95, and 99% LCLs were determined for the critical effect(s) at BMRs of 1, 5, and 10%. All BMDs were calculated using additional risk.

a. Discrete endpoints.. Table 2 summarizes each study in terms of the critical effect(s), species, route of exposure, number of dose levels, and the NOAEL. The MLE and 90, 95, and 99% LCL estimates calculated at BMRs of 1, 5, and 10% are shown in Figs. 3 and 4. Only results from the THRESH model are shown. Tabulated modeling results for both THRESH and THRESHW can be found in Tables 3 and 4. For the majority of data sets, the two BMD models gave similar MLE and similar LCL estimates and provided reasonable fits to most of the data (Tables 3 and 4). However, for many datasets, the BMD models provided perfect fits ($P = 1.0$) to the data since only two dose groups had nonzero responses, and the model perfectly predicted the observations (i.e., the model drew a straight line between two data points). The MLEs were typically within an order of magnitude of the LCLs. Defining the BMD as the 95% LCL at the 10% risk level, the BMDs for all of the discrete endpoints were within an order of magnitude of their respective NOAELs and within two to three times the NOAEL for the majority of data sets.

Summary.

• Model fit was reasonable for some data sets, but unrealistic for many data sets (i.e., $P = 1.0$), since only two doses had nonzero responses.

• For most data sets, the MLE was within an order of magnitude of the LCL.

• The BMDs calculated as a 95% LCL at a 10% risk level were within an order of magnitude of the corresponding NOAELs and within two to three times the NOAEL for the majority of data sets.

b. Continuous endpoints. Table 5 summarizes each study in terms of the critical effect(s), species, route of exposure, number of dose levels, and the NOAEL. The MLE and 90, 95, and 99% LCL estimates at BMRs of 1, 5, and 10% for these studies are shown in Figs. 5–7. Only results from the THC model are shown. Tabulated modeling results for both the THC and THCW can be found in Tables 6 and 7. For this analysis, we assumed that risk level was directly proportional to the percentage of change in the continuous variable

TABLE 2
Toxicological Studies with Discrete Endpoints

Agent	Discrete endpoint(s) dictating NOAEL	Species and sex	Route of exposure	NOAEL	Doses
Acrylamide (acryl)	• Peripheral nerve damage	Male and female rats	Drinking water	1 mg/kg/day	0, 0.05, 0.2, 1, 5, 20
Cyclohexylamine hydrochloride (cyclo)	• Kidney tubular atrophy	Male rats	Diet	600 ppm	0, 600, 2000, 6000
Hexachloroethane (hexa)	• Kidney atrophy	Male and female rats	Diet	1 mg/kg/day	0, 1, 15, 62
1,2-Epoxybutane (epoxy)	• Nasal inflammation	Male and female rats	Inhalation	400 ppm	0, 50, 100, 200, 400, 800
Ethylene oxide (ethyl)	• Rhinitis	Male and female mice	Inhalation	100 ppm	0, 100, 200, 400, 600
TCE (T)	• Liver lesions • Kidney lesions	Male and female mice	Inhalation	100 ppm	0, 100, 200, 400, 800, 1600
1,3,5-Trichlorobenzene (TCB)	• Nasal hyperplasia	Male and female rats	Inhalation	100 mg/m^3	0, 10, 100, 1000

TABLE 3
BMD Analysis of Discrete Endpoints Using a 95% LCL

Agent (sex)	Model fit (P value)		MLE01		LCL01		MLE05		LCL05		MLE10		LCL10	
	T	TW	T	TW	T	TW	T	TW	T	TW	T	TW	T	TW
Acrylamide (M)	1.00	1.00	1.64	2.09	0.09	0.24	2.14	2.49	0.47	0.60	2.46	2.75	0.82	0.88
Acrylamide (F)	1.00	1.00	2.15	2.28	0.09	0.15	2.83	2.87	0.47	0.54	3.23	3.24	0.82	0.88
Cyclohexylamine hydrochloride (M)	0.49	0.49	621	621	391	354	706	706	490	488	817	817	614	614
Hexachloroethane (M)	0.79	0.69	0.16	0.23	0.08	0.07	0.84	1.01	0.39	0.38	1.72	1.95	0.80	0.79
Hexachloroethane (F)	1.00	1.00	2.63	3.32	0.46	0.46	8.42	9.06	2.37	2.37	15.0	15.0	4.87	4.87
1,2-Epoxybutane (M)	1.00	1.00	452	567	166	248	502	611	292	338	530	632	488	385
1,2-Epoxybutane (F)	1.00	1.00	452	567	166	248	502	611	292	338	530	632	488	385
Ethylene oxide (M)	0.96	1.00	125	135	9.0	40.7	147	154	45.3	76.2	161	165	85.4	100
Ethylene oxide (F)	0.79	1.00	102	124	13.8	44.5	120	136	56.3	72.7	130	144	83.6	90.1
TCE (M)—liver	0.99	0.99	52.3	52.3	9.37	9.37	116	116	47.9	47.9	200	200	98.5	98.5
TCE (F)—liver	0.18	0.40	87.5	97.6	6.16	6.15	119	128	31.4	31.4	161	168	64.6	64.5
TCE (M)—kidney	1.00	1.00	112	176	14.1	38.9	126	179	62.9	70.8	136	181	89.0	91.9
TCE (F)—kidney	0.00	0.00	102	102	2.17	2.17	109	109	11.1	11.1	118	118	22.8	22.8
Trichlorobenzene (M+F)	1.00	1.00	560	866	51.5	51.5	826	957	261	263	1000	1000	496	540

Note. F, female; M, male; MLE, maximum likelihood estimate; LCL, lower confidence limit; T, THRESH; TW, THRESHW.

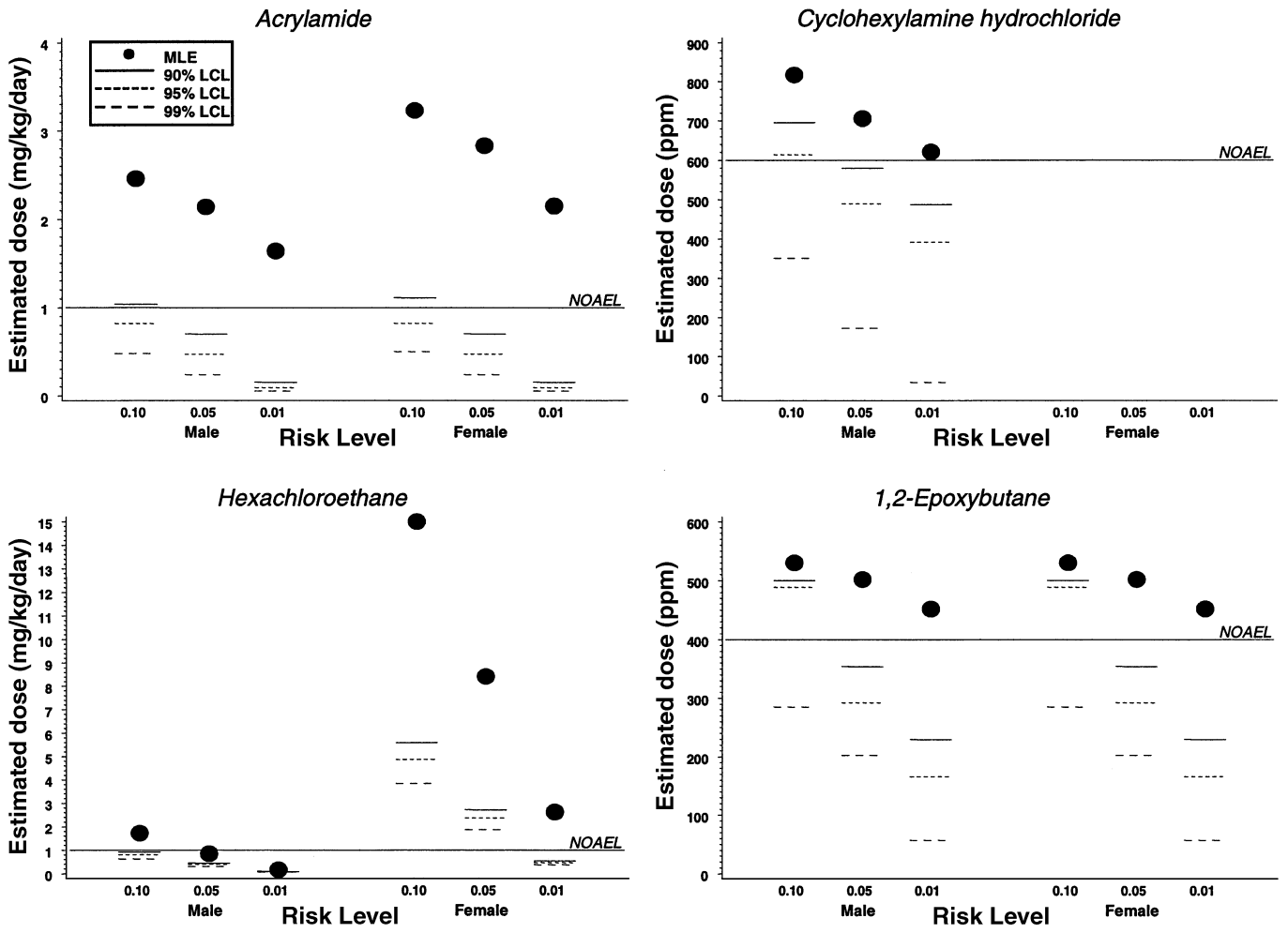


FIG. 3. Modeling results using THRESH for chemicals with discrete endpoints.

TABLE 4
BMD Analysis for Discrete Endpoints at a 10% Risk Level

Agent (sex)	MLE		90% LCL		95% LCL		99% LCL	
	T	TW	T	TW	T	TW	T	TW
Acrylamide (M)	2.46	2.75	1.04	1.04	0.82	0.88	0.48	0.55
Acrylamide (F)	3.23	3.24	1.11	1.15	0.82	0.88	0.50	0.52
Cyclohexylamine hydrochloride (M)	817	817	695	695	614	614	351	348
Hexachloroethane (M)	1.72	1.95	0.92	0.90	0.80	0.79	0.62	0.62
Hexachloroethane (F)	15.0	15.0	5.60	5.60	4.87	4.87	3.85	3.85
1,2-Epoxybutane (M)	530	632	500	404	488	385	285	327
1,2-Epoxybutane (F)	530	632	500	404	488	385	285	327
Ethylene oxide (M)	161	165	107	114	85.4	100	45.9	75.2
Ethylene oxide (F)	130	144	101	101	83.6	90.1	42.3	63.6
TCE (M)—liver	200	200	109	109	98.5	98.5	82.3	82.3
TCE (F)—liver	161	168	70.7	70.5	64.6	64.5	55.2	55.2
TCE (M)—kidney	136	181	136	106	89.0	91.9	41.9	64.9
TCE (F)—kidney	118	118	34.4	34.4	22.8	22.8	18.8	18.8
Trichlorobenzene	1000	1000	566	652	496	540	389	394

Note. F, female; M, male; MLE, maximum likelihood estimate; LCL, lower confidence limit; T, THRESH; TW, THRESHW.

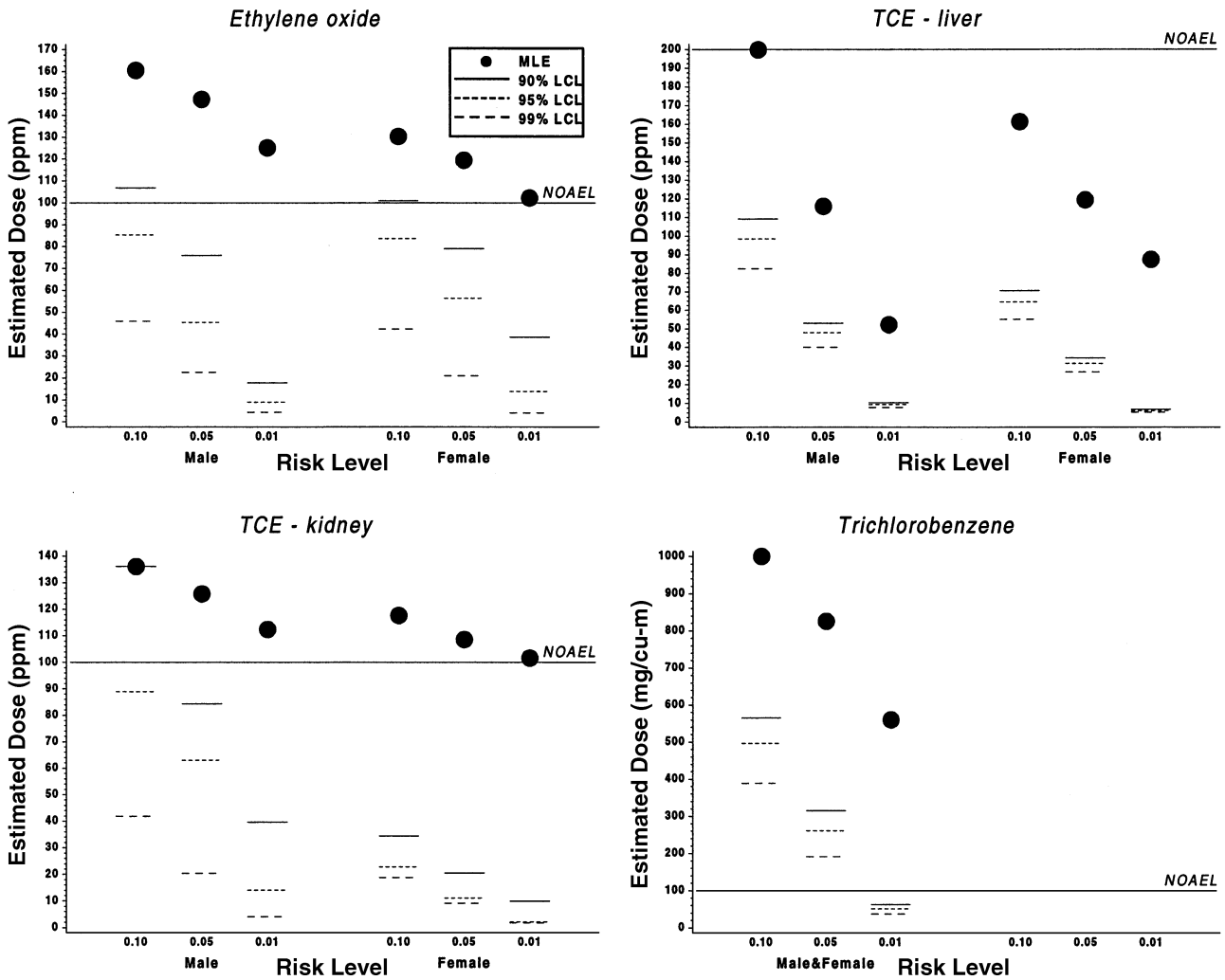


FIG. 4. Modeling results using THRESH for chemicals with discrete endpoints.

TABLE 5
Toxicological Studies with Continuous Endpoints

Agent	Continuous endpoint(s) dictating NOAEL	Species and sex	Route of exposure	NOAEL	Doses
Diethylamine (DEA)	• Body weight	Male and female rats	Inhalation	25 ppm	0, 25, 250
Ethylene glycol monohexyl ether (EGME)	• Enzyme levels (AST, ALT, SDH, and ALP) • Liver/body weight • Kidney/body weight	Female rats	Inhalation	41 ppm	0, 20, 41, 71
1,2-Dichlorobenzene (DCB)	• Liver/body weight	Female rats	Oral (gavage)	25 mg/kg/day	0, 25, 100, 400
Hexachloroethane (hexa)	• Liver/body weight • Kidney/body weight	Male rats	Diet	1 mg/kg/day	0, 1, 15, 62
High-boiling coal liquid (HBCL)	• Liver/body weight	Male and female mice	Inhalation	0.03 mg/L	0, 0.03, 0.14, 0.69
1,3-Butadiene (B)	• Body weight	Male and female mice	Inhalation	1250 ppm (M) 2500 ppm (F)	0, 625, 1250, 2500, 5000, 8000
2,4-Dichlorophenyl-*p*-nitrophenyl ether (DCP)	• Liver/body weight	Male rats	Diet	100 ppm	0, 100, 500, 2500, 12500, 50000
1,2-Epoxybutane (epoxy)	• Body weight	Male and female rats	Inhalation	400 ppm	0, 50, 100, 200, 400, 800
Methyl methacrylate (MM)	• Body weight	Male rats	Inhalation	1000 ppm	0, 500, 1000, 2000, 3000, 5000

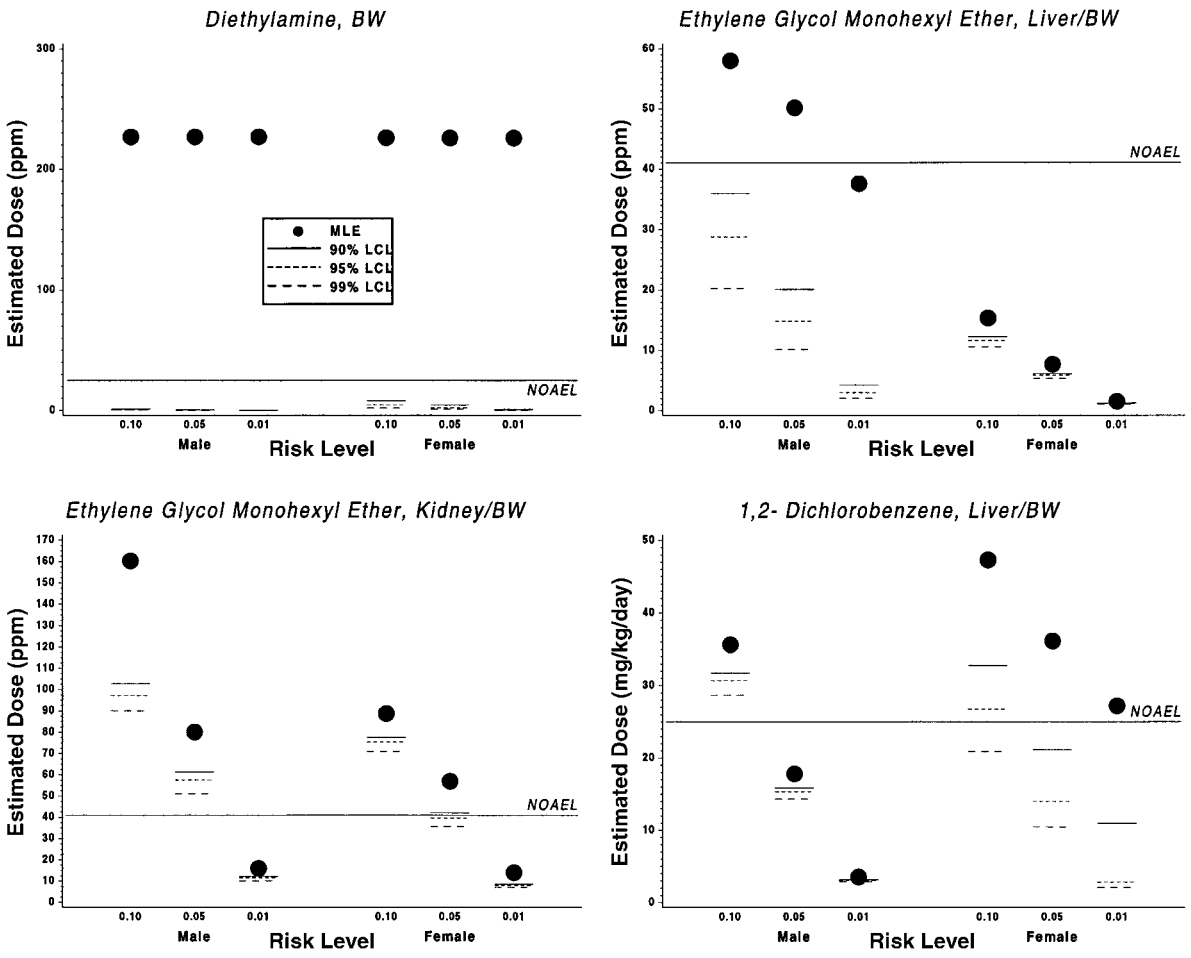


FIG. 5. Modeling results using THC for chemicals with continuous endpoints.

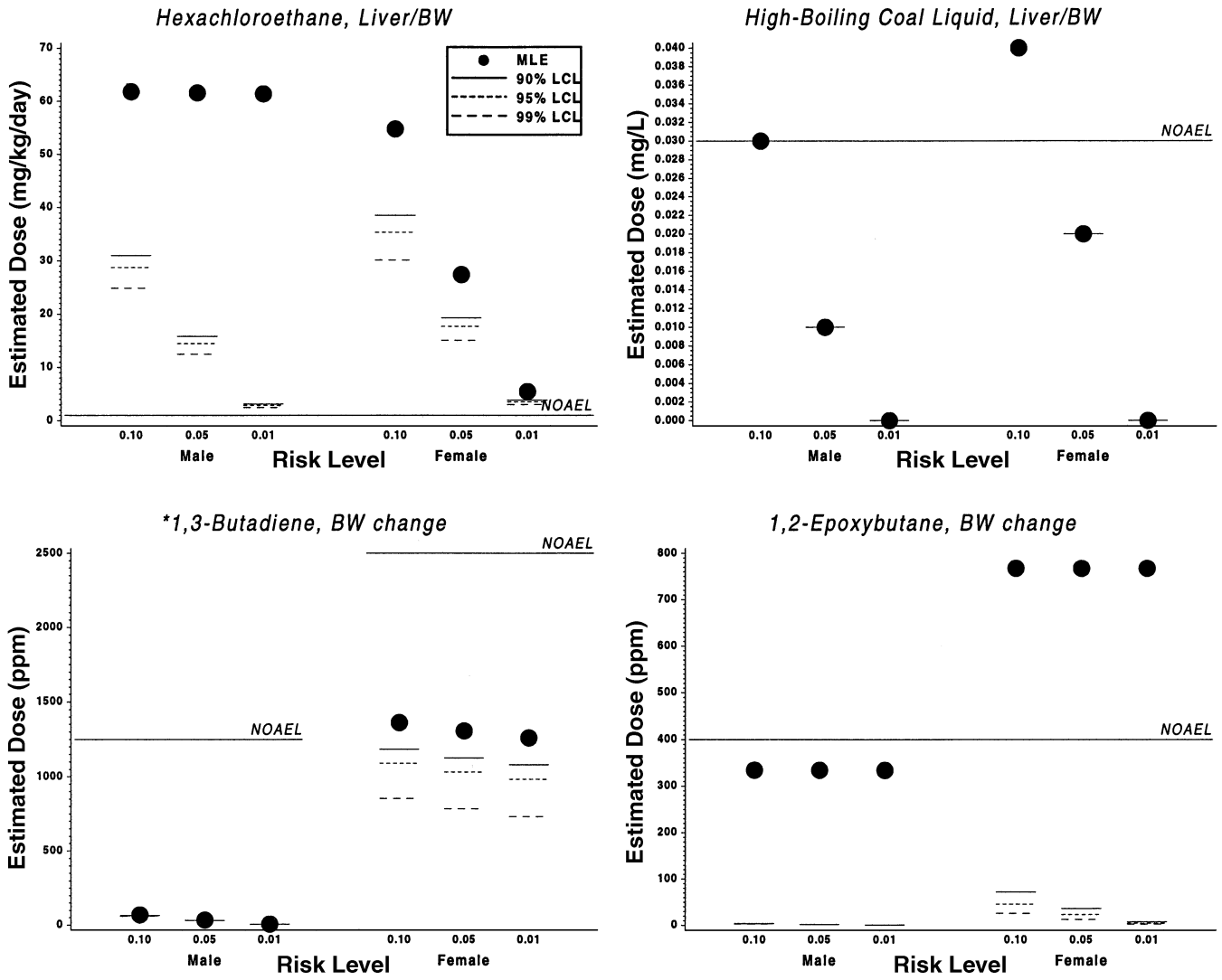


FIG. 6. Modeling results using THC for chemicals with continuous endpoints.

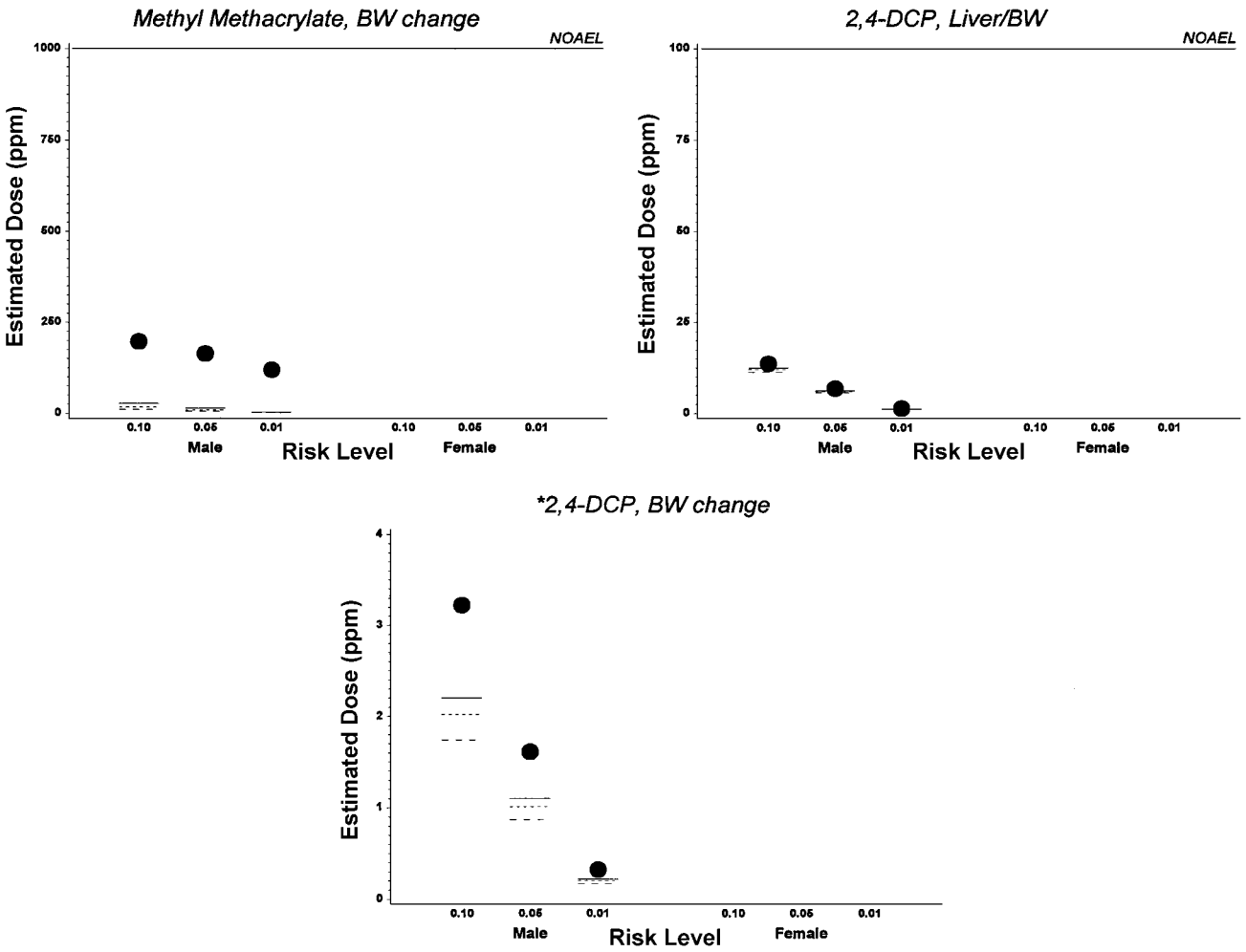
(i.e., a 10% change in body weight represents a 10% risk level).

As with the discrete data, the two BMD models gave similar MLE estimates (Tables 6 and 7). However, the LCL estimates for some of the data sets were highly divergent (i.e., up to a 20-fold difference [see highlighted values in Tables 6 and 7]). Both models provided reasonable fits to some of the data, but for many datasets, the BMD models provided perfect fits ($P = 1.0$) to the data since only one or two dose groups existed above the NOAEL. Defining the BMD as the 95% LCL at the 10% risk level, the BMD was typically an order of magnitude less than the NOAEL.

Summary.

- Model fit ranged from poor (P fit $= 0.00$) to perfect (P fit $= 1.00$) for the various data sets.
- The MLE was typically within an order of magnitude of the LCL at a given risk level.
- The BMDs calculated as a 95% LCL at a 10% risk level were typically an order of magnitude less than the NOAEL.

c. Ratio of NOAEL to MLE or LCL. Figures 8–11 show the ratios of MLEs/NOAELs and LCLs/NOAELs. Discrete and continuous endpoints were used for the analysis; however, only the estimates from the THRESH and THC models were used. In addition, male data were chosen for agents that had both male and female exposures (Figs. 8–10). Comparisons were made using female data if male data were not available (Fig. 11). On average, the MLEs were greater than the NOAEL (MLE/NOAEL ratio greater than 1.0) and the LCLs were less than the NOAEL (LCL/NOAEL ratio of less than 1.0) for discrete endpoints (Fig. 8). For



*NOAEL not shown since it is much greater than the MLEs (NOAEL = 100 ppm)

FIG. 7. Modeling results using THC for chemicals with continuous endpoints.

continuous endpoints, the MLEs and LCLs were highly variable with respect to the NOAEL. No trend was observed for the ratio of MLEs and LCLs to the NOAELs for continuous endpoints (Figs. 9–11).

B. Fabricated Data Sets

1. General Approach

Three fabricated discrete datasets were generated (Table 8). All data sets had the same theoretical NOAEL of 10. Data sets 1 and 2 are considered "ideal" since a number of doses above the NOAEL exist, the group sizes are relatively large (i.e., group sizes of 10), and the data exhibit good dose–response characteristics (i.e., the dose–response curves are relatively smooth; see Fig. 12). In contrast, data set 3 consisted of several nonresponding groups and only one responding group. The fabricated data sets were used to test how the BMD models handled different types and permutations of the data. We hypothesized that if the BMD models could not accurately model different permutations of the ideal data sets, then they would unlikely be able to accurately model real toxicological data sets that often have poor dose–response curves and small sample sizes. We restricted our analysis to discrete data since the analysis of the 90-day studies indicated that the BMD methodology should not be used for continuous data until a more robust methodology is developed. MLEs and LCLs were calculated for the three data sets. The data sets were modified and the MLEs and LCLs recalculated. The modifications consisted of halving the sample size, eliminating various dose levels, eliminating the NOAEL, or lowering a given dose level.

Tables 9–11 show the MLE and 95% LCL estimates for the fabricated discrete data sets at BMRs of 1, 5, and 10% using the THRESH and THRESHW models. The data were mainpulated as indicated in the tables and the MLEs and LCLs recalculated. The data manipulated tested how well the models handled different formats of the data (e.g., small sample size vs large). The

TABLE 6
BMD Analysis of Continuous Endpoints Using a 95% LCL

Agent (sex)	Model fit (*P* value)		MLE01		LCL01		MLE05		LCL05		MLE10		LCL10	
	T	TW	T	TW	T	TW	T	TW	T	TW	T	TW	T	TW
Diethylamine (M)	1.00	1.00	227	205	0.07	0.21	227	205	0.34	0.74	227	205	0.67	1.29
Diethylamine (F)	1.00	1.00	226	199	0.47	1.41	226	199	2.28	4.22	226	199	4.46	6.77
EGME (F)—AST	1.00	1.00	20.5	20.0	0.03	0.09	20.6	20.1	0.16	0.35	20.7	20.2	0.31	0.63
EGME (F)—ALT	1.00	1.00	30.4	21.9	0.05	0.35	30.5	22.4	0.26	1.05	30.6	22.8	0.51	1.68
EGME (F)—SDH	1.00	1.00	24.7	20.8	0.08	0.27	24.8	21.2	0.39	0.94	25.0	21.5	0.78	1.61
EGME (F)—ALP	1.00	1.00	69.3	67.5	0.02	0.02	69.3	67.5	0.11	0.11	69.3	67.5	0.22	0.22
EGME (F)—liver/body weight	0.76	0.95	1.54	1.49	1.17	1.17	7.7	7.43	5.83	5.83	15.4	14.9	11.7	11.7
EGME (F)—kidney/body weight	0.66	0.57	14.0	15.4	7.94	7.91	57.0	55.1	39.7	39.5	88.8	95.4	75.5	76.9
1,2-Dichlorobenzene (F)	0.78	0.78	27.2	27.2	2.83	3.06	36.2	36.2	14.0	14.0	47.3	47.3	26.8	26.8
Hexachloroethane (M)—liver/body weight	1.00	1.00	61.4	61.5	2.89	7.13	61.6	62.1	14.5	35.7	61.8	62.9	28.7	62.9
Hexachloroethane (M)—kidney/body weight	1.00	1.00	61.5	61.4	7.13	2.89	62.1	61.6	35.7	14.5	62.9	61.8	62.9	28.7
HBCL (M)	0.04	0.12	0.00	0.00	0.00	0.00	0.01	0.01	0.01	0.01	0.03	0.03	0.03	0.03
HBCL (F)	0.43	0.31	0.00	0.01	0.00	0.00	0.02	0.03	0.02	0.02	0.04	0.05	0.03	0.03
1,3-Butadiene (M)	0.00	0.00	6.96	6.96	6.22	6.22	34.8	34.8	31.1	31.1	69.6	69.6	62.2	62.2
1,3-Butadiene (F)	0.00	0.00	1261	1261	983	988	1307	1307	1031	1035	1363	1363	1091	1094
2,4-DCPNPE (M)	0.00	0.00	0.32	0.32	0.20	0.20	1.61	1.61	1.01	1.01	3.22	3.22	2.02	2.02
1,2-Epoxybutane (M)	1.00	0.00	334	334	0.34	24.4	334	335	1.69	45.4	335	335	3.38	59.4
1,2-Epoxybutane (F)	1.00	0.00	767	735	4.58	96.6	767	735	22.9	147	767	735	45.8	176
Methyl methacrylate (M)	0.01	0.00	119	511	1.75	28.1	164	531	8.72	66.1	197	547	17.4	95.7

Note. F, female; M, male; MLE, maximum likelihood estimate; LCL, Lower confidence limit; T, THC; TW, THWC.

models were also run using 90 and 99% LCLs; however, those results are not shown since the trends are similar to the trends observed using the 95% LCL. Both models fit the full and modified data sets very well. The THRESH model had *P* values greater than 0.7 for all of the data set manipulations. The THRESHW model had *P* values greater than 0.7 for all but one of the data set manipulations. The two models provided similar MLE and LCL estimates for a given risk level. The MLEs and LCLs for data set 1 for a given manipulation and risk level were typically greater than the MLEs and LCLs for data set 2. This was expected, since data set 2 had a steeper dose–response slope. Surprisingly, most data set manipulations affected the MLE and LCL

TABLE 7
BMD Analysis of Continuous Endpoints Using a 10% Risk Level

Agent (sex)	MLE		90% LCL		95% LCL		99% LCL	
	T	TW	T	TW	T	TW	T	TW
Diethylamine (M)	227	205	1.01	2.20	0.67	1.29	0.41	0.52
Diethylamine (F)	226	199	8.27	10.6	4.46	6.77	2.00	2.84
EGME (F)—AST	20.7	20.2	0.54	1.59	0.31	0.63	0.19	0.19
EGME (F)—ALT	30.6	22.8	0.77	3.06	0.51	1.68	0.32	0.40
EGME (F)—SDH	25.0	21.5	1.23	3.22	0.78	1.61	0.47	0.47
EGME (F)—ALP	69.3	67.5	0.24	0.24	0.22	0.22	0.20	0.20
EGME (F)—liver/body weight	15.4	14.9	12.3	12.3	11.7	11.7	10.6	10.6
EGME (F)—kidney/body weight	88.8	95.4	77.6	79.6	75.5	76.9	71.1	71.2
1,2-Dichlorobenzene (F)	47.3	47.3	32.8	32.8	26.8	26.8	20.9	20.9
Hexachloroethane (M)—liver/body weight	61.8	61.8	31.0	31.0	28.7	28.7	24.9	24.9
Hexachloroethane (M)—kidney/body weight	62.9	62.9	62.9	62.9	62.9	62.9	62.9	62.9
HBCL (M)	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03
HBCL (F)	0.04	0.05	0.03	0.03	0.03	0.03	0.03	0.03
1,3-Butadiene (M)	69.6	69.6	63.6	63.6	62.2	62.2	60.0	59.9
1,3-Butadiene (F)	1363	1363	1184	1184	1091	1094	852	876
2,4-DCPNPE (M)	3.22	3.22	2.20	2.20	2.02	2.02	1.74	1.74
1,2-Epoxybutane (M)	335	335	3.77	64.0	3.38	59.4	3.01	52.2
1,2-Epoxybutane (F)	767	735	72.2	195	45.8	176	25.3	145
Methyl methacrylate (M)	197	547	27.0	105	17.4	95.7	10.4	80.6

Note. F, female; M, male; MLE, maximum likelihood estimate; LCL, lower confidence limit; T, THC; TW, THWC.

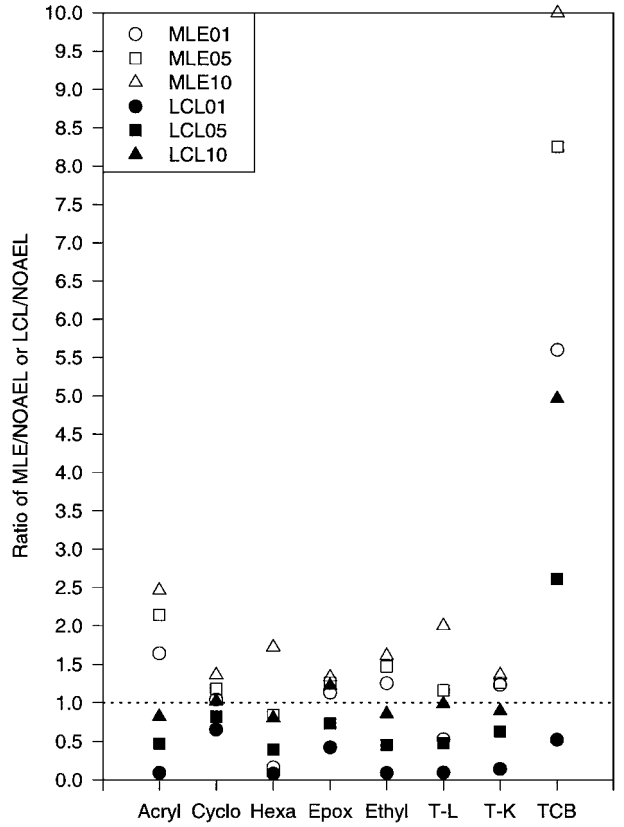


FIG. 8. Ratios of NOAEL to MLE or LCL for discrete endpoints—males.

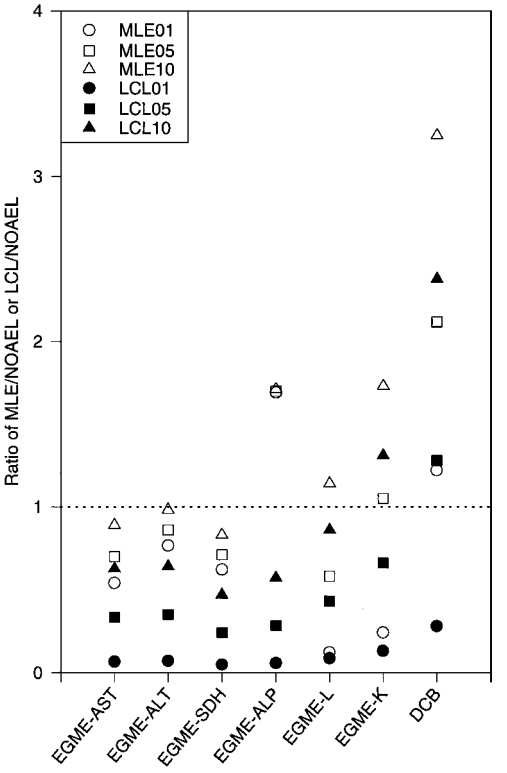


FIG. 10. Ratios of NOAEL to MLE or LCL for continuous endpoints—males.

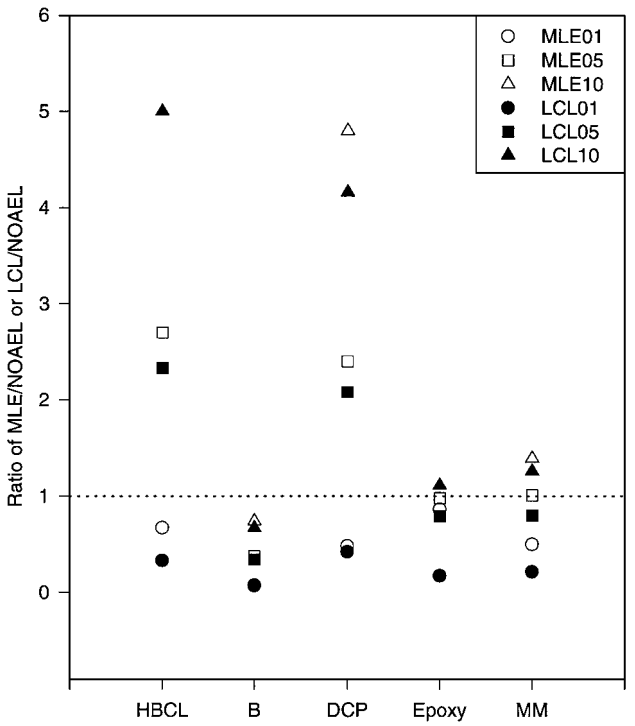


FIG. 9. Ratios of NOAEL to MLE or LCL for continuous endpoints—males.

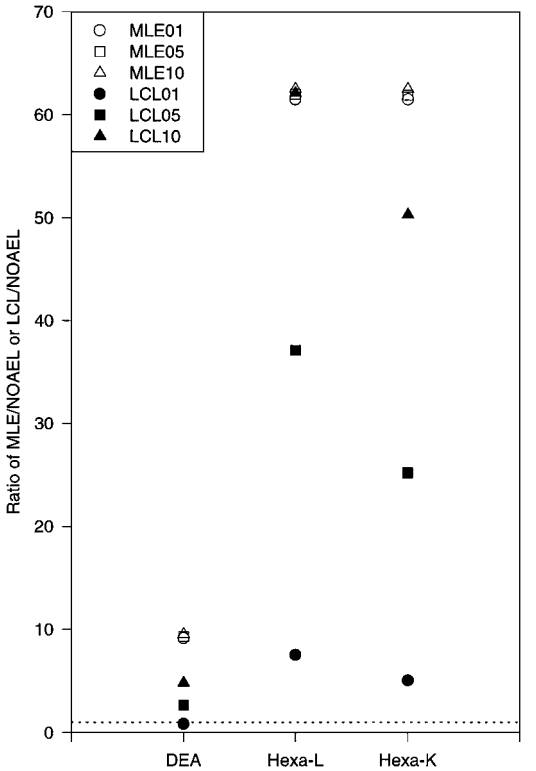


FIG. 11. Ratios of NOAEL to MLE or LCL for continuous endpoints—females.

TABLE 8
Fabricated Discrete Datasets

Dose	Dataset 1D Lesion present	Dataset 1D Reduced dataset	Dataset 2D Lesion present	Dataset 2D Reduced dataset	Dataset 3D Lesion present
0	0/10	0/5	0/10	0/5	0/10
10	0/10	0/5	0/10	0/5	0/10
20	—[a]	—	—	—	0/10
30	—	—	—	—	0/10
40	—	—	—	—	0/10
50	1/10	1/5	3/10	2/5	3/10
100	2/10	1/5	6/10	3/5	—
150	3/10	2/5	7/10	4/5	—
200	5/10	3/5	9/10	5/5	—
300	6/10	3/5	10/10	5/5	—

[a] The particular dose was not part of the given dataset.

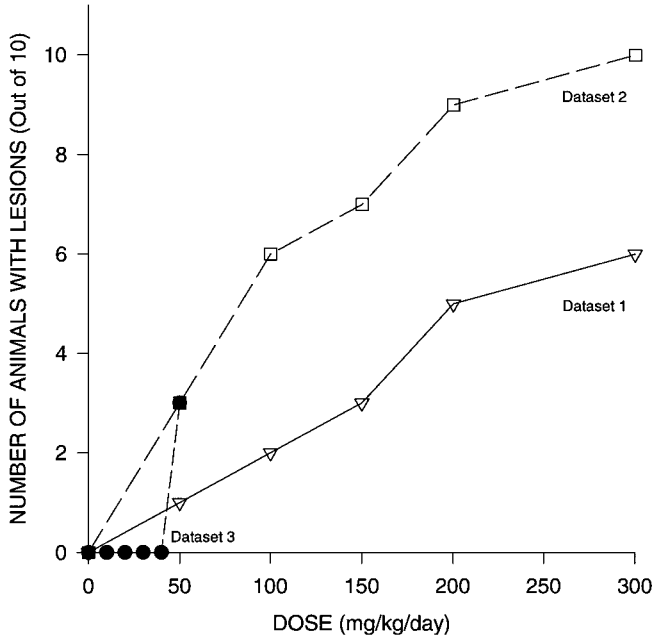


FIG. 12. Graph of fabricated discrete data sets 1, 2, and 3.

estimates very little. The elimination of various doses from data sets 1 and 2 indicated that the higher dose groups contributed very little to the MLE and LCL estimates. Figure 13 shows the effect of eliminating various numbers of dose groups on the MLE and 95% LCL at a 10% risk level for data sets 1; and 2. The dose groups eliminated correspond to those presented in Tables 9 and 10. The most significant effects were noted when four or five dose groups were eliminated. The LCLs decreased for data set 1, whereas the MLEs increased for data set 2 when four or five dose groups were eliminated (Fig. 13).

Data set 3 was used to test how doses below the NOAEL affected the LCL and MLE estimates. Data set 3 used the NOAEL and LOAEL from data set 2, but included several additional nonresponding groups above the NOAEL. Table 11 shows the MLE and 95% LCL estimates for the full and manipulated data set 3 at the 1, 5, and 10% risk levels. Both models fit the data extremely well ($P \geq 0.99$). This would be expected for data sets that consist solely of one responding group and one or more nonresponding groups. The MLEs were insensitive to manipulations of data set 3. In contrast, the LCLs were sensitive to changes in the number of nonresponding dose groups (Fig. 14). In fact, the LCLs were higly dependent on the NOAEL (Fig. 14). The MLEs and LCLs for the full data set 3 were greater than those for the full data set 2.

TABLE 9
BMD Analysis of Discrete Dataset 1 Using a 95% LCL (NOAEL = 10)

Data sampling	Model fit (P value) T	Model fit (P value) TW	MLE01 T	MLE01 TW	LCL01 T	LCL01 TW	MLE05 T	MLE05 TW	LCL05 T	LCL05 TW	MLE10 T	MLE10 TW	LCL10 T	LCL10 TW
Full dataset	0.971	0.973	15.9	15.9	2.65	2.65	32.0	33.7	13.5	13.5	52.2	53.9	27.8	27.8
Eliminated NOAEL	0.888	0.974	15.9	8.25	2.14	2.50	32.0	29.7	10.9	12.7	52.2	52.4	22.5	26.2
Eliminated one dose[a]	0.900	0.902	18.5	15.2	2.49	2.49	32.7	31.9	12.7	12.7	50.9	51.4	26.1	26.1
Eliminated two doses[b]	0.958	0.840	14.0	15.5	2.51	2.50	30.3	33.3	12.8	12.8	51.8	54.0	26.4	26.3
Eliminated three doses[c]	0.977	0.977	13.9	13.9	2.36	2.36	30.0	30.0	12.1	12.1	51.0	51.0	24.8	24.8
Eliminated four doses[d]	1.00	1.00	22.0	43.4	1.71	1.71	37.7	47.9	8.72	8.72	50.0	50.0	17.9	17.9
Eliminated four doses and changed dose 10 to 1	1.00	1.00	23.0	32.0	1.34	1.34	38.2	43.6	6.86	6.86	50.0	50.0	14.1	14.1
Eliminated five doses[e]	1.00	1.00	21.6	43.5	1.71	1.71	33.9	47.9	8.72	8.72	50.0	50.0	17.9	17.9
Reduced data set	0.964	0.964	12.7	12.7	1.86	1.86	23.9	23.9	9.49	9.49	38.5	38.5	19.5	19.5

Note. MLE, maximum likelihood estimate; LCL, lower confidence limit; T, THRESH; TW, THRESHW.
[a] 100 dose eliminated.
[b] 100 and 200 doses eliminated.
[c] 100, 200, and 300 doses eliminated.
[d] 100, 150, 200, and 300 doses eliminated.
[e] 0, 100, 150, 200, and 300 doses eliminated.

TABLE 10
BMD Analysis of Discrete Dataset 2 Using a 95% LCL (NOAEL = 10)

Data sampling	Model fit (*P* value)		MLE01		LCL01		MLE05		LCL05		MLE10		LCL10	
	T	TW	T	TW	T	TW	T	TW	T	TW	T	TW	T	TW
Full dataset	0.984	0.874	11.1	12.7	1.01	1.15	15.7	19.8	5.15	5.62	21.7	27.3	10.6	11.3
Eliminated NOAEL	0.911	0.893	7.25	4.39	0.837	0.812	12.2	13.8	4.27	4.14	18.6	23.0	8.77	8.51
Eliminated one dose[a]	0.970	0.715	11.2	13.0	1.03	1.15	16.4	20.4	5.26	5.60	23.1	28.1	10.8	11.2
Eliminated two doses[b]	0.738	0.425	11.3	12.9	1.05	1.06	16.6	20.3	5.33	5.39	23.6	28.0	11.0	11.0
Eliminated three doses[c]	0.959	0.959	11.2	11.2	1.00	1.00	15.9	15.9	5.12	5.12	22.1	22.1	10.5	10.5
Eliminated four doses[d]	1.00	1.00	29.4	41.0	0.938	0.938	33.1	44.9	4.79	4.79	37.0	46.7	9.83	9.83
Eliminated four doses and changed dose 10 to 1	1.00	1.00	18.5	29.1	0.643	0.643	26.5	37.3	3.28	3.28	32.7	41.6	6.74	6.74
Eliminated five doses[e]	1.00	1.00	36.1	41.0	0.938	0.938	37.7	44.9	4.79	4.79	39.9	46.7	9.83	9.83
Reduced data set	0.973	0.861	11.0	12.5	0.708	0.663	15.0	18.7	3.61	3.38	20.3	25.0	7.42	6.95

Note. MLE, maximum likelihood estimate; LCL, lower confidence limit; T, THRESH; TW, THRESHW.
[a]100 dose eliminated.
[b]100 and 200 doses eliminated.
[c]100, 200, and 300 doses eliminated.
[d]100, 150, 200, and 300 doses eliminated.
[e]0, 100, 150, 200, and 300 doses eliminated.

Summary.

- Model fit was excellent for the full and manipulated data sets.
- Dose groups not immediately surrounding the MLE contributed very little to the LCL and MLE estimates.
- The MLEs and LCLs were insensitive to changes in responses one dose group beyond the MLE.
- The LCL was highly dependent on the NOAEL.
- The BMDs defined as the 95% LCL at a 10% risk level were within an order of magnitude of the corresponding NOAELs.

DISCUSSION

The EPA has endorsed the use of the BMD methodology for noncancer risk assessment (EPA, 1995) and it even recommends the use of BMD models in its revised carcinogen risk assessment guidelines (1996). Even though the BMD offers advantages over the approaches currently used in noncancer risk assessment, the methodology has not been adequately tested. Several studies applied the BMD to numerous developmental toxicity data sets (Kimmel *et al.*, 1995; Gaylor, 1989; Allen *et al.*, 1994). However, it is important to determine how the BMD methodology handles other toxicological data, such as 90-day toxicity studies that are routinely used for the regulation and hazard assessment of chemicals.

1. 90-Day Toxicological Data

A highly touted strength of the BMD methodology is that different models which properly fit the data should give similar BMD estimates for the same data set. There are two basic reasons for the similarity: (1) The BMD is calculated for responses that can be detected in a typical study (i.e., in the observable range),

TABLE 11
BMD Analysis of Discrete Dataset 3 Using a 95% LCL (NOAEL = 10)

Data sampling	Model Fit (*P* value)		MLE01		LCL01		MLE05		LCL05		MLE10		LCL10	
	T	TW	T	TW	T	TW	T	TW	T	TW	T	TW	T	TW
Full dataset	1.00	0.997	42.8	44.0	19.4	24.6	43.9	44.6	33.4	34.5	45.3	47.8	37.9	39.3
Eliminated dose 40	1.00	0.996	35.7	42.6	6.21	11.8	40.3	45.8	22.1	23.3	43.3	47.3	27.9	30.1
Eliminated doses 30 and 40	1.00	1.00	26.3	41.0	2.39	3.44	33.3	44.9	11.8	12.5	38.1	46.7	18.7	20.8
Eliminated doses 20, 30, and 40	1.00	1.00	29.4	41.0	0.938	0.938	33.1	44.9	4.79	4.79	37.0	46.7	9.83	9.83
Eliminated doses 20 and 30	1.00	1.00	40.3	43.9	18.1	19.2	41.6	46.6	30.7	31.0	43.1	47.8	37.6	37.6
Eliminated doses 10, 20, and 30	1.00	1.00	40.3	43.9	24.2	19.0	41.5	46.5	32.4	30.8	43.0	47.8	38.0	37.6
Eliminated doses 0, 10, 20, and 30	1.00	1.00	40.3	43.9	31.6	19.0	41.4	46.5	34.9	30.8	43.0	47.8	38.5	37.6

Note. MLE, maximum likelihood estimate; LCL, lower confidence limit; T, THRESH; TW, THRESHW.

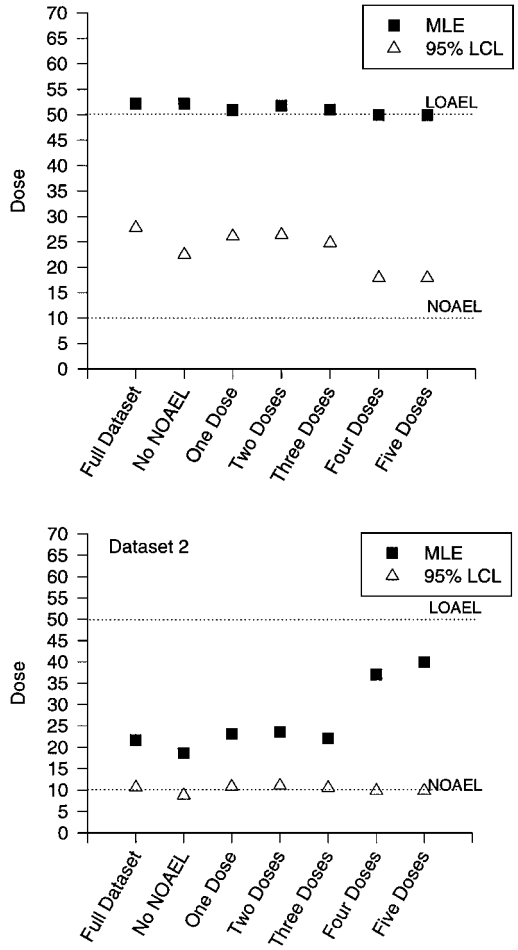


FIG. 13. Effect of eliminating various dose groups on the MLE and 95% LCL at a 10% risk level for data sets 1 and 2.

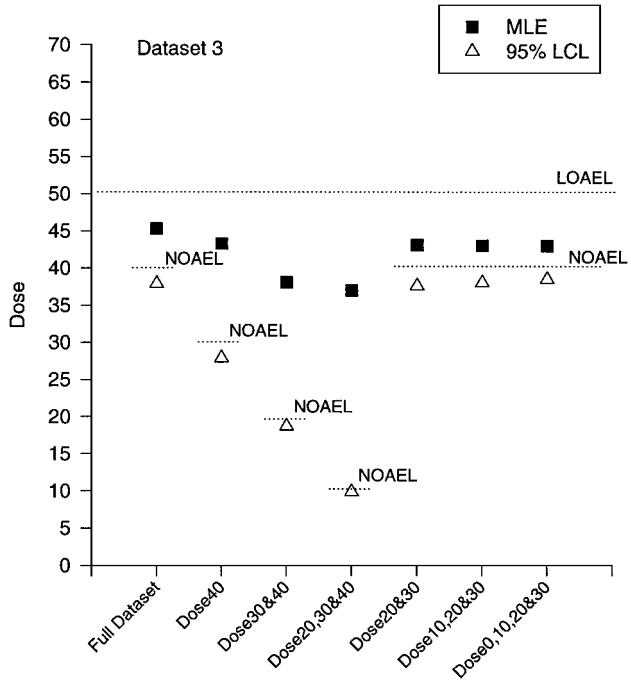


FIG. 14. Effect of eliminating various dose groups on the MLE and 95% LCL at a 10% risk level for data Set 3.

and (2) there are usually only two or three dose groups that affect the result. We found that, in general, when there were models with equally good fit, the different BMD models did give similar BMD estimates for both continuous and discrete data.

We found that the models fit the majority of discrete and continuous data sets very well ($P > 0.5$). Some models fit the data perfectly (i.e., $P = 1.0$); however, this was not surprising since many data sets had only one to three dose levels above the NOAEL (i.e., only one to three dose levels showed positive responses). Since the BMD models use multiple parameters to fit the data, it would be expected that the models would fit these type of data sets perfectly. Using multiple parameters to fit a limited number of responding dose levels is similar to drawing a straight line between two points.

One problem that was frequently encountered was incomplete data. A BMD could not be calculated for many published studies since the data were not presented in an adequate format. In order to calculate a BMD for continuous data, the mean, standard deviation, and sample size are required. The mean was often available, but the standard deviation and/or group size was missing. For discrete endpoints, much of the data were not presented. It was only summarized in the text. This was typically encountered for histopathological lesions. Therefore, if a BMD will be calculated for future studies, it is necessary to summarize the data in a format that is useable.

We found that the BMD estimates for continuous data were highly variable with respect to the NOAEL and LOAEL. This is not surprising, since we assumed that risk was proportional to the percentage of change in the continuous variable. This assumption is acceptable for discrete data since risk and response or often proportional (e.g., a 10% increase in the number of mice with liver necrosis is indicative of a 10% risk level). However, for continuous endpoints, risk and response are rarely directly proportional. For example, a 10% change in serum ALT levels (a marker of liver damage) may represent less than a 1% increase in risk.

A more realistic approach is needed before the BMD can be applied to continuous data. Murrell *et al.* (1998) based the risk level relative to the maximum response expected instead of a percentage change in the response. This approach is more realistic than the one used in this study; however, it is often difficult to determine what the maximum response should be. Therefore, more work needs to be done before the BMD can be applied to continuous data.

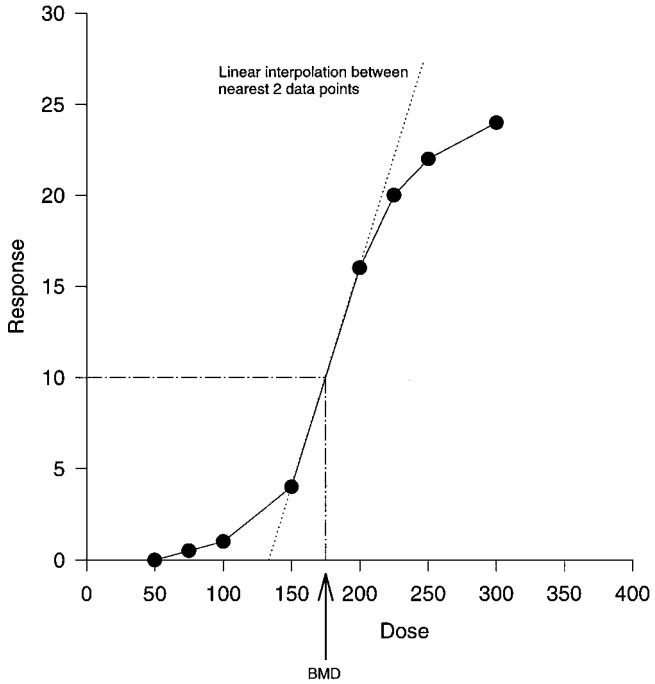


FIG. 15. Dependency of the BMD on the two dose levels immediately surrounding the response rat.

0.0.2. 2. Fabricated Data Sets

Modeling of the fabricated data sets highlighted one potential problem with the BMD methodology—the BMD estimate is highly dependent upon doses immediately surrounding the MLE. Murrell *et al.* (1998) mentioned this potential problem, but they did not provide analytical proof of this behavior. They said that if the risk level is chosen to lie so far into the range of the observed response that the curve is locally linear, then the projection of the BMR to a dose (via the model) relies almost entirely on the response at the nearest two dose levels. Basically, the dose levels immediately above and below the dose corresponding to the BMR will dictate the BMD. This is shown graphically in Fig. 15.

The LCLs and MLEs for BMRs of 1–10% were typically between the NOAEL and LOAEL for the fabricated discrete data sets (Tables 5 and 6 and Fig. 13). These results indicate that the BMD is highly dependent upon the doses immediately surrounding the MLE, which, in this case, happen to be the NOAEL and LOAEL.

Alterations in datapoints greater than one dose group away from the MLE affected the MLE and LCL very little (Fig. 13). This shows that the MLE is highly dependent on the adjacent dose groups. Since the LOAEL in most experiments is associated with low response rates, it is very likely that the NOAEL and LOAEL will be close to the dose associated with 1–10% responses. This will result in the BMD being highly dependent upon the NOAEL and LOAEL. If this is true, then the advantage of using the BMD is lost since the BMD provides very little additional information over the NOAEL and LOAEL.

Surprisingly, nonresponding dose groups below the NOAEL had a greater impact on the LCL than did doses above the NOAEL. As the additional nonresponding group(s) moved closer to the NOAEL, the LCL became greater. This can be seen when doses 20, 30, and/or 40 were removed from data set 3 (Fig. 14). This suggests that nonresponding dose groups below the NOAEL have a significant influence on the BMD.

We did not derive BMDs for endpoints that did not dictate the NOAEL since we were simply interested in determining how the BMD methodology handled typical toxicological data. It is possible that the BMD for an endpoint that did not dictate the NOAEL may result in the lowest BMD. This is because the slope of the dose–response is used in the BMD estimation. Therefore, BMDs should be calculated for all of the endpoints measured in the study. This would more accurately identify the critical effect.

The EPA stated that a BMD should be calculated for risk levels between 1 and 10% (EPA, 1995). In addition, the BMD should be calculated using a LCL in the range of 90–99%, as suggested by the EPA; note that they have generally employed a 95% LCL (EPA, 1995). Unfortunately, the selection of the risk level is not a science-based decision. It is a policy-based decision. Therefore, it is important to present BMDs based on a range of risk and LCL levels so that the risk assessor and/or manager can choose the most appropriate BMD based on the quality of the data, mechanisms of action, etc. From the analysis done in this study, it appears that different BMD models give similar estimates when restricted to responses in the observable range. We did not try extrapolating to extremely low levels of risk; however, it is very likely that the models would diverge greatly in their estimates at low risk levels. This was often observed for models that were used to derive slope factors for carcinogens at extremely low risk levels. The different models often calculated slope factors that differed by several orders of magnitude (EPA, 1988). Therefore, BMDs should not be calculated for extremely low levels of risk, i.e., risks less than 1%.

3. BMD and Risk Assessment

Even though the BMD is believed to provide an accurate representation of the dose-response, the BMD is unlikely to greatly alter the toxicity values used in risk assessment. Depending upon how the BMD is defined (e.g., the MLE or the 95% LCL), the BMD methodology may actually decrease the values used in risk assessment. For discrete data, many of the MLEs at the 10% risk level (least conservative BMD estimation) were at most only two-fold greater than the NOAEL (Figs. 3

and 4). Many of the 95% LCLs at the 10% risk level (more conservative BMD estimation) were less than the NOAEL (Figs. 3 and 4). For continuous data, the MLEs and LCLs were highly variable with respect to the NOAEL (Figs. 5–7).

Although the slope of the dose–response curve is not slated for use in the risk assessment of noncarcinogens, it is possible that it could be used as an additional indicator of safety. The final toxicity value, after applying UFs and any other adjustments to the BMD, could have another number or label associated with it that indicates how steep the dose–response curve is. This would provide an indication of the margin of safety associated with the chemical.

After investigating the application of the BMD methodology, we gathered a series of questions and comments concerning the approach. Many of these have been answered while others remain to be answered. The following summarizes these issues.

a. BMD vs NOAEL.

- The BMD is unrestricted. It can be calculated for any response rate. In contrast, the NOAEL is restricted to the experimental doses. *This is a beneficial property of the BMD methodology.*
- The BMD is less sensitive to dose level selection. *This does not appear to be the case since the BMD is highly dependent upon the NOAEL and LOAEL.*
- The BMD provides a slope of the dose–response curve and the uncertainty associated with the BMD. These values can be used to improve the risk assessment process. *It is unclear whether the slope of the dose–response curve will be used in risk assessments. Currently, the EPA uses the LCL to define the BMD; however, the LCL does not appear to be superior to the MLE since the LCL is highly dependent upon the NOAEL.*
- The BMD encourages better experimentation since larger group sizes should result in higher BMDs when based on the LCL. This is because larger groups decrease the uncertainty associated with the dose–response curve. In contrast, larger group sizes typically decrease the NOAEL since lower response rates can be detected. *Although we did a limited analysis, group size does not appear to have a large effect on the BMD.*
- A BMD can be calculated from an experiment lacking a NOAEL. Therefore, repeat of experiments lacking a NOAEL may not be necessary. *This is a beneficial property of the BMD methodology.*

b. Model and data selection.

- What model(s) should be used to calculate the BMD? *It is unclear which model(s) should be used; however, our analysis indicated that the two discrete and continuous BMD models we tested gave similar BMD estimates.*
- Should a threshold parameter be incorporated into the model? *We did not specifically address this point; however, since noncarcinogens are assumed to have thresholds, a threshold parameter should be incorporated into the model.*
- Should a background parameter be incorporated into the model? *We did not specifically address this point; however, a background parameter should probably be incorporated if it is feasible.*
- How should risk be defined? *We only looked at additional risk estimates. Extra risk places greater weight on common effects (i.e., effects with high background rates); therefore, additional risk may be more appropriate since it does not bias effects with high background rates.*
- What should the BMR be set at? *A 10% BMR appears to be reasonable since a 10% response rate can be observed in most well-designed toxicological studies. A 1% BMR is probably too low, since group sizes would have to be 100 to observe this response rate.*
- Should continuous data be transformed into quantal data prior to modeling? *We did not address this point. It is clear that a better methodology is required before BMDs can be derived for continuous data.*
- How should the risk associated with continuous responses be defined? *See the previous comment.*

c. Uncertainty calculation and model fit.

- How should the uncertainty associated with the dose–response curve be calculated? *We did not specifically address this point; however, the LCL appears to be highly dependent upon the NOAEL and may not be a good indicator of the uncertainty.*
- How important is model fit? *We did not specifically address this point; however, intuitively, the model should provide a reasonable fit to the data.*
- Should data points be eliminated to provide a better fit to the model? *We did not specifically address this point; however, it may be appropriate in certain cases to eliminate high dose groups (e.g., a response becomes saturated at high doses).*
- Can correlated effects be accurately modeled? *We did not address this point.*
- What is the minimum number of responding groups required for calculation of a BMD? *A BMD can be calculated with as few as one responding dose group; however, a minimum of two or three responding dose groups is more reasonable.*

d. Risk assessment and the BMD.

- Should the slope of the dose–response curve be used in risk assessment? *We did not specifically address this point; however, the slope could provide useful information on the size of the population affected by a given change in exposure.*

- Should the BMD be defined as the MLE or the LCL? *The MLE was less sensitive to changes in the NOAEL than the LCL and may be a better indicator of the actual dose-response.*
- Should the BMD be benchmarked against the NOAEL? *The BMD should be compared to the overall dose–response rather than a single number.*
- Should multiple effects be modeled or only the effect with the lowest NOAEL? *Since the effect with the lowest NOAEL may not have the lowest BMD, all effects in a given study should be modeled.*
- Should extrapolation to extremely low BMRs be done? *Although we did not specifically address this point, extrapolation to extremely low BMRs should be avoided since the BMD estimates are likely to be highly model dependent.*
- What uncertainty factors should be applied to the BMD to derive an RfD or RfC? *We did not address this point.*

CONCLUSION

Our review showed that the BMD is typically within an order of magnitude of the NOAEL for discrete data. This suggests that BMD analysis adds very little information about the dose-response that could not be determined by simply drawing a line between the NOAEL and LOAEL. In addition, there are many subjective variables and decisions associated with the calculation of the BMD that may result in more or less conservative BMDs. Therefore, if BMDs are calculated for discrete endpoints, a range of results representing various risk levels and LCLs should be presented so that the risk assessor/risk manager can choose the value(s) he/she feels is most appropriate, given the quality of the data and the application of the RfD/C.

In the case of continuous data, the BMD was highly variable with respect to the NOAEL/LOAEL. This was because we assumed that risk was directly proportional to response. With the approach used in this study, the BMD estimates only reflected the percentage of change in the response, not the percentage of change in risk. For continuous data, the BMD will have to be tailored for each endpoint. A generic approach will not accurately characterize the dose-response. Therefore, BMDs should not be estimated for continuous endpoints until a more realistic approach for relating response to risk is determined.

APPENDIX A

Summary of BMD Models

All of the computer programs used to calculate the BMDs in this study were obtained from ICF Kaiser Engineers, Inc., K.S. Crump Division (Ruston, LA).

I. MODELS FOR CONTINUOUS DATA

A. General Response Models

Multistage model (THC)

$$P(d_i) = q_0 + q_1 * (d_i - d_0) + \cdots + q_k * (d_i - d_0)^k)$$

Weibull model (second form is THCW)

$$P(d_i) = q_0 + (1 - q_0) * [1 - \exp(\beta d)^\gamma]$$

and

$$P(d_i) = q_0 + q_1 * (d_i - d_0)^\gamma$$

Log-logistic model

$$P(d_i) = q_0 + (1 - q_0) * [1 - \{1 + (\beta d)^\gamma\}^{-1}]$$

where $P(d)$ is the probability of a response at dose d, q_0 is the background rate, and k, q_i, β, and γ are model parameters.

B. Reproductive/Developmental (Teratogenic) Models

Considering:

litter size (and intralitter effects);
background rate;
dose;
threshold dose.

Weibull-based model (TERAMOD)

$$P(d_i, s_{ij}) = 1 - \exp[-(\alpha + \theta_1 * s_{ij} + (\beta + \theta_2 * s_{ij})) * (d_i - d_0)^\gamma]$$

Log-logistic-based model (TERALOG)

$$P(d_i, s_{ij}) = \alpha + \theta_1 * s_{ij} + (1 - \alpha - \theta_1 * s_{ij}) * [1 + \exp(\beta + \theta_2 * s - \gamma * \log(d_i - d_0)]^{-1}$$

where d_0 is the threshold dose, d_i is the dose for the ith dose group, s_{ij} is the size of the jth litter of the ith dose, α is the background rate parameter, β is the dose rate parameter, θ_1, θ_2 are litter size parameters, and γ is a distribution parameter.

II. MODELS FOR QUANTAL DATA

A. General Response Models

Linear mean model

$$E(d) = \mu_0 + \beta d$$

Power mean model

$$E(d) = \mu_0 + \beta d^\gamma$$

Multistage or polynomial model (THRESH)

$$P(d_i) = q_0 + (1 - q_0) * [1 - \exp(-q_1 * (d_i - d_0) - \cdots - q_k * (d_i - d_0)^k)]$$

Weibul model (THRESHW)

$$P(d_i) = q_0 + (1 - q_0) * [1 - \exp(-q_1 * (d_i - d_0)^{\gamma}]$$

where $E(d)$ is the expected number of responses at dose d, $P(d_i)$ is the probability of a response at the ith dose, d_0 is the threshold dose, d_i is the administered dose, q_0 is the background rate, and q_i and γ are model parameters.

APPENDIX B

Summary of Studies Used

• *Acrylamide* (Burek *et al.*, 1980). Male and female Fisher 344 rats ($n = 10$/group/sex) were administered acrylamide in the drinking water so that the daily intake was 0, 0.05, 0.2, 1, 5, or 20 mg/kg/day. Body weight, organ weight, and histopathology, hematological and clinical chemistry parameters, and neurobehavioral function were measured. The NOAEL was 1 mg/kg/day. The LOAEL was 5 ppm and was based on histopathological evidence of peripheral nerve damage. For the full range of previously mentioned experimental doses, respectively, the following incidence of histopathological evidence of peripheral nerve damage was measured: males—0, 0, 0, 0, 9/10, 10/10; females—0, 0, 0, 0, 6/10, 10/10.

• *1,3-Butadiene* (NTP TR288, 1984). Male and female B6C3F$_1$ mice ($n = 10$/group/sex) were exposed to 1,3-butadiene vapor concentrations of 0, 625, 1250, 2500, 5000, and 8000 ppm 6 h/day, 5 days/week for 14 weeks. Body weight and organ weight and histopathology were measured. The NOAELs were 1250 and 2500 for males and females, respectively. The LOAELs were 2500 and 5000 for males and females, respectively, and were based on body weight. For the full range of previously mentioned experimental concentrations, respectively, the following body weight changes (g ± SD [No. of animals]) were recorded during the course of the experiment: males—13.6 ± 0.5 [10], 12.1 ± 0.6 [9], 14.1 ± 0.05 [9], 9.5 ± 0.4 [9], 6.5 ± 1.0 [4], and 3.2 ± 1.0 [4]; females—12.3 ± 0.5 [10], 10.9 ± 1.3 [9], 12.7 ± 0.3 [10], 11.0 ± 0.3 [10], 9.2 ± 0.7 [9], 6.4 ± 0.5 [9].

• *Cyclohexylamine hydrochloride* (Gaunt *et al.*, 1974). Male and female CFE rats were fed cyclohexylamine hydrochloride in the diet at 0, 600, 2000, or 6000 ppm for 13 weeks. There were 15 animals/group/sex for the control and 600 ppm groups and 20 animals/group/sex for the 2000 and 6000 ppm groups. Body weight, organ weight, and histopathology, food and oxygen consumption, hematological, and urinalysis parameters and urine concentration ability were measured. The NOAEL was 600 ppm. The LOAEL was 2000 and was based on body weight in females and body weight and testicular histopathology in males. For the full range of previously mentioned experimental doses, respectively, the following body weights (g) were measured in females: 281, 271, 242, 214. For the full range of previously mentioned experimental doses, respectively, the following were measured in males: body weight (g)—471, 474, 435, 371; testicular tubular atrophy and decreased spermatogenesis—0, 0, 11/20, 18/20. No standard deviations were available in the report for body weights.

• *1,2-Dichlorobenzene* (Robinson *et al.*, 1991). Male and female Sprague–Dawley rats ($n = 10$/group/sex) were administered 1,2-dichlorobenze at doses of 0, 25, 100, and 400 mg/kg/day by gavage for 90 days. Body weight, organ weight, and histopathology and hematological and clinical chemistry parameters were measured. The NOAEL was 25 mg/kg/day. The LOAEL was 100 mg/kg/day and was based on increased liver weight for females and increased liver weight and serum ALT levels for males. For the full range of previously mentioned experimental doses, respectively, the following values were measured for males: liver/body weight (percentage ± SD)—2.44 ± 0.19, 2.61 ± 0.15, 2.81 ± 0.18, 3.61 ± 0.30; serum ALT (I.U./L ± SD)—26 ± 4, 26 ± 6, 53 ± 36, 42 ± 15. For the full range of previously mentioned experimental doses, respectively, the following liver/body weights (percentage ± SD) were measured for females: 2.54 ± 0.19, 2.51 ± 0.13, 2.86 ± 0.25, 4.20 ±0.39.

• *2,4-Dichlorophenyl-p-nitrophenyl ether* (Ambrose *et al.*, 1971). Male and female Wistar albino rats ($n = 10$/group/sex) were fed 2,4-dichlorophenyl-*p*-nitrophenyl ether in the diet at 0, 100, 500, 2500, 12,500, and 50,000 ppm for 13 weeks. Body weight, organ weight, and histopathology and hematological and urinalysis parameters were measured. The NOAEL was 100 ppm for males. A NOAEL was not established for females. The LOAEL for males was 500 and was based on liver weight. For the full range of previously mentioned experimental doses, respectively, the following liver/body weights (g/kg ± SD) were measured: 33.7 ± 5.0, 34.9 ± 5.1, 40.9 ± 6.2, 56.7 ± 5.9, 121.7 ± 18.1.

• *Diethylamine* (Lynch *et al.*, 1986). Male and female Fisher 344 rats ($n = 58$–80/group/sex) were exposed to vapor concentrations of 0, 25, or 250 ppm of diethylamine for 6.5 h/day, 5 days/week, for 24 weeks. Body weight, organ weight, and histopathology and hematological and clinical chemistry parameters were measured. The NOAEL was 25-ppm for both sexes. The LOAEL was 250 ppm and was based on body weight changes and nasal pathology. However, nasal pathology was not examined for the 25 ppm group. For the full range of previously mentioned experimental concentrations, respectively, the following mean body weights were recorded: male (weight ± SD [No. of animals])—330 ± 18.1 [61], 331 ± 20.5 [79], 260 ± 15.8 [77]; female (weight ± SD [No. of animals])—191 ± 10.6 [58], 199 ± 10.9 [80], 176 ± 13.1 [70].

• *1,2-Epoxybutane* (NTP TR329, 1988). Male and female rats (Fisher 344) and mice ($B6C3F_1$) were exposed to 1,2-epoxybutane vapor concentrations of 0, 50, 100, 200, 400, and 800 ppm 6 h/day, 5 days/week for 13 weeks. Body weight and organ weight and histopathology were measured. The NOAELs were 50 and 100 ppm for female and male mice, respectively. The NOAEL for rats was 400 ppm. The LOAELs were 100 and 200 ppm for female and male mice, respectively, and were based on nasal inflammation. The LOAEL for rats was 800 ppm and was based on nasal inflammation and body weight. For the full range of previously mentioned experimental concentrations, respectively, the following body weight changes (g ± SD) were measured in rats over the course of the experiment: males—198 ± 3, 195 ± 5, 187 ± 5, 197 ± 5, 185 ± 4, 114 ± 7; females—72 ± 2, 71 ± 3, 78 ± 2 (only 9 animals were in this group), 75 ± 4, 72 ± 3, 40 ± 2. Nasal inflammation was present in all male and female rats at 800 ppm and above and absent in rats exposed to lower concentrations. Nasal inflammation was present in all mice at 200 ppm and above. Below 200 ppm, no male mice exhibited nasal inflammation. For female mice, 7/10 had inflammation at 100 ppm, with no female mice exhibiting nasal inflammation below 100 ppm.

• *Ethylene glycol monohexyl ether* (Klonne *et al.*, 1987). Male and female Fisher 344 rats ($n = 10$/group/sex) were exposed to vapor concentration of 0, 20, 41, or 71 ppm of ethylene glycol monohexyl ether for 6 h/day, 5 days/week, for 13 weeks. Body weight, organ weight, and histopathology, hematological parameters, and liver enzyme levels were measured. The NOAEL was 41 ppm for both sexes. The LOAEL was 71 ppm and was based on liver enzyme levels and body and organ weight changes for females and body and organ weight changes for males. Body weight was only graphed; therefore, these data were not used. For the full range of previously mentioned experimental concentrations, respectively, the following values were measured for females: serum AST (I.U./L ± SD)—82 ± 16, 91 ± 18, 75 ± 15, 59 ± 9; serum ALT (I.U./L ± SD)—55 ± 7, 61 ± 15, 50 ± 15, 33 ± 6; serum SDH (I.U./L ± SD)—23 ± 6, 26 ± 7, 20 ± 9, 12 ± 6; serum ALP (I.U./L ± SD)—105 ± 9, 116 ± 10, 106 ± 19, 128 ± 16; liver/body weight (percentage ± SD)—3.08 ± 0.22, 3.23 ± 0.13, 3.35 ± 0.17, 3.57 ± 0.19; kidney/body weight (percentage ± SD)—0.69 ± 0.03, 0.71 ± 0.05, 0.72 ± 0.05, 0.76 ± 0.03. For the full range of previously mentioned experimental concentrations, respectively, the following values were measured for males: liver/body weight (percentage ± SD)—3.30 ± 0.16, 3.35 ± 0.08, 3.34 ± 0.14, 3.58 ± 0.15; kidney/body weight (percentage ± SD)—0.64 ± 0.03, 0.67 ± 0.03, 0.67 ± 0.03, 0.69 ± 0.02.

• *Ethylene oxide* (NTP TR326, 1987). Male and female $B6C3F_1$ mice ($n = 10$/group/sex) were exposed to ethylene oxide vapor concentrations of 0, 50, 100, 200, 400, and 600 ppm for 6 h/day, 5 days/week for 14 weeks. Body weight, organ weight, and histopathology and hematological, urinalysis, and clinical chemistry parameters were measured. The NOAEL was 100 ppm. The LOAEL was 200 ppm and was based on nasal cavity histopathology. Nasal lesions were not examined in the 50-ppm group. At 0, 100, 200, 400, and 600 ppm, respectively, the following incidence of rhinitis of the nasal cavity was measured in males: 0/10, 0/10, 4/10, 10/10, 10/10. At those same doses, respectively, the following incidence of rhinitis of the nasal cavity was measured in females: 0/9, 0/9, 8/10, 9/9, 10/10.

• *Hexachloroethane* (Gorzinski *et al.*, 1985). Male and female Fisher 344 rats ($n = 10$/group/sex) were fed hexachloroethane in the diet for 16 weeks. The approximate doses received by the rats were 0, 1, 15, and 62 mg/kg/day. Body weight and organ weight and histopathology were measured. The NOAEL was 1 mg/kg/day. The LOAEL was 15 mg/kg/day and was based on kidney pathology and kidney and liver weight changes in males and kidney pathology and liver weight changes in females. For the full range of previously mentioned experimental doses, respectively, the following changes were measured in males: hypertrophy of renal proximal tubules—0/10, 1/10, 7/10, 10/10; degeneration of renal tubules—1/10, 2/10, 7/10, 10/10; kidney/body weight (g/100 g ± SD)—0.73 ± 0.04, 0.70 ± 0.02, 0.73 ± 0.01, 0.77 ± 0.02.; liver/body weight (g/100 g ± SD)—2.65 ± 0.06, 2.58 ± 0.07, 2.64 ± 0.09, 2.77 ± 0.12. For the full range of previously mentioned experimental doses, respectively, the following changes were measured in females: degeneration of renal tubules—1/10, 1/10, 2/10, 6/10; liver/body weight (g/100 g ± SD)—2.63 ± 0.06, 2.73 ± 0.11, 2.69 ± 0.09, 2.76 ± 0.10.

• *High-boiling coal liquid* (Springer *et al.*, 1987). Male and female CD-1 mice ($n = 15$/group/sex) were exposed to aerosol concentrations of 0.0, 0.03, 0.14, or 0.69 mg/L of heavy distillate (a high-boiling coal liquid from the solvent-refined coal process) for 6 h/day, 5 days/week, for 13 weeks. Body weight, organ weight, and histopathology and hematological parameters were measured. The NOAEL was 0.03 mg/L for both sexes. The LOAEL was 0.14 mg/L and was based on increased liver/body weights in male and female mice. For the full range of previously mentioned experimental concentrations, respectively, the following liver/body weights (percentage ± SD) were measured: males—5.55 ± 0.16, 5.78 ± 0.10, 6.12 ± 0.22, 8.18 ± 0.19; females—5.64 ± 0.21, 5.76 ± 0.11, 6.01 ± 0.15, 7.73 ± 0.21.

• *Methyl methacrylate* (NTP TR314, 1986). Male and female F344 rats ($n = 10$/group/sex) were exposed to methyl methacrylate vapor concentrations of 0, 500, 1000, 2000, 3000, and 5000 ppm for 6 h/day, 5 days/week, for 14 weeks. Body weight and organ weight and histopathology were measured. The NOAELs were 500 and 1000 ppm for females and males, respectively.

The LOAELs were 1000 and 2000 ppm for females and males, respectively, and were based on body weight. For the full range of previously mentioned experimental concentrations, respectively, the following body weight changes over the 14 weeks of the experiments were measured: males (g ± SD [No. of animals])—132 ± 4 [10], 126 ± 7 [10], 127 ± 4 [10], 103 ± 3 [9], 67 ± 7 [9], all mice in the 5000 ppm group died; females (g ± SD [No. of animals])—59 ± 3 [10], 53 ± 2 [10], 43 ± 2 [10], 39 ± 2 [7], 10 [1].

• *Tetrachloroethylene* (NTP TR311, 1986). Male and female rats (F344) and mice ($B6C3F_1$) (n = 10/group/sex) were exposed to tetrachloroethylene vapor concentrations of 0, 100, 200, 400, 800, and 1600 ppm for 6 h/day, 5 days/week, for 90 days. Body weight, organ weight, and histopathology and hematological and clinical chemistry parameters were measured. The NOAEL was 200 ppm for rats and 100 ppm for mice. The LOAEL was 400 ppm for rats (based on liver histopathology) and 200 ppm for mice (based on liver and kidney histopathology for males and kidney histopathology for females). For the full range of previously mentioned experimental concentrations, respectively, the following were measured in male mice: liver lesions (mitotic alterations)—0/10, not determined, 3/10, 5/10, 5/10, 1/10; kidney lesions (karyomegaly)—0/10, 0/10, 8/10, 10/10, 10/10, 6/7. For the full range of previously mentioned experimental concentrations, respectively, the following were measured in female mice: liver lesions (mitotic alterations)—0/10, not determined, 0/10, 0/10, 0/10, 0/9; kidney lesions (karyomegaly)—0/10, 0/10, 6/10, 10/10, 10/10, 7/7. The incidence of liver lesions was only measured at 0, 200, 400, 800, and 1600 ppm in rats. At these doses, respectively, the following incidence of liver congestion was measured in male rats: 1/10, 2/10, 3/10, 5/10, 7/10. At those doses, the following incidence of liver congestion was measured in female rats: 0/9, 1/10, 5/10, 5/10, 8/9.

REFERENCES

Allen, B. C., Kavlock, R. J., Kimmel, C. A., and Faustman, E. M. (1994). Dose–response assessment for developmental toxicity: II. Comparison of generic benchmark dose estimates with no observed adverse effect levels. *Fundam. Appl. Toxicol.* **23,** 487–495.

Ambrose, A. M., Larson, P. S., Borzelleca, J. F., Blackwell Smith, R., and Hennigar, G. R. (1971). Toxicologic studies on 2,4-dichlorophenyl-*p*-nitrophenyl ether. *Toxicol. Appl. Pharmacol.* **19,** 263–275.

Beliles, R. P., and Shulz, C. O. (1994). Occupational carcinogens. In *Patty's Industrial Hygiene and Toxicology*, 4th ed. (G. D. Clayton and F. E. Clayton, Eds.), pp. 51–75. Wiley, New York.

Breslow, N. E., and Day, N. E. (1980). *Statistical Methods in Cancer Research, Vol. I—The Analysis of Case–Control Studies*, IARC, Lyon, France.

Burek, J. D., Albee, R. R, Beyer, J. E., Bell, T. J., Carreon, R. M., Morden, D. C., Wade, C. E., Hermann, E. A., and Gorzinski, S. J. (1980). Subchronic toxicity of acrylamide administered to rats in the drinking water followed by up to 144 days of recover. *J. Environ. Pathol. Toxicol.* **4,** 157–182.

Crump, K. S. (1984). A new method for determining allowable daily intakes. *Fundam. Appl. Toxicol.* **4,** 854–871.

Crump, K., and Howe, R. (1985). A review of methods for calculating statistical confidence limits in low dose extrapolation. In *Toxicological Risk Assessment, Vol 1: Biological and Statistical Criteria*, (D. Clayson, D. Krewski, and I. Munroe, Eds.). CRC Press, Boca Raton, FL.

EPA (1988). *EPA Toxicology Handbook*. Government Institutes, Rockville, MD.

EPA (1992). *Proposed Guidelines for Carcinogen Risk Assessment*. EPA/600/P-92/003CS.

EPA (1995). *The Use of the Benchmark Dose Approach in Health Risk Assessment*. Prepared by the EPA's Risk Assessment Forum for the Office of Research and Development, Washington, DC. EPA/630/R-94/007.

Faustman, E. M., Allen, B. C., Kavlock, R. J., and Kimmel, C. A. (1994). Dose–response assessment for developmental toxicity: I. Characterization of database and determination of no observed adverse effect levels. *Fundam. Appl. Toxicol.* **23,** 478–486.

Foster, P., and Auton, T. R. (1995). Application of benchmark dose risk assessment methodology to developmental toxicity: An industrial view. *Toxicol. Lett.* **82/83,** 555–559.

Gaunt, I. F., Sharratt, M., Grasso, P., Lansdown, A. B. G., and Gangolli, S. D. (1974). Short-term toxicity of cyclohexylamine hydrochloride in the rat. *Fd. Cosmet. Toxicol.* **12,** 609–624.

Gaylor, D. W. (1989). Quantitative risk analysis for quantal reproductive and developmental effects. *Environ. Health Perspect.* **79,** 243–246.

Gaylor, D. W., and Slikker, W. (1990). Risk assessment for neurotoxic effects. *Neurotoxicology* **11,** 211–218.

Gorzinski, S. J., Nolan, R. J., McCollister, S. B., Kociba, R. J., and Mattsson, J. L. (1985). Subchronic oral toxicity, tissue distribution and clearance of hexachloroethane in the rat. *Drug Chem. Toxicol.* **8,** 155–169.

Kimmel, C. A., Kavlock, R. J., Allen, B. C., and Faustman, E. M. (1995). The application of benchmark dose methodology to data from prenatal developmental toxicity studies. *Toxicol. Lett.* **82/83,** 549–554.

Klonne, D. R., Dodd, D. E., Pritts, I. M., Troup, C. M., Nachreiner, D. J., and Ballantyne, B. (1987). Acute, 9-day, and 13-week vapor inhalation studies on ethylene glycol monohexyl ether. *Fundam. Appl. Toxicol.* **8,** 198–206.

Kociba, R. J., Keeler, P. A., Park, C. N., and Gehring, P. J. (1976). 2,3,7,8-Tetrachlorodibenzo-*p*-dioxin (TCDD): Results of a 13-week oral toxicity study in rats. *Toxicol. Appl. Pharmacol.* **35,** 553–574.

Larson, P. S., Egle, J. L., Hennigar, G. R., and Borzelleca, J. F. (1979). Acute and subchronic toxicity of mirex in the rat, dog, and rabbit. *Toxicol. Appl. Pharmacol.* **49,** 271–277.

Lynch, D. W., Moorman, W. J., Stober, P., Lewis, T. R., and Iverson, W. O. (1986). Subchronic inhalation of diethylamine vapor in Fischer-344 rats: Organ system toxicity. *Fundam. Appl. Toxicol.* **6,** 559–565.

Murrell, J. A., Portier, C. J., and Morris, R. W. (1998). Characterizing dose–response I: Critical assessment of the benchmark dose concept. *Risk Anal.* **18,** 13–26.

National Toxicology Program (NTP). (1984). NTP technical report on the toxicology and carcinogenesis studies of 1,3-butadiene (case No. 106-99-0) in $B6C3F_1$ mice (inhalation studies). NTP TR288.

National Toxicology Program (NTP). (1986). NTP technical report on the toxicology and carcinogenesis tetrachloroethylene (perchloroethylene) (case No. 127-18-4) in F344/N rats and $B6C3F_1$ mice (inhalation studies). NTP TR311.

National Toxicology Program (NTP). (1986). NTP technical report on the toxicology and carcinogenesis studies of methyl methacrylate (case No. 80-62-6) in F344/N rats and B6C3F$_1$ mice (inhalation studies). NTP TR314.

National Toxicology Program (NTP). (1987). NTP technical report on the toxicology and carcinogenesis studies of ethylene oxide (case No. 75-21-8) in B6C3F$_1$ mice (inhalation studies). NTP TR326.

National Toxicology Program (NTP). (1988). NTP technical report on the toxicology and carcinogenesis studies of 1,2-epoxybutane (case No. 106-88-7) in F344/N rats and B6C3F$_1$ mice (inhalation studies). NTP TR329.

National Toxicology Program (NTP). (1993). NTP technical report on the toxicology and carcinogenesis studies of 1,3-butadiene (case No. 106-99-0) in B6C3F$_1$ mice (inhalation studies). NTP TR434.

Robinson, M., Bercz, J. P., Ringhand, H. P., Condie, L. W., and Parnell, M. J. (1991). Ten- and ninety-day toxicity studies of 1,2-dichlorobenzene administered by oral gavage to Sprague–Dawley rats. *Drug Chem. Toxicol.* **14,** 83–112.

Springer, D. L., Miller, R. A., Wright, C. W., Ragan, H. A., Buschbom, R. L., and Mahlum, D. D. (1987). Effects of subchronic inhalation exposure of mice to a high-boiling coal liquid. *Fundam. Appl. Toxicol.* **9,** 277–286.