

teria Assessment Office in Cincinnati, including particularly the comments of Dr. Roy Albert, Dr. Michael Dourson, Dr. Rick Hertzberg, Dr. Rolf Hortung, Dr. Norton Nelson, Dr. Marvin Schneiderman, and Dr. Jerry Stara. Although the research described in this article has been funded by the U.S. Environmental Protection Agency through Contract 68-03-3111, it has not been subject to the Agency's peer and administrative review and therefore does not necessarily reflect the views of the Agency, and no official endorsement should be inferred.

REFERENCES

- ACGIH (1976). *Threshold Limit Values for 1976*. Amer. Conference of Govt. Industrial Hygienists, Cincinnati, Ohio.
- ALUMOT, E., NACHTOMI, E., MANDEL, E., HOLSTEIN, P., BONDI, A., AND HERZBERG, M. M. (1976). Tolerance and acceptable daily intake of chlorinated fumigants in the rat diet. *Food Cosmet. Toxicol.* 14, 105-110.
- COX, D. R., AND LINDLEY, D. V. (1974). *Theoretical Statistics*. Chapman & Hall, London.
- CRUMP, K. S. (1984). Mechanisms leading to dose-response models. In *Principles of Health Risk Assessment* (P. F. Ficci, ed.), pp. 235-278. Prentice Hall, Inc., Englewood Cliffs, N.J.
- CRUMP, K. S., AND MASTERMAN, M. D. (1979). Assessment of carcinogenic risks from PCBs in food. In *Environmental Contaminants in Food*. Vol. II. *Working Papers*. Prepared for the U.S. Congress Office of Technology Assessment (available from NTIS).
- CRUMP, K. S., AND HOWE, R. B. (1983). Review of methods for calculating confidence limits in low dose extrapolation. In *Technological Risk Assessment* (D. B. Clayson, D. R. Krewski, and I. C. Munroe, eds.), CRC Press, Canada.
- EPA (1980). Water quality criteria documents; Availability. *Fed. Regist.* 45, No. 231 (Friday, Nov. 28), 79317-79379.
- Food Safety Council, Scientific Committee (1978). Proposed procedure for food safety assessment. *Food Cosmet. Toxicol.* 16, 1-132.
- KHERA, K. S. (1974). Teratogenicity and dominant lethal studies on hexachlorobenzene in rats. *Food Cosmet. Toxicol.* 12, 471-477.
- KHERA, K. S. (1973). Ethylenethiourea: Teratogenicity studies in rats and rabbits. *Teratology* 7, 243-252.
- KHERA, K. S., AND RUDDICK, J. A. (1973). Polychlorodibenzo-p-dioxins: Perinatal effects and the dominant lethal test in Wistar rats. In *Advances in Chemistry Series*. No. 120, *Chlorodioxins—Origins and Fate*. Amer. Chem. Soc.
- KOCIBA, R. J., KEYES, D. G., JERSEY, G. C., BALLARD, J. J., DITTENBER, D. A., QUAST, J. F., WADE, C. E., HUMISTON, C. G., AND SCHWETZ, B. A. (1977). Results of a two year chronic toxicity study with hexachlorobutadiene in rats. *Amer. Ind. Hyg. Assoc. J.* 38, 589-601.
- KREWSKI, D., AND VAN RYZIN, J. (1981). Dose response models for quantal response toxicity data. In *Statistics and Related Topics* (J. Skorgo, D. Dawson, J. N. K. Rao, and E. Saleh, eds.), pp. 201-231, North-Holland, New York.
- LOEVE, M. (1963). *Probability Theory*, 3rd ed. Van Nostrand, Princeton, N.J.
- MANTEL, N., BOHIDAR, N. R., BROWN, C. C., CIMINERA, J. L., AND TUKEY, J. W. (1975). An improved Mantel-Bryan procedure for "safety" testing of carcinogens. *Cancer Res.* 35, 865-872.
- MURRAY, F. J., SMITH, F. A., NITSCHKE, K. D., HUMISTON, C. G., KOCIBA, R. J., AND SCHWETZ, B. A. (1979). Three generation reproduction study of rats given 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD) in the diet. *Toxicol. Appl. Pharmacol.* 50, 241-252.
- NAS (1977). *Drinking Water and Health*. Safe Drinking Water Committee, National Academy of Sciences, Washington, D.C.
- WEIL, C. S. (1972). Statistics vs. safety factors and scientific judgement in the evaluation of safety for man. *Toxicol. Appl. Pharmacol.* 21, 454-463.

ADIs does not involve purely statistical methods. Toxicological evaluation of data on numerous species and biological endpoints may be required. Included in the many considerations should be differences in species sensitivities to various chemicals and the need for affording different levels of protection for different toxicological effects. The statistical methods proposed in this paper should be useful in this process but they should not supplant a careful toxicological evaluation of all the data.

APPENDIX

Description of Maximum Likelihood Procedures

Likelihood for Quantal Data

Consider an experiment with g dose levels $d_1, \dots, d_g$, and let N_i and X_i be, respectively, the number of animals tested and the number of animals affected at the i th dose level. Let $P(d)$ be the probability of a response at a dose d . Assuming that X_i has a binomial distribution with parameter N_i and $P(d_i)$, the likelihood of the data can be written as

$$L = \prod_{i=1}^g X_i^{N_i} P(d_i)^{X_i} [1 - P(d_i)]^{N_i - X_i}$$

Likelihood for Continuous Data

Consider an experiment with g dose levels $d_1, \dots, d_g$; let N_i be the number of animals in the i th dose group, and let x_{ij} , $j = 1, \dots, N_i$, $i = 1, \dots, g$ represent the response of the j th animal in the i th dose group. It is assumed that x_{ij} has a normal distribution with mean $m(d_i)$ and variance σ_i^2 . The parameters in the model consist of those involved in the definition of $m(d)$, plus $\sigma_1, \dots, \sigma_g$. Let $\bar{x}_i$ be the sample mean in the i th dose group, i.e.,

$$\bar{x}_i = \sum_{j=1}^{N_i} x_{ij} / N_i$$

and s_i^2 the sample variance, i.e.,

$$s_i^2 = \sum_{j=1}^{N_i} (x_{ij} - \bar{x}_i)^2 / (N_i - 1)$$

Then the likelihood of the data can be written as

$$L = (2\pi)^{-g/2} \prod_{i=1}^g \sigma_i^{-1} \exp[-(N_i - 1)s_i^2 - N_i(\bar{x}_i - m(d_i))^2 / \sigma_i^2]$$

Estimation and Confidence Intervals

The parameters are estimated as the values which maximize the appropriate likelihood. The "likelihood method" (Cox and Lindley, 1974; Crump and Howe, 1983) is used to calculate confidence limits. For example, when using quantal data the lower 95% limit on the dose d corresponding to an extra risk of

$$\frac{P(d) - P(0)}{1 - P(0)} = 0.1$$

is calculated as the smallest d which satisfies

$$\frac{P(d) - \hat{P}(0)}{1 - P(0)} = 0.1$$

and

$$2 \log(L_{\max}/L) = (1.645)^2$$

where $L_{\max}$ is the maximum value of the likelihood L . When using continuous data the same approach is followed except the formula for extra response replaces the one for extra risk.

Computer Programs

These methods require iterative numerical calculations. We have developed computer programs to perform these calculations and intend to have them available for the general public in the near future.

ACKNOWLEDGMENTS

The authors acknowledge the input received at a number of workshops conducted by the EPA Environmental Cri-

teria Asse
the
on.
velsc
Stara, Alt
been fund
through C
the Agenc
does not
no official

ACGIH (C
Confere
Ohio.

ALL-MOT.
BONDI,
and acc
the rat

COX, D.
Statisti

CRUMP, I
sponse
(P. F. F
glewoo

CRUMP, I
ment c
vironm
Papers

ology
P. I
for
trapola
Clayson
Press, (E
EPA (198
Fed. R
79379.

NOEL has
 ducted
 ducted in-
 f no NOEL
 rmining an
 xperiment,
 il costs and
 original ex-
 ble for cal-
 ustrated by
 l.
 ADIs would
 ore leeway
 is possible
 th the latter
 ust be spec-
 ler to ensure
 he agency's
 roach, the
 ed to specify
 m dose and
 ose, because
 not be de-
 ment, how-
 uld be given
 ls and sam-
 study and
 D. the
 oc. If an
 critical, the
 uct a large
 lacement of
 e, a smaller

ponding to

1%

129

1.4

.0010

study may be considered adequate. Any prior information on the shape of the dose-response curve could be used in optimally designing an experiment. Such prior information might come from pilot studies or studies of similar chemicals. Given such choices, proponents of chemicals should be able to design studies more in keeping with their needs and budgetary constraints without compromising safety.

As an example of how experimental design considerations could be put to effective use, suppose a company knows the smallest ADI that would permit the marketing of their product. It would be simple to calculate the benchmark that would produce this ADI. They could then design an experiment that would be optimal under the assumption that this needed benchmark is in fact the true benchmark. If the true benchmark were lower than what they were hoping for, the statistical methods used in calculating the benchmark would insure that human safety would not be compromised. On the other hand, if the benchmark were near that for which the experiment was designed, the extra care that went into the design might allow the marketing of a product that could not have been marketed if a less optimal design had been used.

Since safety factors are largely arbitrary, one rationale for choosing safety factors to use with benchmarks would be to make the resulting ADIs comparable, on average, to those calculated previously using the NOEL-SF method. This could be accomplished by calculating benchmarks for a number of substances for which ADIs have been developed from the NOEL-SF method, and then determining the safety factor that, when applied to the benchmarks, would on average yield the same ADI. Of course, ADIs calculated using the NOEL-SF and BM-SF methods could differ appreciably in specific cases.

Although we have not discussed the use of mathematical models for extrapolation of noncarcinogenesis data to low dose and thus eliminating safety factors, this is another possible application of these methods. Our re-

luctance to recommend this application stems from the uncertainty as to the shape of the dose-response curves at low doses for toxic effects in general. Dose-response curves which are linear at low doses have been used to set upper bounds for low dose cancer risks (EPA, 1980). This approach has been justified on the grounds that cancer mechanisms that would produce linear dose responses at low doses appear quite plausible and those that would produce supralinear responses seem highly implausible. The low dose linearity concept could be used to determine upper limits of risks of noncarcinogenic effects as well. However, many of these effects appear threshold-like. The assumption of a linear response could greatly overestimate risk in cases where a threshold exists. The threshold models discussed in this paper might be used to determine risks at low doses for effects which appear to be threshold-like. However, we have not recommended this in this paper because of both the uncertainty as to the existence of a threshold and because these threshold estimates are apt to differ widely depending upon the specific model used.

The model-fitting techniques proposed here have fairly minimal data requirements. When quantal data are used, the basic needs are the doses, number of animals in each group, and the number of these animals which are affected. With continuous data one needs the doses, number of animals in each group, the average response in each group, and the standard errors of these responses. Some effects, such as cloudy swelling of the liver, are inherently difficult to quantify and are normally classified qualitatively, such as by present/absent or mild/severe. Even for effects which are quantifiable, the data needed to apply dose-response methods are frequently not reported in the literature. Thus, it will not be possible to apply these methods universally. However, the introduction of these methods would encourage more complete presentation of data, as well as generally encouraging the use of quantitative methods in toxicology.

It should be kept in mind that determining

TABLE 13
DOSES CORRESPONDING TO GIVEN LEVELS OF EXTRA
RESPONSE FOR CONTINUOUS TCDD DATA^a

Model	Extra response	Dose (μg/kg/day)	
		MLE	95% lower
CLR	0.1	2.55-3 ^b	1.32-3
CPR		2.55-3	1.32-3
CP (no threshold)		6.37-3	1.32-3
CLR	0.05	1.78-3	6.61-4
CPR		1.78-3	6.61-4
CP (no threshold)		5.53-3	6.61-4
CLR	0.01	1.16-3	1.32-4
CPR		1.16-3	1.32-4
CP (no threshold)		1.21-3	1.32-4
CLR	0.001	1.02-3	1.32-5
CPR		1.02-3	1.32-5
CP (no threshold)		1.96-3	1.32-5

^a Source: Murray *et al.*, 1979.

^b 2.55-3 means $2.55 \times 10^{-3} = .00255$.

A BD is calculated using a mathematical dose-response curve estimated from all of the dose-response data. Thus the benchmark should better reflect the shape of the dose response than the NOEL. Because a benchmark represents a statistical lower limit, larger experiments will tend on average to give larger benchmarks, thus rewarding good experimentation. As we pointed out, NOELs have the opposite tendency. With the NOEL approach,

ADIs cannot be determined until a NOEL has been established. An otherwise well-conducted experiment may therefore be considered inappropriate for calculating an ADI if no NOEL is established. In such a case, determining an ADI could require an additional experiment, resulting in considerable additional costs and delays. On the other hand, the original experiment might be quite acceptable for calculating a BD. This situation is illustrated by the quantal data for HCB (Fig. 6).

A BM-SF approach to setting ADIs would allow proponents of chemicals more leeway in the design of experiments than is possible under the NOEL-SF approach. With the latter method minimum sample sizes must be specified by the regulatory agency in order to ensure that NOELs are established to the agency's satisfaction. With a BM-SF approach, the agency would still in some cases need to specify methods for choosing the maximum dose and the sample size to be used at this dose, because otherwise important effects might not be detected at all. Beyond this requirement, however, proponents of a chemical could be given wide latitude in selecting dose levels and sample sizes. Of course, the larger a study and better designed it is to estimate the BD, the higher the benchmark is liable to be. If an accurate benchmark is considered critical, the experimenters may wish to conduct a large study and consider carefully the placement of the experimental doses; otherwise, a smaller

TABLE 14
COMPARISON OF BENCHMARK DOSES WITH NOELs FOR CONTINUOUS DATA

Data set	Dose units	NOEL	Benchmark doses ^a corresponding to % extra risk		
			10%	5%	1%
Carbon tetrachloride (Alumot <i>et al.</i> , 1976)	ppm	150? ^b	141	134	129
HCBD (Kociba <i>et al.</i> , 1977)					
mean body weights	mg/kg/day	2.0?	14.1	7.0	1.4
TCDD (Murray <i>et al.</i> , 1979)	μg/kg/day	0.001?	.0026	.0012	.0010

^a Benchmark doses = 95% lower limits derived from QPR model.

^b ? indicates that it is doubtful whether a NOEL has been established.

TABLE 10
SUMMARY OF FITS TO MODELS TO CONTINUOUS DATA IN TABLE 9

Data	Model ^a	F statistic	df	p value
Carbon tetrachloride (Alumot <i>et al.</i> , 1976)	CLR	0.29	(1, 20)	NS ^b
	CPR	0.29	(1, 20)	NS
	CP	0.29	(1, 20)	NS
	CP (no threshold)	1.25	(2, 20)	NS
HCBd (Kociba <i>et al.</i> , 1977), mean body weights	CLR	0.14	(2, 206)	NS
	CPR	0.14	(2, 206)	NS
	CP	0.14	(2, 206)	NS
	CP (no threshold)	0.14	(2, 206)	NS
TCDD (Murray <i>et al.</i> , 1979)	CLR	0		NS
	CPR	0		NS
	CP (no threshold)	0		NS

^a Code: CLR = continuous linear regression, CPR = continuous polynomial regression, CP = continuous power.

^b NS = not significant (*p* value greater than 0.1).

to a specific increase in risk between 1 and 10%. It is suggested that such a BD replace the traditional NOEL. We believe this ap-

proach mitigates several of the problems raised in Section II concerning the NOEL-SF method.

TABLE 11

DOSES CORRESPONDING TO GIVEN LEVELS OF EXTRA RESPONSE FOR CONTINUOUS CARBON TETRACHLORIDE DATA^a

Model	Extra response	Dose (ppm)	
		MLE	95% lower
CLR	0.1	141	102
CPR		141	63
CP		141	67.2
CP (no threshold)		95.6	47.1
CLR	0.05	134	94.1
CPR		134	37.6
CP		134	44.7
CP (no threshold)		68.0	29.2
CLR	0.01	129	87.9
CPR		129	9.48
CP		129	17.3
CP (no threshold)		30.8	9.53
CLR	0.001	127	86.5
CPR		127	1.03
CP		127	4.4
CP (no threshold)		9.90	1.89

^a Source: Alumot *et al.*, 1976.

TABLE 12

DOSES CORRESPONDING TO GIVEN LEVELS OF EXTRA RESPONSE FOR CONTINUOUS HCBd DATA ON MEAN BODY WEIGHTS^a

Model	Extra response	Doses (mg/kg/day)	
		MLE	95% lower
CLR	0.1	14.1	9.14
CPR		14.1	9.14
CP		14.1	9.14
CP (no threshold)		14.1	9.14
CLR	0.05	7.03	4.57
CPR		7.03	4.57
CP		7.03	4.57
CP (no threshold)		7.03	4.57
CLR	0.01	1.41	9.14-1 ^b
CPR		1.40	9.14-1
CP		1.40	9.14-1
CP (no threshold)		1.41	9.14-1
CLR	0.001	1.40-1	9.14-2
CPR		1.41-1	9.14-2
CP		1.41-1	9.14-2
CP (no threshold)		1.41-1	9.14-2

^a Source: Kociba *et al.*, 1977.

^b 9.14-1 means $9.14 \times 10^{-1} = 0.914$.

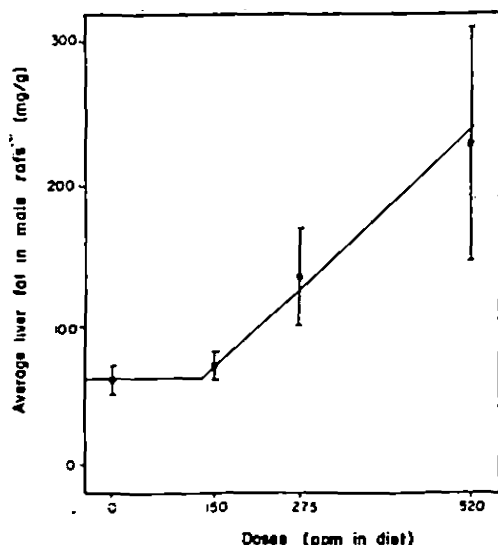


FIG. 8. Mean liver fat in rats exposed to carbon tetrachloride (Alumot *et al.*, 1976), with 90% confidence bars and best-fitting continuous linear regression model.

test was assumed. Also, values reported by Kociba *et al.* as "s.d." were assumed to mean "s.e." As Table 10 shows, all of the models fit each of these data sets adequately.

Tables 11-13 show that the estimates of doses corresponding to given levels of extra response calculated using the four models are

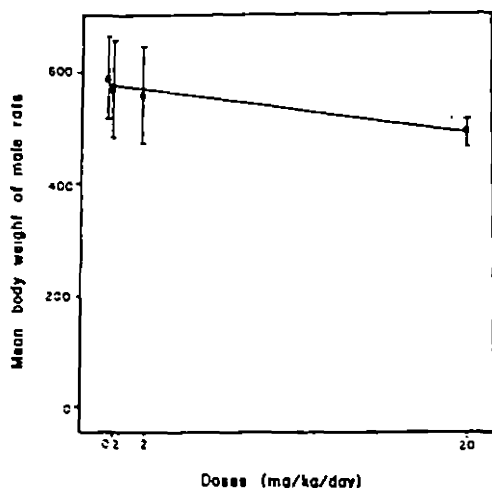


FIG. 9. Mean body weights in rats exposed to HCBD (Kociba *et al.*, 1977), with 90% confidence bars and best-fitting continuous linear regression model.

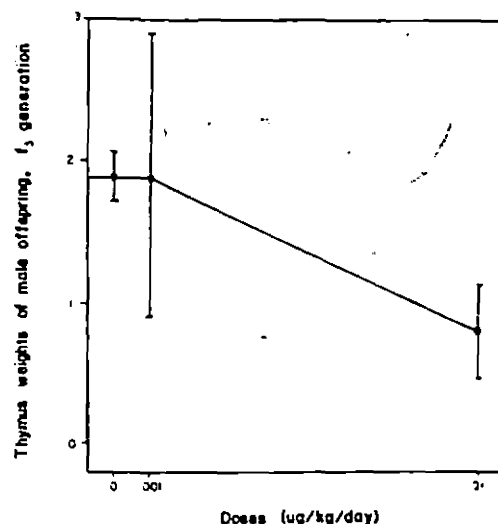


FIG. 10. Mean thymus weights of male offspring, F_3 generation (Murray *et al.*, 1979), with 90% confidence bars and best-fitting continuous linear regression model.

quite similar. In fact, corresponding lower confidence limits are almost identical in Tables 12-14. In Table 11 the 95% lower limits differ by as much as a factor of 2 for an extra response of 0.01 and by larger amounts for smaller values of extra response.

Table 14 compares BDs with NOELs for the continuous data. Question marks are included beside the NOELs because it is not clear when a NOEL has been determined. For example, although for the data for carbon tetrachloride the average liver fat in 150-ppm animals is not statistically different from that of control animals, there is an increase at 150 ppm that appears to be part of a dose-response trend (Fig. 8). For these three data sets, the BD corresponding to an extra response of 1% are roughly comparable to the NOELs.

V. DISCUSSION

In this paper we have examined an alternative to the NOEL-SF approach which involves fitting a mathematical model to toxicological dose-response data. The model is used to define a BD, which represents a statistical lower limit on the dose corresponding

Carbon t
et al.

HCBD (f
mean)

TCDD (f

^a Code

^b NS =

to a sp
10%. It
the tra

DOSES
RESPONS

Mc

CLR
CPR
CP
CP (no t

CLR
CPR
CP
CP (no t

CLR
CPR
CP
CP (no t

CLR
CPR
CP
CP (no t

^a Sour

TABLE 8
COMPARISON OF BENCHMARK DOSES WITH NOELS FOR QUANTAL DATA

mg/kg)	99% lower	Data set	Dose units	NOEL	Benchmark doses ^a corresponding to % extra risk		
					10%	5%	1%
4	16.0	ETU (Khera, 1973)	mg/kg	5	13.9	11.6	7.2
4	16.0	TCDD (Khera and Ruddick, 1973)	µg/kg	0.125	0.32	0.22	0.049
4	16.0	TCDD (Murray <i>et al.</i> , 1979)	µg/kg/day	1.0-3	5.3-3	2.6-3	5.1-4
2	12.2	HCB (Khera, 1974)	mg/kg	ND ^b	17.4	8.5	1.7
5	7.8	Botulinum toxin (Food Research Institute, Univ. of Wisconsin)	ng	0.027	0.029	0.027	0.025

^a Benchmark doses = 95% lower limits derived from QPR model.

^b ND = not determined.

NOEL approach. Since a NOEL was not determined, the NOEL-SF method can not be used with these data to determine an ADI. However, these data would present no difficulty in determining an ADI from a BD.

Examples of Benchmark Doses Calculated from Continuous Data

Table 9 contains dose-response data on liver fat in rats after exposure to carbon tet-

rachloride (Alumot *et al.*, 1976), mean body weights in rats after exposure to hexachlorobutadiene (HCBD) (Kociba *et al.*, 1977), and thymus weights in rats after exposure to TCDD (Murray *et al.*, 1979). Figures 8-10 contain graphs of the responses and 90% confidence intervals, along with the dose-response curve obtained by fitting the continuous polynomial regression (CPR) model to the data. In the Kociba *et al.* data numbers of animals were not provided and the total number on

ELS OF EXTRA VINT DATA^a

% ver	99% lower
4-2	2.8-2
4-2	2.8-2
7-2	2.6-2
4-2	3.0-2
8-2	2.7-2
9-2	2.6-2
2-2	2.2-2
9-2	2.8-2
7-2	2.7-2
5-2	2.4-2
5-2	1.5-2
5-2	2.4-2
7-2	2.6-2
3-2	2.2-2
8-3	1.5-3
6-2	1.4-2

TABLE 9

CONTINUOUS DATA USED TO ILLUSTRATE QUANTITATIVE DOSE-RESPONSE METHODOLOGY

<i>Carbon tetrachloride (Alumot et al., 1976)</i>				
Average liver fat in male rats				
Dose (ppm in diet)	0	150	275	520
Ave. $\pm$ SE (mg/g)	61.0 $\pm$ 6.6	71.0 $\pm$ 6.0	136 $\pm$ 21	229 $\pm$ 49
No. of animals	6	6	6	6
<i>Hexachlorobutadiene (HCBD) (Kociba et al., 1977)</i>				
Mean body weight of male rats				
Dose (mg/kg/day)	0	0.2	2.0	20.0
Ave. $\pm$ SE (gm)	586 $\pm$ 43	568 $\pm$ 53	557 $\pm$ 52	494 $\pm$ 15
No. of animals	90	40	40	40
<i>2,3,7,8-Tetrachlorodibenzo-p-dioxin (TCDD) (Murray et al., 1979)</i>				
Thymus weights of male offspring, f_3 generation				
Dose (µg/kg/day)	0	0.001	0.01	
Ave. $\pm$ SE (g)	0.19 $\pm$ 0.01	0.19 $\pm$ 0.06	0.08 $\pm$ 0.02	
No. of animals	5	5	4	

TABLE 5

DOSES CORRESPONDING TO GIVEN LEVELS OF EXTRA RISK FOR QUANTAL MURRAY POSTNATAL SURVIVAL DATA^a

Model	Extra risk	Dose ($\mu\text{g/kg/day}$)		
		MLE	95% lower	99% lower
QLR	0.1	9.3-3 ^b	5.3-3	4.4-3
QPR			Same as QLR	
QW		9.6-3	5.3-3	4.4-3
LN		9.3-3	4.6-3	3.5-3
QLR	0.05	4.9-3	2.6-3	2.1-3
QPR			Same as QLR	
QW		6.2-3	2.6-3	2.1-3
LN		5.5-3		1.5-3
QLR	0.01	1.6-3	6.1-4	4.2-4
QPR			Same as QLR	
QW		2.3-3	5.1-4	4.2-4
LN		2.0-3	4.1-4	3.2-4
QLR	1×10^{-6}	8.0-4	5.1-8	4.2-8
QPR			Same as QLR	
QW		9.6-6	5.0-8	4.2-8
LN		6.0-5	1.5-6	1.2-6

^a Source: Murray *et al.*, 1979.

^b 9.3-3 means 9.3×10^{-3} .

hibit a very rapid rise in response for doses larger than the NOEL of 27 ng. Neither of the nonthreshold models—QW or LN—fit these data. However, both the QLR and QPR models fit quite adequately (Table 2). All four of the models give comparable doses corresponding to given extra risk levels, even down to levels of extra risk of 10^{-6} .

Table 8 compares NOELs with BDs corresponding to three levels of extra risk. With the exception of the HCB data, the NOELs generally correspond to the BDs for extra risks between 0.01 and 0.05. However, these were all reasonably large studies and involved effects not seen in control animals; for smaller studies or for effects which can occur spontaneously, NOELs are liable to be larger relative to the BDs.

The data for HCB illustrate a particular advantage the benchmark approach has over the

TABLE 6

DOSES CORRESPONDING TO GIVEN LEVELS OF EXTRA RISK FOR QUANTAL HCB DATA^a

Model	Extra risk	Dose (mg/kg)		
		MLE	95% lower	99% lower
QLR	0.1	21.7	17.4	16.0
QPR		22.3	17.4	16.0
QW		22.0	17.4	16.0
LN		21.0	14.2	12.2
QLR	0.05	10.6	8.5	7.8
QPR		11.0	8.5	7.8
QW		10.8	8.5	7.8
LN		11.1	6.1	5.3
QLR	0.01	2.1	1.7	1.5
QPR		2.2	1.7	1.5
QW		2.2	1.7	1.5
LN		3.4	1.3	1.1
QLR	1×10^{-6}	2.1-4 ^b	1.7-4	1.5-4
QPR		2.2-4	1.7-4	1.5-4
QW		2.6-4	1.7-4	1.5-4
LN		5.0-2	4.8-3	4.1-3

^a Source: Khera, 1974.

^b 2.1-4 means 2.1×10^{-4} .

TABLE 7

DOSES CORRESPONDING TO GIVEN LEVELS OF EXTRA RISK FOR QUANTAL BOTULINUM TOXIN DATA^a

Model	Extra risk	Dose (ng)		
		MLE	95% lower	99% lower
QLR	0.1	3.0-2 ^b	2.8-2	2.8-2
QPR		3.0-2	2.9-2	2.8-2
QW		3.3-2	2.7-2	2.6-2
LN		3.3-2	3.1-2	3.0-2
QLR	0.05	2.9-2	2.8-2	2.7-2
QPR		2.8-2	2.9-2	2.6-2
QW		3.0-2	2.2-2	2.2-2
LN		3.1-2	2.9-2	2.8-2
QLR	0.01	2.9-2	2.7-2	2.7-2
QPR		2.7-2	2.5-2	2.4-2
QW		2.4-2	1.6-2	1.5-2
LN		2.9-2	2.5-2	2.4-2
QLR	1×10^{-6}	2.9-2	2.7-2	2.6-2
QPR		2.7-2	2.3-2	2.2-2
QW		7.3-3	1.8-3	1.5-3
LN		2.0-2	1.6-2	1.4-2

^a Source: FSC, 1978.

^b 3.0-2 means 3.0×10^{-2} .

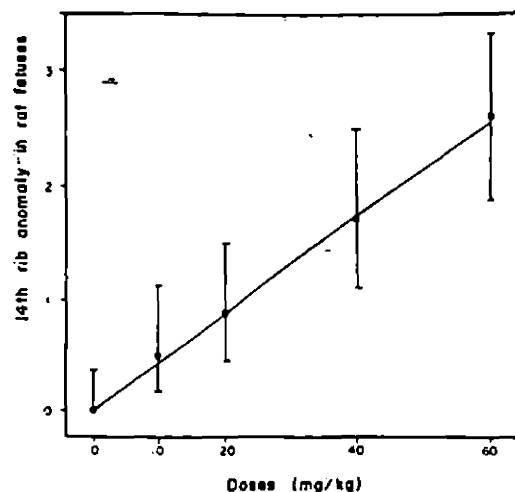


FIG. 6. Probability of 14th rib anomaly from exposure to HCB (Khera, 1974), with 90% confidence bars and best-fitting polynomial regression model.

Both the Murray *et al.* (1979) TCDD data on postnatal survival and the Khera (1974) HCB data on rib anomalies are nearly linear and are described well by all four of the models. Doses calculated from the models corresponding to extra risks of 0.1, 0.05, and 0.01 are also in close agreement.

The data on probability of death after exposure to Botulinum Toxin (FSC, 1978) ex-

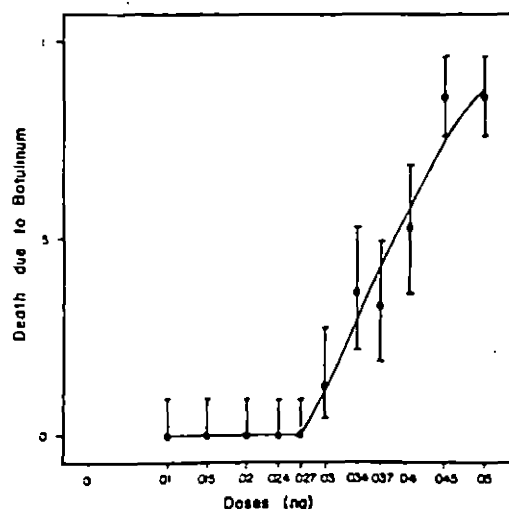


FIG. 7. Probability of death from exposure to botulinum toxin—Type A (Food Research Inst., Univ. of Wisconsin, see FSC, 1978), with 90% confidence bars and best-fitting polynomial regression model.

TABLE 3
DOSES CORRESPONDING TO GIVEN LEVELS OF EXTRA RISK FOR QUANTAL ETU DATA^a

Model	Extra risk	Dose (mg/kg)		
		MLE	95% lower	99% lower
QLR ^b	0.1	12.2	11.8	11.6
QPR		16.4	13.9	13.2
QW		17.9	15.7	14.7
LN		17.1	15.3	14.6
QLR ^b	0.05	11.0	10.6	10.4
QPR		13.2	11.6	11.0
QW		14.5	12.2	11.3
LN		14.8	13.0	12.2
QLR ^b	0.01	10.1	9.7	9.5
QPR		10.2	7.2	6.0
QW		8.9	7.0	6.2
LN		11.2	9.5	8.8
QLR ^b	1×10^{-6}	9.9	9.5	9.2
QPR		9.4	4.0-1	1.5-3 ^b
QW		5.9-1	2.9-1	2.1-1
LN		4.2	3.1	4.7

^a Source: Khera, 1977.

^b 1.5-3 means 1.5×10^{-3} .

TABLE 4
DOSES CORRESPONDING TO GIVEN LEVELS OF EXTRA RISK FOR QUANTAL TCDD DATA^a

Model	Extra risk	Dose (mg/kg)		
		MLE	95% lower	99% lower
QLR	0.1	4.6-1 ^b	3.2-1	2.8-1
QPR		5.0-1	3.2-1	2.8-1
QW		5.2-1	3.6-1	3.0-1
LN		4.9-1	3.5-1	2.9-1
QLR	0.05	3.1-1	2.1-1	1.5-1
QPR		3.3-1	2.2-1	1.5-1
QW		3.6-1	2.0-1	1.5-1
LN		3.5-1	2.1-1	1.4-1
QLR	0.01	2.0-1	5.7-2	3.0-2
QPR		1.7-1	4.9-2	3.0-2
QW		1.6-1	4.3-2	2.9-2
LN		1.8-1	6.4-2	3.0-2
QLR	1×10^{-6}	1.7-1	7.8-3	3.0-6
QPR		1.3-1	4.9-6	3.0-6
QW		1.6-3	4.3-6	2.9-6
LN		2.0-2	5.4-4	1.1-4

^a Source: Khera and Ruddick, 1973.

^b 4.6-1 means 4.6×10^{-1} .

TABLE 2
SUMMARY OF FITS OF MODELS TO QUANTAL
DATA IN TABLE 1

Data	Model*	χ^2	df	p value
ETU	QLR	17	3	0.0007 X
	QPR	0.0	2	1.0
	QW	1.3	2	0.73
	LN	0.46	2	0.93
TCDD (Khera and Ruddick, 1973)	QLR	0.17	2	0.92
	QPR	0.014	1	0.91
	QW	0.32	3	0.85
	LN	0.23	3	0.89
TCDD (Murray <i>et al.</i> , 1979)	QLR	0	0	
	QPR	0	0	
	QW	0	0	
	LN	0	0	
HCB	QLR	0.11	2	0.95
	QPR	0.09	1	0.76
	QW	0.11	2	0.95
	LN	0.31	2	0.86
Botulinum toxin	QLR	7.0	8	0.54
	QPR	4.4	7	0.73
	QW	162	8	0.00001 X
	LN	159	8	0.00001 X

* Code: QLR = quantal linear regression, QPR = quantal polynomial regression, QW = quantal Weibull, LN = log-normal.

The ETU data involve a sizable number of animals and are characterized by a NOEL at 5 mg/kg followed by a steeply rising dose response that reaches 100% response at 80 ppm. Each of the models except the QLR model fits these data quite adequately. Also, except for the QLR model, MLEs and lower confidence limits for doses corresponding to various levels of increased risk computed using the various models agree rather closely for extra risks of 0.1, 0.05, and 0.01. However, doses corresponding to extra risks of 10^{-6} differ by larger amounts.

The Khera and Ruddick (1973) TCDD data on intestinal anomalies involve a dose-related increase in response for doses larger than a NOEL of 0.125 $\mu\text{g}/\text{kg}$. However, the confidence intervals on responses at the experimental doses are wider than those for the ETU

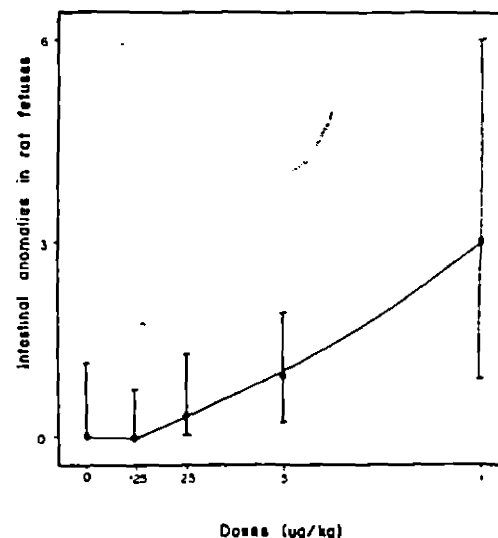


FIG. 4. Probability of intestinal anomaly in rat fetuses from exposure to TCDD (Khera *et al.*, 1973), with 90% confidence bars and best-fitting polynomial regression model.

data (Figs. 3 and 4). As a result, all four of the models fit these data quite well (Table 2). Predictions of the four models also agree closely down to extra risks of 0.01 but differ considerably at extra risks of 10^{-6} .

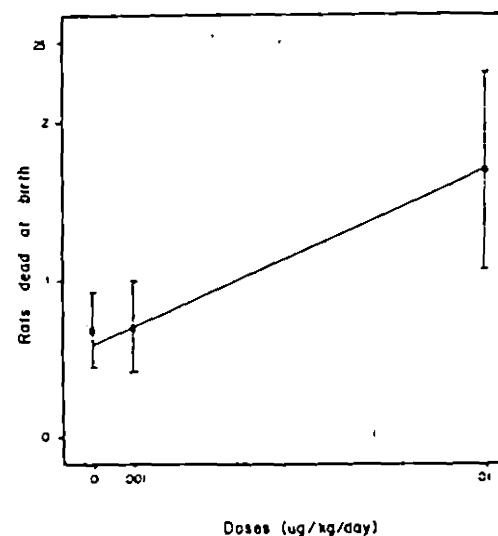


FIG. 5. Probability of fetal death in rats from exposure to TCDD (Murray *et al.*, 1979), with 90% confidence bars and best-fitting polynomial regression model.

TABLE 1

QUANTAL DATA USED TO ILLUSTRATE QUANTITATIVE DOSE-RESPONSE METHODOLOGY

Ethylenethiourea (ETU) (Khera, 1973)

Fetal anomalies in rats

Doses (mg/kg)	0	5	10	20	40	80
No. affected/total No.	0/167	0/132	1/138	14/81	142/178	24/24

2,3,7,8-Tetrachlorodibenzo-p-dioxin (TCDD) (Khera and Ruddick, 1973)

Intestinal anomalies in rat fetuses

Doses (μ g/kg)	0	0.125	0.25	0.5	1.0
No. affected/total No.	0/24	0/38	1/33	3/31	3/10

2,3,7,8-Tetrachlorodibenzo-p-dioxin (TCDD) (Murray et al., 1979)

Rats dead at birth

Doses (μ g/kg/day)	0	0.001	0.01
No. affected/total No.	22/318	16/224	17/100

Hexachlorobenzene (HCB) (Khera, 1974)

14th rib anomaly in rat fetuses

Doses (mg/kg)	0	10	20	40	60
No. affected/total No.	0/80	4/79	8/91	15/87	25/96

Botulinum toxin—Type A (Food Research Inst., Univ. of Wisconsin) (FSC, 1978)

Death due to Botulinum

Doses (ng)	.01	.015	.020	.024	.027	.030
No. dead/total No.	0/30	0/30	0/30	0/30	0/30	4/30
Doses (ng)	.034	.037	.040	.045	.050	
No. dead/total No.	11/30	10/30	16/30	26/30	26/30	

Examples of Benchmark Doses Calculated from Quantal Data

To illustrate application of the benchmark approach to quantal data, we have applied it to five sets of quantal dose-response data, involving exposures to ethylenethiourea (ETU); 2,3,7,8-tetrachlorodibenzo-*p*-dioxin (TCDD) (two data sets); hexachlorobenzene (HCB); and botulinum toxin—Type A (BT-A). The data are listed in Table 1. A summary of the fits of the four models to these data is given in Table 2. Graphs of the data, along with the fit of the QPR model (which was the only model that fit all five data sets adequately) are given in Figs. 3–7. Doses corresponding to various levels of extra risk are furnished in Tables 3–7.

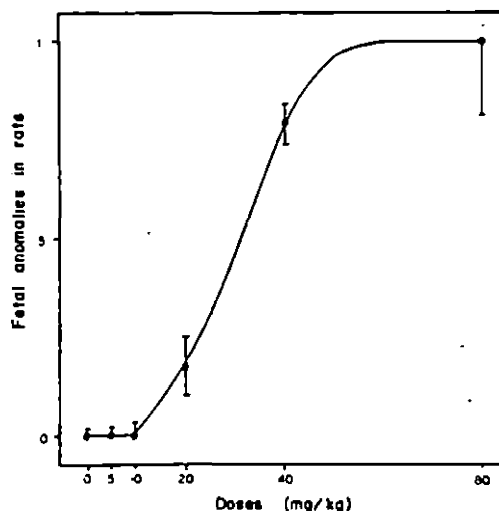


FIG. 3. Probability of fetal anomaly in rats (Khera et al., 1977) from exposure to ETU with 90% confidence bars and best-fitting polynomial regression model.

response models. A standard method for computing confidence limits is to base them upon the asymptotic normal distribution of maximum likelihood estimates. However, these confidence limits have been shown to behave poorly in a low dose extrapolation setting (Crump and Masterman, 1979; Krewski and Van Ryzin, 1981; Crump and Howe, 1983); the upper and lower limits are often too close together to be believable. Further, these limits are not invariant under parameter transformations, and different transformations applied to the same model at low doses can yield vastly different confidence bounds. Cox and Lindley (1974) noted these difficulties in a more general context, and argued that confidence limits based upon the asymptotic distribution of the likelihood ratio statistic "can be expected to behave much more sensibly" than those based upon the asymptotic normality of maximum likelihood estimates. Crump and Howe (1983) reviewed confidence limit procedures for use in dose-response evaluations and recommended limits based upon the distribution of the likelihood ratio statistic as the method of choice. This method for constructing confidence limits is outlined in the Appendix and will be used exclusively throughout this paper.

IV. THE BENCHMARK-SAFETY FACTOR METHOD FOR COMPUTING ADIs

In this section we examine the implications of modifying the NOEL-SF method for calculating ADIs by replacing the NOEL by a "benchmark dose" (BD) calculated using the methods described in the last section. A BD is defined as a lower statistical confidence limit for the dose corresponding to a specified increase in level of health effect over the background level. The increased level of effect upon which the BD is based would be near the lower limit of the experimental range; i.e., near the lower limit of increases in health effects which can be measured with reasonable accuracy in toxicological studies. This value is estimated

to be something on the order of a 10% change from background at typical sample sizes. Benchmark doses calculated in this fashion will have several advantages over NOELs. They will reflect the dose-response pattern to a much greater degree than NOELs. They will also make more reasonable use of sample size (larger experiments will tend to produce larger BDs, which is not true of NOELs). It will not be necessary to define a NOEL in order to determine an ADI. Because these BDs correspond to risks in the experimental range, their value will not depend strongly upon the particular dose-response model used in their calculation.

For quantal data we define the BD to be a dose d which corresponds to a specified value for the extra risk

$$[P(d) - P(0)]/[1 - P(0)].$$

Extra risk can be interpreted as the probability of an effect at dose d given that no effect would have occurred in the absence of the dose. This interpretation is valid irrespective of whether there is independent action between the background and the stimulus. Extra risk places greater weight upon the same increase in rate for a common lesion than for a rare lesion. For example, it takes an increase of 9% above the background level for a lesion with a 10% background rate to attain a 10% extra risk compared to only a 5% increase for a lesion with a 50% background rate. Because of this property, some may prefer using the additional risk $P(d) - P(0)$ to extra risk.

For continuous data we define the BD to be a dose d which corresponds to a specified amount of absolute change in the mean value relative to the mean value in the absence of the dose—i.e., the dose d corresponding to a specific value for the "extra response"

$$\frac{m(d) - m(0)}{m(0)}.$$

Other terms, such as the standard error of the responses in control group, could be used in the denominator in place of $m(0)$ to normalize this expression.

ed are caused by
are poorly
it would
develop dose-
led assumptions
Instead, we pro-
ble generic mod-
ve shall consider
quantal linear

$$-q_1(d - d_0)\} \\ \text{for } d \geq d_0 \\ (1)$$

≥ 0 ; the quantal
model

$$-q_1(d - d_0) \\ \} \text{ for } d \geq d_0 \\ (2)$$

≥ 0 for $i = 1$,
QW) model

$$\exp(-ad^k)\} (3)$$

$k \geq 1$; and the

$$+ b \log d) (4)$$

is the standard

carcinogenesis
recognize (1) and
ions of, respec-
stage models as
a (Krewski and
been modified
se d_0 ; doses be-
no effect. Thus
possibility that
ne effects; how-
plied with the
h we have not
be included in
models.

Note that the restrictions $k \geq 1$ and $b \geq 1$ are assumed for the Weibull and log-normal models, respectively. Some restrictions of this nature seem necessary with these models; otherwise these models can exhibit very extreme and biologically implausible behavior. The restriction $k \geq 1$ was selected for the Weibull model because $k < 1$ corresponds to a supralinear curve shape which is implausible for any biological effect (Crump, 1984). Although the restriction $b \geq 1$ for the slope parameter in the log-normal model does not have a strong theoretical basis, it does have a precedent, as it was recommended by Mantel *et al.* (1975) for cancer data.

Dose-Response Methods for Continuous Data

Continuous dose-response data consist of the dose level and the response level for each animal. With most continuous effects there will be variation about a nonzero value in the control group. There has been little experience in applying dose-response models to such data. It is possible to convert continuous data to quantal data by considering all animals with responses beyond a particular value as "affected" and all others as "unaffected." However, this procedure entails a considerable loss of information as well as requiring the arbitrary choice of a cut-off value. The following method makes more complete use of the data.

The method will be based upon the supposition that the responses in an animal group subject to a dose d_i are normally distributed with mean $m(d_i)$ and variance σ_i^2 . There is both a theoretical reason and a pragmatic reason for assuming the normal distribution. First, by the Central Limit Theorem of probability theory (Loeve, 1963) the sample means will be approximately normally distributed for large samples regardless of the form of the underlying distribution. Second, with the normal distribution, maximum likelihood methods can be applied knowing only the doses $d_1, \dots, d_g$, the numbers of animals at each dose $n_1, \dots, n_g$, and the corresponding sample

means and standard errors $(\bar{x}_1, s_1), (\bar{x}_2, s_2), \dots, (\bar{x}_g, s_g)$. If a non-normal distribution were assumed, maximum likelihood methods would require knowledge of the individual animal responses, which usually are not readily available. The choice of the normal distribution is not a critical decision as this distribution only determines the error structure, and not the dose response. The mean function $m(d)$, which represents the average response at a dose d , determines the dose response. We do not require any assumptions regarding the variances (other than that they are finite); it is not necessary to assume, for example, that the variances in the different dose groups are all equal.

For illustrative purposes, we will consider the following forms for $m(d)$: the continuous linear regression (LCR) model

$$m(d) = c + q_1(d - d_0) \text{ for } d \geq d_0 \\ = c \text{ for } d < d_0 \quad (5)$$

where $d_0 \geq 0$, but c and q_1 are unrestricted; the continuous polynomial regression (CPR) model

$$m(d) = c + q_1(d - d_0) + \dots + q_k(d - d_0)^k \\ \text{for } d \geq d_0 \\ = c \text{ for } d < d_0 \quad (6)$$

where $d_0 \geq 0$ and the q_i 's are restricted to be either all positive (increasing dose response) or all negative (decreasing dose response); and the continuous power (CP) model

$$m(d) = c + q_1(d - d_0)^k. \quad (7)$$

These models are analogous to (1), (2), and (3), which were suggested for use with quantal data.

With both quantal and continuous data, in addition to the selection of a dose-response model, the proper use of statistical confidence limits is also of critical importance. Often, different confidence limit procedures yield different results; this makes it important to use the same procedures when comparing dose-

the costs involved. The NOEL-SF approach does not lend itself to cost-benefit analyses.

III. FITTING DOSE RESPONSE MODELS TO TOXICOLOGICAL DATA

Quantitative toxicological data are basically of two types: quantal and continuous. Quantal or incidence data specify the number of animals affected, but not the degree of harm. The numbers of animals with tumors or some genetic anomaly are examples of quantal data. On the other hand, with continuous data the level of harm is specified for each animal. Organ weights, triglyceride levels in liver, and serum measurements are examples of effects that are usually recorded as continuous data.

Dose Response Methods for Quantal Data

Quantal data from a toxicological experiment can be represented as a collection of triplets (N_i, X_i, d_i) —one triplet for each treatment or control group—where N_i is the number of animals in the i th group, X_i is the number of affected animals, and d_i is the dose. Let $P(d)$ represent a dose-response model applicable to quantal data, where $P(d)$ is the probability that an effect will occur in an animal subject to a dose d . The parameters of the model can be estimated by fitting the model to quantal dose-response data using maximum likelihood procedures. A number of dose-response models have been suggested for use with cancer data. Some of these, such as the one-hit, multistage, multihit, and Weibull models can be derived from detailed assumptions about carcinogenic mechanisms. Other models, such as the probit or logit models, can be thought of as representing the distribution of individual tolerances in a large population (Krewski and Van Ryzin, 1981).

In the next section some ways are suggested for applying mathematical models to non-cancer data. The toxic endpoints to which

these methods could be applied are caused by diverse mechanisms, most of which are poorly understood. Therefore, it seems that it would not be fruitful to attempt to develop dose-response models from detailed assumptions regarding these mechanisms. Instead, we propose the use of relatively simple generic models. For illustrative purposes we shall consider the following models: the quantal linear regression (QLR) model

$$P(d) = c + (1 - c)\{1 - \exp[-q_1(d - d_0)]\} \quad \text{for } d \geq d_0$$

$$= c \quad \text{for } d < d_0 \quad (1)$$

where $0 \leq c \leq 1$, $d_0 \geq 0$, $q_1 \geq 0$; the quantal polynomial regression (QPR) model

$$P(d) = c + (1 - c)\{1 - \exp[-q_1(d - d_0) - \dots - q_k(d - d_0)^k]\} \quad \text{for } d \geq d_0$$

$$= c \quad \text{for } d < d_0 \quad (2)$$

where $0 \leq c \leq 1$, $d_0 \geq 0$, $q_i \geq 0$ for $i = 1, \dots, k$; the quantal Weibull (QW) model

$$P(d) = c + (1 - c)[1 - \exp(-ad^k)] \quad (3)$$

where $0 \leq c \leq 1$, $a \geq 0$, and $k \geq 1$; and the log-normal (LN) model

$$P(d) = c + (1 - c)N(a + b \log d) \quad (4)$$

where $0 \leq c \leq 1$, $b \geq 1$ and N is the standard normal distribution function.

Readers familiar with the carcinogenesis dose-response literature will recognize (1) and (2) as slightly modified versions of, respectively, the one-hit and multistage models as applied to carcinogenesis data (Krewski and Van Ryzin, 1981). Each has been modified here to include a threshold dose d_0 ; doses below this threshold produce no effect. Thus these models allow for the possibility that thresholds could exist for some effects; however, the models could be applied with the threshold fixed to 0. Although we have not done so, thresholds could also be included in the Weibull and log-normal models.



with 95% confidence

d_{2B} for Experi-
methods to be
capable of fuller
trends.

in Entail Unnec-
expense

A company
complements
ing program as
agency involved.
a 2-year chronic
bioassay, and a
and teratology
ee treatment and
and sample sizes
reproduction and
fact, the treated
n the control an-
ated to the fact
base. The 2-year
cts of treatment
eight reduction
he fact that the
al in such high
l was distasteful.
eight loss follows

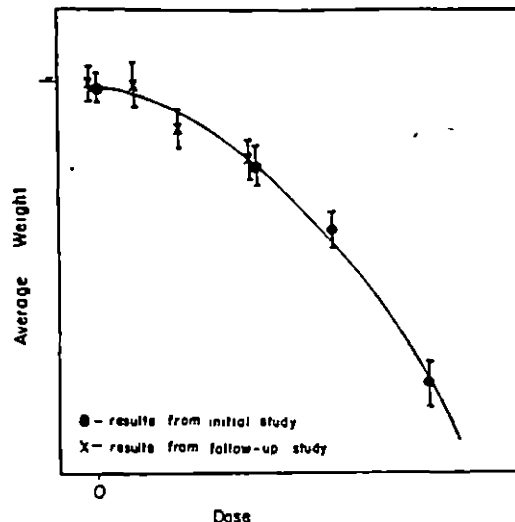


FIG. 2. Hypothetical average weights with 95% confidence limits from two studies (see text for description).

a clear dose-response trend and the effect is statistically significant at the lowest dose tested. The agency rules that a no-effect level has not been determined and thus there is no basis for calculating a ADI. The company is then required to conduct another 2-year study. This study also uses three treatment levels, the highest of which coincides with the lowest used in the previous study. A dose-response trend is also obtained in the follow-up study as illustrated in Fig. 2. The data at the highest dose reproduce almost exactly the results in the earlier study at that dose level. The weight loss at the middle dose in the follow-up study is marginally significant and the average weight at the low dose is comparable to that of the control group. The agency rules that the low dose of the follow-up study is a NOEL; an ADI is then calculated by applying a safety factor, and the company is finally allowed to market its product.

Figure 2 shows that the follow-up study was unnecessary and only verified the dose-response trend of the initial study. If dose-response methods had been used for determining ADIs, the extra expense of the follow-up study and the 2-year delay could have been avoided.

Arbitrariness of Safety Factor

The NAS Safe Drinking Water Committee made the following recommendations for uncertainty factors (safety factors):

1. Valid experimental results from studies on prolonged ingestion by man with no indication of carcinogenicity.

Uncertainty Factor = 0.

2. Experimental results of studies of human ingestion not available or scanty (e.g., acute exposure only). Valid results of long-term feeding studies on experimental animals or in the absence of human studies, valid animal studies on one or more species. No indication of carcinogenicity.

Uncertainty Factor = 100.

3. No long-term or acute human data. Scanty results on experimental animals. No indication of carcinogenicity.

Uncertainty Factor = 1000.

These uncertainty factors are used in every case as a divisor of the highest reported long-term dose which is observed not to produce any adverse effect. (NAS, 1977)

The application of a 100-fold safety factor to results from long-term animal studies is a long-standing practice. It has been interpreted as resulting from the product of two 10-fold safety factors: one factor to account for animal-to-animal variation, and another to translate results from animal to man (Weil, 1972). However, the use of the 100-fold safety factor probably developed simply because some operational basis for setting allowable exposures was needed and a factor of 100 seemed "reasonable" or "prudent." The fact that humans have 10 fingers undoubtedly played a role in the specific factor selected.

When a safety factor is applied it is implicitly assumed that a threshold exists and the resulting ADI is below the threshold and hence safe. However, whether a threshold exists for a specific effect and, if so, whether the ADI is below that threshold are open to question. Also, as economic costs of regulations become more critical, there is increasing need for balancing the level of safety provided with

involve less random variation. However, the NOEL approach has the opposite tendency. A larger study has a better chance of showing a statistically significant result and thus will, on average, produce a smaller ADI. As an illustration, suppose at the control and one treated dose in a study involving 100 rats per dose, the resulting mean liver fat per animal was 15.1 g in the control group and 18.4 g in the treatment group with a standard deviation of 10.0 g in each group. Then the t statistic for a difference between the two groups is 3.3, which is significant at the 1% level. However, if the identical results came from a study involving only 25 rats per group, the t statistic is 1.14, which is not significant at the 10% level. Thus, a NOEL would possibly be estimated for the smaller study but not the larger.

Therefore, rather than encouraging larger studies to demonstrate greater evidence of safety, the NOEL-SF instead penalized proponents of chemicals for conducting large studies. This topsy-turvy state of affairs has made it necessary for regulatory agencies to set minimally acceptable sample sizes. Quite naturally, many studies use these minimal values.

Utilization of Dose Response:

A NOEL is determined solely by information relating to whether or not an effect was observed; the magnitude of positive effects and relationships among the effects at the various doses (i.e., the dose-response trend) is largely ignored. Consider for example the hypothetical data in Fig. 1. Experiment A shows a sharply increasing dose response. Experiment B shows a much flatter dose response, which is, in fact, consistent with a linear response through the origin. It appears, because of the sharp decrease in response with decreasing dose in Experiment A, as opposed to Experiment B, that the NOEL for A should be larger than the NOEL for B. However, because the response at dose d_{2A} was barely significant, the NOEL for Experiment A is d_{1A}

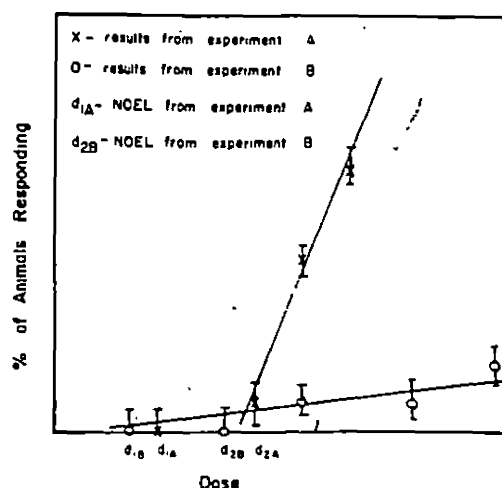


FIG. 1. Hypothetical responses with 95% confidence limits for two experiments.

which is less than the NOEL d_{2B} for Experiment B. The dose-response methods to be discussed in later sections are capable of fuller utilization of dose-response trends.

The NOEL-SF Approach Can Entail Unnecessary Restrictions and Expense

Consider the following scenario: A company wishing to market a new product implements a thorough toxicological testing program as required by the regulatory agency involved. Included in this program is a 2-year chronic toxicity and carcinogenesis bioassay, and a two-generation reproduction and teratology study. Each study involves three treatment and one control group with doses and sample sizes approved by the agency. The reproduction and teratology study is negative. In fact, the treated animals reproduce better than the control animals. This is apparently related to the fact that the control animals are obese. The 2-year study likewise shows no effects of treatment except for a dose-related weight reduction which is apparently due to the fact that the animals were fed the chemical in such high concentrations that their food was distasteful. As illustrated in Fig. 2, this weight loss follows

mathematical and statistical approaches to calculating ADIs for effects other than cancer. In the next section some potential shortcomings of the NOEL-SF approach are discussed. The following section describes some mathematical models and related statistical methods. These methods have two features that are somewhat novel. First, some of the models include the possibility of thresholds below which no effect will occur. Second, methods are suggested for application to "severity" or "continuous" data rather than just on incidence data. In the next section, a recommendation is made for replacing the NOEL in the NOEL-SF approach with a "benchmark dose." This benchmark represents a statistical lower confidence limit on the dose corresponding to a small increase in effect over the background level. The amount of increase in effect used to define the benchmark is small enough so that the estimate of the benchmark dose will reflect the shape of the dose-response curve, and it is large enough so that the lower confidence limit will not depend critically upon the mathematical model used in its calculation. A number of examples are presented illustrating the calculation of benchmark doses and comparing them with NOELs.

II. DIFFICULTIES WITH THE NOEL-SAFETY FACTOR METHOD

Definition of a NOEL

The first problem one faces with the NOEL concept is one of definition: Just what constitutes a NOEL? For effects which are unambiguous because they do not occur in untreated animals, such as acute toxicity or the occurrence of rare tumors, determination of a NOEL can be reasonably straight-forward; if no effect is seen in any animal a NOEL is determined—otherwise a NOEL is not determined. For less well-defined effects, such as necrosis or cloudy swelling of the liver, determination of a NOEL requires the use of judgment. This problem is compounded when

considering effects which have nonzero background levels. Consider, for example, liver weight; all animals have nonzero background levels of this "effect." Liver weights constitute a continuous measure (as opposed to "incidence" or "quantal" data) which can be obtained for each animal. It might happen that the average liver weight in some, or even all, of the treated groups is above that of the control group. Since this could happen by chance, usually the NOEL is taken to be the largest dose for which the increase in liver weight is not statistically significant. However, such a decision can seem rather arbitrary when there is a smooth dose-response trend which overlaps the region where the increase is not statistically significant.

A NOEL must be one of the experimental doses.² This constraint can appear unnecessarily restrictive in some cases. Consider, for example, an experiment to detect liver effects which involves three dose levels. Suppose at the highest dose level there are very severe effects, at the middle there are barely discernible effects, and at the low dose no effects at all are seen. Then the low dose likely will be designated the no-effect level even if the dose-response from the middle to high dose indicates that a much higher dose (one slightly less than the middle dose) would have had no discernible effect. Furthermore, if the data at the lowest dose had not been available, this experiment could not be used at all to define a no-effect level.

Effect of Sample Size

It would be appropriate for larger studies to tend to produce larger ADIs because they

² A NOEL is not an inherent property of the animal system but depends upon the experimental design and outcome. Thus it represents, in statistical terms, a statistic or an estimate of a "true no-effect level." This latter term refers to the highest dose which is absolutely safe and thus is an inherent property of the animal system, or, in statistical terms, a parameter. For an effect for which no threshold exists, the "true no-effect level" is zero.

Fundam.
ADI for
ty factor.
uidelines
r lesions
smaller
ents can
se range
proposes
BD is a
response
el. This
ve. The
method
idesteps
data to
ence or
effect is

ake level"
In this paper,
atter of conve-
n will apply to
tion of an ADI
no-effect level
-SF approach.
arcinogens have
fitting mathe-
l dose-response
to estimate the
specified small
A (1980) used
lity criteria for
L-SF method
omy was based
nogens are un-
nsequently, a
inappropriate
of a threshold,
present some

A New Method for Determining Allowable Daily Intakes¹

KENNY S. CRUMP

K. S. Crump and Company, Inc., 1201 Gaines Street, Ruston, Louisiana 71270

A New Method for Determining Allowable Daily Intakes. CRUMP, K. S. (1984). *Fundam. Appl. Toxicol.* 4, 854-871. The usual method for establishing allowable daily intake (ADI) for a chemical involves determining a no-observed-effect level (NOEL) and applying a safety factor. Even though this method has been used for many years, there appear to be no general guidelines or rules for defining a NOEL. The determination of a NOEL is particularly uncertain for lesions which occur naturally in untreated animals. NOELs also have shortcomings in that smaller experiments tend to give larger values (this should be reversed because larger experiments can provide greater evidence of safety) and that the steepness of the dose response in the dose range where effects occur plays little or no role in the determination of a NOEL. This paper proposes and illustrates the use of a "benchmark dose" (BD) as an alternative to a NOEL. A BD is a statistical lower confidence limit to a dose producing some predetermined increase in response rate such as 0.01 or 0.1. The BD is calculated using a mathematical dose-response model. This approach makes appropriate use of sample size and the shape of the dose-response curve. The BD normally will not depend strongly upon the mathematical model used because the method does not involve extrapolation far below the experimental range. Thus the method sidesteps much of the model dependency often associated with extrapolation of carcinogenicity data to low doses. The method can be applied to either "quantal" data in which only the presence or absence of an effect is recorded, or "continuous" data in which the severity of the effect is also noted. © 1984 Society of Toxicology.

I. INTRODUCTION

A common approach to quantifying permissible human exposure to a toxic agent is to establish a no-effect level using experimental animal data and then to apply a safety factor—or uncertainty factor, as it is sometimes called—to arrive at a permissible exposure level for humans. Allowable daily intakes for chemicals (ADIs), such as were employed by EPA in calculating water quality criteria (EPA, 1980), furnish one example of such calculations. Threshold limit values (TLVs), which are provided by the American Conference of Government and Industrial Hygienists (ACGIH) for many chemicals to which workers are exposed, are calculated in a similar fashion

(ACGIH, 1976). The term "daily intake level" (DIL) has also been employed. In this paper, we will refer to ADIs as a matter of convenience, although the discussion will apply to all such estimates. The calculation of an ADI by applying a safety factor to a no-effect level will be referred to as a NOEL-SF approach.

In recent years, ADIs for carcinogens have sometimes been calculated by fitting mathematical models to experimental dose-response data. These models are used to estimate the dose corresponding to some specified small amount of additional risk. EPA (1980) used this approach to set water quality criteria for carcinogens, and used a NOEL-SF method for noncarcinogens. This dichotomy was based upon the supposition that carcinogens are unlikely to have a threshold; consequently, a NOEL-SF approach would be inappropriate because it assumes the existence of a threshold.

The object of this paper is to present some

¹ Prepared for U.S. Environmental Protection Agency, Environmental Criteria Assessment Office, Contract 68-03-3111, Work Assignment 15.

Monsanto

FROM : NAME-LOCATION-PHONE: Dept. of Medicine & Environmental Health J. M. Kronenberg - G2WD, (4-8832)

DATE : February 11, 1985

cc F. R. Johannsen

SUBJECT : JOURNAL CLUB

REFERENCE :

TO : Toxicology Group

DJL	☐	REN	☒
DO	☐	RO	☐
DWF	☐	EDR	☐
DGH	☐	RDS	☐
FRJ	☐	IAWS	☐
JMK	☐	—	☐
TJL	☐	—	☐
DPM	☐	—	☐

This week's Journal Club meeting will be held on Thursday, February 14th at 12:30 p.m. I will be discussing the basic concepts (not the mathematics) proposed in the attached article on ADIs.

Joel

Joel

JMK/pp
Enclosure

003731