ON REFERENCE DOSE (RfD) AND ITS UNDERLYING TOXICITY DATA BASE

MICHAEL L. DOURSON, LINDA A. KNAUF, AND JEFFREY C. SWARTOUT

Environmental Criteria and Assessment Office
Office of Research and Development
Environmental Protection Agency
Cincinnati, Ohio

The toxicity data of pesticides were summarized and compared amongst different animal species and types of bioassays. These comparisons showed the expected inter-species and inter-bioassay variability. After quantitative and statistical analysis of these data, it was concluded that, on the average, a 2-year dog bioassay detected toxic responses at similar doses as a 2-year rat study, and that both of these bioassays detected toxic responses at lower doses than either a rat 2-generation bioassay, a rat developmental toxicity study, or a 2-year mouse bioassay. Although these chronic dog and rat bioassays were found to detect toxic responses at lower doses than the other studies listed, this analysis does not reflect the seriousness of the effects that were compared. Within the confines of this analysis, then, it appears that a 2-year dog and rat study, reproductive and developmental bioassays are a sufficient data base on which to estimate high confidence Reference Doses (RfDs), and furthermore, that an additional uncertainty factor is needed to estimate RfDs to account for this inter-species and inter-bioassay variability when fewer than this number of bioassays are available.

1. Address all correspondence to: Michael L. Dourson, Environmental Criteria and Assessment Office, Office of Research and Development, Environmental Protection Agency, 26 Martin Luther King Drive, Cincinnati, OH 45268. Tel: (513) 569-7533.

2. Abbreviations: ADI, acceptable daily intake; DG, dog; DV, developmental toxicity; LOAEL, lowest observed adverse effect level; MS, mouse; NOAEL, no observed adverse effect level; RE, reproductive; RfD, Reference Dose; RT, rat.

3. Key words: Reference Dose (RfD), Acceptable Daily Intake (ADI), uncertainty factor, safety factor, pesticides, inter-species comparison of toxic response.

4. Although the research (or other work) described in this article has been founded wholly or in part by the United States Environmental Protection Agency, it has not been subjected to the Agency's required peer and administrative review and, therefore, does not necessarily reflect the view of the Agency and no official endorsement should be inferred.

Toxicology and Industrial Health, 8:3, pp. 171–189
Copyright 1992 Princeton Scientific Publishing Co., Inc.
ISSN:0748-2337

INTRODUCTION

Questions regarding the appropriateness and sufficiency of toxicity data that will provide a basis for meaningful extrapolation from animal to man have remained despite much work on this issue (see for example Roloff, 1987). For noncancer extrapolation, one common method has been the derivation of an Acceptable Daily Intake (ADI). The use of the ADI was initiated by the World Health Organization in 1961 (Lu, 1988) based, in part, on previous work by Lehman and Fitzhugh (1954). The toxicological basis of the ADI has always depended on professional judgment. For example, Clegg (1978) stated that the typical minimum requirements for a data base permitting an estimation of the ADI are the following: two 90-day bioassays in different species, metabolic studies, reproduction studies of at least two generations in at least one species, teratology studies in at least one species different from that used in the reproductive studies, carcinogenicity studies, and special studies related to known toxic properties of related compounds. Vettorazzi (1984) discuss a similar list of studies but cautioned that the type of information required for the assessment of the toxicity of a chemical is a complex matter, and that the interpretation of the design and suitability of a study may be controversial.

The U.S. Environmental Protection Agency (EPA) has suggested the oral Reference Dose (RfD) or inhalation Reference Concentration[1] as a basis of extrapolation for noncancer toxicity (Barnes and Dourson, 1988; Jarabek et al., 1989; EPA, 1989a, 1991a). The RfD is similar in origin and scope to the ADI. Recent deliberations by the EPA on the toxicological basis of RfDs have focused on a somewhat reduced array of tests compared with Clegg (1978) or Vettorazzi (1984). For a high confidence RfD, the EPA (1989a) normally suggests at least the following bioassays: two adequate mammalian chronic toxicity studies in different species, one adequate mammalian 2-generation reproductive toxicity study, and two adequate mammalian developmental toxicity studies in different species. For a low-confidence RfD, the EPA suggests at least a single subchronic bioassay.

One scientific question in relationship to these various bioassays is how much toxicity data is needed in order to generate high confidence in the resulting value (ADI or RfD). It is known, for example, that any single toxicity study cannot adequately address all possible adverse outcomes and that the toxic endpoints vary among different species; these facts argue for more than one type of study and for testing more than one species. Furthermore, reproductive and developmental studies are unique in terms of the types of effects they are able to

[1]The EPA defines the oral RfD as "an estimate (with uncertainty spanning perhaps an order of magnitude) of a daily exposure to the human population (including sensitive subgroups) that is likely to be without appreciable risk of deleterious effects during a lifetime (EPA, 1991a)". An inhalation RfC is the equivalent to the oral RfD, although dosimetry adjustments are incorporated into the analysis.

detect and are, therefore, considered essential. A second scientific question is how, in the absence of such multiple studies, is the scientific uncertainty from the first question addressed in the resulting extrapolation. When toxicity data on only one experimental animal species are available for the estimation of an ADI, the U.S. Food and Drug Administration has used an extra two-fold uncertainty factor to account for the expected inter-species variability in toxic response (Shibko, 1981). The EPA is suggesting the use of a similar uncertainty factor for "incomplete" data bases in estimating RfDs. Are such factors scientifically supportable?

The purpose of this manuscript is to present and quantitatively analyze toxicity data on various endpoints and species in order to focus further discussions on the scientific aspects of both of these questions.

METHODS

Definitions used throughout this paper (see Appendix A) are consistent with the parlance of the EPA (1991b). These definitions are for illustration only; other terms are used in different organizations and countries.

The toxicity data for 69 pesticides were obtained from the EPA's Integrated Risk Information System (IRIS) (EPA, 1991b) (Table 1). The chemicals selected had the most complete toxicity data base possible in 1- to 2-year rat (RT), 1- to 2-year dog (DG) and 1- to 2-year mouse (MS) bioassays, and rat reproductive (RE) and rat developmental toxicity (DV) studies. A total of 296 studies were included in the analysis. Several 90-day studies were included in the data base because they were shown to detect toxic responses at lower doses than an existing 1- to 2-year study in the same species for the same chemical (as determined by a comparison of 90-day with 1- to 2-year no observed adverse effect levels [NOAELs]). In most of the studies included in the data base, both a NOAEL and a lowest observed adverse effect level (LOAEL) were determined. However, the LOAELs do not necessarily refer to the same effect across species.

This comparison of disparate effects is considered appropriate since the purpose of this paper is to compare the relevance of different species and endpoints in order to evaluate their usefulness in determining an RfD. Moreover, the determination of both the NOAEL and LOAEL is important since, in a well-conducted toxicity study, the NOAEL and LOAEL indentify the region of the experimental threshold. All studies were based on the oral route of exposure. Mg/kg/day doses were estimated for dietary and drinking water exposure through the use of standard default consumption assumptions for different species and/or strains.

The data were examined by using ratios of the NOAELs and were analyzed statistically in two sets. Three ratios were used in the first data set: 1) 1- to 2-

TABLE 1
Highest No-Observed-Adverse-Effect Levels (NOAELS) For 69 Pesticides (mg/kg/day)[a]

Chemical	CAS Number	Data Type[b]				
		RT	DG	RE	MS	DV
Acifluourfen	62476-59-9	25[c]	7.5[c]	1.25		20
Alachlor	15972-60-8	2.5		10		150
Amitraz	33089-61-1	2.5[c]	0.25[c]	0.75[c]		25
Andro	67485-29-4	2.5	0.33	2.5	2.75	5
Apollo	74115-25-5	2	1.25	20[d]	75	1000
Assure	76578-14-8		2.5	25	1.5	300
Asulam	3337-71-1	36	60[c]			1500
Baygon	114-26-1	10	5[c]	12.5[c]	105	10
Bayleton	43121-43-3	2.5	2.5	2.5	7.5	50
Baythroid	68359-37-5	2.5	4	2.5		30
Benomyl	17804-35-2	125[d]	12.5	5		30
Carbofuran	1563-66-2		0.5		18.75	
Carbosulfan	55285-14-8	1	12.5[d]			2
Carboxin	5234-68-4	10[c]	15[c,d]			375
Chloramben	133-90-4	50[c,d]	250[c,d]	225[c,d]		25
Chlorathalonil	1897-45-6	3[c,d]	1.5			100
Chlorpropham	101-21-3	100[c]	50[c]	150[d]	—	125
Chlorsulfuron	64902-72-3	5	62.5[d]	25	75	25
Cyhalothrin	68085-85-8	2.5	1	0.5	15	15
Cyromazine	66215-27-8	1.5	0.75	50		5
Daminozide	1596-84-5	150[c,d]	75[c,d]	15[c,d]	750	390
Danitol	39515-41-8	0.05	2.5	1.25	—	6
Dicamba	1918-00-0	125[d]	52[d]	25[d]		3
Dichlorvos	62-73-7	0.5[c]	0.08	25[c,d]		
Dimethipin	55290-64-7	2	7.5	10	12	160
Diphenamid	957-51-7	10[c]	3	10[c]		
Diphenylamine	127-39-4	3.1[c]	2.5[c]	125[c]		
Diquat	85-00-7	0.22	1.7[c]	25[d]		5
Diuron	330-54-1	1.25[c]	0.625	6.25[c,d]		
Dodine	2439-10-3	7[c]	1.25[c]			—
Endothall	145-73-3	125[c,d]	2[c]	5		10
EPTC	759-94-4	5	15	10		100
Fenamiphos	2224-92-6	0.5	0.25	1.5[d]		0.3
Fluridone	59756-60-4	8	75	32.5	15	125
Fluvalinate	69409-94-5	1	5			10
Folpet	133-07-3	10[c]	10	34.5		10
Fosetyl-al	39148-24-8	100	250	300	3000	1000
Glyphosate	1071-83-6	31[d]	20	10		1000
Imazalil	35554-44-0	10[c]	1.25	40[d]	—	10
Imazaquin	81335-37-7	500[d]	25	1000[d]	150	500
Iprodione	36734-19-7	50[d]	4.2	25	1870	60
Isoxaben	83558-50-7	5	10	125	14	320
Lactofen	77501-63-4	25	5	2.5		50
Londax	83055-99-6	35	20	355[d]	226	300
Metalaxyl	57837-19-1	12.5	6.25	62.5[d]		50

TABLE 1 Continued

Chemical	CAS Number	Data Type[b] RT	DG		MS	DV
Methidathion	950-37-8	0.2	0.1	0.25	1.6	2.25
Methomyl	16752-77-5	5	2.5	5[d]	7.5	20
Metolachlor	51218-45-2	15		15	150	360
Metribuzin	21087-64-9	5	2.5	15[c,d]	120	15
Naled	300-76-5	10[d]	0.2	6	15	40
Oryzalin	19044-88-3	15		12.5	75	25
Oxadiazon	19666-30-9	0.5	2.5	5[c]	—	—
Oxyfluorfen	42874-03-3	2	2.5	0.5		10
Paraquat	1910-42-5	25	0.45	7.5[d]	1.87	1
Permethrin	52645-53-1	5	5		3	400
Phosmet	732-4-6	2	10[d]	4[d]		60
Picloram	918-02-1	20	7	50[c]		1000
Propanil	709-98-8	5[c]	15[c]	50[d]		20
Propiconazole	60207-90-1	5	1.25	25	15	30
Savey	78587-05-0	23	2.5	35	37.5	720
Sethoxydim	74051-80-2	18[d]	2	162[d]	18	250
Systhane	88671-89-0	2.5	3.1	9.28	3	29
Terbacil	5902-51-2	12.5	1.25[c]	12.5[d]	7.5	12.5
Tetrachlorovinphos	961-11-5	6.25	3.13	16.5		150
Thiobencarb	28249-77-6		8	10[c,d]		25
Thiophanate-Methyl	23564-05-8	8	50	8	22.85	125
Tridiphane	58138-08-2	3	10	0.33[c]		30
Vinclozolin	50471-44-8	24.3[c]	2.5	72.9[c,d]	72.9	80

[a]Source: EPA, 1991b. NOAELs represented actual doses administered or calculated from dietary exposures. Doses have not been adjusted for either length of exposure or fraction of life span. Please note that data in IRIS are updated on a monthly basis; thus the information in this table may be out of date. Please check IRIS or call IRIS User Support (513/569-7572) for more information.

[b]RT = Chronic Rat, DG = Chronic Dog, RE = Reproductive Rat, MS = Mouse, DV = Developmental Toxicity

[c]Value is from a study that was rated as having poorer study quality. These were deleted to form Subset II (RT, DG and RE data only).

[d]Value is from a study that produced a NOAEL, but no LOAEL. These were deleted to form Subset I (RT, DG and RE data only).

year rat to 1- to 2-year dog (RT/DG), 2) reproductive rat to 1- to 2-year rat (RE/RT), and 2) 1- to 2-year dog to reproductive rat (DG/RE). Calculations made directly from the empirical data were used to determine the probabilities that the ratios were greater than 10^k or less than 10^{-k}, for $k = 0.5, 1.0$, and 1.5. Figure 1 shows the frequency histograms of the logs (base 10) of these first ratios. Note that the values on the horizontal axis correspond to the values of k.[2] For

[2]We are not assuming any distribution with these histograms. They are constructed to reflect the probabilities of interest in this analysis. The intervals, therefore, are as follows: $(X \leq -1.5)$, $(-1.5 < X \leq -1..0)$, $(-1.0 < X \leq -0.5)$, $(-0.5 < X \leq 0.0)$, $(0.0 \leq X < 0.5)$, $(0.5 \leq X < 1.0)$, $(1.0 \leq X < 1.5)$, $(1.5 \leq X)$. Values of $X = 0$ were evenly divided between the two intervals containing this number.

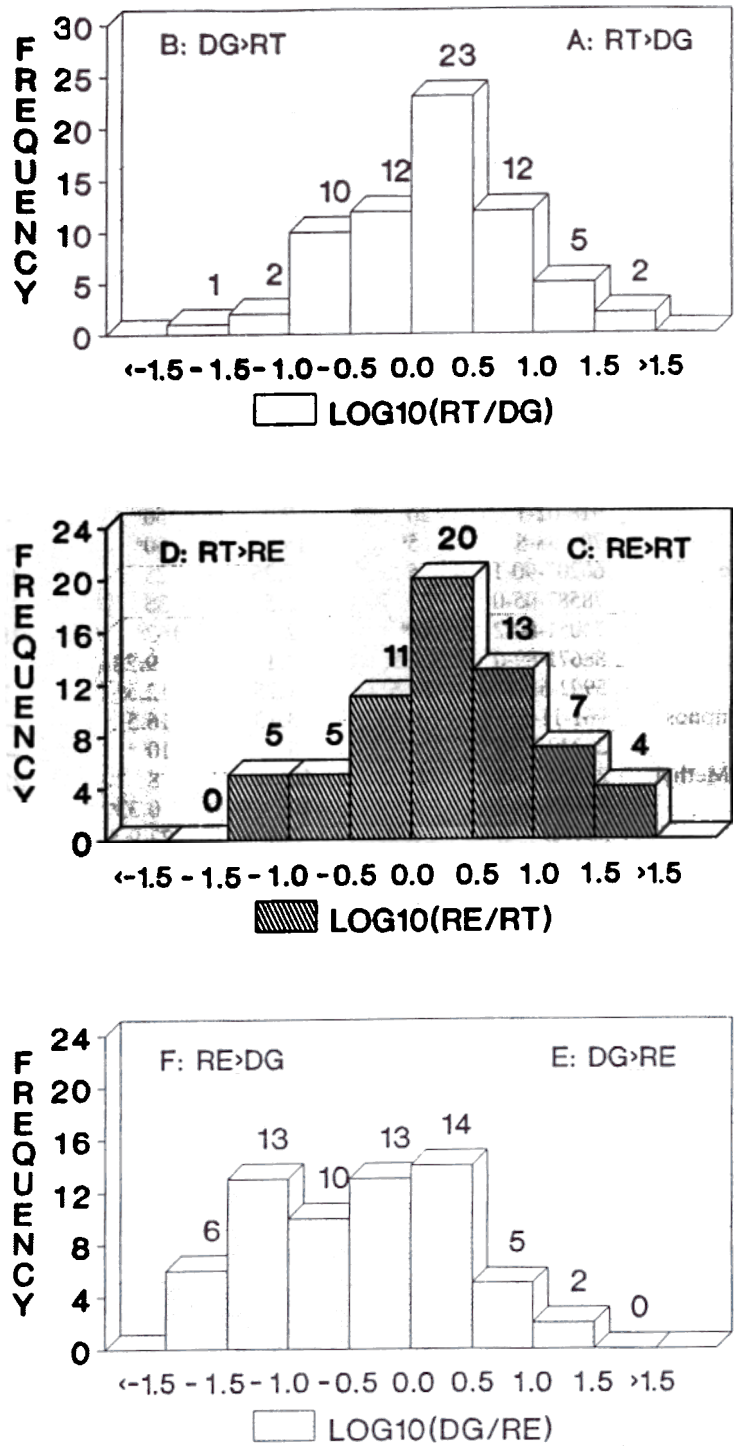

FIGURE 1. Frequency histograms are shown of ratios of NOAELs for chronic rat (RT) to chronic dog (DG), reproductive rat (RE) to chronic rat, and chronic dog (DG) to reproductive rat (RE). Letter designations A through F correspond to Table 2.

example. in the right tail of the top figure, the probability that rat is greater than dog by a factor of 10^1 is 7/67 (~0.10). This probability is calculated by taking the frequency rate of the data points to the right of 10^1. The probability that dog is greater than rat by 10^1 is shown to be 3/67 (~ 0.04) by the left tail of the top figure. This second probability is calculated by taking the frequency rate of the data points to the left of 10^{-1}.

These probabilities were used to estimate the percent of the time that one animal NOAEL is greater than another animal NOAEL by a given amount. These estimations are primarily of importance in determining the potential impact of missing studies on the RfD within the framework of the EPA's current procedure, at least for these pesticides. The analysis can also be used to estimate the relative sensitivity of endpoints. Specifically, for any predefined data base, the frequency with which a missing study or combination of studies might provide a lower NOAEL value than what is actually available can be determined. For example, if the predefined data base includes RT, DG, and RE, but only the RT and RE studies are present, one can estimate the probability that the missing DG study would provide a lower NOAEL value than the available data. In a sense, this provides an estimate of how often the most sensitive endpoint would be missed, given a less than "ideal" set of data.

The calculations described above were repeated for two subsets of the first set of data (69 chemicals, 201 studies). The first subset (SS1) consisted of only studies that produced both a NOAEL and LOAEL (65 chemicals, 157 studies). The second subset (SS2) consisted of only studies that exhibited a better measure of study quality (52 chemicals, 144 studies). These subsets were examined to determine if the probabilities would change due to the absence of a LOAEL or due to better study quality.

In addition, a second data set was formed by combining the first data set with chronic mouse and developmental toxicity data. This second set of data ratios was examined for the purpose of analyzing how much the probabilities are affected by the number of studies in the predefined data base or by the choice of studies in the predefined data base.

Finally, the nonparametric Wilcoxon Signed Rank Test was used to look for statistically significant differences between the NOAEL values for specific pairs of species.

RESULTS

Table 2 shows the probabilities that one animal NOAEL is greater than another animal NOAEL by a factor of 10^k, for all ratios. For example, in event *A*, 28% of the time, the rat NOAEL (RT) was larger than 3.16 ($10^{0.5}$) times the dog NOAEL (DG) for the same chemical, or in event *E*, 3% of the time, the dog

TABLE 2
Probabilities for All Ratios
Event: $NOAEL_1 \geq NOAEL_2 * 10^k$

Event Label	$NOAEL_1$	$NOAEL_2$	Probability of Event		
			$k = 0.5$	$k = 1.0$	$k = 1.5$
A			0.28	0.10	0.03
B			0.19	0.04	0.01
C			0.37	0.17	0.06
D			0.15	0.08	0.00
E			0.11	0.03	0.00
F			0.46	0.30	0.10
G			0.03	0.00	0.00
H			0.67	0.50	0.17
I			0.03	0.00	0.00
J			0.44	0.25	0.06
K			0.13	0.00	0.00
L			0.35	0.19	0.06
M			0.05	0.03	0.00
N			0.66	0.44	0.23
O			0.05	0.03	0.02
P			0.65	0.48	0.19
Q			0.13	0.02	0.00
R			0.50	0.37	0.10
S			0.16	0.09	0.00
T			0.38	0.22	0.06

RT = Chronic Rat. DG = Chronic Dog. RE Reproductive Rat. MS = Mouse. DV = Developmental Toxicity

NOAEL (DG) was larger than 10 times ($k = 1$) the reproductive rat NOAEL (RE) for the same chemical.

Notice the events *A* (RT > DG) and *B* (DG > RT) are mutually exclusive, as are *C* (RE > RT) and D (RT > RE), and also *E* (DG > RE) and *F* (RE > DG). Thus, only certain combinations of events *A* through *F* can occur. For example, events *A*, *C*, and *F* can all occur simultaneously. These three events might combine to form the mathematically true statement. RE > RT > DG. In contrast, events *A*, *C*, and *E* cannot occur simultaneously.

The probabilities of Table 2 can be combined to estimate the probability of NOAEL overestimation. Table 3 presents these results for the first set of ratios. For example, suppose only the rat and reproductive rat studies are available. How likely is it that the missing dog NOAEL would have been smaller than the chronic rat or reproductive rat NOAELs? The probabilities that apply from

TABLE 3
Combined Probabilities for the First Set of Ratios

Study Present	Events Occurring*	Probability Formula	Probability of Events Occurring by a Factor of 10^k		
			$k = 0.5$	$k = 1.0$	$k = 1.5$
RT	$A \cup D$	$P(A)+P(D)-P(A\|D)P(D)$	0.37	0.16	0.03
DG	$B \cup E$	$P(B)+P(E)-P(B\|E)P(E)$	0.24	0.06	0.01
RE	$C \cup F$	$P(C)+P(F)-P(C\|F)P(F)$	0.62	0.39	0.13
RT, DG	$D \cap E$	$P(D\|E)P(E)$	0.06	0.00	0.00
RT, RE	$A \cap F$	$P(A\|F)P(F)$	0.18	0.04	0.00
DG, RE	$B \cap C$	$P(B\|C)P(C)$	0.12	0.02	0.00
RT, DG, RE	None		0.00	0.00	0.00

*Events which must occur if the missing studies are to undercut the lowest NOAEL available ∩ = intersection
∪ = Union

Table 2 are the probabilities of events *A*, (RT > DG), and *F*, (RE > DG); the following scenarios could occur:

Event *A* (RT > DG)	Event *F* (RE > DG)	Order of NOAELs	Smallest NOAEL	Data Available
True	True	RT > DG; RE > DG	DG	RT, RE
True	False	RT > DG ≥ RE	RE	RT, RE
False	True	RE > DG ≥ RT	RT	RT, RE
False	False	DG ≥ RT; DG ≥ RE	Min(RT, RE)	RT, RE

Thus, both events *A* and *F* must occur simultaneously in order for the missing dog NOAEL to be the smallest. This probability is found by taking the intersection of events *A* and *F*. In this case, Table 3 indicates that an 18% chance exists that the missing dog NOAEL is three-fold (i.e., $K = 0.5$) or more lower than both the rat and the reproductive rat NOAELs; a 4% chance exists that it is 10-fold (i.e., $K = 1.0$) or more lower; and a 0% chance exists that it is 30-fold (i.e., $K = 1.5$) or more lower.

Suppose only the rat study is present (Table 3). How likely is it that the missing dog and/or reproductive rat NOAELs would have been less than the rat NOAEL? The probabilities that apply from Table 2 are the probabilities of events *A* (RT > DG), and *D* (RT > RE); the following scenarios could occur:

Event *A* (RT > DG)	Event *D* (RT > RE)	Order of NOAELs	Smallest NOAEL	Data Available
True	True	RT > DG; RT > RE	Min(DG, RE	RT
True	False	RE ≥ RT > DG	DG	RT
False	True	DG ≥ RT > RE	RE	RT
False	False	DG ≥ RT; RE ≥ RT	RT	RT

Thus, if events *A* and *D* occur simultaneously, or if either event occurs by itself, then the rat NOAEL is not the smallest of the three NOAELs. This combined probability is found by taking the union of the two events *A* and *D*. In this case, Table 3 indicates that a 37% chance exists that either the missing dog NOAEL and/or the missing reproductive rat NOAEL is three-fold or more lower than the rat NOAEL; a 16% chance exists that either one or both is 10-fold or more lower; and a 3% chance exists that either one or both is 30-fold lower.

Tables 2 and 3 show the results of the calculations for the first set of ratios, listing the probabilities when only certain studies are present out of the predefined three study data base. It can be seen that the reproductive rat study is not as sensitive as the chronic rat and dog studies. For example, when the reproductive rat study alone is present, a 39% chance exists that the two missing studies would have provided at least a 10-fold lower NOAEL as compared with 16% and 6% chances for the rat study only and dog study only, respectively (Table 3).

Also, any combination of two of the three studies is superior to having only one study available; the best two-study combination is the chronic rat and dog combination. From this data base, there is a zero percent chance that a 10-fold lower NOAEL will be provided by the missing study with a dog and rat combination, a 4% chance with a rat and reproductive rat combination, and a 2% chance with a dog and reproductive rat combination (Table 3). Finally, the probabilities for two-study combinations are highest when estimating the NOAEL within half an order of magnitude (i.e., $K = 0.5$). Table 3 shows a 6% chance of a three-fold lower NOAEL being provided by the missing study with a dog and rat combination, and 18% chance with a rat and reproductive rat combination, and a 12% chance with a dog and reproductive rat combination.

Tables 4 and 5 show the same analysis as Tables 2 and 3, respectively, for the first set of ratios and for the two subsets of these ratios as previously described in "Methods." Few differences exist among the calculated probabilities for each combination of events occurring in each of the subsets when compared with each other or to the primary ratios. Tables 4 and 5 also indicate that having only a reproductive rat study is the worst scenario for the purposes of calculating an RfD, and that having the chronic rat and dog studies available is the best two-study combination. These relationships remain constant across the subsets.

Chronic mouse and developmental toxicity data are also listed in Table 1, and the probabilities for events *G* through *T* that relate to these data and their combinations are given in Table 2. It may be noted from Table 1 that the value of these mouse and developmental toxicity NOAELs are generally larger compared with those of DG, RT, and RE. For example, in Table 2 the mouse NOAEL is larger than 10 times the dog NOAEL 50% of the time (event *H*),

TABLE 4
Probabilities Across Subsets
Event: $NOAEL_1 \geq NOAEL_2 * 10^k$

Event Label	$NOAEL_1$	$NOAEL_2$	Probability of Event								
			$k = 0.5$			$k = 1.0$			$k = 1.5$		
			DB	SS1	SS2	DB	SS1	SS2	DB	SS1	SS2
A			0.28	0.25	0.22	0.10	0.04	0.09	0.03	0.00	0.02
B			0.19	0.17	0.24	0.04	0.02	0.07	0.01	0.02	0.02
C			0.37	0.32	0.34	0.17	0.13	0.14	0.06	0.05	0.05
D			0.15	0.16	0.11	0.08	0.05	0.05	0.00	0.00	0.00
E			0.11	0.13	0.10	0.03	0.03	0.02	0.00	0.00	0.00
F			0.46	0.37	0.45	0.30	0.18	0.31	0.10	0.05	0.10

DB = Complete data base; SS1 = studies with no LOAEL removed; SS2 = studies with poorer study quality removed

and the developmental toxicity NOAEL is larger than 10 times the reproductive rat NOAEL 37% of the time (event *R*).

Table 6 shows that as the number of study types required in the predefined data base increases, the probabilities relating to overestimating the NOAELs will either increase or remain the same, but they will not decrease. This happens because a new study in the data base presents a new opportunity to undercut the "lowest" NOAEL. For example, in line 3 of Table 6, a three-study data base is defined as chronic rat, dog, and reproductive rat studies. If dog is the only study available, then the union of events *B* (DG > RT) and *E* (DG > RE)

TABLE 5
Combined Probabilities Across Subsets

Study Present	Events Occurring*	Probability of Events Occurring								
		$k = 0.5$			$k = 1.0$			$k = 1.5$		
		DB	SS1	SS2	DB	SS1	SS2	DB	SS1	SS2
RT	$A \cup D$	0.37	0.37	0.32	0.16	0.09	0.13	0.03	0.00	0.02
DG	$B \cup E$	0.24	0.23	0.26	0.06	0.05	0.07	0.01	0.02	0.02
RE	$C \cup F$	0.62	0.63	0.69	0.39	0.30	0.41	0.13	0.10	0.13
RT, DG	$D \cap E$	0.06	0.08	0.02	0.00	0.00	0.00	0.00	0.00	0.00
RT, RE	$A \cap F$	0.18	0.08	0.08	0.04	0.01	0.02	0.00	0.00	0.00
DG, RE	$B \cap C$	0.12	0.13	0.11	0.02	0.03	0.05	0.00	0.00	0.02
RT, DG, RE	None	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

DB = Complete data base; SS1 = studies with no LOAEL removed; SS2 = studies with poorer study quality removed

*Events which must occur if the missing studies are to undercut the lowest NOAEL available. $\cap$ = intersection; $\cup$ = union

TABLE 6
Combined Probabilities for Various Predefined Data Base Combinations

Line	Study Present	Studies in the Predefined Data Base	Events Occurring	Probability of Events Occurring		
				k = 0.5	k = 1.0	k = 2.5
	DG	DG	None	0.00	0.00	0.00
2	DG	DG, RT	B	0.19	0.04	0.01
3	DG	DG, RT, RE	$B \cup E$	0.24	0.06	0.01
4	DG	DG, RT, RE, MS	$B \cup E \cup G$	0.24	0.06	0.01
5	DG	DG, RT, RE, DV	$B \cup E \cup M$	0.27	0.09	0.01
6	DG	DG, RT, RE, MS, DV	$B \cup E \cup G \cup M$	0.27	0.09	0.01
7	RT	RT	None	0.00	0.00	0.00
8	RT	RT, DG	A	0.28	0.10	0.03
9	RT	RT, DG, RE	$A \cup D$	0.37	0.16	0.03
10	RT	RT, DG, RE, DV	$A \cup D \cup O$	0.39	0.18	0.05
11	RT	RT, DG, RE, MS	$A \cup D \cup I$	0.40	0.16	0.03
12	RT	RT, DG, RE, MS, DV	$A \cup D \cup I \cup Q$	0.45	0.20	0.05
13	DG, RT	DG, RT	None	0.00	0.00	0.00
14	DG, RT	DG, RT, MS	$I \cap G$	0.00	0.00	0.00
15	DG, RT	DG, RT, DV	$O \cap M$	0.02	0.02	0.00
16	DG, RT	DG, RT, RE	$D \cap E$	0.06	0.00	0.00
17	DG, RT	DG, RT, RE, MS	$(D \cap E) \cup (I \cap G)$	0.06	0.00	0.00
18	DG, RT	DG, RT, RE, DV	$(D \cap E) \cup (O \cap M)$	0.08	0.02	0.00
19	RT, RE	RE, RE	None	0.00	0.00	0.00
20	RT, RE	RT, RE, DV	$O \cap Q$	0.02	0.00	0.00
21	RT, RE	RT, RE, MS	$I \cap K$	0.03	0.00	0.00
22	RT, RE	RT, RE, DG	$A \cap F$	0.18	0.04	0.00
23	RT, RE	RT, RE, DG, MS	$(A \cap F) \cup (I \cap K)$	0.18	0.04	0.00
	RT, RE	RT, RE, DG, DV	$(A \cap F) \cup (O \cap Q)$	0.20	0.04	0.00
25	RT, MS	RT, MS	None	0.00	0.00	0.00
26	RT, MS	RT, MS, RE	$D \cap L$	0.10	0.04	0.01
27	RT, MS	RT, MS, DG	$A \cap H$	0.21	0.05	0.01
28	RT, MS	RT, MS, RE, DG	$(A \cap H) \cup (D \cap L)$	0.25	0.07	0.01

$\cup$ = Union, $\cap$ = intersection

yields the probability of undercutting the dog NOAEL; for a K value of 1.0, this probability is 0.06. In line 5 of Table 6, the predefined data base is comprised of the four study types: rat, dog, reproductive rat, and developmental. Again, if dog is the only study present, the probability that the missing developmental NOAEL (i.e., the new study) would be less than the dog NOAEL, event M from Table 2 (DG > DV), must be added in with events B and E to figure the probability of undercutting the dog NOAEL; for a K value of 1.0, this combined probability is 0.09 as compared to 0.06 with a three study data base.

Table 6 also reveals that the choice of study type is important to the estimation of the smallest possible NOAEL. Lines 26 and 27 of Table 5 are examples of two different predefined three-study data bases with rat and mouse studies present for each case. In line 27, the rat and mouse studies are available from the three-study data base of rat, mouse, and dog. The probability that the missing dog study would yield at least a three-fold (i.e., $K = 0.5$) lower NOAEL is 0.21. However, if the rat and mouse studies are available from the three study data base of rat, mouse, and reproductive rat (line 26), then the probability that the missing reproductive rat study would yield at least a three-fold lower NOAEL is only 0.10.

Hypothesis Testing

Hypothsis tests were performed using the Wilcoxon Signed Rank Test to find and estimate differences in NOAEL values between pairs of species. The assumed model was:

$$\log10(\text{NOAEL}_1) - \log10(\text{NOAEL}_2) = C,$$

where C is some constant value. The corresponding hypothesis test was:

$$H_0 : C = 0 \text{ vs. } H_a : C \neq 0.$$

Furthermore, by looking at differences in the logs of the NOAELs and estimating C (Hollander and Wolfe, 1973), we can find the relationship of the ratios of the NOAELs, with 95% confidence intervals from:

$$\text{NOAEL}_1/\text{NOAEL}_2 = 10^C.$$

We looked at seven relationships with an overall $p = 0.05$; therefore, signficance level for each test was $p = 0.00357$. Results were as follows:

Ratio	*N*	10^C	95% C.I.	*P*	Reject Ho:
RT/DG	46	1.58	(0.99, 2.24)	0.033	No
RE/RT	52	2.45	(1.29, 3.87)	0.00317	Yes
RE/DG	62	2.68	(1.63, 4.47)	0.00009	Yes
MS/RT	31	3.87	(2.24, 6.32)	0.00001	Yes
MS/DG	30	7.07	(3.53, 13.42)	0.00001	Yes
DV/RT	57	7.91	(4.89, 12.91)	1.5E-8	Yes
DV/DG	60	8.66	(5.50, 14.61)	1.4E-9	Yes

From these tests, it was concluded that:

$$\text{DG} = \text{RT} < \begin{matrix}\text{RE}\\ \text{MS}\\ \text{DV}\end{matrix}$$

DISCUSSION

From a practical point of view, the development of credible dose-response assessments such as an RfD or an ADI should allow for scientific distinction among possible minimum toxicity data bases in the number of studies as well as the necessary study types. This issue has plagued previous efforts because one could always object that the most sensitive species was not tested or that the most sensitive effect was not monitored. The data presented in this paper allow for some discussion and perhaps limited resolution of this issue. Minimum data requirements and study sensitivity, and dose-response assessment will be discussed in turn.

Minimum Data Requirements and Study Sensitivity

A primary objective of this research was to define a data set that would provide reasonable assurance that a sensitive effect was not missed in the determination of an RfD (or ADI). This could be viewed as an impossible task because no matter how many data are in hand, that many more are lacking. However, from a practical standpoint, only a few well-defined standard animal bioassays are regularly available. Therefore, we attempted to determine which bioassays are essential.

Quantitatively speaking, a four study base should be adequate within the scope of the data analyzed here. Good chronic dog and rat studies, and reproductive and developmental studies together, cover nearly all of the uncertainty in this set of pesticide data. The chronic mouse data add little, if any, quantitative impact on the probabilities listed in Table 6 (e.g., compare lines 3 & 4, 5 & 6, 9 & 11, 13 & 14, 16 & 17, 19 & 21, 22 & 23). Several bioassays have a large impact on the probabilities listed in Table 6, for example, the chronic dog study (e.g., compare lines 7 & 8, 19 & 22, 25 & 27). Other types of studies have consistent and yet smaller quantitative impact in Table 6, for example, developmental toxicity study (e.g., compare lines 3 & 5, 4 & 6, 9 & 10, 11 & 12, 13 & 15, 16 & 18, 22 & 24).

The data also allow for some conclusions regarding the relative sensitivity of the studies examined. In the context of this discussion, "sensitivity" refers to the ability of a given bioassay to detect any adverse effect, it does not necessarily refer to the innate ability of a particular substance to cause damage in any given species or organ system, nor to the seriousness of the different toxicity evoked among studies. Thus, the distinction between true physiological sensitivity and relative measurement sensitivity cannot be made here.

The data show a general ordering of increasing sensitivity of DV, MS, RE, RT, DG. The DG:RT difference would almost certainly be magnified (with DG being more sensitive) if the DG study design were similar to the RT with respect to the number of animals and relative duration. RT studies generally use about 5-

10 times as many animals per dose group as DG studies, and study duration is generally a much greater proportion of life span for RT than DG (about six-fold for 2-year studies). (Note that NOAELs used in these comparisons were not adjusted for life span differences; see Table 1.)

The trend in sensitivity could be ascribed to differences in genetics, or the phenomenon may be an artifact of animal size. The latter possibility could occur in perhaps one of two ways. The toxicity of these substances, in general, may be proportional to a fractional power of body weight, as are a number of physiological processes, rather than to body weight. An apparent negative correlation of body-weight-based toxicity as measured by NOAEL and animal size would then result when the toxicities were actually equivalent. (That is, animals of larger body weight would tend to have smaller NOAELs. See Dourson and Stara, 1983, for further discussion.) A second artifact of animal size arises from the technical aspects of animal experimentation. Quite simply, the experimenter can investigate more endpoints and obtain more samples, more easily, from larger animals, resulting in a greater likelihood of detecting treatment-related effects.

The lower sensitivities exhibited by the RE and DV studies, as evidenced by the RT:RE and RT:DV differences observed, may not be artifacts of body size because RE and DV are primarily rat studies. The RE and DV endpoints, as measured, are less sensitive than those measured in chronic general-toxicity studies. This may be due in part to the older nature of the RE and DV studies used here when compared to the relatively recent guidelines for analyzing these data for risk assessment pruposes (EPA, 1986, 1989b).

Dose-Response Assessment

The impact of these results on current dose-response assessments, such as the EPA approach for deriving RfDs, depends in part on how one interprets the nature of uncertainty factors. The most common interpretation is essentially empirical. Under this interpretation uncertainty factors are viewed as necessary loose upper-bound values because of missing data. They are not average values, as discussed by Lewis et al. (1990).

Within this interpretation, one attempts to establish a human reference dose sufficiently below the expected population threshold dose for any adverse effect with a given set of data. It follows that, having identified a set of necessary data that have a significant quantitative impact on the uncertainty of the health hazard assessment, one must divide by an uncertainty factor to estimate the value of the missing data set.

Should this adjustment always be in one direction (that is, downward)? The uncertainty factors currently used by the EPA (Barnes and Dourson, 1988) to dervie RfDs are: UF_H (sensitive humans), UF_A (animal to human, UF_S (sub-

chronic to chronic), and UF_L (LOAEL to NOAEL). The unidirectional nature of UF_L and UF_H is obvious; zero expectation exists that an increase in dose rate will provide a NOAEL in the first case and protection for sensitive subpopulations in the second case. The case for UF_S is less unambiguous but conceptually sound. In general, a much greater expectation exists that continued exposure at a given dose rate will produce rather than alleviate effects. A number of mechanisms can be conceived by which effects, not evident after subchronic exposure, may arise following chronic exposure. Repeated insult may result in the accumulation of unrepaired damage not measurable in the short-term, or in the gradual compromise of homeostatic mechanisms, for example.

The argument for the directionality of UF_A is less specific, and perhaps not as compelling as that for the other uncertainty factors. One could argue that the assumption that humans, in general, are inherently more sensitive than the most sensitive laboratory animal species is unreasonable. This argument is particularly cogent when a surfeit of data exist for many species, yet the magnitude of UF_A does not change. It has been suggested previously (Dourson and Stara, 1983) that UF_A, in part, is covering the possibility that the wrong scaling factor for interspecies dose conversion was selected (power of body weight equal to 1 instead of less than 1). However, as UF_A is always constant (10), this reasoning is most applicable when the species is unknown (which of course is never the case).

Another possible interpretation of UF_A is that it accounts for all aspects of the relative insensitivity of laboratory animal experimentation, with respect to real-world human exposure, not already accounted for by other uncertainty factors. This UF, then, would allow for all the measurement error, lack of statistical power, design limitations, and other aspects of uncertainty inherent in the conduct of toxicological studies (such as the inability to measure subjective human endpoints). We would have a higher expectation that this combination of uncertainties would result in a higher value for an experimentally-determined animal threshold than an actual human threshold.

Within this interpretation of uncertainty factors, the results presented in this paper suggest two general conclusions. The first conclusion is that the choice in both the number of species and types of studies in a dose-response assessment can affect the outcome of the specific probabilities of missing relevant data, and more generally, one needs more than a single bioassay to ensure a relatively high confidence in any given RfD or ADI. This conclusion confirms what has been the general thinking of toxicologists for years (for example, Clegg, 1978). This analysis merely expresses the concept in a quantitative fashion. However, a second conclusion is that one may not have to consider very many studies, given that we have looked at all the species and endpoints generally available for this class of compounds, and the data summarized in Table 6 show that the probabilities tend to approach a plateau as the number of studies in the study

base increases. One implication of these two conclusions is that when one or more studies is missing from a predefined "complete" data set, uncertainty is encountered that must be expressed. Whether this uncertainty should be expressed quantitatively (as in the use of an additional uncertainty factor) or qualitatively remains a subject of further research and debates, but in the interim the EPA has used such a factor to account for missing studies (EPA, 1991b).

The limitations of this analysis are several, and some cannot be addressed easily. For example, more meaningful estimates of study sensitivity might be obtained by focusing more on the effect levels than on the no-effect levels. Also, an ideal analysis would include silmilar experimental protocols for the same bioassay for different chemicals—we obtain data for different chemicals from a diverse group of laboratories and throughout a number of years. In addition, the increasing sensitivity from MS to RT to perhaps DG should provoke thought as to the likely biological basis.

Because the data set used in this analysis is not from a random sample of all chemicals, but a collection of readily-available, relatively complete data sets derived from oral exposure (all of which happen to be pesticides), one may not be able to conclude much about other, even larger, categories (solvents, for example) or other routes of exposure. However, in a world of many chemicals and limited toxicity data, generalizations are needed that can be applied to many substances of environmental concern. Research based on sound toxicological principles is continuing to focus towards this end.

ACKNOWLEDGMENT

The authors acknowledge the comments and guidance of Drs. Ellen J. O'Flaherty, George Ghali, and Reto Engler in the preparation of this manuscript.

REFERENCES

BARNES, D.G. and DOURSON, M.L. (1988). Reference Dose (RfD): Description and use in health risk assessment. Reg. Toxicol. Pharmacol. **8**:471–486.

CLEGG, D.J. (1978). Toxicological basis of the ADI—Present and future considerations. In: Pesticide Reviews (H. Frehse and H. Geissbuhler, eds.) Pergamon Press, Oxford, England.

DOURSON, M.L. and STARA, J.F. (1983). Regulatory history and experimental support of uncertainty (safety) factors. Reg. Toxicol. Pharamcol. **3**:224–238.

HEYWOOD, R. (1981). Target organ toxicity. Toxicol. Lett. **8**:349–358.

HEYWOOD, R. (1983). Target organ toxicity II. Toxicol. Lett. **18**:83–88.

HOLLANDER, M. and WOLFE, D.A. (1973). Nonparametric Statistical Methods, John Wiley and Sons, Inc., New York.

JARABEK, A.M., MENACHE, M.G., OVERTON, J.H., DOURSON, M.L., and MILLER, F.J. (1989). Inhalation Reference Dose (RfD): An application of inter-

pecies dosimetry modeling for risk assessment of insoluble particles. Health Physics. **57**:177–183.
LEHMAN, A.J. and FITZHUGH, O.G. (1954). 100-Fold margin of safety. Assoc. Food Drug. Off. U.S.Q. Bull. **18**:33–35.
LEWIS, S.C., LYNCH, J.R., and NIKIFOROV, A.I. (1990). A new approach to deriving community exposure guidelines from no-observed-adverse-effect levels. Reg. Toxicol. Pharmacol. **11**:314–330.
LU, F. (1988). Acceptable Daily Intake: Inception, evolution and application. Reg. Toxicol. Pharmacol. **8**(1):45–60.
ROLOFF, M.V., (ed). (1987). Human Risk Assessment: The Role of Animal Selection and Extrapolation. Taylor and Francis, New York, NY.
SHIBKO, S. (1981). Memorandum to M.L. Dourson, U.S. EPA, Cincinnati, Ohio, December 24. Food and Drug Administation, Washington, DC.
U.S. ENVIRONMENTAL PROTECTION AGENCY (EPA). (1986). Guidelines for the Health Assessment of Suspect Developmental Toxicants. Federal Register. **51**:34028–34040. September 24.
U.S. ENVIRONMENTAL PROTECTION AGENCY (EPA). (1989a). Interim Methods for Development of Inhalation Reference Doses. Office of Health and Evironmental Assessment, Washington DC. EPA/600/8–88/066F.
U.S. ENVIRONMENTAL PROTECTION AGENCY (EPA). (1989b). Proposed Amendments to the Guidelines for the Health Assessment of Suspect Developmental Toxicants: Request for comments. Notice Federal Register. **54**:9386–9403. March 6.
U.S. ENVIRONMENTAL PROTECTION AGENCY (EPA). (1991a). Reference Dose: Description and Use in Health Risk Assessments. Integrated Risk Information System (IRIS): Background Document. Online. Intra-Agency Reference Dose (RFD) Work Group, Office of Health and Environmental Assessment, Envrionmental Criteria and Assessment Office, Cincinnati, OH.
U.S. ENVIRONMENTAL PROTECTION AGENCY (EPA). (1991b). The Integrated Risk Information System (IRIS). Online. Office of Health and Environmental Assessment, Environmental Criteria and Assessment Office, Cincinnati, OH.
VETTORAZZI, G. (1984). Pesticide residues in food in the context of the international activities of the Joint FAO/WHO Meeting on Pesticide Residues and the Codex Committee on Pesticide Residues. Environmental Health Criteria 104. Principles for the Toxicological Assessment of Pesticide Residues in Food. International Programme on Chemical Safety, World Health Organization, Geneva, Switzerland.

APPENDIX A
DEFINITIONS

Critical effect—The first adverse effect, or its known precursor, that occurs as the dose rate increases.

Lowest-observed-adverse-effect level (LOAEL)—The lowest exposure level at which there are statistically or biologically significant increases in frequency or severity of adverse effects between the exposed population and its appropriate control group.

No-observed-adverse-effect level (NOAEL)—An exposure level at which there are no statistically or biologically significant increases in the frequency or severity of adverse effects between the exposed population and its appropriate control; some effects may be produced at this level, but they are considered neither adverse nor precursors to specific adverse effects. In an experiment with several NOAELs, the regulatory focus is primarily on the highest one, leading to the common usage of the term NOAEL as the highest exposure without adverse effect.

Reference dose (RfD)—An estimate (with uncertainty spanning perhaps an order of magnitude) of a daily exposure to the human population (including sensitive subgroups) that is likely to be without appreciable risk of deleterious effects during a lifetime.

Uncertainty factor (UF)—One of several, generally 10-fold factors, used in operationally deriving the dose (RfD) from experimental data. UFs are intended to account for (1) the variation in sensitivity among the members of the human population; (2) the uncertainty in exrapolating animal data to the case of humans; (3) the uncertainty in extrapolating from data obtained in a study that is of less-than-lifetime exposure; (4) the uncertainty in using LOAEL data rather than NOAEL data; and (5) the inability of any single study to adequately address all possible adverse outcomes in man.