Methods for Establishing Oral Reference Doses*

Michael L. Dourson
U.S. Environmental Protection Agency

Empirical observation reveals that as the dose of a chemical is increased, the toxic response generally increases. This increase occurs in both the severity of the response (e.g., with new and more severe effects being observed) and the intensity of the response (i.e., the magnification of an effect of given severity) and in the percentage of the population affected. Such dose-response relationships are well founded in the theory and practice of toxicology and pharmacology.

Dose-response assessment follows hazard identification in the risk assessment process as defined by the National Academy of Sciences (NAS) (1983). Dose-response assessment involves the quantitative evaluation of toxicity data to determine the likely incidence of the associated effects in humans. The information available for dose-response assessment ranges from well-conducted and controlled studies on human exposures, epidemiology studies with large numbers of subjects and well-characterized exposures, and supportive studies in several animal species, to a lack of human and animal toxicity data with only structure-activity relationships to guide the evaluation. In any case, scientists should consider all pertinent studies in this process; a single human case study can provide useful information. However, only data of sufficient quality as judged by experts should be used in the dose-response assessment of a chemical.

Given at least a moderate amount of toxicity data, one goal has been to determine a level of daily exposure that is likely to be without an appreciable risk of deleterious effects during a lifetime. The World Health Organization and many other health agencies have used the "acceptable daily intake" (ADI) concept in this regard. The U.S. Environmental Protection Agency (EPA) has also used the ADI in many of its evaluations. However, EPA has also rethought this approach and developed new, clarifying terminology in keeping with the NAS (1983) risk assessment/risk management paradigm, which calls for a clear demarcation between scientific and nonscientific factors in regulatory

*The views expressed in this paper are those of the author and do not necessarily reflect the views and policies of the U.S. Environmental Protection Agency. Parts of this text are excerpted from Barnes and Dourson (1988).

(RfD)

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 not considered adverse or precursors to specific adverse effects.* In an experiment with several NOAELs, the regulatory focus is primarily on the NOAEL seen at the highest dose. This leads to the common use of the term NOAEL to mean the highest exposure without adverse effect.
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.
Uncertainty factor (UF)	One of several, generally 10-fold, factors used in operationally deriving the reference dose (RfD) from experimental data.
Modifying factor (MF)	A UF that is greater than zero and less than or equal to 10; the default value for the MF is 1.

decisions. In EPA's experience, the older ADI approach included both scientific and nonscientific considerations. By contrast, the EPA's oral reference dose and inhalation reference concentration approaches strive to include scientific-only considerations. The history and rationale for the change in terminology and practice from ADI to oral RfD is described in detail in documentation supporting EPA's Integrated Risk Information System (IRIS) (U.S. EPA 1993, Barnes and Dourson 1988). Texts describing the development of the inhalation reference concentration (RfC) concept are also available (U.S. EPA 1990a, Jarabek et al. 1989, 1990).

EPA defines the 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 an appreciable risk of deleterious effects during a lifetime. The RfD is composed of the no-observed-adverse-effect level (NOAEL) or lowest-observed-adverse-effect level (LOAEL) divided by the composite uncertainty factor (UF) and modifying factor (MF) (Table 1). The following equation is used:

$$\text{RfD} = \frac{\text{NOAEL or LOAEL}}{\text{UF} \times \text{MF}}.$$

The oral RfD and inhalation RfC are useful reference points for gauging the potential effects of other doses. Doses at the RfD (or less) or concentrations at the RfC (or less) are not likely to be associated with any health risks, and are therefore assumed likely to be protective and of little regulatory concern. In contrast, as the amount and frequency of exposures exceeding the RfD or RfC increase, the probability that adverse effects may be observed in a human population also increases. However, the conclusion that all doses below the RfD or RfC are acceptable and that all doses in excess of the RfD or RfC are unacceptable cannot be categorically stated. Moreover, the precision of the RfD or RfC depends in part on the overall magnitude of the composite UFs and MFs used in its calculation. The precision at best is probably one significant figure and more generally an order of magnitude, base 10. As the magnitude of this composite factor increases, the estimate becomes even less precise.

The determination of an RfD or RfC requires scientific judgments as to the appropriate NOAEL of the critical effect and the appropriate UFs and MFs based on database limitations. The logic behind these judgments as well as the assumptions and limitations are briefly discussed below.

Selection of Toxicity Data

After a hazard identification has been conducted and the critical effects have been identified, EPA generally selects an experimental dose rate from a study that represents the highest level tested at which the critical effects were not demonstrated. This level (i.e., the NOAEL) is the key datum gleaned from the toxicologist's review of the chemical's entire database and is the first component in the estimation of an RfD. The use of this NOAEL is based on the assumption that if the critical effect is prevented, then all threshold toxic effects evoked by the chemical at higher doses are prevented. This process also occurs in the estimation of ADIs by other health agencies.

Human toxicity data adequate for use in the estimation of RfDs are seldom available, but if they are available, they are used in the selection of this NOAEL. The use of human data has the advantage of avoiding the problems inherent in interspecies extrapolation.

In the absence of appropriate human data, animal data are closely scrutinized. Animal studies typically reflect situations in which exposure to the toxicant has been carefully controlled and the problems of heterogeneity of the exposed population and concurrent exposures to other toxicants have been minimized. Presented with data from several animal studies, EPA and others first seek to identify the animal model that is most relevant to humans based on the most defensible

data. In the absence of a clearly relevant species, EPA and others generally choose the critical study and species that show an adverse effect at the lowest administered dose.

However, it may be that the toxicity data of one species are clearly atypical compared with other experimental species. If a scientific argument can be maintained that such an outlier response is not expected in humans, this behavior is often discounted. For example, the so-called male rat nephropathy is generally regarded to represent a toxicologic finding in a sensitive animal model that appears to have little relevance for human health hazard assessment (U.S. EPA 1990b).

Confidence In the RfD

When available, adequate data from acceptable human studies are often used as a basis for the RfD because the problem of interspecies extrapolation is avoided and, thus, the confidence in the estimate is often greater. In the absence of such data, RfDs are estimated from studies in experimental animals.

In the absence of adequate human data EPA generally considers a "complete" database, that is, complete for the purpose of calculating a chronic RfD for noncancer health effects, to be composed as follows: 1) two adequate† mammalian chronic toxicity studies by the appropriate route in different species, 2) one adequate† mammalian multigeneration reproductive toxicity study by an appropriate route, and 3) two adequate† mammalian developmental toxicity studies by an appropriate route in different species.

Generally, for such "complete" databases, the likelihood that additional toxicity data may change the RfD is low. Thus, EPA usually has higher confidence in such an RfD; that is, additional toxicity data are not as likely to change the value of a high-confidence RfD compared with a lesser-confidence RfD. However, many chemicals pose only an acute health hazard and general toxicity studies often indicate a need for special studies to assess, for example, neurotoxic or immunotoxic effects. In such cases, the database may not be "complete" without such special studies.

EPA considers a single, well-conducted, subchronic mammalian bioassay by the appropriate route as a minimum database for estimating an RfD. However, EPA generally has less confidence in such an RfD, because additional toxicity data are likely to change (usually increase)

†As determined by professional judgment. EPA and others have published guidelines in this area (e.g., U.S. EPA 1986a,b, 1991, and U.S. Food and Drug Administration [FDA] 1966).

the value of the RfD. Examples of confidence statements with RfDs can be found in EPA's IRIS (U.S. EPA 1993).

Selection of Uncertainty Factors

The choice of appropriate UFs and MFs reflects a case-by-case judgment by experts and should account for each of the applicable areas of uncertainty (described below) and any nuances in the available data that might change the magnitude of any factor. Several reports describe the underlying basis of UFs (Zielhuis et al. 1979, Dourson and Stara 1983) and research into this area (Calabrese 1985, Hattis et al. 1987, Hartley and Ohanian 1988, Lewis et al. 1990, Renwick 1991, 1993, Dourson et al. 1992, Calabrese et al. 1992, Calabrese and Gilbert 1993).

Uncertainty factors are reductions in the dose rate to account for several areas of scientific uncertainty inherent in most toxicity databases. As shown in Table 2, interhuman variability (designated as H) is intended to account for the variation in sensitivity among the members of the human population. Experimental animal to human variability (designated as A) is intended to account for the uncertainty in extrapolating animal data to the case of humans, and is considered to consist primarily of uncertainties in both toxicokinetics and toxicodynamics. Subchronic to chronic variability (designated as S) is intended to account for the uncertainty in extrapolating from less than chronic NOAELs (or LOAELs) to chronic levels. LOAEL to NOAEL variability (designated as L) is intended to account for the uncertainty in extrapolating from LOAELs to NOAELs. Database completeness (designated as D) is intended to account for the inability of any single study to adequately address all possible adverse outcomes.

Each of these five areas is generally addressed by EPA with an order of magnitude factor (i.e., 10). In practice, however, the magnitude of any composite UF is dependent on professional judgment as to the total uncertainty in all areas. If uncertainties in all areas have been resolved, EPA generally uses a 1-fold UF to estimate the RfD. As an example, see the verified RfD for fluoride in EPA's IRIS (U.S. EPA 1993). When uncertainties exist in one, two, or three areas, EPA generally uses a 10-, 100- or 1000-fold UF, respectively. When uncertainties exist in four areas, EPA generally uses a 3000-fold UF.[‡] When a single subchronic animal study that does not define a NOAEL is the only

[‡]This practice of "lumping" several areas of uncertainty is based on the knowledge that each individual factor is generally conservative from the standpoint of the behavior of the average chemical (Dourson and Stara 1983) and that the multiplication of four or five values of 10 is likely to yield unrealistically conservative RfDs.

available information, uncertainties in all five areas exist; this is the minimum database to estimate an RfD. In this minimum case, EPA generally uses a UF of 10,000. Databases that are weaker than a single, animal, subchronic bioassay (that does not define a NOAEL), and that would result in UFs in excess of 10,000, are considered too uncertain as a basis for quantitation. In such cases, the EPA does not estimate an RfD and additional toxicity data are sought or awaited.

The EPA occasionally uses a factor of less than 10, or even a factor of 1, if the existing data reduce or obviate the need to account for a particular area of uncertainty.[§] For example, the use of a 1-year rat study as the basis of an RfD may reduce the need for a 10-fold factor to a 3-fold factor for subchronic to chronic extrapolation, since it can be empirically demonstrated that 1-year rat NOAELs are generally closer in magnitude to chronic values than are 3-month NOAELs (Swartout, personal communication). A recent publication more fully investigates this concept of variable UFs through an analysis of expected values (Lewis et al. 1990).

The composite UF to use with a given database is again strictly a case-by-case judgment by experts, and should be flexible enough to account for each of the applicable five areas of uncertainty and any nuances in the available data that might change the magnitude of any factor. EPA describes its choice of composite UF and subcomponents for individual RfDs in its IRIS (U.S. EPA 1993).

Selection of Modifying Factors

EPA currently uses an additional factor, an MF, as an occasional, necessary adjustment in the estimation of an RfD to account for areas of uncertainty not explicitly addressed by the usual factors. The value of the MF is greater than zero and is less than or equal to 10, but it should generally be developed on a log 10 basis (i.e., 0.3, 1, 3, 10) since its precision is not expected to be any greater than the standard UFs. The default value for this factor is 1.

EPA's reasoning in its use of this MF is that the areas of scientific uncertainty of H, A, S, L, or D, discussed in the preceding section on UFs, do not represent all of the uncertainties in the estimation of an RfD. For example, the fewer the number of animals used in a dosing group, the more likely it is that a NOAEL will be observed. Such a case might argue for modifying the usual 10-fold factors—a 100-fold UF

[§]The usual intermediate factor used is 3, since it is the approximate logarithmic mean of 1 and 10. The choice of 3, instead of a 5, for example, reflects both the expected precision of the UFs (about one digit, log base 10) and the view that it is not generally possible to be more precise than about halfway in considering the nuances of these areas of uncertainty.

Table 2. Description of typical uncertainty and modifying factors in deriving reference doses (RfDs)

Standard uncertainty factors (UFs)	General guidelines[a]
H (interhuman)	Generally uses a 10-fold factor when extrapolating from valid experimental results from studies using prolonged exposure to average healthy humans. This factor is intended to account for the variation in sensitivity among members of the human population.
A (experimental animal to human)	For RfDs, generally uses a 10-fold factor when extrapolating from valid results of long-term studies on experimental animals when results of studies of human exposure are not available or are inadequate. For reference concentrations, this factor is reduced to 3-fold when a NOAEL(HEC) (HEC = human equivalent concentration) is used as the basis of the estimate. In either case, this factor is intended to account for the uncertainty in extrapolating animal data to humans.
S (subchronic to chronic)	Generally uses a 10-fold factor when extrapolating from less than chronic results on experimental animals or humans. This factor is intended to account for the uncertainty in extrapolating from less than chronic NOAELs to chronic NOAELs.
L (LOAEL to NOAEL)	Generally uses a 10-fold factor when deriving an RfD from a LOAEL instead of a NOAEL. This factor is intended to account for the uncertainty in extrapolating from LOAELs to NOAELs.
D (incomplete database to complete)	Generally uses a 10-fold factor when extrapolating from valid results in experimental animals when the data are "incomplete." This factor is intended to account for the inability of any single study to adequately address all possible adverse outcomes.
MF (modifying factor)	Uses professional judgment to determine an additional uncertainty factor, termed a "modifying factor," that is greater than zero and less than or equal to 10. The magnitude of the MF depends on the professional assessment of scientific uncertainties of the study and database not explicitly treated above (e.g., the number of animals tested). The default value for the MF is 1.

Source: Adapted in part from Dourson and Stara (1983), Barnes and Dourson (1988), and Jarabek et al. (1989).

Note: The maximum uncertainty factor used with the minimum confidence database is 10,000. See text for discussion.

[a]Professional judgment is required to determine the appropriate value to use for any given UF. The values listed in this table are nominal values that are frequently used by the EPA.

might be raised to 250 if too few animals were used in a chronic study. Although this increase is scientifically reasonable, it introduces two difficulties: the adjustment of the standard 10-fold values might be perceived as arbitrary, and the precision of some of the resulting UFs (in this example a UF of 250 has an implied precision of two digits) is not appropriate in relation to the variability of the biological response. EPA intends to avoid these difficulties through use of the MF. Thus, EPA might in the former case use a UF of 100 and an MF of 3 to arrive at the best estimation of an RfD. The resulting UF × MF of 300 would avoid the perception of arbitrariness and maintain consistency with the overall precision of the standard UFs.

Assumptions and Limitations

The basic assumption in the development of an RfD is that a threshold exists in the dose rate at or above which an adverse effect will be evoked in an organism. EPA and others consider this assumption to be well founded. It is supported by known mechanisms of toxicity of many compounds, which show that a known physiologic reserve must be depleted and/or the repair capacity of the organism must be overcome before toxicity occurs (Doull et al. 1980).

A second assumption is that the RfD represents an estimate of a population subthreshold dose rate, that is, it adequately protects sensitive humans, since one of the UFs accounts for the variability of individual thresholds. In unusual situations, for some compounds, sufficient evidence of hypersusceptibility or chemical idiosyncrasy exists to warrant some concern over whether the RfD is always sufficiently low to account for the population threshold. In contrast, the larger the UF, the more likely the RfD will be well below this population threshold, because UFs are generally conservative. Thus, as noted previously, while exposures at or below the RfD should not categorically be considered "safe," exposures above the RfD should not categorically be considered "unsafe." Rather, the RfD should be considered a level that is likely to be without an appreciable risk of deleterious effects over a lifetime and whose precision is at best one significant digit but more often one order of magnitude (i.e., 10-fold) or greater.

The third assumption in the development of an RfD relates to the use of the critical effects. EPA assumes that if the critical effects are prevented, other adverse threshold effects that occur at higher doses are prevented. However, since the RfD procedure generally ignores the dose-response slope, and if these other adverse effects have shallower slopes, estimating the RfD on the basis of the critical effect may not be sufficiently protective. For this reason, information on slopes of dose-response curves may be used to determine the critical effects.

Several other assumptions are used as well. For example, it is often assumed that using animals of different ages in experiments does not affect the resulting comparison of NOAELs and LOAELs. This assumption ignores possible differences in toxicity between ages.

The scientific strengths and limitations of this approach are similar to those for the former "ADI = NOAEL/Safety Factor" approach and have been discussed in the literature (Munro and Krewski 1981, Lu 1983, 1985, 1988, Krewski et al. 1984, Crump 1984, 1986, Dourson et al. 1985, 1986, Barnes and Dourson 1988, Kimmel and Gaylor 1988).

The scientific strengths, in brief, are that all toxicity data are reviewed in the choice of the NOAEL of the critical effects, and that uncertainties in the entire database can be factored into the resulting value of the RfD through the use of professional judgment as to the appropriate UFs and MFs. The limitations, in brief, are that the NOAEL is directly influenced by the choice of dose spacing, the number of animals used in an experiment, and factors that influence the precision and quality of the study. That is, studies with wide dose spacing and few animals per dose group can lead to more poorly characterized RfDs compared with studies with tighter dose spacing and more animals per dose group (e.g., Hattis et al. 1987, Leisenring and Ryan 1992). The NOAEL is also not generally influenced by the nature of the dose-response curve. Uncertainty factors, although considered necessary and perhaps accurately showing the potential underlying variability of different areas of uncertainty, are imprecise. Nor does the RfD approach enable one to estimate risks at exposures greater than the RfD. Scientists are developing methods that attempt to address this latter limitation (e.g., DeRosa et al. 1985, Dourson et al. 1985, Hertzberg 1989, Kimmel et al. 1988, Kimmel and Gaylor 1988, Hertzberg and Dourson 1993).

References

Barnes DG, Dourson ML (1988) Reference dose (RfD): description and use in health risk assessments. Regul Toxicol Pharmacol 8:471–86

Calabrese EJ (1985) Uncertainty factors and interindividual variation. Regul Toxicol Pharmacol 5:190–96

Calabrese EJ, Beck BD, Chappell WR (1992) Does the animal to human uncertainty factor incorporate interspecies differences in surface area? Regul Toxicol Pharmacol 15(2):172–79

Calabrese EJ, Gilbert CE (1993) Lack of total independence of uncertainty factors (UFs): implications for the size of the total uncertainty factor. Regul Toxicol Pharmacol 17:44–51

Crump KS (1984) A new method for determining allowable daily intakes. Fundam Appl Toxicol 4:854–71

Crump KS (1986) Letter to the Editor reply. Fundam Appl Toxicol 6:183–84

DeRosa CT, Stara JF, Durkin PR (1985) Ranking of chemicals based upon chronic toxicity data. Toxicol Ind Health 1(4):177–92
Doull J, Klaasen C, Amdur M (1980) Casarett and Doull's toxicology, 2nd ed. MacMillan, New York, p 718
Dourson ML, Hertzberg RC, Hartung R, Blackburn K (1985) Novel approaches for the estimation of acceptable daily intake. Toxicol Ind Health 1(4):23–41
Dourson ML, Hertzberg RC, Stara JF (1986) Letter to the Editor. Fundam Appl Toxicol 6:182–83
Dourson ML, Knauf L, Swartout JC (1992) On reference dose (RfD) and its underlying toxicity data base. Toxicol Ind Health 8(3):171–89
Dourson ML, Stara JF (1983) Regulatory history and experimental support of uncertainty (safety) factors. Regul Toxicol Pharmacol 3:224–38
Hartley WR, Ohanian EV (1988) The use of short-term toxicity data for prediction of long-term health effects. In Hemphil DD (ed), Trace substances in environmental health—22. University of Missouri, Columbia, MO, pp 3–12
Hattis D, Erdreich L, Ballew M (1987) Human variability in susceptibility to toxic chemicals—a preliminary analysis of pharmacokinetic data from normal volunteers. Risk Anal 7(4):415–26
Hertzberg RC (1989) Fitting a model to categorical response data with application to species extrapolation of toxicity. Health Physics 57:405–09
Hertzberg RC, Dourson ML (1993) Using categorical regression instead on a NOAEL to characterize a toxicologist's judgment in noncancer risk assessment. in press
Jarabek AM, Menach MG, Overton JH, et al. (1989) Inhalation reference dose (RfD_i): an application of interspecies dosimetry modeling for risk assessment of insoluble particles. Health Physics 57:177–83
Jarabek AM, Menach MG, Overton JH, et al. (1990) The U.S. Environmental Protection Agency's inhalation RfD methodology: risk assessment for air toxics. Toxicol Ind Health 6(5):279–301
Kimmel CA, Gaylor DW (1988) Issues in qualitative and quantitative risk analysis for developmental toxicity. Risk Anal 8(1):15–20
Kimmel CA, Wellington DG, Farland W, et al. (1988) Overview of a workshop on quantitative models for developmental toxicity risk assessment. Environ Health Perspect 79:209–15
Krewski D, Brown C, Murdock D (1984) Determining "safe" levels of exposure: safety factors or mathematical models? Fundam Appl Toxicol 4:5383–94
Leisenring W, Ryan L (1992) Statistical properties of the NOAEL. Regul Toxicol Pharmacol 15(2):161–71
Lewis SC, Lynch JR, Nikiforov AI (1990) A new approach to deriving community exposure guidelines from no-observed-adverse-effect levels. Regul Toxicol Pharmacol 11:314–30
Lu F (1983) Toxicological evaluations of carcinogens and noncarcinogens: pros and cons of different approaches. Regul Toxicol Pharmacol 3:121–32
Lu F (1985) Safety assessments of chemicals with thresholded effects. Regul Toxicol Pharmacol 5:460–64
Lu F (1988) Acceptable daily intake: inception, evolution, and application. Regul Toxicol Pharmacol 8:45–60

Munro IC, Krewski DR (1981) Risk assessment and regulatory decision making. Food Cosmet Toxicol 19:549–60
National Academy of Sciences (NAS) (1983) Risk assessment in the federal government: managing the process. National Academy Press, Washington, DC
Renwick AG (1991) Safety factors and establishment of acceptable daily intake. Food Addit Contam 8(2):135–50
Renwick AG (1993) Data derived safety factors for the evaluation of food additives and environmental contaminants. in press
U.S. Environmental Protection Agency (U.S. EPA) (1986a) Guidelines for carcinogen risk assessment. Federal Reg 51:33992–34003 (Sept 24)
U.S. EPA (1986b) Guidelines for mutagenicity risk assessment. Federal Reg 51:34006–12 (Sept 24)
U.S. EPA (1990a) Interim methods for development of inhalation reference doses, EPA/600/8-90-066A. Office of Health and Environmental Assessment, Washington, DC
U.S. EPA (1990b) Alpha 2 microglobulin association with renal toxicity and neoplasms in male rats (draft report prepared for the Risk Assessment Forum). Office of Research and Development, Washington, DC
U.S. EPA (1991) Proposed amendments to the guidelines for the health assessment of suspect developmental toxicants; request for comments; notice. Federal Reg 54:9386–9403 (March 6)
U.S. EPA (1993) The Integrated Risk Information System (IRIS) [online]. Office of Health and Environmental Assessment, Environmental Criteria and Assessment Office, Cincinnati, OH
U.S. Food and Drug Administration (U.S. FDA) (1966) Guidelines for reproduction and teratology of drugs. Bureau of Drugs, Rockville, MD
Zielhuis RL, van der Kreek FW (1979) The use of a safety factor in setting health based permissible levels for occupational exposure. Int Arch Occup Environ Health 42:191–201