This document was prepared under contract 3W-0477-NASX and does not represent policies or conclusions of the U.S. Environmental Protection Agency

METHODS FOR IDENTIFYING A DEFAULT CROSS-SPECIES SCALING FACTOR

Prepared by:

Lorenz R. Rhomberg, Ph.D.[1]

Thomas A. Lewandowski, Ph.D., DABT[2]

[1]Gradient Corporation
238 Main Street
Cambridge, MA 02142

[2]Gradient Corporation
600 Stewart Street, Suite 803
Seattle, WA 98101

Prepared for:

Risk Assessment Forum

U.S. Environmental Protection Agency

1200 Pennsylvania Avenue, N.W.

Washington, D.C. 20460

April 2, 2004

TABLE OF CONTENTS

LIST OF FIGURES

1. Introduction – The Need for Default Dose Scaling Factors

As a matter of necessity, the potential for a chemical agent to produce adverse health effects in humans is often investigated in experimental animals, typically rats and mice. The use of these surrogates is premised upon the high degree of physiological, biochemical, and anatomical similarity among mammalian species; toxicological effects observed in the experimental animals may be taken as evidence that humans might show similar responses to equivalent chemical exposures. What constitutes “equivalent" is of course the challenging aspect of this assumption and comprises the topic of this paper.

The importance attached to the concept of toxicologically equivalent doses varies somewhat depending on the stage of the risk assessment process. For qualitative evaluations such as hazard identification, there is a lesser need to ensure that doses are equivalent between humans and experimental animals. In fact, dosing animals at much higher levels than probable human exposures is the default approach, intended to elicit a detectable response in a limited population of experimental animals. When the aim is dose-response extrapolation and setting acceptable levels of human exposure, however, one must be able to specify a quantitative relationship between the dose levels in humans and in animals that is expected to result in the same degree of adverse effect. In this case, it is necessary to take into account the pronounced difference in scale between the tested model organisms and humans. Even if fundamental similarity is presumed, one must allow for the fact that the greater size and lifespan of the human relative to the experimental animal has a significant impact on the amount of chemical intake needed to provoke a given response.

The question of cross-species dose equivalency has traditionally been handled differently in assessment of cancer versus non-cancer toxicity. Although approaches within each realm have varied, the key difference between them is that cancer risk assessment has focused on explicit extrapolation from animal to human doses, using a quantitative adjustment to calculate a presumed "human equivalent" dose, whereas non-cancer assessment has relied on a semiquantitative allowance for uncertainty in the human dose that may be equivalently toxic, accomplished by applying an uncertainty factor that does not pretend to be a precise extrapolation. The dose extrapolation in cancer assessment is (at least by intention) a central

estimate, producing a single best guess at an equivalent human dose while acknowledging that there is a distribution of uncertainty around this extrapolation that rarely receives analytical examination. The "animal-to-human" uncertainty factor of non-cancer assessment, in contrast, is intended as a lower bound on the range of possible human doses that might be toxicologically equivalent (in view of chemical-to-chemical variation in how sensitive humans are vis-à-vis experimental animals), and as such ought to be somewhat conservative on average. A harmonized approach should seek to find commonality between these approaches. This would seem to entail bringing a focus on best-guess extrapolation of equitoxic doses to assessment of non-cancer endpoints and a more explicit allowance for uncertainty around dose-equivalence calculations to the assessment of cancer endpoints.

In the absence of chemical-specific information sufficient to do otherwise, the guidance of the U.S. Environmental Protection Agency (EPA) for carcinogen risk assessment is to apply a default animal-to-human oral dose extrapolation based on presumed toxicological equivalence of daily doses scaled by the ¾-power of body weight (i.e., $mg/kg^{3/4}$/day doses are presumed equivalent) (US EPA 1992, 1999). (Inhalation exposures are assessed according to methods developed for non-cancer assessment, as described in US EPA 1994; Jarabek 1995). When such doses are reexpressed in the traditional units of mg/kg/day, it is seen that a human dose about 7-fold less than a mouse dose or 4-fold less than a rat dose are being presumed to be equivalently carcinogenic.

The current default practice for non-cancer risk assessment (Dourson, 1996) is to apply an animal-to-human uncertainty factor (U_A), reducing the human mg/kg/day dose by 10-fold to allow for the fact that humans might be up to this much more sensitive than the test species to which they are being compared. The factor of 10 is based on a combination of professional judgment, experience, and empirical data suggesting that a larger adjustment is rarely needed (Dourson and Stara, 1983; Dourson et al., 1992, 1996). While this approach has worked well in practice, it has the shortcoming of failing to address the actual central estimate of the expected human equivalent dose. This has the following consequences:

- it is hard to tell how much of the 10-fold is needed adjustment to achieve human dose-equivalency on average versus an allowance for chemical-by-chemical uncertainty in that adjustment. Accordingly, it is hard to know how conservative the assessment is on average;

- when animal tests are being compared (on a mg/kg/day basis) the truly most sensitive species may not be chosen for the critical endpoint if a different scaling of doses is actually more appropriate to judge relative sensitivity. For instance, if doses proportional to a fractional power of body weight tend to be equitoxic, larger species will tend to be more sensitive on a mg/kg/day basis, and these may be chosen inappropriately;
- without explicit assumptions about dose-equivalency, harmonization with cancer assessment methods is problematic.

For these reasons, an examination of animal-to-human dose extrapolation for presumed equitoxicity with respect to non-cancer effects is in order. We note that much of the reasoning, rationale, and empirical evidence cited for cancer risk assessment's use of $mg/kg^{3/4}$/day scaling (US EPA, 1992) would seem to apply equally well to chronic oral exposures for non-cancer effects. In the present analysis, we examine the basis for deriving a default approach to dose extrapolation for non-cancer effects. We leave for discussion in another forum the related question of how cancer and non-cancer assessment should address the uncertainty in animal-to-human dose extrapolation. That is, we are seeking insights into how to choose a default approach for a central estimate, best-guess of the human doses that are equivalent in impact to observed animals doses.

Defining "toxicologically equivalent" doses has been problematic. Alternatives that have been used include scaling daily administered amounts by body weight or by body surface area; scaling cumulative lifetime intake by body weight; and equating exposures to contaminated air, food, or water according to the concentration of toxic agent. Despite considerable study and debate (Pinkel, 1958; Freireich et al., 1966; Mantel and Schneiderman, 1975; Rall, 1977; Hoel, 1977; Hogan and Hoel, 1982; Calabrese, 1983, 1987; Crump et al., 1985; Davidson et al., 1986; Hill et al., 1986; Vocci and Farber, 1988; Travis and White, 1988; Beliles and Parker, 1989; Crump et al., 1989; Jarabek et al., 1989; Travis and Bowers, 1991; USEPA, 1992; Rhomberg, 1994; Baird et al., 1996; Rhomberg, 1995; Jarabek, 1995; Riviere et al., 1997; Rhomberg and Wolff, 1998; Rhomberg and Caprario, 1999; Kalberlah et al., 2002; Meek et al., 2003), no alternative has emerged as the definitive solution, either on empirical or theoretical grounds.

Prior work has looked at cross-species scaling methodologies for use in dose-response assessment for chemical carcinogens (US EPA, 1992; Rhomberg, 1994; Kuo et al., 2002). Ironically, most of the data used for those comparisons were based not on tumor incidence but on acute lethality or other non-cancer endpoints. Past emphasis on developing scaling strategies for

carcinogenic endpoints can be attributed to the driving role carcinogens have played in standard-setting and environmental remediation. This pre-eminent focus on carcinogens has lessened with increasing recognition of the implications of less-than-lifetime exposures for certain equally important health endpoints (e.g., development, neurotoxicity). Currently, there is a desire in the regulatory arena to harmonize the procedures used in dose-response assessment and risk characterization for chemicals across cancer and non-cancer health effects, recognizing that many chemicals fall into both categories and that a logically consistent framework for risk assessment is required (Conolly, 1995; Gaylor and Kodel, 1999; Butterworth and Bogdanffy, 1999; Bogdanffy et al., 2001; Clewell et al., 2002a).

The goal of the present evaluation is to provide a rationale for selecting default cross-species extrapolation factors with which to scale oral doses across species for non-carcinogenic endpoints. Recognizing the importance of incorporating specific chemical data wherever appropriate, a second goal is to describe how chemical-specific data (e.g., metabolic or mechanistic data) can be incorporated into the dose extrapolation process.

This paper is focused primarily on oral doses and oral exposures, as these represent: (1) the most common route of dosing used in toxicology studies, and (2) the route least affected by factors (e.g., differences in respiratory physiology or skin thickness) which are unlikely to follow a clear relationship across species. The oral route is not the only exposure route for which cross-species scaling is of interest but, for the reasons just given, it is considered as the base case upon which further scaling relationships can be built to account for factors specific to other exposure routes.

It should be noted that the term "scaling" as used in this document does not specifically address the issue of inter-species uncertainty in risk assessment. In particular, the default scaling factors discussed below will not yield human equivalent doses that exactly describe the relationship in toxicity between humans and experimental species for each and every chemical. Uncertainty will remain on a chemical-specific basis in terms of species differences (and intra-individual differences) in pharmacokinetics and pharmacodynamics. Only by incorporating chemical-specific data into the scaling process can this uncertainty be reduced.

2. Properties of a Default Scaling Approach

Prior to specifying how one derives a default scaling factor, it is useful to specify what the properties of the ideal scaling factor should be. While it is unlikely that the ideal can be achieved, it is worth articulating the desirable qualities. These qualities include:

(1) As a default, the scaling method should be based on generally applicable arguments. It will not address every chemical equally well, but it should address the majority of chemicals fairly well using a consistent rationale, and there should not be whole groups of chemicals for which it is not suitable.

(2) The goal of a cross-species scaling methodology is not to arrive at true values of equivalent doses under all circumstances (for this is impossible, even in principle). Rather it is to embody correctly and without bias the impact of the information at hand, providing rational estimates that take into account what is known, recognizing that true values vary around this estimate as a result of case-by-case particulars, many of which are either unknown or are known to vary among individuals for whom the projections are being made.

(3) The default factor should also be based on clearly articulated principles about the determinants of equitoxic doses and a conceptual approach as to what governs differences (or similarities) in dose-response across species.

Thus, the logic and principles need to be laid out prior to the identification or selection of specific scaling factor values. While this may seem obvious, it has in fact not often been the practice. Too often scientists working in the risk assessment field are faced with the problem of retroactively developing arguments to support use of specific parameters based on policy decisions. The shortcoming of the retroactive approach is that arguments supporting the chosen values developed after the fact may lack credibility and the process used to develop them be viewed as arbitrary.

To achieve the general applicability described above it is necessary to rely on simplified and overriding schemes of biological variation which, while embodying fundamental species differences, may not encompass all of the details and causes of case-by-case variation. This is not to deny the importance of these details, nor to denigrate the value of case-specific data that show species- or dose-related differences in uptake, metabolism, or physiological actions of putative

toxic agents. To the contrary, the intention is to provide a framework for the use of such data, allowing (and indeed encouraging) one to go beyond the *prima facie* case based on overall trends. In fact, it is highly desirable that a weight-of-evidence approach be employed such that different approaches or bases for developing scaling factors yield roughly similar results. In such a case, it would be preferable to have a less precise scaling factor that comports with several lines of logical argument than one that is only well supported by a single line of argument.

One common misconception is that scaling only needs to address metabolic (*i.e.*, pharmacokinetic) differences between species. There appears to be the incorrect understanding that doses are adjusted to some power of body weight solely because small animals metabolize and eliminate compounds more quickly than larger animals. It is important to bear in mind that other processes that influence disposition of a dose, such as uptake and non-metabolic clearance processes, also share the pattern of faster rates in smaller species. Moreover, pharmacodynamic processes (*e.g.*, damage to macromolecules, cellular repair and regeneration, signaling cascades, sustained proliferative responses) are also encompassed by the cross-species dose scaling factor. To date, while cross-species scaling strategies for pharmacokinetics have been widely discussed (Reitz *et al.*, 1988; Starr, 1990; Casanova et al., 1991; Cox and Ricci, 1992; Conolly and Andersen, 1993; Andersen and Krishnan, 1994; Watanabe and Bois; 1996; Riviere et al., 1997; Morgan, 1997), attention to pharmacodynamic differences has been considerably less. (An exception is work on one chemical, formaldehyde (Monticello and Morgan, 1994; Conolly *et al.*, 2000; Schlosser *et al.* 2003.) The contribution to the default scaling method of pharmacodynamics must still be accounted for, or the scaling factors will be in error. For example, the fact that mice metabolize a DNA-reactive compound more readily to an unreactive metabolite may be offset by the fact that mice may have less efficient DNA repair compared to humans (Hart and Setlow, 1974). Investigating, quantifying and incorporating pharmacodynamic differences between species currently represents one of the frontiers of risk assessment science.

The defaults must not be viewed as an end in themselves. The process of deriving the default must, to the extent possible, identify how chemical-specific data can be incorporated into the process to result in a cross-species extrapolation with reduced uncertainty, owing to the replacement of general scaling arguments with case-specific observations. Similarly, the process must identify where the default factors will not apply. For example, a local effect such as irritation of stomach mucous membranes by corrosive chemicals is likely to be governed by properties such as stomach pH and buffering capacity, properties that may not vary among

species according to any of the commonly used scaling relationships. The toxicity of inhaled or dermally absorbed chemicals also poses unique problems for extrapolation between rodents and humans as discussed later in this document.

3. Possible Approaches for Deriving a Default Scaling Factor

There are two broad and complementary approaches to choosing a cross-species scaling factor. The first is empirical; one seeks cases of agents with experimental data regarding toxicological potency in human and experimental animals. Over many such cases, one then examines the predictive ability of alternative methods for extrapolation of effect levels across species to derive a generally applicable relationship. The second approach is theoretical, and is grounded in the principles of allometry, which is the study of the regular variation in features of anatomy and physiology as a function of overall body size. The strategy for this second approach is to develop a scientific rationale for a particular scaling factor by investigating the cross-species patterns of variation in biological features and processes that constitute the key influences on toxicological potency.

These two approaches need not be mutually exclusive. Each has it own advantages and neither alone is sufficient. The empirical approach offers the advantage of providing "hard" evidence of a scaling relationship for specific chemicals (or groups of chemicals) but suffers from an inability to demonstrate applicability in all situations *(i.e.*, because it is focused on the observable patterns of scaling rather than on the underlying basis). In contrast, the allometric approach is focused on comprehensive understanding of biological phenomena that are the basis of cross species variability (*i.e.*, organ volumes, blood flows, metabolic rates, *etc.*). The allometric approach, by itself, remains hypothetical—it lacks the ability to demonstrate the relevance of the observed allometric relationships to toxicological outcomes (*i.e.*, it is uncertain whether the biological phenomena considered encompass all the relevant ones). Each approach involves an acceptance of a particular uncertainty; either uncertainty regarding the generic nature of empirical observations on a limited number of chemicals or uncertainty regarding the relevance of specific allometric relationships to particular chemical modes of action. However, if both approaches are used in a complementary fashion and both suggest a similar dose scaling method (or at least if they do not yield contradictory conclusions), then we can have greater confidence that the default factors derived are both highly relevant and yet sufficiently general.

4. Empirical Approach

As indicated above, the empirical approach attempts to identify a cross-species scaling factor that is empirically successful in producing good estimates of potency in humans from data on potencies in other species. The underlying reason why such a factor works is a secondary consideration. The advantage of the empirical approach is that, by directly examining toxicological potencies (rather than influences on potency, such as rates of individual pharmacokinetic or pharmacodynamic processes), all relevant factors are included.

The disadvantage of the empirical approach is that the data are few and confined to certain classes of chemicals and certain toxicologic endpoints. The evaluation depends of necessity on having comparable data in multiple species for a wide range of chemicals. Ideally, one would want data for at least several chemicals in each chemical class (metals, chlorinated solvents, non-chlorinated aromatic solvents, aliphatic solvents, tricyclic antidepressants, organophosphorus pesticides, *etc.*) in several mammalian species of differing body size and longevity. Such a breadth of data is not currently available. Except in the case of some pharmaceuticals, non-cancer testing data rarely provide comparable data (similar experimental designs and endpoints) in more than one species. Data collected from multiple species, when available, often consist of experimental data in mice and rats (species that are too close in body size to be of much use by themselves in extrapolation) and very limited data in humans (typically individual poisoning episodes with poorly characterized exposures). Thus, further comparative studies on the potency of toxic agents are needed, but few systematically compiled datasets are available. In employing the empirical approach, one must hope that the idiosyncratic properties of the chemicals (or chemical classes) average out to give a good estimate of the general relationship.

Although data are limited, a number of authors have attempted empirical analyses. Much of the data evaluated has involved the acute (or subacute) toxicity of antineoplastic agents because unlike other chemicals they are intentionally administered to humans at doses associated with some degree of toxicity. Freireich *et al.* (1966), testing and extending the suggestion of Pinkel (1958), examined maximum tolerated doses (MTDs) of 18 antineoplastic drugs in mice, rats, hamsters, dogs, monkeys and humans. LD_{10}s were used for rodents, and were presumed to

be an equivalent level of toxicity to an MTD. Doses from experiments of different length were reexpressed in terms of an exposure regimen of five consecutive days, on the assumption that cumulative dose is proportional to effect. The authors concluded that, when doses were expressed as mg/m^2 body surface area/day, good predictions of human MTDs were obtained from all animal species, but that body weight scaling of doses (*i.e.*, mg/kg BW$^{1.0}$/day) overpredicted human MTDs (*i.e.*, underpredicted potency in humans) by a margin that increased as one extrapolated from smaller and smaller species.[1]

Subsequent to these early attempts to look at the possibility of deriving scaling relationships for pharmaceuticals among species, many authors examined possible relationships of pharmacokinetics among species for specific drugs. Perhaps the most well known is the analysis of Dedrick *et al.* (1970) which examined the pharmacokinetics of methotrexate in mice, rats, dogs, monkeys and humans. Dedrick subsequently performed a similar analysis of the pharmacokinetics of the antineoplastic drug Ara-C (Dedrick, 1973). The number of similar studies conducted by other authors on other pharmaceutical compounds is far too numerous to review here but the reader is referred to a number of reviews that have put the large amount of data in a coherent context (Boxenbaum and Ronfeld, 1983; Andersen *et al*., 1995; Ings, 1990). It is important to note that the majority of these studies examined inter-species differences in pharmacokinetics only – pharmacodynamic differences were not considered. These studies therefore yield a valuable but incomplete picture of interspecies differences in toxicity, the endpoint of regulatory interest.

Collins *et al.* (1986, 1990) found that the human MTD for 16 antineoplastic drugs was well predicted on average by the mouse LD_{10} when doses were expressed as mg/m^2 of body surface area.[2] If these endpoints of acute toxicity are taken as equivalent, scaling doses in proportion to surface area tends to equalize toxicity across species. Moreover, Collins *et al.* (1990) compared the blood levels (in terms of areas-under-the-curve of concentration in plasma as it declines over time, or "C x T") that correspond to equally toxic administered doses and found that these were an even better predictor, in that they displayed less case-by-case variation.

[1] Because an MTD is intended to be a dose causing no lethality, while an LD_{10} causes 10% lethality, the equivalence of these two end points can be questioned. Antineoplastic drugs typically have very steep dose-response curves, however, and survival near the MTD is maintained by close monitoring and intervention, which the rodent LD_{10} determinations lack.

[2] If the MTD is considered to be a less severe end point, in such comparisons potencies in the larger species are overestimated vis-à-vis those in rodents; a bias would then be created that would increase the apparent success of surface area scaling compared to scaling by body weight.

These results illustrate three points: 1) Scaling administered doses in this way tends to equalize blood levels across species; 2) at least for these effects, areas-under-the-curve of blood concentration may serve as a predictive measure of the toxic response to a dose, even across species; and 3) obtaining pharmacokinetic data on internal dose measures can increase the precision of the cross-species prediction of equivalently toxic doses by accounting for case-by-case variation.

Travis and White (1988) reanalyzed the Freireich *et al*. (1966) data set and nearly doubled the number of drugs by adding a similar data set of Schein *et al*. (1979). Instead of simply examining the success of previously proposed scaling methods, they used regression techniques empirically to determine the optimal power of body weight (in the dose measure $mg/kg^{power}/day$) to achieve the best-fitting allometric relationship of MTDs across species. For both data sets individually and for the combined data set, a power of 0.72 to 0.74 led to the best cross-species predictions. In the analysis of the combined data, a power of unity (body weight scaling) was clearly rejected at the 95% level of significance, and a power of 2/3 (so-called surface area scaling[3]) was barely rejected. The authors discuss the history of empirical studies of allometric variation in a number of physiological features, primarily basal metabolism, and argue that their result is part of a general empirical support for scaling daily doses by the ¾ power of body weight.

Travis and Bowers (1991) investigated the anesthetic potency for 11 volatile anesthetics (*e.g*., halothane, nitrous oxide, diethyl ether). Substantial data were available for mice, rats, rabbits, cats, dogs and humans, with less data for monkeys, pigs and horses. They observed that equivalent blood levels of the anesthetics resulted in similar degrees of anesthesia. Since the amount of air intake form breathing scales in proportion to the ¾ power of body weight, this implies that intakes in that proportion produce equivalent effects, although the application of an intake basis for inhaled volatile gases is unclear since uptake may decline as blood levels rise.

Baird *et al*. (1996) described probabilistic methods for setting RfDs. Part of their analysis involved an evaluation of the uncertainty in interspecies extrapolation based on body surface area (*i.e.*, $BW^{2/3}$). Using data previously compiled and published by Dourson *et al*. (1992) on the toxicity of 69 pesticides, they compared NOAELs in mice, rats and dogs using

[3] Scaling daily doses by the 2/3 power of body weight is called "surface area" scaling because the surface area of similarly shaped objects varies as the 2/3 power of their volume.

surface area scaling as a standardization method. They found that the ratio of the NOAELs (*e.g.*, $NOAEL_{rat}/NOAEL_{dog}$) when based on body weight was significantly different from 1.0 in two of the three possible comparisons (mouse/rat and rat/dog; mouse/dog was not significantly different from 1.0). They concluded that the uncertainty involved in extrapolating between mice and rats to humans was likely to be at least as large as the extrapolations between rodents and dogs. They also concluded that surface area scaling was not clearly the best method for cross-species extrapolation and that developing chemical- and species-specific information was highly desirable.

Similarly, Schmidt *et al*. (1997), analyzing a dataset of NOAELs for 56 antineoplastic drugs that represent a further expansion of the dataset used by Travis and White (1988), found that when doses were expressed in units of mg/kg/day, the apparent toxicity generally increased with the species' body weight over the series mouse, rat, monkey, dog, human. This implies that a scaling method with dose proportional to a power of body weight less than one would be appropriate for determining equally toxic doses of these direct-acting agents.

Rhomberg and Wolff (1998) examined patterns in the correspondence of single oral dose LD_{50} values across several species (mouse, hamster, rat, guinea pig, rabbit, cat, monkey, dog and human). In contrast to earlier analyses, they found that scaling by body weight to the power of one (*i.e.*, mg/kg $BW^{1.0}$) best fit the data, rather than scaling to $BW^{3/4}$. One limitation of this analysis was the small number of human datapoints; most of the data were obtained in comparisons among rodents (*i.e.*, mouse, rat, guinea pig, hamster) where the correlation was strongest. Rhomberg and Wolff also observed that some species (notably hamsters) appeared to be more resistant to acute toxicity than predicted by the general scaling rule. Such a species-specific pattern should be noted in development of a default scaling factor, for example, by including a precautionary note that data from a certain species may warrant further consideration when extrapolating to humans. Rhomberg and Wolff suggest that differences with earlier research showing $BW^{3/4}$ being superior may have to do with the dosing regime. The earlier studies were based primarily on subacute (*i.e.*, several-day), repeated dosing, whereas their findings are based on single-dose regimens. Differences in the two systems, they suggest, may be due to the contrast of damage rapidly overwhelming the standing levels of defenses by the single exposures on the one hand and lesser but accumulating damage overcoming ongoing repair or coping processes caused by the repeated exposures on the other. This suggests that the dose-scaling patterns across species should be different depending on pattern of exposure. Because the

exposure patterns EPA deals with are typically longer in duration, this argument supports use of a fractional power of body weight as a scaling factor.

Rhomberg and Caprario (1999) extending the findings of Rhomberg and Wolff (1998), looking at LD_{50} values for other routes of administration (*i.e.*, intravenous, intraperitoneal and intramuscular). Over 3,000 agents were evaluated in pair-wise species comparisons across 8 species (mouse, rat, hamster, guinea pig, rabbit, cat, dog and monkey). Results were similar to the oral LD_{50} analysis, indicating that for severe acute exposures, scaling to $BW^{1.0}$ provided a good extrapolation factor across species.

Studies of the empirically derived data on cross-species differences in chemical potency suggest that scaling to $BW^{3/4}$ represents a reasonably well-supported approach for scaling ongoing oral doses across species. The ¾ power of body weight scaling represents an approximation appropriate to most chemical and chemical classes, in keeping with the desired properties of a default factor identified above. It should nonetheless be noted that support for this approach is not entirely robust. The supporting data are more heavily concentrated in small animals rather than evenly spread across animals of different sizes, including humans. The data have also been largely obtained from acute or subacute exposure scenarios, involving endpoints that could be more closely related to the peak dose than to the AUC. Thus, scaling to other powers (*e.g.*, the 2/3-power, so-called surface-area scaling) is not conclusively rejected but the weight of the available evidence suggests that the ¾ power scaling is an appropriate choice for scaling the chronic exposures which are typically of concern to EPA.

5. Allometric Arguments

The complement to the empirical investigation of potency scaling is a theoretical approach that seeks to identify the biological factors whose variation underlies the variation in a chemical's toxicological potency across species, and then attempts to adjust for their effect. Clearly, these factors are numerous and, for the most part, poorly understood. Fortunately, there are some rather simple and general quantitative patterns in the variation of many features of anatomy and physiology across differently sized mammalian species, representing broad trends in the way the essentially similar mammalian system operates in large and small editions. Although specific processes acting on specific chemicals can (and do) deviate from these broad trends, the

general patterns observed can provide a benchmark that expresses the expectation about a chemical's toxicologic potency in small mammals such as experimental rodents and larger ones such as humans. This expectation can be refined (or refuted) by case-specific biological and mechanistic data, when available, showing how the actual processes involved in producing toxicity differ from the presumptions of the broad trend analysis that serves as the default.

In short, the allometric approach aims at characterizing the impact of scale – how the regular patterns of size, physiological pace, and lifespan across the spectrum of differently sized mammals affects the toxicological process – and then allows for normalization of dose to account for this scale effect.

The effort to quantify physiological patterns among different species is not new; review articles on the subject were published over 50 years ago (Kleiber, 1932; Adolph, 1949). Allometric research in dosimetry has been in large part driven by the need to establish initial doses for human drug pharmaceutical trials based on the findings of animal data (reviewed by Mahmood, 1999). For example, the work of Dedrick et al. (1970), looking at the comparison of methotrexate pharmacokinetics across mice, rats, dogs, monkeys and humans, was geared towards the development of disposition models which could predict drug concentration at the target site for various dosing regimens. Dedrick et al. demonstrated that although the pharmacokinetic profile of methotrexate varied across the species by over 10,000-fold, the data could be brought into concordance via a scaling approach that resulted in a single curve fitting data for multiple species (such "Dedrick plots" scale the amount of drug by dividing the drug concentration by body weight and scale the time by dividing by $BW^{-1/4}$).

In these types of studies, the primary aim of this research has been to identify the exponential coefficient in the well-known scaling relationship,

$$Y = a \cdot BW^{b}$$

where Y is the parameter of interest and b is the power of body weight (BW) to which Y can be related by a proportionality constant (a). (The equation is described in more detail in Appendix A.) Data collected in multiple animal species exist primarily for pharmaceuticals, as a result of FDA requirements. For the most part, use of allometry in the pharmaceutical field has focused on establishing a scaling method to predict serum levels of drugs; i.e., scaling for pharmacokinetic

differences. In general, research in the pharmaceutical field has indicated the effectiveness of interspecies scaling to some power of body weight with the ¾ power being a reasonable choice for a default value (Mahmood, 1999).

The scaling of the myriad physiological processes that underlie the absorption, distribution, metabolism and elimination (ADME) of toxicants and their associated health effects can be drawn together into a single scheme by referring to the concept of physiological time (Dedrick et al., 1970; Dedrick, 1973; Boxenbaum, 1982, 1983, 1984, 1986; Lindstedt and Calder, 1976, 1981; Mordenti, 1986; Lindstedt, 1987; Travis, 1990). This concept proposes that quantitative differences across mammalian species in physiological processes can be seen largely as the consequence of fundamentally similar anatomical and biochemical machinery operating at different rates in differently sized species, smaller species having faster physiological "clocks." A mouse, for example, is carrying out approximately the same set of physiological processes as a human, but each process proceeds at a rate that tends to be approximately 7-times faster. In theory, the various processes stay in proportion to one another, but all of them are relatively accelerated in smaller species. (This obviously excludes cases where metabolic pathways are not comparable between species). If one scales the natural units of time (e.g., minutes) by dividing them by the fourth root of body mass (i.e., minutes × $BW^{-1/4}$), thus adjusting for the physiological time scale, then the time-course of physiological processes becomes comparable across species. This is the basis of the success of Dedrick plots, mentioned earlier. If we could measure time by some internal physiological standard (e.g., heartbeats) rather than in minutes, then the rates of physiological processes, the time course of disposition of a dose, and even overall lifespan would be roughly equal among species.

Owing to their importance, it is worthwhile to examine the starting assumptions that form the basis of the allometric, "physiological time" concept and its predictions. They are: (a) volumes and capacities (organ sizes, blood volumes) retain proportionality to body weight (BW); (b) the absolute rates of physiological processes are proportional to $BW^{3/4}$, these rates include cardiac output, minute volume, glomerular filtration, and the rates of specific metabolic steps; (c) physicochemical and thermodynamic properties of compounds (solubilities in various tissues) are equal in all species; and (d) for metabolic pathways with saturable metabolism, the Michaelis constant (the substrate concentration at which half the maximum reaction velocity is achieved) is invariant, while the maximum velocity scales as $BW^{3/4}$. A corollary to points (a) and (b) is that when rates are figured relative to body size (or to a volume, or in terms of concentration rather

than absolute amount), they scale as $BW^{3/4}/BW=BW^{-1/4}$. For example, the mass of chemical cleared per liter of blood per hour is affected not only by species differences in metabolic rate (which scale by the ¾ power of body weight) but also by species differences in blood volume (which scale directly with body weight).

The concept of physiological time and how it can be used for cross-species scaling can be illustrated by a simple example—one involving repeated administration of an oral dose of a compound to a mouse and a human, and its subsequent blood concentration as it is removed from a single body compartment. If doses are scaled to body weight (mg/kg/day), the blood level takes much longer to reach steady state in the human, owing to slower processing of the compound (Figure 1) and the steady-state that is eventually reached has a much higher blood concentration in the human than in the mouse. The human has a blood volume which is 2000 times that of the mouse (proportional to $BW^{1.0}$), but the chemical must be cleared from this volume by processes (i.e., metabolism and excretion) that operate only 300 times faster (i.e., proportional to $BW^{3/4}$) or seven-fold slower per unit blood volume. As a result, the human blood levels build to a higher level and the area under the blood-concentration curve (or AUC) is seven-fold higher. The AUC has units of [concentration]×[time], e.g., [mg/L] × min.

There are two kinds of scaling one could use to correct doses for this species difference in pharmacokinetic behavior. Rather than give a dose based on the full body weight difference between the two species (mg/kg, Figure 1A), one could give a smaller initial dose of chemical to the human, one that is approximately seven times less than the mouse-dose but equal in terms of $mg/kg^{3/4}$ (Figure 1B). The initial concentration is lower, but this is balanced by the slower removal of the chemical from the body. Over repeated daily dosing, the lesser daily uptake in the human balances the lesser daily removal, and a similar average plasma or tissue concentration is achieved in both species. Alternatively, one could give the same mg/kg doses, but administer successive doses by equal intervals of physiological time (Figure 1C). This second approach can also be achieved by expressing the abscissa or horizontal axis in units of physiological time, by multiplying by $BW^{-¼}$ (Figure 1D). This last figure is similar to those developed by Dedrick et al. (1970) to demonstrate that scaling time in this fashion led to congruity of methotrexate pharmacokinetics among several species. The concentration-time relationships for mice and humans would be identical on such graphs, falling to the same concentration after the same amount of physiological time has elapsed.

It can be shown that the two approaches for scaling, either shrinking doses or converting the time scale to physiological time, are equivalent ways of dealing with cross-species scaling differences in average tissue concentrations as long as saturable pharmacokinetic processes do not figure prominently (O'Flaherty, 1989). (Spacing doses equally in physiological time has the added advantage, however, that the amplitude of blood-level variation in the "sawtooth" curve is conserved across species.) Although the example described is quite simple, the sample principles would apply to more complex situations, those involving multiple compartments, multiple routes of uptake, metabolism and/or elimination.

Thus far we have only discussed the relevance of scaling *via* physiological time to the scaling for pharmacokinetic differences among species; as noted previously, species differences in pharmacodynamics are a significant aspect of the scaling problem that cannot be ignored. Unfortunately, data pertaining to how pharmacodynamic responses scale across species are sorely lacking. It would seem, however, that on a larger scale, the arguments related above for applying physiological time to pharmacokinetics would also prove useful in examining pharmacodynamics. Pharmacokinetic components of toxicity are more readily measured than pharmacodynamic components (*via* biomonitoring of metabolites, use of *in vitro* metabolic systems, *etc*.) The allometric approach suggests, however, that species difference in physiological time should apply equally to pharmacokinetic and pharmacodynamic processes since pharmacodynamic response is also underlain by physiologic distributions of catalytic and synthetic processes that should scale with physiological time.

Most of the assumptions that support scaling by physiological time are well supported by data on comparative anatomy and physiology. The chief uncertainty of the allometric approach lies in the scaling of rates of specific metabolic transformation reactions according to $BW^{3/4}$. The data to support this assumption are rather limited and some individual metabolic enzyme activities have been shown to very considerably across species with no clear relationship to body weight (Gillette, 1987; Calabrese 1986a,b). O'Flaherty (1989) noted that adjusting animal data *via* scaling for physiological time is generally either realistic or conservative for humans, except when first-order elimination and capacity-limited formation of a toxicologically active metabolite occur simultaneously. In this case, she suggested that scaling for physiological time may be either over- or underpredictive of the human dose, depending on the relative dependence of the efficiency of metabolite formation and elimination on body weight.

Such concerns should not preclude scaling by $BW^{3/4}$ as a default scaling factor, one that can be modified in light of available information. There are data that support the proposition of $BW^{3/4}$ scaling in specific cases (*e.g.*, Reitz *et al*., 1988). Furthermore, overall metabolic rates clearly scale with $BW^{3/4}$, this is the premise upon which the field of allometry developed. Since intake of the materials necessary for basal metabolism *(e.g.*, oxygen, oxidizable organic molecules, water) tend to be in proportion to $BW^{3/4}$ across species, calling exposure to environmental media equivalent on a ppm basis (*i.e.*, when they are equally contaminated) produces essentially the same expectation of pharmacokinetic equivalence as scaling by $BW^{3/4}$ (Hattis, 1991). In fact, all variables containing [time] in their units will scale in a way that leads to the human value being bigger by the ratio of body weights to the ¼ power. If these variables are re-expressed in terms of "physiological time units," *i.e.*, $[time]*BW^{-1/4}$, then their values are equal across species. This is illustrated in Figures 2A and 2B. In Figure 2A several physiological and metabolic parameters for the mouse and human are graphed as reported in the scientific literature (*i.e.*, according to chronological time). Clear differences exist. When these same values are scaled according to $[time]*BW^{-1/4}$, the apparent interspecies differences are markedly reconciled[4]. The exception is lifespan where it is known that humans have a lifespan which is somewhat out of proportion to general scaling principles. Humans are an outlier in the general mammalian relationship of lifespan to $BW^{1/4}$, living longer than expected. Some authors have addressed this by including brain weight as an additional factor in the allometric scaling equation (Boxenbaum, 1986).

Consideration of these points leads to the view that $BW^{3/4}$ scaling of the rates of individual metabolic transformation reactions can be viewed as a benchmark around which different species vary from instance to instance. Such variation does not invalidate the general scaling argument, nor does it provide evidence for any different scaling factor. Rather, the variation simply illustrates that any single conception of cross-species scaling can accommodate only the general trends, not the diversity of particular instances. Clearly, when data on metabolic conversion are available in a particular case, they should be used to modify the $BW^{3/4}$ default. In fact, instances of chemical-, dose-, and species-specific variation in metabolic transformation of a chemical may constitute the principal reason for deviation from the allometric default assumptions herein laid out.

[4] Note that for parameters with time in the denominator (*e.g.*, respiration rate, clearance), the value is divided by $BW^{-1/4}$ whereas for parameters where time is the numerator (*e.g.*, half-life), the value is multiplied by $BW^{-1/4}$. Note also that for parameters which contain a volume component (*e.g.*, GFR, cardiac output), the volume itself needs to be scaled to bodyweight.

The use of case-specific information should not, however, be used in a vacuum. It is important to bear in mind that cross-species differences in toxicological potency typically depend not on a single pathway or reaction but on at least several, interrelated processes (*e.g.*, rates of elimination, metabolic activation, DNA or protein damage and repair, compensatory cell proliferation, *etc.*). One must remember that the many quantitative differences across species exert their influences simultaneously and often in interactive ways. It is the interactions and net results of these processes that determine the effects of dose.

6. Use of Chemical-Specific Data to Modify Default Factors

The default approach attempts to account for the very general influence of the organism's scale on the toxic effects of an exposure, presuming that the underlying components are qualitatively the same, and quantitatively different in a way that can be adjusted by accounting for the overarching scale difference between experimental animals and humans in body size and physiological pace. As such, the default approach represents a kind of central tendency around which the actual equitoxic doses will vary from one compound to the next, the variation being attributable to chemical-specific factors that differ between animals and humans over and above those accounted for by the generalized normalization. In part, these deviations result from imprecision in the scaling correction, but the main factor is that the underlying processes have idiosyncratic aspects that cause deviation from the overall scale correction. Some of these are pharmacokinetic, primarily the differences in the activity of the relevant biotransformation pathways that toxify and/or detoxify the compound in question. Others are pharmacodynamic, such as differences in the rates of regenerative or repair processes that undo the specific physiological damage wrought by the toxic agent. To the degree that the particular processes relevant to particular compounds in particular species differ in their impact from that encompassed by the default scaling factor, the actual toxicologically equivalent dose will differ as well, and the default extrapolation may under- or overpredict the toxicity in humans.

If the species-specific differences are qualitative and major (*e.g.*, as in the case of α_2-immunoglobulin and male rat kidney cancers), this may cast doubt on the use of the particular animal model as a surrogate for humans, since no amount of quantitative correction will address

the lack of correspondence of underlying biology. If, however, the differences are ones of degree, then it should be possible (with sufficient data on the relevant quantitative differences) to modify the animal-to-human extrapolation in a chemical-specific way. This should not be regarded as *replacing* the default scaling factor (since the general impact of size and physiological pace is still relevant) but rather as *refining* it to include, in addition, the specific quantitative peculiarities in the particular relevant underlying biological processes as they are manifest for the species and chemical in question.

Since these chemical-specific considerations lead to variation around the central tendency expressed by the default scaling, they should produce a distribution of outcomes over chemicals on either side of the default's prediction. If one has little chemical-specific information about an agent, then this distribution represents uncertainty about the performance of the default scaling approach when applied to a particular chemical. For instance, Rhomberg and Wolff (1998) found a lognormal distribution over chemicals of the ratio of LD_{50} values for a given pair of species. The magnitude of this variation represents a description of the uncertainty about the equitoxic dose for any particular chemical. One can presume that the chemical of interest is a member of the distribution of differing cross-species equivalencies but, not knowing which equivalence applies, the distribution over all chemicals gives a measure of one's uncertainty about extrapolating for any one chemical. This approach is used to provide the empirical underpinnings to uncertainty factors (Dourson and Stara, 1983; Dourson, 1996; Dourson *et al*., 1996), and it can also be used to take a distributional approach to extrapolation (Baird *et al.* 1996, Evans *et al.* 2001).

When one has additional information about the magnitude of a relevant biological factor in a particular case, one can narrow this uncertainty by taking it into account. In essence, this corresponds to finding out where in the distribution of variations a particular chemical lies. For instance, of all the variations on species differences in pharmacokinetics that lead to variations in toxicologically equivalent doses, if one determines a compound's particular reasons for deviating from the overall default scaling, then the extrapolation can account not only for the general effects of size and physiological pace (embodied in the default) but also for the particular deviation from the average pattern. This simultaneously modifies the compound-specific extrapolation and reduces its uncertainty, since the realm of all possible pharmacokinetic differences is replaced by the specific ones that apply in the case at hand.

Over many cases, the accumulated set of chemical-specific adjustment factors should trace out the distribution of uncertainty that applies when no compound-specific information is available. The various case-by-case deviations with their specific causes should describe the uncertainty in the extrapolation when one does not know the specifics. If, for instance, as we accumulate experience with chemical-specific adjustment factors we find that adjusted human risk projections are systematically lower than by the default scaling procedure, then this is evidence that the default scaling (or the logic used in chemical-specific adjustment) is somehow flawed. For instance, the finding by Baird *et al*. (1996) that mg/kg/day scaling led to larger species being systematically more sensitive than smaller species is evidence that simple body-weight scaling of doses fails adequately to adjust for the effects of species body size. Just as residuals around a good linear-regression line should be unbiased and without trends, the distribution of chemical-specific adjustments should be symmetric about the default factor. One has confidence in a default scaling factor that produces unbiased chemical-specific residuals and one should question a scaling approach that leaves unexplained systematic variation among the residuals.

The question, then, is how to apply chemical-specific information. A full discussion is beyond the scope of this paper, but a few of the elements can be reviewed. First, one must decide when the experimental data are of sufficient quality that they constitute a sufficiently reliable characterization of the chemical-specific effect. The chemical-specific data may be considered of acceptable quality when the anticipated error associated with the experimental design (and/or with combining data collected using different experimental designs) is less than the anticipated error associated with the use of generic relationships (Rhomberg, 2000).

An additional question is to address what "portion" of a dose adjustment is obviated by chemical-specific data and what remains. When both scaling and uncertainty are accommodated by a single "uncertainty factor" (say, a 10-fold factor) then the practice has been to reduce the overall factor when chemical-specific information is used, as a way of acknowledging that uncertainty has been reduced (and that, to some ill-defined extent, the "scaling" allowance built into the factor has been partially obviated). The problem with this approach is that it conflates uncertainty with scaling adjustment, and it leaves unaddressed the question of how scaling considerations contributed to the original factor (and thus how a chemical-specific alternative should change it). For example, if one does pharmacokinetic modeling to estimate "internal doses" as a way of replacing default scaling of external doses with scientifically supported and

chemical-specific information, one needs to ask what "internal doses" would be expected to yield equivalent toxicity in different species. Depending on how pharmacodynamic differences across differently sized species play out, there is not an obvious answer, and one has simply replaced a question about equivalent external doses with one about equivalent internal doses.

One pitfall is to assume that a quantitative adjustment of the part serves as a quantitative adjustment for the whole, for instance, assuming that described species differences in the rate of a certain biotransformation step translate directly into differences in the amount of a chemical that is metabolized. In fact, the outcomes depend on the interplay of a number of different rates (absorption, competing metabolism, clearance of the parent compound) that all are expected to differ across species. The impact of quantitative differences in one factor alone may be difficult to judge unless the whole pharmacokinetic context is specified. A second pitfall is to assume that any factor not explicitly specified in the default scaling has been assumed to have equal values across species, and that any difference that is discovered thereby changes the extrapolation accordingly. For example, discovering that humans metabolize and toxify a certain compound more slowly than rats does not necessarily lead to a lower chemical-specific adjustment factor for that compound, since the allometric considerations already incorporated in the default scaling allow for this effect. The real questions are (1) whether the compound-specific metabolic differences that are discovered are greater than or less than the differences already implicitly assumed and incorporated into the default factor, and (2) how variation in that specific metabolic activity translates into variation in biologically effective dose, in view of the simultaneous influence of other factors. In order to sort out what to make of chemical-specific data, one needs a clear articulation of what was implicitly being assumed about the investigated factors under the default.

The default approach proposed herein, scaling daily intakes by the ¾-power of body weight, corresponds to the assumption that all relevant rate processes have this same relation to body weight. One of the chief benefits of physiological time as a unifying general principle for scaling is that it makes these implicit assumptions about the contribution of individual factors explicit. The way to gauge the impact of a rate that does not scale in this way (as discovered by chemical-specific experiments) is to combine this discovered effect with a maintenance of the allometric scaling assumption for all factors that have no chemical-specific data.

A proposed approach for how to consider the use of chemical-specific data on cross-species scaling is summarized in Figure 3.

7. Route-Specific Issues for Cross-Species Extrapolation

Cross-species extrapolation for chemicals posing inhalation or dermal risks represent special cases. Species-specific differences in the gastrointestinal tract may influence absorption and, therefore, toxicity , but the differences would seem to be compound-specific matters of degree (*i.e.*, 50% absorption *vs*. 70% absorption) and constitute the kind of individual-case variation that is not easily captured in a general-purpose scaling method. Given sufficient time in contact with the absorptive surface, comparable levels of uptake and toxic effects are possible. For inhalation and dermal exposures, other factors may be equally or more significant than species differences in absorption at the relevant membrane. For example, differences in respiratory deposition and absorption are likely to be quite significant and may result in different patterns of toxicity depending on cross-species differences in respiratory tract anatomy and physiology. Species-specific differences in response to irritant properties of chemicals can also result in significant differences in absorption or deposition patterns which may not scale well in a proportional fashion. The nasal passages of mice and rats are convoluted and much more efficient at scrubbing respiratory contaminants than those of humans. Rodents are also obligate nasal breathers, unlike humans who, at periods of peak respiration (and maximal exposure), breath through the mouth, which limits the effectiveness of scrubbing inhaled air (Dahl, 1991). Opposite to the more effective scrubbing in the nasopharyngeal section of the rodent airway, chemicals may be more effectively deposited in the tracheobronchial region of the human airway relative to the rodent, because the larger size of the airway in humans results in more turbulent flow. Similarly, regional skin thickness or body hair pattern may substantially affect dermal absorption and may not vary among species according to allometric principles.

The need to adjust for differences in respiratory physiology between rodents and humans has been well studied (Jarabek *et al*., 1989; Beliles and Parker, 1989; Fredrick *et al*., 2001; Fredrick *et al*., 2002; Pauluhn, 2003). Jarabek *et al*. (1989) proposed the use of dosimetry modeling to improve cross-species extrapolation for inhaled risks of soluble particles. The rat-to-human ratio of deposited dose varied by respiratory region and particle size and could not be described based on scaling to some power of body weight. Jarabek (1995) subsequently

expanded on this work and described a basis for using mechanistic information for refining cross-species extrapolations for establishing reference concentrations (RfCs) for both particulate and gaseous inhaled toxicants. Jarabek defined a process whereby specific characteristics (*e.g.*, water solubility) are used to classify the inhaled material into specific categories. Category specific properties as well as species-specific differences in respiratory physiology (*e.g.*, relative sizes of each portion of the respiratory tract) were mathematically combined to yield a dosimetric adjustment factor (DAF). An important distinction made in this paper was between the default *scaling* factors used to derive equivalent doses between animals and humans and default *uncertainty* factors used to account for lack of knowledge about the appropriateness of the scaling factor used. When the derived DAF values were used to convert rodent to human inhaled doses, Jarabek recommended that the inter-species uncertainty factor be set at 3, reflecting the fact that pharmacokinetic but not pharmacodynamic differences across species had been accounted for. If more specific information on the pharmacokinetics of the specific material (*e.g.*, chemical specific dosimetry data) or pharmacodynamic data (*e.g.*, correlations between tissue dose and adduct levels) are available, then the UF could be further reduced. Bogdanffy and Jarabek described how mechanistic data for one compound, vinyl acetate, could be used to supplant default values in the dose-response assessment (Bogdanffy and Jarabek, 1995). This was subsequently done using nasal extraction efficiency studies and PBPK modeling based on analytically determined rates of enzymatic reactions in the nasal mucosa (Bogdanffy *et al*., 1999). This work predicted that the LED_{10} concentration of 8.7 ppm in the rat would be toxicologically equivalent to a LED_{10} concentration of 10 ppm in humans engaged in average levels of activity. Considerable research has also been conducted on incorporating pharmacodynamic data into inhalation dosimetry models for formaldehyde (Casanova *et al*., 1991; Conolly and Andersen, 1993; Conolly *et al*., 2000, 2002).

Detailed modeling of interspecies differences in toxicity from dermal exposures to toxicants has not received even the limited attention given to inhaled toxicants, presumably because dermal exposures are rarely risk drivers in EPA assessments. Nonetheless, the unique features of dermal exposures noted above suggest that more attention should be given to this pathway for deriving an appropriate method of dose scaling.

8. Acute Versus Chronic Toxicity

As noted previously, cross species comparisons of measures of toxicity suggest that there may be different patterns of cross-species response to chemical agents depending on the duration of the exposure. It is a well accepted principle that a single bolus dose (*i.e.*, acute administration) results in a lower safe dose rate than if the dose is spread out over a longer period of time and, in general, the more spread out a dose, the lower the likelihood of an effect due to accumulated damage. In reality, the terms "acute" and "chronic" as used in toxicology and risk assessment are somewhat arbitrary categorization terms. Perhaps a better categorization would be to distinguish single from multiple exposures. The basis for using separate terms to describe single and repeated exposures results from the fact that they often exhibit very different sequellae, which may result from quite different patterns of insult and response. Acute/single exposures exert their toxicity by an immediate and intolerable level of damage to some critical biological pathway. The rate at which the body repairs or compensates for the effects of an acutely toxic dose is not generally relevant because it is the basal level of defense capacity that is rapidly overwhelmed by the single large onslaught before recovery processes can really begin. In contrast, for chronic/repeated exposures, the limiting factor may be the ability of the body to marshal an ongoing response to cope with each dose's incremental effect, and so the *rates* at which repair or compensatory processes proceed is key for whether a given dose *rate* leads to a toxicological effect. This was previously referred to in the discussion of the results of Rhomberg and Wolff (1998).

In general there are three ways in which repeated exposure to an agent may cause toxicity even when the repeated dose rate is lower than the level eliciting toxicity as a single dose:

1. *Chemical accumulation*. In this process chemicals (or their metabolites) may accumulate over time with ongoing exposure, building up slowly even if the individual doses pose no harm. Eventually, the level of chemical will reach a point at which adverse effects will ensue. The use of warfarin as a rodenticide is a good example of this process of chronic toxicity. Individual doses of the warfarin do not manifest toxicity; it is only when a particular threshold is reached after repeated exposures that difficulties in clotting lead to a lethal effect. Many highly halogenated organic compounds (PCBs, PBBs) may follow an essentially similar pattern where low levels are apparently tolerated but accumulated body burdens lead

to adverse symptoms. In this setting, clearance of a compound or a long-lived metabolite is the limiting factor, and dose-rates that are balance by sufficiently rapid clearance can be tolerated since excessive body burden buildup is avoided. Toxicity in this situation is amenable to cross-species scaling of toxicant clearance, which generally scales as $BW^{3/4}$ (although specific chemicals may vary).

2. *Damage accumulation*. Particularly relevant for rapidly eliminated chemicals, this process requires that damage resulting from exposure accumulate over time even if levels of the responsible chemical do not. It requires that (1) the target tissue have a functional reserve capacity, and (2) the rate of repair be slow relative to the rate of damage, such that functionality is not completely restored between insults. Eventually, the level of damage exceeds the functional reserve capacity of the tissue and toxic symptoms become manifest. The effect of ethanol upon the liver is a classic example of this form of chronic toxicity. While ethanol is rapidly eliminated and even large doses do not bioaccumulate, liver damage from excessive ethanol exposure does accumulate, eventually leading to liver impairment. In this situation, cross-species scaling would account for pharmacodynamic rather than pharmacokinetic differences. It is the rate of repair or recovery that is limiting. As previously noted, there are minimal data on how pharmacodynamic processes scale among species, but the general scaling of physiological rates (including synthetic processes that presumably underlie repair) to $BW^{3/4}$ provides some rationale for using this approach in this case.

3. *Accumulated risk*. This process involves a slightly different conceptualization of toxicity, as a stochastic rather than deterministic process. From the viewpoint of accumulated risk, each individual exposure results in an incremental increase in the risk of a deleterious outcome, with the overall risk accumulating over time. This model of chronic toxicity requires the assumption that a key event that may or may not happen is sufficient to precipitate a chain of events leading to a deleterious outcome. While one is exposed, the chance per unit time of such a key event being precipitated is increased. As the exposure period lengthens, the number of chances for the key event to occur accumulates, and hence the lifetime risk of the adverse outcome increases with exposure duration. This view is most consistent with how we currently view genotoxic carcinogens or teratogenic agents but it also applies to heart attacks, strokes, and in general any endpoint in which there is a low functional redundancy and the failure of a single element can precipitate an adverse effect. In this final situation, scaling would be according to lifetime (as well as depending on how the event probability depends

on the balance of physiological processes). The longer lifetime (scaling approximately as $BW^{3/4}$ with the acknowledgement that humans live longer than a typical mammal of their size) over which risk accumulates means that the daily increment must be less.

It should be pointed out that one chemical may exhibit toxicity consistent with more than one of the three processes outlined above. For example, the ability of ethanol to increase the risk of developing certain cancers is consistent with the accumulated risk hypothesis.

The role of recovery in distinguishing between acute and chronic exposures cannot be underemphasized. If recovery of functional capacity is complete between bouts of individual exposures, then ongoing exposure is essentially a series of acute exposures. In such a situation, the safe acute dose will be equal to the safe chronic dose and the tolerable dose is the one that just balances the capacity to undo the effects of each day's new exposure.

9. Dose Scaling for Less Than Chronic Exposures

Traditionally, regulatory agencies have been focused on evaluating the health risks of chronic exposures to chemicals. In part, this has been because levels of exposure that are safe for a lifetime would be expected to be safe for a portion of a lifetime. At least some people will be exposed to environmental chemicals for long periods, and regulatory standards are generally set to ensure that the safety of these maximally exposed individuals rather than the individual receiving the average level of exposure. This practice also reflects the fact that carcinogens, and the interest in assessing lifetime cancer risk, have typically overshadowed non-cancer health risk assessment. With increasing emphasis being placed on specific age groups such as children and the elderly, who may have different susceptibilities to chemicals, concern has increasingly been focused on evaluating the effects of less than chronic exposures to chemicals at particular life stages. Moreover, accidents and terrorism pose threats of short-term high exposures. Thus, subchronic and acute toxicity benchmarks (RfDs, RfCs) are becoming increasingly important in risk assessment. One way such values have been derived in the past is simply to "undo" the effects of the subchronic-to-chronic safety factor and increase the chronic RfD by a factor of 3 or 10. However, because chronic RfDs are typically determined using healthy adult animals (except for the case where developmental endpoints are used), such an approach may not be appropriate for deriving RfDs where sensitive subpopulations are a concern.

Exposures that are intermediate between acute and chronic could probably be tolerated at a somewhat higher dose than chronic exposures, although no dose should exceed the acutely toxic level. The degree of this effect depends on the specifics of the time courses of kinetics and damage-repair (dynamics) for the chemical and on the particulars of the exposure scenarios involved. Case-specific and exposure scenario-specific calculations could be made if PBPK and BBDR models are available, but as a default, the chronic dose rate provides a conservative standard.

Guidance for using RfDs in assessing particular exposure scenarios should include a case-specific assessment of the likely recovery period – if exposure episodes are spaced more widely than this period, then the exposures can be thought of as acute exposures (the chemical-specific expected recovery period is the best criterion for discriminating between acute and chronic exposures rather than a definition based on days). On the other hand, exposures that are continuous or repeated often within the timespan needed for recovery are likely to have accumulating effects and so should be assessed as chronic – *i.e.*, as a dose rate that should not be exceeded because to do so would lead to a long term imbalance of damage and repair.

It should be remembered that tissue damage (*e.g.*, DNA hits, lipid peroxidation, oxidation of reactive sites on enzymes) is only significant if it cannot be repaired in a manner that restores functional capacity to what it was prior to the damage occurring. Exposures to chemicals arising from endogenous or naturally occurring sources (*e.g.*, metabolically generated reactive oxygen species, chemicals produced during the cooking of foods, hormonally active substances produced by plants) result in a substantial level of potentially life-threatening damage to the body. Yet virtually all of this damage is repaired or otherwise addressed before it can lead to significant harm. The body generally has a great ability to intervene between exposure to a potentially harmful substance and the occurrence of significant damage. "Living" is in reality the successful achievement of the balance between the rate of damage and the rate of repair. It is only when exogenous agents perturb the balance, through increasing damage, decreasing repair, or inducing an excessive repair response, that a significant deleterious outcome occurs.

10. Considerations for Extrapolations Related to Childhood and Early Life Stages

As noted previously, increasing attention is being focused on early life stages as windows of susceptibility for chemical exposure (Goldman 1995, 1998; Landrigan, 1999; Goldman and Koduru, 2000; Miller *et al.*, 2002). Animal studies addressing the developmental effects of chemical exposures have typically investigated the period prior to and around the time of birth. Less attention has been paid to the effects of chemical exposures during the time period which lies between infancy and adulthood. In part this gap is attributable to the idea that any increased risks during childhood are primarily related to increased exposure; children receive a higher dose than adults based on the combination of their typically higher contact rate (soil ingestion, inhalation, water ingestion) and their substantially smaller body weight. These differences in exposure and dosimetry have been fairly well studied and are readily adjusted for in the exposure assessment portion of risk assessment.

Issues related to the toxicological susceptibility of early life stages have not been well addressed until relatively recently. It is frequently stated that "children are not small adults". In some cases it may not be sufficient to scale the child's exposure for their higher intake relative to body weight and then assume the physiological response of the child is proportional to the adult's. The basic biochemistry of children may be distinctly different from the adults and must be considered on a chemical-by-chemical basis. Examples of physiological differences between children and adults that may be significant in terms of toxicological response and which may not scale proportionally or continuously with body weight include the following:

- Respiration rate,
- Glomerular filtration rate,
- Active gastrointestinal absorption of nutrients,
- Composition and activity of intestinal flora
- Percentages of body fat and body water,
- Levels of CYP 450 isoforms and other phase I enzymes,
- Glucuronic acid conjugating ability,
- Biliary excretion ability, and

- Rates and patterns of growth in particular organs (bones, brain, immune system, *etc.*)

These and other differences in rates of absorption, distribution, metabolism, and elimination (ADME) have all been well documented (O'Flaherty, 1994; Bruckner , 2000; Renwick *et al*., 2000; Pelekis *et al*., 2001; Miller *et al*. 2002; Clewell *et al*., 2002b; Ginsburg *et al*., 2002, 2004; Hattis *et al*., 2003; Gentry *et al*., 2003; Sarangapani *et al*., 2003).

As indicated above, children have lower levels of certain cytochrome P450s than adults, which may translate into slower elimination of some compounds or less bioactivation of others (Clewell *et al*., 2002b). Children also differ from adults in terms of their expression of Phase II enzymes (*e.g.*, less glucuronic acid conjugation) (Miller *et al*., 1976). However, it would be premature to invoke necessarily greater sensitivity in children due to these factors, because excess capacity in other pathways may be sufficient to handle environmentally relevant doses. In this area, PBPK modeling would be useful to evaluate what specific adult-child differences in metabolism may mean for doses of toxicants at the target site.

In addition to recognizing that children are not small adults, it is also important to realize that all children are not equivalent.[5] Hattis and colleagues (Hattis *et al*., 2003; Ginsburg *et al*. 2002, 2004) analyzed rates of drug metabolism in adults and children of different age groups (*i.e.*, premature neonates, full term neonates, ages 1 week-2 months, 2 to 6 months, 6 to 24 months, 2-12 years, and 12-18 years). Their database was composed of pharmacokinetic data for 44 chemicals, most of which were pharmaceuticals but which also contained the environmentally-relevant chemicals chloral hydrate, dichloroacetic acid, and trichlorethanol. They found that half-lives in infants were 3 to 9 times longer than those in adults and that in the 6 to 24 month age group, drug half-life was often slightly shorter than in adults. By the teenage years, the difference with adult half-lives had largely disappeared. Consistent findings were reported by Renwick and co-workers (Renwick, 1998; Renwick *et al*., 2000; Dorne *et al*., 2001). Ginsberg *et al*. (2002) noted that for children two years of age and beyond, scaling according to $BW^{3/4}$ yields fairly good predictions of clearance. Thus, while the relationship between age and drug clearance is complex and cannot, as a whole, be described by a linear relationship with body weight, there is some

[5] It should be recognized that the concept of "the child" as used in risk assessment represents a somewhat arbitrary, if necessary, construct and imposes a discrete categorization on a population that is actually characterized by continuously varying and evolving exposure and toxicological variables, some of which become "adult-like" early in life and others which do so only gradually during maturation.

evidence that body weight scaling is a reasonable approach with children older than age two. This in turn suggests that it is important to carefully consider age-specific patterns of exposure along with possible age-specific differences in metabolism. Body weight scaling might be appropriate when exposure stems from a scenario involving only older children (*e.g.* trespasser or recreational land use scenarios) but would not be appropriate when exposure to infants is a real possibility (*e.g.*, when exposure to chemicals present in the home is of concern).

While pharmacokinetic differences between children and adults have been clearly documented, pharmacodynamic differences have not been well quantified. Qualitative examples exist however. One important example is the process of brain cell proliferation and differentiation which occurs in children but is apparently absent in adults. Similarly, the developing immune system and the growing bones represent cell populations which, due to a higher level of proliferative activity, may be at increased risk of damage compared to the more static populations one finds in adults. To date, however, no investigators have attempted to develop a comprehensive relationship which quantitatively describes these pharmacodynamic differences or combines them with differences in pharmacokinetics.

11. Conclusions

This paper aims at proposing a very broad generalization about chemical exposures that can be considered of equal risk in experimental animals and humans – one that can be applied as a default for both cancer and non-cancer health endpoints when adequate information is lacking regarding either chemical-specific pharmacokinetics or pharmacodynamics. It attempts to provide a rationale for a default method for extrapolating doses in animals studies to equivalent levels in humans, consistent with the available empirical data regarding the scaling of chemical toxicological potencies between animals and humans and with the known patterns of variation in body size and rates of physiological processes.

To arrive at a generally applicable default, it is necessary to rely on simplified, broad patterns and trends of biological variation, while paying less attention to the details of specific, individual cases. These details are not unimportant. Rather, the intent is to provide a framework for the incorporation of specific data where available, allowing (and indeed, encouraging) users to

go beyond the default assessment based on overall trends to address the impact of specific knowledge about the chemical and its actions.

The empirical data on toxicological potencies estimated in various animal species and in humans demonstrate the large variability involved. Although scaling doses by $BW^{3/4}$, as proposed herein, characterizes the trend fairly well, individual chemicals may deviate from this overall pattern. In the case of the allometric arguments, there are dozens of points in the chain of inference where one could raise counterexamples to simplifying assumptions, arguing that the $BW^{3/4}$ approach would over- or underestimate human risks for that case.

The existence of such underlying variation means that the extrapolation of chemically induced risks observed in one circumstance (say, in mouse lifetime cancer bioassay) to another (say, to people exposed to environmental pollutants) needs to be carefully and properly interpreted. Clearly, the projection of an equivalent dose is not merely a conversion of units, with the resulting human dose achieving an equal factual standing to the original animal observation. The projection is an hypothesis, formulated in the face of uncertainty. In the most basic case – when there is little additional information that may be brought to bear – this hypothesis is framed in terms of the general features of anatomical and physiological scale differences among species that should affect all chemicals. It represents a best guess based on general principles and the recognition of overall trends. This best guess is surrounded by an envelope of considerable uncertainty, owing to the dozens of factors that make each chemical's disposition and toxic effects in various species unique. When applicable pharmacokinetic and pharmacodynamic insights into the particular chemical and its actions are available, they can (and should) be used to refine the projections by identifying and accounting for these chemical-specific factors.

Every projection of human equivalent dose, no matter how sophisticated, will have associated with it both uncertainty and variability. The uncertainty concerns whether the scaling method employed has correctly embodied and utilized the information at hand (be it general cross-species trends over all chemicals or case-specific insights from pharmacokinetics and pharmacodynamic studies). The variability arises because even a sophisticated projection, when applied to a population of cases, will at best predict the mean of an array of actual values that reflect the myriad individual factors that no analysis can completely take into account. The true dose of equivalent risk will vary among exposed humans according to how each individual

deviates from the overall human norm, owing to genetic factors, environmental influences, age, sex, lifestyle, and countless details of personal history.

The goal of cross-species scaling methodology is not to arrive at true values of equivalent doses under all circumstances. Rather it is to embody correctly and without bias the impact of the information at hand, providing rational estimates that take into account what is known, recognizing that true values will vary around this estimate as a result of case-by-case particulars, many of which are either unknown or known to vary among the individuals for whom the projection is being made.

The proposed scaling of daily administered oral doses of chemicals by $BW^{3/4}$ is intended to be such an unbiased projection; *i.e.*, it is to be thought of as a "best" estimate rather than one with some conservativism built in to assure that any error is on the side of being overly protective. It should not be confused with the uncertainty or safety factor normally used as part of cross-species extrapolation. It is to be expected that some individual chemicals will have their human potencies overestimated by this approach, while others will have them underestimated.

This having been said, it must be acknowledged that considerable uncertainty exists as to what exponential power in the allometric scaling equation yields the ultimately superior result. Empirical data suggest that $BW^{3/4}$ may fit data best in most cases of repeated exposure but there are also proponents of scaling by surface area ($BW^{2/3}$), which also leads to reasonable fits to the data. The use of scaling to $BW^{3/4}$ is proposed because it represents a consensus of the empirical data and because it can be related to an explicit rationale based on allometric variation of the underlying anatomy and physiology.

The utility of the concept of physiological time for supporting scaling across species lies in its simplicity and generality. Because organ volumes tend to share a common pattern of allometric variation, while rate of physiological processes share another, the general predictions of cross-species differences is independent of specific hypothesis about target organs or mechanisms of action. One could, for instance, envisage an alternative allometric formulation that, rather than relying on overall patterns for unspecified organs in all mammals, focuses instead on the details of specific organs (*e.g.*, the liver) in specific laboratory animals strains and in humans. The utility of such an approach for a default scaling factor would be doubtful, because the generality of the argument is lost, and the analysis becomes contingent on the details of the

physiological hypothesis being elaborated. If such specificity is possible in an individual instance, it should become part of the case-specific pharmacokinetic-pharmacodynamic analysis that overrides the default methodology.

The primary focus of this paper has been the extrapolation of oral doses across species. Cross-species extrapolation for exposure to toxicants *via* other routes of exposure (*i.e.*, inhalation, dermal) present a situation where the default methodology of scaling to $BW^{3/4}$ needs to be very carefully evaluated on a case-by-case basis. Differences in respiratory or dermal physiology among species may factor more significantly in toxicological response than differences in metabolism and biological processes that scale according to $BW^{3/4}$ and physiological time. The dosimetry adjustment factor (DAF) as described by Jarabek (1995) represents an excellent example of how various types of information pertinent to cross-species scaling can be incorporated in an assessment in lieu of the generic scaling factors. Although the DAF as described by Jarabek and others is intended to address inhalation exposures, an equivalent approach could be used to addressed other exposure routes (*e.g.*, the dermal route).

The allometric patterns relied on by the present argument represent variation among species for adult organisms. Allometric patterns among variously sized individuals of the same species differ from one species to another. Thus, the scaling approach presented here would not logically extend to scaling among differently sized humans. This is particularly true when attempting to scale between adults and infants or very young children, since the specific physiological differences identified in Section 11 may not necessarily scale well with body size (Ginsburg *et al.*, 2002). Body weight scaling for older children appears somewhat more reasonable although the data supporting this approach remain limited. As noted previously, metabolic differences between children and adults (*e.g.*, differences in metabolic pathways) are known to exist and chemical specific data could therefore be incorporated into dose-response assessments. However, pharmacodynamic differences have not been described to a similar degree; thus it would be difficult to completely characterize potential adult/child differences in toxicological susceptibility.

Finally it must be stated that scaling cannot solve high to low dose extrapolation issues which commonly confound the interpretation of the relevance of animal studies to humans. In cases where high doses administered to experimental animals overwhelm metabolic or repair pathways which would not be overwhelmed at lower, more environmentally relevant doses, no

form of a default scaling relationship will be able to compensate or accurately adjust for these factors. In such cases, other sources of data (*e.g.*, from pharmacokinetic studies) must be called upon to supplement the default approach. Such analyses can provide a valuable insight into differences in target tissue doses between rodents at high bioassay exposures and humans at much lower exposures.

In summary, we recommend that toxicological risks be presumed equal when daily amounts of chemicals administered orally are in proportion to body weight raised to the ¾ power. The empirical data on comparative toxicological potencies, though limited in scope, support the general practice of scaling animal potencies to humans (for mammals) and indicate that, on average, scaling to $BW^{3/4}$ performs reasonably well. Theoretical support for scaling toxic doses by the ¾ power of body weight is also available from analysis of the allometric variation of key physiological parameters across mammalian species. Such an analysis has the benefit of providing an articulated rationale for the scaling methodology and of setting out the underlying assumptions explicitly.

12. References

Adolph, E.F. 1949. Quantitative relations in the physiological constitution of mammals. *Science* 109:579-85.

Andersen, M.E., and K. Krishnan. 1994. Physiologically based pharmacokinetics and cancer risk assessment. *Environ. Health Perspect.* 102 (Suppl. 1):103-8.

Andersen, M.E., H. Clewell 3rd, and K. Krishnan. 1995. Tissue dosimetry, pharmacokinetic modeling, and interspecies scaling factors. *Risk Anal.* 15(4):533-7.

Baird, S.J.S., J.T. Cohen, J.D. Graham, A.I. Shlyakhter, and J.S. Evans. 1996. Noncancer risk assessment: A probabilistic alternative to current practice. *Hum. Ecol Risk Assess* 2 (1) :79-102.

Beliles, R.P., and J.C. Parker. 1989. Dose paradigms for inhaled vapors of primary carcinogens and their impact on risk assessment. *Health Phys.* 57 (Suppl. 1):333-40.

Bogdanffy, M.S., and A.M. Jarabek. 1995. Understanding mechanisms of inhaled toxicants: implications for replacing default factors with chemical-specific data. *Toxicol Lett*. 82-83:919-32

Bogdanffy, M.S., R. Sarangapani, D.R. Plowchalk, A. Jarabek, and M.E. Andersen. 1999. A biologically based risk assessment for vinyl acetate-induced cancer and noncancer inhalation toxicity. *Toxicol. Sci.* 51(1): 19-35.

Bogdanffy, M.S., G. Daston, E.M. Faustman, C.A. Kimmel, G.L. Kimmel, J. Seed, and V. Vu. 2001. Harmonization of cancer and noncancer risk assessment: proceedings of a consensus-building workshop. *Toxicol. Sci*. 61(1): 18-31.

Boxenbaum, H. 1982. Interspecies scaling, allometry, physiological time, and the ground plan of pharmacokinetics. *J. Pharmacol. Biopharm.* 10:201-27.

Boxenbaum, H. 1983. Evolutionary biology, animal behavior, fourth-dimensional space, and the raison d'etre of drug metabolism and pharmacokinetics. *Drug Metab. Rev*. 14:1057-97.

Boxenbaum, H., and R. Ronfeld. 1983. Interspecies pharmacokinetic scaling and the Dedrick plots. *Am J Physiol.* 245(6):R768-75

Boxenbaum, H. 1984. Interspecies pharmacokinetic scaling and the evolutionary-comparative paradigm. *Drug Metab. Rev*. 15:1071-121.

Boxenbaum, H. 1986. Time concepts in physics, biology, and pharmacokinetics. *J. Pharmacol. Sci.* 75:1053-62.

Bruckner, J.V. 2000. Differences in sensitivity of children and adults to chemical toxicity: the NAS panel report. *Regul. Toxicol. Pharmacol.* 31(3):280-5.

Butterworth, B.E., and M.S. Bogdanffy. 1999. A comprehensive approach for integration of toxicity and cancer risk assessments. *Regul. Toxicol. Pharmacol.* 29(1): 23-36.

Calabrese, E.J. 1983. *Principles of Animal Extrapolation*. John Wiley & Sons. New York.

Calabrese, E.J. 1986a. Animal extrapolation and the challenge of human heterogeneity. *J. Pharmacol. Sci*. 75:1041-6.

Calabrese, E.J. 1986b. Comparative biology of test species. *In*: Hill, T.A., R.C. Wands, and R.W. Leukroth, Jr. (eds.). 1986. *Biological Bases for Interspecies Extrapolation of Carcinogenicity Data.* Prepared for Food Safety and Applied Nutrition, Food and Drug Administration, Department of Health and Human Services, Washington, D.C. under contract to Federation of American Societies for Experimental Biology, FDA contract No 223-83-2020.

Calabrese, E.J. 1987. Extrapolation from animal data. *In*: Tardiff, R.G., and J.V. Rodricks (eds.). 1987. *Toxic Substances and Human Risk: Principles of Data Interpretation*. Plenum Press. New York.

Casanova, M., K.T. Morgan, W.H. Steinhagen, J.I. Everitt, J.A. Popp, and H.D. Heck. 1991. Covalent binding of inhaled formaldehyde to DNA in the respiratory tract of rhesus monkey: pharmacokinetics, rat-to-monkey interspecies scaling, and extrapolation to man. *Fundam. Appl. Toxicol.* 17(2):409-28.

Clewell, H.J., M.E. Andersen, and H.A. Barton. 2002a. A consistent approach for the application of pharmacokinetic modeling in cancer and noncancer risk assessment. *Environ. Health. Perspect.* 110(1):85-93.

Clewell, H.J., J. Teeguarden, T. McDonald, R. Sarangapani, G. Lawrence, T. Covington, R. Gentry, and A. Shipp. 2002b. Review and evaluation of the potential impact of age- and gender-specific pharmacokinetic differences on tissue dosimetry. *Crit. Rev. Toxicol*. 32(5):329-89.

Collins, J.M., C.K. Grieshaber and B.A. Chabner. 1986. Potential role for preclinical pharmacology in phase I clinical trials. *Cancer Treat. Rep*. 70:73-80.

Collins, J.M., D.S. Zaharko, R.L. Dedrick, and B.A. Chabner. 1990. Pharmacologically guided phase I clinical trials based upon preclinical drug development. *J. Natl. Cancer Inst.* 82:1321-6.

Conolly R.B., and M.E. Andersen. An approach to mechanism-based cancer risk assessment for formaldehyde. 1993. *Environ. Health Perspect.* 101 (6):169-76.

Conolly, R.B. 1995. Cancer and non-cancer risk assessment: not so different if you consider mechanisms. *Toxicology.* 102(1-2):179-88.

Conolly, R.B., P.D. Lilly, and J.S. Kimbell. 2000. Simulation modeling of the tissue disposition of formaldehyde to predict nasal DNA-protein cross-links in Fischer 344 rats, rhesus monkeys, and humans. *Environ. Health Perspect.* 108 (Suppl. 5): 919-24.

Conolly, R.B., J.S. Kimbell, D.B. Janszen, and F.J. Miller. 2002. Dose response for formaldehyde-induced cytotoxicity in the human respiratory tract. *Regul. Toxicol. Pharmacol.* 35(1):32-43.

Cox, L.A. Jr., and P.F. Ricci. 1992. Reassessing benzene cancer risks using internal doses. *Risk Anal*. 12(3): 401-10.

Crump, K.S., A Silvers, P.F. Ricci, and R. Wyzga. 1985. Interspecies comparison for carcinogenic potency to humans. In: Ricci, P.F. (ed.) *Principles of Health Risk Assessment.* Prentice-Hall, Englewood Cliffs, N.J.

Crump, K., B. Allen, and A. Shipp. 1989. Choice of dose measure for extrapolating carcinogenic risk from animals to humans: an empirical investigation of 23 chemicals. *Health Physics* 57 (Suppl. 1):387-93.

Dahl, A.R., J.D. Sun, L.S. Birnbaum, J.A. Bond, W.C. Griffith Jr., J.L. Mauderly, B.A. Muggenburg, P.J. Sabourin, and R.F. Henderson. 1991. *Toxicol. Appl. Pharmacol*. 110(1):9-19.

Davidson, I.W.F., J.C. Parker, and R.P. Beliles. 1986. Biological basis for extrapolation across mammalian species. *Reg. Toxicol. Pharmacol*. 6: 211-37.

Dedrick, R.O., K.B. Bischoff, and D.S. Zaharko. 1970. Interspecies correlation of plasma concentration history of methotrexate (NSC-740). *Cancer Chemother. Rep. Pt. 1*, 54:95-101.

Dedrick, R.L. 1973. Animal scale-up. *J. Pharma. Biopharm.* 1:435-61.

Dorne, J.L., K. Walton, and A.G. Renwick. 2001. Uncertainty factors for chemical risk assessment. Human variability in the pharmacokinetics of CYP1A2 probe substrates. *Food Chem Toxicol*. 39(7): 681-96.

Dourson, M.L., and J.F. Stara, 1983. Regulatory history and experimental support of uncertainty (safety) factors. *Regul Toxicol Pharmacol*. 3(3):224-38.

Dourson, M.L., L.A. Knauf, and J.C. Swartout. 1992. On reference dose (RfD) and its underlying toxicity data base. *Toxicol. Ind. Health*. 8(3): 171-89.

Dourson, M.L., S.P. Felter, and D. Robinson, 1996. Evolution of science-based uncertainty factors in noncancer risk assessment. *Regul Toxicol Pharmacol*. 24(2 Pt 1):108-20.

Dourson, M. 1996. Uncertainty factors in noncancer risk assessment. *Regul Toxicol Pharmacol*. 24(2 Pt 1):107.

Evans, J.S., L.R. Rhomberg, P.L. Williams, A.M. Wilson, and S.J. Baird, 2001. Reproductive and developmental risks from ethylene oxide: a probabilistic characterization of possible regulatory thresholds. *Risk Anal*. 21(4):697-717.

Fredrick, C.B., P.R. Gentry, M.L. Bush, L.G. Lomax, K.A. Black, L. Finch, J.S. Kimbell, K.T. Morgan, R.P. Subramaniam, J.B. Morris, and J.S. Ultman. 2001. A hybrid computational fluid dynamics and physiologically based pharmacokinetic model for comparison of predicted tissue concentrations of acrylic acid and other vapors in the rat and human nasal cavities following inhalation exposure. *Inhal. Toxicol.* 13(5):359-76.

Fredrick, C.B., L.G. Lomax, K.A. Black, L. Finch, H.E. Scribner, J.S. Kimbell, K.T. Morgan, R.P. Subramaniam, and J.B. Morris. 2002. Use of a hybrid computational fluid dynamics and physiologically based inhalation model for interspecies dosimetry comparisons of ester vapors. *Toxicol. Appl. Pharmacol*. 183(1):23-40.

Freireich, E.J., E.A. Gehan, D.P. Rall, L.H. Schmidt, and H.E. Skipper. 1966. Quantitative comparison of toxicity of anticancer agents in mouse, rat, hamster, dog, monkey and man. *Cancer Chemother. Rep.* 50: 219-44.

Gaylor, D.W., and R.L. Kodell. 1999. Dose-response trend tests for tumorigenesis, adjusted for body weight. *Toxicol. Sci.* 49(2):318-23.

Gentry, P.R., T.R. Covington, and H.J. Clewell III. 2003. Evaluation of the potential impact of pharmacokinetic differences on tissue dosimetry in offspring during pregnancy and lactation. *Regul. Toxicol. Pharmacol.* 38 :1-16.

Gillette, J.R. 1987. Dose, species and route extrapolation: general aspects. *In*: National Research Council. *Pharmacokinetics in Risk Assessment: Drinking Water and Health, Vol. 8*. National Academy Press, Washington, DC.

Ginsberg G., D. Hattis, B. Sonawane, A. Russ, P. Banati, M. Kozlak, S. Smolenski and R. Goble. 2002. Evaluation of child/adult pharmacokinetic differences from a database derived from the therapeutic drug literature. *Toxicol Sci*. 66(2): 185-200.

Ginsberg G., D. Hattis, A. Russ, and B. Sonawane. 2004. Physiologically based pharmacokinetic (PBPK) modeling of caffeine and theophylline in neonates and adults: implications for assessing children's risks from environmental agents. *J Toxicol Environ Health A.* 67(4): 297-329

Goldman, L.R. 1995. Children—unique and vulnerable. Environmental risks facing children and recommendations for response. *Environ. Health Perspect.* 103 (Suppl. 6):13-18.

Goldman, L.R. 1998. Linking research and policy to ensure children's environmental health. *Environ. Health Perspect.* 106 (Suppl. 3):857-62.

Goldman, L.R., and S. Koduru. 2000. Chemicals in the environment and developmental toxicity to children: a public health and policy perspective. *Environ. Health Perspect.* 108 (Suppl. 3): 443-8.

Hart, R.W, Setlow, R.B. 1974. Correlation between deoxyribonucleic acid excision-repair and life-span in a number of mammalian species. *Proc Natl Acad Sci U S A.* 71(6):2169-73.

Hattis, D. 1991. Use of biological markers and pharmacokinetics in human health risk assessment. *Environ. Health. Perspect*. 89:230-8.

Hattis D., G. Ginsberg, B. Sonawane, S. Smolenski, A. Russ, M. Kozlak, and R. Goble. 2003. Differences in pharmacokinetics between children and adults—II. Children's variability in drug elimination half-

lives and in some parameters needed for physiologically-based pharmacokinetic modeling. *Risk Anal*. 23(1): 117-42.

Hill, T.A., R.C. Wands, and R.W. Leukroth, Jr. (eds.). 1986. *Biological Bases for Interspecies Extrapolation of Carcinogenicity Data*. Prepared for Food Safety and Applied Nutrition, Food and Drug Administration, Department of Health and Human Services, Washington, DC, under contract to Federation of American Societies for Experimental Biology, FDA contract No 223-83-2020.

Hoel, D.G. 1977. Some problems in low-dose extrapolation. *In*: Hiatt, H.H., J.D. Watson, and J.A. Winsten (eds.) *Origins of Human Cancer: Book C, Human Risk Assessment*. Cold Spring Harbor Conferences on Cell Proliferation, Vol. 4. Cold Spring Harbor Laboratory, Cold Spring Harbor, L.I., N.Y.

Hogan, M., and D.G. Hoel. 1982. Extrapolation to man. *In*: Hayes, A.W. (ed.) *Principles of Toxicology*. Raven Press, New York.

Ings, R.M. 1990. Interspecies scaling and comparisons in drug development and toxicokinetics. *Xenobiotica*. 20(11):1201-31.

Jarabek, A.M., M.G. Menache, J.H. Overton, M.L. Dourson, and F.J. Miller. 1989. Inhalation reference dose (RfDi): An application of interspecies dosimetry modeling for risk assessment of insoluble particles. *Health Phys.* 57 (Supp 1) :177-183.

Jarabek, A.M. 1995. Interspecies extrapolation based on mechanistic determinants of chemical disposition. *Hum. Ecol. Risk Assess.* 1(5) :641-662.

Kalberlah, F., U. Fost, and K. Schneider. 2002. Time extrapolation and interspecies extrapolation for locally acting substances in case of limited toxicological data. *Ann. Occup. Hyg.* 46(2):175-185.

Kleiber, M. 1932. Body size and metabolism. *Hilgardia* 6:315-53.

Kuo, J., I. Linkov, L. Rhomberg, M. Polkanov, G. Gray, and R. Wilson. 2002. Absolute risk or relative risk? A study of intraspecies and interspecies extrapolation of chemical-induced cancer risk. *Risk Anal.* 22:141-157.

Landrigan, P.J. 1999. Risk assessment for children and other sensitive populations. *Ann. N Y Acad. Sci.* 895:1-9.

Lindstedt, S.L., and W.A. Calder. 1976. Body size and longevity in birds. *Condor* 78:91-4.

Lindstedt, S.L., and W.A. Calder. 1981. Body size, physiological time, and longevity of homeothermic animals. *Quart. Rev. Biol.* 56:1-16.

Lindstedt, S.L. 1987. Allometry: body size constraints in animal design. In: National Research Council. *Pharmacokinetics in risk assessment: drinking water and health, Vol. 8*. National Academy Press, Washington, DC.

Mahmood, I. 1999. Allometric issues in drug development. *J. Pharm. Sci.* 88(11):1101-6.

Mantel, N., and M.A. Schneiderman. 1975. Estimating safe levels, a hazardous undertaking. *Cancer Res*. 35: 1379-86.

Meek, B., A. Renwick, and C. Sonich-Mullin. 2003. Practical application of kinetic data in risk assessment—an IPCS initiative. *Toxicol. Lett.* 138 (1-2) :151-160.

Miller, R.P., R.J. Roberts, and L.J. Fischer. 1976. Acetaminophen elimination kinetics in neonates, children, and adults. *Clin. Pharmacol. Ther*. 19(3):284-94.

Miller, M.D., M.A. Marty, A. Arcus, J. Brown, D. Morry, and M. Sandy. 2002. Differences between children and adults: implications for risk assessment at California EPA. *Int. J. Toxicol.* 21(5):403-18.

Monticello, T.M., and K.T. Morgan. 1994. Cell proliferation and formaldehyde-induced respiratory carcinogenesis. *Risk Anal.* 14(3):313-9.

Mordenti, J. 1986. Man versus beast: Pharmacokinetic scaling in mammals. *J. Pharmacol. Sci.* 75:1028-40.

Morgan K.T. 1997. A brief review of formaldehyde carcinogenesis in relation to rat nasal pathology and human health risk assessment. *Toxicol. Pathol.* 25(3):291-307.

O'Flaherty, E.J. 1989. Interspecies conversion of kinetically equivalent doses. *Risk Anal*. 9(4):587-98.

O'Flaherty, E.J. 1994. Physiologic changes during growth and development. *Environ Health Perspect.* 102 Suppl 11:103-6.

Pauluhn, J. 2003. Issues of dosimetry in inhalation toxicity. *Toxicol. Lett*. 140-141: 229-238.

Pelekis, M., L.A. Gephart, and S.E. Lerman. 2001. Physiological-model-based derivation of the adult and child pharmacokinetic intraspecies uncertainty factors for volatile organic compounds. *Regul. Toxicol. Pharmacol.* 33(1):12-20.

Pinkel, D. 1958. The use of body surface area as a criterion of drug dosage in cancer chemotherapy. *Cancer Res*. 18: 853-6.

Rall, D.P. 1977. Species differences in carcinogenesis testing. In: Hiatt, H.H., J.D. Watson, and J.A. Winsten (eds.) *Origins of Human Cancer: Book C, Human Risk Assessment. Cold Spring Harbor Conferences on Cell Proliferation, Vol. 4*. Cold Spring Harbor Laboratory, Cold Spring Harbor, L.I., N.Y.

Reitz, R.H., A.L. Mendrala, and F.P. Guengerich. 1988. In vitro studies of methylene chloride (MEC) metabolism in human and animal tissues: Use in physiologically-based pharmacokinetic (PB-PK) models. *The Toxicologist* 8(1): 156

Renwick, A.G. 1998. Toxicokinetics in infants and children in relation to the ADI and TDI. *Food Addit Contam.* 15 Suppl:17-35.

Renwick, A.G., J.L. Dorne, and K. Walton. 2000. An analysis of the need for an additional uncertainty factor for infants and children. *Regul Toxicol Pharmacol.* 31(3):286-96.

Rhomberg, L.R. 1994. Empirical comparisons of carcinogen potencies across species and their bearing on appropriate scaling of doses for risk assessment. p425-445. *In*: Wang, R.G.M. (ed.) 1994. *Water Contamination and Health: Integration of Exposure Assessment, Toxicology, and Risk Assessment*. Marcel Dekker, Inc., New York.

Rhomberg, L.R. 1995. Use of quantitative modelling in methylene chloride risk assessment. *Toxicology* 102:95-114.

Rhomberg, L.R, and S.K. Wolff. 1998. Empirical scaling of single oral lethal doses across mammalian species based on a large database. *Risk Anal*. 18 :741-753

Rhomberg, L.R., and L.C. Caprario. 1999. Cross-species and cross-route comparisons of acute lethal doses: Further results of analysis of a large database. *In*: Society for Risk Analysis Annual Meeting, December 6, 1999.

Rhomberg, L.R., 2000. Mode of action as a guide to quantitative analytical approaches. Presented at the Society of Toxicology Annual Meeting, Philadelphia, PA. March 21, 2000.

Riviere, J.E., T. Martin-Jimenez, S.F. Sundlof, and A.L. Craigmill. 1997. Interspecies allometric analysis of the comparative pharmacokinetics of 44 drugs across veterinary and laboratory animal species. *J. Vet. Pharmacol. Ther*. 20(6):453-63.

Sarangapani, R., P.R. Gentry, T.R. Covington, J.G. Teeguarden,and H.J. Clewell. 2003. Evaluation of the potential impact of age- and gender-specific lung morphology and ventilation rate on the dosimetry of vapors. *Inhal Toxicol*. 15(10):987-1016.

Schein, P.S., R.D. Davis, S. Carter, J. Newman, D.R. Schein, and D.P. Rall. 1979. The evaluation of anticancer drugs in dogs, and monkeys for the predication of quantitative toxicities in man. *Clin. Pharmacol. Therapeut.* 11:3-40.

Schlosser P.M., P.D. Lilly, R.B. Conolly, D.B. Janszen, and J.S. Kimbell. 2003. Benchmark dose risk assessment for formaldehyde using airflow modeling and a single-compartment, DNA-protein cross-link dosimetry model to estimate human equivalent doses. *Risk Anal.* 23(3):473-87.

Schmidt, C.W., C.A. Gillis, R.E. Keenan, and P.S. Price. 1997. Characterizing inter-chemical variation in the interspecies uncertainty factor (UFa). *Fundam. Appl. Toxicol.* (Suppl.) 36(1 Part 2):208. Abstract 1057.

Starr, T.B. 1990. Quantitative cancer risk estimation for formaldehyde. *Risk Anal*. 10(1):85-91.

Travis, C.C., and R.K. White. 1988. Interspecific scaling of toxicity data. *Risk Anal*. 8:119-25.

Travis, C.C. 1990. Tissue dosimetry for reactive metabolites. *Risk Anal.* 10:317-21.

Travis, C.C., and J.C. Bowers. 1991. Interspecies scaling of anesthetic potency. *Toxicol. Ind. Health.* 7(4):249-60.

US EPA**.** June 5, 1992. Draft report: A cross-species scaling factor for carcinogen risk assessment based on equivalence of $mg/kg^{3/4}$/day. *Fed. Reg.* 57 (109) :24152(22).

US EPA. 1994. *Methods for Derivation of Inhalation Reference Concentrations and Application of Inhalation Dosimetry*. Office of Health and Environmental Assessment, Environmental Criteria and Assessment Office, Research Triangle Park, NC. EPA/600/8-90/006F.

US EPA. 1999. *Guidelines for Carcinogen Risk Assessment* (Review Draft). Risk Assessment Forum, U.S. Environmental Protection Agency, Washington, DC. NCEA-F-0644.

Vocci, F., and T. Farber. 1988. Extrapolation of animal toxicity data to man. *Regul. Toxicol. Pharmacol.* 8:389-98.

Watanabe, K.H., and F.Y. Bois. 1996. Interspecies extrapolation of physiological pharmacokinetic parameter distributions. *Risk Anal.* 16(6) :741-754.

APPENDIX A

Mathematical Relationships For Scaling According To A Power of Body Weig*ht*

Allometric relationships are premised upon the following equation:

$$Y = a \cdot BW^b,$$

where *Y* is the feature (an organ weight, blood flow, metabolic rate, *etc.*), *BW* is body weight, and *b* is a power of weight, and *a* is a proportionality constant.

In the special case where the value of *b* is 1.0, the plot is a straight line and is described as isometry. Organ volumes tend to scale in a direct relationship with body weight and so can be described as isometric. When *b* is a value different than 1.0 (*e.g.*, 0.75), the relationship between *Y* and *BW* is not a straight line and is described as allometric. Many metabolically-related parameters and in general all physiological rate processes (*e.g.*, heart beats, cardiac output, glomerular filtration rate, minute volumes, oxygen consumption) tend to be described by allometric equations.

When the scale of values considered is extremely large (*e.g.*, the body weight range between elephants and mice), the expression can conveniently be expressed in logarithmic form:

$$\log Y = \log (a \cdot BW^b) = \log a + b \cdot \log BW.$$

That is, on a log-log plot, the relationship is a line with slope *b*.

Some parameters have only time as a determinant, such as metabolic half-lives (fraction per hour) or respiration (breaths per hour). Other parameters of interest, such as cardiac output (ml/hr) are more conveniently related in proportion to overall body size. This is accomplished by dividing out the body mass of the organism and then examining the scaling of amount per kilogram. This yields value having units with (mass*time) in the denominator. Thus:

$$Y/BW = (a \cdot BW^b)/M$$

Given that the exponent for body mass is 1.0, if the value of $b = 0.75$, then

$$Y/BW^{1.0} = (a \cdot BW^{0.75})/BW^{1.0}$$

or

$$Y' = a \cdot BW^{-0.25}$$

where *Y*' is the value of interest expressed in a body mass specific manner (liters of blood per hour).

If it is presumed that daily doses are equally toxic in differently sized species when they are proportional to body weight to the power *b*, this means that the raw daily amounts *A* (in mg) can be normalized by dividing by BW^b, *i.e.*,

$$A_1/BW_1^b = A_2/BW_2^b,$$

where 1 and 2 refer to two species. For example, if species 1 is a 35 g mouse and species 2 is a 70 kg human, and if b=3/4, then when A_1 is a 1 mg dose to a mouse, this constitutes 1 mg/(.035kg)$^{3/4}$ = 12.4 mg/kg$^{3/4}$. A toxicologically equivalent dose A_2 to a human would also equal 12.4 mg/kg$^{3/4}$, that is

$$A_2/70^{3/4} = 12.4$$

$$A_2 = 12.4 \cdot 70^{3/4} = 299 \text{ mg.}$$

The total dose is almost 300-fold bigger, which is less than proportional to the body size difference of 2000-fold (= 70/0.035).

If these two mg doses A_1 and A_2 are expressed as mg/kg/day doses D_1 and D_2, then they are

$$D_1 = A_1/BW_1 = 1/.035 = 28.6 \text{ mg/kg/day}$$

$$D_2 = A_2/BW_2 = 299/70 = 4.27 \text{ mg/kg/day.}$$

That is, the human dose is 28.6/4.27 = 6.7-fold lower on a body-weight basis.

If one starts with a mouse dose *already* scaled to body weight (*i.e.*, D_1 in mg/kg/day) and wishes to scale this (using body weight to the ¾ power scaling) to a human mg/kg/day dose D_2, then the "scaling factor" is 6.7 since $D_2 = D_1/6.7$. The size of this factor is given by $(BW_2/BW_1)^{1/4}$, in our case $(70/.035)^{1/4}$ = 6.7. The reason this is so (and the ¼ power is used) is because (using relationships shown above)

$$A_1/BW_1^{3/4} = A_2/BW_2^{3/4}$$

Since $1/BW^{3/4} = BW^{1/4}/BW$, this can be rewritten as

$$(A_1/BW_1) \cdot BW_1^{1/4} = (A_2/BW_2) \cdot BW_2^{1/4}$$

$$D_1 \cdot BW_1^{1/4} = D_2 \cdot BW_2^{1/4}$$

$$D_2 = D_1 \cdot (BW_1^{1/4}/BW_2^{1/4}) = D_1/(BW_2/BW_1)^{1/4}$$

Figure 1A. Plasma toxicant concentrations with human and mouse doses scaled according to bodyweight. Humans and mice receive the same mg/kg body weight dose at the same dose intervals. The faster metabolism in the mouse leads to a more rapid attainment of steady state conditions and a lower cumulative body burden.

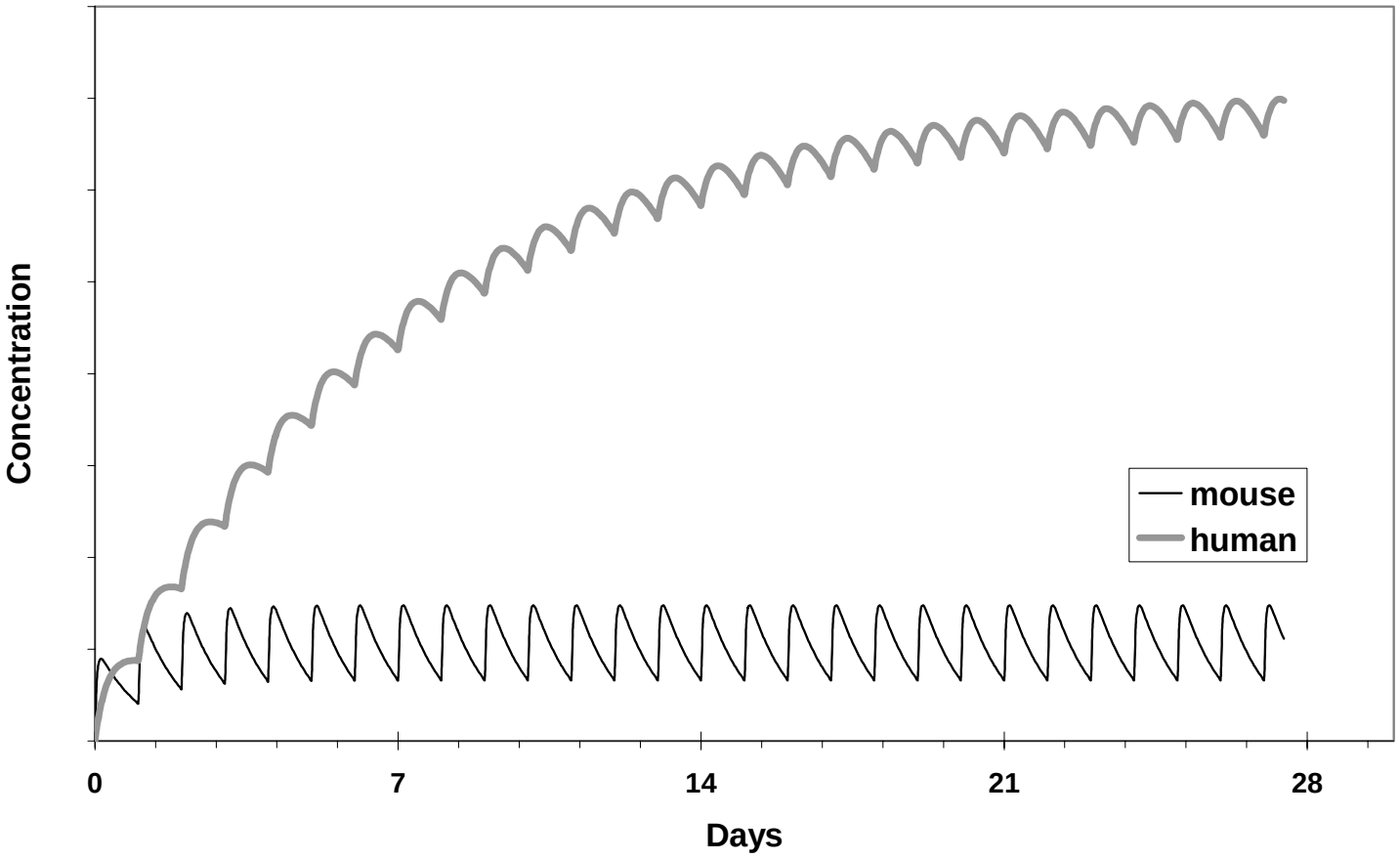


Figure 1B. Plasma toxicant concentrations with human and mouse doses scaled according to the ¾ power of bodyweight. This method of dose scaling leads to the same average steady state concentration and more similar AUCs compared to scaling to $BW^{1.0}$, as shown in Figure 1A.

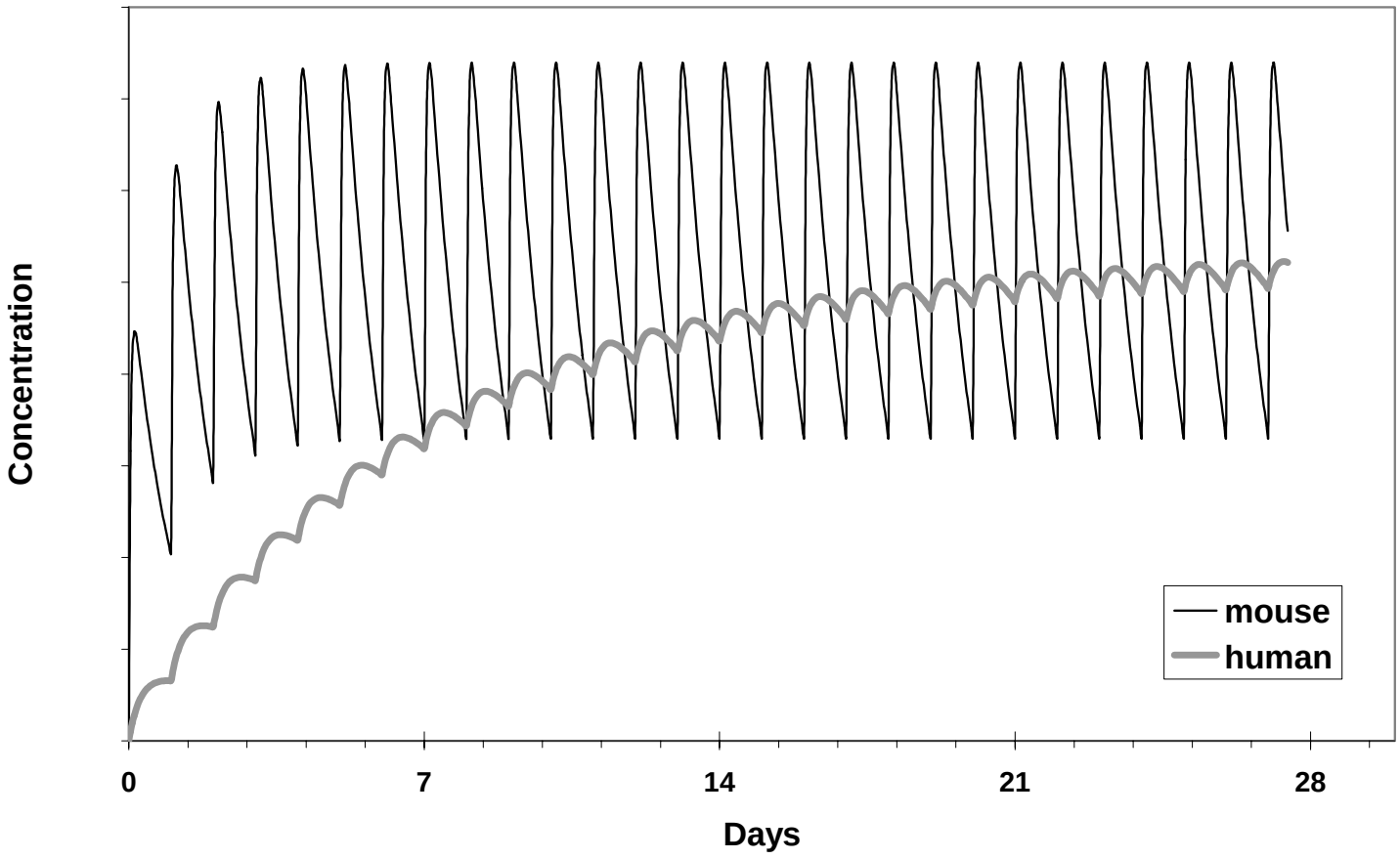


Figure 1C. Plasma toxicant concentrations with the dose interval adjusted according to scaling by the ¾ power of body weight. Plasma concentrations with daily doses, scaled according to $BW^{1.0}$, mice dosed daily and humans dosed every 7 Days. Adjusting the dose interval rather than the dose itself by multiplying the human dosing interval by the body weight difference between the two species raised to the –1/4 power (*i.e.*, $(0.035/70)^{-0.25} = 6.7$), also leads to the same average steady state concentration and similar AUCs.

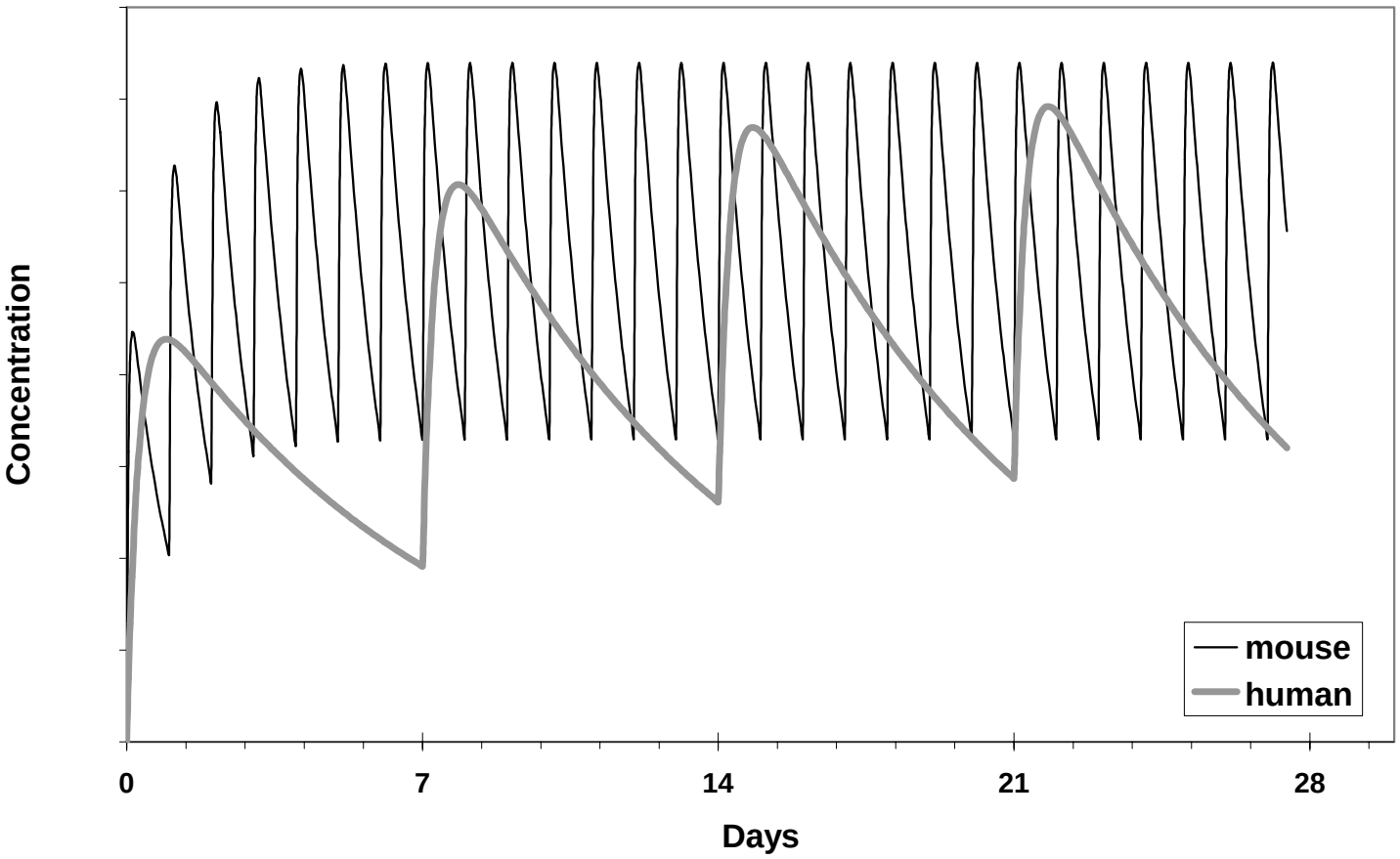


Figure 1D. Plasma toxicant concentrations scaled according to physiological time. Data from Figure 1C shown on an x-axis of physiological time. The x-values for the human and mouse data have been converted from chronological time to physiological time by multiplying the human x-axis data by the body weight difference between the two species raised to the –1/4 power (*i.e.*, $(0.035/70)^{-0.25} = 6.7$). This leads to overlapping of the human and mouse data.

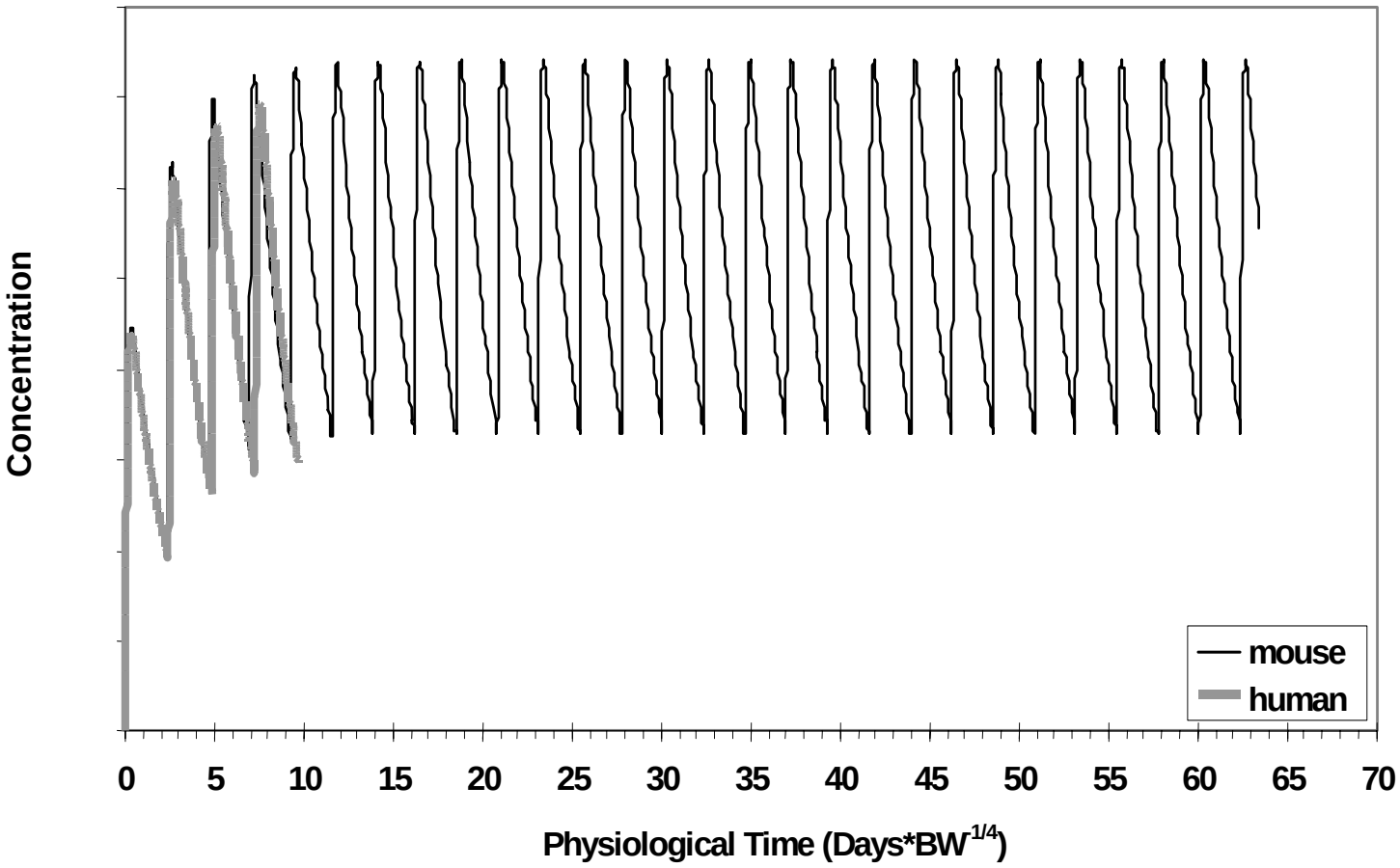


Figure 2. Biological parameters scaled according to chronological and physiological time. (A) Various biological parameters are shown expressed in chronological time. Not surprisingly, data for mice and humans differ by several orders of magnitude. (B) The same data is shown expressed in physiological time. Note that the interspecies differences in the parameter values are markedly reduced when placed in the context of physiological time..

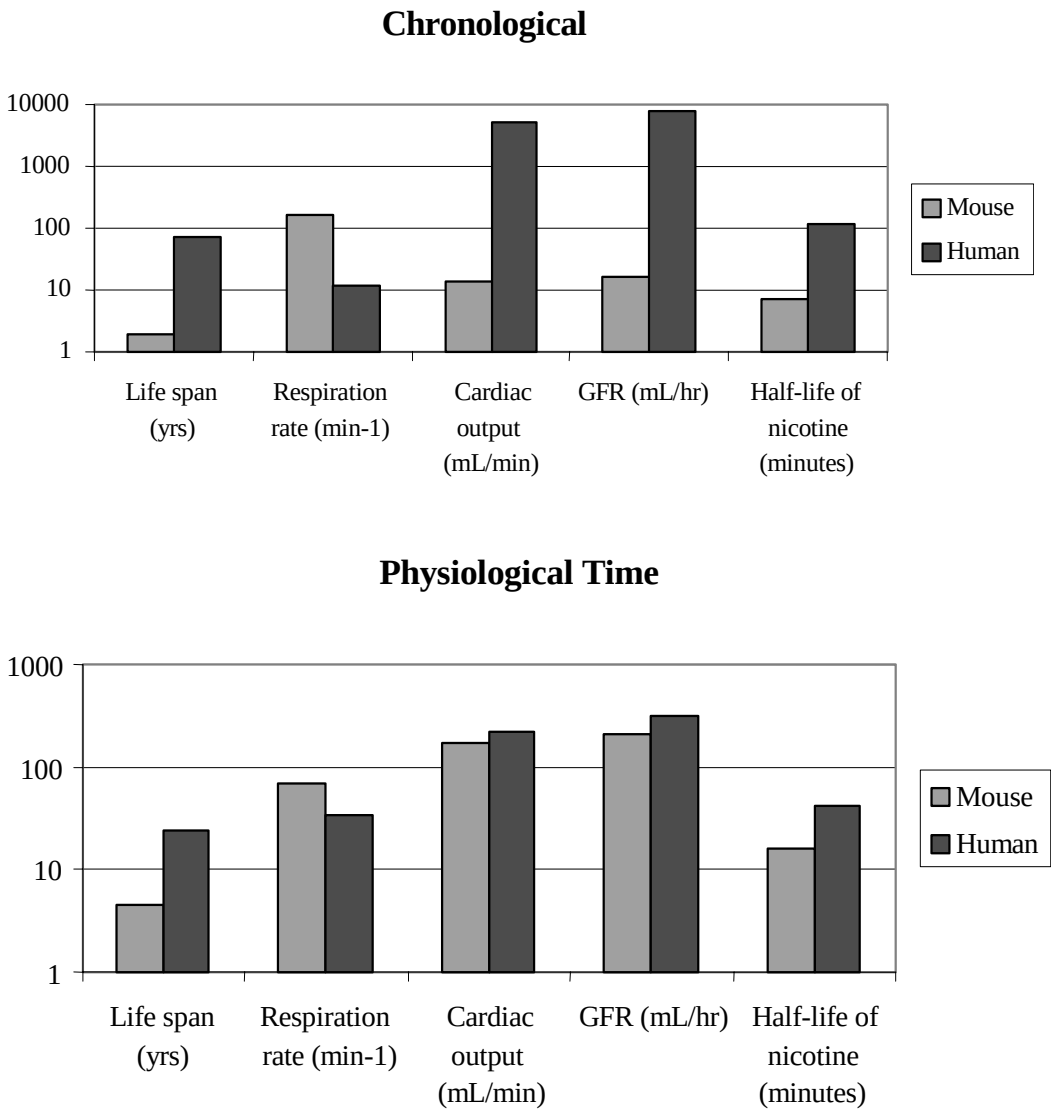
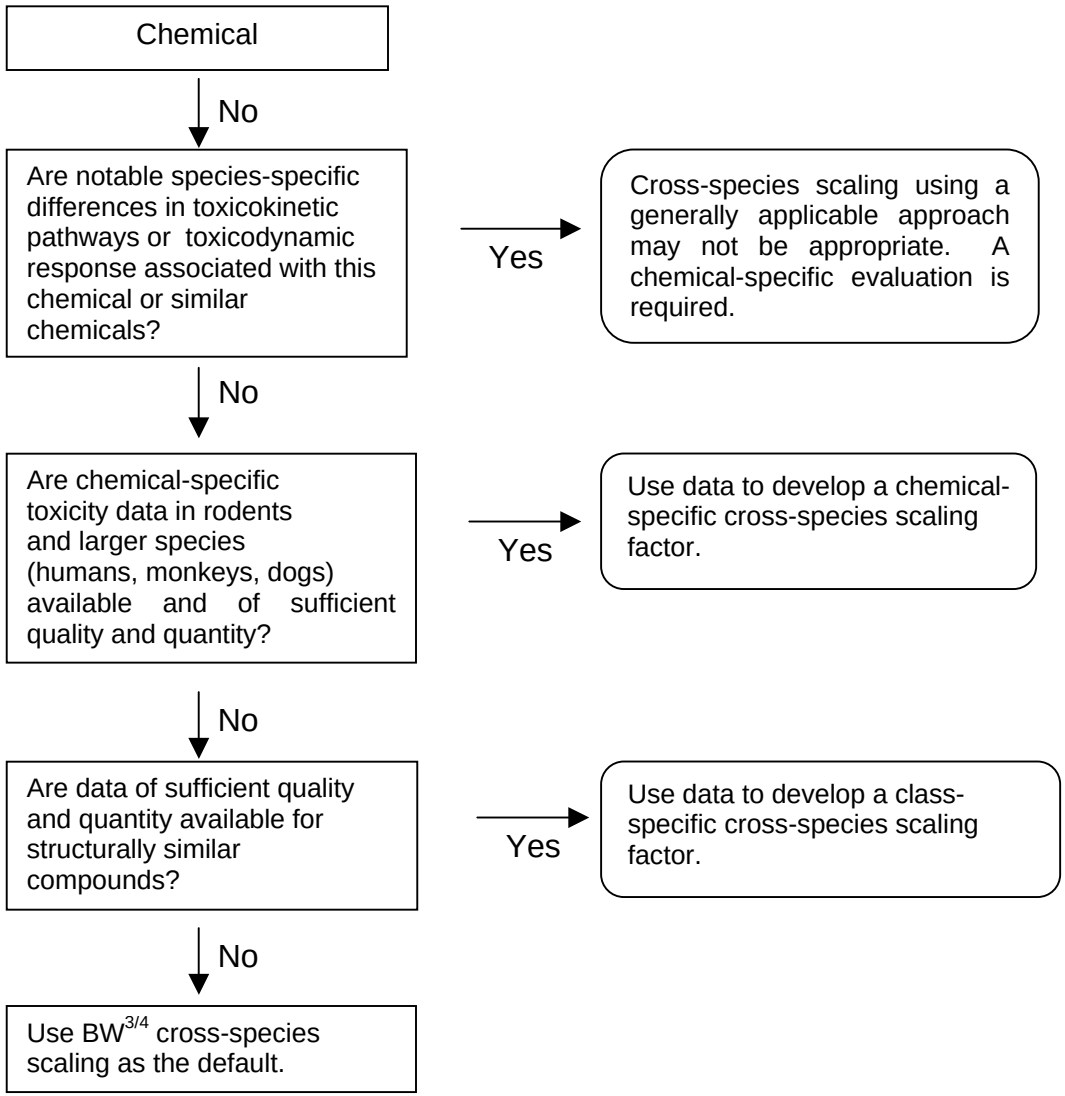


Figure 3. Proposed decision process for considering the use of chemical-specific cross-species scaling data.



Limitations

The existance of chemical- and species-specific differences in metabolic pathways or other factors may not be known *a priori*. This suggests that the scaling may have to be revisited as new data become available.

If data are limited (*e.g.*, acute toxicity only, no data in humans) the validity and applicability of the chemical-specific scaling factor may be questionable.

The degree to which the chemical follows rules for the class will be unknown.