EPA/600/R-21/305
ERASC-017F
December 2021

ALLOMETRIC SCALING OF TERRESTRIAL WILDLIFE ORAL TOXICITY MEASUREMENTS AND COMPARISON OF ECOLOGICAL TO HUMAN HEALTH ASSESSMENT CONTEXTS

Ecological Risk Assessment Support Center
Office of Research and Development
U.S. Environmental Protection Agency
Cincinnati, OH

DISCLAIMER

This document has been reviewed in accordance with U.S. Environmental Protection Agency policy and approved for publication. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

Preferred Citation:

Farrar, D. and M. Kravitz. 2021. Allometric Scaling of Terrestrial Wildlife Oral Toxicity Measurements and Comparison of Ecological to Human Health Assessment Contexts. U.S. Environmental Protection Agency, Ecological Risk Assessment Support Center, Cincinnati, OH. EPA/600/R-21/305.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

AUC – area under a curve, here a curve relating tissue concentration to time

BW – body weight

CL – clearance (mass of tissue cleared of toxicant per unit time)

C_{max} – maximum tissue concentration

LD10, LD50, etc. – dose lethal with probability 10%, 50%, etc. (a parameter in a dose response curve, e.g., probit), or an estimate of such a parameter based on data from a toxicity study. (Whether the reference is to actual parameter value or an estimate will be clarified according to context.)

NOAEL – no observed adverse effect level (of exposure)

LOAEL – low observed adverse effect level (of exposure)

PBPK – physiologically-based pharmacokinetic

PD – pharmacodynamic (equivalent for present purposes to toxicodynamic)

PK – pharmacokinetic (equivalent for present purposes to toxicokinetic)

RAF – Risk Assessment Forum

TD, TK – see PD, PK

U.S. EPA – U.S. Environmental Protection Agency

U.S. FDA – U.S. Food and Drug Administration

AUTHORS AND REVIEWERS

AUTHORS

David Farrar
U.S. Environmental Protection Agency
Office of Research and Development
Center for Public Health and Environmental Assessment
Cincinnati, OH

Michael Kravitz
U.S. Environmental Protection Agency
Office of Research and Development
Center for Environmental Solutions and Emergency Response
Cincinnati, OH

REVIEWERS

Robert Paul Hunter
One Medicine Consulting
Olathe, Kansas

Brad E Sample
Ecological Risk, Inc.
Rancho Murieta, California

ACKNOWLEDGMENTS

The first draft of this document was internally (within EPA) reviewed by Jeff Swartout (EPA Office of Research and Development), and Kristina Garber and Melissa Panger (EPA Office of Pesticide Programs). Programmatic review was conducted by David Charters (EPA Office of Land and Emergency Management), Andrea LaTier (EPA Region 10) and Marc Greenberg (EPA Office of Land and Emergency Management). This document was externally peer reviewed under contract to Versar Inc., 6850 Versar Center, Springfield, VA 22151 (Contract EP-C-17-023). Paul Schlosser (EPA Office of Research and Development) provided helpful input into preparation of the final document.

QUALITY ASSURANCE

This work was conducted under the U.S. EPA Quality Assurance (QA) program to ensure data are of known and acceptable quality to support their intended use. Surveillance of the work by the ERASC Director ensured adherence to QA processes and criteria, as well as quick and effective resolution of any problems. The QA manager and the ERASC Director have determined under the QA program that this work meets all U.S. EPA quality requirements. This was written with guidance from the Center for Public Health and Environmental Assessment (CPHEA) Program Quality Assurance Project Plan (PQAPP), the QAPP titled "PQAPP for the Superfund Health Risk Technical Support Center (STSC) and Ecological Risk Assessment Support Center (ERASC)" QAPP ID L-CPAD-0030721-QP, formally QAPP ID NCEA-16-00003. As part of the QA system, a quality product review is done prior to management clearance. This document received internal peer review by at least two scientists and an independent external peer review by at least three scientific experts managed by Versar under contract EP-C-17-023.

PREFACE

EPA's Ecological Risk Assessment Support Center (ERASC) provides state-of-the-science technical information relevant to ecological risk assessments and cleanups at hazardous waste sites (https://www.epa.gov/risk/erasc). Due to uncertainty surrounding the use of allometry in ecological risk assessments, ERASC was requested by risk assessors in the EPA Office of Land and Emergency Management (OLEM) program to clarify the appropriate use of allometric scaling of toxicity measurements in ecological risk assessments.

1. SUMMARY

The possibility of nonlinear effects of body weights should be considered in any analysis of biological parameters across species with significantly different body sizes. The term "allometry" is used where there is a possibly non-linear relationship of a toxicologically relevant parameter to body weight (BW), particularly when the relationship can be described with a power function BW^b with exponent b. (See Section 2. We also use the term "scaling".) This report discusses scaling defaults for terrestrial wildlife oral toxicity measurements in the form of exposure values (dose or concentration) associated with specified toxicological outcomes. These may be in "dietary" form (e.g., ppm toxicant in feed)[1] and dose form (e.g., mg/kg or mg/kg-d). Extrapolation of dose-based toxicity measurements on a "simple body weight basis" (the case b = 1) is held to apply primarily to lethality measurements based on single doses, or other situations involving appreciable lethality, while in most other situations an allometric adjustment based on b = ¾ is the recommended default. Exceptions for particular situations can be based on direct empirical evidence, mechanistic information, or modeling. The recommendations are consistent with current health assessment policy (U.S. EPA, 2011).

[1] Here the term "dietary toxicity" follows U.S. EPA (2015) and does not refer simply to exposure via the diet. The term means here that the toxicity measurement takes the form of food concentration associated with an effect, not converted to dose on a body weight basis by combining the dietary toxicity with a food intake rate.

It is argued that a biologically consistent approach is to apply allometric adjustments to dose-based measurements, expressed on a body weight basis, but apply no adjustment to "dietary" measurements (assume the same critical feed concentration for test and assessment species). Sufficient assumptions for this approach include that uptake and clearance rates scale to the same power of body weight and that tissue concentrations over time can be appropriately summarized by averaging or cumulation (e.g., AUC) (O'Flaherty, 1989). The effect of the generally higher food intake rate for smaller animals is that tissue concentrations may rise more rapidly initially in small animals, at a given environmental concentration, without necessarily producing a substantial difference in longer-term average exposure (as illustrated by simulation in Appendix A).

2. BACKGROUND

Nonlinear effects of body weight (BW) have been documented for many biological processes (Schmidt-Nielsen, 1984; Mahmood, 2005). "Allometric" curves allow that such effects of BW on biological processes may be nonlinear, for example if the biologically effective mass of a toxicant increases smoothly but not proportionally with species body weight.[2] The possible effect of allometric relationships has been considered in the context of cross-species extrapolations of toxicity for human health and ecological assessments, and also in veterinary toxicology.

The U.S. EPA Risk Assessment Forum proposed defaulting to allometric scaling with three-quarters power of bodyweight ($BW^{3/4}$) for interspecific extrapolation of toxicity data for deriving human health oral reference doses for both cancer and non-cancer endpoints in the absence of other information to support a different interspecific relationship (U.S. EPA 2011). $BW^{3/4}$ scaling is viewed as relating primarily to pharmacokinetic (PK) concerns but possibly also to pharmacodynamic concerns. (In addition to literature cited by U.S. EPA to support the PK arguments, see Boxenbaum, 1980; Boxenbaum and DiLea, 1995). The approach is considered

[2] In practice, a nonlinear allometric relationship of a quantity Y to body weight (BW) is ordinarily expressed using a power function $Y = a \bullet BW^b$ where parameter a is termed the "coefficient" and b the "exponent." Thus $b = 1$ corresponds to proportionality between Y and BW, while $b < 1$ to decelerating curves and $b > 1$ to accelerating curves.

most appropriate for oral exposure, toxicity caused by the parent compound or a stable metabolite, clearance through first-order biological (metabolic) processes, and chronic exposures and effects. It is considered by USEPA (2011) "not generally appropriate "in the case of a single exposure eliciting sudden and severe toxicity resulting from immediate and intolerable damage to some critical biological pathway, and where repair processes (i.e., TD) would be overwhelmed." Nonetheless, U.S. EPA (ibid.) suggests $BW^{3/4}$ scaling may still be useful for acute exposures with non-lethal effects "in which the functional status of physiological processes are comparable to the chronic scenario." Allometric scaling is considered (by U.S. EPA, ibid.) inappropriate when toxicity is attributed to formation of a reactive metabolite or for very high exposures that saturate the relevant metabolic processes.

In contrast, Allard, et al. (2009) proposed that ecological risk assessors should not use allometric dose-scaling with body mass when assessing chronic toxicity between species, stating that "allometric scaling models developed for both human and wildlife risk assessment are all based on acute toxicity data." (An extended quote from this source can be found in Appendix B.) However, Allard et al. did not discuss the pharmacokinetic basis for allometric scaling of chronic toxicity in U.S. EPA (2011).

Due to the uncertainty surrounding the use of allometry in ecological risk assessments, risk assessors in EPA OLEM submitted a request to ERASC for clarification of the appropriate use of allometric scaling of toxicity measurements in ecological risk assessments. Specifically, the problem statement was:

> What is the appropriate use of allometric scaling for characterizing toxicity in ecological risk assessments? 1) What is the default methodology? 2) Can you develop scientifically justified deviations from the default? If so, how?

This document attempts to answer these questions by reviewing the general types of arguments used to support science-based policy in scaling decisions as those arguments may apply to wildlife assessments, leading to recommended defaults, while recognizing that exceptions may be justified in particular situations. Allometric scaling would be considered in

Step 3 (baseline risk assessment problem formulation) of the Ecological Risk Assessment Guidance for Superfund (U.S. EPA, 1997). (It would not be used in the screening steps of an assessment.) It is important to note that while the information provided in this document may be useful to a number of programs, particularly Superfund, it is not meant to be prescriptive.

2.1 Allometric Scaling of Toxicity Measurements

U.S. EPA (1993) documents many allometric relationships involving biological parameters for wildlife. While the primary concern here is the direct scaling of toxicity values, any biological parameter determining toxicity may itself scale allometrically. Toxicity is influenced by a balance of uptake processes with processes of elimination or recovery. Internal distribution of toxicant is also important, but information on distribution is rarely available in an ecotoxicological setting.

For human health assessments, the U.S. EPA (2011) has reviewed the scientific basis for a current approach — "$BW^{3/4}$ scaling" — and recommends that approach as a default for oral reference doses. Keeping in mind that the emphasis in the U.S. EPA (2011) review is human health assessment methodology based on mammalian data, that review may be considered for general toxicological insights on scaling, in combination with analyses more specific to ecological risk assessment.

A formula for $BW^{3/4}$ adjustment is:

$$T_2 = T_1 * (BW_1 / BW_2)^{1/4}$$

where T_2 is the extrapolated toxicity value for Species 2, based on toxicity measurement T_1 in Species 1, both expressed on a BW basis such as mg/kg-d, and BW is indexed 1 or 2 according to species.[3]

[3] Some confusion seems inevitable from the use of a 1/4 power in a formula said to represent a 3/4-power method. The 3/4 power appears in the expression for scaling a critical dose (e.g., mg/d) that has not been normalized relative to BW (see Appendix D). In any case the expressions given can be rationalized pharmacokinetically by assuming that clearance of a chemical (CL) is proportional to the 3/4 power of BW. A quarter power relates to the *fraction* of tissue clear per unit time to BW, because $BW^{3/4} / BW = 1/BW^{1/4}$. See U.S. EPA (2011) for additional discussion of powers of BW in expressions for different types of biological quantities.

More generally, $BW^{3/4}$ can be considered one case of a "BW^{b}" framework, where b is the allometric exponent. The general formula is similar to the one just given, but the power in general is $1 - b$. If we choose $b = 1$ the formula becomes simply $T_2 = T_1$, i.e., the extrapolated toxicity value equals the measured toxicity value when both are expressed on a body-weight basis. This "BW^{1}" option can also be termed "extrapolation on a simple body-weight basis." Various choices of $b \neq 1$ represent nonlinearity. The term "surface area correction" refers most appropriately to the specific choice $b = 2/3$, a choice that, as a default value, has been replaced by 3/4 in U.S. EPA (2011). The qualitative effect of using an inappropriate value for the exponent is to underestimate risk (overestimate effective doses) when extrapolating from a smaller to a larger species when the assumed value of b is too large, or when extrapolating from larger to smaller species when the value assumed for b is too small.

For the present document there will be a general preference for a few particular values of b that seem to have considerable precedent, especially the values 3/4 and 1. Attempts to further refine allometric methodology by recognizing more context-specific values cannot be dismissed and are viewed as areas for possible further study. We note for example a recent analysis of basal metabolic rate (White et al., 2009) suggesting variation of the allometric exponent among evolutionary lineages of mammals. At the same time, at least one serious theory (the West-Brown-Enquist theory, West 2017; West et al., 2002; c.f., Savage et al., 2008) attempts a rationale for an exact value 3/4 for the allometric exponent.

2.2 Some Aspects of the U.S. EPA Viewpoint on Scaling of Toxicity Measurements for Human Health Assessments

The U.S. EPA (2011) review is directly concerned with human health assessments but can provide a departure for more general discussions. Notable features of the review include:

- The scientific rationale, summarized in Section 4 of the document, for moving to $BW^{3/4}$ from "surface area" adjustment ($BW^{2/3}$) is a combination of empirical evidence from comparisons of toxic dose values, general biological considerations, and modeling. General biological considerations involve the scaling of basal metabolic rate and other physiological rates including glomerular filtration as $BW^{3/4}$ across species (U.S. EPA, ibid.,

Table 4.1). A particularly notable analysis of empirical data cited is that of Travis and White (1988), based on toxicity of 27 chemotherapy agents (see Section 3.1).

- $BW^{3/4}$ scaling is viewed as relating primarily to PK concerns but possibly also to PD concerns to some degree (see Rhomberg and Lewandowski, 2006, for further discussion).
- PBPK modeling is viewed as the preferred approach for addressing the role of PK in cross-species extrapolation (for human health assessments). However, the possibility of addressing PK considerations to some degree without such modeling is recognized. PBPK models are mechanistic biological models defined by rates of movement of substances between physiological compartments.
- $BW^{3/4}$ scaling is recommended as the usual default for sublethal oral toxicity measurements including many acute toxicity measurements. $BW^{3/4}$ scaling is considered appropriate particularly when area under a concentration curve (AUC) is an appropriate summary of tissue concentrations over time. It is suggested that $BW^{3/4}$ scaling is most appropriate when exposures are in a range in which critical physiological processes operate in about the same way as without exposure.
- "… $BW^{3/4}$ scaling would apply most appropriately to those exogenous substances for which the unmetabolized parent or a stable metabolite is the relevant toxic species and clearance is according to first-order processes". Under these conditions elimination of the toxic moiety is expected to vary as $BW^{3/4}$, which leads to the conclusion that the concentration of the moiety in the body varies likewise. Conversely, "the applicability of $BW^{3/4}$ scaling is less well supported when toxicity is a consequence of exposure to a very reactive parent compound or metabolite that is not removed from the site of formation by biological processes (e.g., subsequent metabolism) but chemically reacts with cellular constituents." In this case total elimination of the toxic moiety is expected to vary as BW^{1}. (see references in U.S. EPA 2011 for support of these conclusions.)
- Possible exceptions to the proposed default are discussed such as lethal effects and portal-of-entry effects.

- It is allowed that specific assessments may differ from the general recommended defaults for "policy" as well as scientific reasons.

2.3 Ecotoxicological Data and Ecological Assessments

Scaling procedures must be considered for two types of ecotoxicological measurements, namely dose-based and dietary (or food-based) (see U.S. EPA, 2015). For a dose-based toxicity measurement, the measurement units are mass toxicant over body weight (e.g., mg/kg) or dosing rate (e.g., mg/kg-d). For a dietary measurement, the measurement units are represented as toxicant concentration in feed. Dietary measurements can be converted to dose rate using information on feeding rate and body weight. The dietary approach has been used, for example, to report the results of avian reproduction studies. (The terminology is somewhat confusing: in this context "dietary" means something more specific than exposure via the diet.)

Additional relevant features of ecological assessment include:

- Ecological assessments may need to address effects on a species category (e.g., birds in general) or multiple exposed species. The variety of receptors exposed, and differences in terms of taxonomy, behavior, and physiology, is expected to pose a challenge for any effort to implement detailed mechanistic (e.g., PBPK) models, as sometimes used in human health assessments.$BW^{3/4}$ is protective in extrapolating from small to large species, the usual situation in human health assessments. For ecological assessments, in contrast, species tested are not necessarily small compared to species exposed, e.g., laboratory rats are large relative to small mammal species of concern in many ecological assessments.
- Lethality data are ordinarily not considered directly in human health assessments, but commonly considered in ecological assessments (see particularly U.S. EPA, 2011, "executive summary").

3. SCIENTIFIC ARGUMENTS FOR THE MOST APPROPRIATE EXTRAPOLATION METHOD

Different conceptual approaches have been used to justify decisions on BW scaling of toxicity measurements. Two general types of evidence have been most important (Rhomberg and Lewandowski, 2006): A "direct empirical" approach uses collections of toxicity measurements where the same substance is evaluated for multiple species. A second approach uses mechanistic arguments, particularly relating to 1) the most appropriate summary of concentration over time and 2) possible body-weight dependency of physiological rates.

3.1 Direct Empirical Evidence from Toxicity Data

The direct empirical approach relies on toxicity measurements of given substances for multiple species with a range of body weights. Analyses of such data for single-dose LD50s have been carried out by various authors (Mineau et al., 1996[4]; Sample & Arenal, 1999[5]; Rhomberg and Wolff, 1998[6]; Burzala-Kowalczyk and Jongbloed, 2011[7]) (additional details for several reviews provided in Appendix B). For the most part these analyses favor BW^{1} scaling (i.e., extrapolation on a simple BW basis) over $BW^{3/4}$ scaling. However, for avian LD50 data for some pesticides, there is support for allometric scaling with a coefficient greater than unity (Mineau et al., 1996, 2001; Sample and Arenal, 1999). We are somewhat uncertain of the set of chemicals and avian taxa to which these results should be held to apply. The database relied upon is said to be weighted towards cholinesterase inhibitors.

It is not clear how successful the empirical approach, based on toxicity measurements, will be for sublethal toxicity measurements. However, Travis and White (1988), invoked in U.S. EPA

[4] For 37 pesticides contributing avian LD50 data, percentiles calculated here, for allometric exponents tabled are 0.8 (5%), 1.15 (50% - median), and 1.4 (95%).

[5] For a subset of 122 substances with an LD50 for 5 or more bird or mammal species the following percentiles of allometric slope distribution have been calculated: 0.7 (5%), 1.1 (50% - median), 1.7 (95%).

[6] The article is used by U.S. EPA (2011) for essentially the same conclusion as stated here. It is based on over 3,000 mammalian single dose toxicity values (which have not been obtained for inspection). The analysis was based on ratios for species. Data have been re-analyzed using regression methodology by Burzala-Kawalczyk and Jongbloed (2011) with essentially similar results. Both analyses concluded that the best single value for allometric exponent is about 1.

[7] The article presents a reanalysis of data assembled by Rhomberg and Wolff (ibid.) using regression methods, again concluding that an allometric exponent of 1 is a reasonable central tendency, while also reporting wide variation.

(2011) to support $BW^{3/4}$ scaling, applied the direct empirical approach for 27 chemotherapy agents by combining a dataset compiled from maximum tolerated doses for 3 species (human, monkey, dog) with LD10 data from mice and rats.[8] For chronic toxicity, possible obstacles include that the data may be summarized by a NOAEL, which may represent different endpoints in different studies (as discussed in Allard et al., 2009). For example, an avian reproduction study may produce measurements of reproductive output (eggs), and growth and survival of offspring through a series of developmental stages. If feasible, it seems preferable to compare the same endpoint across species, based on effective doses estimated using a statistical curve-fitting approach (e.g., nonlinear regression).

3.2 Mechanistic Arguments

Decisions on scaling depend on the most appropriate summary of internal concentration over time (O'Flaherty, 1989). Subject to various exceptions and qualifications $BW^{3/4}$ scaling is supported by various analyses in situations where the concentrations are appropriately summarized in some type of average or cumulative exposure (O'Flaherty, 1989; Sharma and McNeill, 2009; Rhomberg and Lewandowski, 2006; U.S. EPA, 2011). Particular cases are the AUC concentration from a single dose (emphasized particularly in U.S. EPA, 2011), steady-state concentrations from chronic exposure, and time-weighted average exposures.

The strongest PK arguments for $BW^{3/4}$ scaling of toxicity measurements relate to situations where rates for elimination processes are proportional to $BW^{3/4}$ across species. From a PK perspective, species $BW^{3/4}$ ratios are in effect surrogates for corresponding ratios of species typical clearance rates.[9] This conclusion can be derived in a framework of classical

[8] Two sets of estimates of the allometric exponent were combined. For one set of 14 substances, estimates ranged from 0.53 to 0.87; for a second set of 13 substances estimates ranged from 0.53 to 0.96. The authors calculated a 95% confidence interval (0.69, 0.77) for a single estimate.

[9] Clearances have units of volume [https://www.sciencedirect.com/topics/immunology-and-microbiology/clearance] of tissue cleared of chemical per unit time and as a default may be assumed to scale to the 3/4 power of body weight (U.S. EPA, 2011). Glomerular filtration (GF) in particular, possibly an important mechanism of elimination for some substances, scales to the 3/4 power of body weight in mammals and in birds as well (Edwards, 1975; Schmidt-Nielsen, 1984).

pharmacokinetics, assuming that AUC is the appropriate internal dose summary and is inversely proportional to systemic clearance (CL), in turn proportional to $BW^{3/4}$.

In contrast to AUC or average concentrations, peak tissue concentrations (e.g., C_{max} associated with a single exposure event) appear not to have been associated with a simple approach to allometric scaling (U.S. EPA, 2011). If allometric scaling of doses is to be used in a situation where C_{max} is considered the most appropriate basis for extrapolation, the best allometric exponent has not been identified (it may be 1 so far as is known). Modeling (e.g., Fischer, 2005) may be needed if C_{max} is considered the most appropriate internal dose summary.

The analysis of Kirman et al. (2003) has been used by U.S. EPA (2011) as support of $BW^{3/4}$ scaling on pharmacokinetic grounds. The analysis supports that $BW^{3/4}$ scaling tends to approximate the use of PBPK models in extrapolation, assuming that AUC is the best summary of internal dose. Results from $BW^{3/4}$ scaling were compared to model-predicted AUC concentrations for mice, rats, and humans, for 12 "predominantly volatile and lipophilic" chemicals.[10] A likely basis for including these chemicals in this analysis was availability of PBPK models for mammals involved in health assessments. It may be noted that some of the substances evaluated are encountered in ecological assessments, but the substances studied were not selected to represent those of interest for ecological assessors. The analysis indicates variation across chemicals in how well scaling would approximate the use of current PBPK models to predict AUC concentrations. For example, methylmercury is an example where the approximation is comparatively poor. This analysis suggests that as a rule $BW^{3/4}$ scaling approximates the use of PBPK models which, it should be noted, may loosely follow allometric relationships. Additional analysis is desirable to explore what PK properties of the chemicals reflect greater or lesser success in use of allometric scaling.

Kirman et al. (2003) considered whether the PBPK-based ratios of AUCs comparing pairs of species were consistent with expectations based on $BW^{3/4}$ scaling. Another way to express the

[10] Benzene, ethanol, styrene, carbon tetrachloride, ethylene oxide, tetrachloroethylene, chloroform, methylene chloride, trichloroethylene, diisopropyl fluorophosphates, methylmercury, and vinyl chloride.

results is to calculate an "effective allometric exponent" *b* such that ratios of body weights, scaled using the exponent computed, are equal to PBPK-predicted AUC (see Appendix C). Both sets of computations - those of Kirman et al. along with the effective exponents - are displayed in Appendix C and appear to support default $BW^{3/4}$ scaling as a PK adjustment, particularly preferred over extrapolation on a simple body weight basis (under assumptions of the analysis).

Some support for allometric scaling in human health assessments comes from data for humans and animals on acute toxicity of anti-neoplastic drugs. Allometrically-scaled rodent LD10 measurements are found to be supportable as estimates of human sublethal effects (Travis and White, 1988). A point of interest for purposes of this section is that tissue AUCs have been held to be useful dose summaries for these chemicals. It has been observed, however, that such chemicals are not typical of wildlife assessments – see Allard et al. (2009).

4. SENSITIVITY OF HAZARD ASSESSMENT TO BODY WEIGHT SCALING DECISIONS

To provide some sense of how much difference $BW^{3/4}$ scaling could make, Table 1 displays information on sensitivity to relative body weights (test species versus assessment species), assuming that toxicity and exposure information are combined into a hazard ratio[11]. The table gives examples of ratios of species body weights, associated with different multiplicative factors for adjustment of toxicity measurements when using the $BW^{3/4}$ approach.

The first example illustrates extrapolation from of a larger, roughly rat-sized species to a smaller species, roughly mouse-sized. With BW^1 scaling we would assume the same toxicity value (on a body weight basis) in the assessment species as in the test species. With $BW^{3/4}$ scaling the toxicity value from the test species is multiplied by a factor of 2, i.e., allometric scaling results in the assessment species being judged less sensitive than if extrapolation had been based on BW^1. The second example illustrates extrapolation from a smaller (mouse-sized) to larger (skunk-sized) species. Now, the assessment species is judged more sensitive (lower toxicity

[11] Hazard ratio, or HR, is the estimated environmental or tissue concentration divided by the toxic concentration.

value after scaling) than with BW^1 scaling. Indeed, the toxicity value from the test species is divided by 3.

(Of course, "smaller-to-larger" extrapolation is the general rule in extrapolating toxicity for human health assessment, so that $BW^{3/4}$ scaling is generally more health-protective than BW^1 scaling in that situation. Ecotoxicological extrapolations can be smaller-to-larger or larger-to-smaller.)

More precisely, a toxicity value (e.g., a dose estimated to have no detectable effect, a stipulated magnitude of mortality, etc.) would be "adjusted" (multiplied or divided) by a factor of 2.0 (2-digit accuracy) if the ratio of species body weights (larger / smaller) is in the range 14.5–17.7, or a factor of 3.0 if such a ratio is in the range 75.7–86.5. (Multiply the toxicity value by such a factor when extrapolation is larger-to-smaller, divide by the factor when the extrapolation is smaller-to-larger.)

Table 1. Sensitivity of Hazard Assessment to Allometric Scaling of Toxicity Values[a]

Adjustment Factor Applied to Measurement	Ratio of Species Body Weights (Test/Assessment)	Example Extrapolation	
		Test Species	Assessment Species
2.0	15	300-g rat	20-g deer mouse
1 / 3.0	1 / 86	35-g mouse	3,000-g striped skunk

[a] Example (Row 1): The body weight ratio is 15 = 300/20, and the adjustment factor applied to test-species measurement (on a BW basis) based on $BW^{3/4}$ scaling is $15^{1/4}$ or approximately 2.

5. RECOMMENDED DEFAULT SCALING PROCEDURES

The evidence available is held to support the following recommendations.

A general recommendation is to recognize that nonlinear body-size effects are likely in biological data, toxicological or otherwise, based on species with widely different typical body weights. Examples that are likely to be of toxicological significance are allometry in rates of uptake or clearance of toxicants, with rates tending to be slower relative to body weight in larger species. Scaling relationships for toxicity measurements reflect the combined effect of

scaling relationships involving multiple biological processes that are expected to determine species toxicity.

For purposes of development of specific recommendations for terrestrial wildlife, most toxicity measurements can be classified as single-dose measurements designed to measure lethality (generally as an LD50), repeated-dose, or "dietary" dose levels from longer-term experiments. (Repeated dose and dietary dose levels may be interconverted with assumptions particularly regarding ingestion rate, but the distinction is considered important at least in pesticide assessment.) Repeated-dose studies may be lethal by design (e.g., a feed study used to estimate the LC50). Dose-related lethality may sometimes be observed in studies not designed to be lethal.

The proposed default allometric exponent for single dose LD50s is b = 1 (extrapolation on a simple body weight basis). The primary support for this recommendation is the analyses of Rhomberg and Wolff (1998) and Sample and Arenal (1999), using single-dose LD50 studies. The same is proposed for any dose-based study (single or repeated dose) with substantial lethality (about 50% or greater).

The proposed default for toxicity extrapolation for repeated-dose studies with limited or no lethality is b = 3/4. Similar assumptions lead to a default *no allometric scaling* for "dietary" (or food based") measurements (i.e., measurements reported as toxicant concentrations in food). It may be noted that toxicity results reported as a food dose rate (e.g., mg/kg-d) will generally derive from a food concentration, and so there is an argument for avoiding explicit allometric scaling, but to use the food concentration without conversion and directly assume that equal concentrations are equipotent across species (Sample et al., 2014). Exceptions would include where an extrapolation would be between species not considered to be comparable with any available adjustment. A possible example would be if a test species is monogastric, and the receptor species is a ruminant that consumes and ferments a high volume of plant material.

The U.S. EPA (2011) proposes that the rationale for 3/4 scaling applies when physiological processes function in a similar way with exposure as without exposure. This suggests that the

rationale for such scaling is inapplicable at some level of lethality. (In particular we do not propose b = 0.75 for LD50s.) However, limited lethality does not seem to disqualify doses from scaling with b = 0.75: The latter has in fact been supported by U.S. EPA (2011) based on correlation of rodent LD10 measurements to human doses that were chosen to be generally non-lethal (Travis and White, 1988).

When low levels of lethality are observed only after a week or more in a repeated-exposure scenario, this may be taken as an indication that clearance mechanisms have had enough time to act to reduce toxic effects to some degree, and that $BW^{3/4}$ adjustment or use of a dietary approach may be appropriate. (In discussing single- versus repeated-dosing, Rhomberg and Wolff (1998) discuss the concepts of standing levels and rates of regeneration of toxicological defenses.) Simple PK models are potentially of use in interpreting acute and subacute response data. Fischer (2005), in an ecological context, extrapolated a half-life across species; however the U.S. EPA (2014) discussion of the extrapolation of PK parameters, in a human health context, needs to be taken into account.

It is possible that the biological basis for BW^1 extrapolation of highly lethal doses is applicable as well to some sublethal effects associated with single doses. However, U.S. EPA (2011) generally favors $BW^{3/4}$ for sublethal acute effects.

For dietary sublethal effects it may be unnecessary to account for allometry explicitly (e.g., by converting measurements to dose rates such as mg/kg-day and then applying allometric scaling). No theoretical basis was encountered for a general body-weight dependency of average tissue concentrations on body weight for species exposed to similar concentrations in feed. Thus, as a rule, a critical dietary concentration determined for a test species can be assumed for the assessment species as well. A set of assumptions that appear to support this approach include that internal doses are appropriately summarized by cumulation or averaging, and that rates of assimilation (ingestion plus absorption) and clearance scale in the same way relative to body weight (e.g., as $BW^{3/4}$). Then average internal dose does not change systematically with body size because the allometric effects on assimilation and clearance effectively cancel (Appendices A & D; O'Flaherty, 1989; U.S. EPA, 2005, 3-7). Note that this

argument would not justify extrapolations between monogastric species and species that consume and ferment plant material. These may generally ingest at a high rate relative to body size, and the effect of fermentation on the toxicant would also need to be considered. Probably no simple extrapolations to such species from monogastric species will be biologically defensible.

See Box 1 for a summary of recommended defaults for use of allometric scaling for characterizing toxicity in ecological risk assessments.

Box 1. Recommended Defaults for Use of Allometric Scaling for Characterizing Toxicity in Ecological Risk Assessments. Note: Recommendations are subject to a principle of using as much of the available science as possible in a given situation.

- The default for single dose LD50, or when the duration of dosing before toxicity is observed is less than a chemical's half-life, is $b = 1$ (extrapolation on a simple body weight basis). The same is proposed for any dose-based study (single or repeated dose) with substantial lethality (about 50% or greater).

- The default for toxicity extrapolation for repeated-dose studies, where toxicity is only observed after five half-lives, with units reported as a dose rate such as mg/kg-day, and limited or no lethality is $b = 3/4$.

- The default for a "dietary" or ("food based") toxicity, reported and applied as toxicant concentration in feed, is no allometric scaling.

Arguments for $BW^{3/4}$ scaling as the most common default for sublethal doses include the following: 1) validation with data on sublethal acute toxicity of anticancer drugs (Travis and White, 1988; U.S. EPA, 2011); 2) comparability of $BW^{3/4}$ scaling to application of PBPK-based AUC estimates (Kirman et al., 2003; Appendix C); 3) a tendency – with various exceptions – for allometric exponents to be less than 1 relating clearance to body weight across species, in

compilations such as those of Chiou et al. (1998), Huang et al., (2015), and Tang and Mayersohn (2005); and 4) an apparent, current viewpoint in veterinary medicine that a BW^1 default is particularly unsafe as a general rule (Hunter and Isaza, 2008). (With regard to the final point, however, neither $BW^{3/4}$ scaling nor any other point prediction method is assumed to be safe, in view of the uncertainties. The best that can be done is to attempt to use all available scientific information and recognize the remaining uncertainty. Limitations of these lines of evidence are explored further in the discussion.) These arguments rely to some degree on data involving substances not likely to be subjects of ecological risk assessment; however, some consideration may be given to use of information on such substances for purposes of recognizing general biological patterns.

Risk assessment computations are somewhat simplified if toxicity measurements can be used in a sense "as reported" (without transformations). However, the implications are different for different types of toxicity measurements, depending in particular on whether or not the toxicity is in the form of dose relative to body weight. An "as-reported" use of a dose-based measurement would be to assume that the same dose can apply to test and assessment species if both are expressed on a BW basis. An as-reported use of dietary toxicity can assume that the same critical concentration in feed applies to both test and assessment species. These approaches are similarly appealing with respect to simplicity but are supported by different assumptions and have different implications in practice.

Acceptable implementations of allometric scaling could involve triggers based on relative body weights, comparing test and assessment species. If allometric scaling is considered burdensome relative to value added, such scaling could be implemented when the ratio of body weights (assessment species/test species) exceeds a specified threshold. This approach is not pursued further here, and specific cutoff values based on ratios of body weights are not proposed, but sensitivity analyses like those in the preceding section would be helpful if such an approach is pursued.

Cross-species toxicity extrapolation depends on the identification of groups of species and substances such that extrapolation is to be allowed within but not among groups.

Extrapolation between conventional vertebrate classes (amphibian, reptiles, birds, mammals) has been discouraged (Allard et al., 2009). Here, no effort has been made to develop guidelines on whether or not species are too different for extrapolations to be allowed. Additional criteria for appropriate classification may be based on classification or phylogeny, or on physiological or behavioral traits, such as homeo-/heterothermy, carnivory/herbivory, ingestion of infrequent, large meals by snakes, concentration of urine in arid conditions, seasonal dormancy, long-distance migration of birds, differences in renal physiology, employment of fermentation of plant material by some species but not others, and so on. Further elaboration of allometric methods to provide distinct criteria for more and smaller taxa, or other species groups, would depend on having enough data for each group to support a specific scaling rule. Obtaining toxicity data representing vertebrate classes may be difficult enough, and there could only be greater difficulty in obtaining adequate data for more and smaller groups of species.

6. DISCUSSION AND FUTURE WORK WITH AN EMPHASIS ON UNCERTAINTIES IN EXTRAPOLATION

Assessment methodological development seeks to optimize the use of scientific information, with recognition of uncertainties, allowing that the timely development of assessments requires precise, practical guidance. There is much that can be done to further the development of methods that meet such objectives related to body weight scaling. In the remainder of this section, we recognize some important themes related to these broad objectives.

Using as Much Scientific Information as Practical, and Allowing for Policy. Recommendations presented here on scaling toxicity measurements are subordinate to a broader principle of attempting to fully utilize the science available in a given situation. The applicable science may suggest a specific approach for scaling toxicity measurements, or an approach that does not involve scaling. In some situations, the information might support a simple modeling approach. The information needed will relate to combinations of species and substances; however, it is not feasible here to identify information likely to be available for every important combination.

An important objective has been to clarify the PK basis of $BW^{3/4}$ scaling as promulgated by U.S. EPA (2011) for human health assessments. A central idea is the use of $BW^{3/4}$ as a surrogate for systemic clearance (CL). (Technically, ratios of $BW^{3/4}$ serve as surrogates for ratios of corresponding CL.) The available science could suggest other surrogates. A ratio of glomerular filtration rates might be considered as a surrogate for the ratio of species CL (Lin, 1995).

Some authors report allometric results for blood half-lives instead of CL, considering the former to scale more predictably with body weight (Riviere et al., 1997; Antonissen et al., 2015). However, U.S. EPA (2014) states that "Half-life is not an acceptable basis for" calculation of data-derived extrapolation factors (DDEFs). Therefore, blood half-life should be combined with the volume of distribution to yield CL before use in scaling. (Volume of distribution may be represented with an uncertainty distribution, if desired – see discussion of quantitative uncertainty evaluations.)

Echoing U.S. EPA (2011), it is understood that alternatives to the defaults proposed may be appropriate for particular regulatory entities, based on science or policy, for specific situations. For example, an entity that regulates pesticides may determine that there is data of sufficient quality and quantity to justify routine scaling of toxicity estimates, for some pesticides and receptors, with an allometric exponent developed for the specific situations. An example proposed for avian risk assessments is Mineau et al. (1996). Regulatory entities may choose to develop tiered assessment schemes, in which higher-tier assessments require more data.

Relevant Species and Substances for Developing Ecological Assessment Methods. Species used to evaluate human health effects are a biased sample for ecological assessment purposes. For example, these species are generally not large predators. However, we advocate development of comparative frameworks that address variation across species and substances, without automatically excluding species commonly involved in human health. FARAD (FARAD.org) is a source of PK information on domesticated mammal species, some of which could be relevant to wildlife receptors in a given assessment context (also see Martinez et al., 2006; Mahmood et al., 2006). We have not encountered arguments that domestication as such reduces the relevance of a species for wildlife assessments.

The data on PK is fairly rich for pharmaceuticals, but the relevance to ecological assessments of such data is open to some question. Data on toxicity of anti-neoplastic drugs has played a significant role in development of the human health assessment approach of U.S. EPA (2011). Allometric scaling of toxicity of such chemicals is not necessarily limited to mammals: Antonissen et al. (2015) is an example where a blood half-life for an anti-neoplastic substance scales allometrically in birds (based on applications to four species including a passerine). The questions are, what properties of these substances account for the reported success in scaling their biological properties? And are those properties also characteristic of some substances of concern for ecological assessment?

Risk assessors are understood to require guidance that is transparent and practical. A helpful development might be a list of qualitative "indicators for particular applicability of $BW^{3/4}$ (or BW^{1}) scaling," easy for assessors to use. These indicators may be based largely on intrinsic properties such as lipophilicity that are easily measured in vivo. Criteria based on in vivo PK parameters would be relevant as well, but data for such parameters may often be more difficult to obtain than information on intrinsic properties considering the variety of ecological receptors. It is very important for such information to be brought to bear in the identification of the most appropriate test species to use in a given assessment. Decision trees seem to be viewed favorably by assessors and may be useful for this purpose.

We think there would be consensus on the value of organizing comparative information in a form useful to assessors. This could take the form of groups of species and substances that can be handled in similar ways in an assessment. PBPK modeling, though perhaps infeasible for most specific assessment situations, could have a role in combination with classifications of species and substances if the models suggest that a simpler approach provides an adequate approximation for some groups defined by the classification. The properties of the chemicals considered by Kirman et al. (2003) should be correlated to how well $BW^{3/4}$ scaling approximates PBPK results. For purposes of developing comparative frameworks, we encourage information exchange between human health assessors, ecological risk assessors, and veterinarians.

Extreme phylogenetic extrapolations. Additional guidance is desirable on extrapolations that are to be considered extreme because of involving species with profoundly different biology. Important biological differences between vertebrate groups include the absence of a renal portal system in mammals and major differences between groups in the regulation of body temperature. Birds and mammals are as a rule endotherms, other vertebrates ectotherms. (A related distinction is between homeothermy, poikilothermy, and heterothermy.)

We do not dispute a recommendation (Allard et al., 2009) to avoid extrapolations across vertebrate classes. In view of the difficulty of such extrapolations, they should perhaps be treated as a special topic. However, some preliminary remarks are offered. Criteria for protection of aquatic life are potentially protective of aquatic life stages of amphibia (a possibility that is not reviewed here). Special allometric procedures have been advocated for extrapolating pharmaceutical doses from mammals to reptiles or amphibians (Hunter, 2010). The methodology involves, in addition to a factor based on body weights, multiplicative factors for each of several groups based on tendencies for species in some groups to have higher metabolic rates than similar-size species in other groups (e.g., generally higher metabolic rates for endotherms than for similar-sized ectotherms, higher metabolic rates for passerines than non-passerines). We note, however, that a current phylogenetic framework for animals (tolweb.org) places birds with reptiles. Thus, consideration might be given to a unified PK framework for reptiles and birds. In fact, some similarities relevant to pharmacokinetics are discussed in Hunter (2010). PBPK modeling is expected to be relatively well developed for poultry, at least for pharmaceuticals.

Some Areas of Uncertainty

***Metabolic elimination of toxins*.** Any aspect of absorption, distribution, metabolism or excretion (ADME) may contribute to uncertainty in a given assessment context. Species variation in metabolic elimination of toxins is a biological factor that may not be handled well by allometric methods (e.g., Hutchinson et al., 2014). PBPK modeling as used in human health assessment frequently assumes $BW^{3/4}$ scaling of liver perfusion and metabolic rates are typically expressed allometrically (e.g., metabolic Vmax = $VmaxC*BW^{3/4}$) although the allometric

coefficient (VmaxC) is not kept constant between species. These assumptions result in some tendency for systemic clearance to scale in a similar way. A parameter of interest is intrinsic clearance, which relates the rate of metabolism to the concentration of the parent at the site of metabolism. To evaluate species variation, efforts should be made to make use of information on enzyme activity, e.g., P450 (see Head and Kennedy, 2010; Manning et al., 2013). It may be noted that differences in metabolism related to diet have been reported among populations of a single species (Malenke et al., 2012).

The role of metabolism in determining toxicity to wildlife is reviewed by Hutchinson et al. (2014). In particular they state that:

> "The essential purpose of xenobiotic metabolism is to convert lipid-soluble, non-polar and non-excretable chemicals into water soluble, polar molecules that are readily excreted. ... wildlife species with low metabolic competency may exhibit *zero-order* metabolic (pharmacokinetic) profiles and thus high API [Active Pharmaceutical Ingredient] toxicity, as in the case of diclofenac and the dramatic decline of vulture populations across the Indian subcontinent. A similar threat looms for African Cape Griffon vultures exposed to ketoprofen and meloxicam, recent studies indicating toxicity relates to zero-order metabolism [some technical detail redacted]. While all aspects of ADMET [ADME + Toxicity] are important in toxicity evaluations, these observations demonstrate the importance of methods for predicting API comparative metabolism as a central part of environmental risk assessment." [Italics and bracketed parentheticals added.]

***Trophic ecology as an illustration of uncertainty*.** Two relatively extreme types of diets seem illustrative, namely hyper-predation (e.g., cats), and low-nutrient plant diets.

Low nutrient plant diets are often associated with gut fermentation, which occurs in the foregut for some species and in the hind gut for others. Stevens and Hume (2004) provide a general, comparative account of vertebrate digestive physiology. Use of the best-justified allometric scaling of an oral toxicity measurement is not expected to reduce uncertainty appreciably in an extrapolation between species that do and do not use fermentation. However, we cannot say that clearance of *intravenous* doses, if available, could not be extrapolated on a $BW^{3/4}$ basis. As usual, we underline that the role of scaling is determined by the scientific information relevant to the context.

For mammals with a diet essentially entirely of meat Shrestha et al. (2011) report a tendency towards genetic loss of capability for metabolizing some toxic substances, consistent with low dietary exposure to plant secondary compounds in some evolutionary lineages. (Incidentally we note the use by these authors of reconstructed phylogenies rather than taxonomy in describing evolutionary loss of metabolic function.)

***Body size and gastro-intestinal physiology*.** In developing allometric expressions, it is helpful (other things being equal) to use species with a wide range of body sizes. In general, datasets may be sensitive to the largest-bodied species that have been studied. The largest terrestrial vertebrates will be mostly ruminants that consume and ferment plant material at a high rate. The largest monogastric species tested may be carnivores (most often dogs). There is a substantial amount of veterinary literature on extrapolations of pharmaceutical doses involving large species (Mahmood et al., 2006; Martinez et al., 2006, 2009; Hunter and Isaza, 2008; Sedgwick, 1993). FARAD.org may be a source of relevant PK information.

Additional Uncertainties and General Remarks. There are of course many uncertainties in the use of laboratory toxicity measurements. While a thorough treatment is beyond the scope here, it is well to keep these in mind. Acute lethality is often evaluated using gavage. Relevance of the results to exposure by ingestion with food is obviously a difficult issue. For dietary studies, it may be difficult to state exactly the ingestion rate of toxicant. Some feed will be spilled, and high feed concentrations may elicit aversive responses.

The best summary of internal dose over time might be the subject of further, useful study. The strongest argument for $BW^{3/4}$ as a PK adjustment appears to assume that the most appropriate internal dose metric is inversely proportional to $BW^{3/4}$. This occurs using AUC as the internal dose metric when elimination scales as $BW^{3/4}$. However, C_{max} cannot be excluded as the most biologically relevant summary. The ideal summary of concentration over time could be something different from either of these.

Some literature (e.g., Huang et al., 2014) is disparaging towards adoption of any default allometric approach, whether with exponent 3/4 or some other value, on grounds that not enough biology would be considered and that exponents are variable. Nonetheless,

extrapolation of toxicity across species is often necessary, and even the use of BW^1 involves an assumption about the role of body weight. A comprehensive account of extrapolation uncertainty and variation in species sensitivity will not be attempted here.

Differences in species sensitivity may relate to diet (with associated differences in physiology, especially gastro-intestinal), toxicodynamics, and toxicokinetics. Pharmacokinetic (PK) considerations are sometimes classified as ADME (absorption, distribution, metabolism, and excretion). Components of PK that can impact species extrapolation are differences in binding to molecular components (a factor in *distribution*), and differences in *metabolism* (discussed in more detail below). These may interact in complex ways, e.g., transformation of a substance to compounds that may or may not be as readily excreted as the parent. Calabrese (1991) provides a general discussion. However, despite of the variety of ADME factors that may be important, an analysis of clearance (CL) for 115 xenobiotic substances report generally better support in mammals for $BW^{3/4}$ than BW^1 scaling across four categories of substances, (i) proteins, (ii) compounds eliminated mainly by renal excretion, (iii) compounds eliminated by metabolism, or (iv) compounds eliminated by renal excretion and metabolism combined (Hu and Hayton, 2001). All subgroups except (ii) showed a *b* value statistically "not different from 0.75." For group (ii) the average coefficient "was 0.65, which differed from 0.75 but not from 0.67". (However, we may note that dependence of allometry on mechanisms of elimination may be described at different levels of granularity by different biologists. For example, Walton et al. (2004) summarize allometric effects for subcategories of renal excretion among mammals for purposes of extrapolation to humans.[12]; references alluded to in footnote are in Walton et al. (2004)).

[12] "The prediction of kinetics in humans using allometric scaling of data from a range of animal species has been successful for a number of compounds that are eliminated largely unchanged in the urine (e.g. [references for 5 compounds]), and for the renal clearance aspect of compounds eliminated by both metabolism and excretion ([references for 2 compounds]). However, allometric scaling has been less successful when the compound undergoes active transport in the kidney (e.g. [reference for napsagatran]) or is extensively bound to plasma proteins (e.g. [references for 2 compounds]). Mahmood (1998) analysed the results of allometric scaling of renal clearance for eight drugs and concluded that renal clearance in humans would be under-predicted for drugs cleared largely by tubular secretion."

Species variation in sensitivity to a substance is expected to be determined in part by variation in aspects of species biology that determine pharmacokinetics. Extensive information exists on species variation in PK parameters for pharmaceuticals (e.g., Mahmood, 2005). Situations involving interspecies extrapolation of pharmaceutical effects (efficacy or toxicity) include first uses of a new drug in humans and veterinary pharmacology.

When adequate data is available it is better to estimate an allometric exponent for PK extrapolation from the data than to use a fixed allometric exponent such as 1 or 3/4. However, use of the most appropriate allometric approach does not guarantee confident predictions for every substance. For example, an analysis by Riviere et al. (1997) of comparative pharmacokinetics of 44 drugs noted a large spread of individual values around the regression line. (For a statistic to represent this residual spread in regression, a common choice is the R^2 statistic.)

We note that in collections of PK data representing multiple substances, estimates of the exponent are variable and sometimes far from 0.75 but still cluster closer to 0.75 than to 1 (e.g., Tang et al., 2007 Table 2, for i.v. administrations). Huang et al. (2015) found a mean of 0.82 for serum CL for 85 drugs. On the one hand this variation suggest that extrapolated point estimates are highly uncertain and could pose unacceptable risk, particularly from a standpoint of first use in humans. At the same time, such results point to the possibility of the use of uncertainty distributions or uncertainty factors, rather than point estimates only. The central tendency (e.g., median) of an appropriate distribution might be 0.75, 1, or some other value. A distribution centered on 0.75 could assign appreciable probability to a value of 1 or lower. It is understood that options for collecting additional data may be limited in a site assessment context. However, it is not possible to say with confidence what data may be available in the future.

Qualitative and Quantitative Uncertainty Evaluations. No one expects that every important biological difference between species will be addressed by scaling of oral toxicities. Where an extrapolation on a simple body weight basis would not be viewed as plausible because of a likelihood of profound biological differences between species, the extrapolation based on

changing the exponent to 3/4 would not, as a rule, be plausible either. Therefore, an initial step is to decide whether an extrapolation between a given pair of species is plausible at all.

Assuming that an extrapolation is plausible for a given pair of species, but some likely biological differences are identified, a subsequent question can be, *can species differences be quantified with some proportionality constant (factor) in an expression relating internal dose (e.g., AUC) to exposure?* For example, absorption is customarily quantified as fraction of ingested dose that is absorbed. For pharmaceuticals, some factors affecting absorption of pharmaceuticals have been reviewed by Kararli (1989). If the biological factor can be represented this simply, we only need information on the ratio of the proportionality constant for the two species. (We do not need the two specific values of the constant for each species.) Consideration might be given to a quantitative uncertainty approach (uncertainty factor or distribution). The median should be 1 if there is no evidence to suggest which species has the larger value of the constant. A likely "first cut" for an uncertainty distribution, when the variable is positive and there is no clear upper bound, is a lognormal, with allometric assumptions readily incorporated (e.g., Chiu and Slob, 2015). In particular Monte Carlo simulation is not required for uncertainties for products of random variables modeled as lognormal.

If species sensitivity differences cannot be expressed this simply, then a more complicated approach might be considered. Covering possible models is beyond the scope of this document (but see Fischer, 2005). In particular, at this time we have not investigated whether there is some practical modeling approach for using information on intrinsic clearance (which reflects the metabolic rate at a cellular level).

Data and Methods for Developing Empirical Scaling Expressions. We have no grounds for dismissing any analyses cited here, that have estimated allometric exponents by regressing toxicity measurements on body weight across species. The evidence can be strengthened if a purported difference among groups can be supported by biological arguments.

We observe that guidance for such analyses could relate to quality of toxicity data, appropriate body weights (e.g., possible use of species default body weights), handling of cases of multiple studies per substance and species, number of species, appropriate variety of species (e.g.,

range of body sizes, phylogenetic representation, diversity of physiology), meaningful groupings of species based on life history and/or phylogeny, meaningful groupings of substances, criteria for deciding if allometric differences are similar or different among groups, and statistics (e.g., mean or median) for summarizing results for groups of chemicals or species.

The allometric exponent is usually estimated by ordinary bivariate linear regression, relating the log of the dependent variable to the log of body weight. The assumption of no error in body weights is probably acceptable for applications with species body weights ranging over several orders of magnitude (discussed, for example, by Kilmer and Rodriguez, 2017). However, if the body weights are largely at one extreme (say, most of the species included are small), then results may be very sensitive to measurements for a few species representing the other extreme (say, large species). Ideally, both extremes will be represented by biologically and taxonomically diverse species.

Over-representation of some taxa can in principle be handled with phylogenetic generalized least squares (e.g., Smears and Rohlf, 2016), an essentially straightforward extension of linear mixed modeling. (However, this methodology may not yet be very well known among ecological assessors.) Regression results without such adjustment assume independent observations and are expected to overstate the information in the data to some degree, so that it may be too easy to find differences among groups (in statistical terms, to make Type I errors).

Phylogenetic distance may serve as a surrogate for unrecognized physiological differences among species. Standard taxonomy does not necessarily provide a good reflection of phylogeny. Several uses of reconstructed phylogenies were encountered in the course of preparing this review (Bakken et al., 2004; Shrestha et al., 2011; White et al., 2009). Additional development would be required if such methods are to be practical for ecotoxicologists.

Some investigation may be given to other statistical methods that 1) can address multiple predictors of a parameter of interest such as CL, not only body weight (i.e., multivariate prediction methods), or 2) emphasize accurate prediction. In particular, meta-analyses of allometric exponents computed for multiple species and substances are subject to decisions on "lumping versus splitting." Groups with too few species may not provide enough data for a

reliable summary (e.g., median), while groups that are too large may be so heterogeneous that some species are not accurately represented by a group summary. In the context of predictive accuracy, this is recognized as a tradeoff between bias (with groups too large) and variance (with groups too small), or as a problem of underfitting versus overfitting (as discussed, for example, by Efron and Hastie, 2016). Statistical decision tree methods (e.g., Anderson et al., 2014) will result in a good balance of parsimony and fit, with an objective of accurate prediction. To compare a few alternative groupings, the Akaike Information Criterion may be considered.

The use of quantitative uncertainty distributions (e.g., with Monte Carlo) can assign greater weight to the exponent (0.75 or 1 etc.) that is considered most plausible. Bayesian methods in particular can be used for developing uncertainty distributions, allowing rigorously for values other than a default, where supported by data. A Bayes prior distribution for the exponent may be centered on the single value considered *a priori* most plausible. The uncertainty in the exponent, considering the data, would be expressed in a posterior distribution for the exponent. Such a specialized approach might need to be restricted to higher-tier assessments.

For data on chronic effects, attention may focus largely on reproductive effects. Multiple endpoints may be available from a given study. Studies may report a NOAEL, which may be based on different endpoints in different studies. It may be preferable to focus on a single endpoint and estimate an effective exposure for each study by statistical curve fitting. Perhaps the ideal endpoint for such analyses would have high ecological significance, as well as being amenable to curve fitting.

Model Validation. If current practices for extrapolation of toxicity values are viewed as validated, then the validation criteria should be stated and used to compare alternative extrapolation approaches. Assuming that toxicity values will be extrapolated on some basis, recourse is to the approach with the strongest scientific basis. The current validation status of BW^1 extrapolation seems dubious. In a zoological pharmacological context, Hunter and Isaza (2008) treat BW^1 as a form of extrapolation, in fact a risky one:

> "From practical experience, many drugs do not have a simple linear relationship relative to weight. At the extremes of the weight range, this method tends to *overdose large animals and underdose small animals,* which may be very clinically significant. Again, the simplicity of the calculation tempts many practitioners to use this *potentially dangerous method of extrapolation* without consideration of the consequences. In fact, many *clinicians who read dosage recommendations from formularies are not even aware that an extrapolation is being made* nor the risks of the associated assumptions." [Italics added.]

A famous example of over-dosing a large species, based on a BW^1 extrapolation, is West et al., 1962).

Ecological assessment, as a discipline, may define validation criteria that are meaningful and feasible with data likely to be available. Validating that a risk assessment methodology with a particular approach to toxicity data will protect a species in the field is beyond the scope of our discussion, considering that the assessment will contain other components, e.g., exposure assessment, that are uncertain (and outside the scope of our discussion). More focused validation would relate to the efficacy of scaling as, specifically, a PK adjustment. We suggest that there may be some role for exposure biomarkers if available, to provide a limited validation. These could point to species differences in internal exposure not accounted for by the suggested $BW^{3/4}$ scaling of toxicity values.

It may be helpful to identify validation exercises based on predictions of PBPK models. An impressive validation of a mechanistic (e.g., PBPK) model would be if it is shown to provide accurate predictions when applied in conditions other than those where the model was initially developed, e.g., a different time pattern of dosing, without re-optimizing parameters or complicating the model (e.g., adding more compartments). A simple approach such as $BW^{3/4}$ scaling of toxicity would have a degree of validation, for some combinations of species and substances, if shown to be consistent with a validated PBPK representing those combinations.

7. REFERENCES

Allard, P., Fairbrother, A., Hope, B., Hull, R., Johnson, M., Kapustka, L., Mann, G., McDonald, B., and Sample, B. (2009) Recommendations for the Development and Application of Wildlife Toxicity Reference Values. Integr Env Assess and Man 6:28-37.

Anderson, R.H., Farrar, D.B., and Zodrow, J.M. (2013) Terrestrial metals bioavailability: A comprehensive review and literature-derived decision rule for ecological risk assessment. Human and Ecological Risk Assessment: An International Journal 19(6):1488-1513.

Antonissen, G., Devreese, M., De Baere, S., Hellebuyck, T., Van de Maele, I., Rouffaer, L., Stemkens, H.J., De Backer, P., Martel, A., and Croubels, S. (2015) Comparative pharmacokinetics and allometric scaling of carboplatin in different avian species. PLoS ONE 10(7): e0134177. https://doi.org/10.1371/journal.pone.0134177.

Bakken, B.H., McWhorter, T.J., Tsahar, E., and del Rio, C.M. (2004) Hummingbirds arrest their kidneys at night: diel variation in glomerular filtration rate in *Selasphorus platycercus.* J Exper Biol 207:4383-4391.

Boxenbaum, H. (1980) Interspecies variation in liver weight, hepatic blood flow, and antipyrine intrinsic clearance: extrapolation of data to benzodiazepines and phenytoin. J Pharmacokinet Biopharm 8:165-176.

Boxenbaum, H., and DiLea, C. (1995) First-time-in-human dose selection: allometric thoughts and perspectives. J Clin Pharmacol 35:957–966.

Burzala-Kowalczyk, L., and Jongbloed, G. (2011) Allometric Scaling: Analysis of LD50 Data. Risk Anal 31:523-532.

Calabrese, E.J. (1991). Principles of Animal Extrapolation. Michigan: Ed. Lewis Publishers.

Chiou, W. L., Robbie, G., Chung, S. M., Wu, T. C., and Ma, C. (1998) Correlation of plasma clearance of 54 extensively metabolized drugs between humans and rats: mean allometric coefficient of 0.66. Pharm Res 15:1474-1479.

Chiu, W., and Slob. (2015) A unified probabilistic framework for dose–response assessment of human health effects. Environmental Health Perspectives 123:1241-1254.

Edwards, N.A. (1975) Scaling of renal functions in mammals. Comp Biochem Physiol 52A:63-66.

Efron, B., and Hastie, T. (2016) Computer Age Statistical Inference: Algorithms, Evidence, and Data Science. Cambridge University Press.

Fischer, D.L. (2005) Accounting for differing exposure patterns between laboratory tests and the field in the assessment of long-term risks of pesticides to terrestrial vertebrates. Ecotoxicology 14(8):853-862.

Freireich, E.J., Gehan, E.A., Rall, D.P., Schmidt, L.H., and Skipper, H.E. (1966) Quantitative comparison of toxicity of anticancer agents in mouse, rat, hamster, dog, monkey, and man. Cancer Chemother Rep 50:219-224.

Head, J.A., and Kennedy, S.W. (2010) Correlation between an in vitro and an in vivo measure of dioxin sensitivity in birds. Ecotoxicology 19(2):377-382.

Hu, T-M, and Hayton, W.L. (2001) Allometric scaling of xenobiotic clearance: Uncertainty versus universality. AAPS PharmSci 3(4):30-43.

Huang, Q., and Riviere, J.E. (2014) The application of allometric scaling principles to predict pharmacokinetic parameters across species. Expert Opinions Drug Metab Toxicol 10:1241-1253.

Huang, Q., Gehring, R., Tell, L.A., Li, M., and Riviere, J.E. (2015) Interspecies allometric meta-analysis of the comparative pharmacokinetics of 85 drugs across veterinary and laboratory animal species. J Vet Pharmacol Therap 38:214-226.

Hunter, R.P., and Isaza, R. (2008) Concepts and issues with interspecies scaling in zoological pharmacology. J Zoo Wildlife Med 39:517-526.

Hunter, R.P. (2010) Interspecies Allometric Scaling. In: Cunningham F., Elliott J., Lees P. (eds.) Comparative and Veterinary Pharmacology, Handbook of Experimental Pharmacology, vol 199. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-10324-7_6.

Hutchinson, T.H., Madden, J.C., Naidoo, V., and Walker, C.H. (2014) Comparative metabolism as a key driver of wildlife species sensitivity to human and veterinary pharmaceuticals. Phil Trans R Soc B 369:20130583. https://dx.doi.org/10.1098/rstb.2013.0583.

Kararli, T.T. (1989) Gastrointestinal absorption of drugs. Crit Rev Therap Drug Carr Sys 6: 39-86.

Kilmer, J.T., and Rodríguez, R.L. (2017) Ordinary least squares regression is indicated for studies of allometry. J Evol Biol 30(1):4-12. https://doi.org/10.1111/jeb.12986.

Kirman, C., Sweeney, L., Meek, M., and Gargas, M. (2003) Assessing the dose-dependency of allometric scaling performance using physiologically based pharmacokinetic modeling. Reg Toxicol Pharmacol 38:345-367.

Lin, J.H. (1995) Species similarities and differences in pharmacokinetics. Drug Metabolism Disposition 23:1008-1021.

Mahmood, I. (1998) Interspecies scaling of renally secreted drugs. Life Sciences 63:2365-2371.

Mahmood, I. (2005) Interspecies Pharmacokinetic Scaling: Principles and Application of Allometric Scaling. Pine House Publishers, Rockville, MD.

Mahmood, I., Martinez, M, and Hunter, R.P. (2006) Interspecies allometric scaling. Part I: prediction of clearance in large animals. J vet Pharmacol Therap 29:415-423.

Malenke, M.R., Magnanou, E., Thomas, K., and Dearing, M.D. (2012) Cytochrome P450 2B diversity and dietary novelty in the herbivorous, desert woodrat (*Neotoma lepida*). PLoS ONE 7(8): e41510. https://doi.org/10.1371/journal.pone.0041510.

Manning, G.E., Mundy, L.J., Crump, D., Jones, S.P., Chiu, S., Klein, J., Konstantinov, A., Potter, D., and Kennedy, S.W. (2013) Cytochrome P4501A induction in avian hepatocyte cultures exposed to polychlorinated biphenyls: Comparisons with AHR1-mediated reporter gene activity and in ovo toxicity. Toxicol Applied Pharmacol 266(1):38-47.

Martinez, M., Mahmood, I., and Hunter, R.P. (2006) Interspecies allometric scaling: prediction of clearance in large animal species: Part II: mathematical considerations. J vet Pharmacol Therap 29:425-432.

Martinez, M., Mahmood, I., and Hunter, R.P. (2009) Allometric scaling of clearance in dogs. J vet Pharmacol Therap 32:411-416.

Maxwell, L.K., and Jacobson, E.R. (2004) Allometric scaling of kidney function in green iguanas. Comp Biochem Physiol Part A 138:383-390.

Mineau, P., Collins, B., and Baril, A. (1996) On the use of scaling factors to improve interspecies extrapolation of acute toxicity in birds. Reg Toxicol Pharmacol 24:24-29.

Mineau, P., Baril, A., Collins, B.T., Duffe, J., Joerman, G., and Luttik, R. (2001) Pesticide acute toxicity reference values for birds. Rev Environ Contam Toxicol 170:13-74.

O'Flaherty, E.J. (1989) Interspecies conversion of kinetically equivalent doses. Risk Anal 9:587-598.

Rhomberg, L.R, and Lewandowski, T.A. (2006) Methods for identifying a default cross-species scaling factor. Hum Ecol Risk Assess 12(6):1094-1127. https://doi.org/10.1080/10807030600977269.

Rhomberg, L.R., and Wolff, S.K. (1998) Empirical scaling of single oral lethal doses across mammalian species based on a large database. Risk Anal 18(6):741-753.

Riviere, J.E., Martin-Jimenez, T., Sundlof, S.F., and Craigmill, A.L. (1997) Interspecies allometric analysis of the comparative pharmacokinetics of 44 drugs across veterinary and laboratory animal species. J Vet Pharmacol Therap 20(6):453-463.

Sample, B., and Arenal, C. (1999) Allometric models for interspecies extrapolation of wildlife toxicity data. Bull Environ Contam Toxicol 62:653-663.

Sample, B.E., Fairbrother, A., Kaiser, A., Law, S., and Adams, B. (2014) Sensitivity of ecological soil-screening levels for metals to exposure model parameterization and toxicity reference values. Environ Toxicol Chem 33(10):2386-2398.

Savage, V.M., Deeds, E.J., and Fontana, W. (2008) Sizing up allometric scaling theory. PLoS Comput Biol 4(9): e1000171. https://doi.org/10.1371/journal.pcbi.1000171.

Schmidt-Nielsen, K. (1984) Scaling: Why is animal size so important? Cambridge University Press.

Sedgwick, C.J. (1993) Allometric Scaling and Emergency Care: The Importance of Body Size. In: Fowler, M.E. (ed.). Zoo and Wild Animal Medicine, 3rd ed. W. B. Saunders, Philadelphia, Pennsylvania. Pp. 34-37.

Sharma, V., and McNeill, J.H. (2009) To scale or not to scale: the principles of dose extrapolation. British J Pharmacol 157:907-921.

Shrestha, B., Reed, J.M., Starks, P.T., Kaufman, G.E., Goldstone, J.V., Roelke, M.E., O'Brien, S.J., Koepfli, K.P., Frank, L.G., and Court, M.H. (2011) Evolution of a major drug metabolizing enzyme defect in the domestic cat and other felidae: phylogenetic timing and the role of hypercarnivory. PLoS One 6(3):e18046. https://doi.org/10.1371/journal.pone.0018046.

Smaers, J.B., and Rohlf, F.J. (2016) Testing species' deviation from allometric predictions using the phylogenetic regression. Evolution 70(5):1145-1149.

Stevens, C.E., and Hume, I.D. (2004) Comparative Physiology of the Vertebrate Digestive System. Cambridge University Press.

Tang, H., and Mayersohn, M. (2005) A novel model for prediction of human drug clearance by allometric scaling. Drug Metab Dispos 33:1297-1303.

Tang, H., Hussain, A., Leal, M., Mayersohn, M., and Fluhler, E. (2007) Interspecies prediction of human drug clearance based on scaling data from one or two animal species. Drug Metab Dispos 5:1886-1893.

Toutain, P.L., and Bosquet-Mélout, A. (2004) Plasma clearance. J Vet Pharmacol Therap 27:415-425.

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

U.S. EPA. (1993) Wildlife Exposure Factors Handbook. U.S. Environmental Protection Agency, Washington, D.C., EPA/600/R-93/187.

U.S. EPA. (1997) Ecological Risk Assessment Guidance for Superfund: Process for Designing and Conducting Ecological Risk Assessments. Interim Final. U.S. Environmental Protection Agency, Office of Solid Waste and Emergency Response, EPA 540-R-97-006. https://semspub.epa.gov/work/HQ/157941.pdf.

U.S. EPA. (2005) Guidelines for Carcinogen Risk Assessment. U.S. Environmental Protection Agency, Washington, D.C., EPA/630/P-03/001F. http://www.epa.gov/sites/production/files/2013-09/documents/cancer_guidelines_final_3-25-05.pdf.

U.S. EPA. (2011) Recommended Use of Body Weight$^{3/4}$ as the Default Method in Derivation of the Oral Reference Dose. U.S. Environmental Protection Agency, Washington, D.C., EPA/100/R11/0001. https://www.epa.gov/risk/recommended-use-body-weight-34-default-method-derivation-oral-reference-dose.

U.S. EPA. (2014) Guidance for Applying Quantitative Data to Develop Data-Derived Extrapolation Factors for Interspecies and Intraspecies Extrapolation. U.S. Environmental Protection Agency, Washington, D.C., EPA/100/R-14/002F. https://www.epa.gov/sites/default/files/2015-01/documents/ddef-final.pdf.

U.S. EPA. (2015) (as accessed July 25, 2016) "Terrestrial Models" [For pesticide ecological risk assessments for U.S. Environmental Protection Agency, Office of Pesticide Programs/ Environmental Fate and Effects Division]. https://www.epa.gov/pesticide-science-and-assessing-pesticide-risks/t-rex-version-15-users-guide-calculating-pesticide.

Wajima, T., Fukumura, K., Yano, Y., and Oguma, T. (2002) Prediction of human clearance from animal data and molecular structural parameters using multivariate regression analysis. J Pharm Sci 91:2489-2499.

Walton K, Dorne, J.L., and Renwick, A.G. (2004) Species-specific uncertainty factors for compounds eliminated principally by renal excretion in humans. Food Chem Toxicol 42(2):261-274.

West, G.B. (2017) Scale: The Universal Laws of Growth, Innovation, Sustainability, and the Pace of Life in Organisms, Cities, Economies, and Companies. Penguin, New York.

West, G.B., Woodruff, W.H., and Brown, J.H. (2002) Allometric scaling of metabolic rates from molecules and mitochondria to cells and mammals. Proc Natl Acad Sci 99(Suppl 1):2473-2478.

West, L.J., Pierce, C.M., and Thomas, W.D. (1962) Lysergic acid diethylamide: Its effects on a male Asiatic elephant. Science 138(3545):1100-1103.

White, C.R., Blackburn, T.M., and Seymour, R.S. (2009) Phylogenetically Informed analysis of the allometry of mammalian basal metabolic rate supports neither geometric nor quarter-power scaling. Evolution 63:2658-2667.

Yokota, S.D., Benyajati, S., and Dantzler, W.H. (1985) Comparative aspects of glomerular filtration in vertebrates. Renal Physiol 8:193-221.

Zou, P., Yu, Y., Zheng, N., Yang, Y., Paholak, H.J., Yu., L.X., and Sun, D. (2012) Applications of human pharmacokinetic prediction in first-in-human dose estimation. The AAPS Journal 14:262-281.

Recommended References for Additional Information:

Boxenbaum, H. (1982) Interspecies scaling, allometry, physiological time, and the ground plan of pharmacokinetics. J Pharmacokin Biopharm 10:201-227.

Boxenbaum, H., and Ronfeld, R. (1983) Interspecies pharmacokinetic scaling and the Dedrick plots. Am J Physiol 245:R768-R775.

Chaui-Berlinck, J. G. (2006) A critical understanding of the fractal model of metabolic scaling. J Exp Biol 209:3045-3054.

Deck, A.T., and Johnson, M.S. (2015) Chapter 2 - Methods for Derivation of Wildlife Toxicity Values for Use in Ecological Risk Assessments. In: Williams, M.A., Reddy, G., Quinn, M.J., and Johnson, M.S. (eds.) Wildlife Toxicity Assessments for Chemicals of Military Concern. Elsevier. Pp. 15-22.

Dedrick, R. L. (1972) Animal Scale-Up. In: Teorell, T., Dedrick, R.L., and Condliffe, P.G. (eds.) Pharmacology and Pharmacokinetics. Springer, Boston, MA. Pp. 117-145.

Dedrick, R. L., and Bischoff, K.B. (1980) Species similarities in pharmacokinetics. Federation Proc 39:54-59.

Hayton, W. L. (1989) Pharmacokinetic parameters for interspecies scaling using allometric techniques. Health Physics 57:159-164.

Hunter, R.P., Mahmood, I., and Martinez, M.N. (2008) Prediction of xenobiotic clearance in avian species using mammalian or avian data: How accurate is the prediction? J vet Pharmacol Therap 31:281-284.

Kleiber, M. (1932) Body size and metabolism. Hilgardia 6:315–353.

Kleiber, M. (1961) The Fire of Life. An Introduction to Animal Energetics. John Wiley & Sons, Inc., New York.

Korsman, J.C., Schipper, A.M., and Hendriks, A.J. (2016) Dietary toxicity thresholds and ecological risks for birds and mammals based on species sensitivity distributions. Environ Sci Technol 50(19):10644-10652.

Luttik, R., and Aldenberg T. (1995) Extrapolation Factors to Be Used in Case of Small Samples of Toxicity Data (With a Special Focus on LD50 Values for Birds and Mammals). Report No. 679102029. National Institute of Public Health and Environmental Protection, Bilthoven, The Netherlands.

Luttik R., and Aldenberg T. (1997) Extrapolation factors for small samples of pesticide toxicity data: special focus on LD50 values for birds and mammals. Environ Toxicol Chem 16:1785-1788.

Luttik R., Mineau P., and Roelofs, W. (2005) A review of interspecies toxicity extrapolation in birds and mammals and a proposal for long-term toxicity data. Ecotoxicology 14:817-832.

Martinez, M. N. (2005) Interspecies Differences in Physiology and Pharmacology: Extrapolating Preclinical Data to Human Populations. In: Rogge, M., and Taft, D. (eds.) Preclinical Drug Development (Drugs and the Pharmaceutical Sciences, Vol. 152). Marcel Dekker, New York, New York. Pp. 11-66.

Mayfield, D.B, Johnson, M.S., Burris, J.A., and Fairbrother, A. (2014) Furthering the derivation of predictive wildlife toxicity reference values for use in soil cleanup decisions. Integr Environ Assess Manag 10(3):358-371.

Mealey, K. L., Bentjen, S.A., Gay, J.M., and Cantor, G.H. (2001) Ivermectin sensitivity in collies is associated with a deletion mutation of the mdr1 gene. Pharmacogenetics 11:727-733.

Raimondo, S., Mineau, P., and Barron, M.G. (2007) Estimation of chemical toxicity to wildlife species using interspecies correlation models. Environ Sci Technol 41(16):5888-5894.

Riviere, J. E. 1999. Interspecies Extrapolations. In: Comparative Pharmacokinetics: Principles, Techniques, and Applications. Iowa State University Press, Ames, Iowa. Pp. 296-307.

Sample, B.E., Opresko, D.M., and Suter, G.W.II. (1996) Toxicological Benchmarks for Wildlife: 1996 Revision. Oak Ridge National Laboratory, USDOE. ES/ER/TM-86/R3.

APPENDIX A. BODY SIZE AND TISSUE STEADY STATE CONCENTRATION: A NUMERICAL ILLUSTRATION

Tissue concentration of a toxicant is determined by relative rates of uptake and elimination. Both types of processes are subject to allometric trends across species (Schmidt-Nielsen, 1984; U.S. EPA, 2011). As a rule, larger species will consume less on a weight specific basis, and also eliminate less rapidly (on a weight specific basis). A bias may result if the risk calculations address allometry for one type of process - uptake or elimination - but not the other. While direct measurements of some rates of interest may be available for a species of interest (e.g., U.S. EPA, 1993), this section considers the effect of assuming that rates conform to general expectations based on allometry.

In theory AUC or steady-state concentration should not depend on body weight other things being equal (O'Flaherty, 1989). The table below illustrates the theoretical lack of BW dependency, based on comparing tissue concentration trajectories for a 35 g "mouse-sized" species and a 300 g "rat-sized" species. It is assumed that the environmental concentration is the same for each species and that differences result from allometry in rates of uptake and clearance. The rate for each of the two species are assumed to conform to the same proportionality to $BW^{3/4}$ across species.[13]

It may be helpful to assume that the elimination process is glomerular filtration, simply because for that process the pharmacokinetic definition of "clearance," as volume of tissue cleared per unit time, seems relatively obvious. A general definition of clearance (Toutain and Bosquet-Mélout, 2004) applies to elimination by various mechanisms.

We assume initially (Column 3) that each species ingests a toxicant (in arbitrary mass units) proportional to whole-organism metabolism. If the toxicant is mixed in tissues before any elimination, the result (Column 4) is a lower tissue concentration for the larger animal,

[13] More technically, if the two species belong to a group with the same allometric coefficient and exponent then the coefficient can be ignored for purposes of the illustration. Also, the conclusions do not depend on the value of the allometric exponent *b* (3/4 or otherwise) so long as the same value is assumed for both uptake and clearance.

reflecting the lower weight-specific food consumption (lower mass food intake per unit body mass). Next (Column 5), we assume that a volume of body fluid is removed, again typical for the body weight, and replaced with uncontaminated fluid. We assume that the fraction of ingested toxicant removed is proportional to the ratio of fluid removed to body weight. (Technically, for the argument presented we are using body weight as a surrogate for volume of distribution, which would be used if directly available.) After this elimination event, the tissue concentrations are more nearly equal in the two species (Column 8). Columns 9 and 10 show the results of repeating these computations for two additional cycles of ingestion and elimination, which suffice for the illustration. While the simulation approach is crude, it suggests tissue concentrations converging to the same value in each species, as expected based on theory.

Example of uptake and elimination related to body size. Table columns have been numbered to facilitate cross-referencing with explanations (in text).

1	2	3	4	5	6	7	8	9	10
Species	BW (g)	Toxicant Mass Uptake $BW^{3/4}$	Tissue Conc. $BW^{-1/4}$	Glomerular Filtration (mass in 1 time unit)	% Toxicant Eliminated	% Toxicant Remaining	Tissue Conc.	(Add 1 Step)	(Add 1 Step)
Mouse-sized	35	14.39	0.41	14.39	41%	59%	0.24	0.38	0.47
rat-sized	300	72.08	0.24	72.08	24%	76%	0.18	0.32	0.43

APPENDIX B. ADDITIONAL DETAIL ON SELECTED SOURCES

Allard, P., Fairbrother, A., Hope, B., Hull, R., Johnson, M., Kapustka, L., Mann, G., McDonald, B., and Sample, B. 2009. Recommendations for the Development and Application of Wildlife Toxicity Reference Values. Integr Env Assess and Man 6:28–37.

Type of Source	Assessment Context / Taxonomic Restriction	Substances of Special Interest	Chronicity	Objectives / Approach
Journal article	Ecological/Birds and mammals	General	General	General policy document addressing multiple practices in toxicity reference value derivation

Procedure Described or Recommended: "Don't use allometric dose-scaling with body mass when assessing chronic/subchronic toxicity between species. ... Allometric scaling ...has been used for wildlife risk evaluations despite its multiple limitations. ... is no longer recommended for use in wildlife risk assessment (U.S. EPA, 2005). First, supporting data are limited. Much of the mammalian data are based on anticancer drugs evaluated in Freireich et al. (1966) rather than contaminants typically evaluated in wildlife risk assessments. Second, the allometric scaling models developed for both human and wildlife risk assessment are all based on acute toxicity data. Their applicability to chronic toxicity data is unknown. ... Because modes of action can vary dramatically for the same chemical over acute and chronic exposures (discussed in more detail below), it is likely that interspecific scaling factors based on chronic toxicity data also will differ from those based on acute toxicity data. Additionally, given the variation in cross-species physiological responses in different organ systems, it is reasonable to expect multiple chronic scaling factors for a given chemical, depending on the mode of action considered. In their current forms, neither allometric scaling nor ICE [interspecies correlation estimation] models represent chronic toxicity, and, therefore, their application to chronic data is not recommended. In the absence of suitable models, we favor the use of toxicity information as reported, because it is often unknown whether target species would be more resistant or more sensitive."

Mineau, P., Collins, B., and Baril, A. 1996. On the use of scaling factors to improve interspecies extrapolation of acute toxicity in birds. Reg Toxicol Pharmacol 24:24-29.

Type of Source	Assessment Context / Taxonomic Restriction	Substances of Special Interest	Chronicity	Objectives / Approach
Journal article presents analysis of pesticide lethality to birds	Ecological/Avian	37 pesticides, "heavily weighted" towards cholinesterase inhibitors	LD50	Estimate allometric coefficient using regression

Procedure Described or Recommended: "We used an avian LD50 database to derive empirically the appropriate scaling factor for birds. With a subset of 37 pesticides of varying structures but heavily weighted to cholinesterase inhibitors, we found that the appropriate scaling factor in birds is usually higher than 1 and can be as high as 1.55. Extrapolations on the basis of weight alone or, worse, the use of inappropriate mammalian scaling factors could lead to serious underprotection of small-bodied bird species modeled in the course of risk assessment procedures."

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

Type of Source	Assessment Context / Taxonomic Restriction	Substances of Special Interest	Chronicity	Objectives / Approach
Journal article presents meta-analysis of lethal effects and discusses policy implications	Human health/Data for mammals in multiple orders	Data analysis based on 135,000 substances, largely of occupational human health concern.	Single-dose lethal (LD50)	Compare alternative body weight adjustments based on pairwise species comparisons

Procedure Described or Recommended: "We find a good correspondence of LD50 … across species when the dose levels are expressed in terms of mg … per kg of body mass. … contrast with earlier analyses that support scaling doses by the 3/4-power of body mass to achieve equal subacute toxicity of antineoplastic agents. We suggest that, especially for severe toxicity, single- and repeated-dosing regimes may have different cross-species scaling properties, as they may depend on standing levels of defenses and rate of regeneration of defenses, respectively."

Sample, B., and Arenal, C. 1999. Allometric models for interspecies extrapolation of wildlife toxicity data. Bull Environ Contam Toxicol 62:653-663.

Type of Source	Assessment Context/ Taxonomic Restriction	Substances of Special Interest	Chronicity	Objectives / Approach
Journal article presents meta-analysis of lethal effects and discusses policy implications	Ecological/Birds and mammals	Multiple classes of organic and inorganic compounds	Single-dose lethal (LD50)	Regression analysis of 2,853 lethal oral dose measurements. Objectives were characterization of relationship to body weight and comparison of birds to mammals

Procedure Described or Recommended: "Do not extrapolate from birds to mammals or vice versa. Use a chemical-specific scaling factor or possibly a factor for a chemical group, e.g., chlorinated organics … . Use $BWA^{1.2}$ (birds) or $BWA^{0.94}$ (mammals)." [Note that the avian value suggested by these authors derives from Mineau et al., 1996.]

APPENDIX C. BODY WEIGHT SCALING AS AN APPROXIMATION OF PHARMACOKINETIC MODELING – ADDITIONAL ANALYSIS OF PUBLISHED RESULTS

For the human health assessment context, U.S. EPA (2011) indicates PBPK modeling as the preferred approach for addressing PK considerations. The model-based analysis of Kirman et al. (2003) is cited by U.S. EPA (ibid.) as support for $BW^{3/4}$ scaling. The analysis draws on previous modeling efforts for each of 13 substances. The table below reproduces selected results from that analysis, along with a re-expression in terms of the best allometric exponent b for approximation of PBPK modeling, assuming blood AUC to be the appropriate summary of tissue-level exposure over time.

The authors evaluated $BW^{3/4}$ scaling by comparison to model-based predictions of blood AUC for mouse, rat and human for 12 lipophilic, predominantly volatile chemicals. The results were expressed as equivalent internal doses (EID) comparing two species, which are ratios of AUC (AUC_a / AUC_h) at the same external dose (mg/kg-d). (For various quantities, subscripts h and a will indicate respectively human and non-human animal.) The table below gives the geometric mean EID values for the 9 chemicals as reported in their Table 4, based on model blood concentration of parent.

The approach taken by the authors was to evaluate whether their PBPK-based ratios were reasonably close to $BW^{3/4}$ expectations. Here, instead of taking 3/4 as the point of reference, an effective allometric exponent is computed using the formula

$$b = 1 - \frac{\log \mathrm{EID}}{\log\left({}^{\mathrm{BW}_a}/_{\mathrm{BW}_h}\right)}.$$

Here BW_a and BW_h are species body weights. This approach identifies a value of b that is in a sense best for purposes of approximating the effect of using PBPK modeling (assumed to be the preferred approach).

Overall, these results seem to provide reasonable support for $BW^{3/4}$ scaling as a PK-based adjustment. Extrapolation on a simple body weight basis appears less supported for parent compound. It should be noted that some geometric mean EIDs (bold) are reported by the authors to only one digit of precision.

It might happen that parametrization of a PBPK model makes use of $BW^{3/4}$ scaling in estimation of some parameters. Then, there would be a degree of circularity in using the results to claim that allometric scaling of toxic doses approximates PBPK-based results. Nevertheless, it seems to be still of interest how well $BW^{3/4}$ scaling approximates current PBPK-based estimates.

Allometric exponents for approximation of PBPK model-based ratios of species AUCs for Continuous Oral Exposures.

Dose (mg/kg-d)	mouse – human		rat – human	
	EID	effective *b*	EID	effective *b*
0.0001	0.13	0.74	0.28	0.77
0.001	0.11	0.72	0.25	0.75
0.01	0.10	0.70	0.23	0.73
0.1	0.10	0.70	0.22	0.72
1	**0.09**	**0.69**	0.22	0.72
10	**0.08**	**0.67**	0.19	0.70
100	**0.04**	**0.58**	0.14	0.64
1000	**0.05**	**0.61**	0.29	0.77
10000	0.11	0.72	0.41	0.84

APPENDIX D. PHARMACOKINETIC DERIVATIONS OF THE ALLOMETRIC SCALING PROCEDURES

U.S. EPA (2011) advises in the context of human health assessments that allometric scaling serves primarily to address pharmacokinetic (PK) considerations and to some degree pharmacodynamic (PD) considerations. Two PK derivations are given for $BW^{3/4}$ scaling of doses. Each assumes some form of averaging or cumulation of internal toxicant concentrations, as the toxicologically most appropriate summary of a tissue-time curve. In addition, we show the theoretical independence of equilibrium tissue concentrations from body size. Results in this section have been compiled for convenience, with no claims of originality.

Models here are simplistic, first-order and single-compartment. Lin (1995) can be recommended as an introduction to species PK differences based on simple models (allowing that the source could have a particular focus on extrapolation among mammals). Such models may be taken as a reasonable basis for low-tier assessment methodologies, at least.

Two species are indexed 1 (test species with toxicity measured) and 2 (assessment species with toxicity measurement unavailable). It may be helpful to think of extrapolating from Species 1 to Species 2. Various quantities are subscripted 1 or 2 particularly W_1 and W_2, the body weights of the two species.[14]

The allometric exponent will be assumed to be 3/4 (however, results are easily generalized).

Scaling based on biological equivalence of AUC from single doses. Here we assume a single exposure event recorded as mg toxicant per kg body weight. Tissue concentration eventually diminishes with time as the toxicant is eliminated by some mechanism, as described by a concentration-time curve. The area under a concentration-time curve (AUC) may be considered as a basis for dose equivalence results (e.g., U.S. EPA, 2011). AUC is a measure of cumulative internal exposure, with units concentration*time (say, min*mg/L).

[14] This section uses color– a dark green for quantities associated with the test species (Species 1) and red for the assessment species (Species 2).

The AUC-based extrapolation of toxic dose (mg/kg) from Species 1 to Species 2,

$$\text{toxicity}_2 = \text{toxicity}_1 \times \left(\frac{W_1}{W_2}\right)^{0.25},$$

can be derived from an expression for AUC (Lin, ibid., Expression 4),

$$\text{AUC} = \frac{F \times \text{dose}}{\text{clearance}}.$$

Here dose is mass (mg) of chemical ingested, F (between 0 and 1) is bioavailable fraction, and "clearance" (units volume/time) denotes the volume of tissue cleared of toxicant per unit time (units volume/time). Lin (ibid.) expresses F as a product of factors representing fractions absorbed and surviving breakdown in the liver and gut wall. In any case, such factors are here assumed similar between Species 1 and 2, so that they cancel approximately in the derivation. (However, refinements of extrapolation methodology may be based on known species differences in such factors.)

Suppose that the biological response will depend on AUC in the same way, for Species 1 and 2, that is, a given AUC will or will not produce an effect in Species 1 according as it does or does not produce the effect in Species 2. Differences in AUC for a given dose are assumed to be based on differences in clearance. A first-order clearance process is assumed (i.e., one that is independent of dose), with a rate that scales across species with an exponent 3/4, as with glomerular filtration rate (GFR) or other processes that scale with body weight in the same way as blood flow (Lin, ibid.; U.S. EPA, 2011 particularly Table 4.1). Then from the AUC expression we have

$$\frac{\text{dose}_2^*}{W_2^{3/4}} = \frac{\text{dose}_1^*}{W_1^{3/4}}$$

where dose_i^* is the dose just sufficient for the toxicological effect in Species i (i = 1 or 2).

Rearranging this expression leads to the factor for extrapolating a dose in mass units (not BW-normalized):

$$\text{dose}_2^* = \text{dose}_1^* \times \left(\frac{W_2}{W_1}\right)^{3/4}$$

To derive the factor appropriate for application to a BW-normalized toxicity measurement, first multiply both sides of the previous equation by 1:

$$W_2 \frac{\text{dose}_2^*}{W_2} = W_1 \frac{\text{dose}_1^*}{W_1} \left(\frac{W_2}{W_1}\right)^{3/4}.$$

Then rearrange to yield the scaling expression:

$$\frac{\text{dose}_2^*}{W_2} = \frac{\text{dose}_1^*}{W_1} \frac{W_1}{W_2} \left(\frac{W_2}{W_1}\right)^{3/4}$$

$$= \frac{\text{dose}_1^*}{W_1} \left(\frac{W_1}{W_2}\right)^{1/4}.$$

In summary, while $BW^{3/4}$ scaling of a toxic dose that is *not* BW-normalized (in mass units) involves a factor $(W_2/W_1)^{3/4}$, the corresponding factor for a BW-normalized toxic dose is $(W_1/W_2)^{1/4}$ (also see U.S. EPA, 2011, ix).

Scaling for repeated exposure based on biological equivalence of equilibrium body burden. Now assume that dosing is repeated, and the results are reported (for the sake of concreteness) in units mg toxicant per kg body weight, per day (mg/kg-d). An extrapolation expression can be derived by viewing an average dose as based on a sum of AUCs corresponding to individual meals. A stochastic version of this idea may be useful for probabilistic risk assessment based on exposures that vary in space and time, and feeding behavior more or less unpredictable. For a deterministic approximation of average body burden we may view uptake and elimination as continuous and use steady-state value, found by equating uptake and elimination rates, as an approximation of average body burden. For a single species, consider the amount of toxicant in

any 1-kg volume of tissue and write the steady-state equation for that volume, for a single time unit, as

$$\text{mg toxicant taken in} = \text{mg toxicant eliminated}$$

$$= (\text{mg toxicant in tissue}) \times (\text{fraction eliminated}).$$

Regarding the right-hand side note that the *total volume* of tissue cleared of toxicant in a unit time for the whole organism is assumed proportional to $W^{3/4}$. Assuming the same fraction eliminated per unit time in each 1 kg portion of tissue, that fraction would be $W^{3/4}/W = 1/W^{1/4}$ so,

$$\text{toxicant taken in} \propto \frac{[\text{toxicant in tissue}]}{W^{1/4}}$$

or

$$\text{toxicant in tissue} \propto [\text{toxicant taken in}] \times W^{1/4}.$$

(The proportionality signs indicate neglect of constants that are assumed to be similar across species and cancel in the development of the extrapolation factor.) Finally, to develop the factor for extrapolation from Species 1 to Species 2 we want to know the Species 2 external exposure (mg/kg-d) that will yield the steady-state body burden (mg/kg) sufficient for a toxic effect in Species 1 (in each 1 kg volume of a given species.) Equating body burdens we write:

$$[\text{Species 2 mg/kg-d}] \times W_2^{1/4} = [\text{Species 1 mg/kg-d}] \times W_1^{1/4}$$

which can be solved to yield the usual extrapolation factor.

Steady-state body burden does not depend on body weight when uptake and elimination scale to the same power of body weight (also see O'Flaherty, 1989). The following relates to species assumed to belong to a group which has the same allometric expressions for toxicant uptake and clearance rates. Equilibrium tissue concentration can be identified by setting input rate equal to elimination rate:

toxicant input (mg) = toxicant elimination (mg)

Suppose that metabolic rate and clearance scale to the same power b of body weight (e.g., both are $BW^{3/4}$). Then

$$C_{\text{food}} W^b \propto C_{\text{tissue}}\, W^b$$

where C_{food}, C_{tissue} are respectively concentrations of toxicant in food and (at steady state) tissue. (On the left-hand side, the exponent is based on food ingestion being proportional to metabolic rate.) Therefore, tissue concentration is simply proportional to feed concentration given these relationships. If toxicity is measured as a feed concentration associated with a biological effect, then we know of no PK argument for body weight scaling of food concentrations. With different allometric exponents for uptake and elimination there may be some body weight dependence of critical feed concentration but the effect may be small if the exponents are not too different.