EMPIRICAL TOXICOKINETIC AND DYNAMIC SCALING OR "SAFETY" FACTOR ESTIMATES FROM INTEGRATED TOXICOLOGY AND EPIDEMIOLOGY STUDIES OF PERSISTENT ORGANIC POLLUTANTS

Tom Muir[1*] and Joel E Michalek[2]

[1]Environment Canada – Retired. 70 Townsend Ave, Burlington, Ontario, Canada; [2]Department of Epidemiology and Biostatistics, University of Texas Health Science Center at San Antonio, 314.9 Admin Building, 7703 Floyd Curl Drive, San Antonio TX.

Introduction

Risk assessment (RA) and management are tools of policy and regulation. The present RA policy is to assess individual chemicals in isolation. Use is deemed ”safe”, if exposures (daily intakes or body burdens) are below some estimated "no effect" or “threshold" concentration, or “margin of exposure” (MOE), compared, typically, to an experimental model critical effect concentration without any relation to actual human health effects. This ratio of toxicology over human concentrations (or intakes) is the raw MOE, without any adjustments or modifications. The MOE itself does not usually account for variation or uncertainty of dose or response. “Default uncertainty factors” exist primarily for interspecies differences and intraspecies variability. These defaults are being challenged as perhaps overstating the uncertainty[1,2], however, refinements have been suggested with factors higher and lower[3,4,5]. The suggestion made, is that toxicokinetic and dynamic factors, for example, "biomonitoring equivalents"[1], should be estimated and used to scale the animal to human extrapolation of dose-response used in RA. Our integrative work in this paper, yields empirical estimates of toxicokinetic and dynamic "scaling" or "safety factors" (CSF) derived from the multiple species, endpoints, and chemicals data assembled. Using a weight of evidence approach, we integrate a data-base of comparable internal dose and response effect concentration data, from a number of toxicological (*in vitro* and *in vivo*) and epidemiological studies reporting on a range of POPs chemicals, multiple species, and for multiple toxicological responses or endpoints[2]. We empirically estimate the relative concentration potencies in the three study types, and provide data-derived scaling or uncertainty or safety factors that account, where relevant, for toxicokinetic and dynamic relationships and variation as integrated and expressed in the resulting internal doses and responses of the selected studies. In the *in vitro* assay system, metabolism and other relevant kinetic components are not occurring as they are *in vivo*, but no modifications for this were found in literature. Nonetheless, given the prevalence of *in vitro* data, their comparability to *in vivo* internal dose metrics, and reported future plans, such as USEPA Tox21, the respective doses are relevant. Using the dose distributions of each study type, we provide 95% confidence intervals for the Calculated Safety Factors (CSF), defined as the toxicology dose divided by the epidemiology body burden, by category of effect and basis of measurement (lipid weight, wet weight). The confidence interval expresses the uncertainty or scaling or "safety" factors in terms that can apply directly to the real world of RA. We name these CSF because the human doses compared with the experimental doses are based on epidemiological reports of significant effects, and not just exposure or biomonitoring assessment with no regard to effects.

Methods and Materials

We selected 68 relevant POPs *in vitro* (n= 40) and *in vivo* (n= 28) studies, and 53 epidemiological studies. We made the selection to include studies of BFRs, FRs and POPs with published internal dose potencies and specification of the effect. Animal species included mouse, rat, monkey, kestrel, rainbow trout, flounder, and fathead minnow. We stratified by basis (lipid weight, wet weight), study (*in vivo* toxicology, *in vitro* toxicology, epidemiology), chemical (in 22 categories), and effect (in multiple categories or markers (n=102), aggregated to DNT (n=22), thyroid (n=35), and NTE (n=45) due to sample size constraints). The chemical category number (22) was not selected by us, but emerged from the studies used. We expressed the internal dose in a common Molar metric expressed in log base 10. We assessed the statistical significance of variation in reported or minimum internal dose observed to be associated with an effect with study type (in vitro (toxicology), in vivo (toxicology), epidemiology), basis (wet, lipid), and effect category (non-thyroid endocrine (NTE), developmental neurotoxicity (DNT), thyroid). We contrasted toxicology with epidemiology and in vivo toxicology with in vitro toxicology with regard to the

mean $\log_{10}$ (Molar) using analyses of variance and, for each contrast, a 95% confidence interval for the mean difference (toxicology mean dose minus the epidemiology mean body burden). We provide 95% confidence intervals for the Calculated Safety Factors (CSF), defined as the toxicology effect dose distribution over the epidemiology effect dose distribution, by category of effect and basis of measurement (lipid weight, wet weight). We applied the Tukey method to correct multiple pairwise comparisons. All statistical testing was two-sided with a nominal experimentwise significance level of 5%. We used SAS Version 9.2 for Windows (SAS Institute, Cary, NC) throughout[6].

Results and Discussion

Table 1 shows the sample sizes by basis, study design, and effect category for the all toxin chemical category. We summarized 652 dose measurements in all studies (Lipid weight: Epidemiology 136, in vivo toxicology 29, in vitro toxicology 0, Wet weight: Epidemiology 141, in vivo toxicology 64, in vitro toxicology 282).

Table 1. Sample sizes by basis, effect, and study

	Lipid Weight			Wet Weight		
		Toxicology			Toxicology	
Effect Category	Epidemiological	in vivo	in vitro	Epidemiological	in vivo	in vitro
DNT	21	11	0	24	35	66
NTE	42	8	0	32	17	133
Thyroid	73	10	0	85	12	83
Total	136	29	0	141	64	282

Reflecting the variance in the mean dose by basis (lipid weight, wet weight), effect category (DNT, NTE, Thyroid), and study type (*in vivo*, *in vitro*, epidemiology), the Calculated Safety Factor (CSF) varied considerably (Table 2). The CSF (toxicology effect dose distribution over the epidemiology effect dose distribution) for the in vivo and epidemiological results are the empirical estimates of the cross-species pharmacokinetic and dynamic uncertainty or safety factors embodied in the database of results integrated here. The in vitro results do not contain metabolic or relevant kinetic components, and this must be kept in mind in comparisons.

Table 2 Calculated Safety Factor (CSF) and 95% Confidence Interval by Basis and Effect

		CSF (95% Confidence Interval): Epidemiology Relative to		
Basis	Effect	In vivo	In vitro	All Toxicology
Lipid	DNT	3.8 (0.5, 27)		3.8 (0.5, 27)
	NTE	16.8 (2.6, 110.5)		16.8 (2.6, 110.5)
	Thyroid	3.7 (1.1, 12.4)		3.7 (1.1, 12.4)
Wet	DNT	182 (59.6, 555.8	4053.1 (1518.6, 10817.8)	1382.8(477.6, 4003.7)
	NTE	23.1 (6.0, 89.5)	225.4 (87.3, 581.9)	174.1 (67.5, 449)
	Thyroid	6357 (1229.9, 32856.5)	185.5 (76.8, 448.4)	289.9 (120.3, 698.9)
All	All	39.6 (19.1, 82.0)	669.7 (380.5,1178.9)	67.3 (42.3, 106.9)

Relative to all toxicology studies across both bases and all three effect categories, the overall CSF was 67.3 with CI 42.3 to 106.9, indicating a 67-times increase in potency in humans relative to animals and assays combined. The CSF varied from 3.8 CI (0.5, 27) for DNT effects in epidemiology relative to *in vivo* in studies reporting lipid weight results, indicating a 3.8 times increase in potency in humans relative to animals, to 6357 CI (1229.9, 32856.5) for Thyroid effects in epidemiology relative to *in vivo* studies reporting wet weight results, indicating a 6357-times increase in potency in humans relative to animals, on

the average. Where the data were most abundant, in DNT, NTE, and Thyroid effects in studies reporting wet weights results, the CSF was at least 23.1 CI (6.0, 89.5). Safety Factors calculated for the upper bound of 95% confidence range from 1 to 5 orders of magnitude for *in vivo* versus epidemiology contrasts, and 2 to 5 orders of magnitude for *in vitro* versus epidemiology. Thyroid results in particular stand out for apparent higher sensitivity of response in humans.

It is noteworthy, that in comparing human and rat dose-response models for individual PCB congeners and thyroid hormones, from several integrated studies, that Parham et al 2012[7] found predicted rat dose-responses generally orders of magnitude lower than those for humans. Overall, the human dose-response ranged from about 1 order of magnitude to between 5 and 6 orders higher than the rat. For different metrics of the sum of PCBs comparisons were much closer to the lower end of this range, to one instance of almost equality out of 79 comparisons. Explanations for greater human sensitivity, both in the present paper, and in Parham et al, are several, including model types and designs, context and other factors. These and other possible physiological reasons are beyond the present scope. The results also support the importance of considering, together, mixtures of chemicals that can affect the same common adverse outcome or pathway.

Please note that we suspect possible measurement error in lipids analysis of relatively leaner epidemiology matrices compared to in vivo. In the overall results, there is a gain of 1.5 orders by the lipid weight contrasts compared to wet weight. We recommend the wet weight factors rather than lipid based.

The summary statistics of mean difference and 95% confidence interval of the contrasts between study types and effect categories provides further background to the CSF presented above. These statistical results, in Tables 3a and 3b, generally parallel the variation shown in the CSF.

Table 3 Contrasts between Toxicological and Epidemiological Studies on mean Log_{10}(Dose or Body Burden in Molar units)

a) Wet weight [N, mean±SD)]

Effect	Toxicology in vivo	in vitro	All	Epidemiology	p-value	95% CI
DNT	35	66	101	24		
	-6.61±0.71	-5.26±0.79	-5.73±1	-8.87±1.15)		
	•			•	<0.001	(1.78, 2.74)
		•		•	<0.001	(3.18, 4.03)
			•	•	<0.001	(2.68, 3.6)
NTE	17	133	150	32		
	-6.88±0.74	-5.89±1.06	-6±1.07	-8.24±1.08)		
	•			•	<0.001	(0.78, 1.95)
		•		•	<0.001	(1.94, 2.76)
			•	•	<0.001	(1.83, 2.65)
Thyroid	12	83	95	85		
	-5.23±1.05	-6.76±1.33	-6.57±1.39	-9.03±1.18)		
	•			•	<0.001	(3.09, 4.52)
		•		•	<0.001	(1.89, 2.65)
			•	•	<0.001	(2.08, 2.84)

Among studies reporting DNT, NTE or Thyroid effects in wet weight (Table 3a), the mean dose was significantly decreased ($p<0.001$ for all contrasts) in epidemiology relative to wet weight *in vivo* and wet weight *in vitro* toxicology [DNT: in vivo toxicology -6.61±0.71, in vitro -5.26±0.79, epidemiology -8.87±1.15, CI (1.78, 2.74; 3.18, 4.03, respectively), NTE: in vivo toxicology -6.88±0.74, in vitro -5.89±1.06, epidemiology -8.24±1.08, CI (0.78, 1.95; 1.94, 2.76, respectively), Thyroid: in vivo toxicology -5.23±1.05, in vitro -6.76±1.33, epidemiology -9.03±1.18, CI (3.09, 4.52; 1.89, 2.65, respectively)].

Table 3 Contrasts between Toxicological and Epidemiological Studies on mean Log_{10}(Dose or Body Burden in Molar units)

b) Lipid weight [N, mean±SD)]

Effect	in vivo Toxicology	Epidemiology	p-value	95% CI
DNT	11	21		
	-5.89±1.12	-6.46±1.13	0.18	(-0.28, 1.43)
NTE	8	42		
	-5.56±0.45	-6.79±1.13	0.004	(0.41, 2.04)
Thyroid	10	73		
	-6.61±1.02	-7.18±0.75	0.04	(0.04, 1.09)

Corresponding contrasts in lipid weight (Table 3b) were in the same direction, but were generally smaller, and did not reach significance for studies expressing DNT effects. The lipid weight contrast gains compared to wet weight appear again in the individual effect strata, with thyroid > DNT> NTE.

The possible variation seen across species and basis become more evident when the three effect categories are stratified and appear separately for both toxicology model types, and in specific contrast with epidemiology. The apparent much higher sensitivity of response in humans versus animals for the thyroid outcome stands out, although higher human sensitivity appears general in the results. In the thyroid, however, the specific animal species and/or strains used may be partial explanations. For example, for in vivo thyroid, the animals used are rat, fathead minnow, rainbow trout, and flounder. Chemicals tested are TBBPA, PBDE-47, and HBCD. For the human epidemiology, there are many more chemicals with significant effects reported, and indeed there are a whole host of factors not present in the animal models.

To our knowledge, this is the first combined quantitative review of toxicological and epidemiological data organized in this way, in the same Molar metric and as internal dose, and the first empirical display of variability in relative potency. Such data, as the CSF results presented here, can refine or supplant the current "default uncertainty factors", as noted above.

References

1. Hayes S, Becker R, Leung A, Aylward L, Pyatt D. 2007. *Regulatory Toxicology and Pharmacology* **47** (2007) 96–109
2. Rhomberg L, and Lewandowski T. 2004. Gradient Corporation, Prepared for Risk Assessment Forum, U.S. Environmental Protection Agency, Washington, DC. Contract 3W-0477-NASX
3. Calabrese E, and Baldwin L. 1994. *Environ Health Perspect* 102:14-17.
4. Dorne J, and Renwick A. 2005. *Toxicological Sciences* 86(1), 20–26.
5. Swedish National Chemicals Inspectorate. 2003. Human Health Risk Assessment: Proposals for the use of assessment (uncertainty) factors. Stockholm, Sweden.
6. Muir T, and Michalek J. 2012. Manuscript in preparation
7.Parham F, Wise A, Axelrad, D, Guyton K, Portier C, Zeise L, Zoeller T, Woodruff T. 2012. Environmental Research. on-line edition, April 4, 2012.