A FRAMEWORK FOR EVALUATING RELATIVE POTENCY DATA IN THE DEVELOPMENT OF TOXICITY EQUIVALENCY FACTORS (TEFS)

Kevin Connor[1] and Brent Finley[2]

[1]Exponent, 21 Strathmore Road, Natick, MA 01760
[2]Exponent, 631 1st Street, Santa Rosa, CA 95404

Introduction

The toxic equivalency (TEQ) scheme is an approach to the human health risk assessment of polychlorinated dibenzo-*p*-dioxins (PCDDs), dibenzofurans (PCDFs), and biphenyls (PCBs), and is based on structure-activity information collected over the past 25 years. Recent attention has been focused on the toxic equivalency factors (TEFs), which represent a fractional potency of each congener relative to 2,3,7,8-tetrachlorodibenzo-*p*-dioxin (TCDD). The importance of the TEFs is underscored by the fact that >85% of the background TEQ body burden in humans is thought to be comprised of congeners other than TCDD[1].

The current TEFs established by the World Health Organization (W.H.O.) were derived by an *ad hoc* work group[2]. A database of 936 relative potency (REP) values was the basis for the workgroup's evaluation, which contained data from published and unpublished studies[3]. Not surprisingly, the REP data represent a wide variety of animal or cell culture models, dosing regimens, time courses, and measurement endpoints. As a result, there is a high degree of variability in the REP values, and in many cases the REP values for a given congener span several orders of magnitude. Further, while the W.H.O. workgroup placed greater emphasis on REP data from chronic and *in vivo* studies (versus acute or *in vitro* studies), there was no systematic or quantitative weighting scheme employed during the TEF derivation process. As a result, it is not possible to reproduce the current TEFs from the underlying REP data. Because of these and other factors, the TEFs often introduce a very high degree of uncertainty into the PCB and PCDD/F risk assessment process.

The objective of this work is to develop a quantitative weighting scheme for evaluating individual REP values, so that TEFs can be established in a more consistent, reproducible and transparent manner. We believe that a more systematic approach to REP evaluation will permit a more informed discussion of the uncertainties present in the TEFs (and the health risk estimates derived using them). In this analysis, we propose a detailed framework for evaluating seventeen different metrics of quality and relevance of REP study data. PCB 126 and 1,2,3,7,8-PeCDF are evaluated as a case study.

Methods

This analysis utilized the data assembled in the W.H.O. 1997 REP database. We used the data at face value; however, we first conducted an "audit" of the REP database and found it necessary to eliminate numerous erroneous or repetitive REP values from consideration. This resulted in the removal of 7 REPs for 1,2,3,7,8-PeCDF and 49 REPs for PCB 126. These values were removed for one or more of the following reasons: 1) identical data were published more than once, 2) identical data were represented with different units, 3) REPs were provided by study authors and also calculated by W.H.O., 4) the REPs did not meet the WHO inclusion criteria, or 5) REPs could not be linked to unique, qualified data in the cited reference (e.g., data-entry errors).

Seventeen different study elements were identified as critical measures of REP quality and relevance. A grading scheme for each study element is proposed (Tables 1 and 2). Relevance refers to the congruence of the study element to human exposure/risk assessment with cancer as an endpoint, and this was the most important criterion for grading study elements such as species/strain, route of exposure, measurement endpoint, and exposure duration. Quality refers to an objective measure of study reliability or validity and was the most important criterion for grading study elements such as the number of dose levels tested, chemical purity, and appropriate controls.

Table 1. Quantitative Rating Scheme for Tier 1 Study Elements

10	7	4	1
Cell Culture			
Primary cultures derived from relevant human tissues; well-characterized/ relevant/reliable models (e.g., human hepatocytes or splenocytes)	Cells/tissue slices from common lab strains (hepatocytes, lymphocytes, thymii from monkey, pig, rat); transformed cell lines derived from human tissues (HepG2; MCF-7)	Transformed cell lines derived from non-human mammals and relevant tissue type/ conventional/less relevant/reliable models (Hepa 1c1c7; H4IIE)	Less conventional choice/less reliable/relevant models; not chosen as Ah-responsive model (JEG 3; SVK14; cytosol)
Route of administration			
Oral (diet or gavage)	*ip* injection	*iv*, *im*, *sc* injection; *in utero*, lactational	Dermal (topical); or uncertain
Chemical Purity			
>99%	>98%	>95% or less; or not stated	A specific problem is known/suspected
Exposure duration			
in vivo studies			
7-14 days	3-6 days	Single dose or <3 days	Other, or uncertain
in vitro (cell culture) studies			
24 hours (most cell lines)	48 hours	72 hours	Other (<6 hours; >72 hours), or uncertain
Delay between Final Dose and Measurement of Effect (in vivo studies only)			
1 day	2-3 days	>3 days	>7 days or uncertain
Measurement Endpoint			
Tumor promotion/formation of neoplastic foci; CYP1A1 induction; immunosuppression (PFC response w/ TNP-LPS as antigen).	Vitamin A levels, GST, UDPGT, antiestrogen-icity, DRE-driven reporter gene activity, thymic involution; lymphoid development; hydronephrosis, cleft palate; CYP1A2 induction.	Immunosuppression (PFC response w/ SRBCs as antigen), porphyria, thyroid hormone changes, e.g., fT4, or UGT activity; intercellular communication; terminal differentiation (keratinocytes).	Acute lethality; AhR binding; body weight loss; liver weight or lipid levels, weight changes/histopathol-ogy in other organs; protein content; aromatase activity (CYP19).

Each study element was placed into one of 3 tiers, in accordance with the rank order importance: Tier 1 > Tier 2 > Tier 3. Tier 1 elements (Table 1) included the most critical measures of study relevance and reliability, such as route of administration, chemical purity, and measurement endpoint and were graded on a four-point scale: 1, 4, 7, and 10. Tier 2 elements are related to the selection of species/strain, characterization of the dose-response curve and the derivation of the REP (Table 2). In general, tier 2 elements were roughly judged as being optimal, adequate, and inadequate, and given a score of 6, 3, or 0, respectively. Tier 3 elements (Table 2) are important elements of any study design, but were found to contribute very little to the difference between studies, and thus, would not contribute to quality-based distinctions between studies. The studies that furnished information on these elements were given either 1 or 2 points; and those that provided no information were given a score of zero.

Table 2. Quantitative Rating Scheme for Tier 2 and 3 Study Elements

Tier 2 Elements		
6	**3**	**0**
Species/Strain		
Highly conventional/ responsive model; or similar biology to humans (C57BL/6, B6C3F1 mice, SD, LE, F344 rats; monkeys)	Conventional/ responsive model (Hartley guinea pigs, Hamsters, H/W rats)	Less conventional/ less responsive model (DBA/2 mice)
Number of dose levels tested		
>3	3	<3
Tissue type (in vivo only)		
Liver, lung, skin and thyroid	Thymus, reproductive organs (or fetuses), spleen	Brain, CNS, serum (enzymes), whole body (e.g., body weight), kidney, muscle
Maximal Response Attained?		
Maximum response clearly achieved	Near-maximum response achieved; some uncertainty	Maximum response not achieved or can not be determined with data provided
Method of Derivation; quality of dose-response modeling		
Comparison of ED_{50}s or similar metrics; full dose-response curves for both test and reference compound	REP calculated by IEM and/or linear interpolation; calc-ulations not transparent; or significant shortcomings in dose-response curves	Crude estimate only; methods and calculations not provided.
Tier 3 Elements		
2	**1**	**0**
Vehicle		
Corn oil, feed (*in vivo*) DMSO or isooctane (*in vitro*)	Any other choice	Not stated
Animal age		
Young adult animals	Immature or older adult animals	No information provided
Number of animals per treatment group		
>3	3	<3
Controls		
Appropriate and stable	Unstable, “shared”, or irregular	No information provided
Reference Compound		
TCDD	--	PCB 126
Animal sex		
Female	Male	Not stated

Each REP was graded by giving a score to each of the applicable study elements (16 *in vivo* study elements, 10 *in vitro* study elements) and summing the individual scores to obtain a total score. Overall, we applied this type of evaluation to 98 REPs contained in 39 different studies (14 studies for 1,2,3,7,8-PeCDF and 25 studies for PCB 126). The maximum possible score for an *in vivo or in vitro* study REP was 92 and 64, respectively. The total score assigned to each REP was then converted to a weighting factor based on the percentage of the maximum points possible (Table 3). Lastly, an additional weighting factor of 3 was applied to all *in vivo* studies.

Table 3. Assignment of Weightings to the REPs

Overall Score (as %Total)	Weighting Factor
85—100	10
75—84	5
65—74	2.5
<64	1
Disqualified	0

Results and Discussion

A Monte Carlo simulation was used to generate weighted distributions of REP values for PCB 126 and 1,2,3,7,8-PeCDF. The mean, 50th, and 95th percentiles are summarized in Table 4. The overall results are consistent with the current W.H.O. TEFs, i.e., for both congeners, the W.H.O. TEF approximates the 50th percentile of the distribution. In both cases, the 95th percentile is several-fold greater than the W.H.O. TEF. The data cleanup step had the greatest impact at the 95th percentile of the 1,2,3,7,8-PeCDF distribution; prior to removal of the duplicative and erroneous values, the 95th percentile was 0.8, versus 0.2 after data cleanup (unweighted distributions). We are currently evaluating the influence of different weighting factors and weighting schemes.

Table 4. Statistical descriptors of weighted REP distributions

Statistic	**Value**
1,2,3,7,8-PeCDF (WHO TEF = 0.05)	
Mean	0.05
50^{th} %-ile	0.02
95^{th} %-ile	0.1
PCB 126 (WHO TEF = 0.1)	
Mean	0.17
50^{th} %-ile	0.1
95^{th} %-ile	0.7

REP distributions such as those derived here can be used directly in a PCDD/F or PCB probabilistic assessment. Alternatively, or in addition, point estimates from the distributions (e.g., the 50^{th} and 95^{th} percentiles) can be used in deterministic assessments. We believe the use of this framework for data cleanup and evaluating REP data will lead to a more consistent derivation of PCB and PCDD/F TEFs. Further, use of a consistent weighting scheme will permit a more informed and quantitative analysis of the uncertainties in the TEFs. Over 50 papers have been published since 1997 that are likely to contain new REP data; we are continuing work aimed at developing distributions for other key congeners, with the intention of incorporating these new data.

References

1. Patterson, D., Todd, G., Turner, W., Maggio, V., Alexander, L. and Needham, L. 1994. Environ. Health Perspect. Suppl. 102(1):195-204.
2. Van den Berg, M., Birnbaum, L., Bosveld, A., Brunstrom, B., Cook, P., Feeley, M., Giesy, J., Hanberg, A., Hasegawa, R., Kennedy, S., Kubiak, T., Larsen, J., van Leeuwen, F., Liem, A., Nolt, C., Peterson, R., Poellinger, L., Safe, S., Schrenk, D., Tillitt, D., Tysklind, M., Younes, M., Waern, F., and Zacharewski, T. 1998. Environ. Health Perspect. 106: 775-92.
3. The WHO REP database was provided by Dr. Frederik Waern (Institute of Environmental Medicine (IEM) at the Karolinska Institute (Sweden).