AN ALTERNATIVE METHOD FOR ESTABLISHING TEFs FOR DIOXIN-LIKE COMPOUNDS. PART 2. DEVELOPMENT OF AN APPROACH TO QUANTITATIVELY WEIGHT THE UNDERLYING POTENCY DATA

Haws LC[1], DeVito MJ[2], Birnbaum LS[2], Walker NJ[3], Scott PK[4], Unice KM[4], Harris MA[5], Farland WH[6], Finley BL[7], and Staskal DF[1]

[1]ChemRisk, Austin, TX; USA, [2]USEPA, ,ORD, NHEERL, RTP, NC; USA, [3]NIEHS, NIH, RTP, NC; USA, [4]ChemRisk, Pittsburgh, PA; USA, [5]ChemRisk, Houston, TX; USA, [6]USEPA, ORD, Washington, DC; USA, [7]ChemRisk, San Francisco, CA, USA

Introduction

The current approach for evaluating potential health risks associated with exposure to mixtures of polychlorinated dibenzo-p-dixoins (PCDDs), polychlorinated dibenzofurans (PCDFs) and dioxin-like polychlorinated biphenyls (PCBs) [hereafter referred to as "dioxin-like compounds"] is based on the toxic equivalency factor (TEF) methodology. In accordance with this methodology, each PCDD, PCDF, and PCB congener believed to exhibit dioxin-like activity has been assigned a TEF based on comparison to 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD). The current TEFs represent consensus-based values recommended the World Health Organization (WHO)[1]. In assigning TEFs to each congener, the WHO expert panel employed scientific judgment and a qualitative weighting scheme whereby individual relative estimates of potency (REPs) from *in vivo* studies were given greater weight than *in vitro* studies and/or quantitative structure activity relationship (QSAR) data; chronic studies were given greater weight than subchronic studies, which were given greater weight than subacute studies, which were given more weight than acute studies; and Ah-mediated toxic responses were given more weight than biochemical responses (e.g., enzyme induction)[1].

The TEF methodology has been re-evaluated in a variety of forums over the past 20 years. Recently, investigators proposed basing risk estimates on the distribution of REP values for each congener to allow for better characterization of the uncertainty and variability inherent in the risk estimates that are based on the TEFs[2,3]. We believe that such an approach is important given that the underlying REPs for most congeners are derived from a heterogeneous data set, and the values themselves often span several orders of magnitude[1,2,4,5].

Recently, Haws and coworkers published a refined database of REPs and presented distributions of REP values for each congener[3]. However, those distributions were based on treating all REPs equally, despite the many differences between the studies from which the REP values were obtained (e.g., different species, study designs, endpoints, REP calculation methods, etc.). The development of a framework to quantitatively assess differences in study quality and relevance would allow one to place greater emphasis on those REP values believed to be more well-suited for purposes of human risk assessment. In this paper, we present a possible quantitative weighting scheme for consideration.

Methods

The first step in developing a quantitative weighting scheme involved evaluating different decision analysis methods to identify the most suitable approach for aggregating subjective decision criteria to rank REP values with respect to quality and relevance as described by Scott and coworkers[6]. These authors concluded that the Analytical Hierarchy Process (AHP) was the preferred framework as it can incorporate both multiple value comparison scales (different levels of better or worse) and a binary scale (better or not) and is well documented in the scientific literature. The next step, which is the subject of this paper, involved selecting the specific study elements to include in our quantitative weighting scheme, as well as the specific numerical values that would be applied to each of the study

elements. It was determined that the focus should be on those study elements that most impacted REP quality and relevance, while keeping the weighting scheme as simple as possible.

The original qualitative criteria employed by the WHO expert panel in 1997 (described above) were used as a starting point. It was determined that study type was an important criterion and should be retained, and that *in vivo* and *in vitro* REP values should be evaluated separately, as well as combined. In the case where *in vivo* + *in vitro* REPs were combined, it was concluded that those REPs based on *in vivo* studies should be given more weight than those based on *in vitro* studies or QSAR analyses, as the *in vivo* studies are believed to be more relevant for human risk assessment. With respect to weighting *in vivo* studies of longer duration more than those of shorter duration, it was concluded that the primary purpose of this criterion was to address potential differences in time to reach pharmacokinetic steady state. As such, this criterion was refined, focusing instead on whether differences in pharmacokinetics (PK) were accounted for in the study design. In cases where congeners were believed to exhibit PK properties that were similar to TCDD (i.e., 2,3,4,7,8,pentachlordibenzo-p-dioxin and PCB126), it was concluded that study design would not dramatically affect the relative potency estimate, and, as such, those congeners were assumed to satisfy the PK criterion. Additionally, it was concluded that all congeners would be close to achieving PK steady state following subchronic or chronic exposures and, as a result, all such studies would be assumed to satisfy the PK criterion. For all other congeners and study durations, differences in PK properties would have to be explicitly addressed in the study design to satisfy the PK criterion. With regards to endpoints, it was concluded that this criterion should be retained and that toxic endpoints should be given more weight than biochemical endpoints.

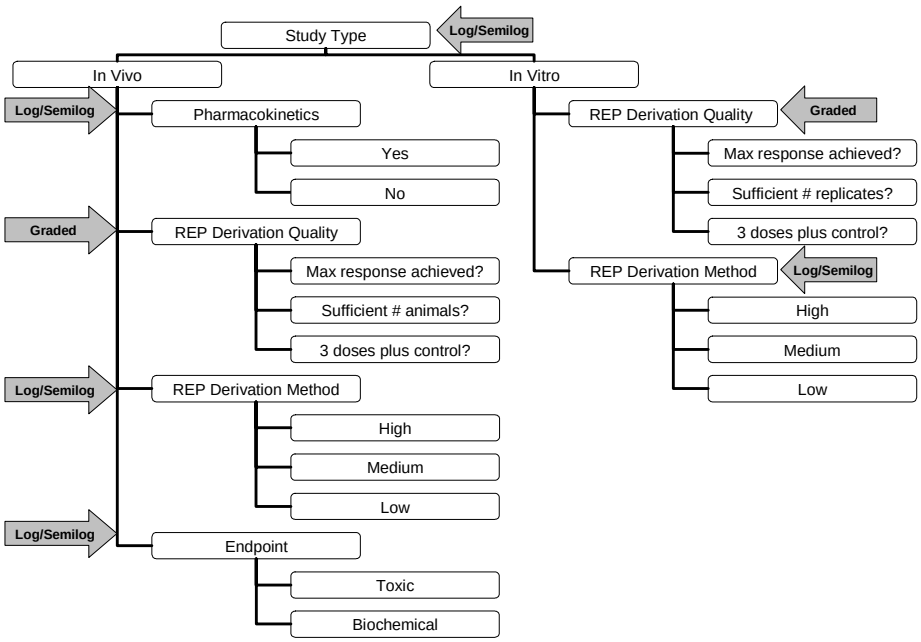
In addition to the original WHO criteria, several other factors were determined to be important contributors to REP quality. These included the quality of the underlying dose response data ("REP derivation quality") and the specific method used to derive the REP value ("REP derivation method"). The underlying dose response data was determined to be of high quality when the following criteria were met: 1) there was a sufficient number of dose levels (i.e., at least 3 dose levels + control); 2) there was a sufficient number of animals for the *in vivo* studies (i.e., cancer [N=20]; tumor promotion [N=10]; immunotoxicity & developmental toxicity endpoints [N=6]; histopathology, body weight, organ weights, endocrine endpoints [N=5]; all others [N=3]) or a sufficient number of replicates for the *in vitro* studies (i.e., at least 2 intra- or inter-study replicates); and 3) a maximum response was achieved or it was not necessary to achieve an observed maximum response because of the REP derivation method used. As indicated by Haws and coworkers, a wide variety of methods were relied upon to calculate the REP values, ranging from statistically based non-linear dose response modeling to linear graphical techniques and LOEL/NOEL ratios[3]. For purposes of weighting, the REP calculation methods were grouped into one of three possible categories (high, medium, low) based on their perceived ability to accurately estimate a REP. As an example, those REP derivation methods that involved using statistical models and included evaluation of parallelism of dose response (a fundamental assumption in the use of TEFs) were categorized as "high". In cases where statistical models were employed but parallelism was not assessed, the REP derivation method was categorized as "medium". Other types of derivation methods classified as "medium" included ED50 or EC50 ratios, promotion index ratios, development of dose response graphs, etc. Those that involved crude estimates, NOEL/LOEL ratios, or response ratios were categorized as "low".

Results and Discussion

Our initial quantitative weighting scheme is depicted in Figure 1 and is based on a non-parallel framework where *in vivo* REP quality was determined based on consideration of pharmacokinetics, the quality of the underlying dose response data ("REP derivation quality"), the perceived accuracy of the REP derivation method, and the nature of the endpoint upon which the REP value was based. For the *in vitro* studies, REP quality was determined based on the quality of the underlying dose response data and the perceived accuracy of the REP derivation method. In accordance with this scheme, both log and semi-log scales were evaluated for all study elements except for the REP

derivation method, which was evaluated based on a graded scale where a value of +1 was assigned for each criterion satisfied.

Figure 1: Schematic of the Quantitative Weighting Scheme



A numerical value was then assigned to each study element and the REP values were then compared against one another to determine the weight for each specific study element. This is illustrated in Figure 2A, where each cell in the AHP matrix represents the relative judgments of quality or importance for paired REP comparisons. The REP weights were determined using an algebraic solution (i.e., the eigenvector) such that the ratios for all paired comparisons were taken into account simultaneously. The weights for each study element were then combined as illustrated in Figure 2B, with each study element being given equal weight, to calculate the overall weight for the REP value. The overall weights for each individual REP were then combined to prepare a cumulative distribution for each congener, which in turn was used to determine the 10^{th}, 25^{th}, 50^{th}, 75^{th}, and 90^{th} percentiles of the weighted distributions.

In conclusion, we believe that the quantitative weighting scheme described in this paper provides an approach for placing greater emphasis on those REP values that are believed to be more well suited for health risk assessment purposes. Applying such a framework to the REP values underlying each assigned TEF will allow for development of weighted distributions of REP values that should ultimately facilitate better characterization of the variability and uncertainty inherent in the health risk estimates for this class of compounds. The use of distributions will also give risk managers the flexibility to tailor the desired level of protection to the specific situation under consideration and will facilitate establishment of a consistent level of protection for all congeners. Finally, the development and application of a quantitative weighting scheme will also yield a more transparent, reproducible, and consistent method for deriving TEFs from the underlying REP data.

Figure 2:
Determining the Overall Weight for Each REP Value-An Example Calculation

A) .

REP	A (High)	B (High)	C (Med)	D (Low)	Weight
A (High)	1	1	10	100	0.474
B (High)	1	1	10	100	0.474
C (Med)	0.1	0.1	1	10	0.047
D (Low)	0.01	0.01	0.1	1	0.005

Example: REP "D" Derivation Quality Rated 100 times lower than REP "A" Quality

B)

PK

REP	A (Y)	B (Y)	C (Y)	D (Y)	Weight
A (Y)	1	1	1	1	0.25
B (Y)	1	1	1	1	0.25
C (Y)	1	1	1	1	0.25
D (Y)	1	1	1	1	0.25

$x \frac{1}{3}$ +

REP - Quality

REP	A (N/Y/Y)	B (N/Y/Y)	C (N/Y/Y)	D (Y/Y/Y)	Weight
A (N/Y/Y)	1	1	1	0.75	0.231
B (N/Y/Y)	1	1	1	0.75	0.231
C (N/Y/Y)	1	1	1	0.75	0.231
D (Y/Y/Y)	1.33	1.33	1.33	1.33	0.308

$x \frac{1}{3}$ +

REP - Derivation

REP	A (High)	B (High)	C (Med)	D (Low)	Weight
A (High)	1	1	10	100	0.474
B (High)	1	1	10	100	0.474
C (Med)	0.1	0.1	1	10	0.047
D (Low)	0.01	0.01	0.1	1	0.005

$x \frac{1}{3}$

REP	Final Vivo Weight
A	0.318
B	0.318
C	0.176
D	0.187

Acknowledgements

The research presented in this document was funded in part by Tierra Solutions, Inc. Drs. Birnbaum, DeVito, and Farland were supported by the Office of Researched and Development, USEPA, while Dr. Walker was supported by the Intramural program at the NIEHS, NIH. The contents of this paper reflect the opinions and views of the authors and do not represent the official views of NIEHS, NIH, or USEPA. The mention of trade names or commercial products does not constitute endorsement or recommendation for use. We would like to extend our thanks to Drs. Richard Peterson, Angelika Tritscher, and Martin Van den Berg for their input on the development of a quantitative weighting scheme.

References

1. Van den Berg M, Birnbaum L, Bosveld ATC, Brunstrom B, Cook P, Feeley M, Giesy JP, Hanberg A, Hasegawa R, Kennedy S, Kubiak T, Larsen JC, van Leeuwen FXR, Liem AKD, Nolt C, Peterson RE, Poellinger L, Safe S, Schrenk D, Tillitt D, Tysklind M, Younes M, Waern F, and Zacharewski T (1998) *Environ. Health Perspect*. 106(12), 775.
2. Finley BL, Connor KT, and Scott PK (2003) *J. Toxicol. Environ. Health. Part A*. 66, 533.
3. Haws LC, Su SH, Harris M, DeVito MJ, Walker NJ, Farland WH,, Finley B, and Birnbaum LS (2006) *Toxicological Sciences.* 89, 4.
4. U.S. Environmental Protection Agency (2000) SAB Review Draft. Chapter 9. EPA/600-P-00/001BE.
5. Birnbaum LS, Emond C, and DeVito MJ (2004) *The Toxicologist* 78, 362.
6. Scott PK, Haws LC, Staskal DF, Birnbaum LS, Walker NJ, DeVito MJ, Harris MA; Farland WH, Finley BL, and Unice KM (2006) *Organohalogen Comp*, submitted.