AN ALTERNATIVE METHOD FOR ESTABLISHING TEFs FOR DIOXIN-LIKE COMPOUNDS. PART 1. EVALUATION OF DECISION ANALYSIS METHODS FOR USE IN WEIGHTING RELATIVE POTENCY DATA

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

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

Introduction

A number of investigators have recently examined the utility of applying probabilistic techniques in the derivation of toxic equivalency factors (TEFs) for polychlorinated dibenzo-p-dioxins (PCDDs), polychlorinated dibenzofurans (PCDFs) and dioxin-like polychlorinated biphenyls (PCBs) [hereafter referred to as "dioxin-like compounds"].[1,2] The current mammalian TEFs established by the World Heath Organization (WHO)[3] represent single assigned point values despite the fact that they are derived from underlying distributions of individual REP values. It is believed that the use of a distributional approach will allow for better characterization of the uncertainty and variability inherent in the health risk estimates that are based on the TEF values assigned by the WHO. During their most recent re-evaluation of the TEF methodology in June 2005, the WHO expressed interest in discussing a probabilistic approach to deriving TEFs for dioxin-like compounds

In a recent publication by Haws and coworkers[2], the authors presented a refined database of mammalian REPs (relative estimates of potency) for dioxin-like compounds, along with information concerning the distribution of REP values for each congener. However, in developing these distributions, all REPs were treated equally regardless of underlying differences in study quality or relevance. A logical next step in the development of REP distributions is the application of a quantitative weighting scheme to place greater emphasis on those REP values that are of greater quality or are more relevant to humans.

Preliminary REP weighting schemes have been fairly simple, evaluating from 2 to 4 different study factors. For example, Finley et al.[1] developed a weighting scheme based on study type (*in vivo* or *in vitro*) and endpoint using weighting scale based on powers of 10 while Harris et al.[4] developed a weighting scheme based on study type, study duration, multiple versus single dosing, and exposure route using a linear weighting scale.

Another approach for developing REP weighting schemes is to use existing decision analysis methods to develop weights based on multiple decision factors. In this analysis, we will evaluate two decision analysis methods – the Paired Comparison Technique (PCT) described by Dean and Nishry[5] and the Analytical Hierarchy Process (AHP) developed by Saaty.[6] These methods provide an objective framework for aggregating subjective comparisons of the multiple REPs based on a set of decision factors related to REP derivation quality and method. In this analysis we apply these methods to the REP database for PCB 126 based on the work of Haws et al.[7] and compare the weighted distributions generated by them.

Methods

Preliminary analyses were conducted using mock data to identify the preferred decision analysis framework. For purposes of discussion in this paper, we evaluated the different decision methods using the REP database for PCB 126 from Haws et al.[2] and the quantitative weighting scheme outlined in the companion paper by Haws and

coworkers.[7] The weighting criteria, or decision factors, used in each decision method in this study are presented in detail in Haws et al.[7] Briefly, these factors were segregated based on study type (*in vivo* versus *in vitro*). For *in vivo* studies, pharmacokinetics, REP derivation quality, and REP derivation method were evaluated and for *in vitro* studies, only REP derivation quality and REP derivation method were considered. In a separate paper by Staskal and coworkers, we evaluate the impact of different quantitative weighting schemes based on the AHP framework on the distribution of REP values for PCB126[8]. The Haws et al database contains 115 REP values for PCB 126 from 38 studies, with 86 REPs based on *in vivo* studies and 29 REPs based on *in vitro* studies[2].

REP distributions for PCB 126 were evaluated according to two decision analysis methods, PCT and AHP. The PCT method consists of three steps – the ranking of each REP based on each decision factor, the ranking of each decision factor relative to each other, and the combination of the rankings for each REP based on each decision factor to produce an overall weight for each REP.[5,9] For the first step each REP is compared relative to every other REP in a pairwise manner for each decision factor. The more desirable REP of the pair is given a rank of 1, and the least desirable REP is given a rank of 0. If the REPs have equal desirability, a rank of 0.5 is assigned to both REPs. The overall weight for each REP for each decision factor is the sum of the rankings across all of the other REPs to which the initial REP is compared divided by the total ranks for all of the studies. For the second step, the decision factors are compared to each other in the same manner as the REPs were compared for each decision factor. The overall weights for all decision factors are the sum of the products of the decision factor weight and the REP weight for the decision factor across all decision factors.

The AHP method consists of similar steps – the ranking of each REP based on each decision factor, the ranking of each decision factor relative to each other, and the combination of the rankings for each REP based on each decision factor to produce an overall weight for each REP.[6,10,11] The biggest differences between the AHP and the PCT methods are that the AHP method can use any relative scale to develop ranks,[6] that the AHP develops weights for each REP based on each decision factor by estimating the eigen values associated with the matrix of the associated comparisons,[6] and that there is an extensive literature that discusses the use of AHP for a variety of applications, including engineering feasibility studies, environmental risk ranking, and environmental impact analysis.[11,12] Because of the more sophisticated mathematical design of AHP, there is much flexibility in the application of the method, the consistency of the pairwise comparisons can be easily evaluated, and relative scales based on either subjective judgment or objective data can be used together to develop weights.[10]

For AHP, three different scales were evaluated. The first scale, the binary scale, was similar to the one used for the PCT using a value of 2 if the REP was more desirable than the other REP, 0.5 if the REP was less desirable, and 1 if they were equally desirable. The second and third scales were based on the decision factor hierarchy presented in Hawes et al.[7] For the semi-log scale, decision factors for which REP values were ranked low, medium, and high were given individual ranks of 1, 3, and 10, respectively. For the log scale, decision factors for which REP values were ranked low, medium, and high were given individual ranks of 1, 10, and 100, respectively.

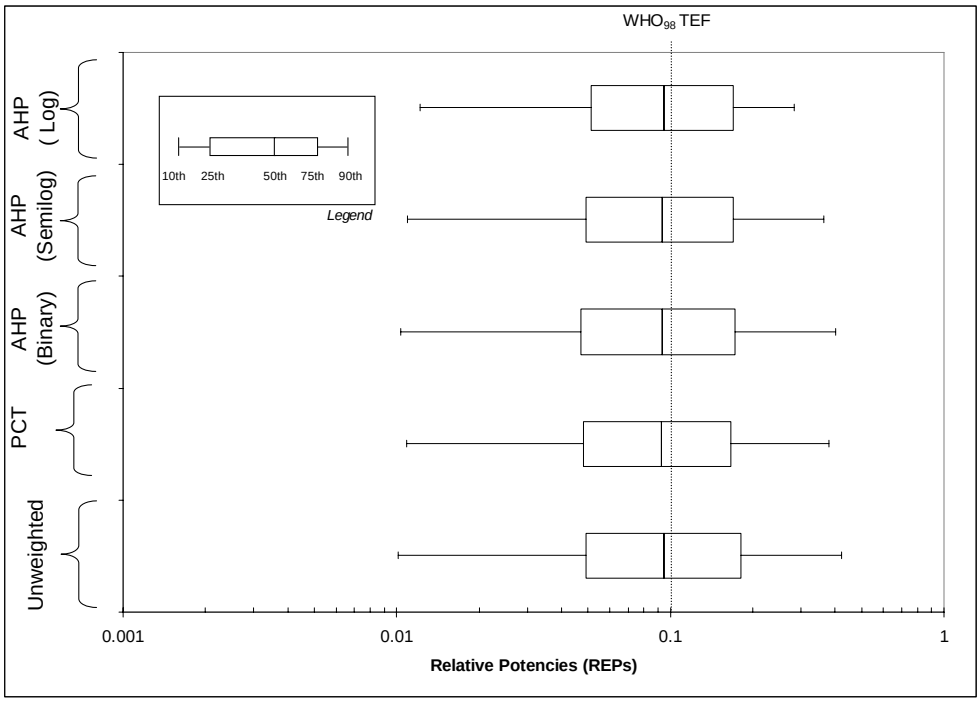
For these two scales, the rank used in the pairwise comparison for each decision factor is the ratio of these ranks for the two REP values. For example, if REP A was high for REP derivation method and REP B was low, the pairwise comparison of A versus B for the semi-log scale would have a rank of 10/1 or 10 and for the log scale a rank of 100. Conversely, the ranks for the semi-log and log scales for the comparison of B versus A would have ranks of 1/10 or 0.1, and 1/100 or 0.01, respectively.

Results and Discussion

Figure 1 presents the box plots for the unweighted and the four weighted REP distributions. In general, there are only slight differences in the REP distribution for PCB 126 across the four weighting methods. All of the weighting methods produce similar medians and lower percentiles that are not different than those for the unweighted

distribution. However, for the three AHP methods, the 95th percentile appears to decrease as the ranking scale increases from the binary to semi-log and from semi-log to log scales. Comparing PCT to the AHP method using a binary scale, there appears to be no difference in the two REP distributions. In addition to the reduction in the 95th percentiles for the four weighting methods, there is a noticeable decrease in the variability of the REP distributions as rankings move from a binary to log scale (Figure 2).

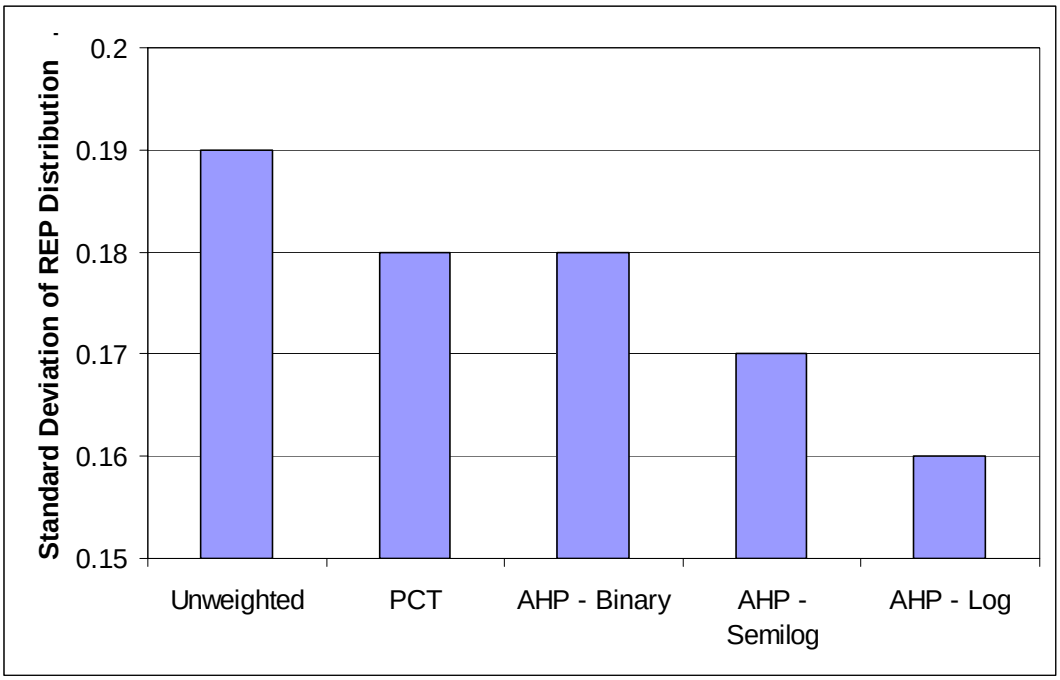
Figure 1. Summary of REP distributions for Each Weighting Method



While the REP distributions for the PCT and the AHP methods using the same ranking scale show no substantial difference from the unweighted distribution, the AHP method is preferred because several different relative ranking scales can be used. In general, as the ratio between the worst and best REP for any decision factor increases for the AHP method, the variability in REP value decreases and the upper bound values (95th percentiles) decrease. Use of a log scale with the AHP method to develop a REP distribution appears to lead to the highest degree of difference from the unweighted distribution and a reduction in the variability for the REP distribution. For these reasons, the AHP method will be used to develop weighted REP distributions as described in Haws et al.[7] and Staskal et al.[8]

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.

Figure 2. Relationship Between Weighting Method and Variability of REP Distribution



References

1. Finley BL, Connor KT, and Scott PK (2003) *J. Toxicol. Environ. Health. Part A.* 66, 533.
2. Haws LC, Su SH, Harris M, DeVito MJ, Walker NJ, Farland WH,, Finley B, and Birnbaum LS (2006) *Toxicological Sciences.* 89, 4.
3. 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.
4. Harris M, Finley B, and Scott P (2005). *Organohalogen Comp.* 67, 2479.
5. Dean BV and Nishry MJ (1965). *J. Operations Res. Soc. Am.* 13, 550.
6. Saaty, TL (1990). *European Journal of Operational Research.* 48, 9.
7. Haws LC, DeVito MJ, Birnbaum LS,, Walker NJ, Scott PK, Unice KM, Harris MA, Farland WH, Finley BL,, and Staskal DF (2006). *Organohalogen Comp.* Submitted.
8. Staskal DF, Unice KM, Walker NJ, DeVito MJ Birnbaum LS, Scott PK, Harris MA, Farland WH, Finley BL and Haws LC (2006). *Organohalogen Comp.* Submitted.
9. Clifford PA, Atkinson SF., Hinckley DA, Suedel BC, Ludwig DF, and Dillon TM (1998). *Human and Ecological Risk Assessment.* 4, 1189.
10. Saaty TL (1994). *European Journal of Operational Research.* 74, 426.
11. Ramanathan R (2001). *J. Environ. Manage.* 63, 27.
12. Saaty TL (1987). *Risk. Anal.* 7, 159.