

Use of Evidence Integration Techniques to Derive Toxicity Reference Values



U.S. ARMY PUBLIC HEALTH CENTER

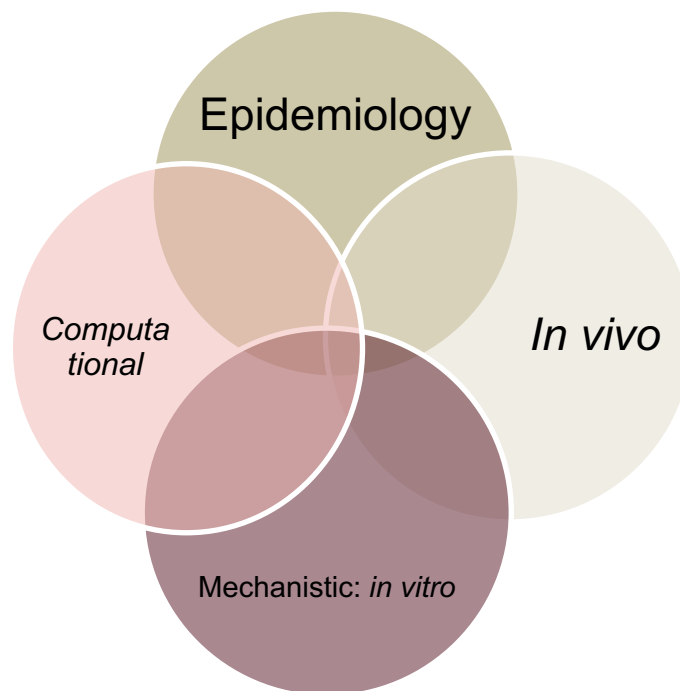
Mark S. Johnson, Ph.D., DABT, Fellow ATS
Toxicology Directorate, Defense Health Agency
Defense Public Health-Aberdeen

Disclaimer

The views and assertions expressed in this presentation are the private views of the presenter and do not necessarily reflect, nor should they be construed as reflecting, the views or official policy of the Department of Defense, the Department of the Army, the U.S. Army Medical Department, or the U.S. Federal Government.

The mention of any non-federal entity and/or its products is not to be construed or interpreted, in any manner, as federal endorsement of that non-federal entity or its products.

- Epidemiology (field) -
 - Pros: Target species, realistic exposure scenarios
 - Cons: Confounders (mixtures other stressors); lack of cause/effect
- *In vivo* animal
 - Pros: Models include toxicokinetics and dynamics
 - Cons: Animals are different than humans
- *In vitro*
 - Pros: Can investigate low doses in human cell lines
 - Cons: Missing interconnectivity, kinetics, adaptation
- Computational
 - Pros: Relatively fast, based on shared molecular moieties/Modes of Action
 - Cons: Models often incorrect



- Animal studies
 - Controlled
 - Includes absorption, disposition, metabolism, excretion (ADME)
- Human data
 - Typically correlational evidence; confounders and exposure issues
- *In vitro*
 - In vitro/In vivo extrapolation (IVIVE) –
 - Lacks (ADME) but can model this
 - Only see what you are looking for
 - Lack interorgan/cellular effects
- Mechanistic/Mode of action
 - Addresses biological plausibility

sch et al., 1997.pdf - Adobe Acrobat Pro 2020

edit View Window Help

Tools Klimisch et al., 199... x

1 / 7

REGULATORY TOXICOLOGY AND PHARMACOLOGY 25, 1-5 (1997)
ARTICLE NO. RT961076

A Systematic Approach for Evaluating the Quality of Experimental Toxicological and Ecotoxicological Data¹

H.-J. KLIMISCH,² M. ANDREAE, AND U. TILLMANN
BASF Aktiengesellschaft, D-67056 Ludwigshafen, Germany

Received November 7, 1996

The evaluation of the quality of data and their use in hazard and risk assessment as a systematic approach is described. Definitions are proposed for reliability, relevance, and adequacy of data. Reliability is differentiated into four categories. Criteria relating to international testing standards for categorizing reliability

supplied data must be evaluated considering their quality and adequacy for a risk assessment. Some general guidelines on data evaluation were published by the European Commission in a "Technical Guidance Document" (EU, 1994, 1995), based on general principles for data evaluation of the International Coordination of Criteria Document Production (IPCS, 1993).

Code	Category
1	Reliable without restriction
2	Reliable with restrictions
3	Not reliable
4	Not assignable

A	B	C	D
1	Reliability assessment of in vitro toxicity studies		
2			
3	Study under evaluation		
4	Authors:		
5			
6	Titel:		
7			
8	Testing facility, year, sponsor, study no. or bibliographic reference:		
9			
10			
11	Explanations are available for most criteria and show up, when the cursor is moved over the criteria field. Please read carefully! Red criteria: the maximum score is needed for these criteria to achieve reliability category 1 or 2 (see worksheet Explanations): Please evaluate with special care!		
12			
13			
14	Criteria		Evaluator's explanations, comments on criteria, etc.
15	No. Criteria Group I: Test substance identification	Score	
16	1 Was the test substance identified?		
17	2 Is the purity of the substance given?		
18	3 Is information on the source/origin of the substance given?		
19	4 Is all information on the nature and/or physico-chemical properties of the test item given, which you deem <u>indispensable</u> for judging the data (see explanation for examples)?		
20		0	
21			
22	Criteria Group II: Test system characterisation		
23	5 Is the test system described?		
24	6 Is information given on the source/origin of the test system?		
25	7 Are necessary information on test system properties, and on conditions of cultivation and maintenance given?		
26		0	
27			
28	Criteria Group III: Study design description		
29	8 Is the method of administration given (see explanations for details)?		
30	9 Are doses administered or concentrations in application media given?		
31	10 Are frequency and duration of exposure as well as time-points of observations explained?		
32	11 Were negative controls included (give also point, if not necessary, see explanations)?		
33	12 Were positive controls included (give also point, if not necessary, see explanations)?		
34	13 Is the number of replicates (or complete repetitions of experiment) given?		
35		0	
36			
37	Criteria Group IV: Study results documentation		

←
→
Rater
Explanations
in_vivo
in_vitro
+

- Rooney et al. 2014
- Study quality and utility
- Added a PF step (PECO statement)
- Document literature review
- Clarifies decision criteria for study exclusion
- Risk of bias tool
- Qualitative
- Other tools (GRADE, Distiller SR, HAWK)

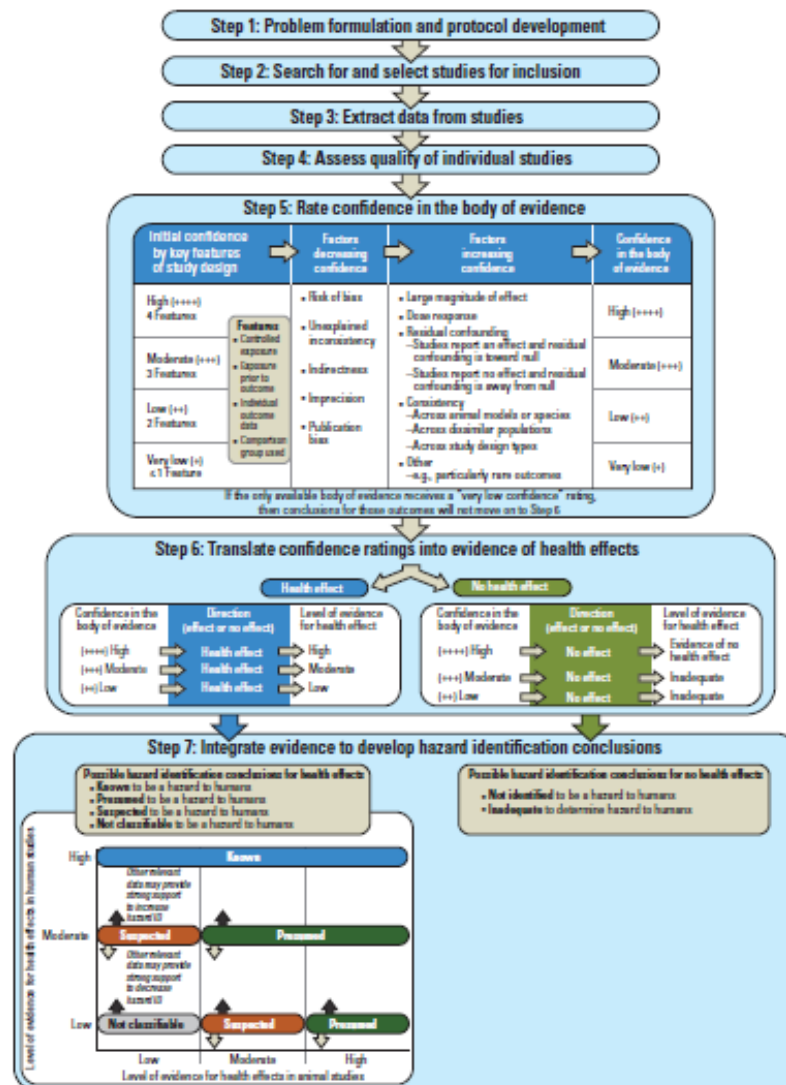


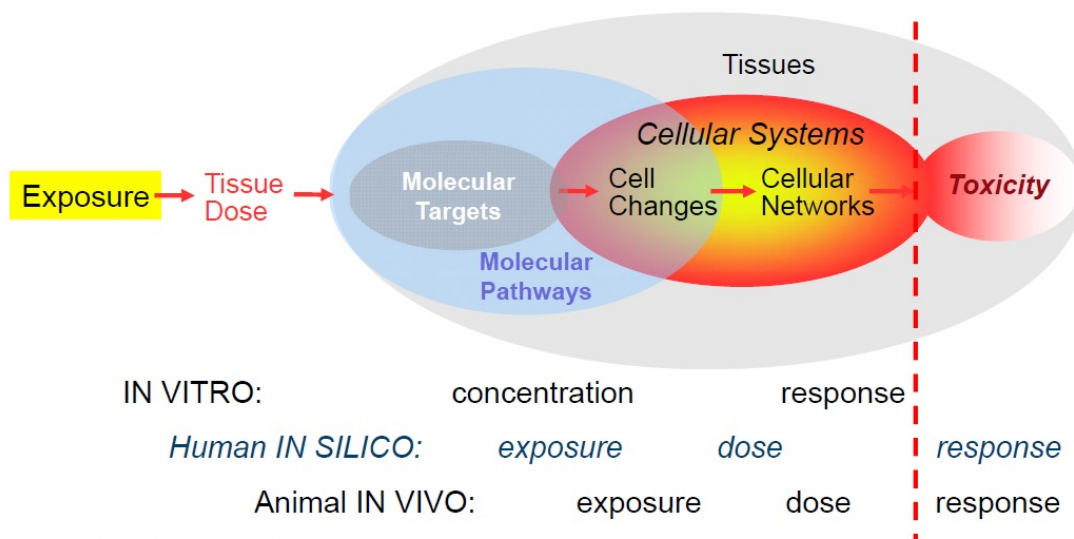
Figure 1. The OHAT Approach for systematic review and evidence integration for literature-based environmental health science assessments.

- “We know accurately only when we know little; with knowledge, doubt increases” (Johann Wolfgang von Goethe)*
- Also “...uncertainty can prevail in situations where a lot of information is available new information can either decrease or increase uncertainty. New knowledge on complex processes may reveal the presence of uncertainties that were previously unknown or were understated. In this way, more knowledge illuminates that our understanding is more limited or that the processes are more complex than thought before” (Walker et al. 2003).

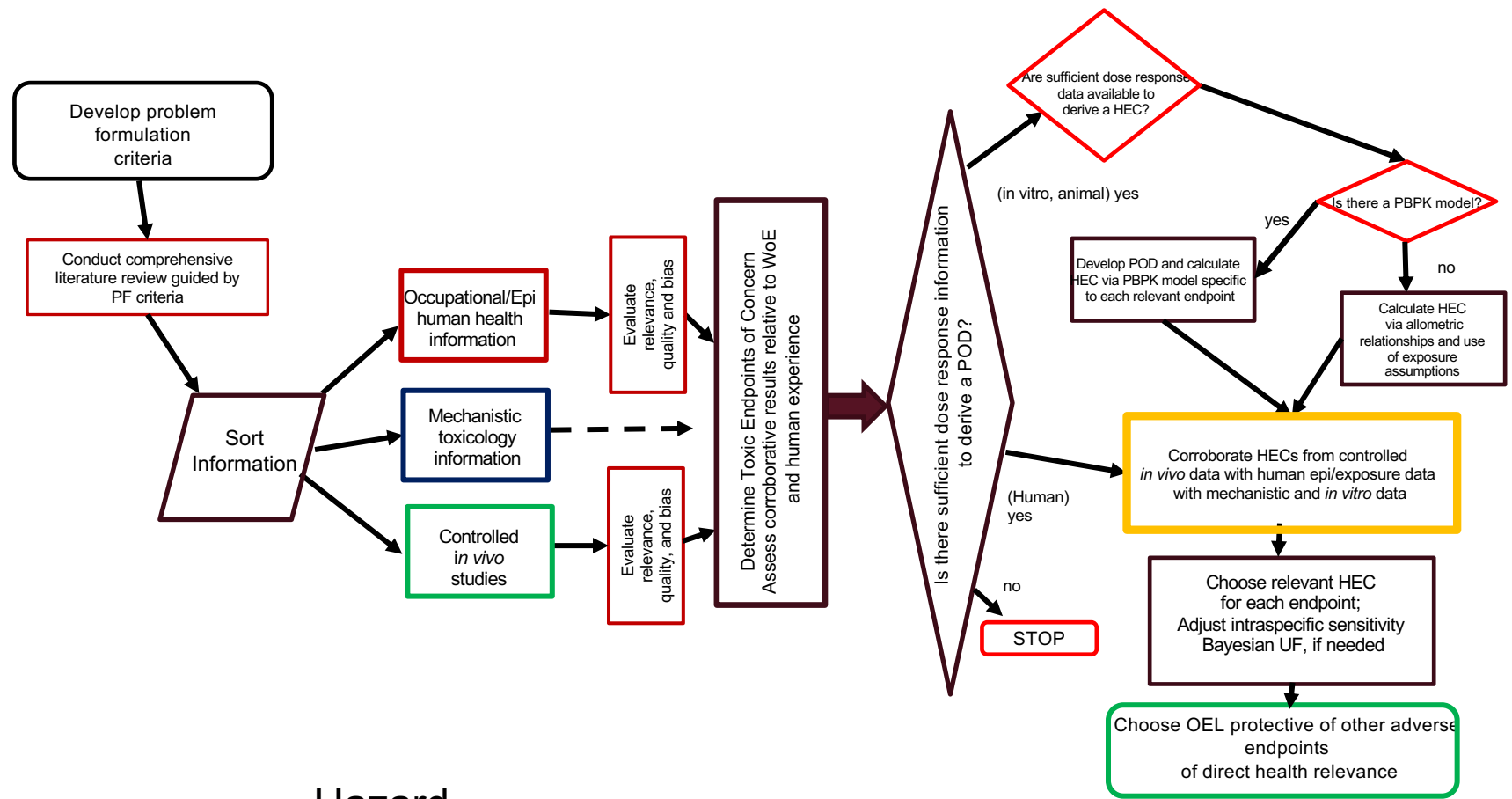
- Data quality
 - Study design
 - Analytical chemistry – confirmation of exposure/dose
- Risk of bias
 - Selection bias
 - Performance bias (confounding)
 - Detection bias
 - Reporting bias
 - Negative publication bias
- Type I Error (false positives)
- Type II Error (false negatives)
- More data – more issues

- Computational modeling
- Biochemical/molecular endpoints
- Physiological endpoints
- Functional endpoints

HIGH THROUGHPUT SCREENING and CHARACTERIZATION OF HAZARD & RISK



- Problem Formulation
 - PECO – (populations, exposures, comparators, outcomes)
- Literature Review – document procedure and outcome
 - Sorting – e.g., human (epi), laboratory animal, in vitro/mechanistic, in silico/read-across
- Hazard Identification
 - Toxicity profile
- Dose response evaluation
 - POD derivation only on toxic endpoints with sufficient information (lab animal, IVIVE, in silico).
- Risk Assessment
 - Derive HECs (PBPK models), implement Bayesian UFs, candidate OELs – compare with human data.
 - Select OELs.



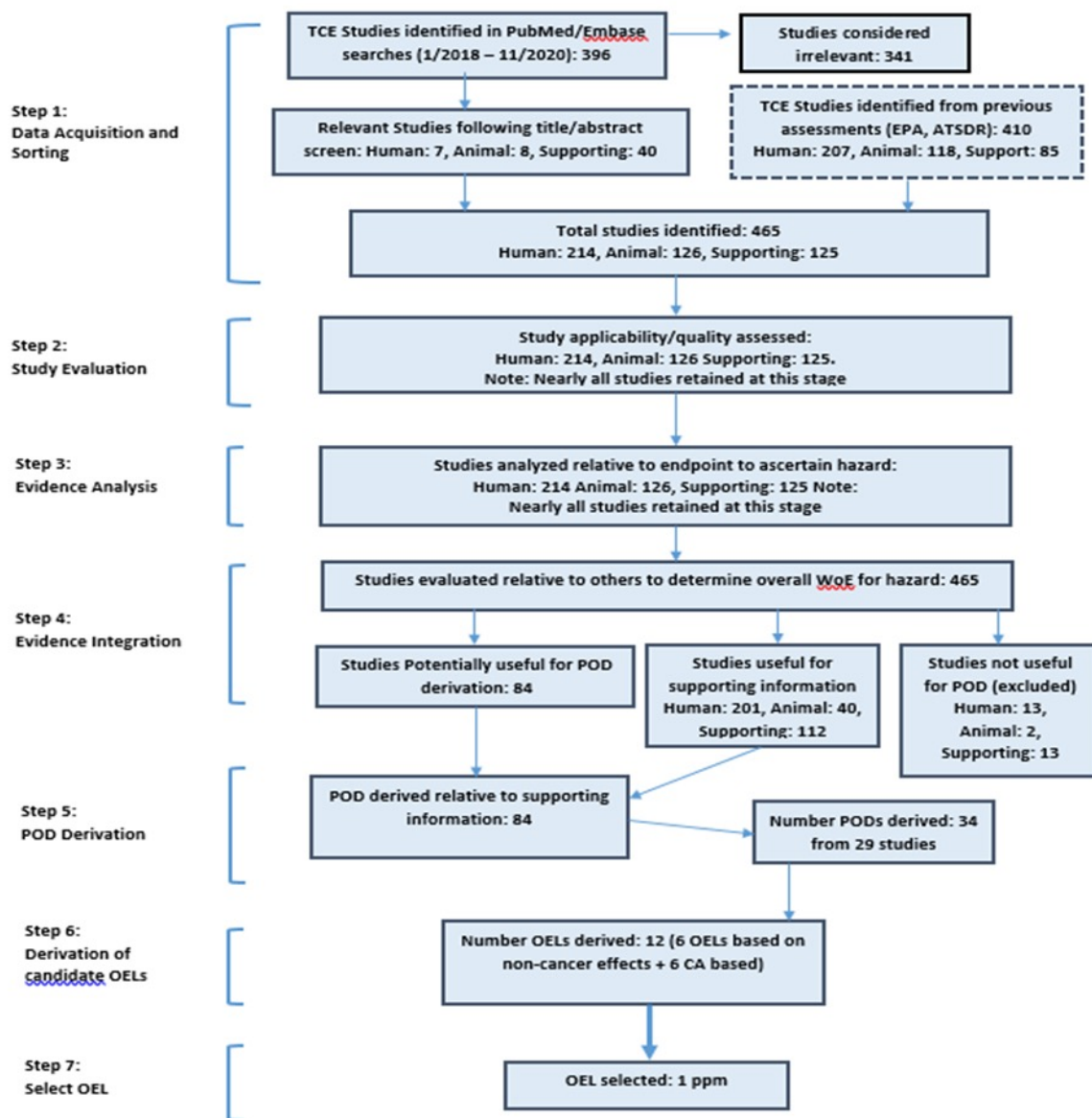


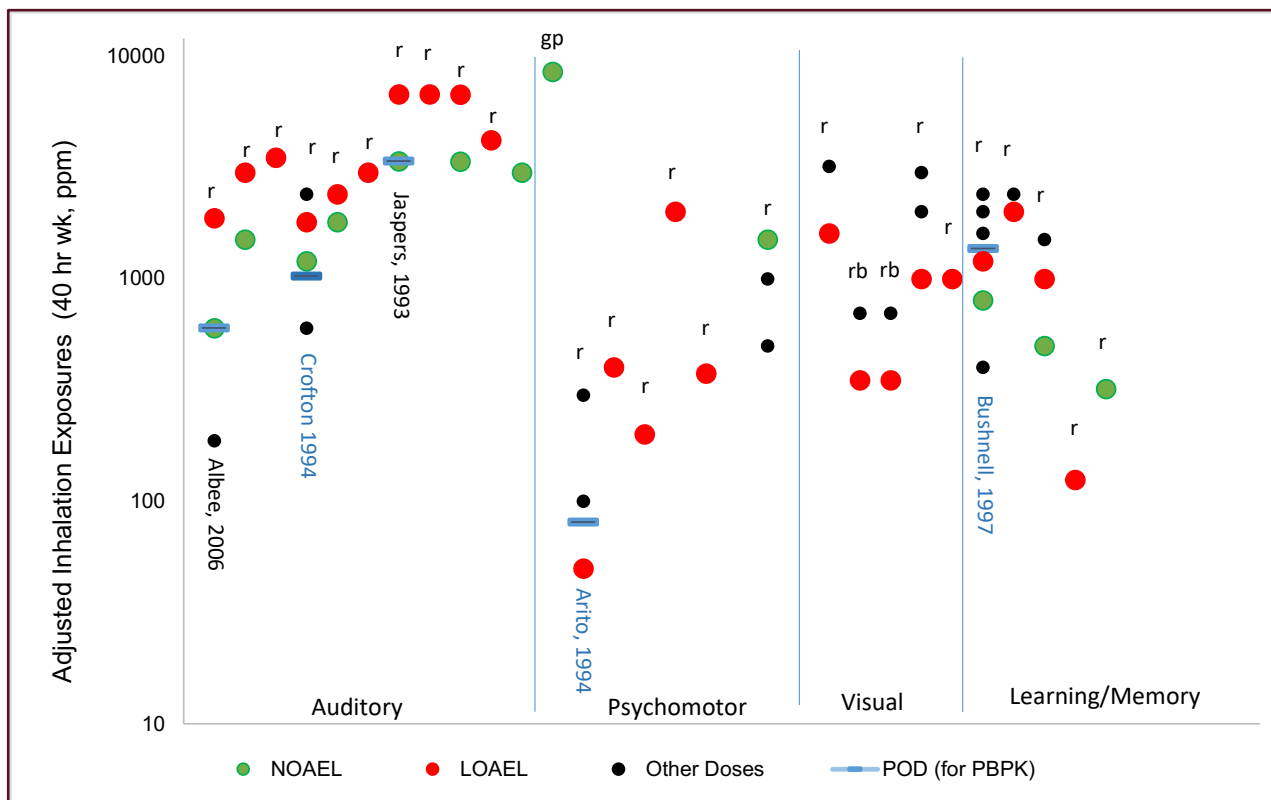
Table 2. Summary of evidence integration for target organ systems across categorical data streams

Target Organ System	Human	Animal	Mechanistic*	Comments
Neurological	Sufficient	Sufficient	Suggestive	Historical <u>anesthetic</u> use. Auditory, psychomotor, and visual effects.
Kidney	Sufficient	Sufficient	Suggestive	Biomarkers of kidney injury in animals and humans suggests tubular injury (high doses)
Hepatic	Sufficient (high exposures)	Sufficient	Suggestive	Weak human evidence
Reproductive	Sufficient (males)	Sufficient (males)	Suggestive (males)	No change in fertility in rodents
Developmental	Insufficient	Sufficient (growth) Insufficient (<u>terrata</u>)	Indeterminate (inhalation) Suggestive (oral)	Perinatal survival and growth at high exposures in animals.
Immune	Sufficient	Sufficient	Suggestive	Autoimmunity
Kidney cancer	Sufficient	Insufficient	Suggestive	High doses.
Non-Hodgkin's Lymphoma	Insufficient	Insufficient	Indeterminate	
Liver cancer	Insufficient	Sufficient	Indeterminate	
Repro. and other cancers	Insufficient	Sufficient	Indeterminate	Testicular/lung in specific animal strains

*Terms "suggestive" and "indeterminate" were used to apply to mode of action/mechanistic qualitative information.

Table 3. Summary of evidence integration for target organ systems

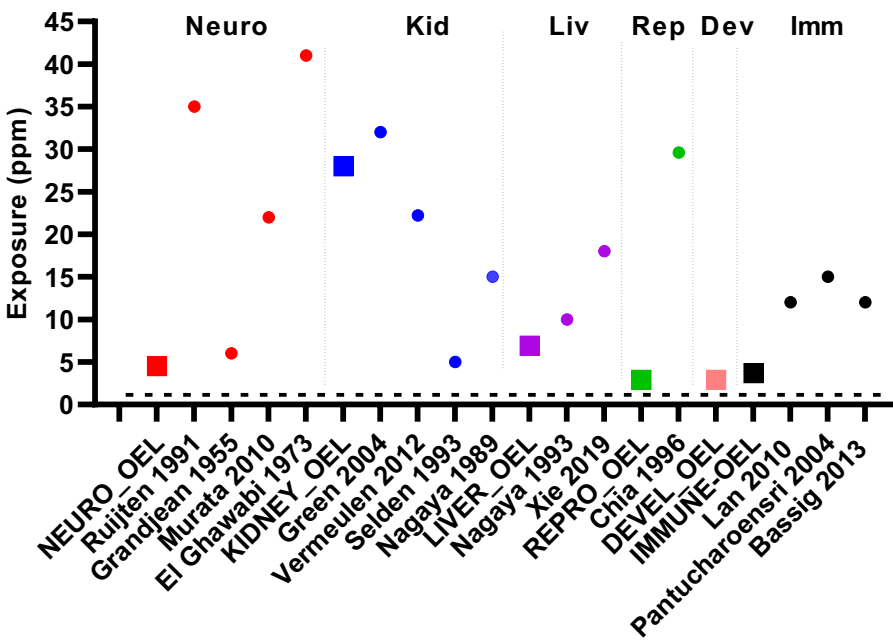
Target Organ System	Overall	Comments
Neurological	Sufficient	Evidence across all three data streams; historical use as an anesthetic.
Kidney	Sufficient	Evidence across all three data streams; GSH conjugation to toxic metabolites consistent with proximal tubule effects.
Hepatic	Sufficient	Evidence across all three data streams; weak human evidence.
Reproductive	Sufficient – (males)	Plausible testicular effects from all three data streams; none for females.
Developmental	Sufficient (weak)	Evidence from all three data streams for growth/development, Evidence for <u>terrata</u> considered insufficient (no evidence of <u>terrata</u> from inhalation exposures (human or animal); growth and reductions in offspring observed (oral).
Immune	Sufficient	Autoimmunity, immunosuppression
Kidney cancer	Sufficient	Site of cancer (proximal tubule) and potential carcinogen concordant.
Non-Hodgkin's Lymphoma	Insufficient	Consistency lacking in overall evidence.
Liver cancer	Insufficient	Relevance of animal evidence questionable.
Repro., lung, and other cancers	Insufficient	Variation in metabolism of metabolites in lung of rodents.



Scatterplot of Studies for Neurological Effects. Comparison of exposures for neurological effects in a range of rodent inhalation studies with those selected as PODs. Exposure concentrations were time-adjusted for equivalence to a 40 h/wk for auditory and psychomotor effects, with the exception of visual and learning/memory effects which demonstrated greater reliance on exposure concentration rather than duration; therefore, concentrations were not adjusted to 40 h/wk for these effects. NOAEL values are represented with green circles, LOAEL values are red circles, and other exposure concentrations are black dots.

Comparing Human Equivalent Concentrations with Human Data

OELs Compared to Human Occ. Exposures
(OEL as squares; human exposures as circles)



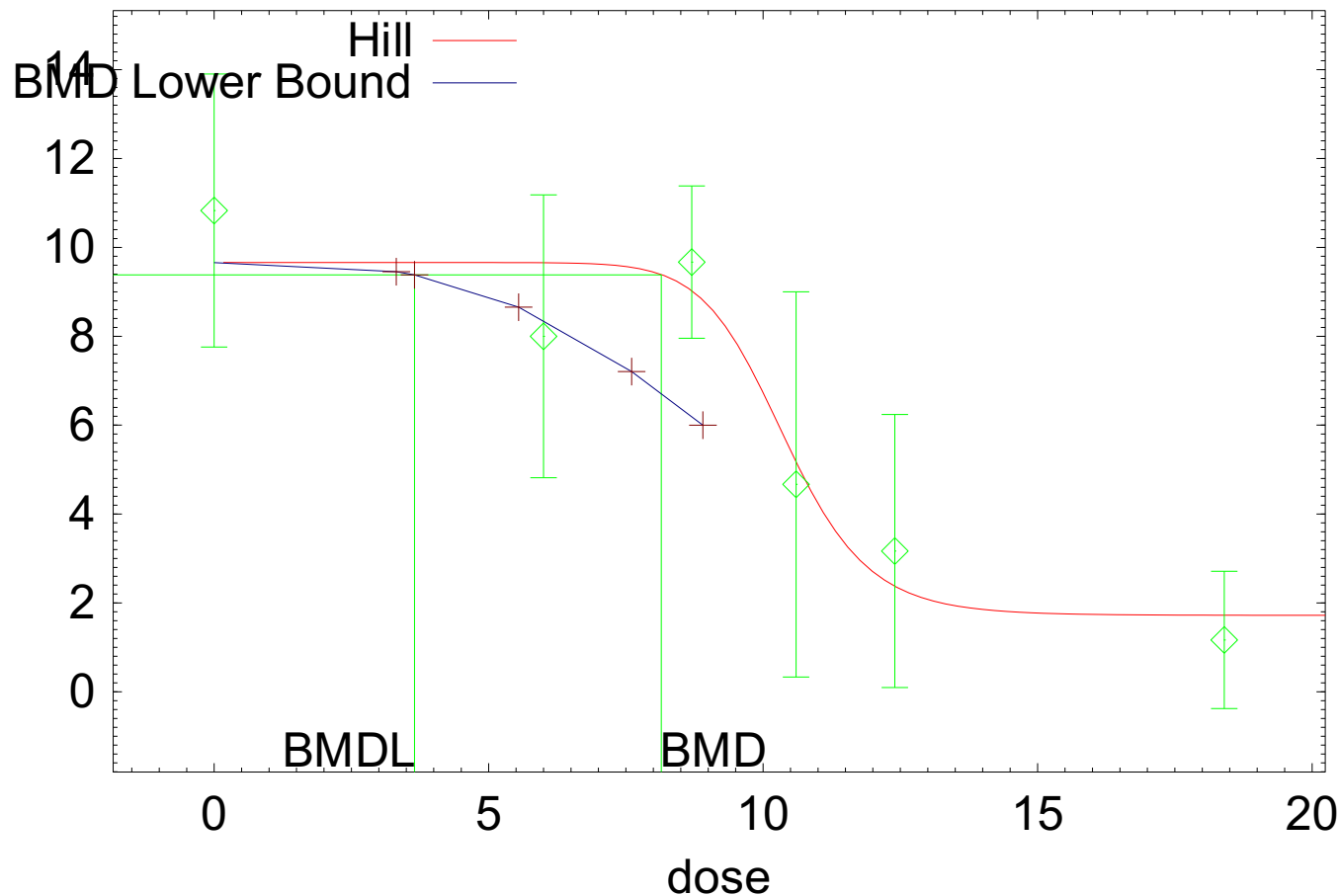
Comparison of Derived Occupational Exposure Levels (OELs) for six health effect domain and TCE studies cited that had available exposure levels. Where several air levels or a range was available for a study, the lowest one was used. OELs are the large squares, while exposure levels for human studies of the same domain are filled circles of the same color.

- Rigor of the literature review
- Process objective and transparent (hazard analysis across all three lines of evidence)*.
 - *...but not cookbook
- More data is more problematic than fewer data
 - False negatives vs. false positives
- Protective vs. predictive?
 - Screening vs. remedial action?
- Stay away from NOAELs/LOAELs
- How much data are needed?
- Individual vs. population vs. ecosystem services?
- Risk communication

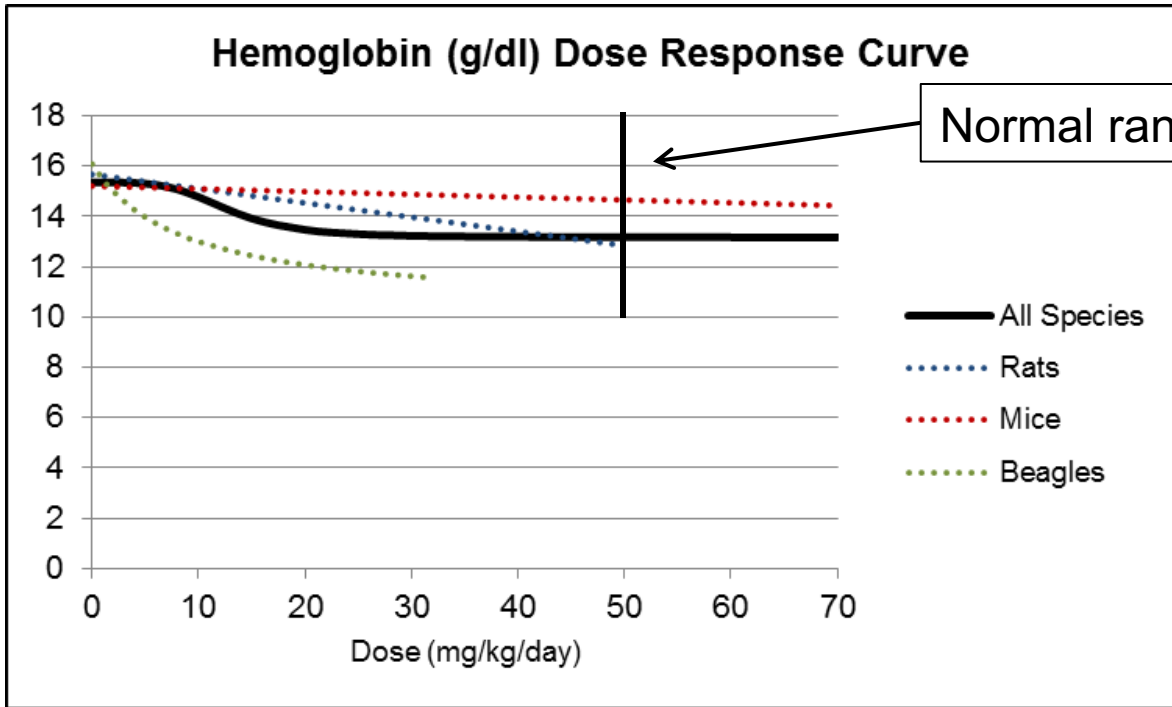
- The NOAEL/LOAEL is highly dependent on dose selection since the NOAEL/LOAEL is limited to being one of the doses included in a study.
- The NOAEL/LOAEL is highly dependent on sample size. The ability of a bioassay to distinguish a treatment response from a control response decreases as sample size decreases¹, so that the NOAEL for a compound (and thus the POD) will tend to be higher in studies with smaller numbers of animals per dose group.
- More generally, the NOAEL/LOAEL approach does not account for the uncertainty in the estimate of the dose-response which is due to the characteristics of the study design.
- NOAELs/LOAELs do not correspond to consistent response levels for comparisons across studies/chemicals/endpoints and for use as PODs for the derivation of RfCs.
- The slope of the dose-response curve is not taken into account in the selection of a NOAEL or LOAEL and is not usually considered unless the slope is very steep or very shallow.
- A LOAEL cannot be used to derive a NOAEL when a NOAEL does not exist in a study. Instead, a tenfold uncertainty factor has been routinely applied to the LOAEL to account for this limitation.
- While the NOAEL has typically been interpreted as a threshold (no-effect level), simulation studies and reanalyses of developmental toxicity bioassay data have demonstrated that the rate of response above control at doses fitting the criteria for NOAELs, for a range of study designs, is about 5-20% on average, not 0%.

- St. J. Warne M, van Dam R. 2008. NOEC and LOEC data should no longer be generated or used. *Australasian Journal of Ecotoxicology* 14:1-5.
- Laskowski R. 1995. Some good reasons to ban the use of NOEC, LOEC and related concepts in ecotoxicology. *Oikos* 73(1):140-144.
- Kodell RL. 2009. Replace the NOAEL and LOAEL with BMDL₀₁ and BMDL₁₀. *Environ. Ecol. Stat.* 16:3-12.
- Allen A, Gentry R, Shipp A, Van Landingham CV. 1998. Calculation of benchmark doses for reproductive and developmental toxicity observed after exposure to isopropanol. *Reg. Toxicol. Pharmacol.* 28(1):38-44.
- Yanagawa T, Kikuchi Y. 2001. Statistical issues on the determination of the no-observed-adverse-effect levels in toxicology. *Environmetrics* 12(4):319-325.
- Leisenring W, Ryan L. 1992. Statistical properties of the NOAEL. *Reg. Toxicol. Pharmacol.* 15(2):161-171.
- Bokkers BGH, Slob W. 2005. A comparison of ratio distributions based on the NOAEL and benchmark approach for subchronic-to-chronic extrapolation. *Toxicological Sciences* 85(2):1033-1040.
- Chapman PM, Caldwell RS, Chapman PF. 2009. A warning: NOECs are inappropriate for regulatory use. *Environ. Toxicol. Chem.* 15(2):77-79.
- Kooijman SALM. 1996. An alternative for NOEC exists, but the standard model has to be abandoned first. *Oikos* 75(2):310-316.
- Fox DR. 2009. Is the ec_x a legitimate surrogate for a NOEC? *Integrated Environ/ Assess. Mngt.* 5(2):351-353.

Hill Model with 0.95 Confidence Level



14:06 08/09 2001



Species	g/dl
Rat	12-17
Mouse	10-19
Dog	11-18

Species	BMD (g/dl)	BMDL (g/dl)
Rats	40.6	29.2
Mice	196.2	87.0
Beagles	2.2	1.2
Average	79.6	39.1
All Species (combined)	50.4	11.9

- Uses data distributions to refine expected variation
- Less data demanding than DDUFs
- Reflected as the log-normal distributions of the geometric mean (μ) and geometric standard deviation (σ) of the composite UF.
- Recommended by NASEM (Simon et al. 2016)
- Subchronic to chronic:

$$\ln(OEL) = \ln(HEC) - \sum \mu_{UF} - Z_{\alpha} \sqrt{\sigma_{UFS}^2 + \sigma_{UFA}^2 + \sigma_{UFH}^2 + \sigma_{UFL}^2}$$

Simon, T. W., Zhu, Y., Dourson, M. L. and Beck, N. B. (2016). Bayesian methods for uncertainty factor application for derivation of reference values. *Regul Toxicol Pharmacol*, 80, 9-24. doi:10.1016/j.yrtph.2016.05.018

Conversion of HECs to candidate occupational exposure levels via application of Bayesian uncertainty factors.

Study	Human Equivalent Concentration	ln(HEC)	Uncertainty Factor (UF)												Composite UF (Standard)	Composite UF (Bayesian)	Notes	ln(OEL)	Occupational Exposure Level
			UF _S	α _{UFS}	UF _A	α _{UFA}	UF _H	α _{UFH}	UF _L	μ _{UFL}	α _{UFL}	UF _D	α _{UFD}						
Neurological																			
Albee 2006	1493 ppm	7.309	10	1.4	3	0.668	3	0.668	1	0	0	1	0	100	16.1	UF _S =1; effect is not time dependent and is reversible with 24h	4.53	93 ppm	
Arito 1994	21.2 ppm	3.052	1	0	3	0.668	3	0.668	1	0	0	1	0	10	4.7		1.50	4.5 ppm	
Bushnell 1997	1109 ppm	7.011	10	1.4	3	0.668	3	0.668	1	0	0	1	0	100	16.1		4.23	69 ppm	
Crofton 1997	2007 ppm	7.604	10	1.4	3	0.668	3	0.668	1	0	0	1	0	100	16.1		4.83	125 ppm	
Jaspers 1993	3554 ppm	8.176	10	1.4	3	0.668	3	0.668	1	0	0	1	0	100	16.1		5.40	221 ppm	
Kidney																			
Adams 1951	187 ppm	5.231	1	0	3	0.668	3	0.668	1	0	0	1	0	10	4.7	UF _S =1; see Kjellstrand 1983b	3.68	40 ppm	
Maltoni 1988	168 ppm	5.124	1	0	3	0.668	3	0.668	1	0	0	1	0	10	4.7		3.57	36 ppm	
Woolhiser 2006	132 ppm	4.883	1	0	3	0.668	3	0.668	1	0	0	1	0	10	4.7		3.33	28 ppm	
Liver																			
Adams 1951	293 ppm	5.680	1	0	3	0.668	3	0.668	1	0	0	1	0	10	4.7	UF _S =1; Rel Liver/Kidney weights at 120d not higher than at 30d.	4.13	62 ppm	
Kjellstrand 1983b	184 ppm	5.215	1	0	3	0.668	3	0.668	10	1.504	0.531	1	0	100	26.8		1.93	6.9 ppm	
Kumar 2001a	303 ppm	5.714	1	0	3	0.668	3	0.668	10	1.504	0.531	1	0	100	26.8	UF _S =1; see Kjellstrand 1983b	2.43	11 ppm	
Ramadhan 2010	1310 ppm	7.171	10	1.4	3	0.668	3	0.668	1	0	0	1	0	100	16.2		4.39	81 ppm	
Woolhiser 2006	197 ppm	5.281	1	0	3	0.668	3	0.668	1	0	0	1	0	10	4.7		3.73	42 ppm	
Immunological																			
Blossom 2017	3.73 ppm	1.316	1	0	3	0.668	2	0.421	1	0	0	1	0	6	3.7	UF _H =2; autoimmune prone strain	0.02	1.0 ppm	
Gilbert 2014	4.3 ppm	1.459	1	0	3	0.668	2	0.421	1	0	0	1	0	6	3.7	UF _H =2; autoimmune prone strain	0.16	1.2 ppm	
Griffin 2000	23.7 ppm	3.165	1	0	3	0.668	2	0.421	1	0	0	1	0	6	3.7	UF _H =2; autoimmune prone strain	1.87	6.5 ppm	
Kaneko 2000	571 ppm	6.347	10	1.4	3	0.668	2	0.421	10	1.504	0.531	10	1.4	6000	166.9	UF _H =2; autoimmune prone strain, UFD=10; only 2 inhalation studies	1.23	3.4 ppm	
Sanders 1982	23.0 ppm	3.135	1	0	3	0.668	3	0.668	10	1.504	0.531	1	0	100	26.8	UFD=10; only 2 inhalation studies	-0.15	0.9 ppm	
Woolhiser 2006	136 ppm	4.910	10	1.4	3	0.668	3	0.668	1	0	0	10	1.4	1000	37.0		1.30	3.7 ppm	
Reproductive																			
George 1985	2696 ppm	7.900	1	0	3	0.668	3	0.668	10	1.504	0.531	1	0	100	26.8	UF _S =1; sperm assessed at the end of the cohabitation period	4.61	101 ppm	
Kumar 2000a	192 ppm	5.257	1	0	3	0.668	3	0.668	10	1.504	0.531	1	0	100	26.8	UF _S =1; exposed throughout sensitive period of oocyte maturation	1.97	7.2 ppm	
Land 1981	324 ppm	5.779	10	1.4	3	0.668	3	0.668	1	0	0	1	0	100	16.1		3.00	20 ppm	
Wu 2007	136 ppm	4.913	1	0	3	0.668	3	0.668	10	1.504	0.531	1	0	100	26.8		1.63	5.1 ppm	
Zenick 1984	46.9 ppm	3.848	10	1.4	3	0.668	3	0.668	1	0	0	1	0	100	16.1		1.07	2.9 ppm	
Developmental																			
Blossom 2017	3.73 ppm	1.316	1	0	3	0.668	3	0.668	10	1.504	0.531	1	0	100	26.8		-1.97	0.14 ppm	
Carney 2006	511 ppm	6.236	1	0	3	0.668	3	0.668	1	0	0	1	0	10	4.7		4.68	108 ppm	
Cosby 1992	282 ppm	5.642	10	1.4	3	0.668	3	0.668	1	0	0	1	0	100	16.1		2.86	18 ppm	
Fisher 2001	853 ppm	6.749	1	0	3	0.668	3	0.668	1	0	0	1	0	10	4.7		5.19	180 ppm	
Healy 1982	77.7 ppm	4.353	1	0	3	0.668	3	0.668	10	1.504	0.531	1	0	100	26.8		1.07	2.9 ppm	
Isaacson 1989	32.7 ppm	3.487	1	0	3	0.668	3	0.668	10	1.504	0.531	1	0	100	26.8		0.20	1.22 ppm	
Manson 1984	224 ppm	5.412	1	0	3	0.668	3	0.668	1	0	0	1	0	10	4.7		3.86	47 ppm	
Narotsky 1995	34.9 ppm	3.552	1	0	3	0.668	3	0.668	1	0	0	1	0	10	4.7		2.00	7.4 ppm	
Schwetz 1975 (m)	1070 ppm	6.975	1	0	3	0.668	3	0.668	1	0	0	1	0	10	4.7		5.42	226 ppm	
Schwetz 1975 (r)	445 ppm	6.098	1	0	3	0.668	3	0.668	1	0	0	1	0	10	4.7		4.54	94 ppm	

Note: $\ln(OEL) = \ln(HEC) - \sum \mu_{UF} - Z_{\alpha} \sqrt{\sigma_{UFS}^2 + \sigma_{UFA}^2 + \sigma_{UFH}^2 + \sigma_{UFL}^2}$; $OEL = 2.71828^{\ln(OEL)}$

- Include all of the evidence.
- Hazard assessment before POD derivation.
- Consider PODs only for toxicity endpoints with sufficient evidence.
- Support draft TRVs with field, IVIVE, other evidence.
- HQs are screens
 - Consider relevance to assessment endpoints, PF, and consequences of remedial action/mitigation.

Questions/Discussion