Data Fusion-based Human Health Risk Assessment Framework: Illustrative Examples

Workshop III: Alliance for Risk Assessment
May 5, 2011

Asish Mohapatra
Health Canada (Alberta/Northern Region)

Rehan Sadiq, Amin Zargar, M. Shafiqul Islam, Roberta Dyck
University of British Columbia

Disclaimer: The views presented in this presentation does not represent views of agencies of authors which they belong. This is a work in progress and draft information has been shared solely for this workshop discussion and is subject to further analysis, validation and correction.

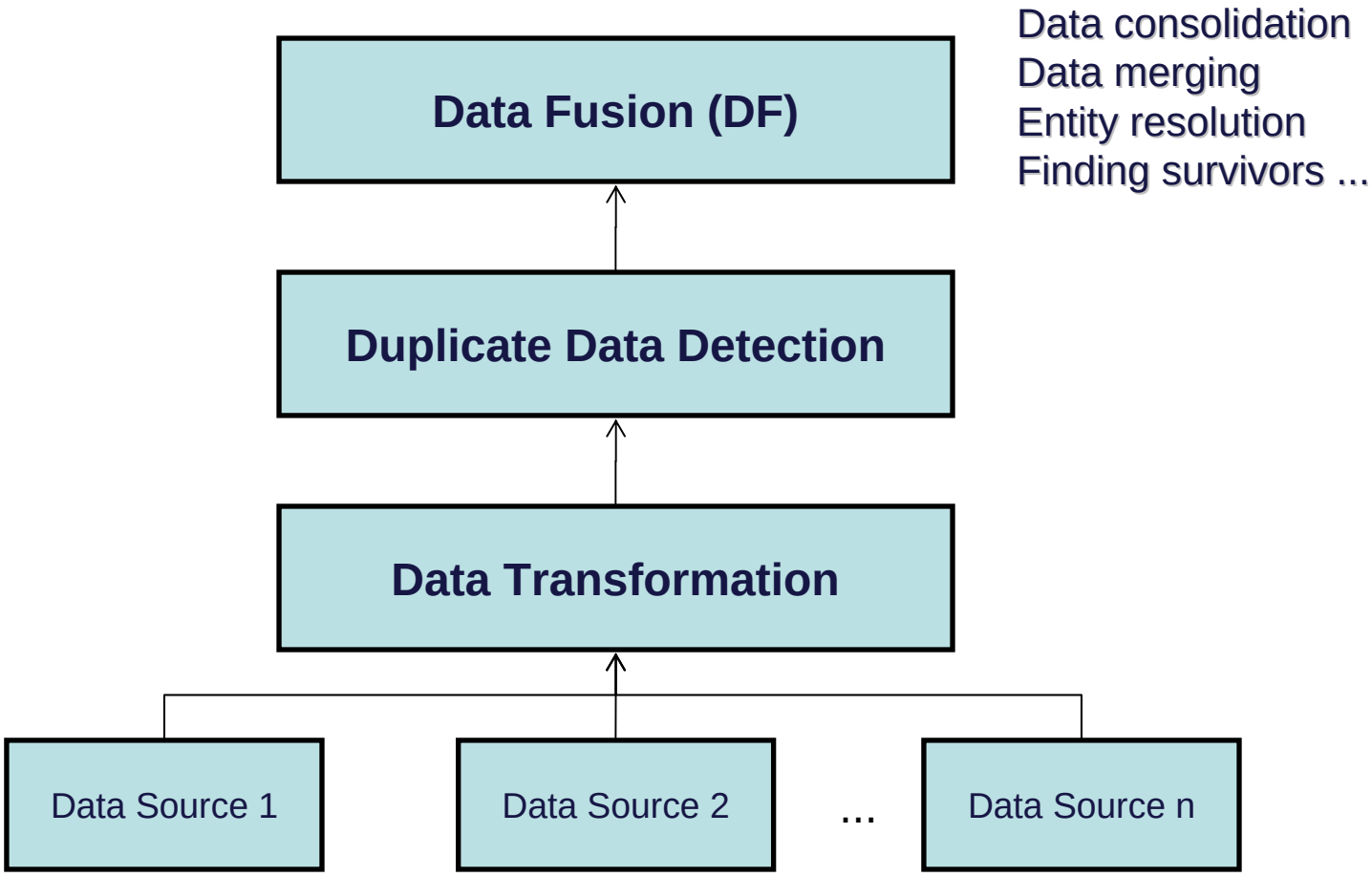
Presentation Outline

- Data Fusion Human Health Risk Assessment Framework (DF – HHRA)

- Benzene
- F1

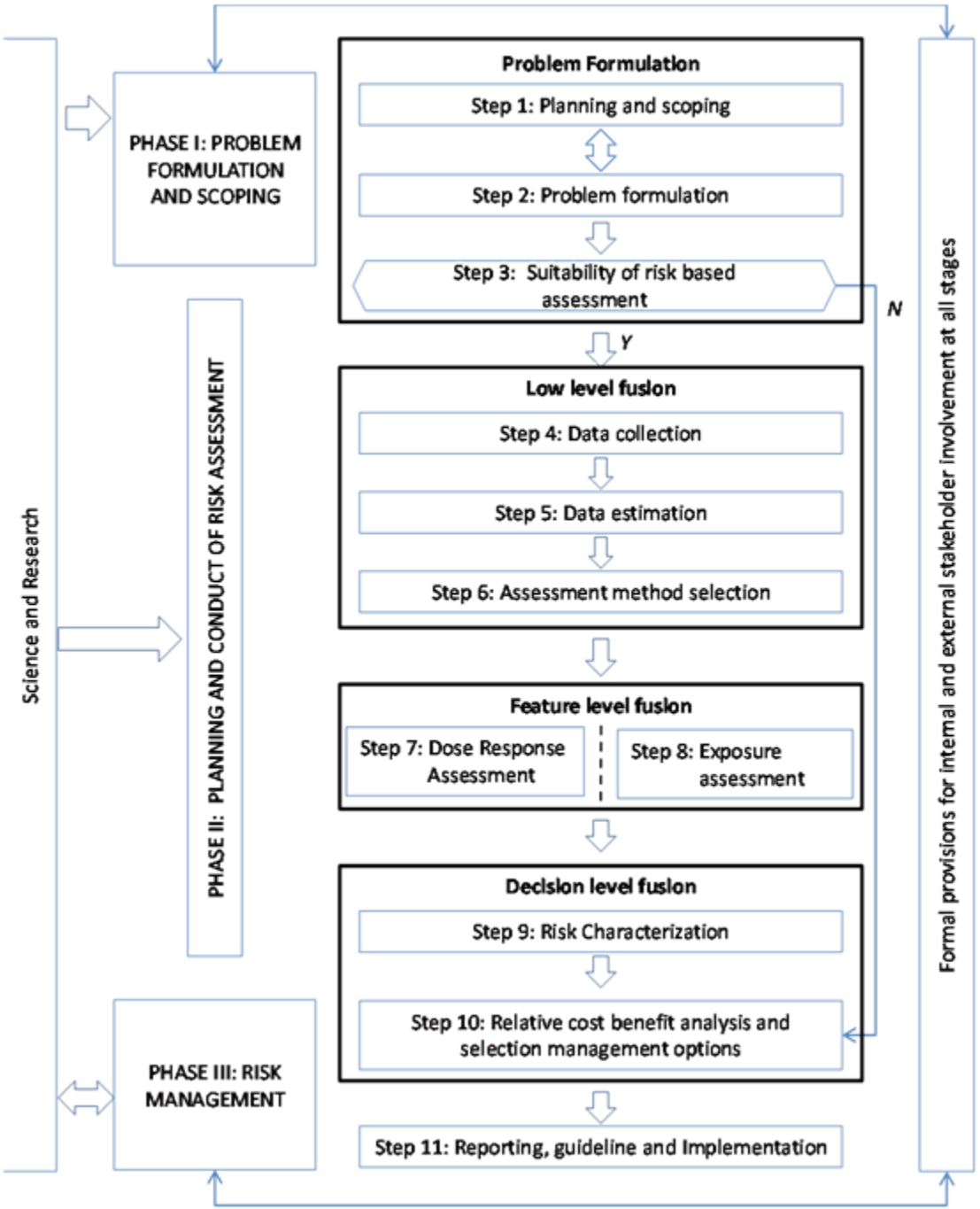
Data Integration

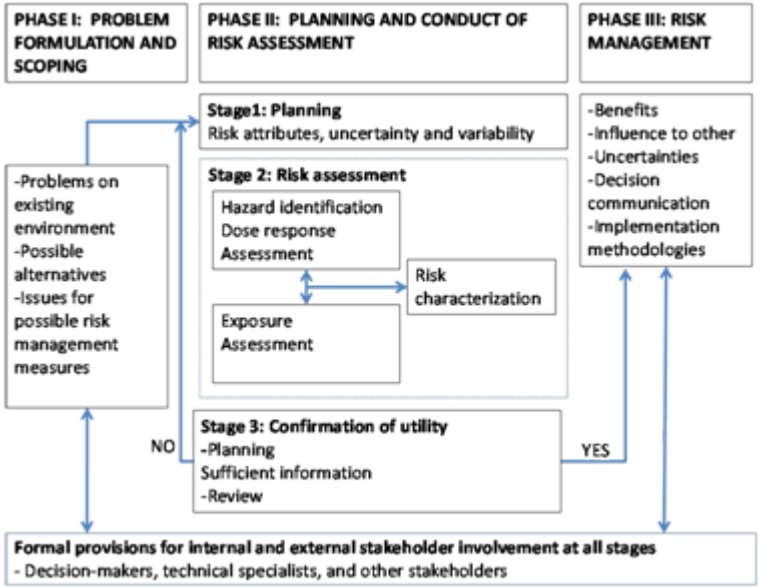
Data refinement and improving data quality
Additional inferences and increasing benefit from data
Improving understanding and decision

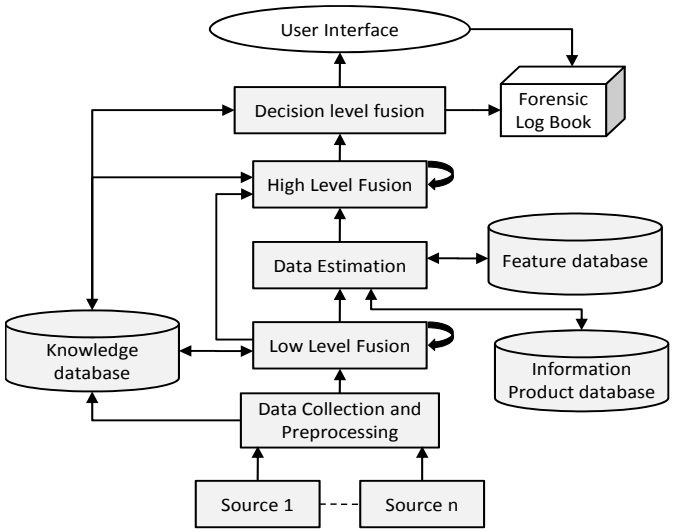


Data consolidation
Data merging
Entity resolution
Finding survivors ...

Proposed DF Framework







DF Techniques

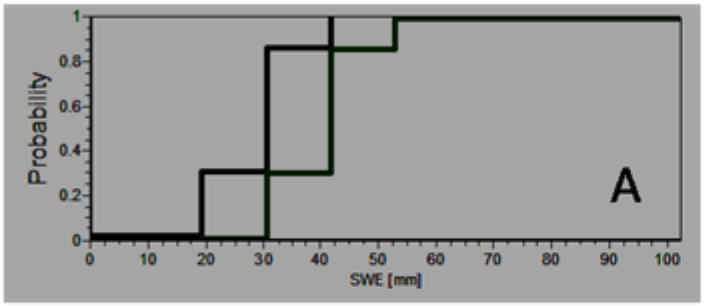
Fusion technique	Identity fusion	Feature-level fusion	Decision-level fusion
Cluster Analysis	X	X	
Classical Inference	X		X
Bayesian Inference	X	X	X
Dempster-Shafer Theory	X	X	X
Voting Strategies			X
Expert Systems	X	X	X
Logical Templates		X	X
(Adaptive) Neural Networks	X	X	X
Fuzzy Logic	X		X
Blackboard			X
Contextual Fusion			X
Syntactic Fusion			X
Estimation theory	X		
Entropy	X		
Figure of Merits	X		
Templates	X		
Generalized evidence processing theory			X

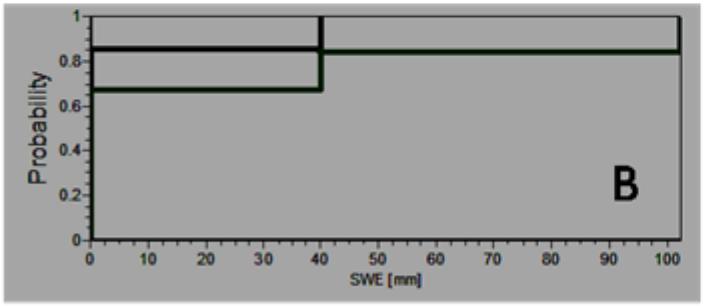
DF in the Context of HHRA

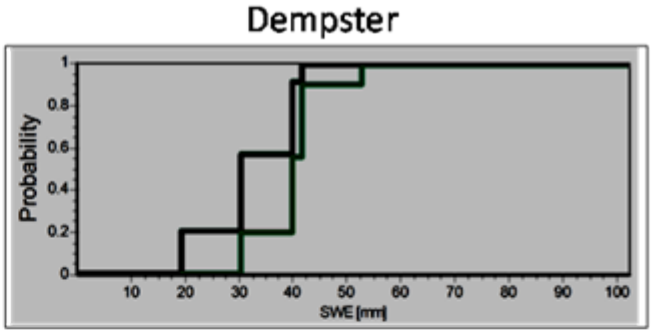
Data fusion technique	Application area(s)	Methods	HHRA area	Sources
Statistical and kernel inference	Genomic data fusion	Kernel-based statistical-learning; different data types/formats are transformed into kernels; to combine kernels, it uses semi-definite programming to minimize the statistical loss function	TA	(Lanckriet, *et al.* 2004a)
	Transcription factor target gene prediction	Statistical inference coupled with additional sources	TA	(Xiaofeng *et al.* 2010)
	Biomedical data fusion	Optimization of the L_2-norm of multiple kernels	TA	(Yu *et al.* 2010)
Bayesian inference (BI)	Multi-study and multi-endpoint BMD	Combines mechanistically informed model results with empirical data to derive several endpoints; combines multi-endpoint BMDs to derive BMDL	TA	(Schmitt 2006)
	Wide-area assessment of UXO contamination	Generates PDFs of features extracted from survey maps, uses BI methods to combine features with auxiliary information and data quality features	EA	(Johnson *et al.* 2009)
	Syndrome surveillance	Uses Bayesian conditional autoregressive (CAR) models to combine symptom data collected from a network for early outbreaks detection	TA	(Banks *et al.* 2009)
Dempster-Shafer theory (DST)	Risk assessment of water treatment	Transferable belief models (TBM) input diverse data such as fuzzy, interval probabilities and statistical data to produce a belief network		(Demotier *et al.* 2006)
	Drinking water quality (WQ)	Uses disjunctive operator for the interpretation of overall WQ in the distribution system and the development of a WQ index	EA	(Sadiq and Rodriguez 2005)
	Microbial water quality in distribution network	Four DST fusion rules are applied to fuse weak information from two microbial water quality data sources, results in four p-boxes	EA	(Sadiq *et al.* 2006)
	Prediction of breast cancer tumours	Fuses the outputs of multiple classifiers from different diagnostic sources	TA	(Raza *et al.* 2006)
Artificial neural networks (ANN)	Surface WQ estimation	Combines optical data and microwave data to estimate surface WQ	EA	(Zhang *et al.* 2002)
Fuzzy sets theory	Analysis of gene expression data	Transforms gene expression values into qualitative descriptors that are then evaluated using a set of heuristic rules	TA	(Woolf and Wang 2000)

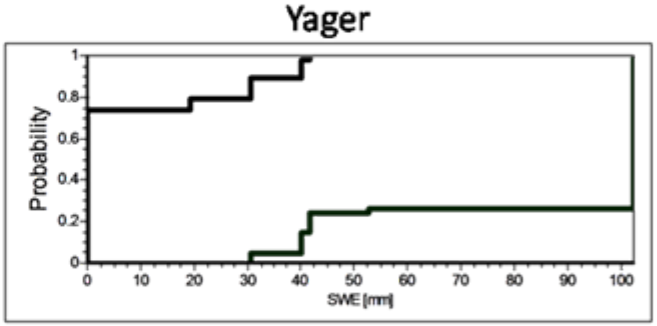
TA: toxicity assessment and EA: exposure assessment

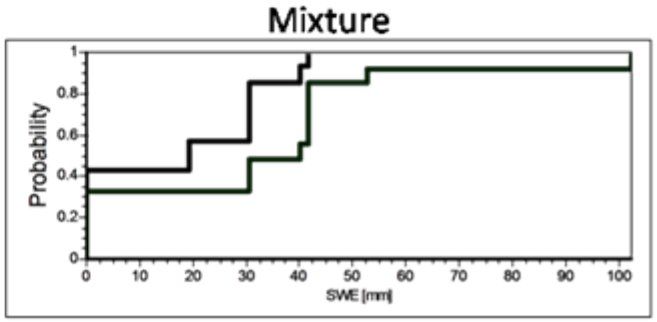
DF Techniques



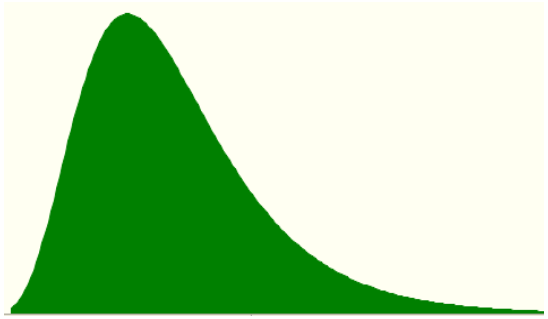


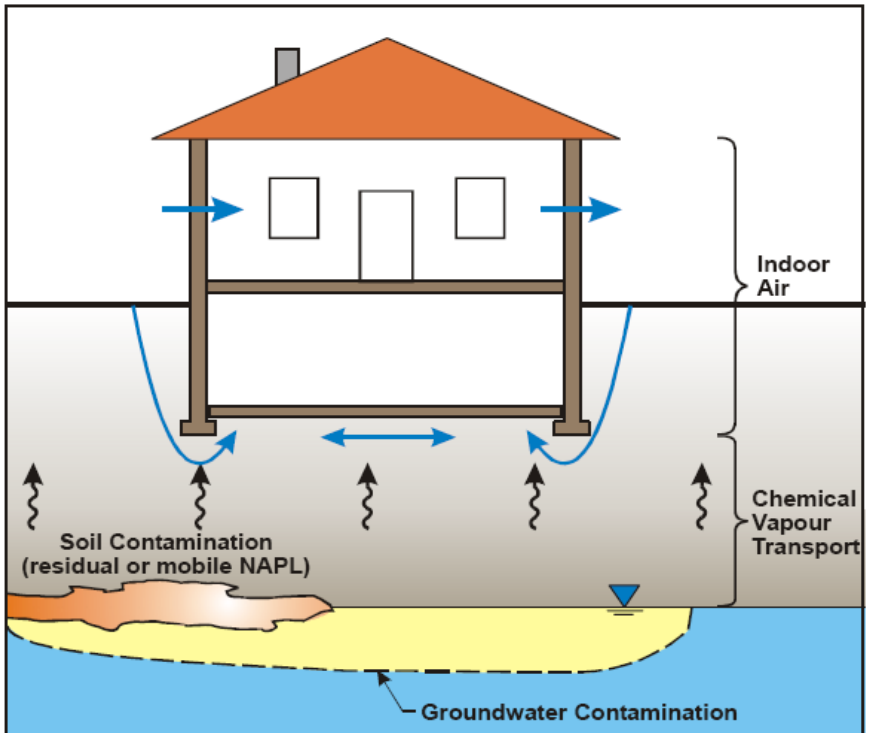






Exposure Scenarios





Benzene Example

Standardized Mortality Ratios (SMRs) for leukemia among Pliofilm workers based on the estimated cumulative exposures

Table 6. Standardized mortality ratios for leukemia in Pliofilm workers[a] by cumulative exposure at all locations.

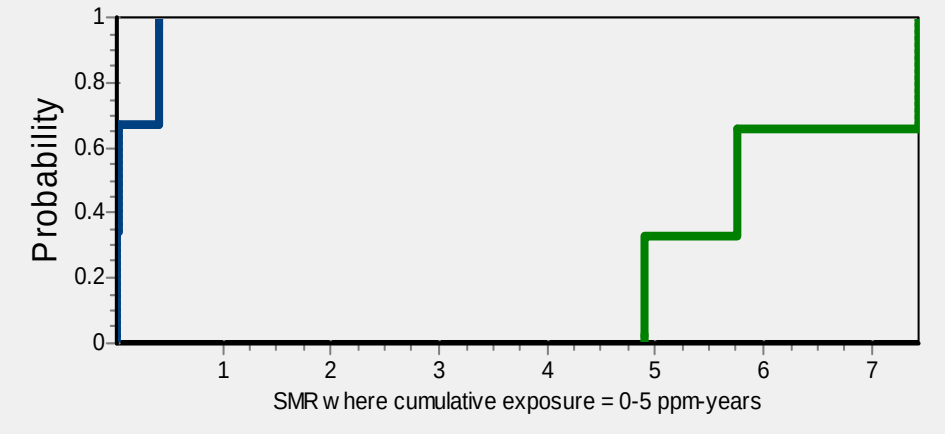
Exposure estimates	Cumulative exposure, ppm-years	Person-years	Observed	Expected	SMR[b]	95% CI
Rinsky	0–5	18,178	3	1.52	1.97	0.41–5.76
	>5–50	13,456	3	1.31	2.29	0.47–6.69
	>50–500	8,383	7	1.01	6.93**	2.78–14.28
	>500	328	1	0.05	20.00	0.51–111.4
Crump	0–5	12,974	1	1.14	0.88	0.02–4.89
	>5–50	13,951	4	1.23	3.25	0.88–8.33
	>50–500	11,448	6	1.23	4.87*	1.79–10.63
	>500	1,972	3	0.29	10.34**	2.13–30.21
Paustenbach	0–5	9,645	1	0.75	1.33	0.03–7.43
	>5–50	12,882	2	1.12	1.79	0.22–6.45
	>50–500	14,095	4	1.43	2.80	0.76–7.16
	>500	3,723	7	0.59	11.86**	4.76–24.44

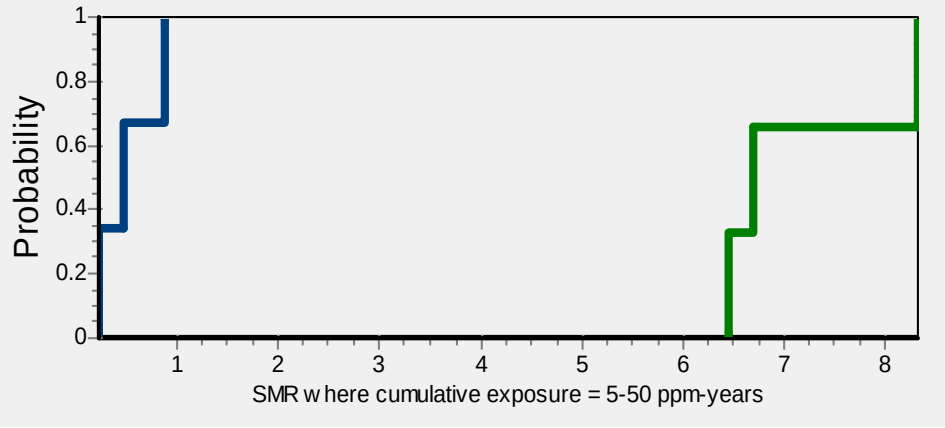
[a]White male wetside workers. [b]*p*-Value by two-sided Poisson test: * $p<0.05$; ** $p<0.01$.

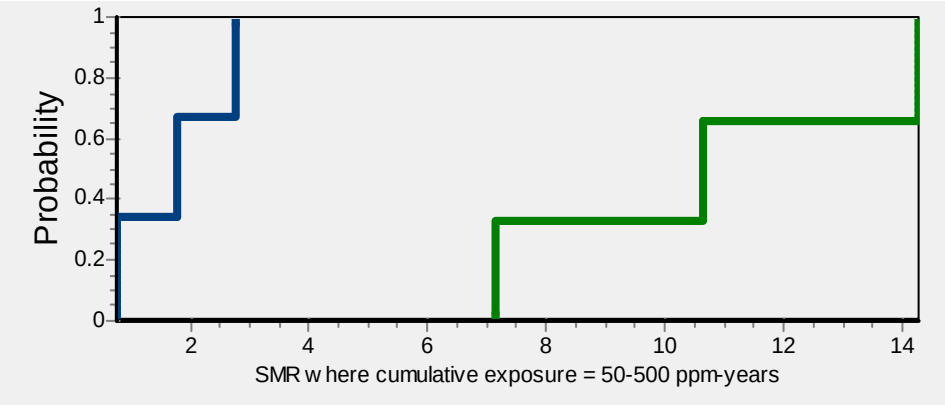
Leukemia Risk Associated with Benzene Exposure in the Pliofilm Cohort

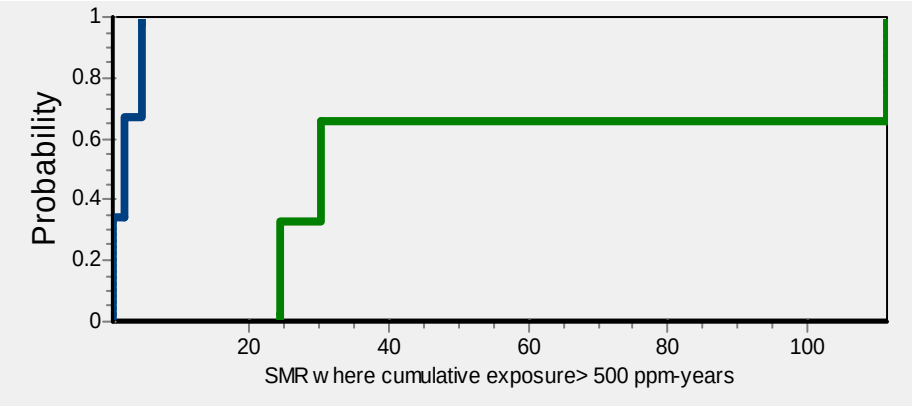
Mary Burr Paxton

Benzene Example





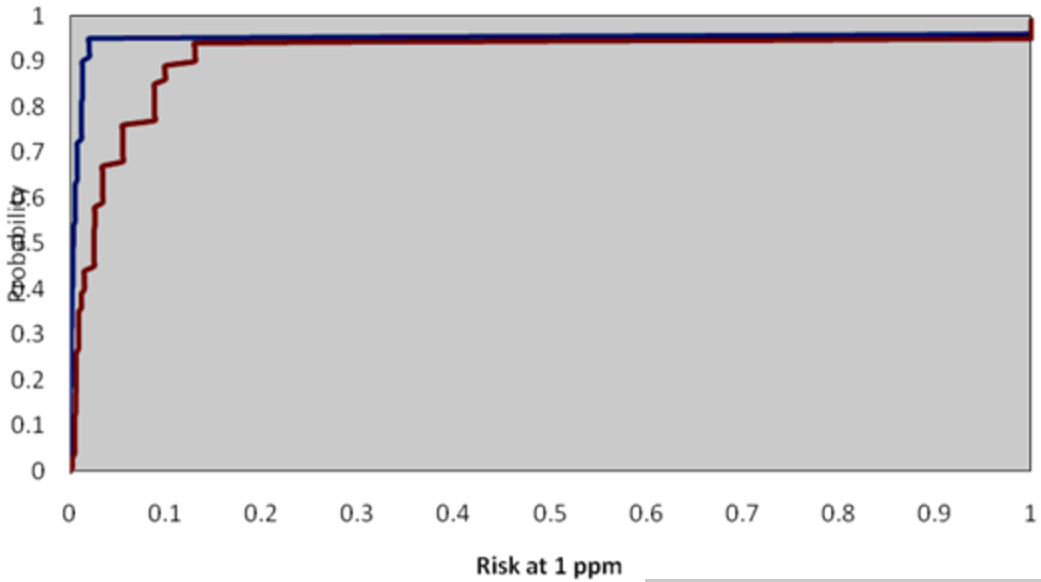


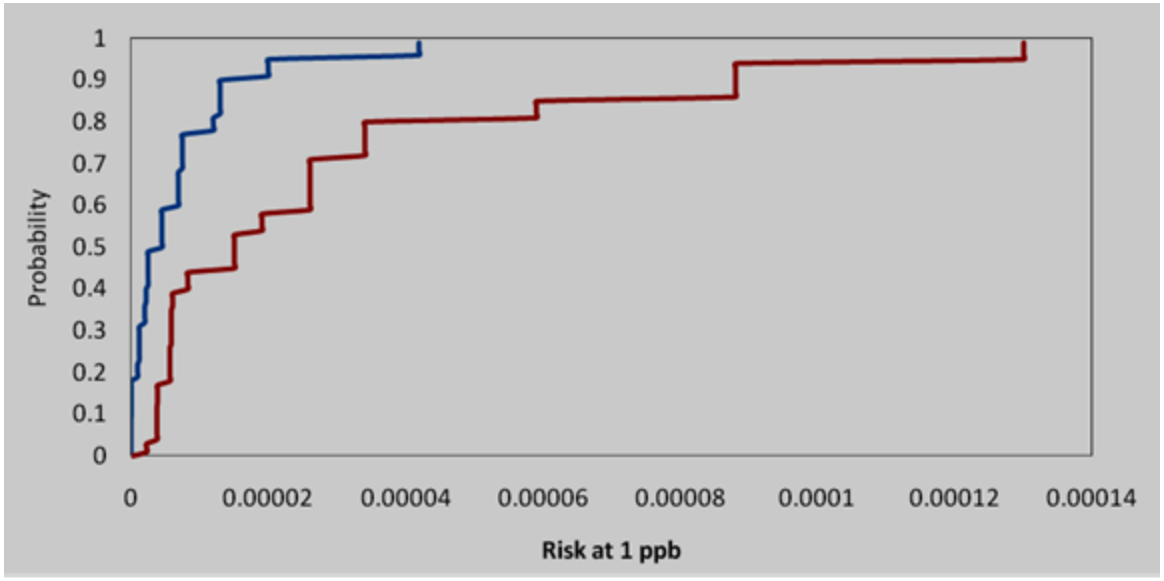


Benzene Example

Source	Risk at 1 ppm	Risk at 1 ppb	Reference & model
US EPA (1985)	0.018 (7.5E-3, 3.4E-2)	0.000018 (7.5E-6, 3.4E-5)	Crump and Allen, additive risk
	0.041 (1.3E-2, 8.8E-2)	0.000041 (1.3E-5, 8.8E-5)	Crump and Allen, relative risk
Brett *et al.* (1989)	4.0E-3 (1.0E-3, 1.2E-2) to 2.5E-2 (2.5E-3, 9.9E-2)	3.6E-6 (9.5E-7, 6.9E-6) to 1.1E-5 (2.2E-6, 1.9E-5)	Crump and Allen, conditional logistic
	2.2E-1 (1.2E-2, 1.0) to 8.4E-1 (1.5E-2, 1.0)	2.4E-5 (6.9E-6, 4.2E-5) to 3.4E-5 (8.2E-6, 5.9E-5)	Rinsky, conditional logistic
Paxton (1992)	0.0022 (3.8E-5, 4.9E-3)	0.0000019 (3.7E-8, 3.7E-6)	Crump and Allen, proportional hazard
	0.0046 (1.3E-3, 9.0E-3)	0.0000035 (1.2E-6, 5.8E-6)	Paustenbach, proportional hazard
	0.018 (3.0E-3, 5.5E-2)	0.0000089 (2.5E-6, 1.5E-5)	Rinsky, proportional hazard
Crump (1992; 1994)	1.1E-2 (2.2E-3, 2.0E-2) to 2.5E-2 (6.0E-3, 1.3E-1)	1.1E-5 (2.2E-6, 2.0E-5) to 2.5E-5 (6.0E-6, 1.3E-4)	Crump and Allen, linear
	5.4E-3 to 2.5E-2	4.5E-6 to 2.6E-5	Crump and Allen, nonlinear
	7.1E-3 (2.0E-3, 1.2E-2) to 1.5E-2 (3.8E-3, 2.6E-2)	7.2E-6 (2.0E-6, 1.2E-5) to 1.6E-5 (3.8E-6, 2.6E-5)	Paustenbach, linear
	8.6E-5 to 6.5E-3	8.6E-11 to 5.6E-6	Paustenbach, nonlinear

Benzene Example

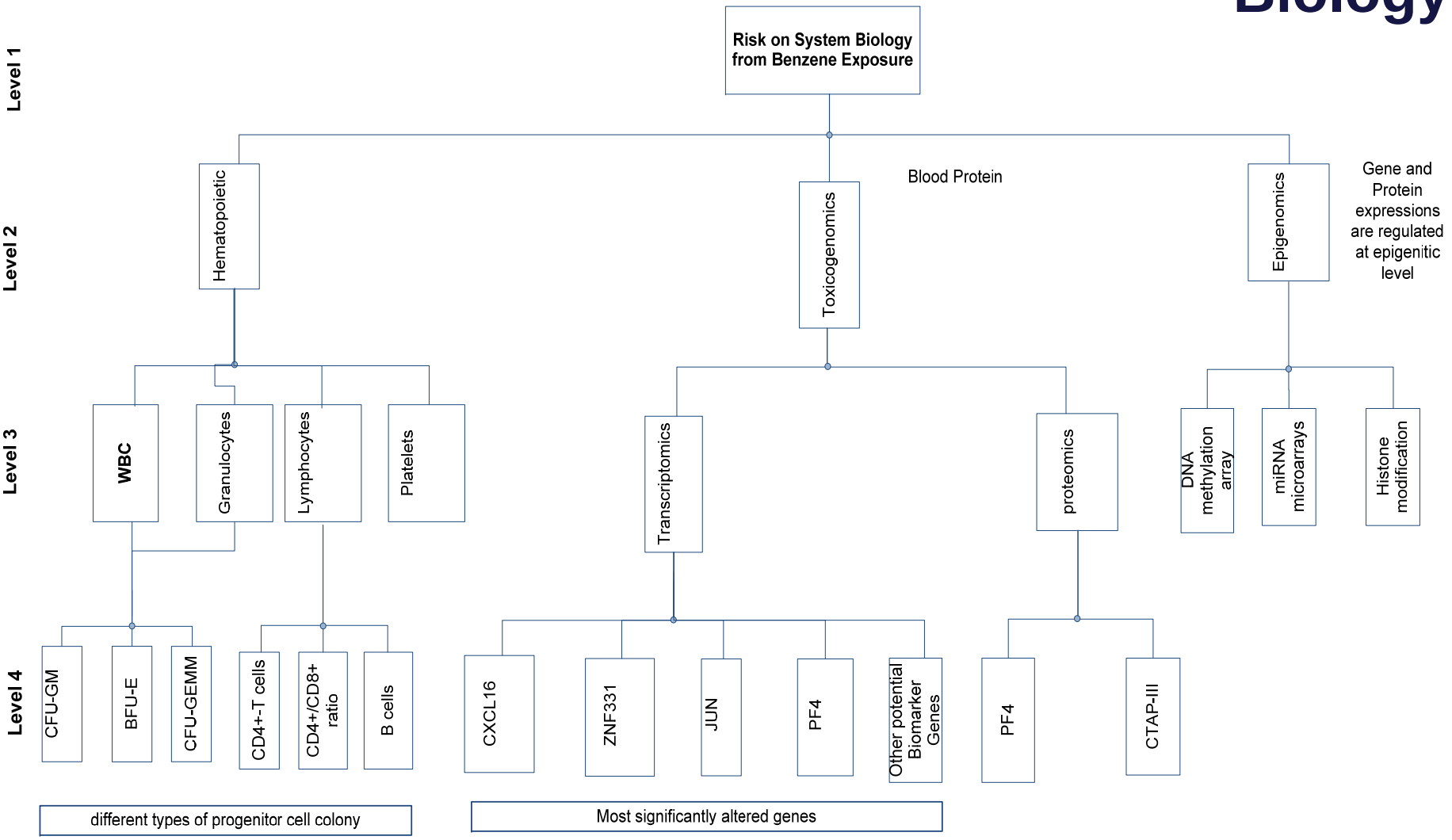




Benzene: System Biology

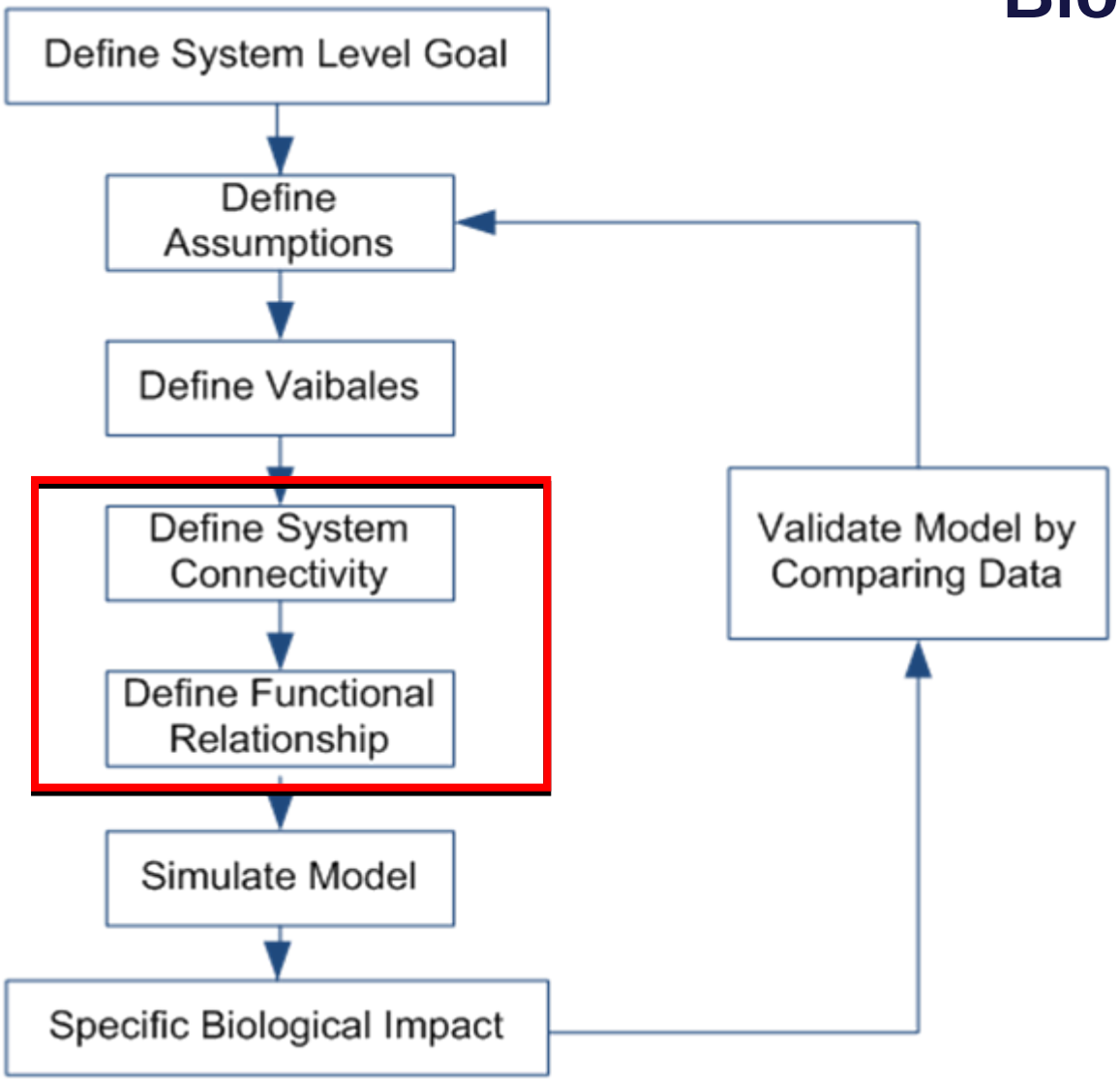
- Comparative Toxicogenomics Database
- 400 interacting genes - at least a dozen highly interacting genes
- Six most altered genes (based on Benzene (gene-cell-tissue-disease) Problem Formulation (with a disease focus – Leukaemia)
- Literature Extraction Process – 115 peer reviewed publications
- Overall objective: Probability of failure of biological systems identified in the Benzene System Biology flowchart (Overall impacts to Hematopoietic components).

Benzene: System Biology



Benzene: System Biology

Mathematical Modelling of a Biological System



Benzene: System Biology challenges

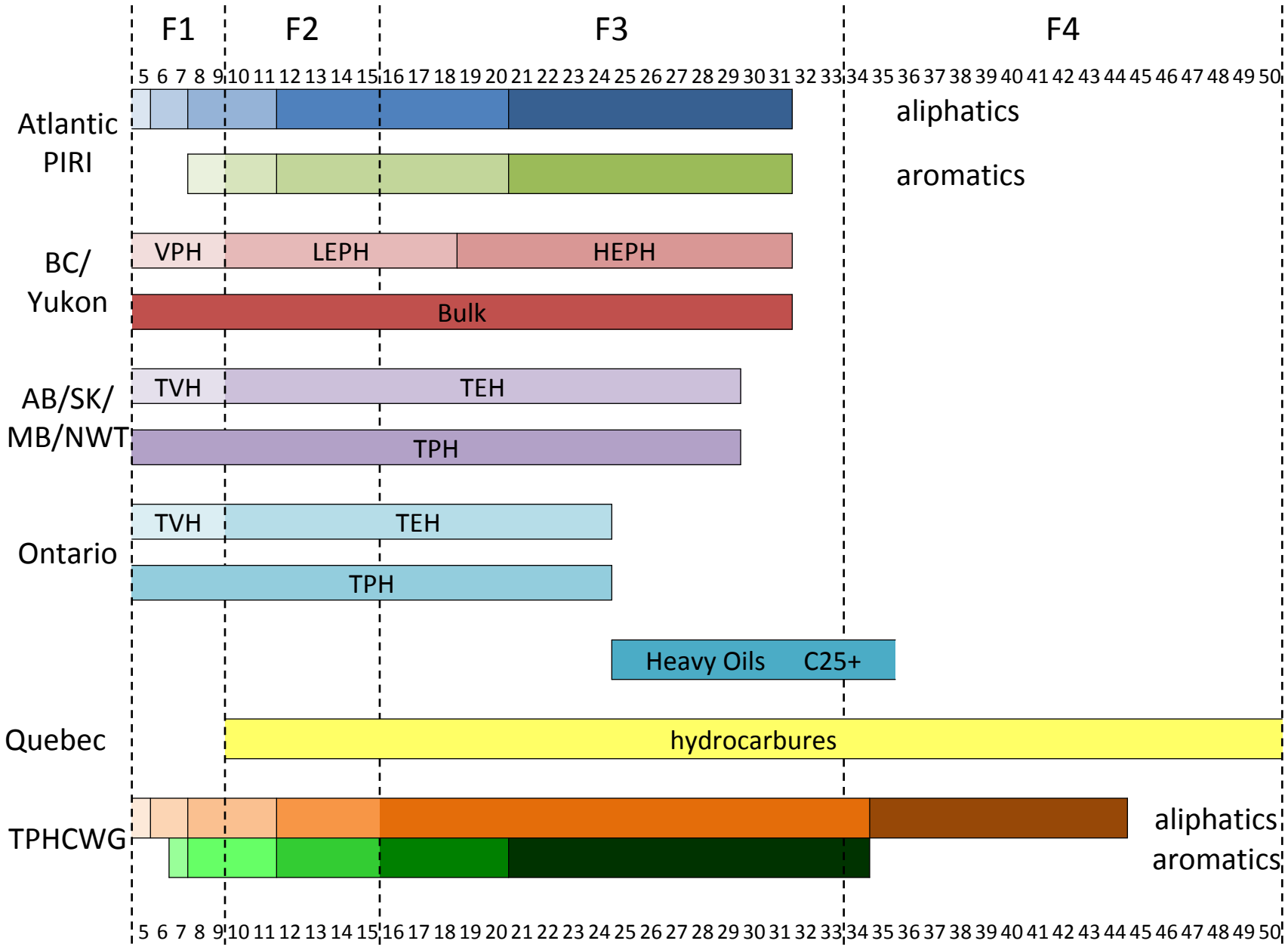
- Huge amount of sequence data
- Huge amount of genomics data
- Complex connectivity
- Understanding toxicological interactions
- Prediction of protein-coding genes
- Cell-cell interaction
- Cell-tissue-gene level interactions
- Genome has a multi-dimensional structure

F1-Hydrocarbon Example

- F1 hydrocarbon mixture
  - **55%C6-C8 aliphatics**
    (n-hexane may vary between 3% to 12% or more?)
  - **36% C8-C10 aliphatics**
  - **9% C8-C10 aromatics**

- F1 PHC = [F1 –BTEX]
- n-hexane is used as a surrogate

F1-Hydrocarbon Example



F1-Hydrocarbon Example

Fraction	Equivalent Carbon #	Corresponding TPHCWG subfractions	TDI (mg/kg·d)	RfC (mg/m³)	Critical Effect used by TPHCWG to derive criteria
F1	C_6 to C_{10}	aromatics $C_{>7}$-C_8	-[a]	-[a]	-[a]
		$C_{>8}$-C_{10}	0.04	0.2	hepatotoxicity, neurotoxicity
		aliphatics C_6-C_8	5.0	18.4	neurotoxicity
		$C_{>8}$-C_{10}	0.1	1.0	Liver and blood changes
F2	$C_{>10}$ to C_{16}	aromatics	0.04	0.2	decreased body weight
		$C_{>10}$-C_{12}	0.04	0.2	decreased body weight
		$C_{>12}$-C_{16}	0.1	1.0	Liver and blood changes
		aliphatics $C_{>10}$-C_{12}	0.1	1.0	Liver and blood changes
		$C_{>12}$-C_{16}			
F3	$C_{>16}$ to C_{34}	aromatics	0.03	NA[b]	nephrotoxicity
		$C_{>16}$-C_{21}	0.03	NA[b]	nephrotoxicity
		$C_{>21}$-C_{34}	0.1	1.0	hepatic granuloma
		aliphatics $C_{>16}$-C_{21}	2.0	NA[b]	hepatic granuloma
		$C_{>21}$-C_{34}			
F4	$C_{>34}$ to C_{50}	aromatics	0.03	NA[b]	nephrotoxicity
		$C_{>34}$	20.0	NA[b]	hepatic granuloma
		aliphatics $C_{>34}$			

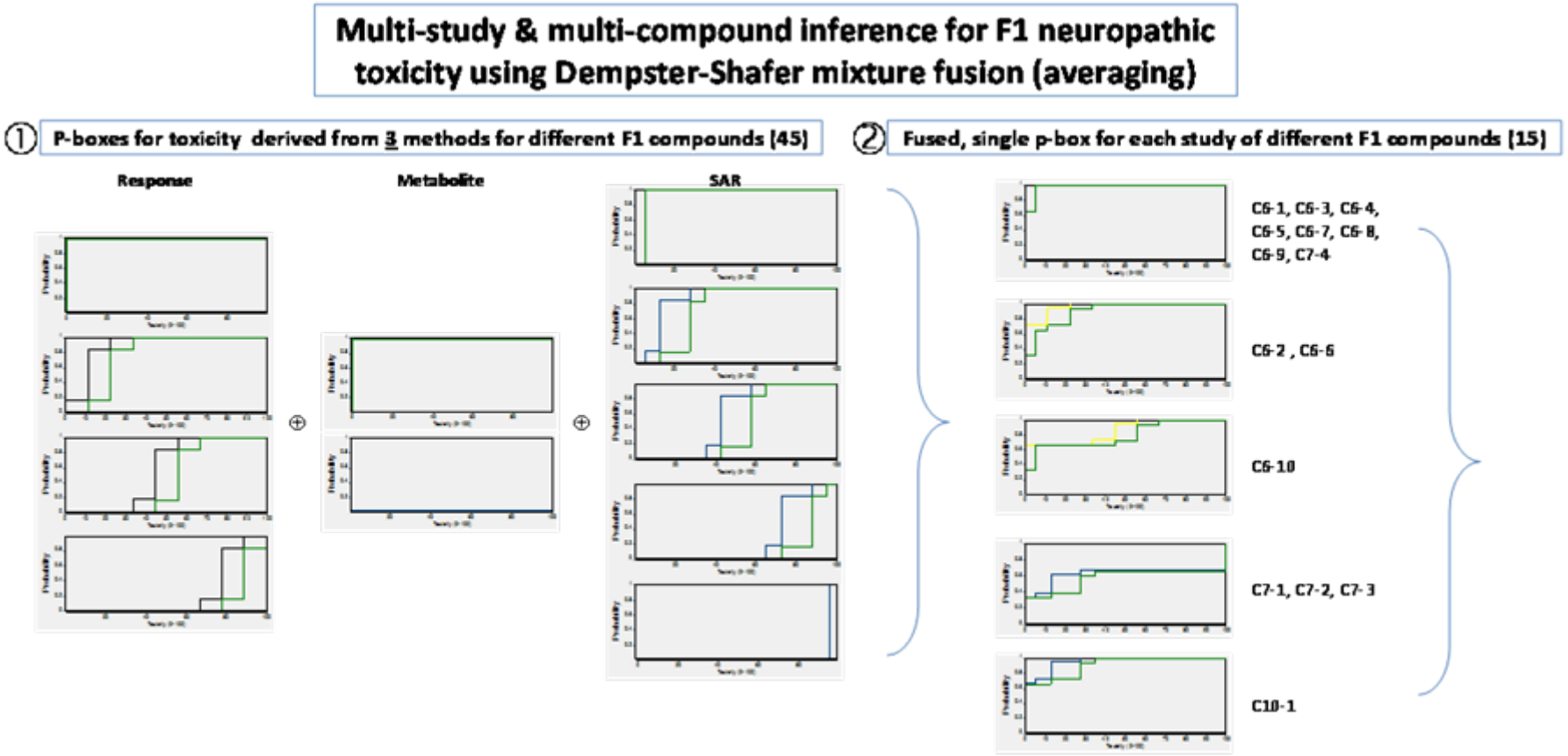
CCME (2008) & Edwards (1997)

F1-Hydrocarbon Example

Review of neurotoxicity studies for F1

Compound	Author	Subjects	Duration	Delivery	Dose	Effects	Response	med
Heptane	(Takeuchi et al. 1981)	Rat		12h/d,7d/w, 16w		3000	no histopathological signs of neurotoxicity	no
	(Frontali et al. 1981)	Rat		9h/d,5d/w 30 wks		1500 ppm	no evidence of histopathological neurotoxicity	no
	(Bahima et al. 1984)	female rat		6h/d, 5d/w, 12 wks		2000 ppm	no clinical signs of neurotoxicity	no
2-methyl Hexane	(Perbellini *et al*. 1985; Sayre *et al*. 1986)	human/rat					neurotoxic metabolites detected	no
3-methyl hexane	(Valentini et al. 1994)	Human	8-10 hr		case study exposure	36ppm heptane 16ppm 3-methyl hexane	peripheral neuropathy, induced by MEK?	med*
Methyl cyclo hexane	(Parnell et al. 1988)	Rats	every second day for 14d		0.8g/kg by gavage		Histopathologic examination of the rat kidney slices indicated only very slight traces of nephropathy,	NA
C7 Mixtures	(MacEwen and Vernot 1985)	dogs, rats, mice, hamsters	.Year-long exposures			0, 400, 2000 ppm	mean body wt depression in hamsters and male rats. Only significant lesions noted was progressive renal nephropathy seen in virtually all of the male rats	NA

F1-Hydrocarbon Example

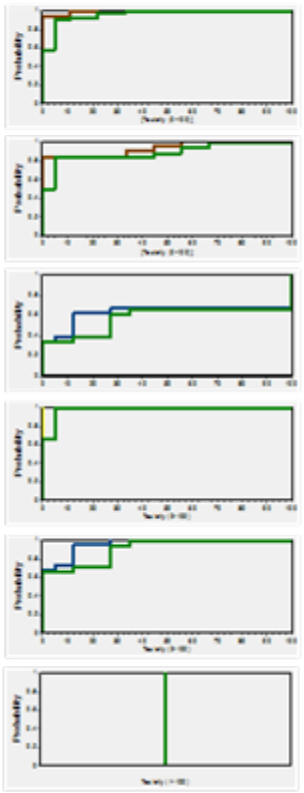


F1-Hydrocarbon Example

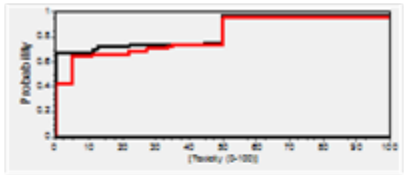
③ Fused p-boxes for 7 F1 compounds (various)

④ Assign weights as per 7-compound mass composition (percentage)

⑤ Fused p-box for F1



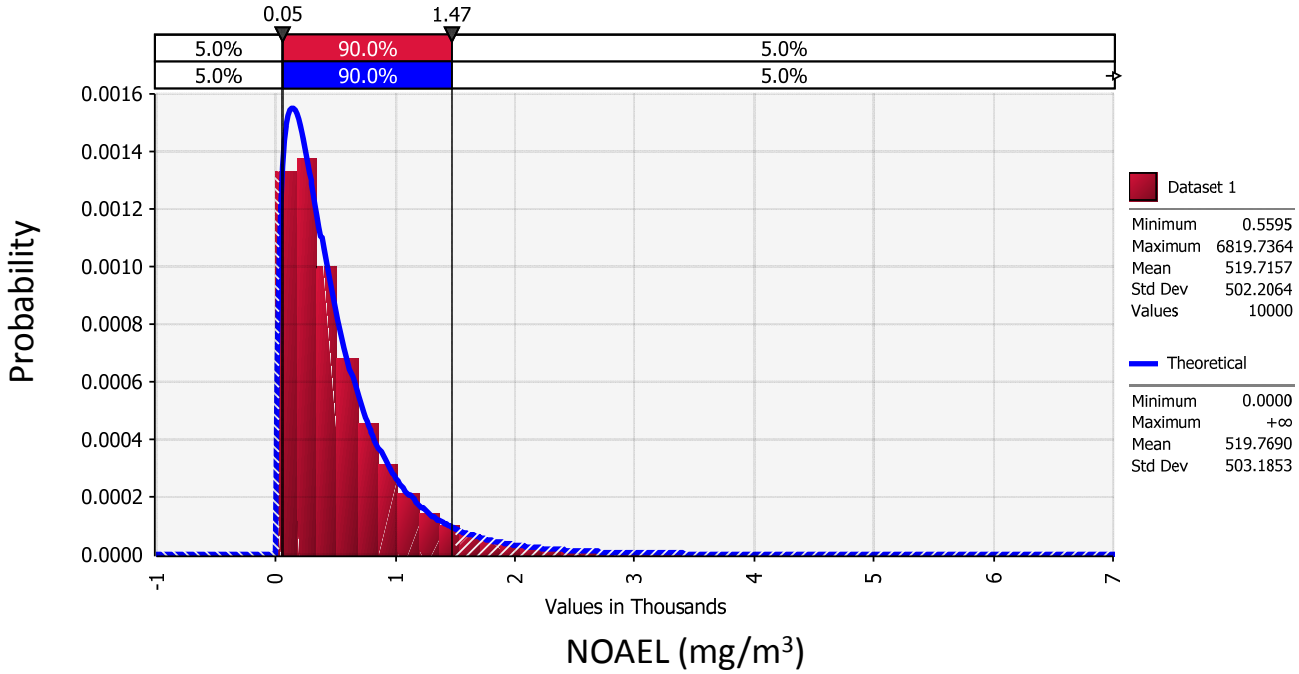
2-methylpentane, 3-methylpentane	28.72% 18.35%
cyclohexane	6.19%
heptane	9.78%
2-methylhexane	12.35%
n-decane	2.23%
n-hexane	22.38%



F1-Hydrocarbon Example

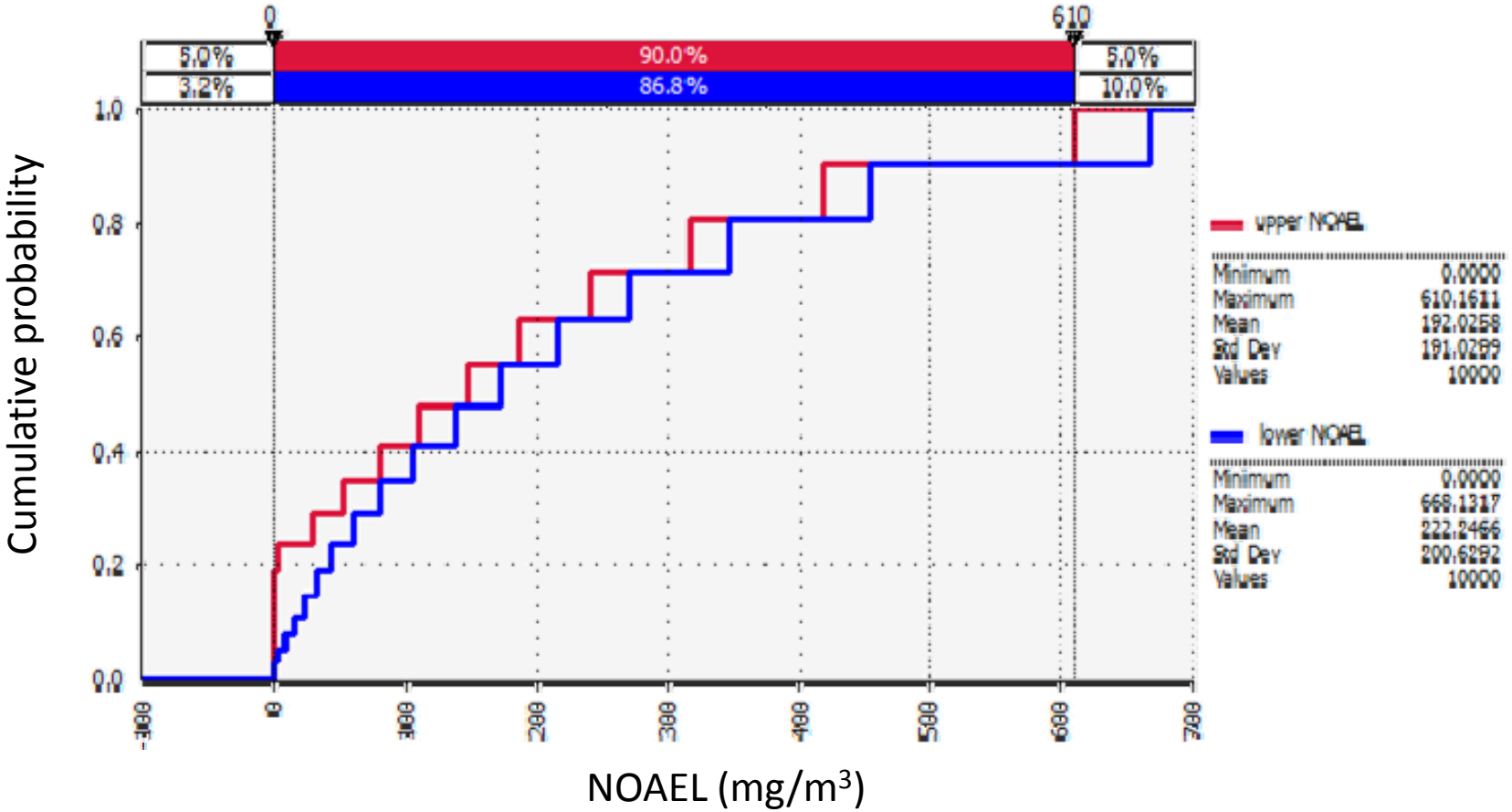
Feature Level Data Fusion: Dose-Response assessment

The toxicity of each compound was applied to the probability density function of the NOAEL concentrations from studies on n-hexane, for which there was much more toxicity data.



PDF of NOAEL from n-hexane subchronic neurotoxicity studies

F1-Hydrocarbon Example



p-box for neurotoxicity NOAEL for all of F1

F1-Hydrocarbon Example

Decision Level Data Fusion: Risk Characterization

The NOAEL from the dose-response assessment applies for rats in a sub-chronic study. Where NOAEL values were not available, the LOAEL values were divided by an uncertainty factor of 10. Other uncertainty factors that can be applied include:

- 10 for inter-species differences
- 10 for intra-species differences
- 3 for deficiencies in the data set.

No uncertainty factor is being used for the severity of toxic effects: a factor was included in calculating the combined NOAEL for F1.

F1-Hydrocarbon Example

Alternative Endpoints

Whether Current Inhalation Reference Concentrations are protective against irritancy for C_6-C_8 aliphatics?

Is this the most sensitive end point? Other health effect endpoints are being evaluated

Limited preliminary analysis of system biology datasets

Paradox of Risk Management

"You always got to be prepared, but you never know for what"

"Sugar Baby" Bob Dylan