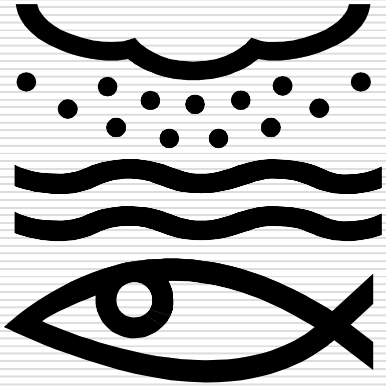
Representation of variability and uncertainty in multi-media exposure models

Consequences for risk assessments of contaminated land

Tomas Öberg
University of Kalmar
Email: tomas.oberg@hik.se



Presentation vid Toxikologiska rådets seminarium om matematisk-statistiska metoder i hälso- och miljöriskbedömning 2007-05-08

Introduction

- Environmental risk assessment relies heavily on mathematical models
- Uncertainty arises from
  - choice and structure of these models
  - model parameterization

$$DF_{gw} = \frac{L \cdot I \cdot W}{K \cdot i \cdot d_{mix} \cdot (2 \cdot y_{mix} + W) + (W + y_{mix}) \cdot (L + X) \cdot I}$$

$$d_{mix} = \sqrt{0.0112 \cdot (L + X)^2} + d_a \cdot \left[1 - \exp\left(-\frac{(L + X) \cdot I}{K \cdot i \cdot d_a} \right) \right]$$

$$y_{mix} = \sqrt{0.0112 \cdot (L + X)^2}$$

Separation of uncertainty

- Variability (natural variation)

  treated separately from

- Uncertainty due to lack of knowledge (epistemic uncertainty)
  - In this presentation, knowledge uncertainty will be denoted simply as ‘uncertainty’

Interval analysis

- A simple approach to carry uncertainties through all the steps of a calculation
- Interval arithmetic

Assume that $x_1 < x_2$ *and* $y_1 < y_2$:

$x + y = [x_1 + y_1, x_2 + y_2]$

$x - y = [x_1 - y_2, x_2 - y_1]$

$x * y = [\min(x_1y_1, x_1y_2, x_2y_1, x_2y_2), \max(x_1y_1, x_1y_2, x_2y_1, x_2y_2)]$

$x / y = [x_1/y_2, x_2/y_1], y \neq 0$

Example: Intake of arsenic $I_{As}=C_s*SI/BW$, *where Cs = [10,100] mg/kg, SI=[10,200] mg/day, and BW=[10,30] kg, then*

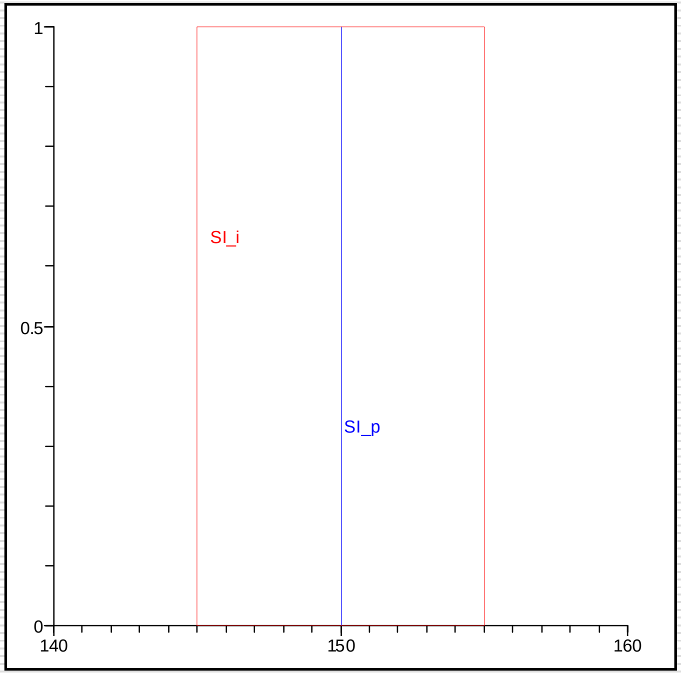
$I_{As}=([10,100]*[10,200]/[10,30])*10^{-6}$ *mg/kg/day =*

$=([10*10,100*200]/[10,30])*10^{-6}=([100,20000]/[10,30])*10^{-6}$

$= [100/30,20000/10]*10^{-6}=[3.3,2000]*10^{-6}$ *mg/kg/day*

Significant digits

- Parameters given as scalars (e.g., in point estimates)
  - Precise number, e.g. 150 mg soil/day
  - Imprecise number (interval), e.g. [145,155] mg soil/day
- The multiplicative effect!!
  - 1.05^20≈2.7
  - 0.95^20 ≈0.36

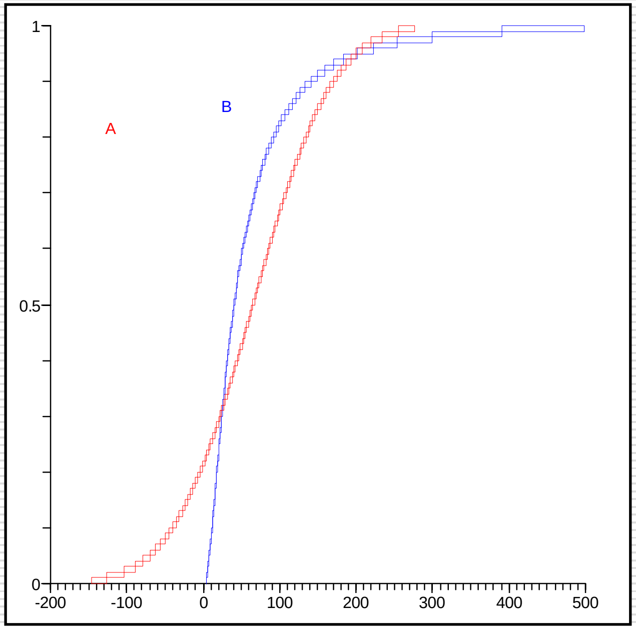


Intervals are not optimal

- Additional information is not used
  - Central tendency (mean, median, mode)
  - Spread (percentiles, standard deviation)
  - Shape (positive values, symmetry, unimodal)
- Probabilistic methods, an alternative framework
  - Precise parametric distributions (normal, lognormal)
  - Empirical distributions

Distribution selection is critical

- Selection
  - Graphical
  - Statistical tests
  - Expert judgment
- The uncertainty in distribution shape is seldom evaluated
  - Difficult when simulating inputs through random sampling from precise distributions (Monte Carlo)

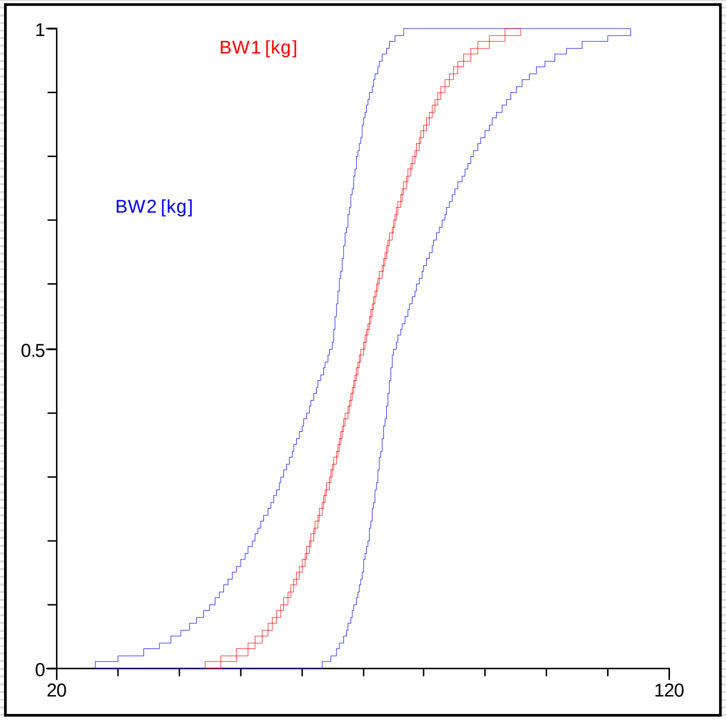


Example: *Daily soil intake: mean=64 and std dev=82 mg/day*

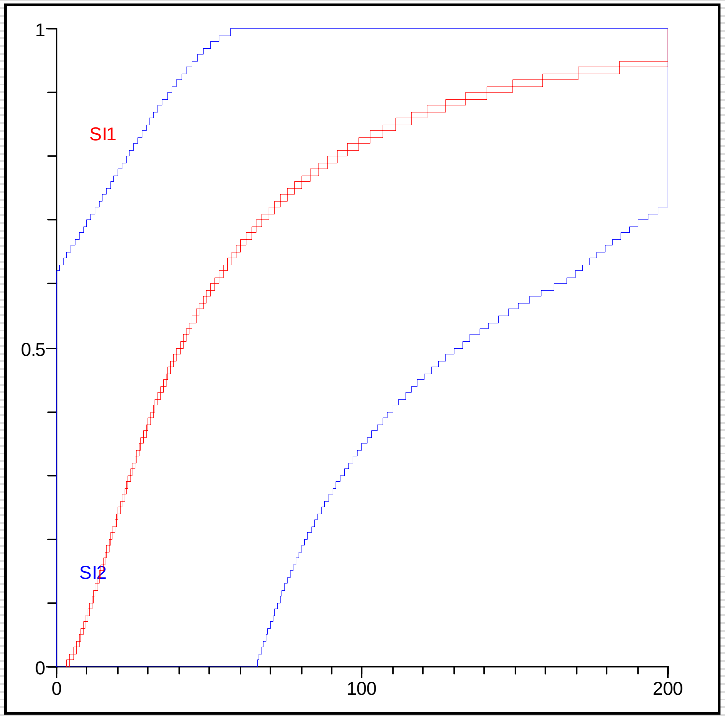
Probability bounds analysis

- Intervals and precise distributions combined
  - Use available information, but no additional assumptions
- Possible distributions are enclosed
  - P-box: A class of distribution functions $F(x)$ bounded by two cumulative distribution functions $F_1(x)$ and $F_2(x)$ such that $F_1(x) \leq F(x) \leq F_2(x)$ for all x.
- Converge to precise distributions when all uncertainty is removed
- Fast calculation algorithms available

P-boxes, two examples



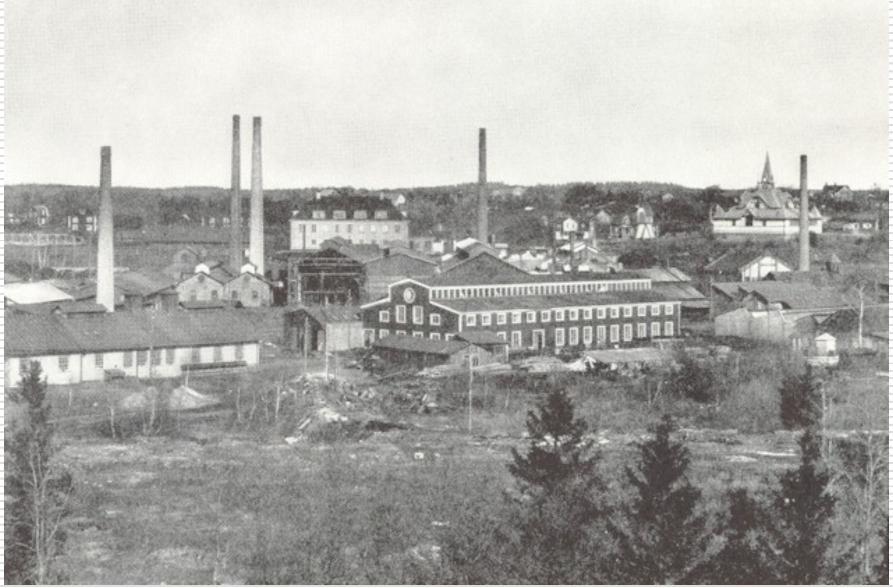
Normal distribution: Mean =70 and std dev=10 kg, or [65,75] and [5,15].



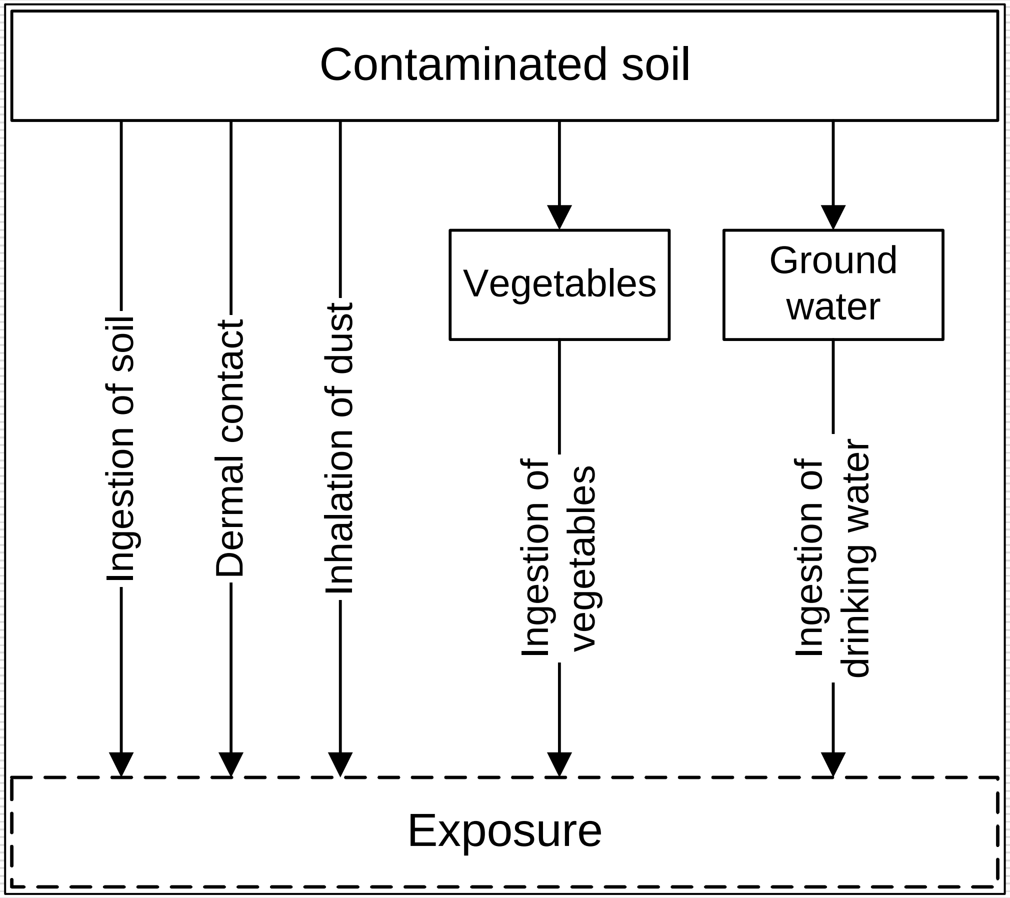
Lognormal(65,82), max=200, or p-box posmeanstddev(65,82), max=200.

A case study

- Site contaminated by metallurgical industry for over two centuries
- Cadmium selected for this evaluation
  - Intake ≈ TDI, choice of method matters
- Sensitive land use scenario
- Exposure of children



The exposure model



Transport sub-models

- Transport to groundwater

$$C_{gw} = C_t \cdot \left[K_d + \frac{\theta_w}{\rho_b} \right]^{-1} \cdot DF_{gw}$$

- Transport to plants

$$K_{pl} = BCF_{stem} \cdot f_{stem} \cdot r_{stem} + BCF_{root} \cdot f_{root} \cdot r_{root}$$

Intake estimates

$$I_{is} = \frac{C_s \cdot SI \cdot t_{is}}{365}$$

$$I_{du} = \frac{C_s \cdot SE \cdot A \cdot f_{du} \cdot t_{du}}{365}$$

$$I_{id} = \frac{C_s \cdot (C_{d,in} \cdot f_{d,in} \cdot f_{t,in} + C_{d,out} \cdot f_{d,out} \cdot f_{t,out}) \cdot BR \cdot LR \cdot t_{id}}{365}$$

$$I_{ig} = \frac{C_s \cdot K_{pl} \cdot R_{ig} \cdot f_h \cdot t_{ig}}{365}$$

$$I_{iw} = \frac{C_{gw} \cdot WC \cdot t_{iw}}{365}$$

$$I_{tot} = \frac{I_{is} + I_{du} + I_{id} + I_{ig} + I_{iw}}{BW}$$

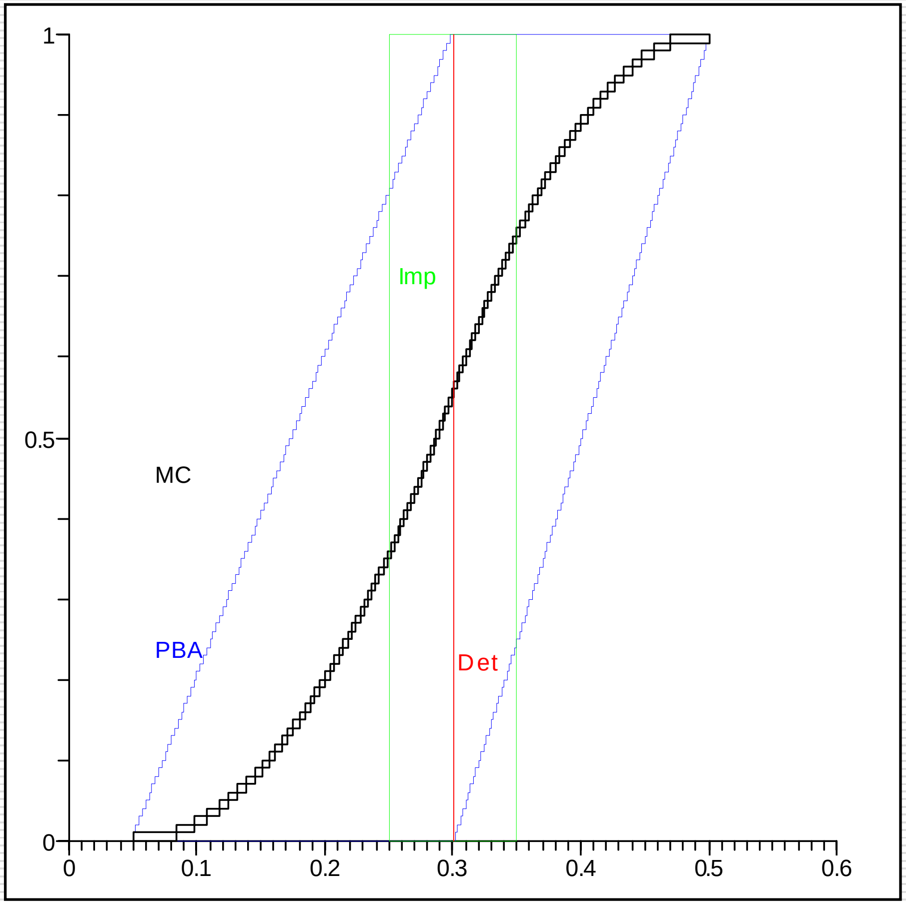
Parameterization

- Deterministic
  - Swedish EPA
- Probabilistic methods
  1. Exposure Factors Handbook (U.S. EPA)
  2. Exposure Factors Sourcebook for European Populations (ECETOC)
  3. Revised CSOIL parameter set
  4. Statistics from SCB and Swedish National Food Administration

Variables	Symbol	Det.	Det. with error	Monte Carlo	PBA	PBA with error
General						
Concentration in soil 0-1 m, mg/kg	C_s	4.95	4.95	LN(3.1,1.0)	LN(3.1,1.0)	LN(3.1,1.0)
Concentration in soil total, mg/kg	C_t	17.4	17.4	LN(6.9,5.7)	LN(6.9,5.7)	LN(6.9,5.7)
Body weight child, kg	BW	15	[14.5,15.5]	N(18,2.66), min=5, max=25	symmeanstddev(18,2.66), min=5, max=25	symmeanstddev(18,2.66), min=5, max=25
Transport to groundwater						
Distribution soil-water, dm^3/kg	K_d	100	[50,150]	100	100	[50,150]
Soil water content, dm^3/dm^3	θw	0.3	[0.25,0.35]	triangular(0.05,0.3,0.5)	minmaxmode(0.05,0.5,0.3)	minmaxmode(0.05,0.5,0.3)
Soil bulk density, kg/dm^3	ρ_b	1.5	[1.45,1.55]	triangular(0.25,1.2,1.6)	minmaxmode(0.25,1.6,1.2)	minmaxmode(0.25,1.6,1.2)
Dilution soil-groundwater	DF_{gw}	0.08	[0.022,0.27]*	0.08	0.08	[0.022,0.27]*
Transport to plants						
Bioconcentration stem and leaf, (mg/kg dry)/(mg/kg soil)	BCF_{stem}	0.7	[0.65,0.75]	LN(0.85,0.61)	posmeanstddev(0.85,0.61)	posmeanstddev(0.85,0.61)
Fractional consumption, stem and leaf	f_{stem}	0.5	[0.45,0.55]	LN(0.38,0.13), max=1	posmeanstddev(0.38,0.13), max=1	posmeanstddev(0.38,0.13), max=1
Dry to fresh weight stem and leaf, kg/kg	r_{stem}	0.117	0.117	0.117	0.117	0.117
Bioconcentration root, (mg/kg dry)/(mg/kg soil)	BCF_{root}	0.15	[0.145,0.155]	LN(0.35,0.33)	posmeanstddev(0.35,0.33)	posmeanstddev(0.35,0.33)
Fractional consumption, root	f_{root}	$1- f_{stem}$	$1- f_{stem}$	$1- f_{stem}$	$1- f_{stem}$	$1- f_{stem}$
Dry to fresh weight root, kg/kg	r_{root}	0.202	0.202	0.202	0.202	0.202
Direct oral intake						
Daily soil intake, mg/kg	SI	150	[145,155]	LN(65,82), max=200	posmeanstddev(65,82), max=200	posmeanstddev(65,82), max=200
Exposure time, day	t_{is}	365	365	365	365	365

Variables	Symbol	Det.	Det. with error	Monte Carlo	PBA	PBA with error
Dermal uptake						
Dermal soil exposure, $mg/m^2/day$	SE	5100	[5050,5150]	LN(2000,990), max=7400	posmeanstddev(2000,900), max=7400	posmeanstddev(2000,900), max=7400
Exposed skin area, m^2	A	0.28	[0.275,0.285]	LN(0.18,0.017), max=0.24	posmeanstddev(0.18,0.017), max=0.24	posmeanstddev(0.18,0.017), max=0.24
Relative dermal absorption	f_{du}	0.14	[0.135,0.145]	0.14	0.14	[0.135,0.145]
Exposure time, day	t_{du}	80	[75,85]	80	80	[75,85]
Inhalation uptake						
Concentration of respirable dust indoors, mg/m^3	$C_{d,in}$	0.052	[0.0515,0.0525]	triangular(0.037,0.069, 0.1)	minmaxmode(0.037,0.1,0.069)	minmaxmode(0.037,0.1,0.069)
Fraction of dust indoors from contaminated area	$f_{d,in}$	0.8	[0.75,0.85]	triangular(0.5,0.65, 0.8)	minmaxmode(0.5,0.65,0.8)	minmaxmode(0.5,0.65,0.8)
Fraction of time spent indoors	$f_{t,in}$	0.88	[0.875,0.885]	uniform(0.75,1)		
Concentration of respirable dust outdoors, mg/m^3	$C_{d,out}$	0.07	[0.065,0.075]	triangular(0.05,0.075, 0.1)	minmaxmode(0.05,0.075,0.1)	minmaxmode(0.05,0.075,0.1)
Fraction of dust outdoors from contaminated area	$f_{d,out}$	0.5	[0.45,0.55]	0.5	minmax(0.75,1)	minmax(0.75,1)
Fraction of time spent outdoors	$f_{t,out}$	1- $f_{t,in}$	1- $f_{t,in}$	1- $f_{t,in}$	1- $f_{t,in}$	1- $f_{t,in}$
Breathing rate, m^3/day	BR	7.6	[7.55,7.65]	7.6	7.6	[7.55,7.65]
Lung retention	LR	0.75	[0.745,0.755]	0.75	0.75	[0.745,0.755]
Exposure time, day	t_{id}	365	365	365	365	365
Intake from vegetables						
Daily consumption of vegetables, kg/day	R_{ig}	0.15	[0.145,0.155]	N(0.13,0.04), min=0	symmeanstddev(0.13,0.04), min=0	symmeanstddev(0.13,0.04), min=0
Fraction of consumed vegetables grown on site	f_h	0.3	[0.25,0.35]	triangular(0,0.13, 0.3)	minmaxmode(0,0.3,0.13)	minmaxmode(0,0.3,0.13)
Exposure time, day	t_{ig}	365	365	365	365	365
Intake with drinking water						
Daily water consumption, dm^3/day	WC	1	[0.5,1.5]	N(0.87,0.49), min=0.1, max=3	symmeanstddev(0.87,0.47), min=0.1, max=3	symmeanstddev(0.87,0.47), min=0.1, max=3
Exposure time, day	t_{iw}	365	365	365	365	365

Example: Soil water content



Det.=0.3

Imp.=[0.25,0.35]

MC=triangular(0.05,0.3,0.5)

PBA=minmaxmode(0.05,0.5,0.3)

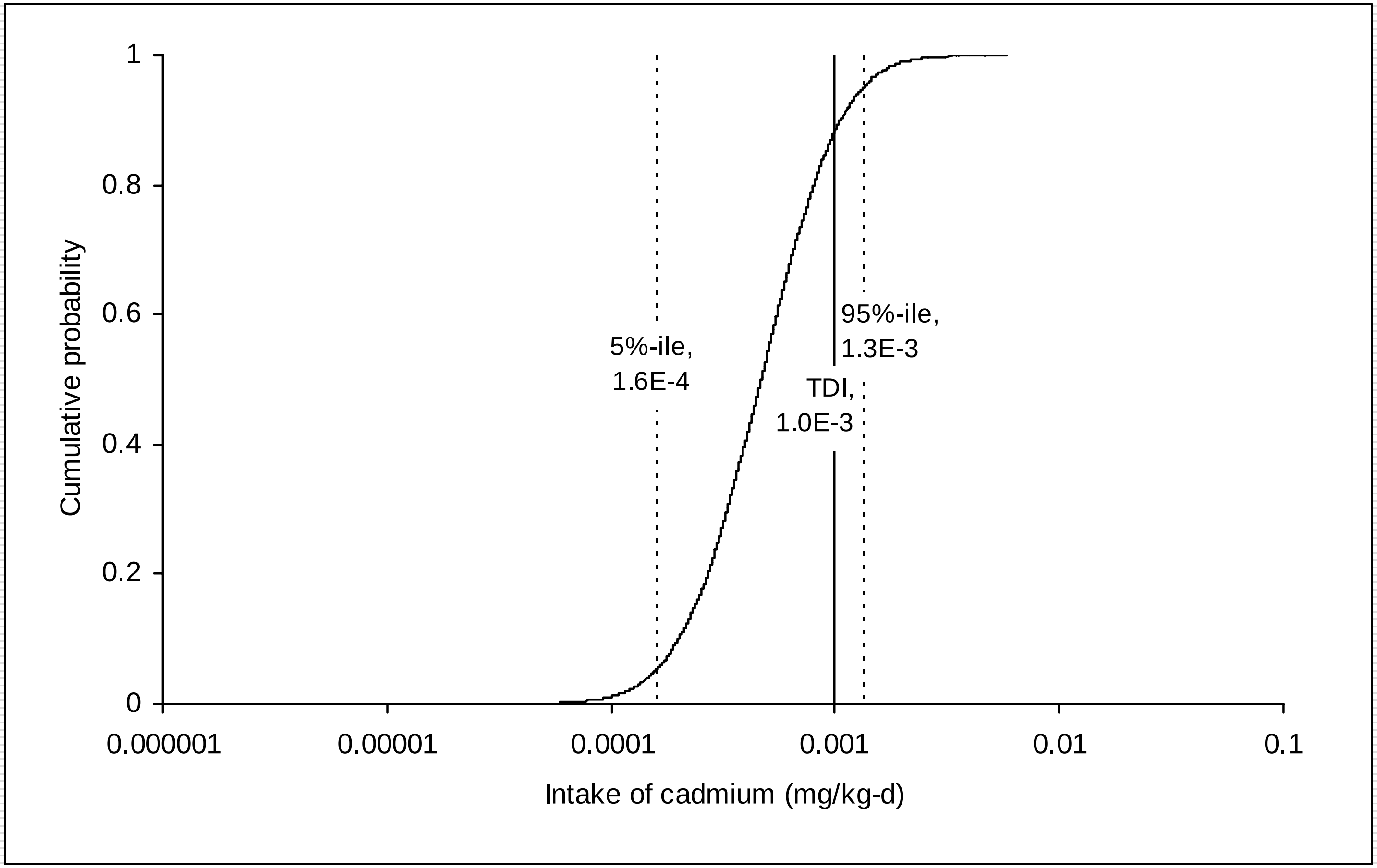
Evaluation

1. Deterministic point estimate
2. Deterministic point estimate with rounding errors
3. Monte Carlo-simulation
4. Probability bounds analysis (PBA)
5. Probability bounds analysis (PBA) with error

Deterministic calculations

- Point estimate
  - 1.8×10^{-3} mg/kg/day
- Deterministic with rounding errors
  - 6.9×10^{-4} to 1.1×10^{-2} mg/kg/day
- Compare to the tolerable daily intake (TDI) applied by SEPA
  - 1.0×10^{-3} mg/kg/day
  - 31% below to 1000% above

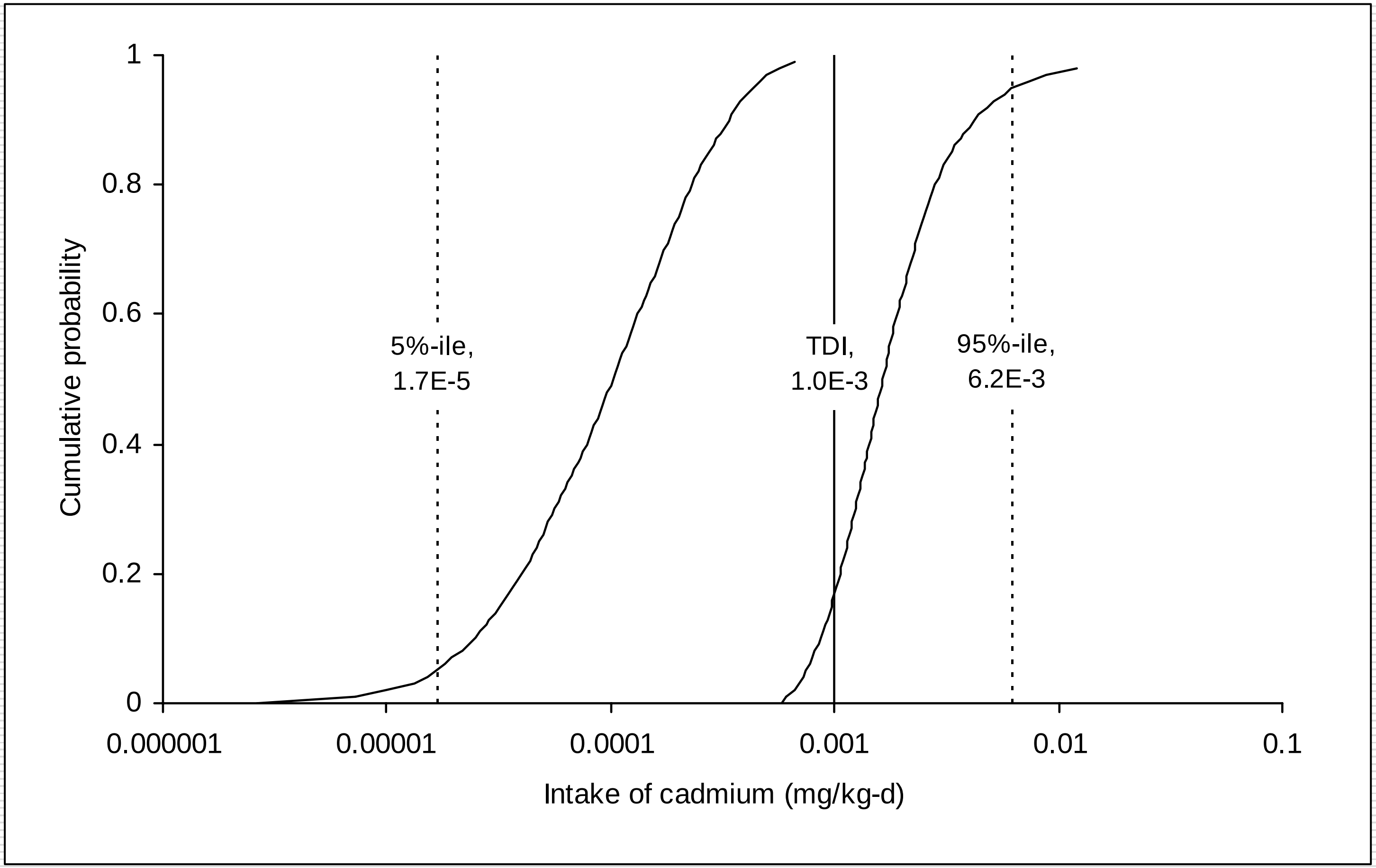
Monte Carlo-simulation



Monte Carlo-simulation, cont.

- Percentiles
  - 5th 1.6×10^{-4} mg/kg/day
  - 50th 4.7×10^{-4} mg/kg/day
  - 95th 1.3×10^{-3} mg/kg/day
- Sensitivity analysis (Spearman rank correlation)
  - Total soil concentration ($r_s=0.54$)
  - Average water consumption ($r_s=0.37$)
  - Fraction of consumed vegetables grown on site ($r_s=0.33$)
  - Three other variables related to intake, surface soil concentration and body weight ($r_s>0.2$)

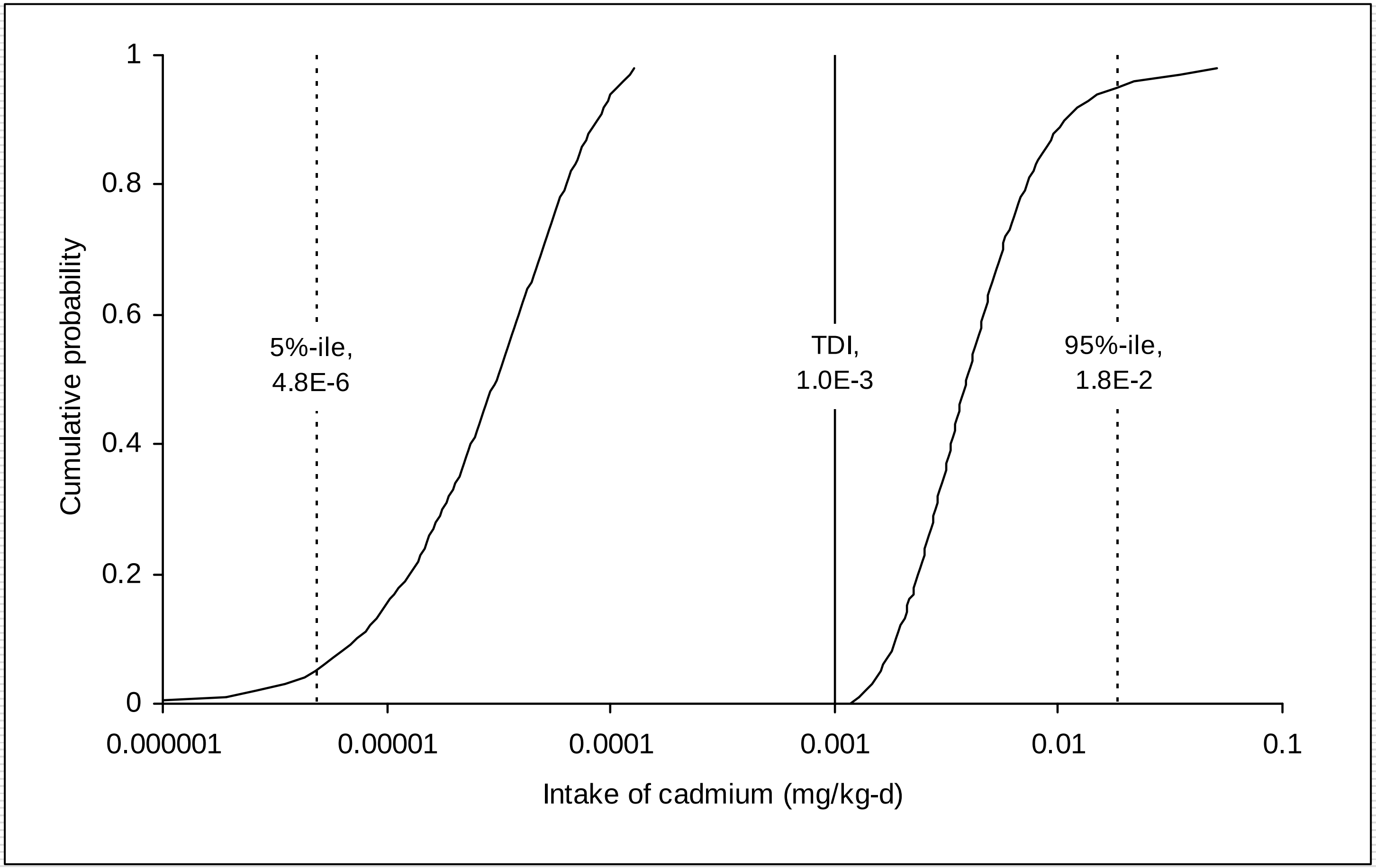
Probability bounds analysis



Probability bounds analysis, cont.

- Percentiles
  - Lower 5th 1.7×10^{-5} mg/kg/day
  - 50th 1.0×10^{-4} to 1.6×10^{-3} mg/kg/day
  - Upper 95th 6.2×10^{-3} mg/kg/day
- Sensitivity analysis (’pinching’), with same results as in the MC simulation
- Both the MC and PBA indicate that we could eliminate the oral, dermal, and inhalation exposure pathways

PBA with rounding errors



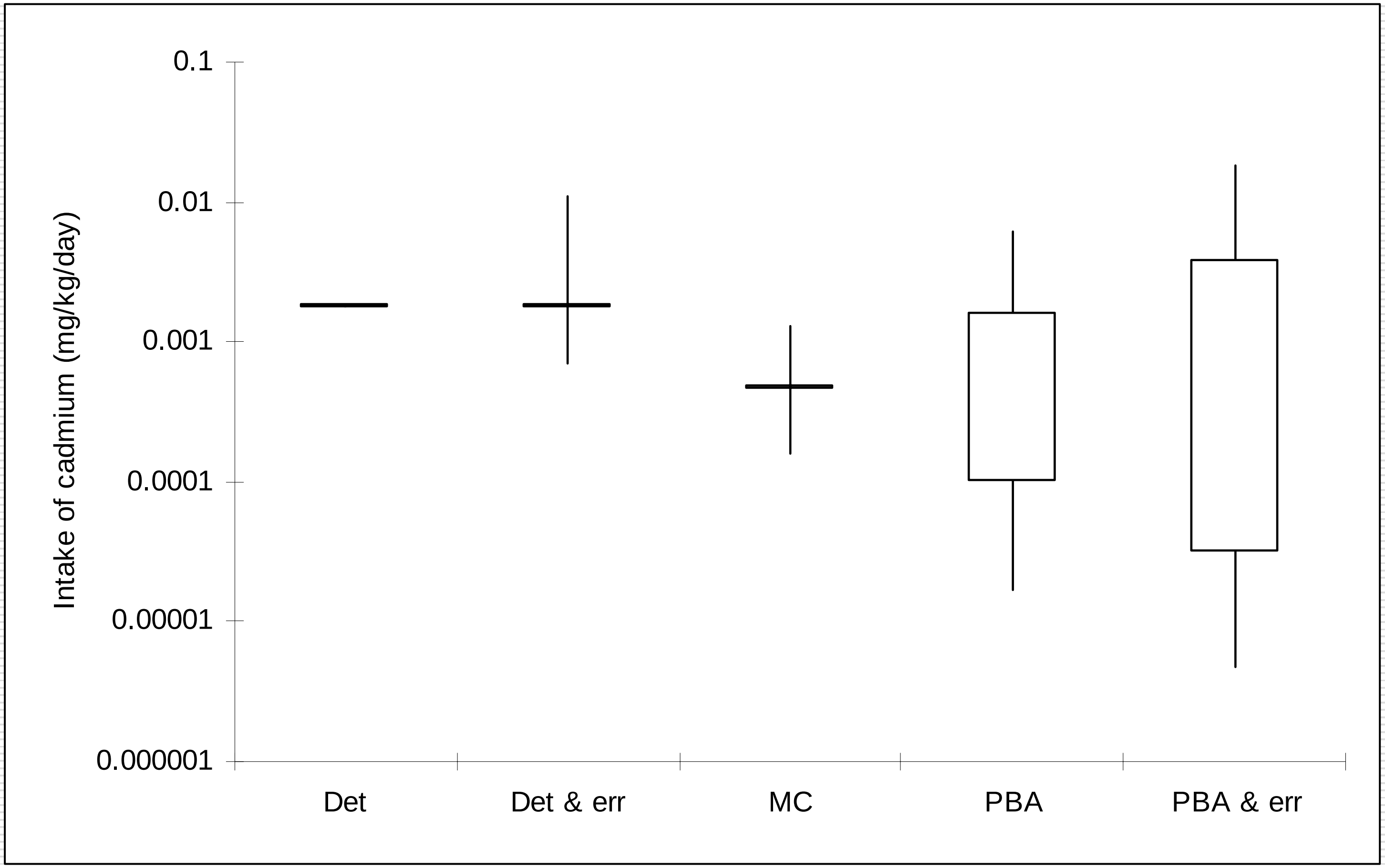
PBA with rounding errors, cont.

- Percentiles
  - Lower 5th 4.8×10^{-6} mg/kg/day
  - 50th 3.1×10^{-5} to 3.9×10^{-3} mg/kg/day
  - Upper 95th 1.8×10^{-2} mg/kg/day
- All evaluations (MC & PBA) have so far assumed independence between the input variables
- What will happen if this assumption is removed

PBA with rounding errors, cont.

- We removed several of the major independence assumptions
- Percentiles
  - Lower 5th 1.7×10^{-6} mg/kg/day
  - 50th 1.3×10^{-5} to 9.4×10^{-3} mg/kg/day
  - Upper 95th 6.7×10^{-2} mg/kg/day
- The probability bounds widen substantially, but less than an order of magnitude
- Our results thus corroborate observations by others, that the choice of input distributions is more important

A comparison of the estimates



Is uncertainty overstated?

- It is reasonable to ask, but the answer is no!
- The method only evaluates the information available to the analyst
  - P-boxes are free of distribution assumptions

    but
  - a probability distribution can be bound with varying degree of detail

    and
  - when unceratinty diminish, then PBA converge to the same results as a MC simulation

The role of the risk analyst

- Information supplied by the analyst may be
  - too precise
    or
  - too imprecise
- Extensive studies the last 25 years seem to suggest that over-confidence is more common

Can we improve?

- Fitting distributions is not enough!
- Better to
  - evaluate a range of reasonable distributions
  - evaluate variables assigned as numbers for rounding errors
- Only focus on the upper-bound estimate is not useful!
- Better to
  - disclose a complete set of information
  - acknowledge the limitations of science

Some final points

- Decision-making needs to be risk-informed rather than risk-based
- Scientific uncertainty could then be treated as knowledge
- The ability of non-experts to understand uncertainty is often underestimated
- Uncertainty can be an advantage since it opens up a negotiation window
- Strictly regulated procedures limit the possibilities for a meaningful risk dialogue

Work in progress

- Identification of sources for data on exposure factors
  - Swedish population
  - Swedish environment
  - (International)
- Evaluation of data to support risk assessments
  - Deterministic

  and
  - Probabilistic

Acknowledgement

- The results reported here are from studies funded by the Swedish EPA and the research program 'Hållbar Sanering'.

- Reports (available as PDF at www.naturvardsverket.se):
  - Filipsson, M., Bergbäck, B., Öberg, T. Report in preparation.
  - Öberg, T., Sander, P., Bergbäck, B. *Probabilistisk riskbedömning fas 2.* Rapport 5621. Naturvårdsverket, 2006.
  - Öberg, T. *Probabilistisk riskbedömning fas 1.* Rapport 5532. Naturvårdsverket, 2006.

- Papers (available from author):
  - Sander, P., Bergbäck, B., Öberg, T. *Risk Analysis* **26**, 1363-1375 (2006)
  - Sander, P., Öberg, T. *Journal of Soils and Sediments* **6**, 55-61 (2006).
  - Öberg, T., Bergbäck, B. *Journal of Soils and Sediments* **5**, 213-224 (2005).