APPENDIX B

PROBABILISTIC MODELLING

PROBABILISTIC MODELLING

The probabilistic modelling has been conducted using a simple Excel™ spreadsheet (“CLEA_in”) linked to the existing CLEA model (v1.06). The CLEA_in spreadsheet contains distributions of possible parameter values for the key uncertain parameters identified from the sensitivity analysis. The CLEA_in spreadsheet contains a simple macro that chooses parameter values at random from the distributions and enters them into the CLEA model (via the “Advanced Settings” worksheets in CLEA). This is repeated 10,000 times in a “Monte Carlo” analysis with the macro recording the “ADE to HCV” ratios in a results worksheet of CLEA_in each time. These ratios are multiplied by the HCVs within CLEA to convert back to 10,000 estimates of ADE. Note that because the relationship between soil concentration and ADE in CLEA are linear (i.e. if soil concentration is doubled, ADE will double) all the probabilistic runs have been conducted assuming a soil concentration of 1 mg.kg^{-1} and the calculated ADEs multiplied by the pC4SL to derive the distribution of ADEs for that soil concentration.

Key uncertain parameters were identified from the sensitivity analysis (Section 3.4 of the main report) as those parameters that contributed most to uncertainty in the assessment criteria derived by deterministic modelling. Although there is uncertainty in the remaining parameters, the sensitivity analysis demonstrated that this does not give rise to significant uncertainty in the CLEA model outputs and these remaining parameters have not therefore been modelled probabilistically.

The parameters modelled probabilistically varied between land-uses and contaminants. For example, the soil to plant concentration factor is not used for the residential without consumption of homegrown produce or the commercial land-uses and so is not relevant for these scenarios. Another example is dermal absorption, which is assumed negligible for lead, and therefore parameters associated with the dermal contact pathways (such as adherence factor and maximum exposed skin fraction) have not been modelled probabilistically for this contaminant. Key parameters modelled probabilistically for each land-use/substance are shown in Table 1 of this appendix.

Table 1 Parameters modelled probabilistically

Parameter	Substances	Generic Land-use				Correlation
		Residential		Allot-ments	Comm-ercial	
		With home grown prod.	Without home grown prod.			
Body weight	All	✓	✓	✓	✓	Correlated between age classes, i.e. a heavy one year old is assumed to become a heavy six year old. Body weight is also correlated with inhalation rate, i.e. a child in the upper percentile body weight will also have an upper percentile inhalation rate
Soil ingestion rate	All	✓	✓	✓	✓	Correlated between age classes
Exposure Frequency skin contact outdoors	As, BaP,Cd, CrVI	✓	✓	✓		Correlated between age classes
Soil to skin adherence factor outdoors	As, BaP,Cd, CrVI	✓	✓	✓		Correlated between age classes
Maximum exposed skin fraction outdoors	As, BaP,Cd, CrVI	✓	✓	✓		Correlated between age classes
Dermal absorption factor	BaP					Not correlated with other parameters
Inhalation rate	All	✓	✓		✓	Correlated between age classes and with body weight
Subsurface soil to indoor air correction factor	Benzene	✓	✓		✓	Not correlated with other parameters
Dust loading factor	As, BaP,Cd, CrVI, Pb	✓	✓		✓	Not correlated with other parameters
Soil to dust	As, BaP,Cd,	✓	✓		✓	Not correlated with other

Parameter	Substances	Generic Land-use				Correlation
		Residential		Allot-ments	Comm-ercial	
		With home grown prod.	Without home grown prod.			
transport factor	CrVI, Pb					parameters
Produce consumption rate	As, BaP,benzene ,Cd, Pb	✓		✓		Correlated between age classes. Also, consumers of homegrown produce assumed to be within the upper quartile of consumers of fruit and vegetables
Homegrown fraction	As, BaP,benzene ,Cd, Pb	✓		✓		Correlated between produce types, i.e. an individual who consumes potatoes, most of which are homegrown will also consume mostly homegrown root and green vegetables and fruit
Soil to plant concentration factors	As, BaP,Cd, Pb	✓		✓		Correlated between produce type, i.e. if a soil allows high plant uptake for potatoes, it will also allow high plant uptake for the remaining produce types

The distributions of possible values for the parameters modelled probabilistically have been derived using a probability density function (PDF) for each parameter. The type of distribution (e.g. normal, log normal, beta etc.) and associated attributes (e.g. mean, standard deviation or 95^{th} percentile) selected for each parameter have been chosen to best represent the range of distribution families considered. The PDF type and associated attributes are summarised for non contaminant specific parameters in Tables 2 to 4 of this appendix and contaminant specific parameters in substance specific Appendices C to H.

A separate Excel™ spreadsheet has been used to derive the distributions for each parameter. The PDFs have been used to derive a set of 999 values for each parameter that correspond to its corresponding PDF as follows:

- **Normal**: Excel function = NORMINV(P_c, a ,b), where a = mean, b = standard deviation and P_c = cumulative probability (varies from 0.001 to 0.999 in 0.001 increments)

- **Log normal**: Excel function = LOGINV(P_c,LN(a),b), where a = geomean and b = standard deviation of ln transformed data

- **Beta**: Excel function = BETAINV(P_c,α,β), where α = alpha and β = beta

- **Triangular**: Excel function based on:

  For x ≤ mode, $x = a + \sqrt{P_c(b-a)(c-a)}$

  For x > mode, $x = a + \sqrt{P_c(b-a)(c-a)} x = b - \sqrt{(1-P_c)(b-a)(b-c)}$

  Where,
  x = value
  a = minimum value
  b = maximum value
  c = mode

In the case of body weight, actual empirical distributions have been used as the input distribution, i.e. the measured body weights for groups of individuals in each age class.

As discussed above the Monte Carlo simulation is conducted by choosing a value at random for each parameter. Correlations between parameters (where applicable – see Table 1) are achieved by ensuring that the same cumulative probability is selected when choosing the value at random. So, for example, for correlation of the soil to plant concentration factor between produce types if the value randomly selected for green vegetables for one Monte Carlo simulation corresponds to a cumulative probability from the PDF of 0.900 (i.e. the 90th percentile soil to plant concentration factor), the corresponding cumulative frequency values (i.e. the 90th percentile values) are selected for the remaining produce types.

The results of the probabilistic modelling are presented graphically as:

- Reverse cumulative frequency (RCFs), i.e. graphs of the reverse cumulative frequency versus ADE for alternative pC4SLs derived deterministically using alternative sets of exposure parameters. The LLTC and background exposure are also marked on these graphs to provide an indication of the probability of the ADE to a random individual within the critical receptor group exceeding the LLTC or background exposure from soil concentrations equal to the alternative pC4SL. For example, as shown in Figure 1, there is an 11% probability that exposure would exceed the LLTC if the average soil concentration that the receptor is exposed to equals 5 mg.kg^{-1}. The RCF graphs are a useful indication of the range of exposures predicted by the probabilistic modelling; and

- Probability of exceedence versus soil concentration graphs. These show how the probability of the ADE exceeding the LLTC varies with soil concentration. This graph has the advantage over the RCF that it can show the probability of total or oral/dermal exposure exceeding the $LLTC_{oral}$ and the probability of inhalation exposure exceeding the $LLTC_{inhal}$. For example, as shown in Figure 2, for benzo(a)pyrene for residential land-use (with consumption of homegrown produce) the probability of inhalation exposure exceeding the $LLTC_{inhal}$ is negligible relative to the probability of total exposure exceeding the $LLTC_{oral}$. This graph also shows how the probability of exceedence varies with soil concentration.

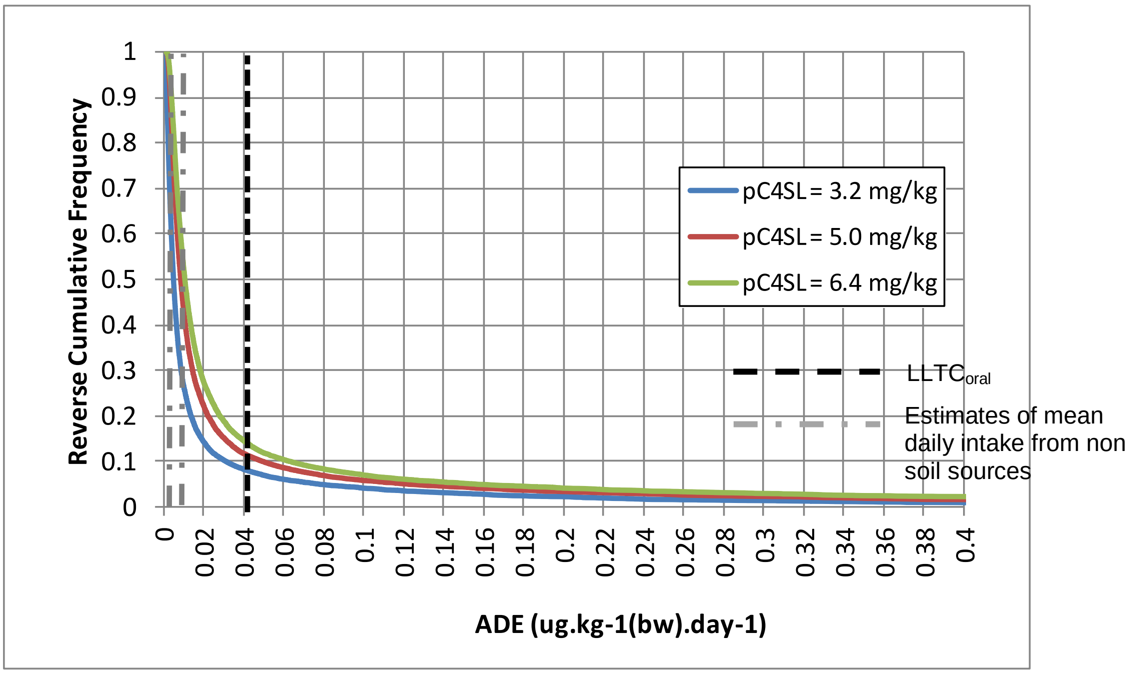

Figure 1: Reverse cumulative frequency graph of ADE for alternative values of pC4SL for BaP for residential (with consumption of homegrown produce) land-use

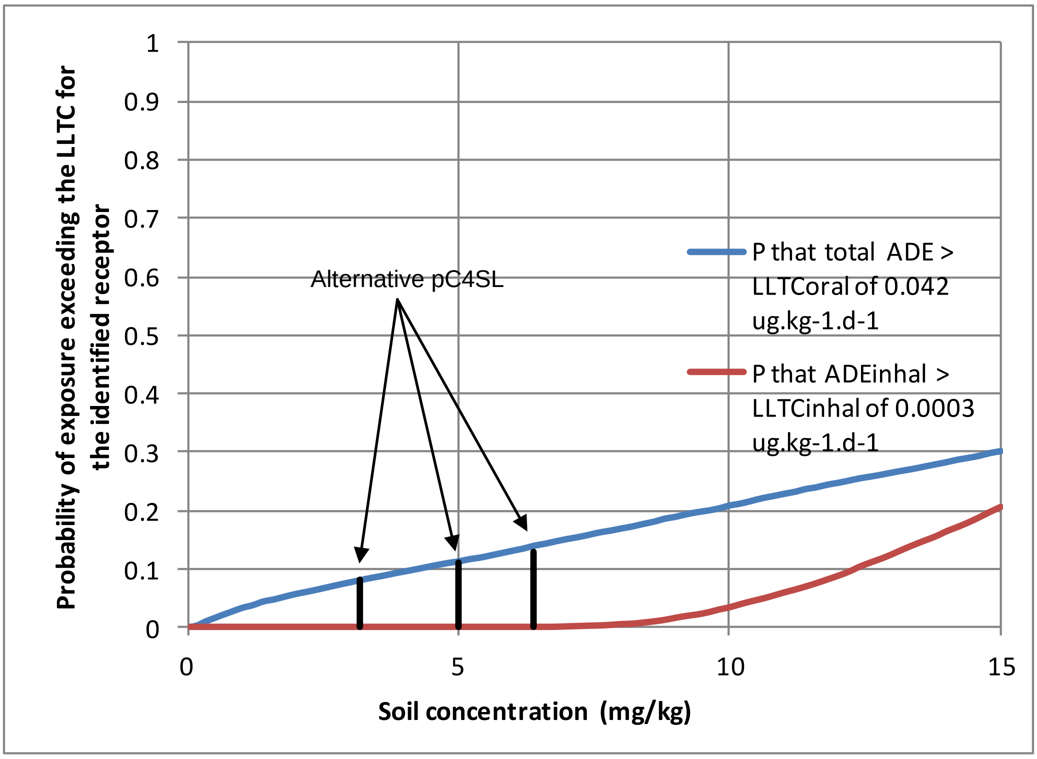

Figure 2: Probability of exposure exceeding LLTC with alternative values of pC4SL for BaP for residential (with consumption of homegrown produce) land-use

The probabilistic modelling has been conducted for four land-uses (residential, allotments and commercial) for all 6 test substances.