WP1: Generation of an *in vivo* database (P4, P24, P29 and P36)

Selection of reference chemicals

The project management opted for 97 chemicals to be selected as test items, based on statistical estimation of worthy sample size and feasibility of *in vitro* experimental testing according to available resources and projected schedule. A principal selection criterion, limiting the scope for eligibility, was availability of documented cases of acute poisoning in humans (accidental ingestion, suicidal overdose, etc.) including clinical/forensic blood concentration measurements from patients/victims. The chemical selection also incorporated nominations from project partners according to expediency of respective organ specificity research interests, including biokinetic modelling. The chemicals were readily available from regular laboratory suppliers, facilitating direct purchase of test items by project partners.

The 97 reference chemicals (Table in Appendix II) cover complementary representation of GHS toxicity categories (Fig 2a) and include different generic use classes (Fig 2b).

Compilation of animal and human data

Search and review of information has been comprehensive with systematic compilation into the project database, AcutoxBase, in standard format. Acute toxicity animal *in vivo* data relevant to the 97 reference chemicals were derived from published literature. Over 2200 LD_{50} values were found (Table 1), from studies of rodents (rat, mouse) and other mammals (e.g., guinea pig) including various administration routes (oral, intravenous, etc.). As available from individual studies, key attributes were extracted (i.e., species, strain and sex of animal, duration of exposure, route of administration, dose, volume applied) supplemented with clinical and necropsy reports as synoptic text. The information was compiled directly into AcutoxBase, an on-line database application available to project partners (Kinsner-Ovaskainen *et al.*, 2009).

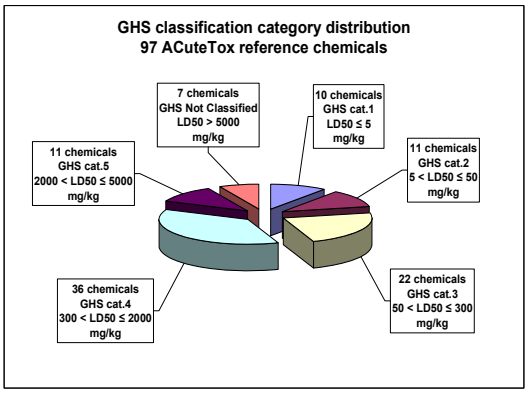

Figure 2a

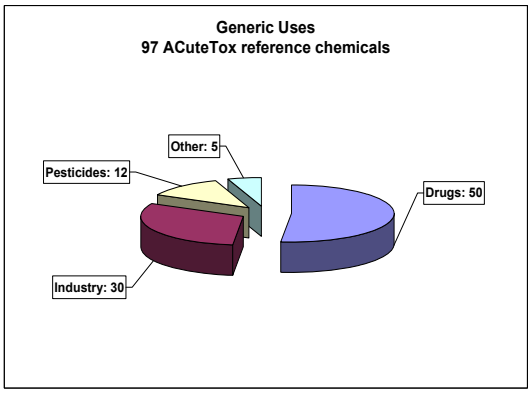

Figure 2b

Principal sources of LD_{50} data, supported by original references, were internet databases (e.g., ChemIDplus, HSDB, etc.) linked to the US National Library of Medicine Toxicology Data Network, TOXNET (http://toxnet.nlm.nih.gov/). Supplementary data sources included specific compendia (e.g., NICEATM/ICCVAM database (2006); Sax's handbook (Lewis 2004)).

Nearly 2800 human cases were compiled, comprising three categories of single dose/exposure acute poisoning, each with/without time information, *viz*: sub-lethal, lethal, and post-mortem (Table 1). The main sources of data were international journals, toxicology handbooks, MEIC monographs, Swedish Poison Information Centre, Göttingen (Germany) Poison Information Centre, Poisindex, Thomson Micromedex, and several on-line databases (e.g. HSDB, ChemPlus, RTECS etc.).

Table 1. Totals of animal studies and human cases compiled for 97 ACuteTox reference chemicals

Animal acute toxicity: Published experimental LD_{50} studies				**Human acute poisoning: Reported cases, (a) time related, (b) without time**					
Rat	Mouse	Other	Total LD_{50} studies	Sub-lethal (a)	Sub-lethal (b)	Lethal (a)	Lethal (b)	Post-mortem (a)	Post-mortem (b)
921	907	376	2204	908	377	436	372	215	472

Each animal *in vivo* study was allocated a reliability rank, graded on a scale of 1 to 4 (Klimisch *et al*., 1997). The concept is based on availability of supporting documentation on method conditions and experimental observations, assumed to be indicative of LD_{50} study quality. The four grades of reliability which have been defined are: ‘reliable without restrictions’ (grade 1); ‘reliable with restrictions’ (grade 2); ‘not reliable’ (grade 3); and ‘not assignable’ (grade 4). In practice, criteria corresponding to the grades are: (1) no deviations from test guideline; (2) close to guideline, with only minor deviations or details missing on method; (3) no necropsy data or clinical symptoms reported, and/or most details of method missing; (4) original publication or secondary reference (review article) not available, or no details on experiments reported, i.e., no criteria to judge LD_{50} validity. Compliance with guideline of statistical method for LD_{50} calculation was also noted, as available, for each *in vivo* study. Moreover, any erroneous quantities detected (e.g., average lethal doses quoted as LD_{50} values) were either omitted or corrected if available from a study report.

Availability of human cases was frequently limited by improbable occurrence of acute human exposure (i.e., accidental ingestion, suicidal overdose, etc.) and/or unknown clinical/forensic monitoring of a patient/victim with reported blood concentration measurements. In particular, eleven reference chemicals remain without any reported acute poisoning cases. The chemicals include nominations by other work packages, selected according to expediency of respective *in vitro* assay research interests (e.g., relating to target organ specific toxicity). Six of the chemicals (*viz*: phenanthrene, pyrene, 1,2,3,4-tetrachlorobenzene, pentachlorobenzene hexachlorobenzene, and benz(a)anthracene) were selected by WP5 specifically for biokinetic modelling, without priority consideration of potential for human exposure by acute poisoning and availability of corresponding blood analyses. Similarly, tert-butylhydroperoxide selected by WP4, is an industrial intermediate normally processed in controlled isolation. Two other chemicals without cases (*viz*: glycerol and sodium lauryl sulphate) typically occur in a variety of products with insignificant toxicity. No cases were found for cycloheximde. Although this chemical has therapeutic and pesticide applications, and is known to be toxic in humans, documented monitoring of blood analyses from acute poisoning was not obtainable. No cases

were found for tetracycline hydrochloride. The chemical is used as a mild therapeutic antibiotic, with low toxicity, were acute poisoning would be anomalous. Two other chemicals selected by WP7.1 (*viz*: dichlorvos and physostigmine) and where only one case each was found, are susceptible to degradation following ingestion/injection. Therefore, ready availability of blood concentration monitoring data would not be expected.

Data evaluation

Comprehensive statistical analyses of the *in vivo* animal data to evaluate variability and reliability, interspecies correlation, predictive capacities with regard to official acute oral toxicity categories, and deduction of performance criteria for *in vitro* methods were the responsibility of WP1. The analyses have also recently been published (Hoffmann *et al*., 2010). Analysis of the compiled human data, in relation to results from the *in vitro* testing has been published independently (Sjöström *et al*., 2008).

Statistical evaluation of animal in vivo data

Preliminary data reduction resulted in 504 individual rat LD_{50} studies from 62 chemicals and 300 individual mouse LD_{50} studies from 51 chemicals, eligible according to applicable selection criteria, for conservative statistical review.

Standard deviation of logarithmically transformed LD_{50} best described variability respective of a chemical, also independent of actual toxicity. Excluding chemicals cited with extreme data ranges (4 for rat, 6 for mouse) median standard deviation was estimated as ~0.2. For the majority of chemicals (rat: 74%; mouse: 61%) variability among rodent data generally followed a log-normal distribution, with 95% confidence interval (±2*SD) for log-transformed LD_{50} values of ~0.8, indicating good reproducibility.

The data reduction yielded 40 chemicals eligible for rat and mouse inter-species comparison of LD_{50} studies. Ordinary regression showed high correlation, supported by robust and weighted regression techniques, and limiting bias by taking account of intra-species variability. Coefficients of determination (R^2) were high, ranging between 0.8 and 0.9, with substance-specific differences only significant for two chemicals, warfarin and cycloheximide. For eight chemicals, sufficient data were available from at least one other species (guinea pig, hamster, rabbit: i.e., only one non-rodent) no substantial differences of LD_{50} were observed. This high inter-species correlation precludes inferences for performance criteria of *in vitro* predictive capacity.

ACuteTox compilation of human data has built on the MEIC dataset (Ekwall *et al* 1998), including extension to non-MEIC chemicals. Regression of human acute lethal doses from the MEIC study with rat oral LD_{50} data from AcutoxBase (possible for 30 eligible chemicals) resulted in a coherent correlation similar to the previous MEIC study, implying similar relevance of the current rat oral LD_{50} dataset.

Klimisch reliability (scores on scale of 1 to 4) rationalizes LD_{50} study quality by degree of supporting documentation available on method conditions and experimental observations. Considering all studies of defined reliability (scores 1, 2, 3) only ~20% of 43 eligible chemicals qualified as 'reliable', and only ~20% of these presented an extreme value

(maximum or minimum among the values available for a given chemical) confirming absence of any trend associating chemical LD_{50} range with presumed study quality rank. Significantly, exclusion of LD_{50} values with 'not assignable' reliability (score 4) reduced the number of duplicate values apparent in AcutoxBase from 54 out of 504 studies (~11%) to 11 out of 269 remaining studies (~4%) suggesting unintentional repetition of data by secondary citation of original primary sources.

For the 62 eligible chemicals, EU/GHS toxicity classifications/categories were allocated, corresponding to respective maximum and minimum LD_{50} values. At face value, 57 (92%) and 53 (85%) of the chemicals (EU and GHS, respectively) display individual ranges of LD_{50} limited to two adjacent classification categories. For 4 (6%) and 7 (11%) of the chemicals (EU and GHS, respectively) LD_{50} values span three different classifications. For 1 (1.6%) and 2 (3%) of the chemicals (EU and GHS, respectively) cited LD_{50} values allow scope for more than three different classifications.

Finally, LD_{50} variability has been related to toxicity classification thresholds for 95 chemicals where at least one rat or mouse LD_{50} was available. Modelling LD_{50} variability among the reference chemicals, employing estimated normal distribution parameters and translating the results to GHS classification categories, showed that ~54% of the chemicals would fall (with at least 90% probability) into only one GHS class. Another ~44% would fall (with 90% probability) within two adjacent classes. This partitioning provides a preliminary indication of applicable scope relating predictive capacity expected of any alternative test/testing strategy for the conventional *in vivo* acute toxicity test.