WP8 and WP9: Optimisation of the testing strategy (P4, P6, P16, P25, and P26) and Prevalidation (P4, P8, P9, P11, P13, P15, P16, P18, P19, P20, P27, P34, P35, subcontractor DKFZ)

The selection of the *in vitro*/*in silico* methods for the prevalidation exercise from the total number of assays performed in the ACuteTox project was based on an analysis of variability, repeatability and reproducibility of the single assays, as well as the assessment of preliminary predictive capacity using univariate and multivariate CART analyses.

The first step of the analysis consisted in the extraction of raw data from the database Acutoxbase (about 10,000 files), followed by a first data quality check done by visual inspection of response variability, especially of control responses. The next step of the analysis was the statistical evaluation of the concentration-response curves and the computation of a characteristic value (EC20, EC50 or LOEC) for every experiment performed (combination of a single assay and a single chemical (57 chemicals in total)). First, a one-way ANOVA, followed when appropriate, by a post hoc test (Dunnett contrast), was used to assess whether the chemical had any effect on response in a given assay. If an effect was found, a four parametric log-logistic model was used to calculate the respective EC20, EC50 or LOEC value, depending on the test method under consideration.
The assays were also compared with respect to their repeatability and reproducibility.
Furthermore, an overview on the bi-variate dependency between different assays was shown by pair wise scatter plots, and Pearson correlations were calculated for the summarized ECxx and LOEC values for all chemicals tested in all the assays.

The concentration-response recalculations, the assessment of within-assay variability and variability in comparison to other assays, as well as information about the dependence between assay results was discussed with ECVAM (P4) to exclude some assays from further analysis due to their insufficient performance, to lack of data for all 56 chemicals (not all chemicals tested in an assay) or to other reasons (e.g. many uncensored values obtained with an assay, failure to identify positive compounds).

Since the ultimate aim of the statistical analysis for ACuteTox is the classification of chemicals into the official acute oral toxicity categories (EU and GHS1), the Classification and Regression Trees (CART) was used as the classification algorithm of choice. After derivation of a CART tree, misclassification rate was estimated from the data. This misclassification rate is expected to be too optimistic because the tree is built on the basis of the same data. An unbiased estimate of the misclassification rate will be obtained from the application of the CART tree to independent data, such as the data of the challenging exercise (prevalidation) that involved testing of a new set of chemicals.

CART analyses were performed with untransformed summary characteristic values for the *in vitro* assays. In addition, for the neurotoxicity assays, a blood brain barrier (BBB) transformation was performed and another approach was used in which the recalculated ECxx and LOEC values from the *in vitro* assays were first transformed to rat LD50 values using the transformation suggested by WP5.

The overall analysis performed before the start of the prevalidation exercise resulted in the selection of 8 methods (Table 10) and the proposal of different combinations of the selected

assays (e.g. testing strategies). The proposed combinations were challenged with the new data generated during the prevalidation in order to identify the best performing strategy.

Table 10. The best performing in vitro assays, which were selected as candidates for the final tiered testing strategy, and further evaluated in the prevalidation phase.

Selected assay	**Target (workpackage involved)**
The neutral red uptake assay using the 3T3 fibroblast cell line (3T3/NRU)	General basal cytotoxicity (WP2)
The cytokine release assay using human whole blood (IL-1, IL-6, TNF-alpha)	Haemotoxicity (WP 4)
Cell differentiation in human cord blood-derived cells (CBC/CFU-GM)	Haemotoxicity (WP 4)
Gene expression (GFAP, HSP-32, MBP and NF-H) in primary rat brain aggregate cultures	Neurotoxicity (WP 7.1)
Uridine incorporation measuring the total mRNA synthesis in primary rat brain aggregate cultures	Neurotoxicity (WP 7.1)
Cytomic panel measuring oxidative stress (intracellular peroxidative activity, intracellular levels of superoxide anion, oxidized DNA base 8-oxoguanine) in HepG2, SH-SY5Y and A.704 cells	New endpoints (WP 4)
Cytomic panel for cytotoxicity screening (intracellular Ca^{2+} levels, mitochondrial membrane potential, plasma membrane potential) in HepG2, SH-SY5Y and A.704 cells	New endpoints (WP 4)
The MTT assay using primary rat hepatocytes	Metabolism (WP 6)
Kinetic parameters: volume of distribution, protein binding, clearance, and oral absorption (Caco-2 cells) for the estimation of the oral dose from the effective concentration observed *in vitro*	Biokinetics (WP5)
The estimation of compound passage through the blood-brain barrier using neuronal networks (for neurotoxicity assays)	Biokinetics (WP5)

The work performed during the prevalidation focused mainly on the assessment of the predictive capacity of the proposed tiered testing strategies and the identification of the combination of methods that gave the best prediction, in terms of classification of compounds in the official acute oral toxicity categories.

During this prevalidation study the methods identified as promising building blocks for the testing strategy on the basis of the so-call training set of compounds (57 compounds), were challenged under blind conditions with a new set of 32 test compounds. The use of reference compounds which were not included in the training set was essential in order to properly assess the predictive capacity of the tiered testing strategy.

Due to time constrains the assessment of transferability of the methods to a second independent (naïve) laboratory was not feasible, as it requires extensive training of the naïve

laboratory. Therefore, laboratories which were involved in the development and optimisation of the selected test methods, took part in the testing exercise.

The participant laboratories tested the coded chemicals in a pre-defined order according to the study plan and reported the data on a regular basis. Some delays in data reporting were encountered by P9, due to problems related with fitting the data in the assigned raw data templates and that they run a lot of experiments, and by partners involved in the kinetics assays (P34, P27, P15, P35), due to problems related to the analytical methods used for detection/measurement of compounds in the samples generated *in vitro*.

The evaluation of raw data was performed in the same manner as for the optimisation phase of the project. The highest numbers of censored summary values were obtained with the 17 cytomic panel endpoints (50-90%). However, when the minimal ECxx values were calculated (SH-SY5Y, HepG2, A704, oxidative stress panel, and cytotoxicity screening panel) the percentages of chemicals with censored values decreased considerably (3-13%).

In addition to the CART methodology used in the first phase of the project, Random Forest model was used for the classification task during the prevalidation. Two classification approaches were studied in detail: single step procedures and two steps tiered testing strategies. Based on the classification analysis carried out, 10 *in vitro* testing strategies have been selected, five using the full set of 89 compounds and another 5 ignoring in the analysis compounds with a logP value larger than 5. The strategy that uses the Random Forest model including 9 endpoints (7 assays) in a single step procedure, gave the best correct classification rate (69.26%) and resulted in only 2 compounds with underestimated toxicity (Brucine, Paraquat), as compared to the official acute toxicity classification.

However the overall results showed that the incorporation of additional endpoints did not improve significantly the outcome of the 3T3/NRU cytotoxicity assay alone in terms of classification of compounds for acute oral toxicity. Only compounds with LD_{50} > 2000 mg/kg are best classified while the other toxicity categories are misclassified.

These results lead us to question the scientific motivation for the current classification systems that are based on arbitrary cut-off for rat oral LD_{50} values used to estimate human acute oral toxicity and whether a revision of the GHS/EU CLP systems would be advisable.

As shown in WP7 a number of assays were identified that were able to flag compounds as neurotoxicants and nephrotoxicants, both in the training (56 compounds) and test sets (32 compounds). Therefore, those *in vitro* assays could be used to alert on tissue specific toxicity for compounds that are identified as toxic (predicted LD_{50} < 2000 mg/kg) with the 3T3/NRU assay.