JRC SCIENTIFIC AND POLICY REPORTS

An investigation into the use of computational and in vitro methods for acute systemic toxicity prediction

Hedvig Norlén,
Elisabet Berggren,
Maurice Whelan and
Andrew Worth

2012

Report EUR 25473 EN

Joint
Research
Centre

European Commission
Joint Research Centre
Institute for Health and Consumer Protection (IHCP)

Contact information
Andrew Worth
Address: Joint Research Centre, Via Enrico Fermi 2749, TP 202, 21027 Ispra (VA), Italy
E-mail: andrew.worth@ec.europa.eu
Tel.: +39 0332 78 9566
Fax: +39 0332 78 9963

http://ihcp.jrc.ec.europa.eu/
http://www.jrc.ec.europa.eu/

This publication is a Reference Report by the Joint Research Centre of the European Commission.

JRC72736
EUR 25473 EN

ISBN 978-92-79-26058-2 (print)
ISBN 978-92-79-21708-1 (pdf)

ISSN 1018-5593 (print)
ISSN 1831-9424 (online)

doi:10.2788/44144

Luxembourg: Publications Office of the European Union, 2012

Printed in Italy

Executive Summary

Information on the acute mammalian toxicity of chemicals, referring to the adverse effects caused by either a single exposure to a chemical substance or multiple exposures within 24 hours, is required under multiple pieces of EU legislation aimed at protecting consumers and workers. Presently all regulatory methods for determining acute oral toxicity are based on animal tests. In these tests, the acute lethal dose to 50% of the treated animals (LD_{50} value) is typically used as the basis for hazard assessment and regulatory classification. The most widely used classification scheme is the Globally Harmonized System of Classification and Labelling of Chemicals (GHS) system managed by the United Nations.

Due to animal welfare and cost considerations, alternatives to animal experiments are being sought, and regulatory frameworks are increasing providing an opportunity or obligation to use such methods. Most of these alternatives are based on in vitro test methods or computational models such as Quantitative Structure-Activity Relationships (QSARs). To date, most studies have focussed on the abilities of individual in vitro tests or QSAR models to predict reference data from acute toxicity tests in rodents, with relatively few attempts to explore the combined use of in vitro and computational methods.

In this study, we used a reference dataset of 180 compounds for which in vitro and in vivo data were already available from international validation studies in order to assess the abilities of five alternative approaches to predict acute oral toxicity. The in vitro data are considered to be of high quality, having been generated and quality controlled as part of the previous validation studies. The in vivo data showed considerable variability for some compounds, with about 20% of the compounds crossing two or three toxic classes. We included four QSAR models (ToxSuite, TOPKAT, TEST and ADMET Predictor), which were available in-house, and one in vitro method, the Neutral Red Uptake (NRU) basal cytotoxicity assay performed in a rodent cell line (BALB/3T3) and using the in vitro prediction model of Halle. We characterised the predictive performance of each alternative method when used alone (both for LD_{50} prediction and acute toxicity classification into three categories), as well as multiple test combinations (batteries) and stepwise testing strategies (for acute toxicity classification into three categories).

When used individually, the alternative methods showed an ability to predict LD_{50} with correlation coefficients in the range from 49% to 84%, and to classify into three toxicity groups with accuracies in the range from 41% to 72%. Among the QSAR models, the best performing models were ToxSuite and TEST, with correlation coefficients of approximately 80% in LD_{50} prediction, and accuracies of approximately 70% in acute toxicity classification. The in vitro 3T3 NRU method, based on the use of the Halle prediction model, had a correlation coefficient with LD_{50} of approximately 50%, and a classification accuracy of approximately 41%.

When the QSAR models were combined in batteries, the overall accuracies were between 62% and 74%. While these figures are not much higher than the individual QSAR models alone, the sensitivities for the different toxic classes were considerably higher. On the other hand, the differences between the specificities for the different toxic classes were relatively small.

When the alternative methods were used in a stepwise testing strategy the overall accuracy could reach 76%. Different test combinations could be used to optimise overall accuracy, sensitivity or specificity, according to the end-user's requirements.

On the basis of our results, we conclude that:

a) the variability in LD_{50} values has an impact on classification, which means that the use of average LD_{50} values as a reference standard has to be used with care.
b) The 3T3 NRU in vitro test, used with the prediction model of Halle, has a lower predictive performance than the QSAR models. It is possible, however, that the in vitro test has a broader domain of applicability compared to the QSARs. It would be useful to explore whether the predictive performance of the in vitro system could also be increased by using an alternative prediction model.
c) the overall accuracies for the test combinations (model batteries) or testing strategies are not much higher than the AMs used alone, but may be optimised in terms of overall predictivity, sensitivity and specificity according to the end-user’s requirements.

Further studies, based on more extensive and high quality datasets (e.g. as generated by High Throughput Screening), would be valuable in the search for optimal strategies for assessing acute toxicity.

An investigation into the use of computational and in vitro methods for acute systemic toxicity prediction

Hedvig Norlén, Elisabet Berggren, Maurice Whelan and Andrew Worth

Table of contents

1 Background and objective of study

1.1 Animal tests for acute oral toxicity

Acute toxicity describes the adverse effects caused by either a single exposure to a chemical substance or multiple exposures within 24 hours. The acute lethal dose to 50% of the treated animals (LD_{50} value) is the basis for the hazard assessment and classification of chemicals and is widely used for regulatory purposes. Presently all accepted methods for regulatory requirements for determining acute oral toxicity are based on animal (in vivo) tests. There are three approved in vivo tests which are modifications of the classical median lethal dose, LD_{50} test (the Organisation for Economic Cooperation and Development (OECD) Test Guideline (TG) 401 (OECD, 1987), which was deleted in 2002. The three modified tests defined in the OECD TGs: 420 Fixed Dose Procedure (FDP), 423 Acute Toxic Class Method (ATC), and 425 Up and Down Procedure (UDP) (OECD, 2001a,b,c) are utilizing the principles of reduction and refinement. These tests are sequential tests where the outcome of the previous step/dose determines the next dose to be tested and the number of animals used per test can then be considerably reduced (from 25 to a minimum of 5 animals per test). The FDP and the ATC identifies a lowest fixed dose causing evident toxicity and they provide estimated LD_{50} intervals whereas the UDP estimates a LD_{50} value.

There are requirements of acute oral toxicity testing for agrochemicals, biocides and also for industrial chemicals. The acute toxicity testing in animals of cosmetic ingredients and products has been banned in the EU since 2009 (EC, 2003). For food additives, flavourings, food-contact materials, pharmaceuticals, and veterinary medical products there are no obligations for oral toxicity testing in EU (Seidle et al., 2010).

1.2 The Globally Harmonised System of Classification and Labelling (GHS)

The Globally Harmonized System of Classification and Labelling of Chemicals (GHS) is an internationally established system, implemented by the United Nations (UN, 2007). It is designed to replace the various classification and labelling standards used in different countries by using consistent criteria for classification and labelling on a global level. Two of the main purposes of the GHS have been to reduce the need for testing and evaluation of chemicals and to facilitate international trade in chemicals whose hazards have been properly assessed and identified on an international basis.

The Classification, Labelling and Packaging (CLP) Regulation (EC No 1272/2008; EC, 2008) is the European Union regulation which aligns, since 2009, the European Union system of classification, labelling and packaging of chemical substances and mixtures to the GHS. It complements the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH) Regulation (EC No 1907/2006; EC, 2006) and replaces the system contained in the Dangerous Substances Directive (67/548/EEC; EC, 1967) and the Dangerous Preparations Directive (1999/45/EC; EC, 1999). For

oral acute toxicity, chemicals are classified in one of four toxicity categories based on their oral acute toxicity estimates, as illustrated in Figure 1.

Table 3.1.1

Acute toxicity hazard categories and acute toxicity estimates (ATE) defining the respective categories

Exposure Route	Category 1	Category 2	Category 3	Category 4
Oral (mg/kg body-weight) See Note (a)	ATE ≤ 5	5 < ATE ≤ 50	50 < ATE ≤ 300	300 < ATE ≤ 2 000

Table 3.1.3

Acute toxicity label elements

Classification	Category 1	Category 2	Category 3	Category 4
GHS Pictograms				
Signal Word	Danger	Danger	Danger	Warning
Hazard Statement: — Oral	H300: Fatal if swal-lowed	H300: Fatal if swal-lowed	H301: Toxic if swal-lowed	H302: Harmful if swal-lowed

Figure 1. The CLP Regulation (EC No 1272/2008) for oral acute toxicity and the four acute toxicity hazard categories.

1.3 Alternatives to animal testing and Integrated Testing Strategies

Due to animal welfare and cost considerations, alternatives to animal experiments are being sought, and regulatory frameworks are providing an increasing opportunity or obligation to use such methods. The modified LD_{50} tests are still debated among toxicologists, animal welfare organizations, legislators and the public primarily due to the ethics of using animals for experimental purposes and evaluating mortality as an endpoint. The two major alternatives to in vivo animal testing are the in silico and the in vitro methods.

According to the REACH chemicals legislation, Quantitative Structure-Activity Relationships (QSARs) can be used as alternatives to animal testing. QSAR models may identify chemical hazards and improve the safe use of chemicals. Laboratory testing may be avoided by using QSAR models to predict chemical effects directly from chemical structure and simulating adverse effects in cells, tissues, laboratory animals and the environment. Estimation of LD_{50} values presents some drawbacks when used for QSAR modelling. First, acute toxicity effects may result from a wide spectrum of biokinetic, cellular and molecular events. Converting the complex, whole-body phenomena related to acute toxicity into a simple number necessarily leads to a loss of information.

Second, available data are highly variable, having been generated by different laboratories, protocols, animal species and strains. This undermines the reliability and repeatability of acute toxicity measurements. These facts complicate the modelling process and may explain why there are relatively few QSAR models and expert systems for predicting oral acute toxicity, in comparison with other endpoints. An overview of the different QSAR models used in the assessment of acute systemic toxicity is given by Lapenna et al. (2010).

In vitro cytotoxicity methods have been evaluated as alternatives to the use of animals in toxicity testing over the past four decades. Many international projects have evaluated the relationship between in vitro cytotoxicity and acute in vivo toxicity. The results of three major projects; MEIC (Multicentre Evaluation of In Vitro Cytotoxicity, Clemedson et al., 1996), the Halle RC (Registry of Cytotoxicity, Halle, 2003) and the NICEATM/ECVAM international validation study (NICEATM-ICCVAM, 2006) have all shown a linear correlation of around 60-70% between in vitro IC_{50} cytotoxicity data and oral rat LD_{50} values. For an overview of the use of in vitro cytotoxicity assays to predict acute oral toxicity see (NICEATM-ICCVAM, 2006). The OECD has established a Guidance Document (GD No 129; OECD, 2010) based on the outcome from the NICEATM/ECVAM validation study that describes how to estimate starting doses for acute oral systemic toxicity tests by first conducting cytotoxicity tests.

In principle, QSARs or in vitro methods can be used as replacements for in vivo acute toxicity, tests, provided they are sufficiently validated. In practice, due to current limitations in predictive methods, both in silico and in vitro approaches are likely to be used in combination in the context of Integrated Testing Strategies (ITS), in order to replace, reduce or refine animal testing. The concept of ITS and its application to regulatory toxicology is discussed elsewhere (van Leeuwen et al., 2007; Bassan et al., 2008). More generally, there is a trend in predictive toxicology to develop a new paradigm of toxicological assessment ("Toxicology in the 21st Century") based on the integrated use of multiple methodologies, including computer-based modelling, high-throughput and high-content screening technologies (NRC, 2007**;** Collins et al., 2008**;** Dix et al., 2007).

1.4 Objective of study

In this study we investigate the predictive performances of five alternative approaches for the assessment of acute oral toxicity. We consider the ability of four QSAR models (ToxSuite, TOPKAT, TEST and ADMET Predictor) and one in vitro method (3T3 NRU using the prediction model of Halle) for prediction of LD_{50} values. The predictive performance of each method when used alone (both for LD_{50} prediction and acute toxicity classification into three categories), as well as multiple test combinations (batteries) and stepwise testing strategies (for acute toxicity classification into three categories are being calculated and compared. To assess the predictive performances of the alternative methods, a test set containing in vitro and in vivo data for 180 compounds is being considered.

2 Materials and methods

2.1 Acute Systemic Toxicity data set

The data set in this study originates from three large alternative method studies. The studies are the following:

(1) NICEATM/ECVAM international validation study (NICEATM-ICCVAM, 2006)
(2) ACuteTox project (Acutetox, 2010)
(3) ECVAM follow-up study (Kinsner-Ovaskainen et al., 2009)

In the first study, the National Toxcicology Program Interagency Centre for the Evaluation of Alternative Toxicological Methods (NICEATM) and the European Centre for the Validation of Alternative Methods (ECVAM) conducted a joint validation study during 2002-2005 of the Neutral Red Uptake (NRU) basal cytotoxicity assay performed in two standard cell systems: a human cell system (normal human keratinocytes, NHK), and a rodent cell system (BALB/3T3 cell line). The study involved 72 chemicals, 12 chemicals from each GHS toxicity category, including non-classified chemicals. Most of the chemicals were pharmaceuticals (35%), pesticides (22%), solvents (10%) or consumer/industrial products (5%). The results of this study showed that the overall accuracy of the 3T3 NRU test method for correctly predicting each of the GHS acute oral toxicity classification categories was only about 30%.

The second study, "An In Vitro Test Strategy for Predicting Human Acute Toxicity" (Acutetox), was an integrated project within the sixth framework programme during 2005-2010**.** The main objective was to develop an *in vitro* test strategy sufficiently robust and powerful to replace *in vivo* testing of acute toxicity. The study involved 97 chemicals tested in diverse in vitro assays. Also in this study most of the chemicals were pharmaceuticals (52%), industrial products (31 %) or pesticides (12%). The study showed that by using a Random Forests model with seven in vitro assays and nine endpoints the classification rate was 69% (of classifying the chemicals classified into the official acute toxicity categories).

The aim the third study was to assess the predictive capacity of a cytotoxicity test to determine if a test chemical correctly falls into one of the two categories, non-classified (LD_{50} > 2 000 mg/kg), or classified ($LD_{50} \leq$ 2000 mg/kg). The study involved 56 industrial chemicals tested with the NRU basal cytotoxicity assay performed in the rodent cell system (BALB/3T3 cell line) tested in three laboratories. The results (accuracy 64-67%, sensitivity 92-96% and specificity 40-44%) of this validation study showed that the 3T3 NRU test method can be regarded as a valuable test method to screen-out the negative test chemicals (unclassified) when the method is used as a first step in a tiered approach for acute oral toxicity testing.

Forty five chemicals are duplicates in studies (1) and (2), so the total number of chemicals is 180 in the complete data set (see Annex). 26 chemicals in the complete data set are inorganic compounds or salts. Inorganic and organometallic compounds, salts, and compound mixtures are often removed

prior to QSAR analysis because many software tools for calculating chemical descriptors are not suitable for these molecules, because molecular graphs for these substances are not defined.

In vivo data databases containing LD_{50} values had been produced for each study, so no new animal in vivo tests were performed. The chemicals selected in the studies were chosen to represent the complete range of *in vivo* acute oral toxicity ranges and are relevant with regard to human exposure potential. Principal sources of LD_{50} data, supported by original references, were internet databases, e.g. ChemIDplus and the Hazardous Substances Data Bank (HSDB). For our study only rodent (for 175 of 180 chemicals rat and f**o**r the other mouse) data was considered with the oral administration route (administration by gavage (stomach tube) was regarded as equivalent to oral). The data set is very heterogeneous when it comes to sample size for the compounds, as can be seen in Figure 2. The sample size varies from one to 28 observations per compound. About one fifth of the compounds (40 compounds) have only one observation. The majority of the compounds have three or fewer observations.

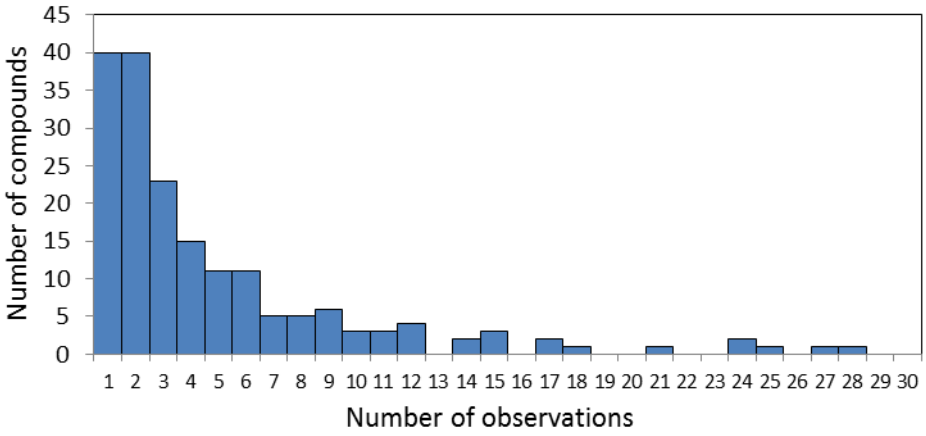


Figure 2. Sample size for Acute Systemic Toxicity data set.

2.2 Analysis based on average LD_{50} values

Figure 3 shows a scatter plot of the LD_{50} values for the 180 compounds. The compounds are ordered according to average LD_{50} (blue dots), starting with the most toxic (lowest LD_{50}) compound to the left to the least toxic (highest LD_{50}) compound to the right. The GHS category borders have also been inserted in the figure. More than one third of the compounds are crossing two GHS categories or even three GHS categories.

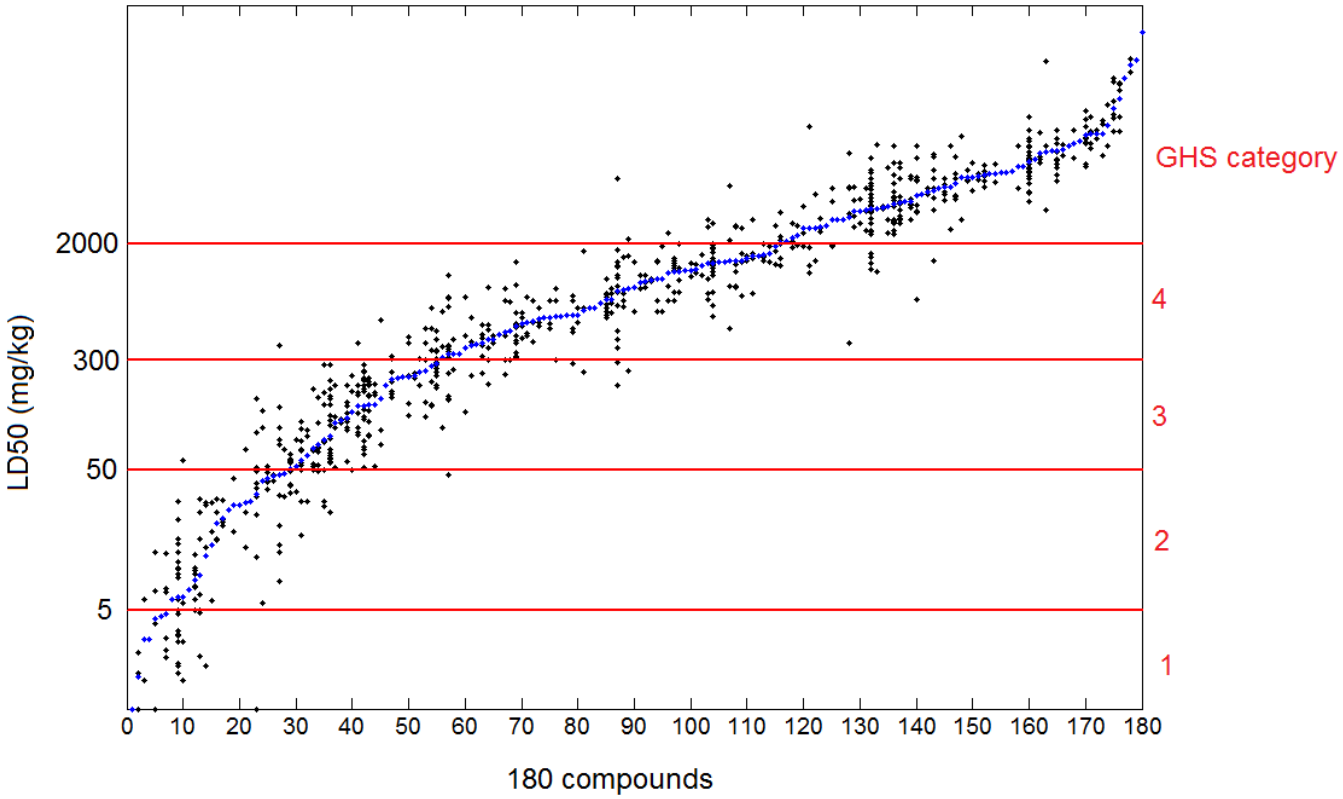


Figure 3. Scatter plot of the LD_{50} values for the 180 compounds.

The number of compounds based on average LD_{50} which fall in the different GHS categories is shown in Figure 4 below.

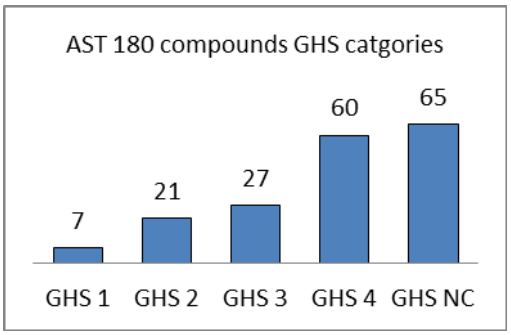
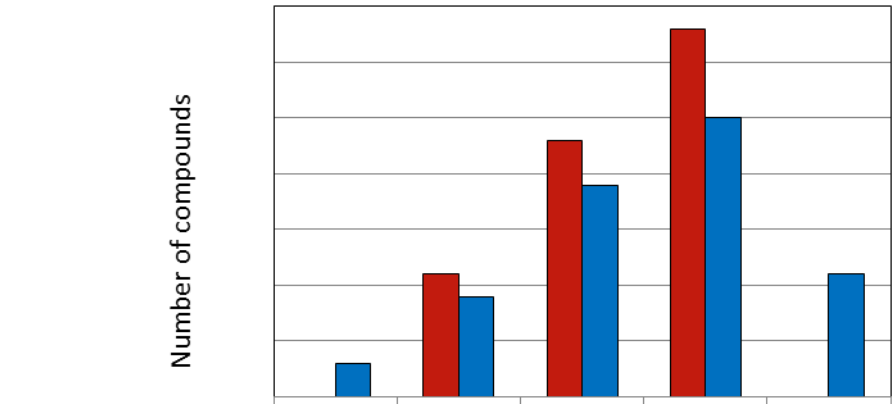


Figure 4. The number of compounds based on average LD_{50} which fall in the different GHS categories. GHS NC (non-classified) means compounds which have an LD_{50} above 2000 mg/kg.

2.3 Comparison of CLP and average LD_{50} classifications

67 of the 180 chemicals have official classifications for acute oral toxicity in Table 3.1, Annex VI to Regulation EC 1272/2008 CLP Regulation (EC, 2008). Another 27 chemicals are classified in the table according to acute dermal, acute inhalation toxicity, reproductive toxicity, aquatic toxicity or carcinogenicity.

Figure 5 shows how the 67 compounds are actually classified in the CLP Regulation compared with average LD_{50} classifications according to the GHS system. NC in the table stands for "non-classified" and refers to compounds which have an average LD_{50} > 2000 mg/kg.



	GHS 1	GHS 2	GHS 3	GHS 4	NC
actually classified	0	11	23	33	
average LD50	3	9	19	25	11

Figure 5. Comparison of actual CLP, blue bars and average LD_{50} classifications (red bars) for 67 chemicals.

43 of 67 compounds are classified in the same category. In 8 cases the average LD_{50} category is lower than the CLP Regulation classification. In 16 cases the average LD_{50} category is higher than the CLP Regulation classification. In these cases the difference in category is one, apart from two cases. Table I lists these two extreme cases.

Table I. Compounds where the CLP and average LD_{50} classifications in the GHS diverge with more than one category.

Chemical	JRC number	CAS number	CLP GHS classification	Average LD_{50} GHS classification
Methanol	JRC-000019	67-56-1	3	NC
Carbon tetrachloride	JRC-000257	56-23-5	3	NC

Methanol (JRC-000019, CAS number 67-56-1) is classified in the GHS category 3 in the CLP Regulation and in the NC (not classified) category for the average LD_{50} method. The sample of methanol LD_{50} values has 15 observations, with an average of 9591 mg/kg and a standard deviation of 2566 mg/kg. Also carbon tetrachloride (JRC-000257, CAS number 56-23-5) is classified in the GHS category 3 in the CLP Regulation and NC (not classified) category for the average LD_{50}. The sample of LD_{50} values contains 17 observations, with an average of 4219 mg/kg and a standard deviation of 2099 mg/kg. These average LD_{50} values differ greatly from the cut-off in GHS category 3, 300 mg/kg. Differences for these two compounds could be due to volatility.

2.4 Definition of Toxic Classes

We introduce our definition of Toxic Classes (TC), namely a grouping of the GHS categories 1 to 3 into TC1, GHS category 4 into TC2 and taking the compounds which are not classified in the GHS system (that are compounds which have LD_{50} above 2000 mg/kg) into TC3. The three classes contain roughly the same number of chemicals. 55 chemicals fall in the first class, 60 in the second and 65 chemicals in the third class (Figure 6).

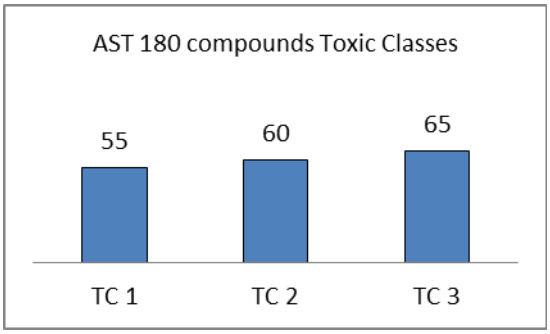


Figure 6. The number of compounds based on average LD_{50} which fall in the different TCs.

About 20% of the compounds cross two or three (in two cases) toxic classes. The compounds that cross three toxic classes are malathion (JRC-000012, CAS number 121-75-5, with 17 LD_{50} observations in the range from 200 to 5800 mg/kg bodyweight) and phenytoin (JRC-000028, CAS number 57-41-0, with 3 LD_{50} observations in the range from 250 to 2200 mg/kg bodyweight). Figure 7 shows a scatter plot of the LD_{50} values for the 180 compounds with the TC borders.

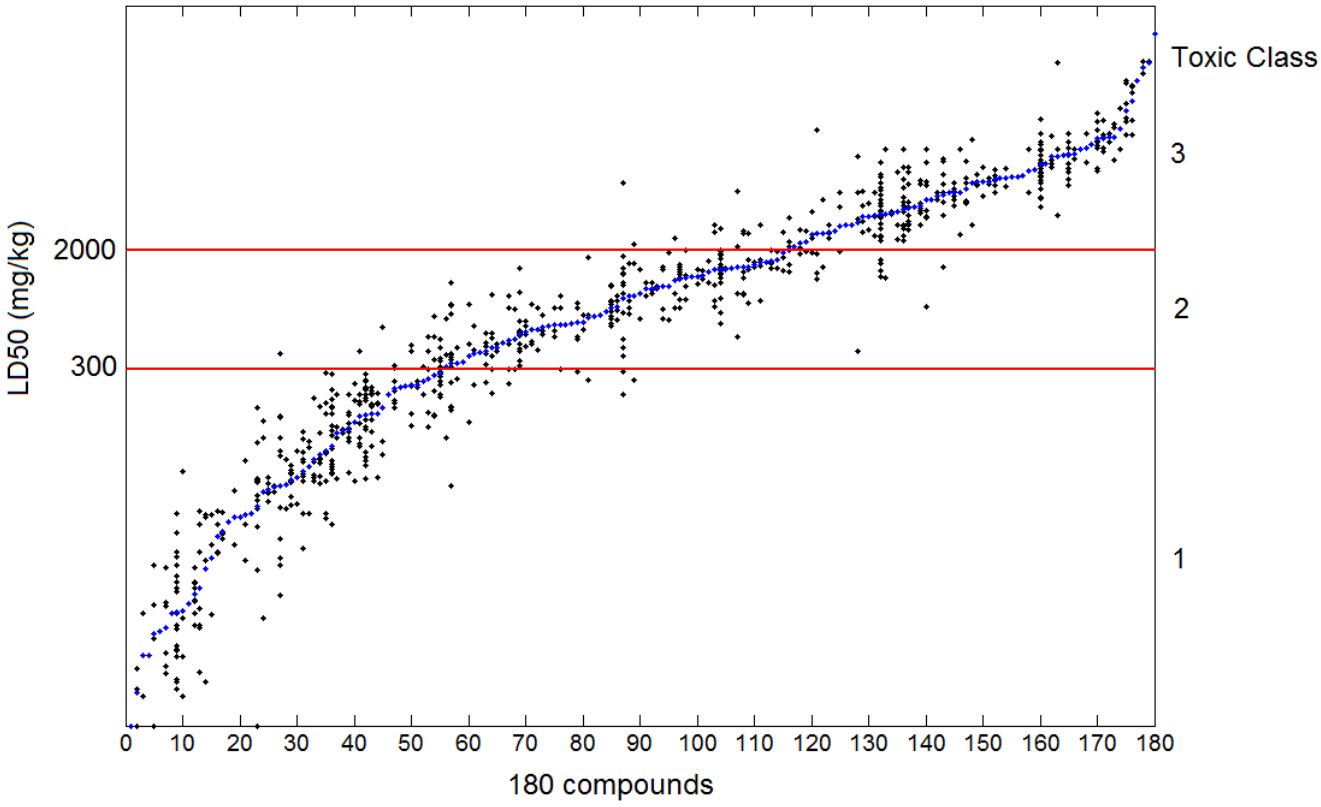


Figure 7. Scatter plot of the LD_{50} values for the 180 compounds with the TC borders.

2.5 Distribution of log LD_{50}

The distribution of log LD_{50} values is not symmetric but negatively (or left) skewed and the Generalised Extreme Value (GEV) distribution fits the data reasonably well. Figure 8 illustrates this, the empirical and the fitted GEV probability density functions to the left and the GEV probability plot to the right. The probability density function for the GEV distribution with location parameter μ, scale parameter σ, and shape parameter $k \neq 0$ is

$$y = f(x|k,\mu,\sigma) = \sigma^{-1} exp\left[-\left[1 + k\frac{(x-\mu)}{\sigma}\right]^{-\frac{1}{k}}\right]\left[1 + k\frac{(x-\mu)}{\sigma}\right]^{-1-\frac{1}{k}} \quad \text{(Eq. 1)}$$

For

$$1 + k\frac{(x-\mu)}{\sigma} > 0 \quad \text{(Eq. 2)}$$

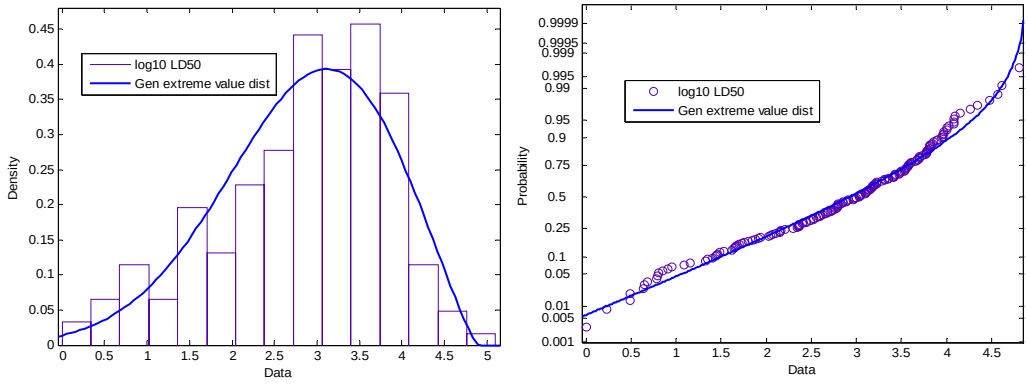


Figure 8. (a) left, shows the empirical and the fitted GEV probability density functions. (b) right, illustrates the GEV probability plot.

2.6 QSAR software models for predicting rodent oral acute toxicity

We have used four QSAR software tools to predict rat oral LD_{50}, three commercial solutions and one freely available solution. The four QSAR tools are:

1. ACD/Tox Suite, version 2.95
2. Accelrys/TOPKAT (Toxicity Prediction by Komputer Assisted Technology), version 6.2
3. U.S. EPA/ T.E.S.T. (Toxicity Estimation Software Tool), version 4.0
4. ADMET Predictor, version 5.5

The first software, ACD/Tox Suite (TOXS), predicts potential toxicity (LD_{50}) in two species (mouse and rat) for various administration routes (intraperitoneal, intravenous, subcutaneous, and oral administrations). The module "Acute Toxicity (LD_{50}, Rat/Oral)" has been used. Predictions are based on a combination of expert knowledge of various basal and extra-cellular effects (e.g., inhibition of cholinesterase and ATP synthesis, CNS and PNS disruption), and SAR/QSAR analysis of more than 100,000 compounds. Predictions are provided with reliability estimations (reliability index). Compounds are classified into one of 5 toxicity categories (corresponding to the GHS categories).

The second software, Accelrys/TOPKAT (TOPK), assesses the toxicity of chemicals solely from their 2D molecular structures and uses a range of robust, cross-validated Quantitative Structure-Toxicity Relationship (QSTR) models for assessing specific toxicological endpoints. The model "Rat Oral LD_{50}" has been used.

The freely available software U.S. EPA/ T.E.S.T. (TEST) includes models for estimating toxicity for several endpoints using different QSAR methodologies. The model "Oral rat 50% lethal dose" was applied for this comparison study.

The fourth software, the ADMET (Absorption, Distribution, Metabolism, Elimination and Toxicity) Predictor (ADMET) can be used for predictive modeling of ADMET properties. The toxicity module "Acute toxicity in rats" has been applied. These software tools are further described in Lapenna et al. (2010).

2.7 An in vitro assay for predicting rodent oral acute toxicity

From the three alternative method studies (Section 2.1 above), in vitro data have been generated from the Neutral Red Uptake (NRU) cytotoxicity assay. It is a cell survival/viability chemosensitivity test based on the ability of viable cells to incorporate and bind the supravital dye neutral red (NICEATM-ICCVAM, 2006). One of the cell models used in the studies was the CCL-163, 3T3 BALB/c immortalised mouse fibroblast, cell line, clone 31 from the American Type Culture Collection (ATCC, Manassas, VA., USA). The in vitro assay and cell model will be referred to the 3T3 NRU test method.

For 7 out of the 180 chemicals the 3T3 NRU test method did not yield the result of an IC_{50} value, mainly because of solubility problems. The problematic chemicals were the inorganic compound ferrous sulfate (JRC-000018, CAS number 7720-78-7), pentobarbital sodium (JRC-000030, CAS number 57-33-0), diphenhydramine (JRC-000033, CAS number 58-73-1), 1,3,5-trioxane, 2,4,6-trimethyl-, paraldehyde (JRC-000038, CAS number 123-63-7), acrylamide (JRC-000040, CAS number 79-06-1), acetaldehyde (JRC-000042, CAS number 75-07-0) and carbon tetrachloride (JRC-000257, CAS number 56-23-5).

From Study 1 (NICEATM-ICCVAM, 2006) the in vitro data originate from three laboratories: the U.S. Army Edgewood Chemical Biological Center (ECBC, USA), the Institute for In Vitro Sciences (IIVS, USA) and the FRAME Alternatives Laboratory (FAL, UK). The in vitro data in Study 2 (Acutetox, 2010) come from three laboratories: FAL, the Advanced In Vitro Cell Technologies, (Advancell, Spain) and from our robotic HTS laboratory (IHCP, JRC, Italy). From Study 3 (Kinsner-Ovaskainen et al., 2009) the in vitro data come from the Health and Safety Laboratory (HSL, UK), the IIVS and from the JRC-IHCP HTS laboratory.

Figure 9 shows a scatter plot of the IC_{50} values for the 173 compounds from the three studies. The compounds are ordered according to average IC_{50} (blue dots), starting with the most cytotoxic (lowest IC_{50}) chemical to the left to the least cytotoxic (highest IC_{50}) compound to the right.

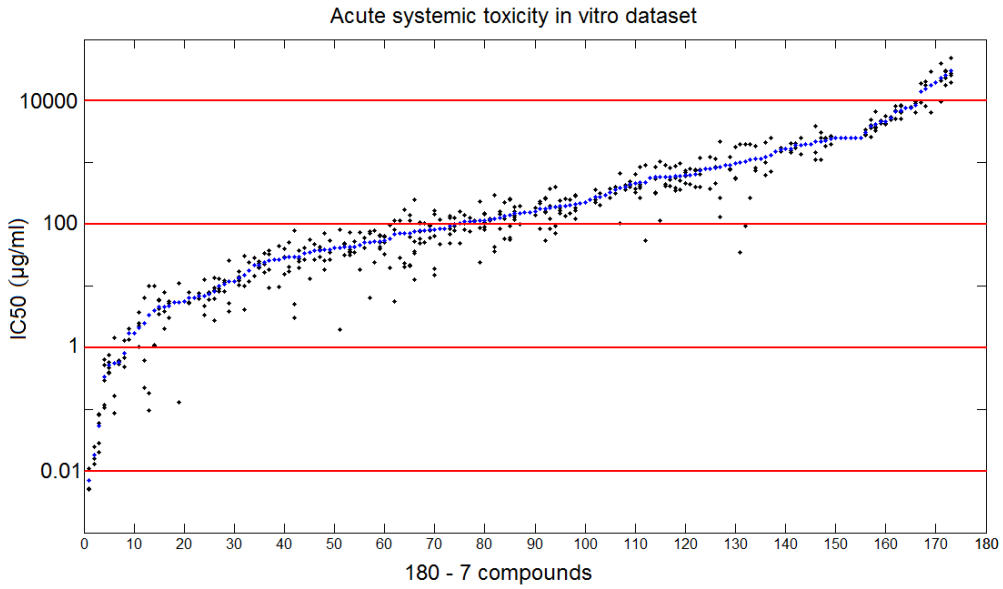


Figure 9. Scatter plot of IC_{50} values for the 173 compounds.

The conclusion from the NICEATM/ECVAM Validation Study was that the 3T3 NRU test method was not good enough for classifying the test compounds in the different hazard categories but could be used to aid in setting the starting dose for sequential rodent acute oral toxicity test methods.

2.8 Evaluation of the predictive capacity of an alternative methods

The predictive capacity of an alternative method (a QSAR software prediction or an in vitro assay) is frequently expressed by comparing the results obtained with the alternative method being tested on a sample of chemicals with the results of reference values on the same chemicals where the reference test results are taken as the true values. We assume that our true reference values are the rodent oral acute toxicity data, the average LD_{50} values.

In the simple case when the test method being evaluated and reference values are expressed with a binary outcome (positive or negative), the result of the test method study can be displayed in a 2 x 2 contingency table whose columns represent the reference results and whose rows represent the test method results. The 2 x 2 contingency table displaying the results of test method studies are often summarised by various characteristics of the test method and the population of chemicals. In medical and toxicological applications they are sometimes referred to as “Cooper statistics” (Cooper et. al, 1979). In our case, we have an extension of the binary classification since we have three toxic classes (TC1, TC2 and TC3) and hence 3 x 3 contingency tables, see Table II and Table III for the definitions of three main Cooper statistics.

Table II. A 3 x 3 contingency table

		In vivo Ref Class			Row
		TC1	TC2	TC3	Totals
Predicted	TC1	a	b	c	nR1
Class	TC2	d	e	f	nR2
	TC3	g	h	i	nR3
Column Totals		nC1	nC2	nC3	n
					Grand Total

Table III. Definitions of three main Cooper statistics for 3 x 3 contingency tables

Statistic	Definition	Calculation
Accuracy (concordance)	The proportion of chemicals that the test method classifies correctly	(a+e+i)/n
Sensitivity for TC1 chemicals	The proportion of chemicals that are classed in TC1 in vivo which the alternative method predicts to be in TC1	a/nC1
Specificity for TC1 chemicals	The proportion of chemicals that are classed in TC2 or T 3 in vivo which the alternative method predicts to be in TC2 or TC3	(e+h+f+i)/(nC2+nC3)
For TC2 and TC3 chemicals the calculations are made in corresponding way		

Sensitivity and specificity can also be easily calculated with Bayes' theorem (also called the inverse probability law). The sensitivity for TC1 chemicals is the probability that the test method predicts chemicals to be in TC1 given that the chemicals are really in that class.

$$P(TC1\,AM|TC1\,REF) = \frac{P(TC1\,REF|TC1\,AM)P(TC1\,AM)}{P(TC1\,REF)} \qquad \text{(Eq. 3)}$$

where AM stands for alternative method class and REF stands for reference class. The specificity for TC1 chemicals is the probability that the software program predicts chemicals to be in TC2 or TC3 given that the chemicals are really in those classes.

$$P(TC2\,AM \cap TC3\,AM \mid TC2\,REF \cap TC3\,REF) = \ldots$$

$$\ldots \frac{P(TC2\,REF \cap TC3\,REF \mid TC2\,AM \cap TC3\,AM)P(TC2AM \cap TC3\,AM)}{P(TC2\,REF \cap TC3\,REF)} \qquad \text{(Eq. 4)}$$

For TC2 and TC3 chemicals the calculations are made in corresponding way.

3 Results

3.1 QSAR software prediction results

In the subsequent analysis the 26 inorganic compounds and salts have been removed from the complete data set. The total number of compounds in the remaining data set is thus 180-26=154. Figure 10 illustrates the scatter plots of the LD_{50} values in the data set and the results from the QSAR software predictions. There seems to be a trend for all four QSAR software that, for more toxic chemicals (for TC1 chemicals) the programs overestimate the LD_{50} (hence underestimate the toxicity) and for less toxic chemicals (for TC3 chemicals) the programs underestimate the LD_{50} (overestimate the toxicity). For the compounds in between there is no trend.

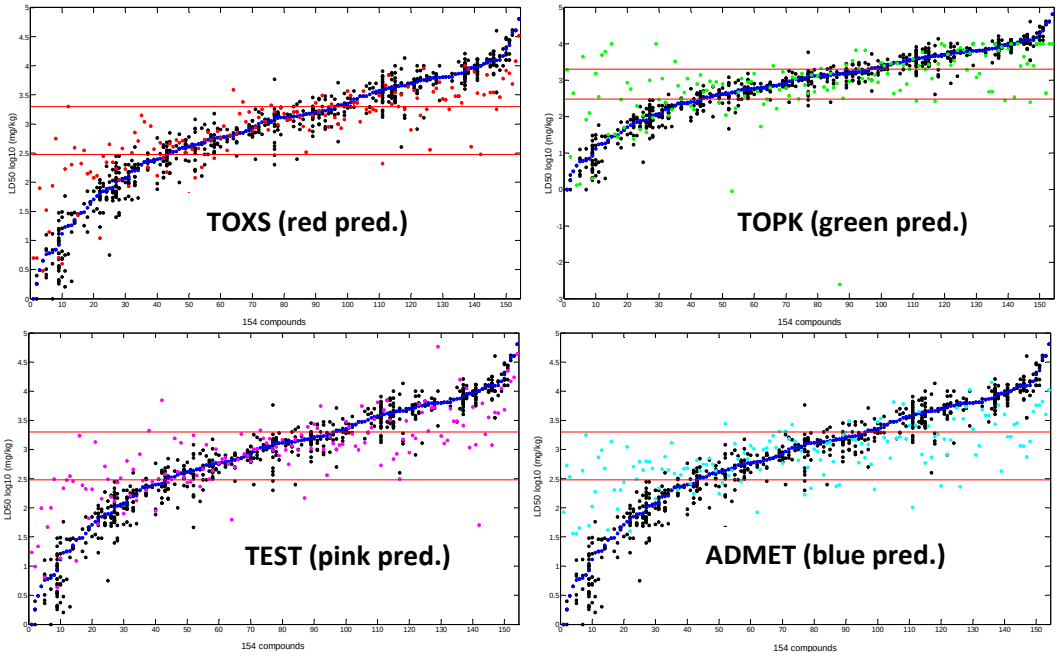


Figure 10. Scatter plots of LD_{50} values and the QSAR software predictions.

A nonparametric statistical method is applied to look at these differences. Few assumptions about the form of the distributions are made. The three TCs will be tested separately and the sample sizes are 43 compounds in TC1, 53 compounds in TC2 and 58 compounds in TC3 taken from the reference LD_{50} data set. We assume that the data from the ith QSAR software form a random sample from a continuous cumulative distribution function F_i, $i = 1, \ldots, 5$ (including the reference LD_{50} data set) and the random samples are mutually independent. Therefore, the following null and alternative hypotheses to be tested are:

H_0: $F_1 = F_2 = \ldots = F_5$ vs. H_1: $F_i < F_j$ for some $i \neq j$. (Eq. 5)

The ranking test called Kruskal-Wallis is applied for each of the three TCs and show the difference between the methods is significant for 1 and TC3 but not for TC2. This fact can also be seen in Figure 11, which depicts the boxplots of the LD_{50} distributions for the reference data and the four

QSAR software programs for each toxic class. For the distributions of TC1 and TC3, all pairwise differences between the reference LD_{50} and the 4 QSAR software are significant (using a method to control Type I family wise error (FWE) rate) but not between all QSAR-methods.

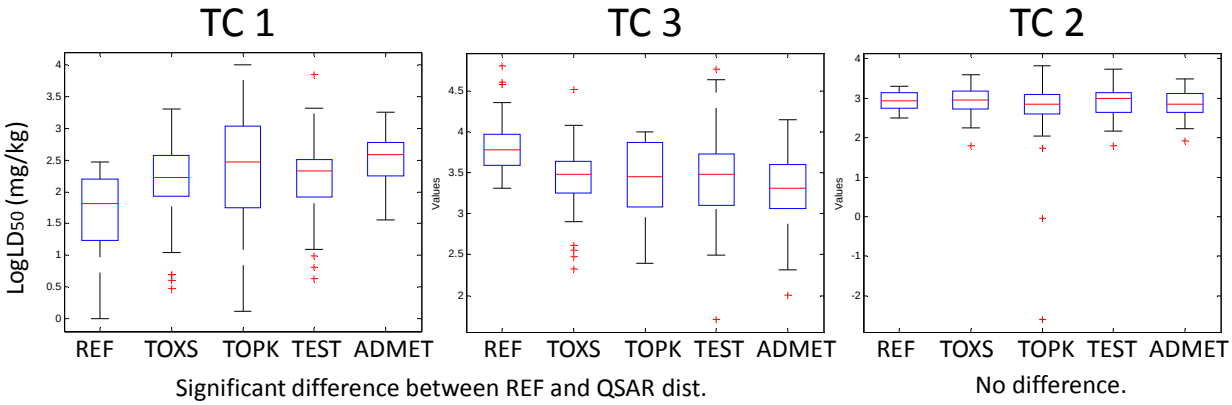


Figure 11. Boxplots of the LD_{50} distributions for the reference data (REF) and the 4 QSAR software programs (TOXS, TOPK, TEST, ADMET) for each TC (1, 3 and 2). The difference between the methods (REF and QSAR software) is significant for TC1 and TC3 but not for TC2.

Figure 12 illustrates scatter plots of the reference LD_{50} values versus the 4 QSAR software predictions on a 10-log scale. The strongest linear relationship (the Pearson correlation coefficient is $\rho = 0.84$) between the predictions of the TOXS software and the reference LD_{50} values. The predictions of TOPK and the reference LD_{50} values have the lowest correlation (ρ=0.49).

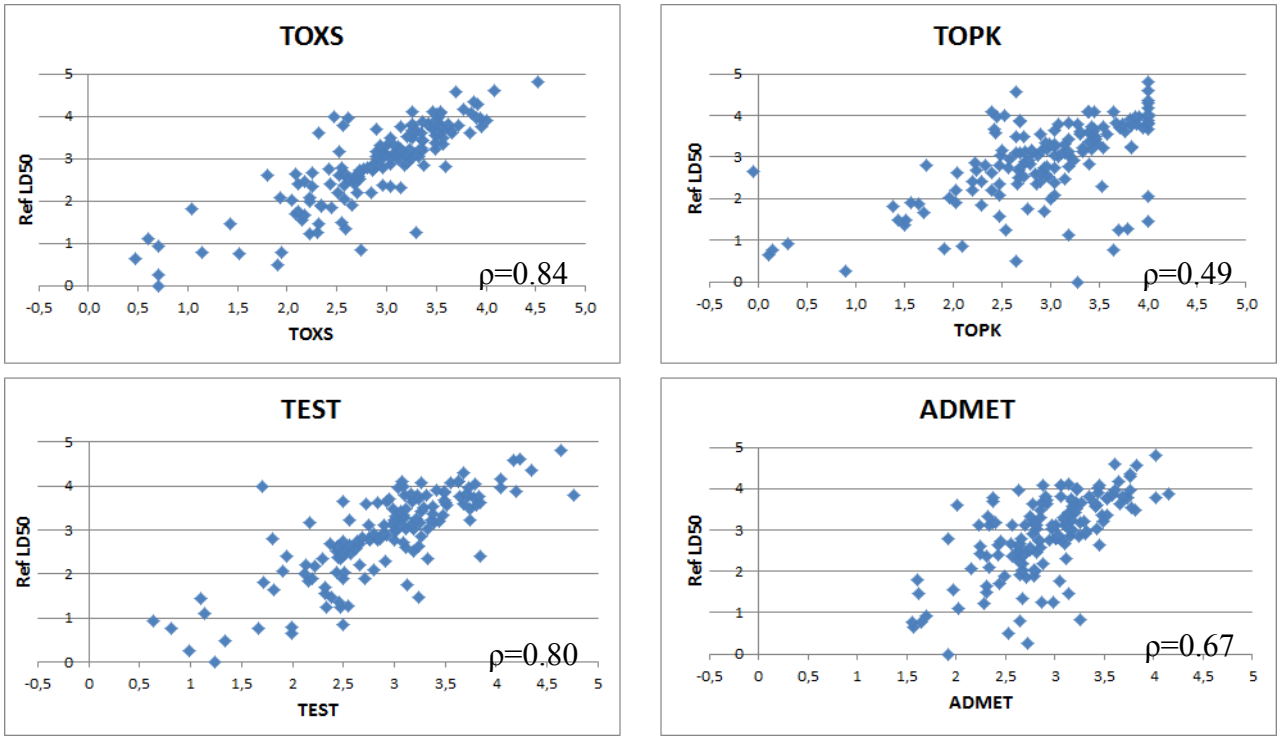


Figure 12. Scatter plots of the predicted LD_{50} values from the 4 QSAR software tools versus the reference LD_{50} values on a 10-log scale.

3.2 In vitro assay prediction results

To obtain the predicted LD_{50} values from the IC_{50} values (µg/ml) from the in vitro test experiments, the regression models from the Halle RC (RC reg) were used:

Millimole regression model: $\log LD_{50}$ (mmol/kg) = 0.439· $\log IC_{50}$ (mM) + 0.621 (Eq. 6)

Weight regression model: $\log LD_{50}$ (mg/kg) = 0.372 ·$\log IC_{50}$ (µg/mL) + 2.024 (Eq. 7)

Figure 13 illustrates a scatter plot of the predicted LD_{50} values from the weight regression model (Eq. 7) versus the reference LD_{50} values on a 10-log scale. Most compounds in the regression model are classified in the middle toxic class and few compounds are classified in the most and least toxic classes. The correlation coefficient is 0.53.

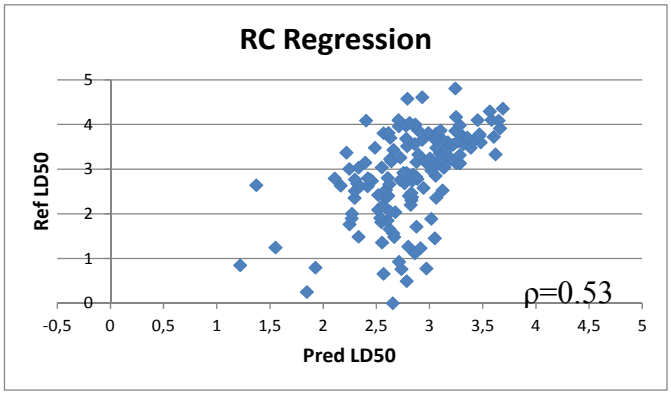


Figure 13. Scatter plots of the predicted LD_{50} values versus the reference LD_{50} values on a 10-log scale.

3.3 Evaluation of the predictive capacity of an alternative test method

The heat map in Figure 14 gives a graphical representation of the five alternative test method predictions, in columns 2 - 6, compared to the reference LD_{50} values, in column 1. The compounds are ordered in decreasing toxicity of the reference values. The compounds that are classified in TC1 are in red, TC2 compounds in yellow and TC3 compounds in green. The software TOPK (column 3) was not able to predict a LD_{50} value for the compound fentin hydroxide (JRC-000241, CAS number 76-87-9) and this compound is therefore grey coloured. For the RC reg predictions (column 6) there are six grey compounds corresponding to the six inorganic problematic chemicals listed in Section 2.7. From an overview of these results we make some observations:

- the worst classification error of predicting TC1 chemicals in the TC3 is made by the TOPK software six times and by the TEST software twice,
- the TOXS and TEST software tools seem to yield similar and the best results (apart from the two TC1 errors),
- the ADMET software and the RC reg tend to predict chemicals in class TC2.

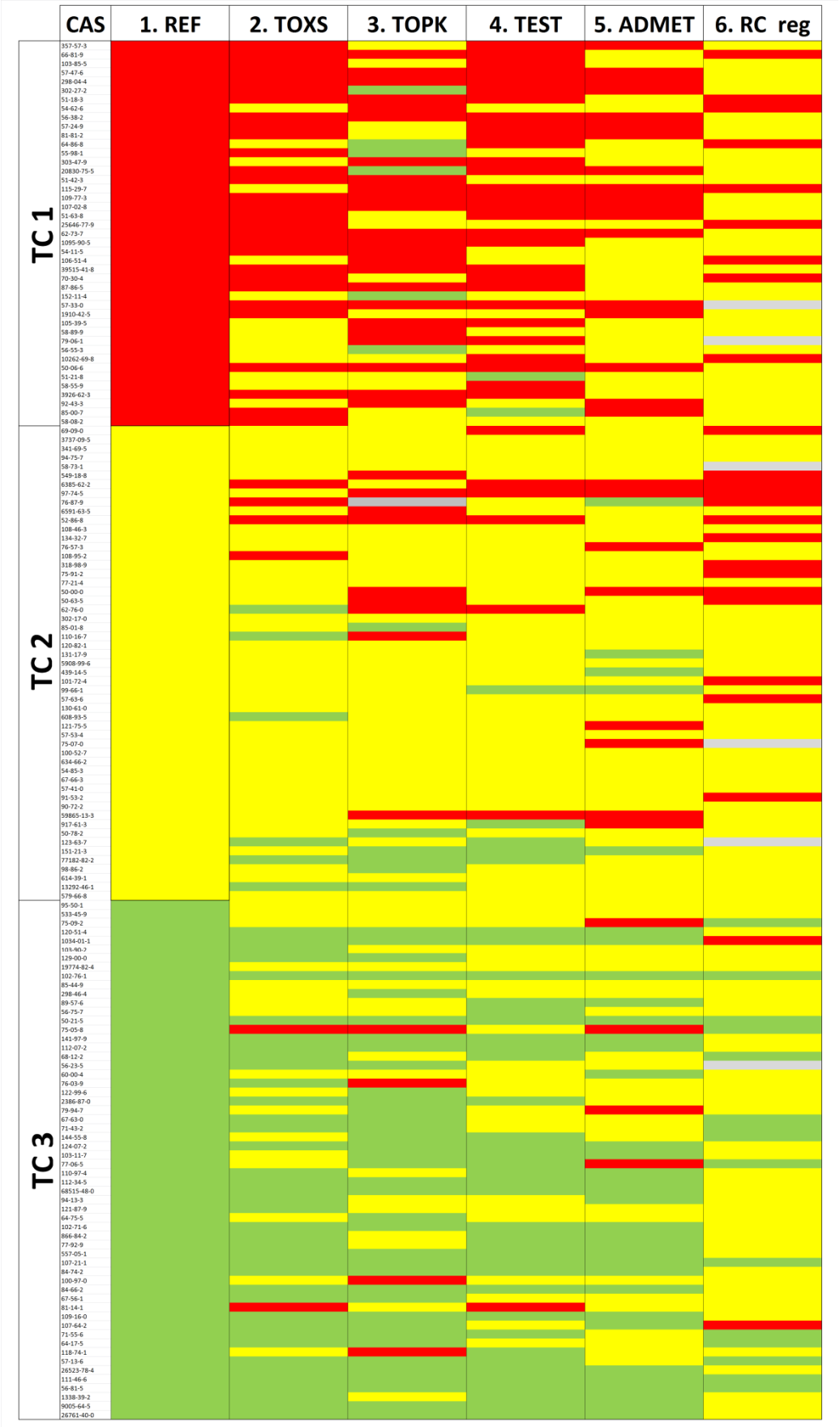


Figure 14. Heat map of the alternative method predictions (columns 2-6) compared to the reference LD_{50} values (column 1).

We also calculated the accuracy, sensitivity and specificity for the five alternative test methods. Table IV and Table V go through the calculations of these statistics for one of the methods, the TOXS software (column 2). The accuracy is relatively high, 72%. The sensitivity for TC2 is higher that the sensitivities for TC1 and TC3 which are almost the same (81% compared to 67% in the other two cases). As a consequence, the specificities for TC1 and TC3 are almost the same and very high (95% and 94%) and larger than the specificity for TC2 (69%).

Table IV. The 3 x 3 contingency table with the results from the TOXS software compared to the reference LD_{50} values.

Software TOXS					
		In vivo Ref Class			Row
		TC1	TC2	TC3	Totals
Predicted	TC1	29	4	2	35
Class	TC2	14	43	17	74
	TC3	0	6	39	45
Column Totals		43	53	58	154
					Grand Total

Table V. Calculations of the accuracy, sensitivity and specificity for the TOXS software.

Accuracy		(29+43+39)/154 ≈ 72.1%
Sensitivity for	**TC1**	29/43 ≈ 67.4%
	TC2	43/53 ≈ 81.1%
	TC3	39/58 ≈ 67.2%
Specificity for	**TC1**	(43+6+17+39)/(53+58) ≈ 94.6%
	TC2	(29+0+2+39)/(43+58) ≈ 69.3%
	TC3	(29+14+4+43)/(43+53) ≈ 93.8%

Table VI gives an overview of the predictive capacities of the five alternative methods. TOXS has the highest accuracy, which can also be seen in the heat map in Figure 14. It also had the highest correlation coefficient for the individual predicted values, as can be seen in section 3.1. TEST has the second highest accuracy (68%), followed by TOPK (63%) and ADMET (57%). The RC reg model has the lowest accuracy, only 41%.

Table VI. Overview of the predictive capacity for the five alternative methods**.**

Alternative Test Method		TOXS	TOPK	TEST	ADMET	RC reg
Accuracy		72.1%	63.4%	68.2%	56.5%	40.5%
Sensitivity for	TC1	67.4%	55.8%	67.4%	41.9%	22.0%
	TC2	81.1%	69.2%	79.2%	75.5%	72.0%
	TC3	67.2%	63.8%	58.6%	50.0%	26.3%
Specificity for	TC1	94.6%	88.2%	93.7%	89.2%	85.0%
	TC2	69.3%	70.3%	65.3%	50.5%	26.5%
	TC3	93.8%	86.3%	92.7%	94.8%	100.0%

Figure 15 illustrates the sensitivities and specificities for the five alternative test methods:

- for all methods the sensitivities for TC2 are higher than for TC1 and TC3. Thus the specificities for TC2 for all methods are lower than for TC1 and TC3,
- the Cooper statistics for TOXS and TEST are similar,
- for the RC reg the specificity for TC3 is 100%, meaning that there are no misclassified TC3 chemicals in the groups TC1 and TC2, but the sensitivity for TC3 is only 26%.

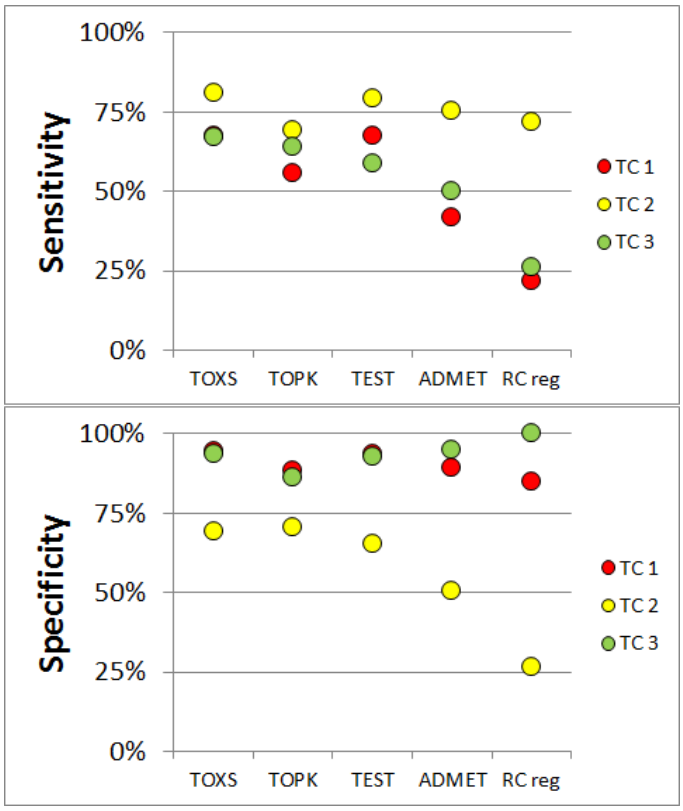


Figure 15. Sensitivity and specificity for the five alternative test methods**.**

3.4 Combination of QSAR predictions

In this section we combine the predictions from the different QSAR models and use the information to draw conclusions from a combined testing scheme. Ten different test batteries were chosen, as described in Table VII.

Table VII. Descriptions of the different ways of combining QSAR predictions (test combinations)

Abbr	Description
C1	The most conservative classification (TC) was picked among the four software tools.
C2	The most conservative classification (TC) was picked among the three software; TOXS, TOPK and TEST.
C3	The most common classification (TC) was picked among the four software tools. When the result is ambiguous (two classifications in two TC), the result is left blank.
C4	The most common classification (TC) was picked among the four software tools. When the result is ambiguous (two classifications in two TC), the most conservative classification (TC) was chosen.
C5	The most common classification (TC) was picked among the three software tools; TOXS, TOPK and TEST. When the result is ambiguous (one classification in each TC), the result is left blank.
C6	The most common classification (TC) was picked among the three software tools; TOXS, TOPK and TEST. When the result is ambiguous (one classification in the each TC), the most conservative classification (TC) was chosen.
C7	For TOXS, TOPK and TEST, an average LD_{50} was calculated and the TC assigned thereafter.
C8	For TOXS and TEST, an average LD_{50} was calculated and the TC assigned thereafter.
C9	The most conservative classification (TC) was picked among the two software TOXS and TEST.
C10	The least conservative classification (TC) was picked among the two software TOXS and TEST.

Figure 16. Accuracy, sensitivity and specificity for ten different test combinations (C1-C10)**.**

A heat map of the results is given in Figure 17.

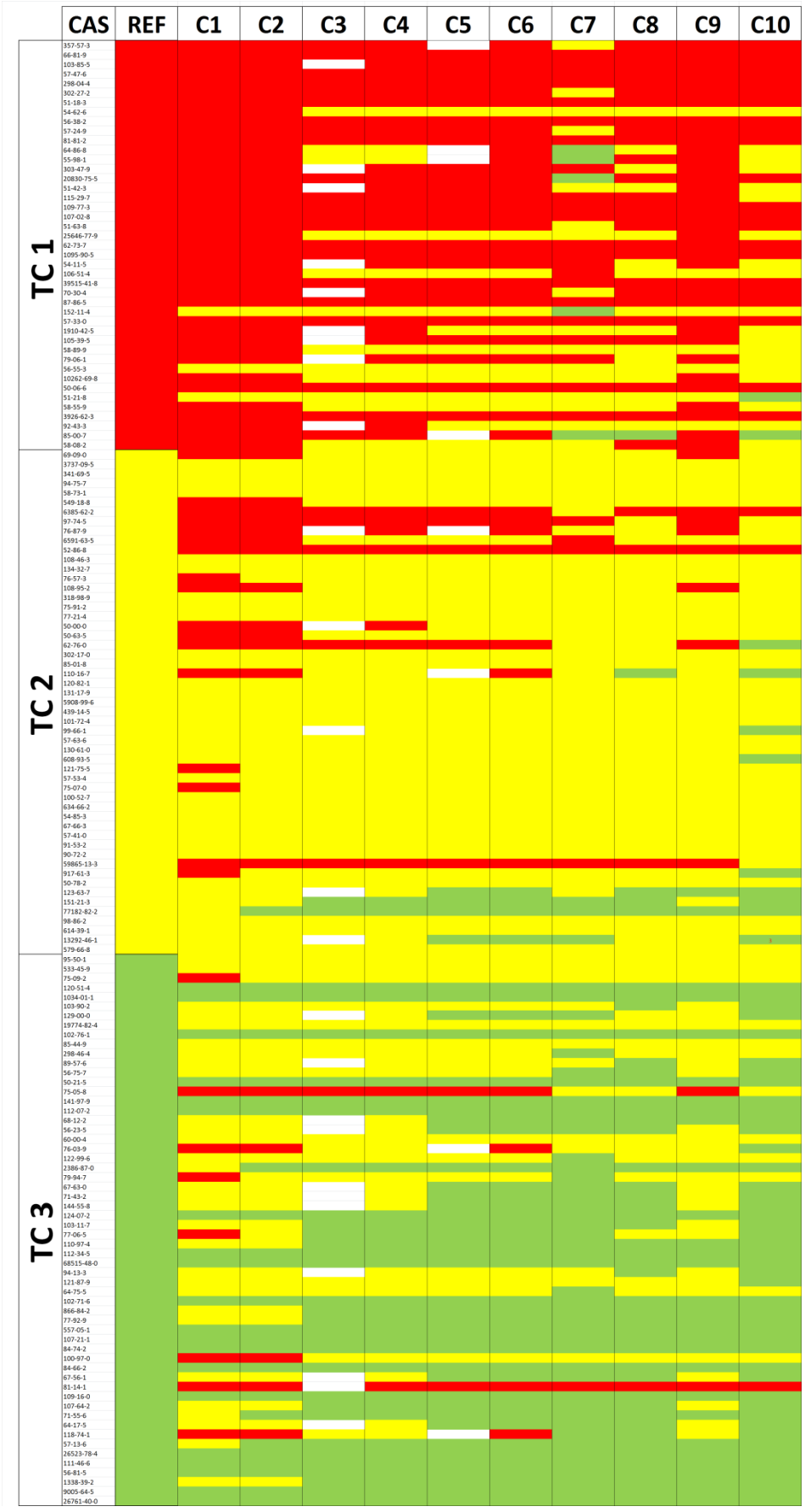


Figure 17. Heat map of the test combinations (C1-C10) compared to the reference LD_{50} values (REF). For the test combinations C3 and C5, 25 and 8 ambiguous (inconclusive) results were found (white compounds), thus the adjusted number of chemicals in these test sets is 131 and 148.

Figure 18 depicts a scatter plot of the accuracy, sensitivity and specificity for the ten combinations of alternative methods. The complete list of statistics are given in Table VII. The accuracy lies between 62-74%. The test combinations C1 and C2 are good at identifying TC1 chemicals (the sensitivities for TC1 chemicals are 93% in both cases), since they are conservative methods. The trade-off is that they are bad at identifying low toxicity chemicals (the sensitivities for TC3 chemicals are only 35% and 40%). A comfort is that when these test batteries pick out TC3 chemicals we can be sure they really are low toxicity chemicals (the specificities for TC3 chemicals are 100% and 99%, respectively). The test combinations C7 and C8 are good at identifying TC2 chemicals (the sensitivities are in both cases 87%). The test combination C10 is good at identifying TC3 chemicals with a sensitivity and specificity of 78% and 89%, respectively.

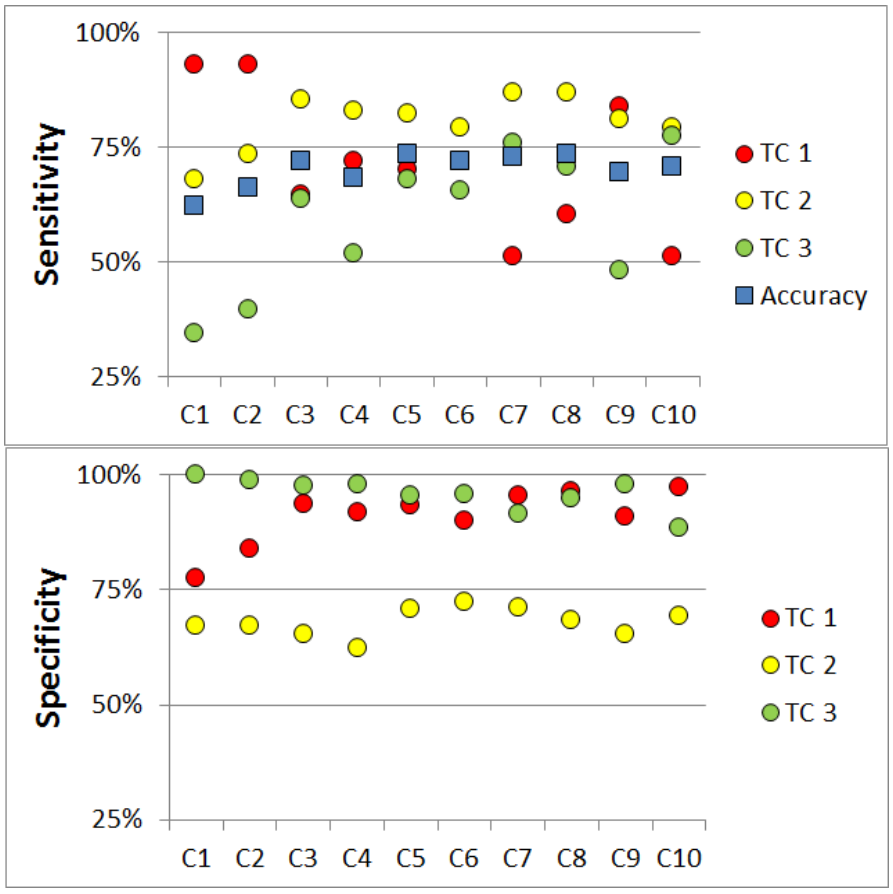


Figure 18. Accuracy, sensitivity and specificity for ten different test combinations (C1-C10)**.**

Table VIII. Predictive performance statistics for the ten test combinations (C1-C10)**.**

Alternative Method		C1	C2	C3	C4	C5
Accuracy		62.3%	66.2%	72.1%	68.2%	73.5%
Sensitivity for	**TC1**	93.0%	93.0%	64.7%	72.1%	70.0%
	TC2	67.9%	73.6%	85.4%	83.0%	82.4%
	TC3	34.5%	39.7%	63.8%	51.7%	67.9%
Specificity for	**TC1**	77.5%	83.8%	93.7%	91.9%	93.5%
	TC2	67.3%	67.3%	65.4%	62.4%	70.8%
	TC3	100.0%	99.0%	97.6%	97.9%	95.6%
Alternative Method		**C6**	**C7**	**C8**	**C9**	**C10**
Accuracy		72.1%	72.7%	73.4%	69.5%	70.8%
Sensitivity for	**TC1**	72.1%	51.2%	60.5%	83.7%	51.2%
	TC2	79.2%	86.8%	86.8%	81.1%	79.2%
	TC3	65.5%	75.9%	70.7%	48.3%	77.6%
Specificity for	**TC1**	90.1%	95.5%	96.4%	91.0%	97.3%
	TC2	72.3%	71.3%	68.3%	65.3%	69.3%
	TC3	95.8%	91.7%	94.8%	97.9%	88.5%

3.5 Integrated testing strategies

By using the methodology of ITS, we combine both the in silico and in vitro approaches presented in this report. In total, ten different ITS were investigated, as described in Table IX and illustrated by the heat map in Figure 20. The main predictive performance statistics are illustrated in Figure 21, and detailed in full in Table X.

Table IX. Description of ten different ITS. An explanation of the structure in the table, for ITS1, first the test combination C2 is applied to identify TC1 chemicals. Then the RC reg model is used to identify TC2 chemicals and finally the chemicals which are left are classified as TC3 chemicals.

Abbr.	Description	Abbr.	Description
ITS1	1. C2 (TC1 chemicals) 2. RC reg (TC2 chemicals) 3. TC3 chemicals	**ITS6**	1. C10 (TC3 chemicals) 2. C9 (TC2 chemicals) 3 TC1 chemicals
ITS2	1. C2 (TC1 chemicals) 2. C9 (TC2 chemicals) 3. TC3 chemicals.	**ITS7**	1. C8 (TC2 chemicals) 2. C2 (TC1 chemicals) 3 TC3 chemicals
ITS3	1. C2 (TC1 chemicals) 2. C8 (TC2 chemicals) 3. TC3 chemicals.	**ITS8**	1. C9 (TC2 chemicals) 2. C8 (TC3 chemicals) 3 TC1 chemicals
ITS4	1. C8 (TC3 chemicals) 2. C9 (TC2 chemicals) 3. TC1 chemicals	**ITS9**	1. C2 (TC1 chemicals) 2. C10 (TC3 chemicals) 3 TC2 chemicals
ITS5	1. C10 (TC3 chemicals) 2. C8 (TC2 chemicals) 3 TC1 chemicals	**ITS10**	1. C10 (TC3 chemicals) 2. C2 (TC2 chemicals) 3 TC1 chemicals

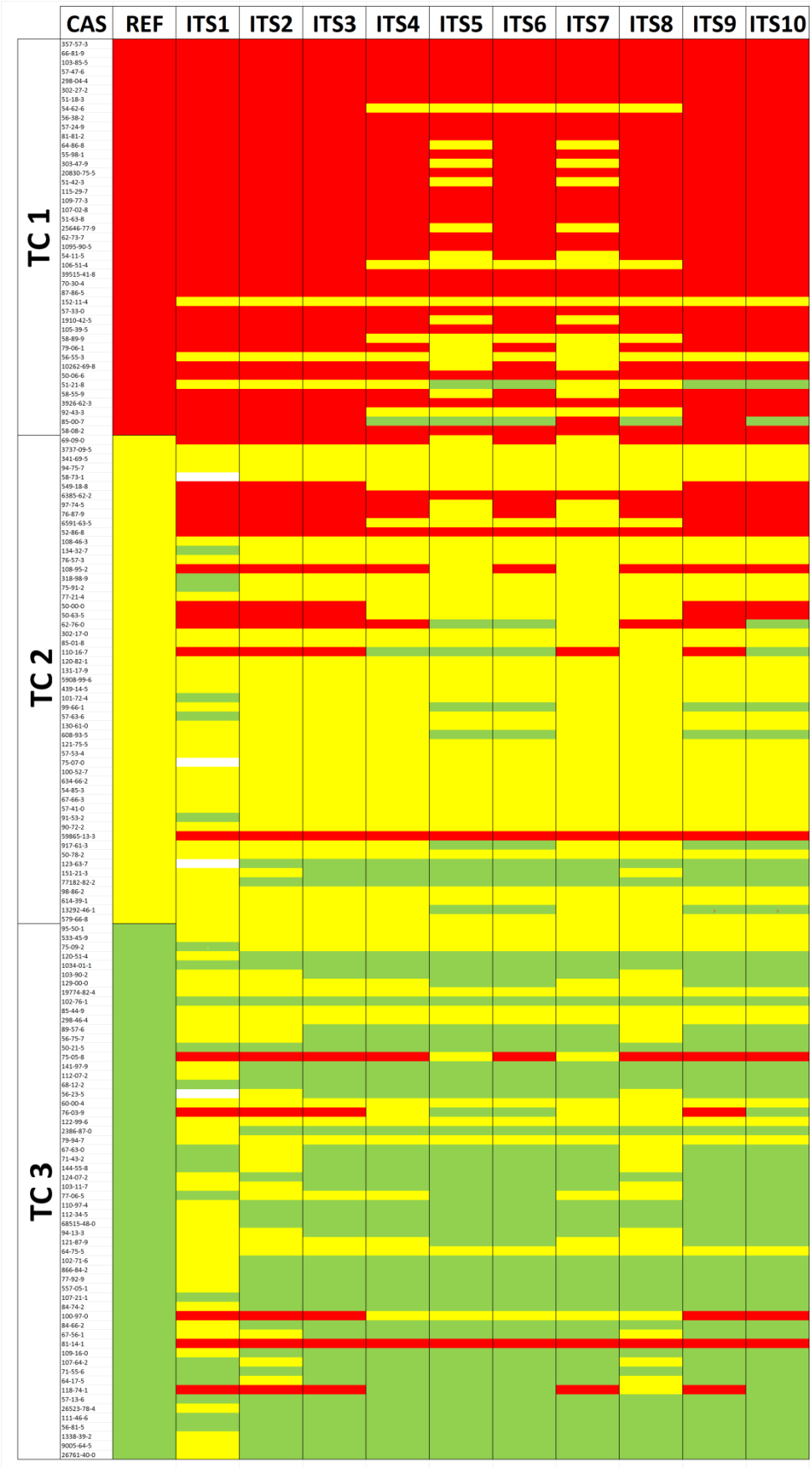


Figure 20. Heat map of ten ITS (ITS1-ITS10) compared to the reference LD_{50} values (REF).

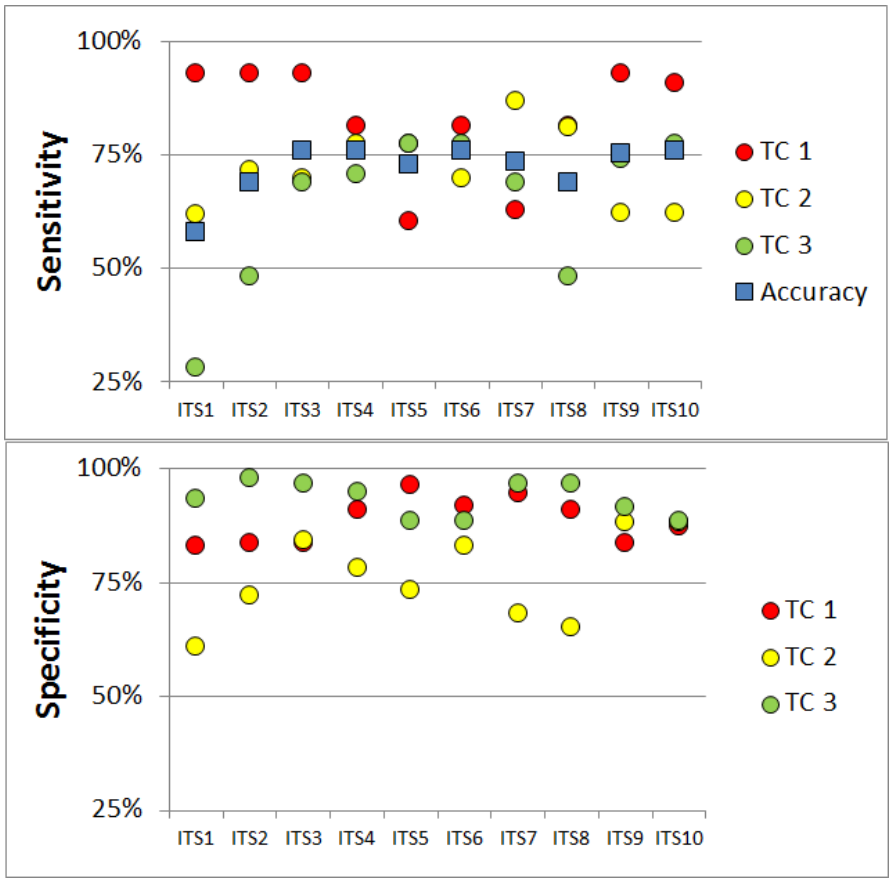


Figure 21. Scatter plots of accuracy, sensitivity and specificity for ten ITS.

Table X. Overview of the predictive capacity for ten ITS (ITS1-ITS10).

Alternative Method		ITS1	ITS2	ITS3	ITS4	ITS5
Accuracy		58.0%	68.8%	76.0%	76.0%	72.7%
Sensitivity for	**TC1**	93.0%	93.0%	93.0%	81.4%	60.5%
	TC2	62.0%	71.7%	69.8%	77.4%	77.4%
	TC3	28.1%	48.3%	69.0%	70.7%	77.6%
Specificity for	**TC1**	83.2%	83.8%	83.8%	91.0%	96.4%
	TC2	61.0%	72.3%	84.2%	78.2%	73.3%
	TC3	93.5%	97.9%	96.9%	94.8%	88.5%
Alternative Method		**ITS6**	**ITS7**	**ITS8**	**ITS9**	**ITS10**
Accuracy		76.0%	73.4%	68.8%	75.3%	76.0%
Sensitivity for	**TC1**	81.4%	62.8%	81.4%	93.0%	90.7%
	TC2	69.8%	86.8%	81.1%	62.3%	62.3%
	TC3	77.6%	69.0%	48.3%	74.1%	77.6%
Specificity for	**TC1**	91.9%	94.6%	91.0%	83.8%	87.4%
	TC2	83.2%	68.3%	65.3%	88.1%	88.1%
	TC3	88.5%	96.9%	969%	91.7%	88.5%

For illustrative purposes, we present in details just two of the 10 ITS. Figure 22 illustrates the work flow of the first ITS. The test combination C2, presented in the previous section, is used in the first step, since it was shown to be good at picking out toxic chemicals (TC1). The sensitivity and specificity for this test combination for TC1 chemicals were 96% and 84%, respectively. We use this property and begin by identifying TC1 chemicals using the test combination C2. For the remaining chemicals, predicted either as TC2 or TC3 chemicals, we use the in vitro method RC reg to identify TC2 chemicals. For this method the sensitivity was 72% for TC2 chemicals. The remaining chemicals in the set are predicted to be low toxic chemicals, in TC3. The total number of chemicals is reduced to 152 since the in vitro method did not produce experimental results for all chemicals. The overall accuracy for ITS1 is 58%, considerably higher than the accuracy of the in vitro test alone (41%) but lower than the individual QSAR tests (63-72%). The high sensitivity and specificity for TC1 chemicals is preserved from the combined test C2 (93% and 83%), the statistics for TC2 chemicals are lower (62% and 61%). For TC3 chemicals the sensitivity is only 28% while the specificity is 93%.

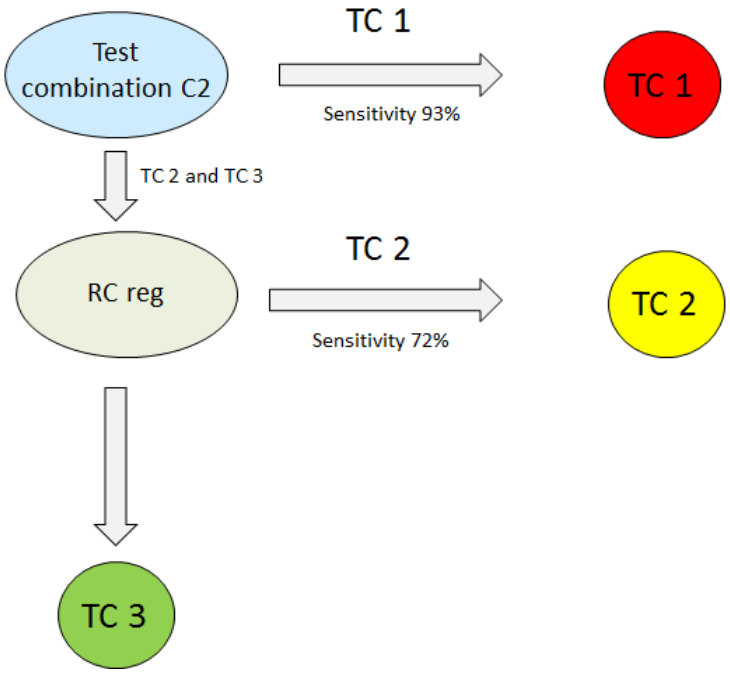


Figure 22. Work flow for ITS1; first the test combination C2 is applied to identify TC1 chemicals, then the RC reg model is used to identify TC2 chemicals and finally the chemicals which are left are predicted to be TC3 chemicals.

For ITS3, which is illustrated in Figure 23, the test combination C2 is first applied to identify TC1 chemicals. Then the test combination C8 is used to identify TC2 chemicals and the chemicals which are left are predicted as TC3 chemicals. For this combination the overall accuracy is relatively high (76%). The Cooper statistics for TC2 and TC3 chemicals are higher than for ITS1 (the sensitivity is 70% in both cases and the specificity is 84%, and 97%, respectively).

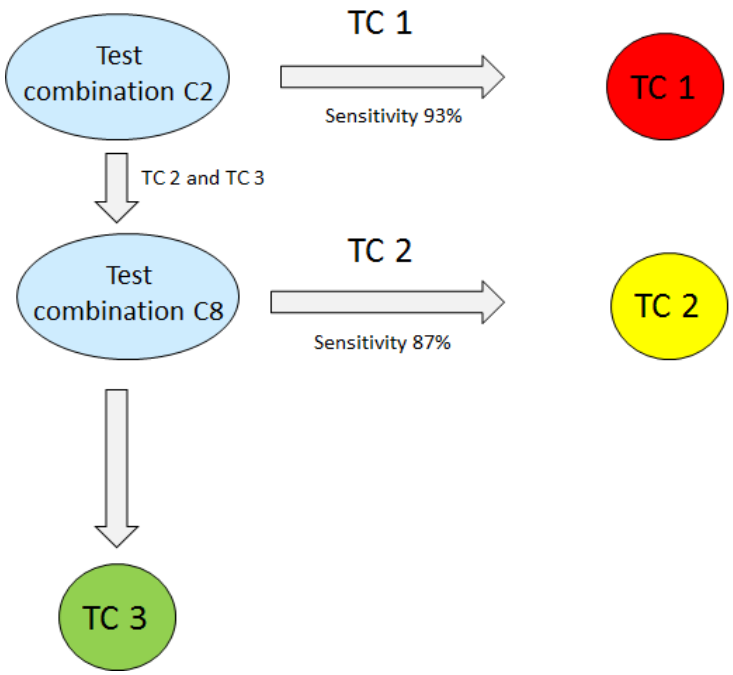


Figure 23. Work flow for ITS3; first the test combination C2 is applied to identify TC1 chemicals, then the test combination C8 is used to identify TC2 chemicals and finally the chemicals which are left are predicted to be TC3 chemicals.

4 Discussion and conclusions

In this study, we have investigated the predictive performances of five alternative approaches for the assessment of acute oral toxicity, a toxicological endpoint required in multiple pieces of legislation on chemicals and consumer products. In particular, we investigated the ability of four QSAR models (ToxSuite, TOPKAT, TEST and ADMET Predictor) and one in vitro method (3T3 NRU using the prediction model of Halle, RC reg). We characterised the predictive performance of each method when used alone (both for LD_{50} prediction and acute toxicity classification into three categories), as well as multiple test combinations (batteries) and stepwise testing strategies (for acute toxicity classification into three categories). To assess the predictive performances of the alternative methods, we compiled a test set containing in vitro and in vivo data for 180 compounds. The in vitro data are considered to be of high quality, being derived from international validation studies on in vitro tests.

In the assessment of QSAR model performance, it should be noted that the statistics for predictivity do not reflect full external predictivity, since the test chemicals had been included to some extent in the training sets of one or more of the models. For consistency in the model comparisons, it was decided not to exclude any training set chemicals, since this is only partially known for the QSAR models. Similarly, the in vitro prediction model had been calibrated by using data for some of the test chemicals, but these were not excluded when assessing the performance of the in vitro prediction model (RC reg).

The in vivo data showed considerable variability for some compounds. About 20% of the compounds cross two or three (in two cases) toxic classes. The compounds that cross three toxic classes are malathion (JRC-000012, CAS number 121-75-5) and phenytoin (JRC-000028, CAS number 57-41-0). Sixty seven of the 180 chemicals have official classifications for acute oral toxicity in Table 3.1, Annex VI to Regulation EC 1272/2008 CLP Regulation (EC, 2008). When comparing the actual CLP classification with classifications derived from the average LD_{50} values according to the GHS system, we found that in 43 cases the chemicals are classified in the same category, in 8 cases the average LD_{50} category was lower than the CLP Regulation classification and in 16 cases the average LD_{50} category was higher than the CLP Regulation classification. In these cases the difference in category was one, apart from two cases, for methanol (JRC-000019, CAS number 67-56-1) and for carbon tetrachloride (JRC-000257, CAS number 56-23-5). According to the CLP classification, these chemicals should be classified in category 3 while with according to their average LD_{50} values, they would not be classified, with average $LD_{50} > 2000$ mg/kg bodyweight. A probable reason for these differences could be due to volatility and loss of the test chemicals.

Overall, the alternative methods, when used individually, showed an ability to predict LD_{50} with correlation coefficients in the range of 49% to 84%, and to classify into three toxicity groups with accuracies in the range 41% to 72%. Among the QSAR models, the best performing models were ToxSuite and TEST, with correlation coefficients of approximately 80% in LD_{50} prediction, and accuracies of approximately 70% in acute toxicity classification. The in vitro 3T3 NRU method,

based on the use of the Halle prediction model, had a correlation coefficient with LD_{50} of approximately 50%, and a classification accuracy of approximately 41%.

When the QSAR models are combined in batteries, the overall accuracies were between 62% and 74%. While these figures are not much higher than the individual QSAR models alone, the sensitivities for the different toxic classes are considerably higher. For example, the highest sensitivity for the most toxic class was 93% in one test combination compared to 67% for an individual QSAR model. The corresponding sensitivity figures for the other toxic classes are 87% compared to 81% (TC2) and 78% compared to 67% (TC3). On the other hand, the differences between the specificities for the different toxic classes are relatively small. The highest specificity for the most toxic class is 97% for a test combination compared to 95% for an individual QSAR model. The highest specificities in other toxic classes are 72% compared to 70% (TC2) and 100% compared to 95% (TC3)

When the alternative methods are used in a stepwise testing strategy the overall accuracy could reach 76%. Different test combinations can be optimised according to the end-users requirements: for example, to maximise overall accuracy, a suitable choice would be ITS3 (or ITS4, ITS10 described in the Annex); to maximise sensitivity for toxic chemicals (at the expense of a higher false positive rate), a suitable choice would be ITS1 (or ITS2, ITS3, ITS9); whereas to maximise specificity for non-toxic chemicals (at the expense of a higher false negative rate), a suitable choice would be ITS5.

On the basis of these results, it can be concluded that:

d) the variability in LD_{50} values has an impact on classification, which means that the use of average LD_{50} values as a reference standard has to be used with care. A detailed analysis of the reference in vivo (LD_{50}) data, to characterise the variability in these data and the impact on the ability to predict in vivo toxicity, is given elsewhere (Norlén et al., 2012).
e) the in vitro test, 3T3 NRU used with the prediction model of Halle, has a lower predictive performance than the QSAR models. It is possible, however, that the in vitro test has a broader domain of applicability compared to the QSARs. It would be useful to explore whether the predictive performance of the in vitro system could also be increased by using an alternative prediction model.
f) the overall accuracies for the test combinations or testing strategies are not much higher than the AMs used alone, but may be optimised in terms of overall predictivity, sensitivity and specificity according to the end-user's requirements.

A similar comparison study of acute toxicity classification into three categories by four QSAR models (ToxSuite, TOPKAT, TEST and ADMET Predictor) and one in vitro method (3T3 NRU using the prediction model of Halle) has to our knowledge not been done before. Sedykh and coworkers (Sedykh et al., 2011) adopted the same toxicity classification scheme with three categories when evaluating their own QSAR and hybrid models. In contrast to our study, Sedykh and coworkers made the simplification to exclude the "marginal" compounds (corresponding to TC2 chemicals) and hence made a clear binary classification of "toxic" and "nontoxic" chemicals.

They used k nearest neighbor classification and random forest QSAR methods to model LD_{50} data using chemicals descriptors alone or combined with biological descriptors derived from concentration-response quantitative high-throughput screening (qHTS) data. The performance of their hybrid models were shown to be superior to TOPKAT.

In another study, Zhu et al. (2009a) divided the ZEBET (database on alternatives to animal experiments on the Internet) dataset into two groups, i.e. compounds with a good or a bad IC_{50}/LD_{50} correlation. The LD_{50} prediction accuracy of the resulting models proved superior to TOPKAT models applied to the same external test set of rodent acute toxicity data (RTECS chemicals). In addition to these local models, a number of QSAR models for rat oral acute toxicity have been developed using large datasets (global models) have been reported by Zhu and coworkers (Zhu et al., 2009b). These models were built by using a combinatorial QSAR modelling approach, including several sets of descriptors and employing several statistical modelling methods (e.g. nearest neighbour methods, the random forest method, and the FDA MDL QSAR method). Ultimately, consensus models were developed by averaging the predicted LD_{50} for every compound using all five models, which afforded higher prediction accuracy as compared to individual models. However, as a result of using a large number of descriptors, which are often sparsely populated, the multidimensional space defined by each of these models is complex and fragmented. As a result of the high complexity of the modelling procedure, these models are difficult to reproduce, even by a specialist, and thus they are not easily transferable and practically useful.

A study by Raevsky and coworkers (Raevsky et al., 2010) proposed the so-called Arithmetic Mean Toxicity (AMT) modelling approach, which produces local models based on a *k*-nearest neighbors approach. The authors showed that LD_{50} values could be predicted with r^2 values up to 0.78, depending on the selection of nearest neighbours (analogues), which is significantly better than the statistics associated with in vitro-in vivo correlation (typically r^2 values less than 0.5). The approach is transparent and reproducible, but would need to be implemented in a software tool for ease of application. It can be thought of as an automated read-across approach.

The results obtained in this study could be used as the basis for further investigations. For example, the Cooper statistics obtained for the AMs could be used as input parameters to explore Bayesian approaches for combining the results obtained by different methods. The results could also be used in the context of cost-benefit analyses in which the cost relates to the time, expense and difficulty of applying a method, and is compared with the benefit in terms of predictive ability and reduction in animal testing.

Further studies, based on more extensive and high quality datasets (e.g. as generated by High Throughput Screening), would also be valuable in the search for optimal strategies for assessing acute toxicity.

5 References

Accelrys/TOPKAT, version 6.2. Available at: http://accelrys.com/solutions/scientific-need/predictive-toxicology.html

ACD/Tox Suite, version 2.95. Available at: http://www.acdlabs.com/products/pc_admet/tox/tox/

ACuteTox. Available at: http://www.acutetox.eu/

ADMET Predictor, version 5.5. Available at: http://www.simulations-plus.com/Products.aspx?grpID=1&cID=11&pID=13

Bassan A & Worth AP (2008). The Integrated Use of Models for the Properties and Effects of Chemicals by means of a Structured Workflow. QSAR & Combinatorial Science 27: 6-20.

ChemIDplus. Available at: http://chem.sis.nlm.nih.gov/chemidplus

Clemedson C et al. (1996). MEIC evaluation of acute systemic toxicity. Alternatives to Laboratory Animals 24: 249-272.

Collins FS, Gray GM & Bucher JR (2008). Transforming Environmental Health Protection. Science 319: 906-907.

Cooper JA, Saracci R & Cole P (1979). Describing the validity of carcinogen screening tests. British Journal of Cancer 39: 87-89.

Dix DJ, Houck KA, Martin MT et al. (2007). The ToxCast Program for Prioritizing Toxicity Testing of Environmental Chemicals. Toxicological Sciences 95: 5-12.

EC (1967). Council Directive 67/548/EEC of 27 June 1967 on the approximation of laws, regulations and administrative provisions relating to the classification, packaging and labelling of dangerous substances. Available at: http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:31967L0548:EN:HTML

EC (1999). Directive 1999/45/EC of the European Parliament and of the Council of 31 May 1999 concerning the approximation of the laws, regulations and administrative provisions of the Member States relating to the classification, packaging and labelling of dangerous preparations. Official Journal of the European Communities L200. Available at: http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:1999:200:0001:0068:EN:PDF

EC (2003). Directive 2003/15/EC of the European Parliament and the Council of 27 February amending Directive 76/786/EEC on the approximation of the laws of the member States relating to

cosmetic products. Official Journal of the European Communities L66, 26–35. Available at: http://ec.europa.eu/consumers/sectors/cosmetics/files/doc/200315/200315_en.pdf

EC (2006). Regulation (EC) No 1907/2006 of the European Parliament and of the Council of 18 December 2006 concerning the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH), establishing a European Chemicals Agency, amending Directive 1999/45/EC and repealing Council Regulation (EEC) No 793/93 and Commission Regulation (EC) No 1488/94 as well as Council Directive 76/769/EEC and Commission Directives 91/155/EEC, 93/67/EEC, 93/105/EC and 2000/21/EC. Official Journal of the European Union L396. Available at: http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2006:396:0001:0849:EN:PDF

EC (2008). Regulation (EC) No 1272/2008 on classification, labelling and packaging of substances and mixtures, amending and repealing directives 67/548/EEC and 1999/45/EC, and amending Regulation (EC) No 1907/2006 (REACH). Official Journal of the European Union L353. Available at:
http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2008:353:0001:1355:en:PDF

Halle W (2003). The Registry of Cytotoxicity: Toxicity Tesing in Cell Cultures to predict Acute Toxicity (LD50) and to reduce Tesing in Animals. Alternatives to Laboratory Animals. 31: 89-198.

Hazardous Substances Data Bank (HSDB). Available at: http://www.nlm.nih.gov/pubs/factsheets/hsdbfs.html

Kinsner-Ovaskainen A, Bulgheroni A, Hartung T & Prieto P (2009). ECVAM's ongoing activities in the area of acute oral toxicity. Toxicology in Vitro 23(8): 1535-40.

Lapenna S, Fuart-Gatnik M & Worth A (2010). Review of QSAR Models and Software Tools for predicting Acute and Chronic Systemic Toxicity. JRC Scientific and Technical Report. EUR 24639 EN - 2010. Available at: http://publications.jrc.ec.europa.eu/repository/

NICEATM-ICCVAM. (2006). Background Review Document (BRD): Validation of Neutral Red Uptake Test Methods NIH /In Vitro Cytotoxicity Test Methods for Estimating Acute Oral Systemic Toxicity. Publication No. 07-4518, November 2006. Available at:
http://iccvam.niehs.nih.gov/methods/acutetox/inv_nru_brd.htm

Norlén H & Worth A (2012). The variability in LD_{50} values and its impact on classification of chemicals. JRC Scientific and Technical Report, in preparation.

NRC, National Research Council (2007). Toxicity Testing in the 21st Century: A Vision and a Strategy. National Academies Press, Washington, DC, USA.

OECD (1987). OECD Guideline for the Testing of Chemicals, Test Guideline 401, Acute Oral Toxicity. OECD, Paris, France. Available at**:** http://www.oecd.org/dataoecd/63/10/45122762.pdf.

OECD (2001a). Guideline for Testing of Chemicals, 420, Acute Oral Toxicity-Fixed Dose Method. Paris, France. Available at: http://browse.oecdbookshop.org/oecd/pdfs/free/9742001e.pdf.

OECD (2001b). Guideline for Testing of Chemicals, 423, Acute Oral Toxicity-Acute Toxic Class Method. Paris, France. Available at: http://browse.oecdbookshop.org/oecd/pdfs/free/9742301e.pdf.

OECD (2001c). Guideline for Testing of Chemicals, 425, Acute Oral Toxicity – Up-and-Down Procedure. Paris, France. Available at: http://www.oecd.org/dataoecd/17/51/1948378.pdf

OECD (2010) Guidance Document No 129 on using Cytotoxicity Tests to Estimate Starting Doses for Acute Oral Systemic Toxicity Tests. Paris, France.
Available at: http://iccvam.niehs.nih.gov/SuppDocs/FedDocs/OECD/OECD-GD129.pdf

Raevsky OA et al. (2010). Prediction of acute toxicity to mice by the arithmetic mean toxicity (AMT) modelling approach. SAR and QSAR in Environmental Research 21(1): 265-275.

RTECS (Registry of Toxic Effects of Chemical Substances). Available at: http://accelrys.com/products/databases/bioactivity/rtecs.html

Sedykh A, Zhu H, Tang H, Zhang L, Richard A, et al. (2011). Use of in Vitro HTS-Derived Concentration–Response Data as Biological Descriptors Improves the Accuracy of QSAR Models of in Vivo Toxicity. Environmental Health Perspectives 119(3): 364-370.

Seidle T et al. (2010). Cross-sector review of drivers and available 3Rs approaches for acute systemic toxicity testing. Toxicological Sciences. 116(2): 382-396.

UN (2007). Globally Harmonized System of Classification and Labelling of Chemicals (GHS), ST/SG/AC.10/30/Rev.2. Available at:
http://www.unece.org/trans/danger/publi/ghs/ghs_rev02/02files_e.html

U.S. EPA / T.E.S.T. (Toxicity Estimation Software Tool) version 4.0, Available at: http://www.epa.gov/nrmrl/std/qsar/qsar.html

van Leeuwen CJ, Patlewicz GY and Worth AP (2007). Intelligent Testing Strategies, in Risk Assessment of Chemicals. An Introduction. 2nd Edition. CJ van Leeuwen & TG Vermeire (Eds). Springer, Heidelberg, Germany, pp. 467-509.

ZEBET (Zentralstelle zur Erfassung und Bewertung von Ersatz- und Ergänzungsmethoden zum Tierversuch) database. Available at: http://www.dimdi.de/static/en/db/dbinfo/zt00.htm

Zhu H, Ye L, Richard A, Golbraikh A, Wright FA, Rusyn I & Tropsha A (2009a). A novel two-step hierarchical quantitative structure-activity relationship modeling work flow for predicting acute toxicity of chemicals in rodents. Environmental Health Perspectives 117(8): 1257-1264.

Zhu H, Martin TM, Ye L, Sedykh A, Young DM & Tropsha A (2009b). Quantitative Structure−Activity Relationship Modeling of Rat Acute Toxicity by Oral Exposure. Chemical Research in Toxicology 22(12): 1913–1921.

6 Annex. List of 180 chemicals used in the study.

Nr	JRC nr	CAS	Chemical name	SMILES[1]
1	JRC-000002	103-90-2	acetaminophen	CC(=O)Nc1ccc(cc1)O
2	JRC-000003	50-78-2	acetylsalicylic acid	CC(=O)Oc1ccccc1C(=O)O
3	JRC-000004	5908-99-6	atropine sulfate monohydrate	C3C(CC2CCC3N2C)OC(C(c1ccccc1)CO)=O
4	JRC-000005	58-08-2	caffeine	Cn1cnc2c1c(=O)n(C)c(=O)n2C
5	JRC-000006	298-46-4	carbamazepine	NC(=O)N1c3ccccc3C=Cc2c1cccc2
6	JRC-000007	64-86-8	colchicine	COc3c(OC)cc2c(c3OC)c1ccc(c(=O)cc1C(CC2)NC(=O)C)OC
7	JRC-000008	66-81-9	cycloheximide	CC2CC(C)C(=O)C(C2)C(CC1CC(=O)NC(=O)C1)O
8	JRC-000009	439-14-5	diazepam	Clc3ccc2c(c3)C(=NCC(=O)N2C)c1ccccc1
9	JRC-000010	20830-75-5	digoxin	O=C3OCC(=C3)C4CCC8(C4(C)C(O)CC5C8CCC2C5(C)CCC(C2)OC6CC(O)C(C(O6)C)OC7CC(O)C(C(O7)C)OC1CC(O)C(C(O1)C)O)O
10	JRC-000011	67-63-0	propan-2-ol	CC(O)C
11	JRC-000012	121-75-5	malathion	CCOC(=O)CC(C(=O)OCC)SP(=S)(OC)OC
12	JRC-000013	7487-94-7	mercury dichloride	-
13	JRC-000014	87-86-5	pentachlorophenol	Clc1c(O)c(Cl)c(c(c1Cl)Cl)Cl
14	JRC-000015	50-06-6	phenobarbital	CCC2(C(=O)NC(=O)NC2=O)c1ccccc1
15	JRC-000016	151-21-3	sodium lauryl sulfate	CCCCCCCCCCCCOS(=O)(=O)O[Na]
16	JRC-000017	1069-66-5	sodium valproate	-
17	JRC-000018	7720-78-7	ferrous sulfate	-
18	JRC-000019	67-56-1	methanol	CO
19	JRC-000020	10377-48-7	lithium salt	-
20	JRC-000021	58-55-9	theophylline	Cn1c(=O)n(C)c2c(c1=O)ncn2
21	JRC-000022	130-61-0	thioridazine hydrochloride	N4(c2c(Sc3c4cccc3)ccc(c2)SC)CCC1N(CCCC1)C
22	JRC-000023	81-81-2	warfarin	CC(=O)CC(c3c(=O)oc2c(c3O)cccc2)c1ccccc1
23	JRC-000024	67-66-3	chloroform	ClC(Cl)Cl
24	JRC-000025	54-85-3	isoniazid	NNC(=O)c1ccncc1
25	JRC-000026	75-09-2	dichloromethane	ClCCl
26	JRC-000027	50-63-5	chloroquine bis(phosphate)	c12c(nccc1NC(CCCN(CC)CC)C)cc(Cl)cc2
27	JRC-000028	57-41-0	phenytoin, 5,5-diphenylhydantoin	O=C2NC(=O)NC2(c3ccccc3)c1ccccc1
28	JRC-000029	94-75-7	2,4-dichlorophenoxyacetic acid	OC(=O)COc1ccc(cc1Cl)Cl
29	JRC-000030	57-33-0	pentobarbital sodium	C1(C(=O)NC(=O)NC1=O)(C(CCC)C)CC
30	JRC-000031	10262-69-8	maprotiline	CNCCCC42CCC(c3c4cccc3)c1c2cccc1
31	JRC-000032	3737-09-5	disopyramide	CC(N(C(C)C)CCC(c2ccccn2)(c1ccccc1)C(=O)N)C
32	JRC-000033	58-73-1	diphenhydramine	CN(CCOC(c2ccccc2)c1ccccc1)C
33	JRC-000034	533-45-9	clomethiazole	Cc1ncsc1CCCl
34	JRC-000035	6591-63-5	quinidine sulfate	c12c(C(C3N4CC(C(C3)CC4)C=C)O)ccnc1ccc(c2)OC
35	JRC-000036	76-57-3	codeine	COc5ccc2c3c5OC1C43CCN(C(C2)C4C=CC1O)C

[1]The SMILES used in the fourth column describes the structure of chemical molecules used in the computational form for the QSAR software programs. The SMILES for inorganic compounds are omitted.

Nr	JRC nr	CAS	Chemical name	SMILES
36	JRC-000037	69-09-0	chlorpromazine hydrochloride	N1(c2c(Sc3c1cccc3)ccc(c2)Cl)CCCN(C)C
37	JRC-000038	123-63-7	paraldehyde	CC1OC(C)OC(O1)C
38	JRC-000039	144-55-8	sodium bicarbonate	O=C(O)O
39	JRC-000040	79-06-1	acrylamide	NC(=O)C=C
40	JRC-000041	50-00-0	formaldehyde	C=O
41	JRC-000042	75-07-0	acetaldehyde	CC=O
42	JRC-000043	107-02-8	acrolein	C=CC=O
43	JRC-000044	19774-82-4	amiodarone hydrochloride	c3(c2c(oc3CCCC)cccc2)C(=O)c1cc(c(c(c1)I)OCCN(CC)CC)I
44	JRC-000045	51-63-8	dexamphetamine sulphate	c1(ccccc1)CC(C)N
45	JRC-000046	71-43-2	benzene	c1ccccc1
46	JRC-000047	56-55-3	benz[a]anthracene	c3ccc4c(c3)cc1c(c4)ccc2c1cccc2
47	JRC-000048	59865-13-3	cyclosporine	CC=CCC(C(C1C(=O)NC(CC)C(=O)N(C)CC(=O)N(C)C(CC(C)C)C(=O)NC(C(C)C)C(=O)N(C)C(CC(C)C)C(=O)NC(C(=O)NC(C(=O)N(C(C(=O)N(C(C(=O)N(C(C(=O)N1C)C(C)C)C)CC(C)C)C)CC(C)C)C)C)O)C
48	JRC-000049	57-63-6	ethinylestradiol	C#CC3(O)CCC4C3(C)CCC1C4CCc2c1ccc(c2)O
49	JRC-000050	1095-90-5	methadone hydrochloride	CCC(=O)C(c2ccccc2)(c1ccccc1)CC(N(C)C)C
50	JRC-000051	64-75-5	tetracycline hydrochloride	C1C2C(C(=C3C(c4c(cccc4C(C13)(C)O)O)=O)O)(C(C(C(N)=O)=C(C2N(C)C)O)=O)O
51	JRC-000052	341-69-5	ethanamine	c1(C(c2ccccc2)OCCN(C)C)c(cccc1)C
52	JRC-000053	152-11-4	verapamil hydrochloride	c1(C(CCCN(CCc2cc(c(OC)cc2)OC)C)(C(C)C)C#N)cc(c(OC)cc1)OC
53	JRC-000054	13292-46-1	rifamycin	COC1C=COC5(C)Oc3c(C5=O)c2c(O)c(C=NN4CCN(CC4)C)c(c(c2c(c3C)O)O)NC(=O)C(=CC=CC(C(C(C(C(C(C1C)OC(=O)C)C)O)C)O)C)C
54	JRC-000055	111-46-6	ethanol	OCCOCCO
55	JRC-000056	15663-27-1	platinum	-
56	JRC-000057	85-00-7	diquat dibromide	BrN13CCCCC3C2N(CC1)(Br)CCCC2
57	JRC-000058	303-47-9	ochratoxin A	CC3OC(=O)c2c(C3)c(Cl)cc(c2O)C(=O)NC(C(=O)O)Cc1ccccc1
58	JRC-000059	608-93-5	pentachlorobenzene	Clc1cc(Cl)c(c(c1Cl)Cl)Cl
59	JRC-000060	85-01-8	phenanthrene	c3ccc2c(c3)c1ccccc1cc2
60	JRC-000061	129-00-0	pyrene	c1cc2ccc3c4c2c(c1)ccc4ccc3
61	JRC-000062	118-74-1	benzene, 1,2,3,4,5,6-hexachloro-	Clc1c(Cl)c(Cl)c(c(c1Cl)Cl)Cl
62	JRC-000063	77182-82-2	glufosinate-ammonium	P(=O)(CCC(C(=O)O)N)(C)O
63	JRC-000064	51-21-8	5-fluorouracil	Fc1cnc(=O)nc1=O
64	JRC-000065	75-91-2	tert-butyl hydroperoxide	OOC(C)(C)C
65	JRC-000066	10108-64-2	cadmium chloride	-
66	JRC-000067	634-66-2	1,2,3,4-tetrachlorobenzene	Clc1c(Cl)ccc(c1Cl)Cl
67	JRC-000068	54-11-5	nicotine	CN2CCCC2c1cccnc1
68	JRC-000069	58-89-9	cyclohexane	ClC1C(Cl)C(Cl)C(C(C1Cl)Cl)Cl
69	JRC-000070	64-17-5	ethanol	CCO
70	JRC-000071	56-38-2	parathion	CCOP(=S)(Oc1ccc(cc1)N(=O)=O)OCC
71	JRC-000072	62-73-7	dichlorvos	COP(=O)(OC=C(Cl)Cl)OC
72	JRC-000073	57-47-6	physostigmine	CNC(=O)Oc3ccc2c(c3)C1(C)CCN(C1N2C)C
73	JRC-000074	7681-49-4	sodium fluoride	-
74	JRC-000075	1910-42-5	paraquat dichloride	c1(c2ccn(H)(cc2)C)ccn(H)(cc1)C
75	JRC-000076	56-81-5	glycerol	OCC(CO)O

Nr	JRC nr	CAS	Chemical name	SMILES
76	JRC-000077	68-12-2	N,N-dimethylformamide	O=CN(C)C
77	JRC-000078	549-18-8	amitriptyline hydrochloride	C1(c2c(CCc3c1cccc3)cccc2)=CCCN(C)C
78	JRC-000079	107-21-1	ethylene glycol	OCCO
79	JRC-000080	108-95-2	phenol	Oc1ccccc1
80	JRC-000081	7647-14-5	sodium chloride	-
81	JRC-000082	1330-20-7	benzene, dimethyl-, xylene	-
82	JRC-000083	151-50-8	potassium cyanide (K(CN))	-
83	JRC-000084	52-86-8	haloperidol	Fc2ccc(cc2)C(=O)CCCN3CCC(CC3)(O)c1ccc(cc1)Cl
84	JRC-000085	318-98-9	2-propanol	c12c(OCC(CNC(C)C)O)cccc1cccc2
85	JRC-000086	1327-53-3	arsenic oxide (As_2O_3)	-
86	JRC-000087	7446-18-6	dithallium sulphate	-
87	JRC-000088	70-30-4	hexachlorophene	Clc2cc(Cl)c(c(c2O)Cc1c(O)c(Cl)cc(c1Cl)Cl)Cl
88	JRC-000089	56-75-7	chloramphenicol	OCC(C(c1ccc(cc1)N(=O)=O)O)NC(=O)C(Cl)Cl
89	JRC-000090	7447-40-7	potassium chloride	-
90	JRC-000091	302-17-0	chloral hydrate	OC(C(Cl)(Cl)Cl)O
91	JRC-000092	57-53-4	meprobamate	CCCC(COC(=O)N)(COC(=O)N)C
92	JRC-000093	57-24-9	strychnine	O=C5CC7OCC=C3C4C7C2N5c6ccccc6C12CCN(C1C4)C3
93	JRC-000094	77-21-4	glutethimide	CCC2(CCC(=O)NC2=O)c1ccccc1
94	JRC-000095	614-39-1	procainamide hydrochloride	c1(C(NCCN(CC)CC)=O)ccc(N)cc1
95	JRC-000096	13410-01-0	selenic acid	-
96	JRC-000097	75-05-8	acetonitrile	CC#N
97	JRC-000098	51-42-3	epinephrine hydrogen tartrate	c1(cc(c(O)cc1)O)C(CNC)O
98	JRC-000099	99-66-1	valproic acid	CCCC(C(=O)O)CCC
99	JRC-000106	90-72-2	2,4,6-tris(dimethylaminomethyl)phenol	CN(Cc1cc(CN(C)C)c(c(c1)CN(C)C)O)C
100	JRC-000107	98-86-2	acetophenone	CC(=O)c1ccccc1
101	JRC-000108	141-97-9	butanoic acid	CCOC(=O)CC(=O)C
102	JRC-000109	100-97-0	1,3,5,7-Tetraazatricyclo[3.3.1.13,7]decane	C1N2CN3CN1CN(C2)C3
103	JRC-000110	112-34-5	ethanol, 2-(2-butoxyethoxy)-	CCCCOCCOCCO
104	JRC-000111	106-51-4	2,5-cyclohexadiene-1,4-dione	O=C1C=CC(=O)C=C1
105	JRC-000112	2386-87-0	7-oxabicyclo[4.1.0]heptane-3-carboxylic acid	O=C(C3CCC4C(C3)O4)OCC1CCC2C(C1)O2
106	JRC-000113	9005-64-5	polysorbate 20	OCCOC(C1OC(CC1OCCO)OCCO)COCCOC(=O)CCCCCCCC=CCCCCCCCC
107	JRC-000114	105-39-5	acetic acid, 2-chloro-, ethyl ester	CCOC(=O)CCl
108	JRC-000115	81-14-1	4'-tert-butyl-2',6'-dimethyl-3',5'-dinitroacetophenone	O=N(=O)c1c(C)c(C(=O)C)c(c(c1C(C)(C)C)N(=O)=O)C
109	JRC-000116	91-53-2	ethoxyquin	CCOc1ccc2c(c1)C(=CC(N2)(C)C)C
110	JRC-000117	110-97-4	2-propanol	CC(CNCC(O)C)O

Nr	JRC nr	CAS	Chemical name	SMILES
111	JRC-000118	3926-62-3	acetic acid, 2-chloro-, sodium salt (1:1)	C(CCl)(O)=O
112	JRC-000119	917-61-3	sodium cyanate	[Na]OC#N
113	JRC-000120	68515-48-0	1,2-benzenedicarboxylic acid	CCCCCCCCCOC(=O)c1ccccc1C(=O)OCCCCCCCCC
114	JRC-000121	101-72-4	1,4-benzenediamine	CC(Nc2ccc(cc2)Nc1ccccc1)C
115	JRC-000122	7778-80-5	potassium sulfate	-
116	JRC-000123	92-43-3	1-phenyl-3-pyrazolidone	O=C2CCN(N2)c1ccccc1
117	JRC-000124	1338-39-2	sorbitan	CCCCCCCCCCCC(=O)OCC(C1OCC(C1O)O)O
118	JRC-000125	60-00-4	edetic acid	OC(=O)CN(CC(=O)O)CCN(CC(=O)O)CC(=O)O
119	JRC-000126	357-57-3	strychnidin-10-one, 2,3-dimethoxy-	COc5cc6c(cc5OC)N1C3C76CCN2C7CC4C3C(CC1=O)OCC=C4C2
120	JRC-000127	120-82-1	benzene, 1,2,4-trichloro-	Clc1ccc(c(c1)Cl)Cl
121	JRC-000128	97-74-5	tetramethylthiuram monosulphide	CN(C(=S)SC(=S)N(C)C)C
122	JRC-000129	95-50-1	1,2-dichlorobenzene	Clc1ccccc1Cl
123	JRC-000130	1034-01-1	octyl ester	CCCCCCCCOC(=O)c1cc(O)c(c(c1)O)O
124	JRC-000131	108-46-3	1,3-benzenediol	Oc1cccc(c1)O
125	JRC-000132	25646-77-9	ethanol	OCCN(c1ccc(c(c1)C)N)CC
126	JRC-000133	1314-13-2	zinc oxide (ZnO)	-
127	JRC-000134	109-16-0	2-propenoic acid	O=C(C(=C)C)OCCOCCOCCOC(=O)C(=C)C
128	JRC-000135	107-64-2	1-octadecanaminium	CCCCCCCCCCCCCCCCCCCN(H)(CCCCCCCCCCCCCCCCCC)(C)C
129	JRC-000136	85-44-9	phthalic anhydride	O=C1OC(=O)c2c1cccc2
130	JRC-000137	79-94-7	2,2',6,6'-tetrabromo-4,4'-isopropylidenediphenol	CC(c2cc(Br)c(c(c2)Br)O)(c1cc(Br)c(c(c1)Br)O)C
131	JRC-000138	26523-78-4	phenol, nonyl-, 1,1',1"-phosphite	CCCCCCCCCc2ccc(cc2)OP(Oc3ccc(cc3)CCCCCCCCC)Oc1ccc(cc1)CCCCCCCCC
132	JRC-000139	112-07-2	ethanol, 2-butoxy-, 1-acetate	CCCCOCCOC(=O)C
133	JRC-000140	121-87-9	benzenamine, 2-chloro-4-nitro-	O=N(=O)c1ccc(c(c1)Cl)N
134	JRC-000141	124-07-2	octanoic acid	CCCCCCCC(=O)O
135	JRC-000142	120-51-4	benzoic acid, phenylmethyl ester	O=C(c2ccccc2)OCc1ccccc1
136	JRC-000143	131-17-9	1,2-benzenedicarboxylic acid	C=CCOC(=O)c1ccccc1C(=O)OCC=C
137	JRC-000144	26761-40-0	1,2-benzenedicarboxylic acid	CC(CCCCCCCOC(=O)c1ccccc1C(=O)OCCCCCCCC(C)C)C
138	JRC-000145	122-99-6	ethanol, 2-phenoxy-	OCCOc1ccccc1
139	JRC-000146	102-71-6	ethanol, 2,2',2"-nitrilotris-	OCCN(CCO)CCO
140	JRC-000147	134-32-7	1-naphthalenamine	Nc1cccc2c1cccc2
141	JRC-000148	110-16-7	2-butenedioic acid (2Z)-	OC(=O)C=CC(=O)O
142	JRC-000149	10361-37-2	barium chloride (BaCl2)	-
143	JRC-000150	579-66-8	2,6-diethylaniline	CCc1cccc(c1N)CC
144	JRC-000151	302-27-2	aconitane-3,8,13,14,15-pentol	COCC15CN(CC)C6C4(C5C(OC)C6C2(C7C4CC(C7OC(=O)c3ccccc3)(C(C2O)OC)O)OC(=O)C)C(CC1O)OC
145	JRC-000152	12125-02-9	ammonium chloride ((NH4)Cl)	-

Nr	JRC nr	CAS	Chemical name	SMILES
146	JRC-000153	7758-98-7	copper sulphate	-
147	JRC-000154	557-05-1	zinc distearate	CCCCCCCCCCCCCCCCCC(=O)O[Zn]OC(=O)CCCCCCCCCCCCCCCCC
148	JRC-000155	102-76-1	1,2,3-propanetriol, 1,2,3-triacetate	CC(=O)OC(COC(=O)C)COC(=O)C
149	JRC-000156	103-11-7	2-propenoic acid, 2-ethylhexyl ester	CCCCC(COC(=O)C=C)CC
150	JRC-000157	100-52-7	benzaldehyde	O=Cc1ccccc1
151	JRC-000158	7779-90-0	trizinc bis(orthophosphate)	-
152	JRC-000159	109-77-3	propanedinitrile	N#CCC#N
153	JRC-000160	866-84-2	tripotassium citrate	C(CC(O)=O)(CC(O)=O)(C(O)=O)O
154	JRC-000161	57-13-6	urea	NC(=O)N
155	JRC-000189	84-74-2	dibutyl phthalate	CCCCOC(=O)c1ccccc1C(=O)OCCCC
156	JRC-000235	71-55-6	1,1,1-trichloroethane	CC(Cl)(Cl)Cl
157	JRC-000236	89-57-6	benzoic acid	Nc1ccc(c(c1)C(=O)O)O
158	JRC-000237	54-62-6	L-glutamic acid	OC(=O)CCC(C(=O)O)NC(=O)c3ccc(cc3)NCc1cnc2c(n1)c(N)nc(n2)N
159	JRC-000238	10043-35-3	boric acid	-
160	JRC-000239	55-98-1	busulfan	CS(=O)(=O)OCCCCOS(=O)(=O)C
161	JRC-000240	77-92-9	citric acid	OC(=O)C(CC(=O)O)(CC(=O)O)O
162	JRC-000241	76-87-9	fentin	O[Sn](c2ccccc2)(c3ccccc3)c1ccccc1
163	JRC-000242	84-66-2	diethyl phthalate	CCOC(=O)c1ccccc1C(=O)OCC
164	JRC-000243	298-04-4	phosphorodithioic acid	CCSCCSP(=S)(OCC)OCC
165	JRC-000244	115-29-7	endosulfan	O=S3OCC2C(CO3)C1(C(C2(Cl)C(=C1Cl)Cl)(Cl)Cl)Cl
166	JRC-000245	39515-41-8	cyclopropanecarboxylic acid	N#CC(c2cccc(c2)Oc3ccccc3)OC(=O)C1C(C1(C)C)(C)C
167	JRC-000246	77-06-5	gibberellic acid	OC(=O)C1C2C3(C4C51CC(=C)C(C5)(O)CC4)C=CC(C2(C)C(=O)O3)O
168	JRC-000247	50-21-5	lactic acid	OC(=O)C(O)C
169	JRC-000248	554-13-2	lithium carbonate	-
170	JRC-000249	103-85-5	phenyl-2-thiourea	NC(=S)Nc1ccccc1
171	JRC-000250	94-13-3	benzoic acid	CCCOC(=O)c1ccc(cc1)O
172	JRC-000251	7784-46-5	arsenenous acid, sodium salt (1:1)	-
173	JRC-000252	7789-12-0	chromic acid (H2Cr2O7)	-
174	JRC-000253	7681-52-9	sodium hypochlorite	-
175	JRC-000254	62-76-0	ethanedioic acid, sodium salt (1:2)	[Na]OC(=O)C(=O)O[Na]
176	JRC-000255	76-03-9	trichloroacetic acid	OC(=O)C(Cl)(Cl)Cl
177	JRC-000256	51-18-3	tretamine, triethylenemelamine	C2CN2c3nc(nc(n3)N4CC4)N1CC1
178	JRC-000257	56-23-5	carbon tetrachloride	ClC(Cl)(Cl)Cl
179	JRC-000258	7758-99-8	cupric sulfate	-
180	JRC-000259	6385-62-2	dipyrido[1,2-a:2',1'-c]pyrazinium	c3ccn2(H)c(c3)c1ccccn1(H)CC2

European Commission
EUR 25473 – Joint Research Centre – Institute for Health and Consumer Protection

Title: An investigation into the use of computational and in vitro methods for acute systemic toxicity prediction

Author(s): Hedvig Norlén, Elisabet Berggren, Maurice Whelan, Andrew Worth

Luxembourg: Publications Office of the European Union

2012 – 40 pp. – 21.0 x 29.7 cm

EUR – Scientific and Technical Research series – ISSN 1018-5593 (print), ISSN 1831-9424 (online)

ISBN 978-92-79-26058-2 (print)
ISBN 978-92-79-21708-1 (pdf)
doi:10.2788/44144

Abstract

We have assessed the abilities of five alternative (non-animal) approaches to predict acute oral toxicity, a toxicological endpoint relevant to multiple pieces of legislation on chemicals and consumer products. In particular, we have investigated four QSAR models (ToxSuite, TOPKAT, TEST and ADMET Predictor) and one in vitro method (3T3 NRU). Based on a test set of in vitro and in vivo data for 180 compounds, we have characterized the predictive performance of each method when used alone (both for LD50 prediction and acute toxicity classification into three categories), as well as multiple test combinations (batteries) and stepwise testing strategies (for acute toxicity classification into three categories). When used individually, the alternative methods showed an ability to predict LD_{50} with correlation coefficients in the range from 49% to 84%, and to classify into three toxicity groups with accuracies in the range from 41% to 72%. When the alternative methods were combined into batteries or testing strategies, the overall accuracy of prediction could reach 76%. We also illustrate how different combinations of methods can be used to optimize sensitivity or specificity.

LB-NA-25473-EN-C

As the Commission's in-house science service, the Joint Research Centre's mission is to provide EU policies with independent, evidence-based scientific and technical support throughout the whole policy cycle.

Working in close cooperation with policy Directorates-General, the JRC addresses key societal challenges while stimulating innovation through developing new standards, methods and tools, and sharing and transferring its know-how to the Member States and international community.

Key policy areas include: environment and climate change; energy and transport; agriculture and food security; health and consumer protection; information society and digital agenda; safety and security including nuclear; all supported through a cross-cutting and multi-disciplinary approach.

Publications Office

ISBN 978-92-79-26058-2

9 789279 260582