[This Page Intentionally Left Blank]

6.0 ACCURACY OF THE 3T3 AND NHK NRU TEST METHODS

This section discusses the accuracy of the 3T3 and NHK NRU test methods for predicting the rodent acute oral toxicity (the LD_{50}) of chemicals. Accuracy, the agreement between a test result and an accepted reference value, is a critical component of the evaluation of the validation status of a method (ICCVAM 2003). Although the 3T3 and NHK NRU test methods are not suitable as replacements for acute oral toxicity assays, the rationale for evaluating the accuracy of LD_{50} predictions from the *in vitro* IC_{50} values is that the animal savings produced by using these *in vitro* test methods to predict starting doses for acute oral toxicity assays will be greatest when the starting dose is as close as possible to the "true" LD_{50} value (see **Section 10** for the evaluation of the potential reduction of animal use).

The ability of the 3T3 and NHK NRU test methods to correctly predict rodent acute oral toxicity is based on the validity of the *in vivo – in vitro* (i.e., IC_{50}-LD_{50}) regression model. The IC_{50}-LD_{50} regression establishes the relationship between the *in vitro* IC_{50} values and the LD_{50} values that will be used to set the starting doses for the computer-simulated acute oral toxicity assays in this study (see **Section 10**). The regressions generated by the three laboratories for each NRU test method were not statistically different, and the data from the 3T3 and NHK NRU test methods were combined (using a geometric mean IC_{50} of the three individual laboratory geometric mean IC_{50} values) into single regressions (see **Section 6.1**). Only rat LD_{50} data were used for these regressions to reduce the variation that would be produced by combining data from multiple species. **Table 6-1** describes the datasets used for the analyses in **Sections 6.1** through **6.4**.

To test the assumption in the *Guidance Document* that the RC millimole regression can be obtained using a basal cytotoxicity method with a single cell type and cytotoxicity endpoint (ICCVAM 2001b), the regressions for each NRU test method (3T3 and NHK) were compared with regressions for the same substances that were calculated using the RC IC_{50} and LD_{50} values (see **Section 6.1**). Because the 3T3 and NHK regressions were not statistically different from the RC regressions for the same chemicals, the RC data were used to develop a regression to predict LD_{50} values from the NRU-generated IC_{50} values because this regression was based on a larger number of substances than the NICEATM/ECVAM regressions (see **Section 6.3**).

The RC millimole regression was used to identify outlier substances (i.e., those that did not fit the regression within the established acceptance limits; see **Section 6.2**) tested in the validation study because:

- Acceptance limits for the RC millimole regression had been established
- The 3T3 and NHK NRU IC_{50} – rat oral LD_{50} regressions were not significantly different from the RC regressions calculated for the same substances
- Use of the RC regressions allow a comparison of the outlier substances determined using RC data to those determined using the 3T3 and NHK data

Table 6-1 Datasets Used for Accuracy Analyses[1]

Use	3T3 NRU[1]	NHK NRU[1]	Characteristics of Dataset
Testing with NRU test methods	72	72	Substances tested; 58 substances were common to the RC
Comparison of laboratory IC_{50}-LD_{50} regressions to one another	47	51	RC substances with IC_{50} values from all laboratories and reference rat oral reference LD_{50} values
Comparison of combined-laboratory IC_{50}-LD_{50} regressions to a regression calculated with RC data	47	47	RC substances with IC_{50} values for both test methods from all laboratories and reference rat oral LD_{50} values
RC millimole regression	NA	NA	RC IC_{50} (mM) and RC oral LD_{50} (mmol/kg) values for 347 substances (282 rat and 65 mouse LD_{50} values)
RC rat-only millimole regression	NA	NA	RC IC_{50} (mM) and RC oral LD_{50} values (mmol/kg) for 282 substances with rat oral LD_{50} data
RC rat-only weight regression	NA	NA	RC IC_{50} (µg/mL) and RC oral LD_{50} values (mg/kg) for 282 substances with rat oral LD_{50} data
Analysis of outliers for the RC millimole regression	70	71	Substances with IC_{50} values from at least one laboratory
Prediction of GHS accuracy using IC_{50} values in RC rat-only regressions	67	68	Substances with IC_{50} values from at least one laboratory and rat oral LD_{50} referene values

Abbreviations: RC=Registry of Cytotoxicity; 3T3=BALB/c 3T3 fibroblasts; NA=Not applicable; NHK=Normal human epidermal keratinocytes; NRU=Neutral red uptake.
[1]Number of substances.

To improve upon the RC millimole regression's[1] ability to accurately predict LD_{50} values from IC_{50} values, and to also make this approach relevant to the testing of mixtures and substances without known molecular weights, two regressions were calculated (see **Section 6.3**). The first regression – the RC rat-only millimole regression – uses the 282 (of 347) substances in the RC dataset that had reported rat LD_{50} values. The LD_{50} data for the regression were limited to one species to decrease the variability in LD_{50} values that would occur if the data from more than one species were combined. Rats were selected because they are the preferred species for acute oral toxicity testing (EPA 2002b; OECD 2001a; OECD 2001d) (see **Section 6.3.1**). The RC rat-only millimole regression was transformed to one based on weight units (mg/kg body weight for LD_{50} and µg/mL for IC_{50}) in order to make the regression equation more generally applicable to the testing of mixtures and substances of unknown molecular weights.

[1] The RC millimole regression was created using rat and mouse oral LD_{50} values from RTECS® and IC_{50} values from *in vitro* cytotoxicity assays using multiple cell lines and cytotoxicity endpoints for 347 substances with known molecular weights (Halle 1998, 2003)

The ability of the 3T3 and NHK NRU IC_{50} data to correctly predict rat acute oral LD_{50} values based on using the RC rat-only millimole regression and the RC rat-only weight regression, was evaluated by determining the extent to which the appropriate GHS acute oral toxicity category was identified for each reference substance (see **Section 6.4**). The rationale for evaluating the accuracy of LD_{50} predictions is that the acute oral toxicity test methods (i.e., UDP, FDP, and ATC) call for starting doses to be placed as close as possible and just below the true LD_{50}. When the starting dose is close to the true LD_{50} for a test substance, fewer animals are needed. When the starting dose is below the true LD_{50}, there is reduced pain and suffering because doses tend to be lower, and the test bias is more conservative. This approach permits an assessment of accuracy that is specific to each GHS hazard classification category. The discordant reference substances from the predictions of GHS category are presented in **Appendix L2**.

The remainder of **Section 6** discusses physical, chemical, and biological, characteristics of substances that may have an impact on the accuracy of the 3T3 and NHK methods.

6.1 Accuracy of the 3T3 and NHK NRU Test Methods for Predicting Rodent Acute Oral Toxicity

The rat LD_{50} values provided in **Section 4.2** are used as the reference values for assessing the ability of the 3T3 and NHK test methods to accurately predict acute oral toxicity[2]. The accuracy of the two *in vitro* cytotoxicity test methods is assessed in two ways: (1) by the goodness of fit of the *in vitro* IC_{50} data to the rat LD_{50} data in linear regression analyses, and (2) by the concordance (i.e., extent of agreement) between the GHS acute oral toxicity categories (UN 2005) assigned based on rat LD_{50} data and those predicted using *in vitro* IC_{50} values.

6.1.1 Linear Regression Analyses for the Prediction of Rat Acute Oral LD_{50} Values from *In Vitro* IC_{50} Values

As described in **Section 5.5.4.3**, linear regressions for each laboratory and *in vitro* method were calculated using log IC_{50} values (mM) versus the corresponding reference log LD_{50} values (mmol/kg) identified in **Table 4-2**. The reference substances used to calculate each of the laboratory regressions met the following criteria for each test method:

- The substance was included in the RC
- All three laboratories reported IC_{50} values
- There was an associated rat acute oral LD_{50} reference value (see **Table 4-2**).

There were 47 and 51 reference substances that fit these criteria for the 3T3 and NHK test methods, respectively. The slopes for the all of the laboratory-specific regressions were statistically significantly different from zero ($p < 0.0001$), which indicates a significant correlation between *in vitro* IC_{50} values and the corresponding rat acute oral LD_{50} values. Comparison of the individual laboratory regressions to one another using the goodness of fit F-test for regression slopes and intercepts described in **Section 5.5.4.3** indicated that the

[2] Toxicity is inversely proportional to LD_{50}. High LD_{50} values reflect low toxicity and low LD_{50} values reflect high toxicity

laboratory-specific regressions for either NRU method were not significantly different from one another. For the 3T3 method, p=0.605 for the slope comparisons and p=0.947 for the intercept comparisons. For the NHK method, p=0.792 for the slope comparisons and p=0.999 for the intercept comparisons.

Because the individual laboratory regressions were not significantly different, the laboratory data were combined into a single regression for each method using the geometric mean of the mean IC_{50} values determined by each laboratory for each substance (see the "Combined-laboratory" regressions in **Table 6-2** and **Figure 6-1**). The combined-laboratory 3T3 regression yielded a better fit to the reference LD_{50} data (R^2=0.579) than the NHK regression (R^2=0.463).

Table 6-2 Linear Regression Analyses of the 3T3 and NHK NRU and Rat Acute Oral LD_{50} Test Results[1]

Laboratory	N	Slope	Intercept	R^2
3T3 NRU				
ECBC[2]	47	0.573	0.541	0.613
FAL[2]	47	0.539	0.373	0.519
IIVS[2]	47	0.552	0.507	0.586
Combined-laboratory[3]	47	0.561	0.475	0.579
NHK NRU				
ECBC[2]	51	0.491	0.412	0.480
FAL[2]	51	0.428	0.407	0.422
IIVS[2]	51	0.483	0.416	0.478
Combined-laboratory[3]	51	0.470	0.413	0.463

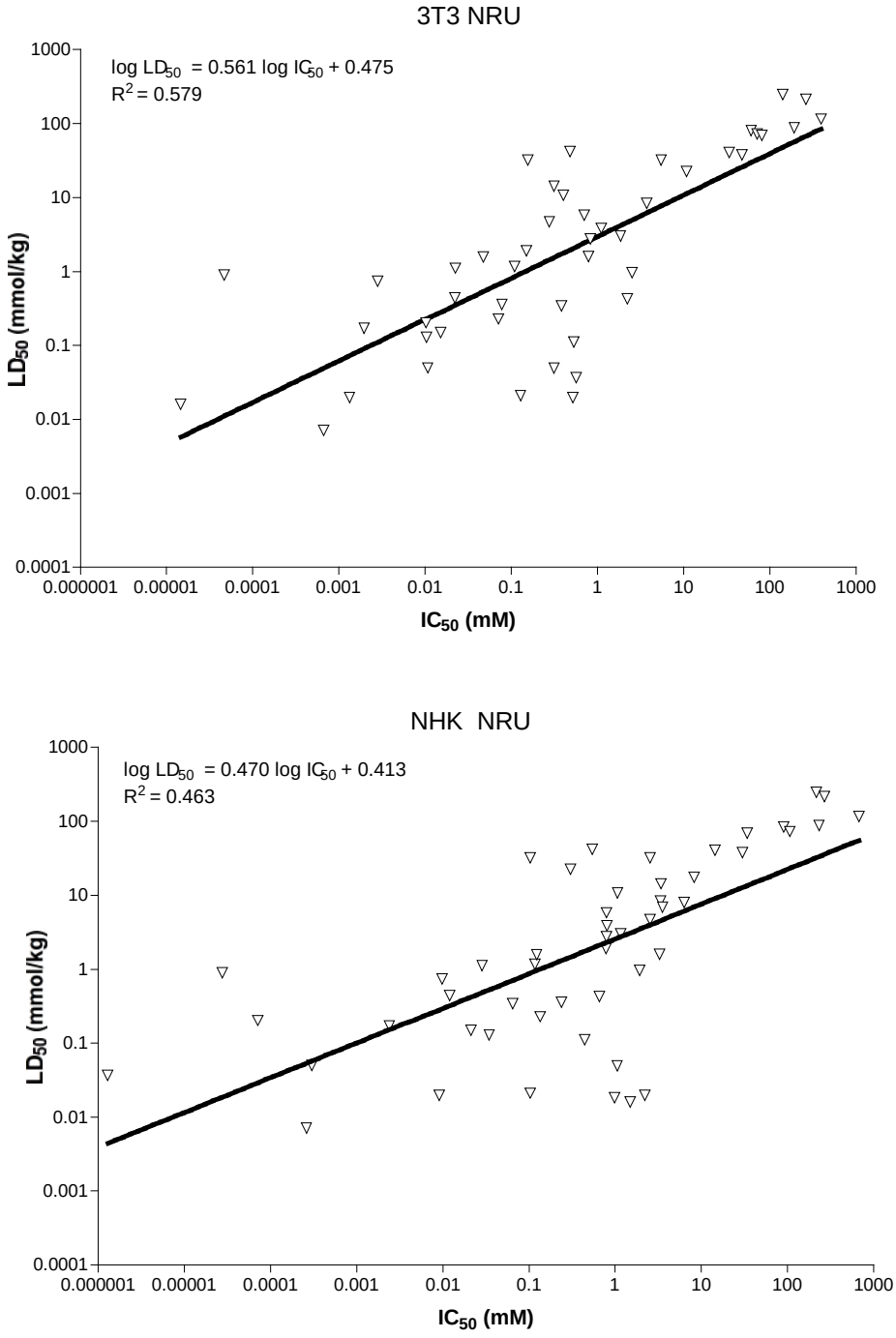
Abbreviations: ECBC=Edgewood Chemical Biological Center; FAL=Fund for the Replacement of Animals in Medical Experiments Alternatives Laboratory; IIVS=Institute for *In Vitro* Sciences; 3T3=BALB/c 3T3 fibroblasts; N=Number of substances used to calculate the regression; NHK=Normal human epidermal keratinocytes; NRU=Neutral red uptake; R^2=Coefficient of determination.

[1]Log IC_{50} in mM; log LD_{50} in mmol/kg.

[2]Regression based on a single point per substance (i.e., the geometric mean of the within laboratory replicate IC_{50} values and the reference rat acute oral LD_{50} from **Table 4-2**).

[3]Regression based on a single point per substance (i.e., the geometric mean of the geometric mean IC_{50} values obtained for each laboratory and the reference rat acute oral LD_{50} from **Table 4-2**).

Figure 6-1 Combined-Laboratory 3T3 and NHK NRU Regressions



Abbreviations: 3T3=BALB/c 3T3 fibroblasts; NHK=Normal human epidermal keratinocytes; NRU=Netural red uptake; R^2=Coefficient of determination.
Points show the geometric means of the laboratory geometric mean IC_{50} values and the reference rat acute oral LD_{50} values (from **Table 4-2**) for 47 reference substances for the 3T3 and 51 reference substances for NHK test methods. Solid lines show the combined-laboratory regressions for each method (see **Table 6-2**).

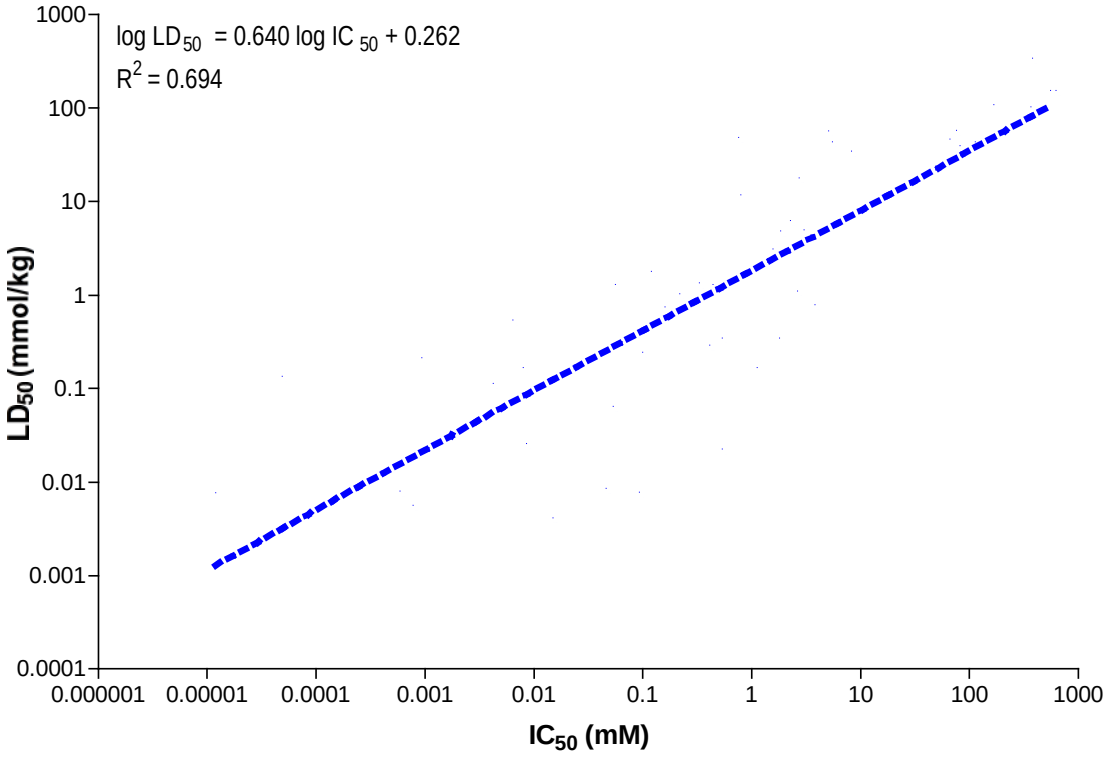
6.1.2 Comparison of the Combined-Laboratory 3T3 and NHK Regressions to the RC Millimole Regression

The validation study tested 58 RC substances using the 3T3 and NHK NRU test methods (see **Figure 3-1**). The resulting method regressions for each cell type were compared to the RC regressions for the same substances to test the assumption in the *Guidance Document* that the RC millimole regression can be obtained with a basal cytotoxicity test method using a single cell type and endpoint (ICCVAM 2001b). The 47 substances used to calculate these regressions met the following criteria:

- The substance was included in the RC
- All three laboratories reported IC_{50} values for both the 3T3 and NHK NRU test methods
- There was an associated rat oral reference LD_{50} value (see **Table 4-2**)

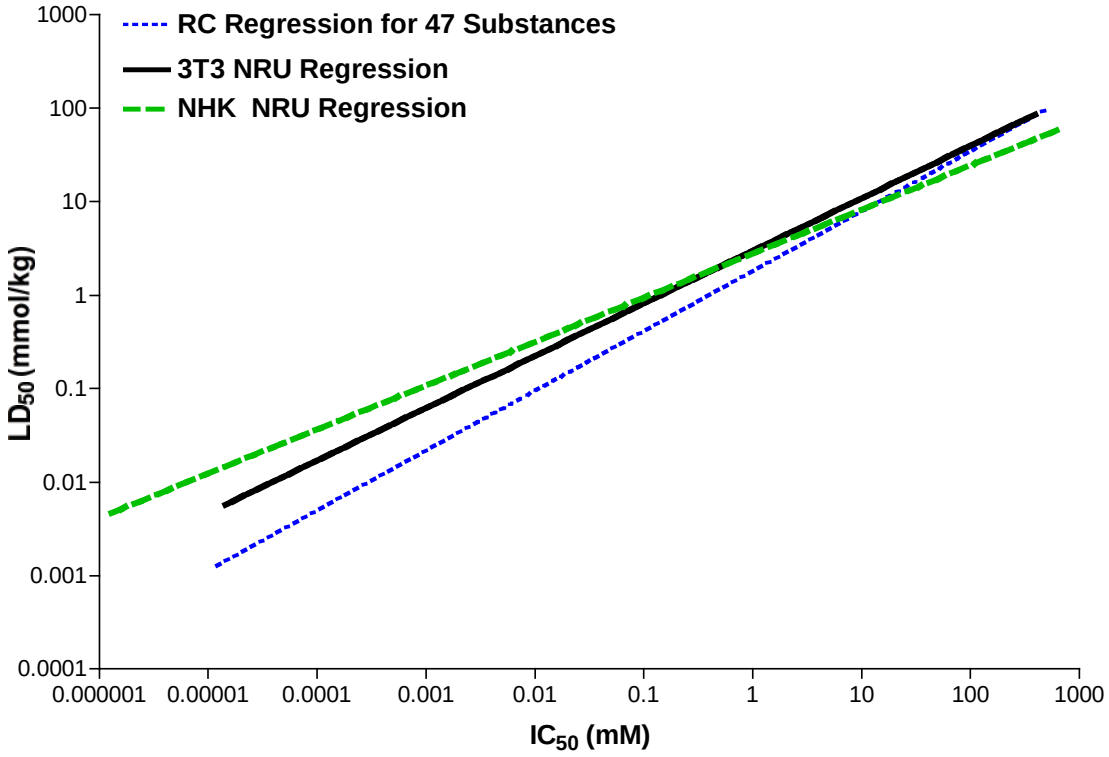
The regression calculated for the 47 substances using the RC IC_{50} and LD_{50} data is shown in **Figure 6-2**. A graphic comparison of the RC regressions and the 3T3 and NHK combined-laboratory regressions is in **Figure 6-3**. A statistical comparison of slope and intercept (simultaneously) using an F test showed that neither the 3T3 regression (p=0.612) nor the NHK regression (p=0.759) was significantly different from the 47 RC substance regression.

Figure 6-2 Regression for 47 RC Substances Using RC Data



Abbreviations: RC=Registry of Cytotoxicity; R^2=Coefficient of determination.
Points show the IC_{50} values and the reference rodent (rat and mouse) acute oral LD_{50} values from the RC for 47 reference substances. The dashed line shows the calculated regression.

Figure 6-3 Regression for 47 RC Substances with the 3T3 and NHK Regressions



Abbreviations: RC=Registry of Cytotoxicity; 3T3=BALB/c 3T3 fibroblasts; NHK=Normal human epidermal keratinocytes; NRU=Neutral red uptake.
The regression for 47 RC substances using RC data is log LD_{50} = 0.640 log IC_{50} + 0.262 (R^2=0.694). The combined-laboratory 3T3 regression for the same 47 substances, is log LD_{50} = 0.561 log IC_{50} + 0.475 (R^2 = 0.579) (from **Table 6-2**). The combined-laboratory NHK regression for the same 47 substances, is log LD_{50} = 0.471 log IC_{50} + 0.445 (R^2 = 0.487).

6.2 Analysis of Outlier Substances for the RC Millimole Regression

The RC millimole regression and each *in vitro* NRU test method were used to identify outliers among the reference substances tested in the validation study (i.e., those for which the rodent LD_{50} was not accurately predicted by the *in vitro* IC_{50}). The outlier substances were then evaluated to determine if they had common characteristics that could assist in identifying the types of substances that are not suited for use in the 3T3 or NHK NRU test methods for determining starting doses for acute oral toxicity assays.

The RC millimole regression was used to determine the outlier status of reference substances because:

- The RC millimole regression had associated acceptance limits (Halle 1998, 2003): a difference greater than 0.699 (or log 5) for log-observed LD_{50} (in mmol/kg) from the log-predicted LD_{50} identifies a substance as an outlier
- The 3T3 and NHK IC_{50} – rat oral LD_{50} regressions were not significantly different from the RC regressions calculated for the same substances

- Use of the RC millimole regression allows a comparison of the outlier substances determined using RC IC_{50} values to those determined using the 3T3 and NHK NRU IC_{50} values.

6.2.1 Identification of Outlier Substances

For each *in vitro* NRU test method, the predicted LD_{50} values for the reference substances were determined using the geometric mean IC_{50} values of the three geometric mean laboratory values in the RC millimole regression. Outliers were identified using the RC method (Halle 1998): a difference greater than 0.699 (or log 5) for log-observed LD_{50} (in mmol/kg) minus the log-predicted LD_{50} identifies a substance as an outlier (see **Appendix J1** for the 3T3 NRU test method and **Appendix J2** for the NHK NRU test method for the predicted LD_{50} values). For the best comparison with the RC outlier results, the outlier evaluation for the 3T3 and NHK NRU test methods used same observed LD_{50} values as those used in the RC database for the 58 reference substances that were included in the RC database (see **Table 3-2)**. For the non-RC substances, the observed values (in **Table 3-2**) were obtained from other databases such as RTECS® or Hazardous Substances Database (NLM 2002). The outlier analysis included all the reference substances that yielded IC_{50} values from at least one laboratory in the validation study whether the *in vivo* LD_{50} values were from rats or mice. Thus, 70 substances were used for the 3T3 NRU outlier analysis and 71 substances were used for the NHK NRU outlier analysis. **Table 6-3** lists the outlier substances for the RC millimole regression when using the RC IC_{50} values and the 3T3 and NHK NRU IC_{50} values.

Table 6-3 Outlier Substances for the RC and the 3T3 and NHK NRU Methods When the RC Millimole Regression is Used[1]

Substances Included in the RC Identified as Outliers in:		
RC[2]	**3T3[3]**	**NHK[4]**
	Acetaminophen (+)	
	Arsenic III trioxide (–)	*Arsenic III trioxide (–)*
		Aminopterin (–)
5-Aminosalicylic acid (+)		5-Aminosalicylic acid (+)
Busulfan (–)	**Busulfan (–)**	**Busulfan (–)**
Caffeine (–)		Caffeine (–)
Cycloheximide (–)	**Cycloheximide (–)**	**Cycloheximide (–)**
Dibutyl phthalate (+)	***Dibutyl phthalate (+)***	***Dibutyl phthalate (+)***
	Diethyl phthalate (+)	***Diethyl phthalate (+)***
Digoxin (–)	***Digoxin (–)***	
Disulfoton (–)	***Disulfoton (–)***	***Disulfoton (–)***
Epinephrine bitartrate (–)	Epinephrine bitartrate (–)	Epinephrine bitartrate (–)
Ethanol (+)	**Ethanol (+)**	**Ethanol (+)**
Lindane (–)	*Lindane (–)*	
Mercury II chloride (–)	Mercury II chloride (–)	Mercury II chloride (–)
		Methanol (+)

Table 6-3 Outlier Substances for the RC and the 3T3 and NHK NRU Methods When the RC Millimole Regression is Used[1]

Substances Included in the RC Identified as Outliers in:		
RC[2]	**3T3[3]**	**NHK[4]**
Nicotine (–)	Nicotine (–)	Nicotine (–)
Paraquat (–)		Paraquat (–)
Parathion (–)	***Parathion* (–)**	***Parathion* (–)**
Phenobarbital (–)	Phenobarbital (–)	Phenobarbital (–)
Phenylthiourea (–)	Phenylthiourea (–)	Phenylthiourea (–)
Potassium cyanide (–)	Potassium cyanide (–)	Potassium cyanide (–)
Propylparaben (+)	Propylparaben (+)	Propylparaben (+)
		Sodium oxalate (–)
Thallium I sulfate (–)	Thallium I sulfate (–)	
Triethylenemelamine (–)	Triethylenemelamine (–)	Triethylenemelamine (–)
1,1,1-Trichloroethane (+)		
Verapamil HCl (–)	**Verapamil HCl (–)**	***Verapamil HCl* (–)**
		Xylene (+)
Outliers That Were Not Included in the RC		
	Dichlorvos (–)	Dichlorvos (–)
	Endosulfan (–)	Endosulfan (–)
	Fenpropathrin (–)	*Fenpropathrin* (–)
	Physostigmine (–)	Physostigmine (–)
	Sodium hypochlorite (+)	Sodium hypochlorite (+)
	Sodium selenate (–)	Sodium selenate (–)
	Strychnine (–)	*Strychnine* (–)

Abbreviations: RC=Registry of Cytotoxicity; 3T3=BALB/c 3T3 fibroblasts; NHK=Normal human epidermal keratinocytes; NRU=Neutral red uptake; (–)=Toxicity was underpredicted by the IC_{50} and RC millimole regression (i.e., the LD_{50} value predicted by the IC_{50} was higher than the *in vivo* LD_{50} value); (+)=Toxicity was overpredicted by the IC_{50} and RC millimole regression (i.e., the LD_{50} value predicted by the IC_{50} was lower than the *in vivo* rodent LD_{50} value).
[Note: Empty cells indicate that the substance was not an outlier for that particular IC_{50} value.]
[1]Log LD_{50} (mmol/kg) = 0.435 log IC_{50} (mM) + 0.625. Log LD_{50} (mmol/kg) values for outlier substances were >0.699 from the RC millimole regression.
[2]Using RC IC_{50} in the RC millimole regression for the 58 RC substances tested in the validation study.
[3]Using the 3T3 NRU IC_{50} in the RC millimole regression for the 70 reference substances that yielded IC_{50} values from any laboratory in the validation study.
[4]Using the NHK NRU IC_{50} in the RC millimole regression the RC for the 71 reference substances that yielded IC_{50} values from any laboratory in the validation study.
Bolded substances have active metabolites *in vivo* (see **Table 3-7**).
Substances that showed evidence of insolubility (i.e., precipitates) during testing (see **Table 5-11**) are identified by italics.

When the RC millimole regression and the RC method of identifying outlier substances were used (Halle 1998, 2003), there were 28 outliers for the 3T3 NRU test method and 31 for the NHK NRU test method. The top part of **Table 6-3** shows a comparison of the 22 RC substances that were identified by the RC as outliers (see **Table 3-2**) and the RC reference substances that were identified as outliers using either the 3T3 or NHK NRU IC_{50} values with the RC millimole regression. For the 58 RC substances that were tested in the validation

study, 18 of the 22 RC outliers also responded as outliers in both NRU test methods, but some of the substances were outliers only in one of the two NRU test methods. The RC regression outliers, 5-aminosalicylic acid, caffeine, paraquat, and 1,1,1-trichloroethane were not outliers when 3T3 data were used, and the RC outliers, digoxin, lindane, thallium sulfate, and 1,1,1-trichloroethane, were not outliers when the NHK NRU test method was used. In contrast the 3T3 NRU test method identified three substances as outliers that were not identified by the RC: acetaminophen, arsenic trioxide, and diethyl phthalate, and the NHK NRU test method identified six: aminopterin, arsenic trioxide, diethyl phthalate, methanol, sodium oxalate, and xylene. Seven additional substances, that were not included in the RC database, were identified as outliers using the NRU IC_{50} values in the RC millimole regression: dichlorvos, endosulfan, fenpropathrin, physostigmine, sodium hypochlorite, sodium selenate, and strychnine.

6.2.2 Evaluation of Outlier Substances

A number of physico-chemical and toxicologic characteristics were evaluated for their frequency of occurrence among the 28 and 31 outlier substances in the 3T3 and NHK NRU test methods, respectively, to identify attributes that may have contributed their outlier status. This section provides a summary of these analyses based on the RC millimole regression and outlier criteria. The frequency of outliers versus the total number of reference substances for each physico-chemical and toxicologic category examined is shown in **Appendix L1**.

6.2.2.1 *Physical Characteristics*

A number of physical characteristics were evaluated for their frequency of occurrence in the set of outlier substances versus the complete set of reference substances. The characteristics chosen were those that were assumed to be readily available, or relatively easy to measure, for new substances that may be tested in these NRU assays. The characteristics examined included chemical class, molecular weight, boiling point, IC_{50}, pH, and log K_{ow} (i.e., log octanol:water partition coefficient). Unfortunately, these attributes were not available for all substances. For example, log K_{ow} was available for 50 of the 70 (71%) substances evaluated for the 3T3 NRU test method and for 51 of the 71 (72%) substances evaluated for the NHK NRU test method. Boiling point was available for only 24 of 70 (34%) substances evaluated for the 3T3 NRU test method and for 25 of the 71 (35%) substances evaluated for the NHK NRU test method. For substances with log K_{ow} >3.00, 8/13 (62%) were outliers for both the 3T3 and NHK test methods. For molecular weights >400 g/mole, 4/7 (57%) substances were outliers using the 3T3 NRU test method and 3/7 (43%) were outliers using the NHK NRU test method. For substances with boiling points >200^{o}C, 9/13 (69%) were outliers using the 3T3 NRU test method and 8/13 (62%) were outliers using the NHK NRU test method.

6.2.2.2 *Chemical Class*

Examination of outliers by chemical class for the RC millimole regression showed that all of the chemical classes that contained at least three reference substances also contained at least one outlier for one test method. Two classes contained 100% outliers for both test methods: organophosphates (3/3) and organic sulfur compounds (5/5). The remaining classes with higher frequencies of outliers included: 2/3 (67%) amines were outliers for both test methods, 7/14 (50%) heterocylics were outliers for the 3T3 NRU and 10/14 (71%) heterocyclics were outliers for the NHK NRU, 2/5 (40%) chlorine compounds were outliers for both test methods, 2/6 (33%) sodium compounds were outliers for both test methods, 3/9 (33%) alcohols were outliers for the 3T3 NRU and 4/10 (40%) alcohols were outliers for the NHK

NRU, and 4/14 (29%) carboxylic acids were outliers for the 3T3 NRU and 6/14 (43%) carboxylic acids were outliers for the NHK NRU.

6.2.2.3 *Solubility*

Another attribute that may cause a substance to be an outlier is the lack of solubility in the test system. Because the SMT expected the toxicity of insoluble substances to be underpredicted in the *in vitro* assays, substances that formed precipitates in the tests were noted and compared with the outlier substances. However, insolubility was not consistently associated with the outlier substances for which toxicity was underpredicted. For example, eight of the 22 (36%) underpredicted substances identified by applying the 3T3 results to the RC millimole regression exhibited signs of insolubility in at least one laboratory. NHK results showed that seven of 23 (30%) underpredicted substances exhibited signs of insolubility in at least one laboratory (see **Table 5-11** for substances that had precipitates in the assays). Additionally, there was evidence of insolubility in the 3T3 and NHK NRU test methods of dibutyl phthalate and diethyl phthalate, but toxicity was overpredicted for both substances, rather than underpredicted. This overprediction may be a characteristic of the phthalates, but more substances would have to be tested before a general rule could be adopted.

There were 25 substances that showed evidence of insolubility in the 3T3 test method in at least one laboratory, and 11 (44%) of these were outliers. Of the 24 substances showed evidence of insolubility in at least one NHK laboratory, 11 (46%) were outliers.

6.2.2.4 *Metabolism*

It was anticipated that the toxicity of substances metabolized *in vivo* to active compounds (see **Section 3.3.4.3** and **Table 3-7**) would be underpredicted *in vitro* by 3T3 and NHK cells, which have little or no metabolic capability (Babich 1991; INVITTOX 1991). Of the 72 reference substances, 19 (26%) are known to have active metabolites *in vivo*, and 10 (45%) of these were classified as outliers for 3T3. Of these 10 substances, which accounted for 36% of the 28 outlier substances, the toxicity of six (60%) was underpredicted, while the toxicity of four (40%) was overpredicted. Among the 31 outliers in the NHK NRU test method, nine (29%) are metabolized to active metabolites. Nine of the 19 substances known to produce active metabolites *in vivo* were discordant for the NHK NRU test method. NHK cells underpredicted the toxicity of five (56%) of these nine substances and overpredicted the other four (44%). These nine outlier substances accounted for 29% of the 31 outliers in the NHK NRU test method. Thus, the fact that a substance has active metabolites that are not expected to be produced in the *in vitro* tests does not necessarily indicate that its toxicity will be underpredicted by *in vitro* basal cytotoxicity test methods.

Similarly, Halle (1998, 2003) noted that the RC substances that required metabolic activation to produce *in vivo* toxicity were not necessarily outliers with respect to their fit to the RC millimole regression. They found that eight (50%) of the 16 substances that required metabolic activation to product toxicity were outliers (see **Table L3-3** in **Appendix L3**).

6.2.2.5 *Mechanism of Toxicity*

Substances whose mechanisms of toxicity would not be detected in the 3T3 or NHK cells would be expected to fit the RC millimole regression poorly. In particular, toxic mechanisms that include, for example, specific actions on the central nervous system (CNS) or the heart

are not expected to be active in the 3T3 or NHK cells. Neurotoxic mechanisms would include, for example, cholinesterase inhibition, CNS nicotinic receptor blockade or activation, or any activity other than membrane destabilization such as that produced by a solvent, or disturbance of energy utilization such as interruption of oxidative phosphorylation. Representative cardiotoxic mechanisms would include calcium channel blockage and beta-adrenergic receptor activation or blockage.

The 72 reference substances used to validate the 3T3 and NHK NRU test methods included 16 (22%) that had specific CNS toxicity (see **Table 6-4**). Of these 16 substances, 10 (63%) were outliers in both *in vitro* NRU test methods. Three of the six (50%) reference substances that are cardiotoxic were outliers in the 3T3 NRU test method and two (33%) were outliers in the NHK NRU test method. When all the reference substances with mechanisms that are not expected to be active in the 3T3 and NHK cells (i.e., in **Table 6-4**) are summed, 13/22 (59%) are outliers for the 3T3 NRU and 12/22 (55%) are outliers for the NHK NRU. These substances represented 13/28 (46%) and 12/31 (39%) of the total outlier substances for the 3T3 and NHK NRU test methods, respectively. Halle (1998, 2003) reported similar findings for the RC database (i.e., approximately half of the substances expected to be outliers based on their mechanisms of toxicity were outliers) (see **Appendix L3**).

Table 6-4 Substances With Mechanisms of Toxicity Not Expected to Be Active in the 3T3 or NHK Cells in Culture

Substance	Mechanism of Toxicity[1]	3T3 Outlier[2]	NHK Outlier[2]
Neurotoxic			
Atropine sulfate	Antimuscarinic; anticholinergic action; competitive antagonism of anticholinesterase at cardiac and CNS receptor sites.	No	No
Caffeine	Inhibition of phosphodiesterase leading to AMP accumulation; translocation of intracellular Ca^{++}; adenosine receptor antagonism; neurotoxic.	No	Yes
Carbamazepine	Therapeutically decreases firing of noradrenergic neurons.	No	No
Chloral hydrate	Potentiation of $GABA_A$ receptor activity; inhibition of N-methyl-D-aspartate activity; modulation of 5-hydroxytryptamine$_3$ receptor-mediated depolarization of the vagas nerve[3].	No	No
Dichlorvos	Inhibition of acetylcholinesterase resulting in acetylcholine accumulation in CNS and effector organs.	Yes	Yes
Disulfoton	Inhibition of acetylcholinesterase resulting in acetylcholine accumulation in CNS and effector organs.	Yes	Yes
Endosulfan	Affects brain neurotransmitter levels[4].	Yes	Yes
Fenpropathrin	Delays closure of sodium channel causing persistent depolarization of membrane.	Yes	Yes
Glutethimide	CNS depression; anticholinergic activity.	No	No
Haloperidol	Blocks dopamine receptors.	No	No
Lindane	CNS depression through inhibition of GABA receptor linked chloride channel at the picrotoxin binding site, leading to blockade of chloride influx into neurons.	Yes	No
Nicotine	Cholinergic block causing polarization of CNS and PNS synapses.	Yes	Yes
Parathion	Inhibition of acetylcholinesterase resulting in acetylcholine accumulation in CNS and effector organs.	Yes	Yes
Phenobarbital	CNS depression through inhibition of GABA synapses; inhibits hepatic NADH cytochrome oxidoreductase.	Yes	Yes
Physostigmine	Inhibition of acetylcholinesterase resulting in acetylcholine accumulation in CNS and effector organs.	Yes	Yes
Strychnine	Increases glutamic acid in the CNS.	Yes	Yes
Cardiotoxic			
Amitriptyline HCl	Blocks norepinephrine, 5-hydroxytryptamine, and dopamine presynaptic uptake; prevents reuptake of heart norepinephrine.	No	No
Digoxin	Impairs ion transport and increases sarcoplasmic calcium by binding to Na^+/K^+ ATPase, increasing automaticity of cardiac cells.	Yes	No

Table 6-4 Substances With Mechanisms of Toxicity Not Expected to Be Active in the 3T3 or NHK Cells in Culture

Substance	Mechanism of Toxicity[1]	3T3 Outlier[2]	NHK Outlier[2]
Epinephrine bitartrate	Adrenergic receptor stimulation.	Yes	Yes
Potassium chloride	Disturbs cardiac membrane potential and electrical activity.	No	No
Procainamide HCl	Slows impulse conduction in the heart[5].	No	No
Verapamil HCl	Inhibition of transmembrane Ca^{++} flux in excitatory tissues; alpha-adrenergic blockade.	Yes	Yes

Abbreviations: NA=Not available or information not found; CNS=Central nervous system; GABA=Gamma aminobutyric acid; PNS=Peripheral nervous system; NADH=Nicotine adenine dinucleotide (reduced).
[1]From Ekwall et al. (1998) or Hazardous Substances Data Bank (NLM 2001, 2002) unless otherwise noted.
[2]As shown in **Table 6-3**.
[3]EPA (2000b).
[4]ATSDR (2000a).
[5]Hardman et al. (1996).

6.3 Improving the Prediction of *In Vivo* Rat Oral LD_{50} Values from *In Vitro* IC_{50} Data

Because the 3T3 and NHK IC_{50} – rat oral LD_{50} regressions were not significantly different from the RC regression for the same substances, the next step was an attempt to improve the RC millimole regression for the prediction of LD_{50} values from IC_{50} values. Because the validation study provided results similar to the RC, and because the RC database has more than 3.5 times the number of substances tested in the validation study, the RC rat data (282 substances) were used to determine the relationship between IC_{50} and LD_{50}. The RC data were used to develop two new regressions, the RC rat-only millimole regression and the RC rat-only weight regression. For reference, the original RC millimole regression, log LD_{50} (mmol/kg) = 0.435 x log IC_{50} (mM) + 0.625 (Halle 1998, 2003), is shown in **Table 6-5**.

6.3.1 The RC Rat-Only Millimole Regression

The first regression used the RC data for the 282 substances with rat LD_{50} data and the original units of mM for IC_{50} and mmol/kg for LD_{50} (see **Table 6-5** and **Figure 6-9**). Only rat data were used because:

- Rats and mice are not always equally sensitive to all substances
- The majority of acute oral LD_{50} data used in the RC millimole regression were from studies using rats (282 rat data points versus 65 mouse data points) (Halle 1998, 2003)
- Most acute oral toxicity testing is performed with rats.

The RC rat-only millimole regression is applicable to substances of known molecular weight that are relatively pure.

Table 6-5 Linear Regression Analyses to Improve the Prediction of Rodent Acute Oral LD_{50} Values from *In Vitro* NRU IC_{50} Using the RC Database[1]

Data Used	Slope	Intercept	R^2
347 RC substances (282 rat and 65 mouse LD_{50} values) – millimole units[2]	0.435	0.625	0.452[3]
282 RC substances with rat LD_{50} data – millimole units[2]	0.439	0.621	0.452
282 RC substances with rat LD_{50} data – weight units[4]	0.372	2.024	0.325

Abbreviations: NRU=Neutral red uptake; RC=Registry of Cytotoxicity; R^2=Coefficient of determination.
[1]Slopes of all regressions were significantly different ($p < 0.05$) from zero at $p < 0.0001$.
[2]IC_{50} in mM; LD_{50} in mmol/kg.
[3]Calculated from RC data (i.e., not reported by Halle [1998, 2003]).
[4]IC_{50} in μg/mL; LD_{50} in mg/kg.

Table 6-5 shows that the RC millimole regression using only rat acute oral LD_{50} data was essentially identical to the original regression that used both rat and mouse data. The slope changed from 0.435 to 0.439 and the intercept changed from 0.625 to 0.621; these changes were not statistically significantly different.

6.3.2 The RC Rat-Only Weight Regression

The second regression used the same RC rat acute oral LD_{50} data for the 282 substances but was calculated using weight units rather than millimolar units (see **Table 6-5** and **Figure 6-**

4b). Weight units (i.e., mg/kg for the LD_{50} and μg/mL for the IC_{50}) were selected for the units of measurement because

- Millimole units are not applicable to mixtures and substances with unknown structures or molecular weights.
- They are the most practical, i.e., hazard classification in all regulatory systems is based on LD_{50} values expressed in mg/kg (see **Table 1-2**).

The RC rat-only weight regression is applicable for use with complex mixtures, substances whose structures or molecular weights are unknown, and substances that are relatively impure (i.e., mixtures that are primarily composed of a named substance).

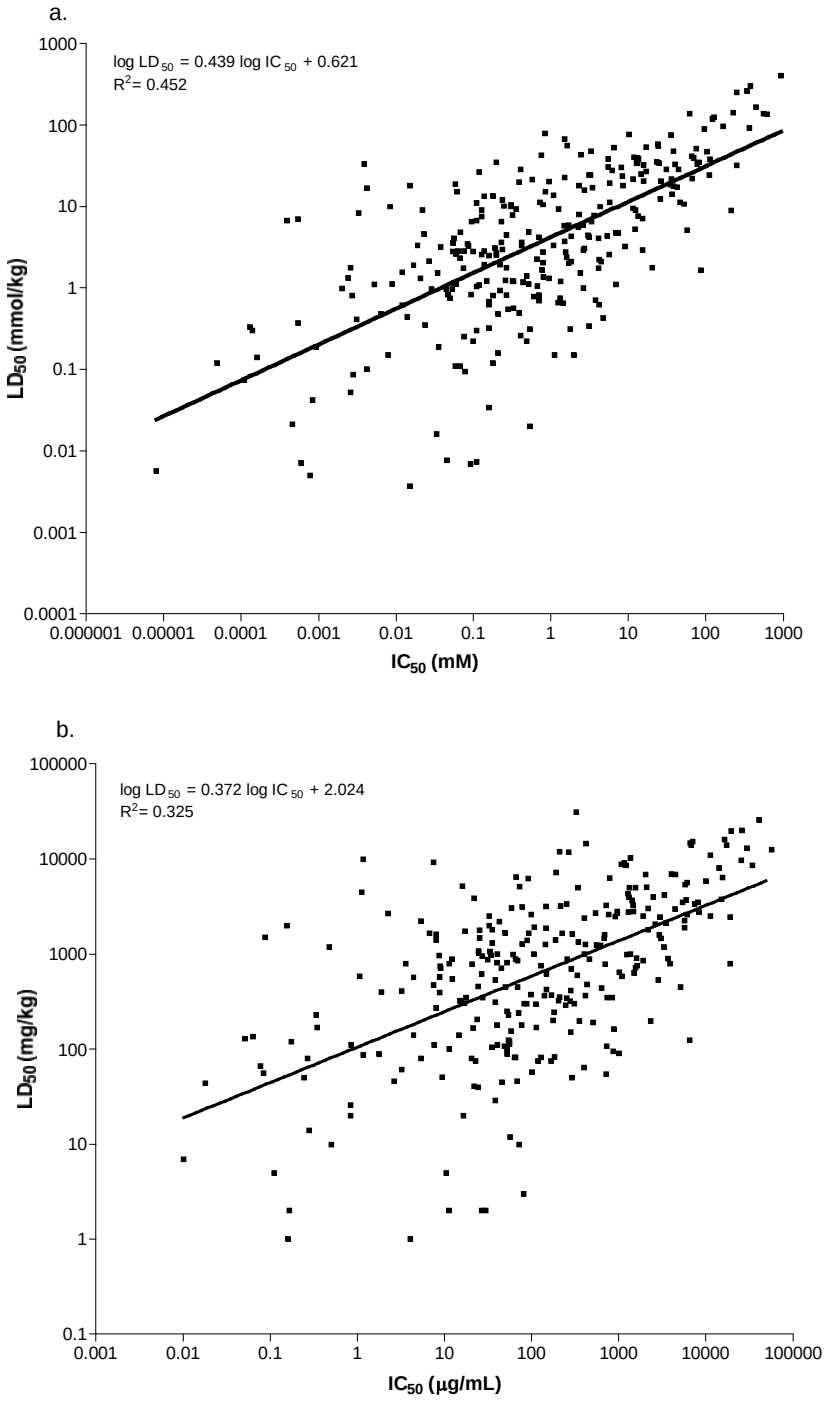
6.4 Accuracy of the 3T3 and NHK NRU Test Methods for Predicting GHS Acute Oral Toxicity Categories

Based on the correlations/regressions obtained between the 3T3 and NHK NRU IC_{50} values and the rat LD_{50} values, it is clear that these *in vitro* methods are not suitable as replacements for rodent acute oral toxicity tests. The use of *in vitro* methods to reduce animal use for rodent acute oral toxicity assays (i.e., to assist in determining the starting doses for *in vivo* assays) also depends upon their accuracy for the prediction of LD_{50} values. However, this latter (adjunct) use does not require the same precision in LD_{50} prediction as complete replacement would.

The NRU-predicted LD_{50} values were determined using the *in vitro* NRU IC_{50} values in the RC rat-only regressions presented in **Table 6-5.** The predicted LD_{50} values were used to assign each substance to a predicted GHS acute oral toxicity category (UN 2005). The accuracy of the 3T3 and NHK NRU test methods for predicting GHS acute oral toxicity categories was determined by comparison with categorization based on rat acute oral LD_{50} data. The rationale for evaluating the accuracy of LD_{50} predictions was that the animal savings produced by using these *in vitro* NRU test methods to predict starting doses for rodent acute oral toxicity assays would be greatest when the starting dose is as close as possible to the LD_{50}. This approach was used because regulatory authorities use rodent acute oral toxicity test results for hazard classification and labelling of products to protect handlers and consumers.

The *in vitro* NRU test methods were evaluated for their ability to predict GHS acute oral toxicity categories using the two regressions presented in **Section 6.3**, the RC rat-only millimole regression and the RC rat-only weight regression. The same reference substances were evaluated for each regression. Sixty-seven and 68 substances were evaluated using the 3T3 and NHK NRU test methods, respectively. Of the original 72 reference substances tested, epinephrine bitartrate, colchicine, and propylparaben were excluded because they had no rat acute oral LD_{50} reference data (see **Table 4-2**). Carbon tetrachloride and methanol were excluded from the 3T3 evaluations because no laboratory attained sufficient toxicity in any test for the calculation of an IC_{50} (see **Table 5-4**). Carbon tetrachloride was excluded from the NHK evaluations because no laboratory attained sufficient toxicity in any test for the calculation of an IC_{50} (see **Table 5-5**).

Figure 6-4 RC Rat-Only Millimole Regression (a) and RC Rat-Only Weight Regression (b)



Abbreviations: RC=Registry of Cytotoxicity; R^2=Coefficient of determination.
Regressions calculated using IC_{50} and rat oral LD_{50} datapoints for 282 substances from the RC (see **Table 6-5**).

For comparison with the NRU test method results and RC rat-only regressions, **Section 6.4.1** provides the accuracy analysis for the RC database used with the RC millimole regression. **Sections 6.4.2** and **6.4.3** provide the accuracy information for the 3T3 and NHK NRU test methods for the RC rat-only millimole regression and RC-rat only weight regression, respectively. A summary of predictivity[3] is provided for each predicted toxicity category, along with the percentage of substances whose toxicity was underpredicted or overpredicted.

6.4.1 Prediction of GHS Acute Oral Toxicity Category by the RC IC_{50} Values Using the RC Millimole Regression

Table 6-6 shows the concordance of the observed (i.e., *in vivo*) and predicted GHS acute oral toxicity categories (UN 2005) for the 347 RC IC_{50} values in the RC millimole regression, log LD_{50} (mmol/kg) = 0.435 x log IC_{50} (mM) + 0.625 (Halle 1998, 2003). Accuracy is the agreement of the *in vitro* category predictions with those based on the 347 rodent (282 rat and 65 mouse) oral LD_{50} values used in the RC database (Halle 1998, 2003). Substances for which the *in vitro* toxicity category prediction did not match the *in vivo* category were considered discordant for the GHS acute oral toxicity category predictions.

The overall accuracy of the RC IC_{50} values for correctly predicting GHS acute oral toxicity classification category using the RC millimole regression was 40% (140/347substances) (**Table 6-6**). Rodent acute oral toxicity was overpredicted for 34% (118/347) and underpredicted for 26% (89/347) of the substances. For this analysis, with respect to the predictions of each GHS category:

- None (0%) of the 12 substances with LD_{50} <5 mg/kg (GHS Category I) was correctly predicted.
- Four (15%) of 26 substances in the 5 < LD_{50} ≤50 mg/kg category (GHS Category II) were correctly predicted.
- Twenty (29%) of 69 substances in the 50 < LD_{50} ≤300 mg/kg category (GHS Category III) were correctly predicted.
- Ninety-seven (69%) of 140 substances in the 300 < LD_{50} ≤2000 mg/kg category (GHS Category IV) were correctly predicted. This toxicity category was also predicted for 106 other substances (52%; 106/203) that did not fall in this category. Thus, the overall predictivity for this category was 48% (97/203 substances predicted for this category matched the *in vivo* category).
- Fourteen (25%) of the 56 substances in the 2000 < LD_{50} ≤5000 mg/kg category (GHS Category V) were correctly predicted.
- Five (11%) of the 44 substances with LD_{50} >5000 mg/kg (GHS Unclassified) were correctly predicted.

.

[3] Proportion of correct *in vivo* category matches for all substances with *in vitro* predictions for a particular category. Predictivity is one of the measures of test accuracy (ICCVAM 2003).

Table 6-6 Prediction of GHS Acute Oral Toxicity Category by the RC IC_{50} Values and the RC Millimole Regression[1]

***In Vivo* Rodent Oral LD_{50}[2] (mg/kg)**	**IC_{50}-Predicted GHS Category (mg/kg)[3]**						**Total**	**Accuracy**	**Toxicity Over-predicted**	**Toxicity Under-predicted**
	LD_{50} <5	**5 < LD_{50} ≤50**	**50 < LD_{50} ≤300**	**300 < LD_{50} ≤2000**	**2000 < LD_{50} ≤5000**	**LD_{50} >5000**				
LD_{50} < 5	0	5	3	4	0	0	12	0%	0%	100%
5 < LD_{50} ≤50	0	4	13	9	0	0	26	15%	0%	85%
50 < LD_{50} ≤300	0	9	20	38	2	0	69	29%	13%	58%
300 < LD_{50} ≤2000	0	4	24	97	14	1	140	69%	20%	11%
2000 < LD_{50} ≤5000	0	1	5	36	14	0	56	25%	75%	0%
LD_{50} >5000	0	0	1	19	19	5	44	11%	89%	0%
Total	0	23	66	203	49	6	347	40%	34%	26%
Predictivity	0%	17%	30%	48%	29%	83%				
Category Overpredicted	0%	61%	45%	27%	39%	0%				
Category Underpredicted	0%	22%	24%	25%	33%	17%				

Abbreviations: GHS=Globally Harmonized System of Classification and Labelling of Chemicals (UN 2005); RC=Registry of Cytotoxicity. Shaded cells are those containing the correct predictions; RTECS®=Registry of Toxic Effects for Chemical Substances®.

[1]The RC millimole regression is log LD_{50} (mmol/kg) = log IC_{50} (mM) x 0.435 + 0.625. Numbers in table represent numbers of substances.

[2]Rat (282 values) and mouse (65 values) oral LD_{50} values, mostly from the 1983/84 RTECS® that were converted to mmol/kg for used in the RC (Halle 1998, 2003).

[3]IC_{50} values from the RC are geometric mean IC_{50} values from *in vitro* cytotoxicity assays using multiple cell lines and cytotoxicity endpoints (Halle 1998, 2003). GHS categories were predicted by using the IC_{50} values to calculate predicted LD_{50} values with the RC millimole regression equation. Predicted LD_{50} values in mmol/kg for each substance were converted to mg/kg and used to classify the substance in the appropriate predicted GHS acute oral toxicity category.

The highest accuracy, 69%, for the RC IC_{50} values in the RC millimole regression were obtained for substances in the 300 < LD_{50} ≤2000 mg/kg category (GHS Category IV). The lowest accuracy, 0%, was obtained for substances with LD_{50} <5 mg/kg (GHS Category I). Although the 11% accuracy was low for substances with LD_{50} >5000 mg/kg (GHS Unclassified), the highest predictivity, 83%, was obtained for substances in this group. The RC millimole regression generally underpredicted toxicity for substances in the highest toxicity (i.e., lowest LD_{50}) categories and overpredicted for substances in the lowest toxicity (i.e., highest LD_{50}) categories (see **Table 6-6**).

Rodent acute oral toxicity was overpredicted for 34% (118) and underpredicted for 26% (89) of the 347 RC substances. Thus, there was a total of were 207 discordant substances. GHS category was overpredicted for 57% (118/207) of the discordant substances and underpredicted for 43% (89/207) of the discordant substances.

6.4.2 Prediction of GHS Acute Oral Toxicity Category by the 3T3 and NHK NRU Test Methods Using the RC Rat-Only Millimole Regression

Table 6-7 shows the concordance of the observed (i.e., *in vivo*) and predicted GHS acute oral toxicity categories (UN 2005) for each *in vitro* test method using the geometric mean IC_{50} values (of the three laboratories) in the RC rat-only millimole regression, log LD_{50} (mmol/kg) = 0.439 x log IC_{50} (mM) + 0.621. Accuracy is the agreement of the *in vitro* category predictions with those based on the rat acute oral LD_{50} reference values in **Table 4-2**. Substances for which the *in vitro* toxicity category prediction did not match the *in vivo* category were considered discordant for the GHS acute oral toxicity category predictions.

6.4.2.1 *In Vitro – In Vivo Concordance Using the RC Rat-Only Millimole Regression*

The overall accuracy of the 3T3 NRU test method for correctly predicting GHS acute oral toxicity classification category using the RC rat-only millimole regression was 31% (21/67 substances) (**Table 6-7**). Rat acute oral toxicity was overpredicted for 34% (23) and underpredicted for 34% (23) of the substances. For this analysis, with respect to the predictions of each GHS category:

- None (0%) of the six substances with LD_{50} <5 mg/kg (GHS Category I) was correctly predicted.
- One (9%) of 11 substances in the 5 < LD_{50} ≤50 mg/kg category (GHS Category II) was correctly predicted.
- Five (42%) of 12 substances in the 50 < LD_{50} ≤300 mg/kg category (GHS Category III) were correctly predicted.
- Thirteen (81%) of 16 substances in the 300 < LD_{50} ≤2000 mg/kg category (GHS Category IV) were correctly predicted. This toxicity category was also predicted for 32 other substances (71%; 32/45) that did not fall in this category. Thus, the overall predictivity for this category was 29% (13/45 substances predicted for this category matched the *in vivo* category).
- None (0%) of the 10 substances in the 2000 < LD_{50} ≤5000 mg/kg category (GHS Category V) were correctly predicted.
- Two (17%) of the 12 substances with LD_{50} >5000 mg/kg (GHS Unclassified) were correctly predicted.

Table 6-7 Prediction of GHS Acute Oral Toxicity Category by the 3T3 and NHK NRU Test Methods and the RC Rat-Only Millimole Regression[1]

Reference Rat Oral LD_{50}[2] (mg/kg)	3T3 -Predicted GHS Category (mg/kg)						Total	Accuracy	Toxicity Over-predicted	Toxicity Under-predicted
	LD_{50} <5	5 < LD_{50} ≤50	50 < LD_{50} ≤300	300 < LD_{50} ≤2000	2000 < LD_{50} ≤5000	LD_{50} >5000				
LD_{50} < 5	0	2	0	4	0	0	6[3]	0%	0%	100%
5 < LD_{50} ≤50	0	1	6	3	1	0	11[4]	9%	0%	91%
50 < LD_{50} ≤300	0	0	5	7	0	0	12	42%	0%	58%
300 < LD_{50} ≤2000	0	1	2	13	0	0	16	81%	19%	0%
2000 < LD_{50} ≤5000	0	0	0	10	0	0	10[5]	0%	100%	0%
LD_{50} >5000	0	0	0	8	2	2	12[6,7]	17%	83%	0%
Total	0	4	13	45	3	2	67	31%	34%	34%
Predictivity	0%	25%	38%	29%	0%	100%				
Category Overpredicted	0%	25%	15%	40%	67%	0%				
Category Underpredicted	0%	50%	46%	31%	33%	0%				
Reference Rat Oral LD_{50}[2]	**NHK -Predicted Toxicity Category (mg/kg)**						**Total**	**Accuracy**	**Toxicity Over-predicted**	**Toxicity Under-predicted**
	LD_{50} <5	**5 < LD_{50} ≤50**	**50 < LD_{50} ≤300**	**300 < LD_{50} ≤2000**	**2000 < LD_{50} ≤5000**	**LD_{50} >5000**				
LD_{50} <5	0	1	2	3	0	0	6[3]	0%	0%	100%
5 < LD_{50} ≤50	0	2	5	3	1	0	11[4]	18%	0%	82%
50 < LD_{50} ≤300	0	1	6	5	0	0	12	50%	8%	42%
300 < LD_{50} ≤2000	0	1	2	12	1	0	16	75%	19%	6%
2000 < LD_{50} ≤5000	0	0	0	10	0	0	10[5]	0%	100%	0%
LD_{50} >5000	0	0	0	7	6	0	13[7]	0%	100%	0%
Total	0	5	15	40	8	0	68	29%	40%	31%
Predictivity	0%	40%	40%	30%	0%	0%				
Category Overpredicted	0%	40%	13%	43%	75%	0%				
Category Underpredicted	0%	20%	47%	28%	25%	0%				

Abbreviations: GHS=Globally Harmonized System of Classification and Labelling of Chemicals (UN 2005); 3T3=BALB/c 3T3 fibroblasts; NHK=Normal human epidermal keratinocytes; NRU=Neutral red uptake; RC=Registry of Cytotoxicity. Shaded cells are those containing the correct predictions.

[1]The RC rat-only millimole regression is log LD_{50} (mmol/kg) = log IC_{50} (mM) x 0.439 + 0.621. Numbers in table represent numbers of substances.

[2]Reference rat oral LD_{50} values in mg/kg from **Table 4-2**.

[3]Epinephrine bitartrate excluded because no rat reference acute oral LD_{50} was identified (see **Table 4-2**).

[4]Colchine excluded because no rat acute oral LD_{50} was identified (see **Table 4-2**).

[5]Carbon tetrachloride excluded because no laboratory attained sufficient toxicity for the calculation of an IC_{50}.

[6]Methanol excluded because no laboratory attained sufficient toxicity for the calculation of an IC_{50}.

[7]Propylparaben excluded because no rat acute oral LD_{50} was identified (see **Table 4-2**).

The overall accuracy of the NHK NRU test method for correctly predicting the GHS acute oral toxicity classification, when the prediction was based on the RC rat-only millimole regression, was 29% (20/68 substances) (see **Table 6-7**). Toxicity was overpredicted for 40% (27) and underpredicted for 31% (21) of the 68 substances. The pattern of concordance between *in vitro* and *in vivo* results for the NHK NRU test method with the RC rat-only millimole regression was similar to that for the 3T3 NRU test method with the exception that none of the substances with a toxicity of LD_{50} >5000 mg/kg were correctly predicted. For this analysis, with respect to the predictions of each GHS category:

- None (0%) of the six substances with LD_{50} <5 mg/kg (GHS Category I) were correctly predicted.
- Two (18%) of 11 substances in the 5< LD_{50} ≤50 mg/kg category (GHS Category II) were correctly predicted.
- Six (50%) of 12 substances in the 50< LD_{50} ≤300 mg/kg category (GHS Category III) were correctly predicted.
- 12 (75%) of 16 substances in the 300< LD_{50} ≤2000 mg/kg category (GHS Category IV) were correctly predicted; however, this category was also predicted for 28 (70%; 28/40) substances that did not match the category. Thus, the overall predictivity for this category was 30% (12/40).
- None (0%) of the 10 substances in the 2000< LD_{50} ≤5000 mg/kg category (GHS Category V) were correctly predicted.
- None (0%) of the 13 substances with LD_{50} >5000 mg/kg (GHS Unclassified) were correctly predicted.

The RC rat-only millimole regression generally underpredicted toxicity for substances in the highest toxicity (i.e., lowest LD_{50}) categories and overpredicted toxicity for substances in the lowest toxicity (i.e., highest LD_{50}) categories (see **Table 6-7**). Although substances at the very low and high ends of the toxicity range were poorly predicted, those in the middle range (i.e., 300 < LD_{50} ≤2000 mg/kg) were predicted much better, with 75 to 81% accuracy. The pattern of accuracy for the GHS categories was similar to the pattern seen with the RC IC_{50} and LD_{50} values and the RC millimole regression (see **Table 6-6**) (i.e., lowest accuracy for very toxic and very nontoxic substances and highest accuracy for substances with 300 < LD_{50} ≤2000 mg/kg).

6.4.2.2 *Discordant Substances in the Prediction of GHS Acute Oral Toxicity Category by the 3T3 and NHK NRU Test Methods and the RC Rat-Only Millimole Regression*

Appendix L2 identifies the discordant substances, that is, those for which the *in vitro* predicted GHS acute oral toxicity category did not match the GHS acute oral toxicity category assigned based on the reference rat acute oral LD_{50} data in **Table 4-2**. Of the total number of substances used for this evaluation (67 for 3T3, 68 for NHK), the 3T3 test method underpredicted the GHS category for 23 (50%) and overpredicted for 23 (50%) of the 46 discordant substances. The NHK test method underpredicted toxicity for 21 (44%) and overpredicted for 27 (56%) of the 48 discordant substances.

6.4.3 Prediction of GHS Acute Oral Toxicity Category by the 3T3 and NHK NRU Test Methods Using the RC Rat-Only Weight Regression

Table 6-8 shows the concordances of the observed and predicted GHS acute oral toxicity categories for each *in vitro* NRU method using the geometric mean IC_{50} values from the

three laboratories and the RC rat-only weight regression (**Table 6-5)**. The regression formula for the RC rat-only weight regression was log LD_{50} (mg/kg) = log IC_{50} (μg/mL) x 0.372 + 2.024. Accuracy is the agreement of the GHS acute oral toxicity category predictions made using the *in vitro* NRU data with those based on the reference rat acute oral LD_{50} values (**Table 4-2)**.

6.4.3.1 *In Vitro – In Vivo Concordance Using the RC Rat-Only Weight Regression*

The overall accuracy of the 3T3 NRU test method with the RC rat-only weight regression was 31% (21/67) (**Table 6-8**). The toxicity was overpredicted for 33% (24) and underpredicted for 36% (22) of the substances. For this analysis, with respect to the predictions of the GHS category:

- None (0%) of the six substances with LD_{50} <5 mg/kg (GHS Category I) were correctly predicted.
- One (9%) of 11 substances in the 5< LD_{50} ≤50 mg/kg category (GHS Category II) was correctly predicted.
- Four (33%) of 12 substances in the 50< LD_{50} ≤300 mg/kg category (GHS Category II) were correctly predicted; however, because 10 other substances were also predicted to be in this category, the overall predictivity was 29% (4/14).
- Twelve (75%) of 16 substances in the 300< LD_{50} ≤2000 mg/kg category (GHS Category IV) were predicted correctly. Because a total of 40 substances were predicted to be in this category, the overall predictivity was 30% (12/40).
- Four (40%) of 10 substances in the 2000< LD_{50} ≤5000 mg/kg category (GHS Category V) were correctly predicted; however, because a total of 11 substances were predicted to be in this category, the overall predictivity was 36% (4/11).
- None (0%) of the 12 substances with LD_{50} >5000 mg/kg (GHS Unclassified) were correctly predicted.

The overall accuracy of the NHK predictions using the RC rat-only weight regression was 31% (21/68) (see **Table 6-8**). The *in vivo* GHS toxicity categories were overpredicted for 37% (22) and underpredicted for 32% (25) of the substances. For this analysis, with respect to the predictions of the GHS category:

- None (0%) of the six substances with LD_{50} <5 mg/kg (GHS Category I) were correctly predicted.
- One (9%) of 11 substances in the 5 < LD_{50} ≤ 50 mg/kg category (GHS Category II) was correctly predicted.
- Five (42%) of 12 substances in the 50 < LD_{50} ≤300 mg/kg category (GHS Category III) were correctly predicted; however, because six other substances were also predicted to be in this category, the overall predictivity was 33% (3/9).
- Thirteen (81%) of 16 substances in the 300 < LD_{50} ≤2000 mg/kg category (GHS Category IV) were predicted correctly; however, because 29 other substances were also predicted to be in this category, the overall predictivity was 31% (13/42).

Table 6-8 Prediction of GHS Acute Oral Toxicity Category by the 3T3 and NHK NRU Test Methods and the RC Rat-Only Weight Regression[1]

Reference Rat Oral LD_{50}[2] (mg/kg)	3T3 -Predicted Toxicity Category (mg/kg)						Total	Accuracy	Toxicity Over-predicted	Toxicity Under-predicted
	LD_{50} <5	5 < LD_{50} ≤50	50 < LD_{50} ≤300	300 < LD_{50} ≤2000	2000 < LD_{50} ≤5000	LD_{50} >5000				
LD_{50} <5	0	0	2	4	0	0	6[3]	0%	0%	100%
5 < LD_{50} ≤50	0	1	5	5	0	0	11[4]	9%	0%	91%
50 < LD_{50} ≤300	0	0	4	8	0	0	12	33%	0%	67%
300 < LD_{50} ≤2000	0	1	3	12	0	0	16	75%	25%	0%
2000 < LD_{50} ≤5000	0	0	0	6	4	0	10[5]	40%	60%	0%
LD_{50} >5000	0	0	0	5	7	0	12[6,7]	0%	100%	0%
Total	0	2	14	40	11	0	67	31%	33%	36%
Predictivity	0%	50%	29%	30%	36%	0%				
Category Overpredicted	0%	50%	21%	28%	64%	0%				
Category Underpredicted	0%	0%	50%	43%	0%	0%				
Reference Rat Oral LD_{50}[2] (mg/kg)	**NHK -Predicted Toxicity Category (mg/kg)**									
	LD_{50} <5	**5 < LD_{50} ≤50**	**50 < LD_{50} ≤300**	**300 < LD_{50} ≤2000**	**2000 < LD_{50} ≤5000**	**LD_{50} >5000**	**Total**	**Accuracy**	**Toxicity Over-predicted**	**Toxicity Under-predicted**
LD_{50} <5	0	1	2	3	0	0	6[3]	0%	0%	100%
5 < LD_{50} ≤50	0	1	5	5	0	0	11[4]	9%	0%	91%
50 < LD_{50} ≤300	0	1	5	6	0	0	12	42%	8%	50%
300 < LD_{50} ≤2000	0	1	2	13	0	0	16	81%	19%	0%
2000 < LD_{50} ≤5000	0	0	0	9	1	0	10[5]	10%	90%	0%
LD_{50} >5000	0	0	0	6	6	1	13[7]	8%	92%	0%
Total	0	4	14	42	7	1	68	31%	37%	32%
Predictivity	0%	25%	36%	31%	14%	100%				
Category Overpredicted	0%	50%	14%	36%	86%	0%				
Category Underpredicted	0%	25%	50%	33%	0%	0%				

Abbreviations: GHS=Globally Harmonized System of Classification and Labelling of Chemicals (UN 2005); 3T3=BALB/c 3T3 fibroblasts; NHK=Normal human epidermal keratinocytes; NRU=Neutral red uptake; RC=Registry of Cytotoxicity.

[1]The RC rat-only weight regression is log LD_{50} (mg/kg) = log IC_{50} (µg/mL) x 0.372 + 2.024.

[2]Reference rat oral LD_{50} values in mg/kg from **Table 4-2**.

[3]Epinephrine bitartrate excluded because no rat acute oral LD_{50} was identified (see **Table 4-2**).

[4]Colchine excluded because no rat acute oral LD_{50} was identified (see **Table 4-2**).

[5]Carbon tetrachloride excluded because no laboratory attained sufficient toxicity for the calculation of an IC_{50}.

[6]Methanol excluded because no laboratory attained sufficient toxicity for the calculation of an IC_{50}.

[7]Propylparaben excluded because no rat acute oral LD_{50} was identified (see **Table 4-2**).

- One (10%) of 10 substances in the 2000 < LD_{50} ≤5000 mg/kg category (GHS Category V) was correctly predicted.
- One (8%) of 13 substances with LD_{50} >5000 mg/kg (GHS Unclassified) was correctly predicted.

The RC rat-only weight regression generally underpredicted toxicity for substances in the highest toxicity (i.e., lowest LD_{50}) categories and overpredicted toxicity for substances in the lowest toxicity (i.e., highest LD_{50}) categories (see **Table 6-8**). Although substances at the very low and high ends of the toxicity range were poorly predicted, those in the middle range (i.e., 300 < LD_{50} ≤2000 mg/kg) were predicted much better, with 75 to 81% accuracy. The pattern of accuracy for the GHS categories was similar to the pattern seen with the RC IC_{50} and LD_{50} values and the RC millimole regression (see **Table 6-6**) and with the NRU IC_{50} and rat oral LD_{50} values and the RC rat-only millimole regression (see **Table 6-7**) (i.e., lowest accuracy for very toxic and very nontoxic substances and highest accuracy for substances with 300 < LD_{50} ≤2000 mg/kg).

6.4.3.2 *Discordant Substances in the Prediction of GHS Acute Oral Toxicity Category by the 3T3 and NHK NRU Test Methods and the RC Rat-Only Weight Regression*

Appendix L2 shows the substances for which the *in vitro* predicted GHS acute oral toxicity category using the RC rat-only weight regression did not match those that were based on the rat acute oral LD_{50} reference data. The two *in vitro* NRU test methods over- and under-predicted the GHS acute oral toxicity category for similar numbers of substances, compared with the GHS acute oral toxicity categories for the rat acute oral LD_{50} reference values in **Table 4-2**. The 3T3 NRU test method overpredicted the GHS acute oral toxicity category for 22 (48%) of 46 discordant substances, and underpredicted of 24 (52%) substances. The NHK NRU test method overpredicted the GHS acute oral toxicity category for 25 (53%) of 47 discordant substances, and underpredicted 22 (47%) substances.

6.4.4 Summary of the Regressions Evaluated

Table 6-9 summarizes the regressions evaluated in **Section 6.4** for accuracy in predicting the GHS acute oral toxicity categories (UN 2005), and the proportion of over- or under-predictions. Prediction accuracy using the RC IC_{50} and LD_{50} values and the RC millimole regression was higher that that for the NRU test methods with the RC rat-only regressions (i.e., 40% for the RC vs. 29% to 31% for the NRU test methods). Prediction accuracy was slightly higher for the 3T3 NRU test method compared with the NHK NRU (i.e., 31% for 3T3 vs. 29% for NHK) using the RC rat-only millimole regression, and the same as the NHK NRU test method (i.e., 31%) using the RC rat-only weight regression. The proportion of discordant substances using the RC IC_{50} values and the RC millimole regression (60%) was lower than that using the *in vitro* NRU test methods and the RC rat-only regressions (69% to 71%). The proportion of discordant substances from the 3T3 test method, 69%, was the same whether it was determined with the RC rat-only millimole regression or the RC rat-only weight regression. The proportion of discordant substances for the NHK test method was slightly lower with RC rat-only weight regression than with the RC rat-only millimole regression (69% vs. 71%). The RC IC_{50} values and the RC millmole regression were expected to perform better than the *in vitro* NRU methods and the RC rat-only regressions since the IC_{50} and LD_{50} values used to evaluate the performance of the RC millimole regression were exactly the same as those used to calculate the linear regression formula. The

NRU IC_{50} values and the reference oral LD_{50} values used to evaluate the RC rat-only regressions were different from those used to calculate the RC rat-only regressions.

Table 6-9 Comparison of Regressions and *In Vitro* NRU Test Methods for Their Performance in Predicting GHS Acute Oral Toxicity Categories

Regression	N[1]	R^2 Statistic	Accuracy	Discordant Substances[2]
RC millimole[3]	347	0.452	RC IC_{50} – 40%	RC IC_{50} – 207/347 (60%)
RC rat-only millimole[3]	282	0.452	3T3– 31% NHK– 29%	3T3– 46/67(69%) NHK– 48/68 (71%)
RC rat-only weight[3]	282	0.325	3T3– 31% NHK– 31%	3T3– 46/67 (69%) NHK– 47/68 (69%)

Abbreviations: GHS=Globally Harmonized System of Classification and Labelling of Chemicals (UN 2005); 3T3=BALB/c 3T3 fibroblasts; NHK=Normal human epidermal keratinocytes; NRU=Neutral red uptake; RC=Registry of Cytotoxicity; R^2=Coefficient of determination.
[1]Number of substances used in regression.
[2]Proportion of discordant substances.
[3]From **Table 6-5**.

The accuracy of the GHS category predictions using the *in vitro* NRU test methods with the RC rat-only regressions obtained for the reference substances from this validations study may or may not be applicable to other substances. A number of reasons may explain the low accuracy for the reference substances. One is the skewness of the substances selected for testing with respect to fit to the RC millimole regression (see **Figure 3-1**). **Table 3-4** shows that 22 (38%) of the 58 RC substances selected for testing were known to poorly fit the RC millimole regression (i.e., the predicted LD_{50} was outside the RC acceptance interval). Toxicity was underpredicted for 17 (77%) of these outlier substances and overpredicted (i.e., predicted LD_{50} was lower than measured *in vivo* LD_{50}) for the remaining five (23%). **Table 6-3** shows that 40% (28/70 for 3T3) and 44% (31/71 for NHK) of the reference substances that yielded IC_{50} values were outliers. Other reasons for the low accuracy for GHS acute oral toxicity prediction, such as those discussed in **Section 1.2.3**, include the major differences between cell cultures and whole animals regarding the absorption, distribution (including binding to serum proteins), availability, metabolism, and excretion of reference substances.

6.5 Correlation of NRU Concentration-Response Slope with Rat Lethality Dose-Response Slope

Because the slope calculations available for the NRU concentration-response curve analyses were based on the Hill function, the SMT determined whether the Hill Slope correlated with the rodent dose-mortality slope. If the two were correlated, the Hill Slope from the NRU test methods could be used to estimate the dose-mortality slope, which could, in turn, be used to estimate the most appropriate dose progression for UDP testing in rodents. A more immediate use for the validation study results, however, would be for the computer simulation modeling of animal testing for the UDP and ATC acute oral toxicity methods (described in **Sections 10.2** and **10.3**).

Dose-mortality slope information was available for 22 of the 72 reference substances, as shown in **Table 6-10**. Hill function slopes were available for 20 and 21 of the 22 substances

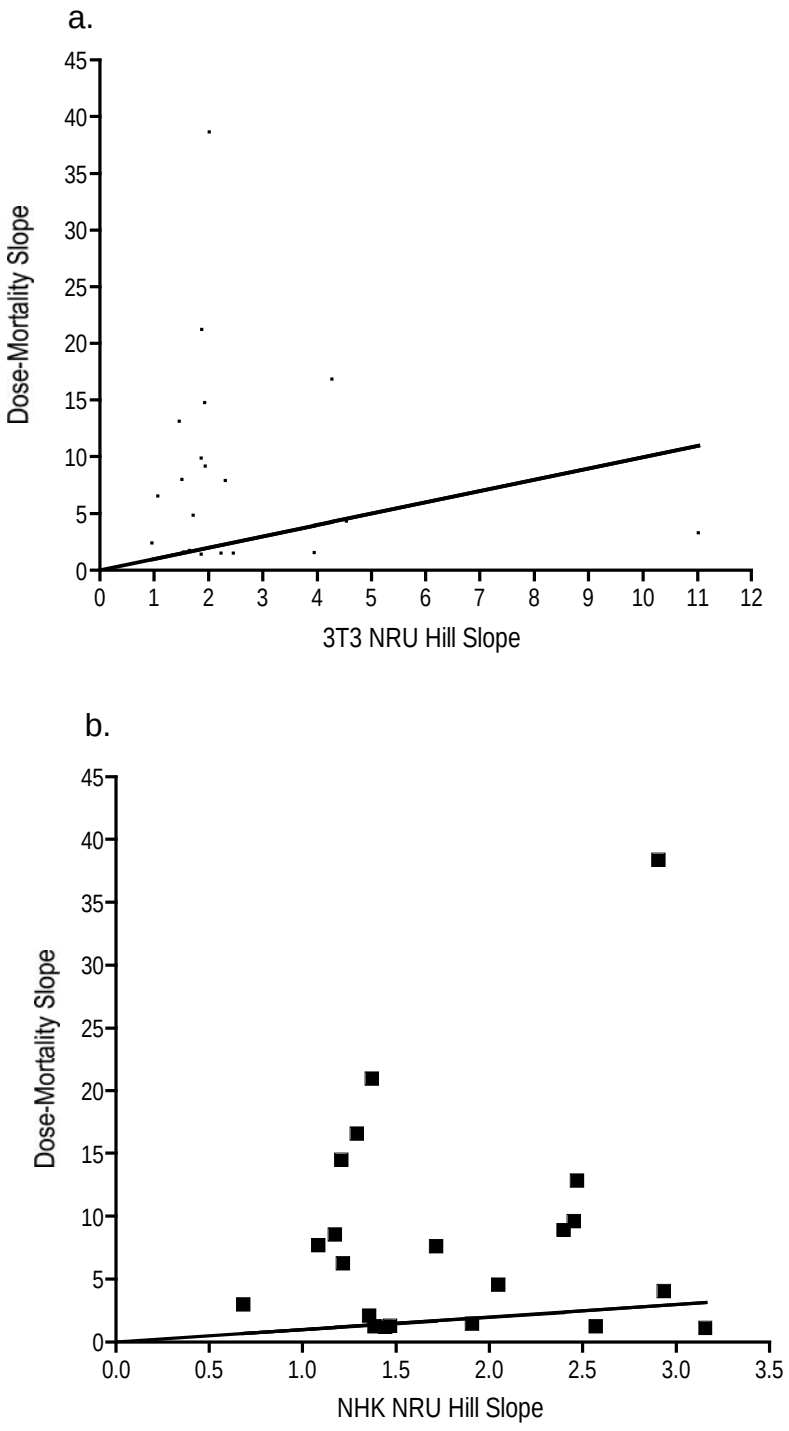
for the 3T3 and the NHK NRU test methods, respectively. The Hill function slopes were transformed to absolute values because geometric means cannot be calculated for negative numbers, and geometric mean Hill function slopes were calculated for the acceptable NRU tests for each reference substance. When there was more than one dose-mortality slope available for a substance, a geometric mean was calculated from the available values. The absolute values of the geometric mean Hill function slopes are plotted against the geometric mean dose-mortality slopes in **Figure 6-5**. To determine whether there was a relationship between the absolute value of the Hill Slope and the dose-mortality slope, Spearman correlation analyses and least squares linear regression analyses were performed for each method. Both analyses showed that the absolute value of the *in vitro* Hill function slope was not related to the dose-mortality slope. The Spearman correlation analysis yielded nonsignificant correlations for both *in vitro* NRU test methods (3T3 r_s=-0.051 with p=0.831, and NHK r_s=-0.142 with p=0.541). Linear regression analyses for the prediction of dose-mortality slope by the absolute value of the Hill function slope also showed that the slopes of the regressions were not significantly different from zero (3T3 p=0.774, and NHK p=0.994). Because there was no relationship between Hill function slope and dose-mortality slope, the Hill function slope was not used to predict the dose-mortality slope for the simulation modeling of animal testing for the UDP and ATC acute oral toxicity methods in **Sections 10.2** and **10.3**.

Table 6-10 Reference Substances with Dose-Mortality and NRU Hill Slopes

Reference Substance	Dose-Mortality Slope[1]	3T3 Hill Slope[2]	NHK Hill Slope[2]
Acetylsalicylic acid	1.45	1.658	1.906
Boric acid	7.70	1.511	1.083
Caffeine	6.27	1.069	1.215
Carbon tetrachloride	2.06	NA	NA
Dichlorvos	1.24	2.240	1.383
Dimethylformamide	1.11	1.875	3.157
Diquat dibromide	16.57	4.273	1.289
Ethanol	4.57	1.725	2.049
Ethylene glycol	38.38	2.016	2.904
Glycerol	8.90	1.941	2.398
Hexachorophene	12.84	1.466	2.470
Lactic acid	4.04	4.541	2.934
Methanol	8.53	NA	1.173
Nicotine	3.00	11.019	0.682
Parathion	1.31	1.551	1.467
Potassium cyanide	14.50	1.931	1.207
Sodium arsenite	7.60	2.317	1.717
Sodium I fluoride	1.26	3.952	2.569
Trichloroacetic acid	20.97	1.883	1.369
Triethylene melamine	2.10	0.963	1.355
Valproic acid	1.20	2.467	1.440
Xylene	9.60	1.871	2.452
Carbon tetrachloride	2.06	NA	NA

Abbreviations: 3T3=BALB/c 3T3 fibroblasts; NHK=Normal human epidermal keratinocytes; NRU=Neutral red uptake; NA=Not available.
[1]Geometric mean if there was more than one value for each substance (from **Appendix H2**).
[2]Geometric mean of absolute values from acceptable *in vitro* NRU tests.

Figure 6-5 Correlation of Dose-Mortality Slope to Hill Function Slope



Abbreviations: 3T3=BALB/c 3T3 fibroblasts; NHK=Normal human epidermal keratinocytes; NRU=Neutral red uptake. Hill function slopes and dose-mortality slopes for the reference substances shown in **Table 6-10** for (a) the 3T3 data and (b) the NHK data. The solid line indicates the theoretical, one-to-one correspondence of Hill function slope with dose-mortality slope. Spearman's correlation coefficients were r_s=-0.051 (p=0.831) for the 3T3 and r_s=-0.142 (p=0.541) for the NHK data.

6.6 Strengths and Limitations of the Use of *In Vitro* Cytotoxicity Test Methods with the IC_{50}-LD_{50} Regressions for Prediction of Rodent Acute Oral Toxicity

6.6.1 *In Vitro* Cytotoxicity Methods

The NRU basal cytotoxicity methods tended to underpredict the toxicity of the most toxic substances and to overpredict the toxicity of the least toxic substances for each regression evaluated. The 3T3 and NHK NRU test methods were best at predicting the toxicity of substances with $300 < LD_{50} \leq 2000$ mg/kg. The accuracy of the *in vitro* prediction of this GHS category using the RC rat-only millimole regression and the RC rat-only weight regression was 75-81%. GHS toxicity categories of substances with higher or lower LD_{50} values were correctly predicted with less than 50% accuracy. The worst accuracy, 0%, was observed for:

- Substances with $LD_{50} \leq 5$ mg/kg in both *in vitro* test methods and regressions
- Substances with $2000 < LD_{50} \leq 5000$ mg/kg using 3T3 with the RC rat-only millimole regression
- Substances with $2000 < LD_{50} \leq 5000$ mg/kg or $LD_{50} > 5000$ mg/kg using NHK with RC rat-only millimole regression
- Substances with $LD_{50} > 5000$ mg/kg using 3T3 with RC rat-only weight regression

Some substances with low toxicity and low solubility could not be tested in the *in vitro* NRU test methods because the concentration of dissolved substance was inadequate to obtain an IC_{50} value. None of the laboratories obtained adequate toxicity in any of the 3T3 tests of carbon tetrachloride or methanol, and at least one laboratory failed to achieve adequate toxicity with gibberellic acid or xylene. No laboratory achieved adequate toxicity in any of the NHK experiments with carbon tetrachloride, and at least one laboratory could not achieve adequate toxicity with methanol, 1,1,1-trichloroethane, or xylene. Another limitation of use of the *in vitro* test methods is in the testing of substances that come out of solution by forming a film on the medium surface or plastic well wall (i.e., "film out"), and for substances that etch the laboratory ware plastics (ICCVAM 2006). Substances that etch plastics can be detected by looking for the presence of etched rings in the 96-well plates after exposure. Some substances that produce films in medium also etch plastic.
The prediction of rodent acute oral toxicity (and the starting doses for acute oral toxicity tests) by the *in vitro* NRU methods is expected to be poor for substances with mechanisms of toxicity that are not effective in the 3T3 and NHK cells. Such toxic mechanisms include specific, receptor-mediated actions on the CNS or the heart.

The evaluation of the 3T3 and NHK NRU test methods for predicting starting doses for rodent acute oral toxicity testing with its potential to reduce and refine animal use is provided in **Section 10**.

6.6.2 Use of Mole-Based vs. Weight-Based Regressions for the Prediction of Toxicity for Low and High Molecular Weight Substances

The ICCVAM ATWG expressed concern that the RC rat-only weight regression may less accurately predict the toxicity of low and high molecular weight substances than the RC rat-only millimole regression. Using the RC IC_{50} and LD_{50} values for the 282 RC substances with rat oral LD_{50} data, analyses were performed to:

- Determine the difference in the over and under-prediction of rodent acute oral toxicity (i.e., LD_{50}) from IC_{50} values between low molecular weight substances (i.e., ≤100 g/mole) and substances with molecular weights >100 g/mole
- Determine the difference in the over and under-prediction of rodent acute oral toxicity from IC_{50} values between high molecular weight substances (i.e., ≥400 g/mole) vs. substances with molecular weights <400 g/mole.
- Compare the RC rat-only millimole regression with the RC rat-only weight regression with respect to the over- and under-prediction of the toxicity of low and high molecular weight substances

This analysis used the RC data rather than the validation studies data because the RC contains data for many more substances. The analysis assumes that the regressions either underpredicted or overpredicted the toxicity of all of the substances evaluated. In other words, there was a difference between the LD_{50} predicted by the regression and the *in vivo* LD_{50} used to calculate the regression even if it was a tiny fraction (i.e., no substances fit the regression exactly). The complete analysis and discussion are presented in **Appendix J7**. Of the 282 RC substances with rat acute oral LD_{50} values, there were 51 with molecular weights ≤100 g/mole and 231 with molecular weights >100 g/mole. For the 51 substances with molecular weight ≤100 g/mole, the RC rat-only millimole regression underestimated the toxicity of 20/51 (39%) substances and overestimated the toxicity of 31/51 (61%) substances. The RC rat-only weight regression underestimated the toxicity of 24/51 (47%) substances and overestimated the toxicity of 27/51 (53%) substances. Fisher's exact test indicated that there was no difference between the millimole and weight regressions with respect to the under or over-prediction of toxicity for the low molecular weight substances (two-tailed p=0.549) (see **Table 6-11**).

For the 231 substances with molecular weights >100 g/mole, the RC rat-only millimole regression underestimated the toxicity of 108/231 (47%) substances and overestimated the toxicity of 123/231 (53%). The RC rat-only weight regression underestimated the toxicity of 101/231 (44%) substances and overestimated the toxicity of 130/231 (57%). Fisher's exact test indicated that there were no significant differences between the millimole and weight regressions for the under- and over-prediction of toxicity for the 231 substances with molecular weight >100 g/mole (two-tailed p=0.575). Fisher's exact test also showed that there were no significant differences in the under- and over-prediction of the toxicity of the 51 substances with molecular weight ≤100 g/mole compared to the under- and over-prediction of the toxicity of the 231 with molecular weight >100 g/mole (two-tailed p=0.756 for the RC rat-only weight regression, and two-tailed p=0.355 for the RC rat-only millimole regression).

Table 6-11 Over- and Under- Prediction of Toxicity for Low and High Molecular Weight Substances Using RC Rat-Only Weight and Millimole Regressions

Comparison	For	Fisher's Exact Test[1]
RC rat-only millimole vs. RC rat-only weight regression	Under- and over-prediction of toxicity for 51 substances with molecular weight ≤100 g/mole	0.549
RC rat-only millimole vs. RC rat-only weight regression	Under- and over-prediction of toxicity for 231 substances with molecular weight >100 g/mole	0.575
51 Low molecular weight (≤100 g/mole) substances vs. 231 other substances (>100 g/mole)	RC rat-only millimole regression	0.355
51 Low molecular weight (≤100 g/mole) substances vs. 231 other substances (>100 g/mole)	RC rat-only weight regression	0.756
RC rat-only millimole vs. RC rat-only weight regression	Under- and over-prediction of toxicity for 20 substances with molecular weight ≥400 g/mole	0.480
RC rat-only millimole vs. RC rat-only weight regression	Under- and over-prediction of toxicity for 262 substances with molecular weight <400 g/mole	NT
20 High molecular weight substances (≥400 g/mole) vs. 262 other substances (<400 g/mole)	RC rat-only millimole regression	0.362
20 High molecular weight substances (≥400 g/mole) vs. 262 other substances (<400 g/mole)	RC rat-only weight regression	0.033

Abbreviations: RC=Registry of Cytotoxicity; NT=Not tested because the proportions were the same. Toxicity was underpredicted for 121/262 (46%) substances and overpredicted for 141/262 (54%) substances.
[1]P-values.

Of the 282 RC substances with rat acute oral LD_{50} values, there were 20 with molecular weights ≥400 g/mole and 262 with molecular weights <400 g/mole. The RC rat-only millimole regression underestimated the toxicity of 7/20 (35%) of the ≥400 g/mole substances and overestimated 13/20 (65%). The RC rat-only weight regression underestimated the toxicity of 4/20 (20%) of the substances and overestimated 16/20 (80%). Fisher's exact test indicated that there were no differences between the millimole and weight regressions for the under- and over-prediction of toxicity for the 20 high molecular weight substances (two-tailed p=0.4801).

For the remaining 262 substances with molecular weights <400 g/mole, both the RC rat-only millimole and the RC rat-only weight regressions underestimated the toxicity of 121/262 (46%) substances and overestimated 141/262 (54%). Thus, there were no statistical differences in the under- and over-esimation of toxicity for the 262 substances with molecular weights <400 g/mole regardless of which regression was used. Fisher's exact test also showed that there was no statistical difference in the under- and over-prediction of the toxicity of substances with high molecular weight (≥400 g/mole) compared with the under- and over-prediction the lower molecular weight substances using the RC rat-only millimole regression (two-tailed p=0.362). In contrast the use of the RC rat-only weight regression, resulted in a small but statistically significant difference in the under- and over-prediction of

the toxicity of substances with high molecular weight (>400 g/mole) compared with the under- and over-prediction of the toxicity of substances with lower molecular weight (two-tailed p=0.033). The weight-based regression significantly overestimated the toxicity of the high molecular weight substances (compared with substances with lower molecular weight) while the millimole regression did not.

6.7 Salient Issues of Data Interpretation

One of the most important considerations for the 3T3 and NHK NRU test methods, as for any test method, is the ability to generate good concentration-response results. In addition to technical difficulties with these test methods, such as occasional poor cell growth and the formation of NRU crystals, this validation study yielded non-monotonic concentration-response curves for certain substances.

A number of substances produced non-monotonic concentration-response curves in the 3T3 and/or the NHK NRU range finding or definitive tests. Because the *in vitro* NRU test methods, and the calculation of IC_{50} values from the resulting concentration curves, presume that the toxic response is linear, the data from non-linear responses (e.g., biphasic curves), as seen with aminopterin, do not always permit an IC_{50} determination by the standard Hill function analysis. In such cases, the lowest concentration that killed approximately 50% of the cells in the range finding test was used to set the concentration range for the definitive test. The definitive test used more closely spaced concentrations in an attempt to obtain a monotonic concentration-response curve. However, 100% toxicity (or 0%) viability was often unattainable in such definitive tests that exhibited a plateau of toxicity well over 0% viability (e.g., 20%). Care must be used in the calculation of the IC_{50} for curves for which toxicity plateaus to assure that the value reflects the concentration at 50% inhibition of the VC value rather than simply the midpoint of the highest and lowest response.

Because of low toxicity and/or low solubility, some substances did not produce sufficient toxicity for the calculation of an IC_{50} value. Carbon tetrachloride, methanol, xylene, gibberellic acid, lithium carbonate, and 1,1,1-trichloroethane failed to yield acceptable IC_{50} results in at least one laboratory because of insufficient toxicity. All of these substances, with the exception of methanol, produced precipitate in the cell culture medium.

6.8 Comparison of NRU Test Results to Established Performance Standards

The *Guidance Document* method of evaluating *in vitro* basal cytotoxicity assays for predicting starting doses for rodent acute oral toxicity assays provides the existing performance standard (ICCVAM 2001b) for the 3T3 and NHK NRU test methods. The *Guidance Document* recommends testing 10 to 20 reference substances from the RC in an *in vitro* basal cytotoxicity assay for predicting starting doses for rodent acute oral toxicity testing (ICCVAM 2001b). These substances should cover a wide range of toxicity and fit the RC millimole regression as closely as possible. The *Guidance Document* recommends using the IC_{50} results for the selected reference substances from the candidate method to calculate a new regression line with the LD_{50} values used by the RC. If the resulting regression is parallel to the RC millimole regression and within the $\pm \log 5$ (i.e., ± 0.699) prediction interval for the RC, candidate assay may be considered effective for predicting starting doses for substances in rodent acute oral toxicity assays.

One goal of the testing in Phases Ib and II of this study was to establish whether the results from the 3T3 and NHK NRU test methods were consistent with the RC millimole regression. As discussed in **Section 3.3.5**, two of the major criteria for selecting the 12 coded substances tested from the 72 reference substances were:

(a) Two substances must be included from each of the unclassified and classified GHS acute oral toxicity categories, and
(b) The substances must fit as closely to the RC millimole regression as possible.

Unfortunately, the SMT could not identify 12 substances that fit both criteria because there was only one substance, aminopterin, in the LD_{50} <5 mg/kg category that fit the RC millimole regression. The other substance chosen from that toxicity category was sodium selenate. Because sodium selenate was not included in the RC, there was no indication of how closely it would fit the RC millimole regression, and it was therefore not included in the Phases Ib and II regression analyses. The other 10 substances selected for testing in Phases Ib and II were colchicine, arsenic trioxide, cadmium chloride, sodium fluoride, propranolol, lithium carbonate, potassium chloride, chloramphenicol, 2-propanol, and ethylene glycol.

The geometric mean log IC_{50} (mM) values from the 3T3 and NHK NRU test methods from each laboratory were used with the oral log rodent LD_{50} (mmol/kg) values from the RC (see **Appendices J1** and **J2**) for the least squares linear regression analyses (see **Section 5.5.3.3**) for the substances tested in Phases Ib and II. The slopes for all regressions were significantly different from zero at $p < 0.0001$, which indicated that there was a significant relationship between IC_{50} and LD_{50}. The R^2 values for the regressions from each laboratory, shown in **Table 6-12**, show that the 3T3 NRU test method produced better-fitting regressions than the corresponding NHK NRU test method (R^2 = 0.940 to 0.953 vs. 0.577 to 0.621). The relatively low R^2 values for the NHK NRU test method were attributed to the much lower toxicity of aminopterin in those cells (see **Figures 6-6** to **6-8** and **Tables 5-3** and **5-4**). All test method and laboratory-specific regressions were consistent with the RC millimole regression. **Table 6-12** shows that all joint comparisons of slopes and intercepts with the RC millimole regression were not significant (i.e., $p > 0.01$). The RC millimole regression slope and intercept were used as constants for this comparison.

A graphic comparison of the IC_{50} regressions with the RC millimole regression as suggested by the *Guidance Document* (ICCVAM 2001b) demonstrated that they were generally within the RC millimole regression acceptance limits (see **Figures 6-6, 6-7,** and **6-8**). According to the *Guidance Document* (ICCVAM 2001b), *in vitro* basal cytotoxicity assays providing such consistency with the RC millimole regression are acceptable for predicting starting doses for rodent acute oral toxicity assays.

As an additional analysis, a regression for the 11 substances tested in Phases Ib and II (the RC-11 millimole regression*),* was calculated using the log RC IC_{50} (mM) and log LD_{50} (mmol/kg) values (see **Table 6-12**). Each of the laboratory regressions for each test method was then compared to the RC-11 regression using an F test for a joint comparison of slope and intercept. None of the regressions were significantly different from the RC-11 regression (p values ranged from 0.755 to 0.933).

Table 6-12 Linear Regressions for 11 Substances Tested in Phases Ib and II

	3T3 Regression[1]				
Laboratory	**Intercept**	**Slope**	**R^2 Statistic**	**Test Against RC Regression[2]**	**Test Against RC-11 Regression[3]**
ECBC	0.793	0.584	0.940	0.040	0.829
FAL	0.709	0.598	0.953	0.024	0.909
IIVS	0.710	0.584	0.949	0.041	0.933
	NHK Regression[1]				
Laboratory	**Intercept**	**Slope**	**R^2 Statistic**	**Test Against RC Regression[2]**	**Test Against RC-11 Regression[3]**
ECBC	0.401	0.530	0.577	0.620	0.805
FAL	0.429	0.548	0.621	0.569	0.853
IIVS	0.373	0.549	0.590	0.538	0.755

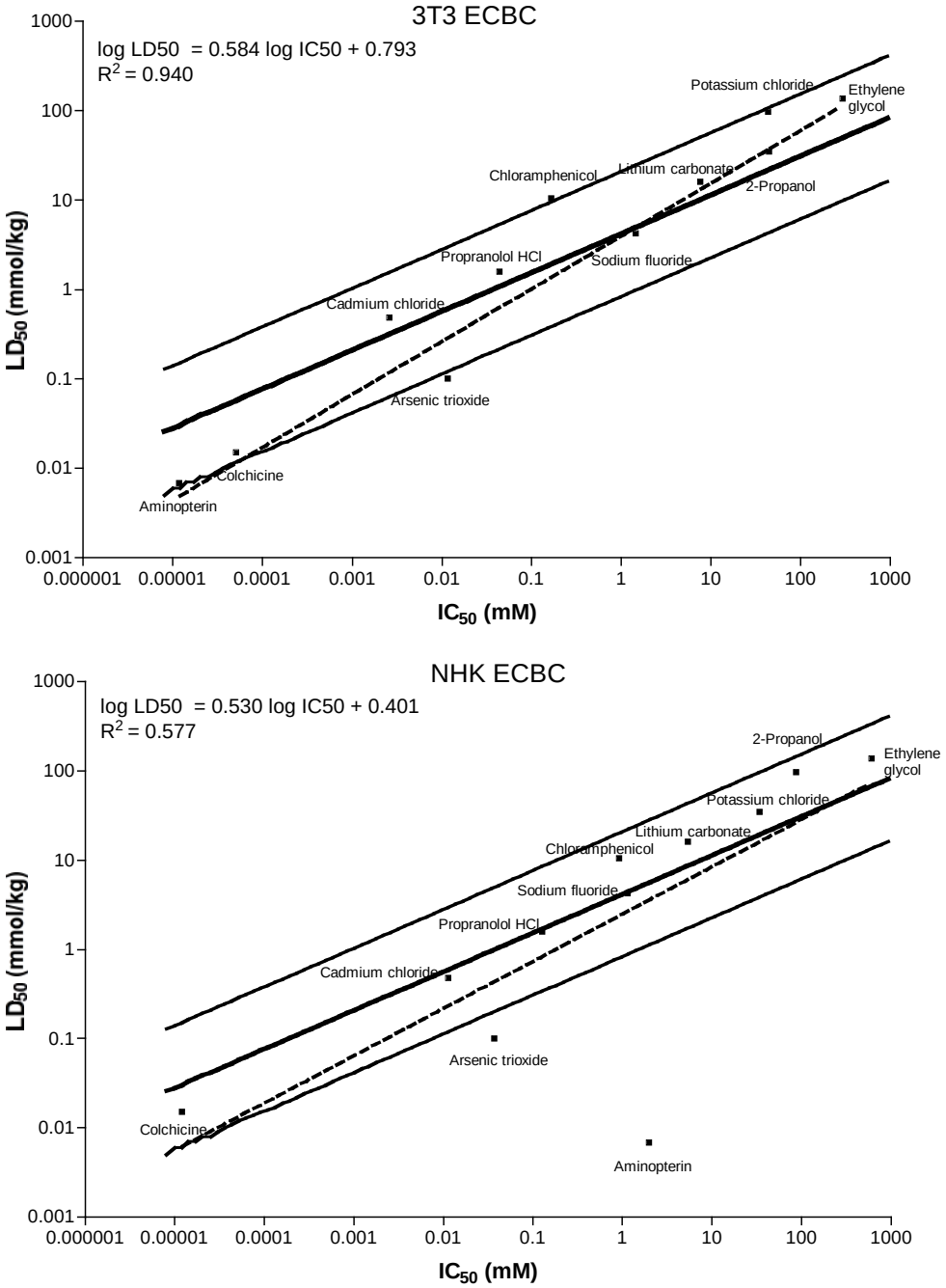
Abbreviations: ECBC=Edgewood Chemical Biological Center; FAL=Fund for the Replacement of Animals in Medical Experiments Alternatives Laboratory; IIVS=Institute for *In Vitro* Sciences; RC=Registry of Cytotoxicity; 3T3=BALB/c 3T3 fibroblasts; NHK=Normal human epidermal keratinocytes; R^2=Coefficient of determination.

[1]Laboratory and test method regressions were calculated after log transforming the NRU IC_{50} in mM and the RC LD_{50} in mmol/kg for the 11 RC substances tested in study Phases Ib and II (shown in **Figures 6-6** through **6-8)**.

[2]Simultaneous comparison of slope and intercept with RC millimole regression: log LD_{50} (mmol/kg) = 0.435 x log IC_{50} (mM) + 0.625; R^2=0.452; the reported values are p values of the statistic.

[3]Simultaneous comparison of slope and intercept with RC-11 regression (defined as a regression on the 11 substances): log LD_{50} (mmol/kg) = 0.552 x log IC_{50} (mM) + 0.602; R^2=0.971; the reported values are p values of the statistic.

Figure 6-6 *In Vitro – In Vivo* Regressions[1] for Phases Ib and II for ECBC

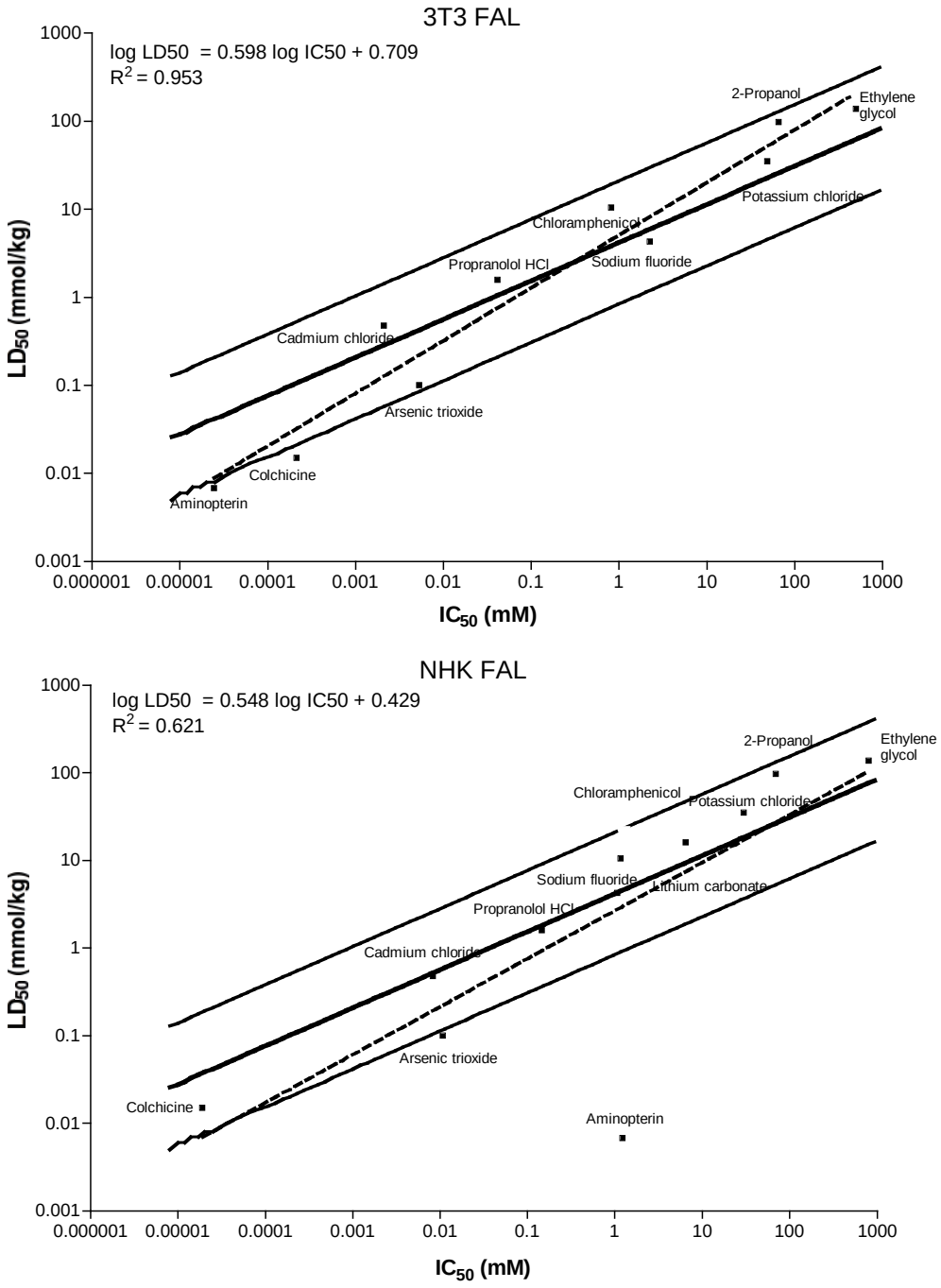


Abbreviations: ECBC=Edgewood Chemical Biological Center; RC=Registry of Cytotoxicity; 3T3=Neutral red uptake using BALB/c 3T3 fibroblasts; NHK= Neutral red uptake using normal human epidermal keratinocytes; R^2=Coefficient of determination.

[1]Regressions of substances tested in study Phases Ib and II do not include sodium selenate because it was not included in the RC. Regressions were calculated using the NRU IC_{50} values and the RC LD_{50} values.

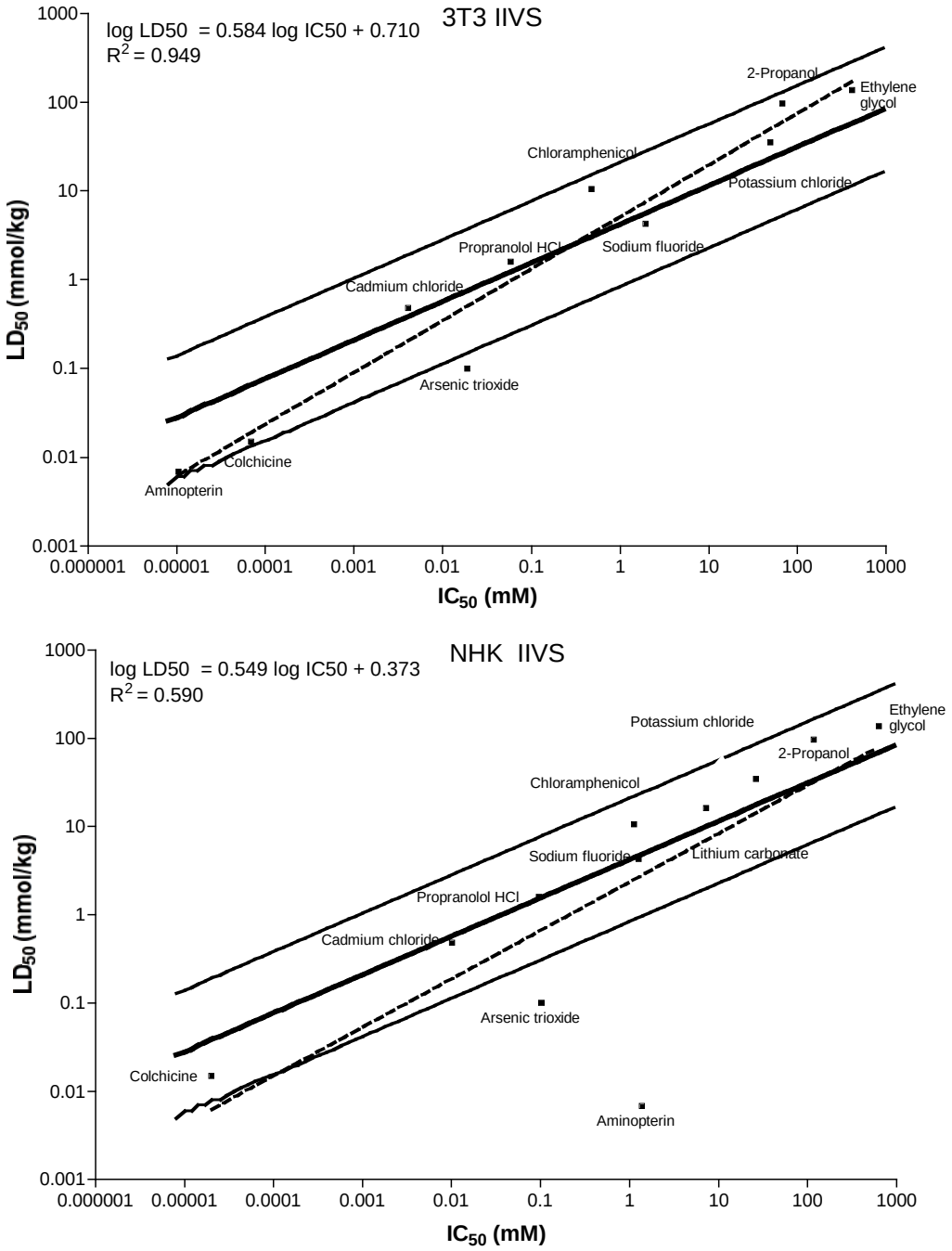
The solid lines show RC millimole regression (bold) and acceptance limits (lighter). The dashed shows the ECBC regressions.

Figure 6-7 *In Vitro – In Vivo* Regressions[1] for Phases Ib and II for FAL



Abbreviations: FAL=Fund for the Replacement of Animals in Medical Experiments Alternatives Laboratory RC=Registry of Cytotoxicity; 3T3=Neutral red uptake using BALB/c 3T3 fibroblasts; NHK=Neutral red uptake using normal human epidermal keratinocytes; R^2=Coefficient of determination.
[1]Regressions of substances tested in study Phases Ib and II do not include sodium selenate because it was not included in the RC. Regressions were calculated using the NRU IC_{50} values and the RC LD_{50} values.
The solid lines show RC millimole regression (bold) and acceptance limits (lighter). The dashed shows the FAL regressions.

Figure 6-8 *In Vitro – In Vivo* Regressions[1] for Phases Ib and II for IIVS



Abbreviations: IIVS=Institute for *In Vitro* Sciences; RC=Registry of Cytotoxicity; 3T3=Neutral red uptake using BALB/c 3T3 fibroblasts; NHK=Neutral red uptake using normal human epidermal keratinocytes; R^2=Coefficient of determination.

[1]Regressions of substances tested in study Phases Ib and II do not include sodium selenate because it was not included in the RC. Regressions were calculated using the NRU IC_{50} values and the RC LD_{50} values.
The solid lines show RC millimole regression (bold) and acceptance limits (lighter). The dashed shows the IIVS regressions.

6.9 Summary

The millimole regressions developed using the validation study IC_{50} and LD_{50} values were not significantly different from the regressions for the same 47 RC substances using the RC data (F test; p=0.612 for the 3T3 regression and p=0.759 for the NHK regression). Because this validation study provided results similar to the RC, which has more than 3.5 times the number of substances, the 282 RC substances with rat LD_{50} values were used to determine the relationship between the IC_{50} and LD_{50} data. One linear regression was developed using millimole units for the measurement of substances, the RC rat-only millimole regression, and one was developed using weight units (which are more practical in a routine testing situation), the RC rat-only weight regression. The RC rat-only millimole regression is applicable to substances of known molecular weight while the RC rat-only weight regression is applicable for use with complex mixtures, substances whose molecular weight is unknown.

Characteristics that seemed promising for characterizing the RC millimole regression outliers were chemical class, boiling point, molecular weight, and log K_{ow}. Different chemical classes behaved differently with respect to being outliers; ranging from 5/5 (100%) for the organic sulfur compounds for both test methods to 4/14 (29%) for carboxylic acids for the 3T3 NRU. Of the reference substances with boiling points >200°C, 9/13 (69%) were outliers for the 3T3 NRU and 8/13 (62%) were outliers for the NHK NRU. With respect to molecular weights, 4/7 (57%) substances with molecular weight >400 g/mole were outliers using the 3T3 data, and 3/7 (43%) were outliers using the NHK data. When log K_{ow} was used, 8/13 (62%) substances with a log K_{ow} >3 were outliers for both test methods.

The lack of fit of individual substances to the RC millimole regression was not consistently related to insolubility or to the fact that the test method systems had little to no metabolic capability. Of the substances that exhibited precipitation, 11/25 (44%) were outliers in the 3T3 NRU assays and 11/24 (46%) were outliers in the NHK NRU assays. However, although the 3T3 and NHK cells have little to no metabolic capability, the toxicity of substances known to produce active metabolites *in vivo* was not underpredicted by these assays. Of the 19 substances known to produce active metabolites *in vivo*, 10 (53%) were outliers in the 3T3 NRU test method; the toxicity of six (60%) was underpredicted while the toxicity of four (40%) overpredicted. These 10 substances accounted for 36% of the 28 outliers identified by the 3T3 NRU test method. Similarly, nine (47%) of the 19 substances known to produce active metabolites *in vivo* were outliers in the NHK NRU test method. Of these nine, the NHK NRU test method underpredicted the toxicity of five (56%) and overpredicted four (44%). These nine outliers accounted for 29% of the 31 outliers identified by the NHK NRU test method.

The examination of outliers based on mechanisms of toxicity showed that 10/16 (63%) substances with specific neurotoxic mechanisms were outliers in both the 3T3 and NHK NRU test methods. Three of the six (50%) cardiotoxic substances were outliers in the 3T3 NRU test method and two (33%) were outliers in the NHK NRU test method. When all the reference substances with mechanisms of toxicity that are not expected to be active in the 3T3 and NHK systems (i.e., in **Table 6-3**) were summed, 13/22 (59%) were outliers for the 3T3 NRU and 12/22 (55%) were outliers for the NHK NRU.

The accuracy of the 3T3 and NHK NRU test methods for predicting the GHS acute oral toxicity categories was 31% (21/67) and 29% (20/68), respectively, when used with the RC rat-only millimole regression. The corresponding accuracy with the RC rat-only weight regression was 31% for both methods (21/67 for 3T3, and 21/68 for NHK). Accuracy was highest for substances in the $300< LD_{50} \leq 2000$ mg/kg range. The accuracies of the regressions, with respect to the GHS categories, were similar for both regressions (millimole and weight) and all three laboratories.

- 0% for substances with $LD_{50} \leq 5$ mg/kg (GHS Category I)
- 9% to 18% for substances with $5< LD_{50} \leq 50$ mg/kg (GHS Category II)
- 33% to 50% for substances with $50< LD_{50} \leq 300$ mg/kg (GHS Category III)
- 75% to 81% for substances with $300< LD_{50} \leq 2000$ mg/kg (GHS Category IV)
- 0% to 40% for substances with $2000< LD_{50} \leq 5000$ mg/kg (GHS Category V)
- 0% to 17% for substances with $LD_{50} >5000$ mg/kg (GHS Unclassified)

The overall accuracy for prediction of GHS category prediction using the RC IC_{50} and LD_{50} values and the RC millimole regression was higher that that for the NRU test methods with the RC rat-only regressions (i.e., 40% for the RC vs. 29% to 31% for the NRU test methods and RC rat-only regressions). However, the pattern of accuracy for the GHS categories was similar. For all the accuracy analyses, the lowest accuracy was obtained for very toxic and very nontoxic substances and highest accuracy was obtained for substances with $300 < LD_{50} \leq 2000$ mg/kg.

The accuracy of GHS acute oral toxicity category predictions using the *in vitro* NRU test methods with the RC rat-only regressions obtained for the reference substances may or may not be broadly applicable to substances that might require acute oral toxicity testing. The reasons for the low accuracy obtained in this validation study include: the differences between cell cultures and whole animals regarding the absorption, distribution, availability, metabolism, and excretion of reference substances, and the presence or absence of toxicity targets; the skewness of the selection of substances for testing (with respect to fit to the regression); and the structure of the GHS acute oral toxicity categories.

[This Page Intentionally Left Blank]