[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

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

F-test for regression slopes and intercepts described in **Section 5.5.4.3** indicated that the 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

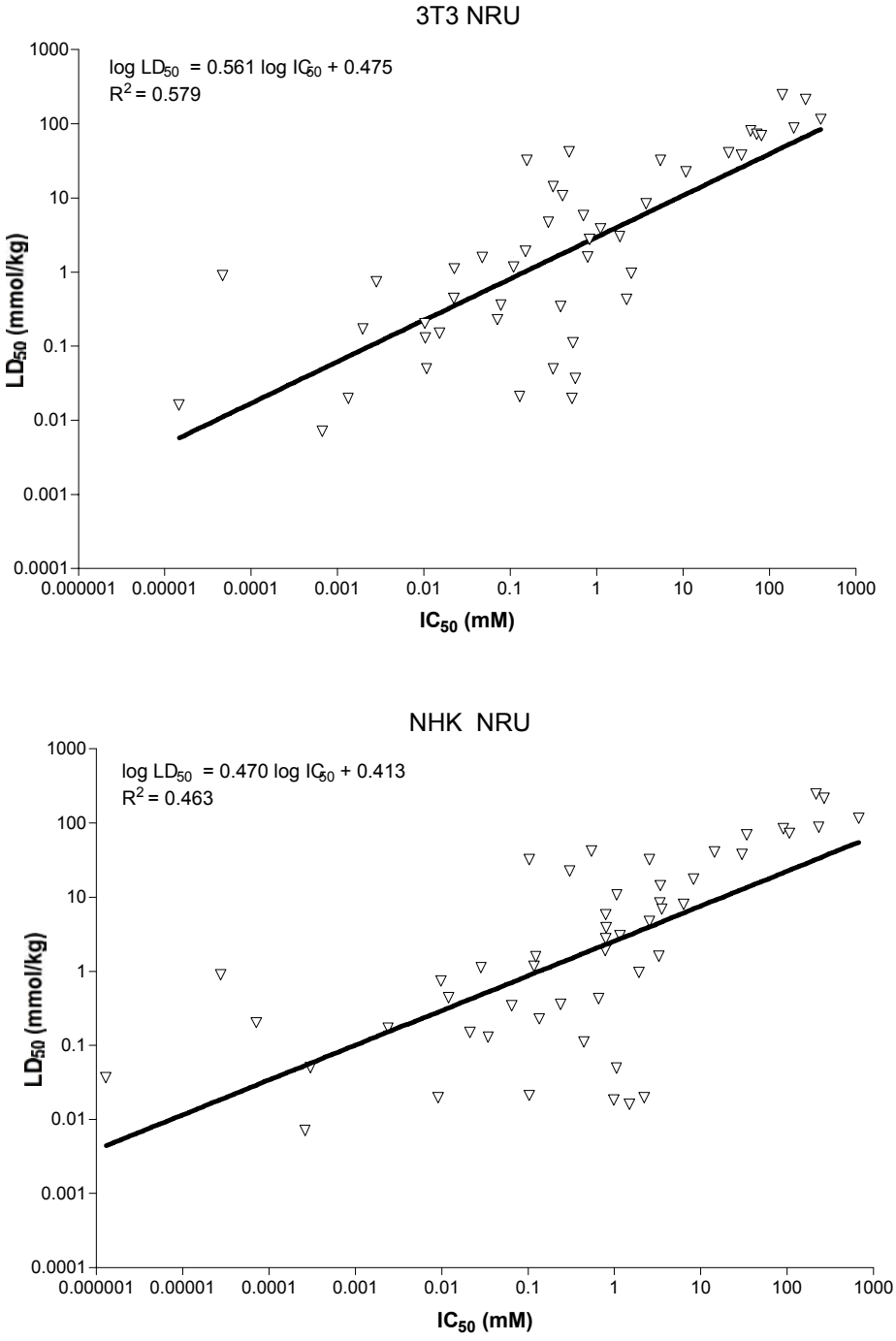
Abbreviations: ECBC=Edgewood Chemical Biological Center; FAL=Fund for the Replacement of Animals in Medical Experiments Alternatives Laboratory; IIVS=Institute for *In Vitro* Sciences; 3T3=BALB/c 3T3 fibroblasts; N=Number of substances used to calculate the regression; NHK=Normal human epidermal keratinocytes;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

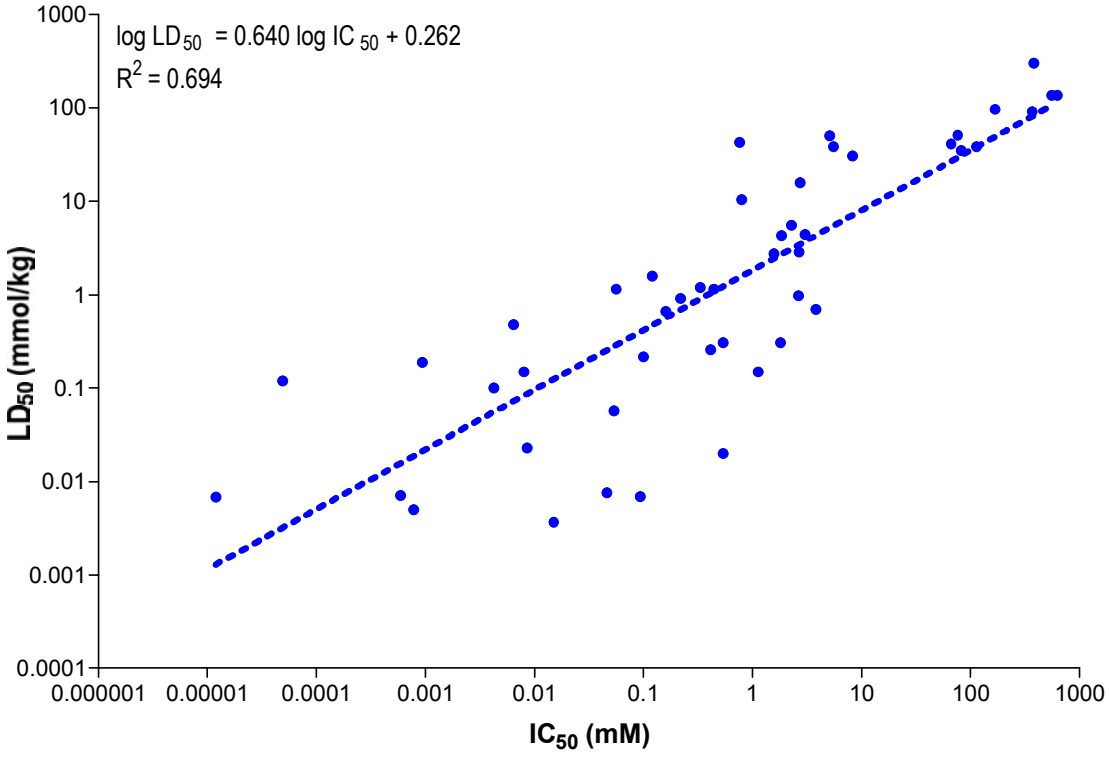
6.1.2 Comparison of the Combined-Laboratory 3T3 and NHK Regressions to the RC Millimole Regression

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**)

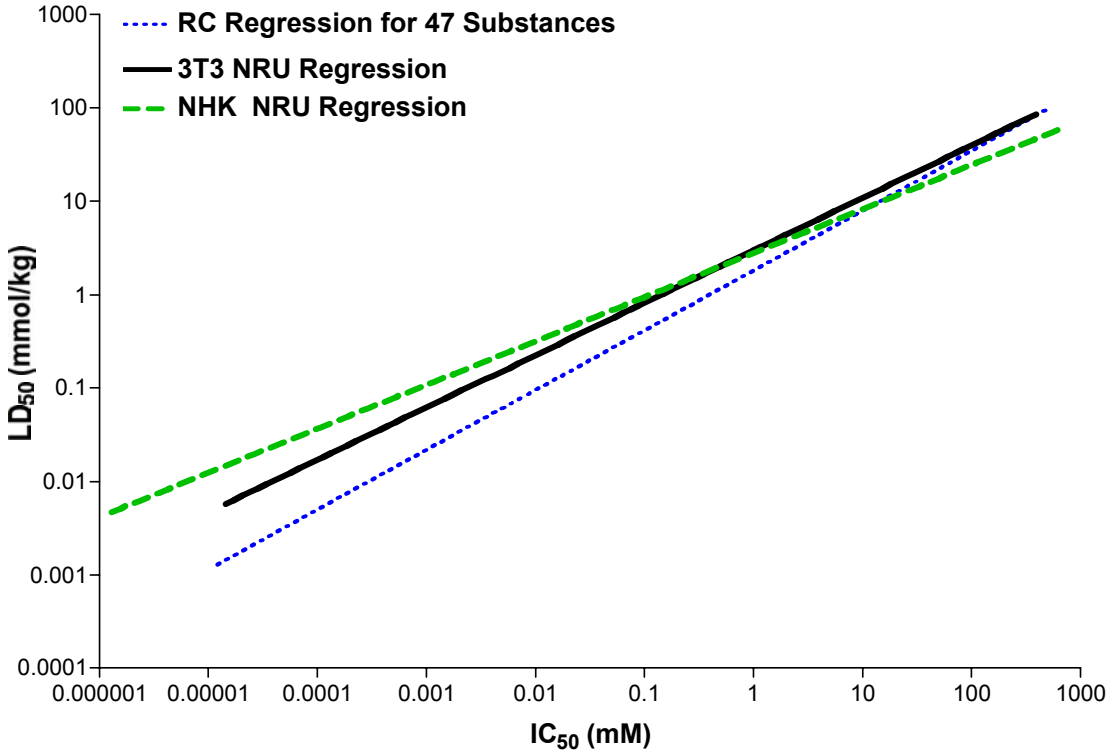
The regression calculated for the 47 substances using the RC IC_{50} and LD_{50} data is shown in **Figure 6-2**. A graphic comparison of the RC regressions and the 3T3 and NHK combined-laboratory regressions is in **Figure 6-3**. A statistical comparison of slope and intercept (simultaneously) using 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).

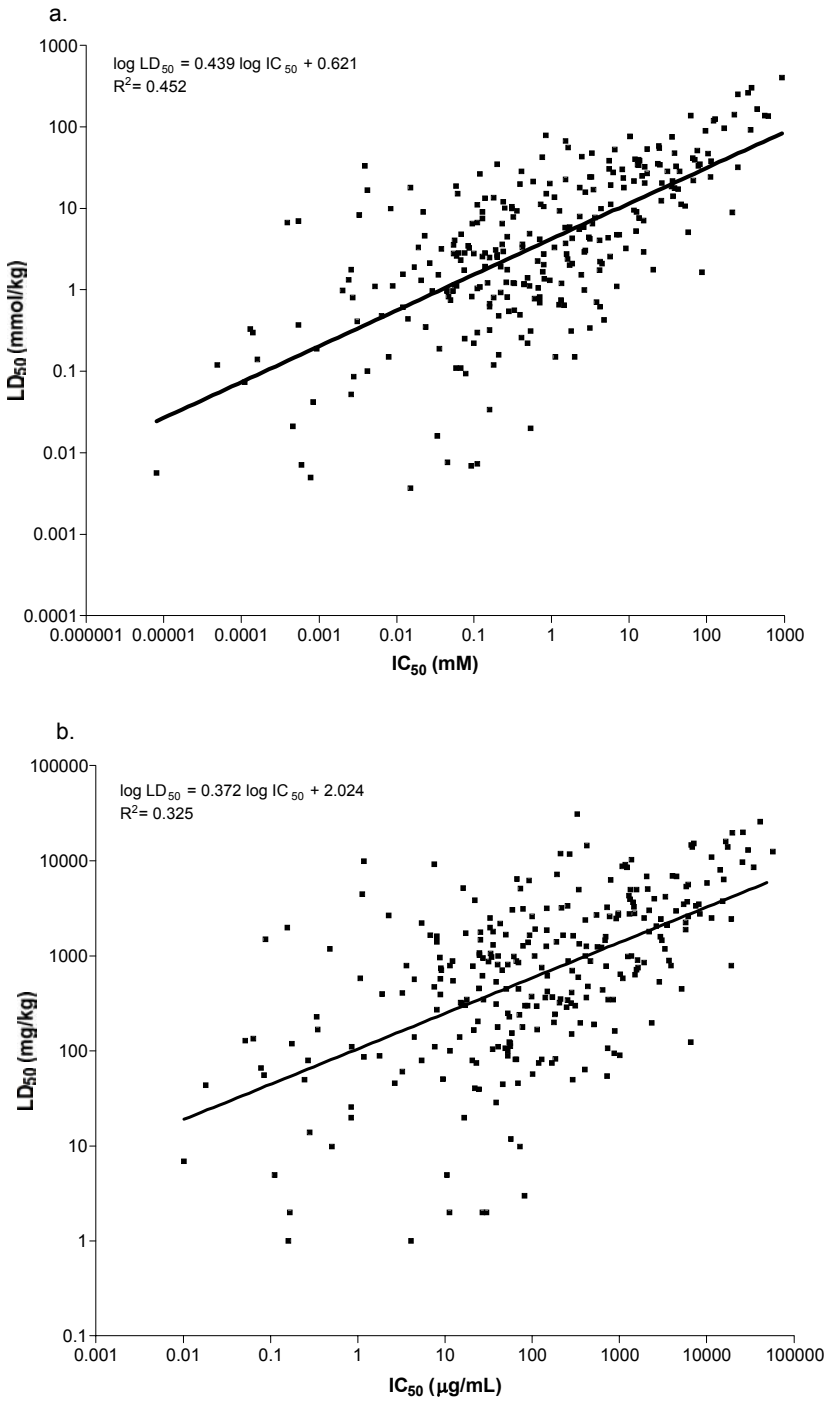
6.4 Accuracy of the 3T3 and NHK NRU Test Methods for Predicting GHS Acute Oral Toxicity Categories

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

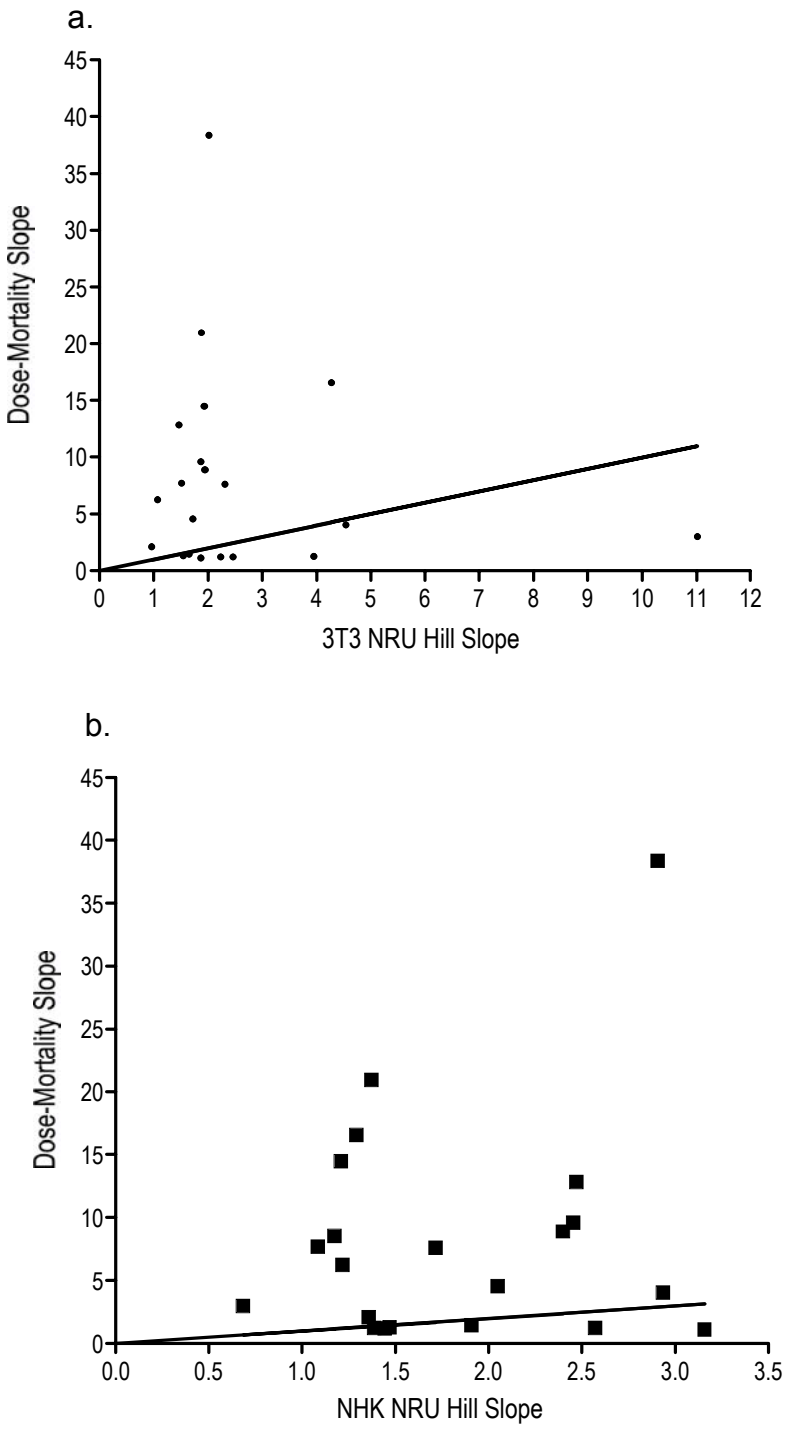
for the 3T3 and the NHK NRU test methods, respectively. The Hill function slopes were transformed to absolute values because geometric means cannot be calculated for negative numbers, and geometric mean Hill function slopes were calculated for the acceptable NRU tests for each reference substance. When 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

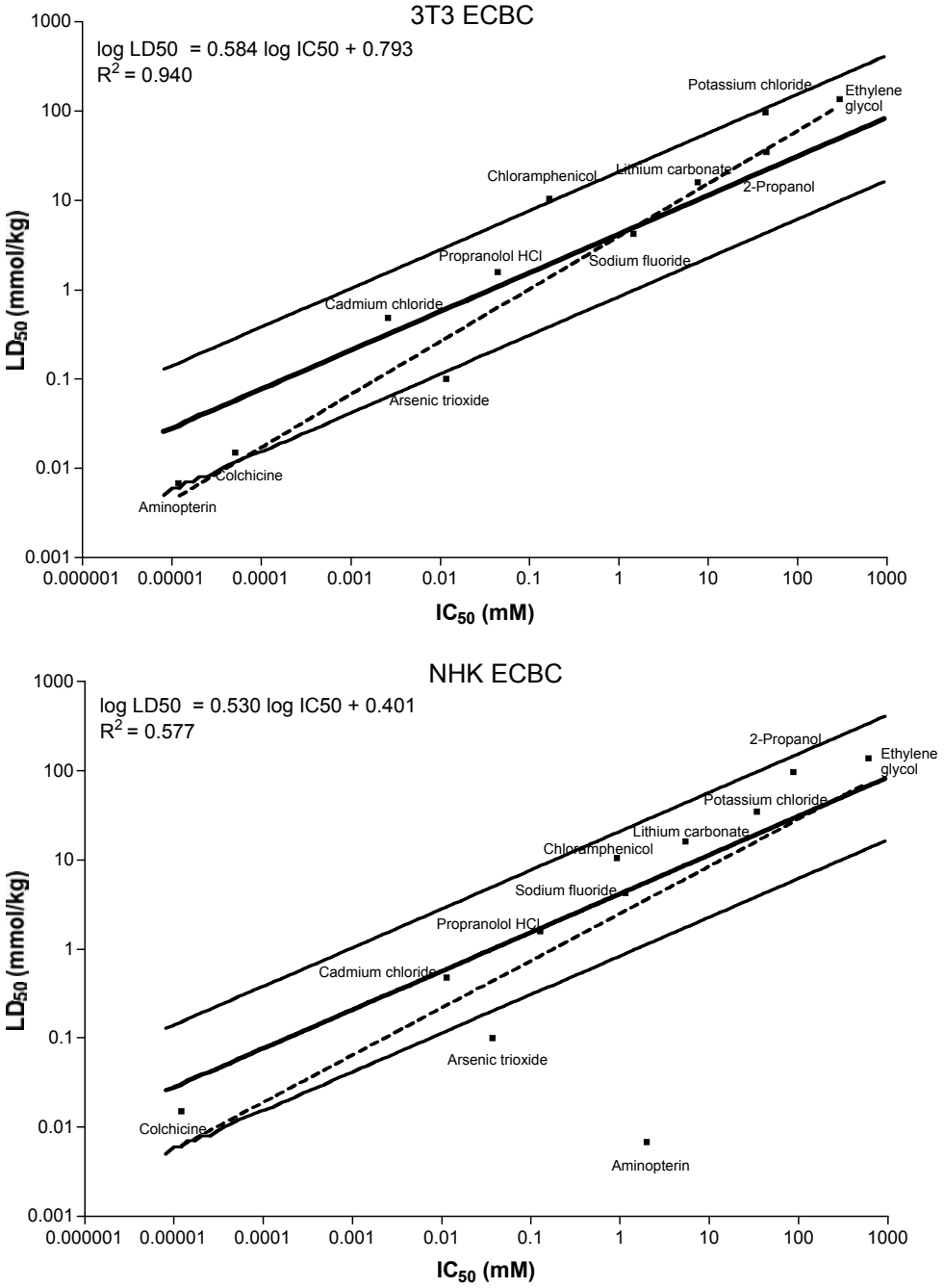
Abbreviations: ECBC=Edgewood Chemical Biological Center; FAL=Fund for the Replacement of Animals in Medical Experiments Alternatives Laboratory; IIVS=Institute for *In Vitro* Sciences; RC=Registry of Cytotoxicity; 3T3=BALB/c 3T3 fibroblasts; NHK=Normal human epidermal keratinocytes; R^2=Coefficient of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

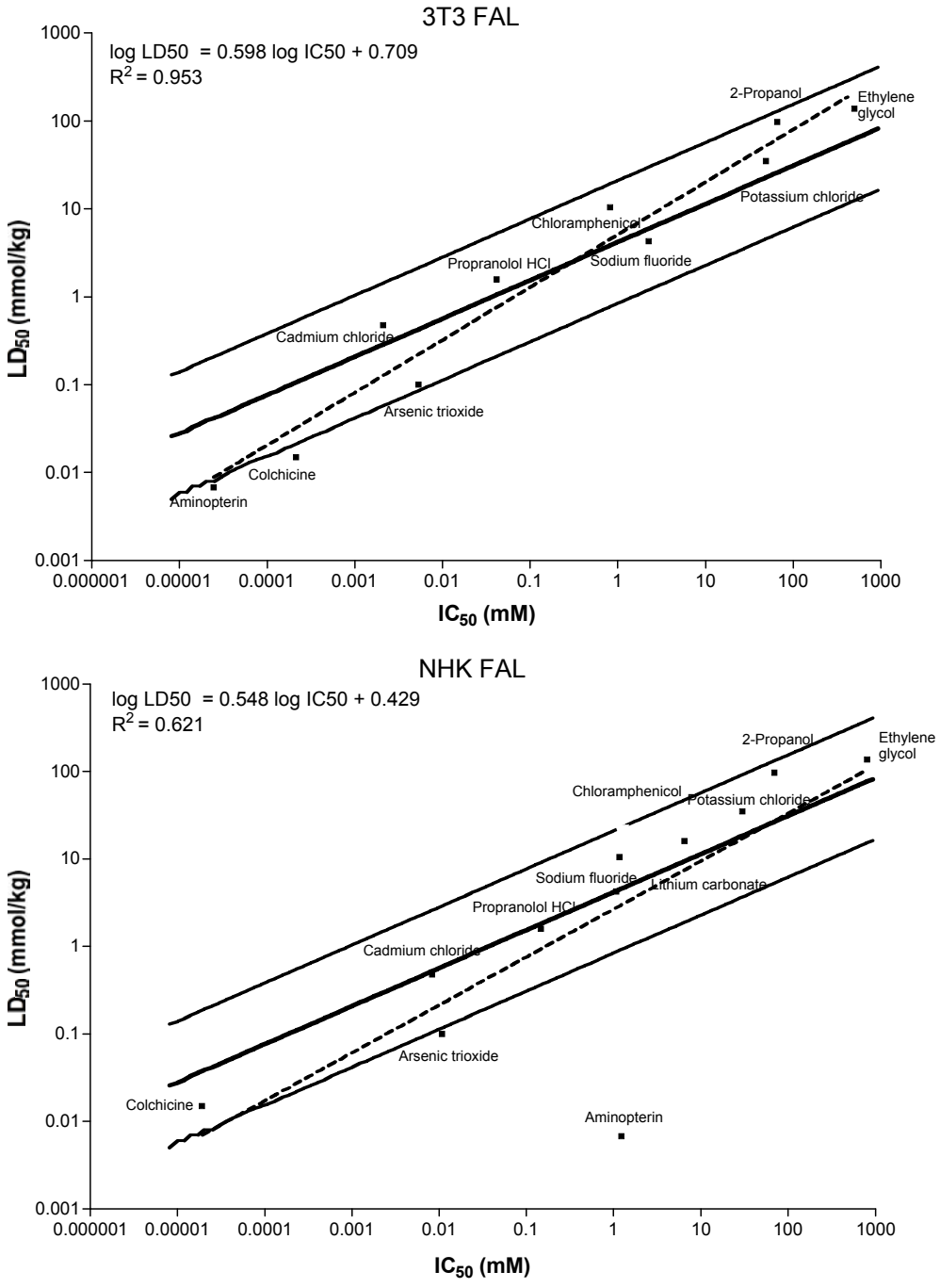


Abbreviations: ECBC=Edgewood Chemical Biological Center; RC=Registry of Cytotoxicity; 3T3=Neutral red uptake using BALB/c 3T3 fibroblasts; NHK= Neutral red uptake using normal human epidermal keratinocytes; R^2=Coefficient of determination.

[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.

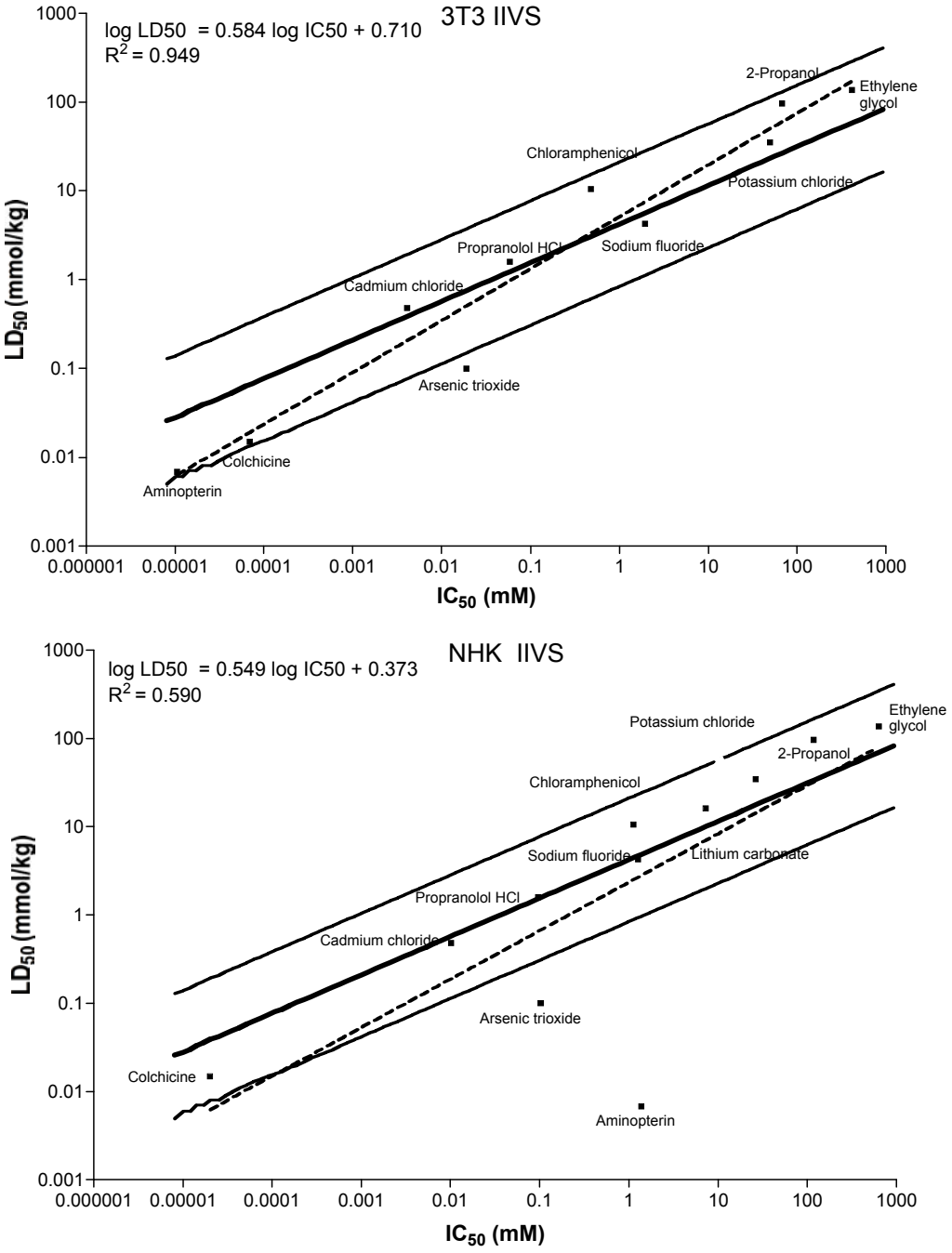
The solid lines show RC millimole regression (bold) and acceptance limits (lighter). The dashed shows the ECBC regressions.

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]

[This Page Intentionally Left Blank]

7.0 RELIABILITY OF THE 3T3 AND NHK NRU TEST METHODS

The reliability of the 3T3 and NHK NRU test methods was assessed by determining intra- and inter-laboratory reproducibility. Intralaboratory reproducibility is the agreement of results produced when people in the same laboratory perform the method using the same test protocol at different times (ICCVAM 2003). Interlaboratory reproducibility is the agreement of results among different laboratories using the same protocol and reference substances. Interlaboratory reproducibility indicates the extent to which a method can be successfully transferred among laboratories. Repeatability, usually applied to results within a laboratory, is the closeness of agreement between test results obtained when the procedure is performed on the same substance under identical conditions within a given time. This study was not designed to assess intralaboratory repeatability.

The interlaboratory reproducibility of the test results was assessed by comparing the laboratory-specific IC_{50}-LD_{50} regressions for the 3T3 and NHK NRU test methods to the mean (i.e., across-laboratory mean) laboratory regressions (see **Section 7.2.1**). This comparison is relevant because the 3T3 and NHK NRU test methods are intended for use with IC_{50}-LD_{50} regressions to determine starting doses for acute oral toxicity tests. Interlaboratory reproducibility of the 3T3 and NHK NRU test methods was also determined using ANOVA, CV analysis, and comparison of maximum:minimum IC_{50} ratios calculated using laboratory mean values **(**see **Sections 7.2.2**, **7.2.3**, and **7.2.4**, respectively), as discussed in **Section 5.5.2.2**. Inter- and intra-laboratory reproducibility of the PC (SLS) was determined using ANOVA, CV analysis, and/or linear regression over time (see **Section 7.3**). The extent of laboratory concordance in selecting the solvent to be used for each test substance (described in **Section 2.10**) is provided in **Section 7.4**.

7.1 Reference Substances Used to Determine the Reliability of the 3T3 and NHK NRU Test Methods

The validation study was designed for the purpose of using the IC_{50} results of 72 reference substances (see **Table 3-2)** to determine the reliability of the IC_{50} values from the 3T3 and NHK NRU test methods. The number of reference substances used for the reproducibility analysis was not the same as the number of reference substances used for the accuracy analyses in **Section 6.4**. In the former case, only reference substances for which all three laboratories reported replicate IC_{50} values were used, while in the latter case, substances with rat acute oral LD_{50} data only and at least one laboratory reporting replicate IC_{50} values were used. **Table 7-1** lists the reference substances that failed to yield sufficient toxicity for the calculation of an IC_{50} in each laboratory, and the number of remaining reference substances with replicate IC_{50} values. The laboratories obtained acceptable IC_{50} values for 66 to 68 reference substances using the 3T3 NRU test method, and for 69 to 70 substances using the NHK NRU test method. When only reference substances with IC_{50} values from all three laboratories are considered, 64 and 68 substances were available to evaluate the reliability of the 3T3 and NHK NRU test methods, respectively. The substances that were excluded from the 3T3 reliability analysis were carbon tetrachloride, disulfoton, gibberellic acid, lithium carbonate, methanol, 1,1,1-trichloroethane, valproic acid, and xylene. The substances that were excluded from the NHK reliability analysis were carbon tetrachloride, methanol, 1,1,1-trichloroethane, and xylene.

Table 7-1 Reference Substances Excluded from Reproducibility Analyses Because of Insufficient Cytotoxicity

Laboratory	3T3 NRU Test Method		NHK NRU Test Method	
	Reference Substances Lacking IC_{50} Results	N[1]	Reference Substances Lacking IC_{50} Results	N[1]
ECBC	Carbon tetrachloride Methanol 1,1,1-Trichloroethane Xylene	68	Carbon tetrachloride Methanol Xylene	69
FAL	Carbon tetrachloride Disulfoton Gibberellic acid Lithium carbonate Methanol Xylene	66	1,1,1-Trichloroethane Carbon tetrachloride Xylene	69
IIVS	Carbon tetrachloride Lithium carbonate Methanol Valproic acid	68	Carbon tetrachloride 1,1,1-Trichloroethane	70

Abbreviations: 3T3=BALB/c 3T3 fibroblasts; NHK=Normal human epidermal keratinocytes; NRU= Neutral red uptake; ECBC=Edgewood Chemical Biological Center; FAL=Fund for the Replacement of Animals in Medical Experiments Alternatives Laboratory; IIVS=Institute for *In Vitro* Sciences; N=Number of substances.
[2]Number of substances with replicate IC_{50} values.

Despite the fact that IC_{50} values were not obtained by all the laboratories for all reference substances, **Table 7-2** shows that the complete range of LD_{50} responses, as defined by the GHS classification for acute oral toxicity in **Table 3-1**, was covered by the reference substances for which replicate IC_{50} values were obtained. The 3T3 NRU IC_{50} values ranged from 0.005 to 38,878 μg/mL, while the NHK values covered a larger range, from 0.00005 to 49,800 μg/mL (see **Tables 5-4** and **5-5**).

Table 7-2 Number of Reference Substances Tested vs Number of Reference Substances Yielding IC_{50} Values from Each Laboratory, by GHS Acute Oral Toxicity Category

GHS Category[1] (mg/kg)	Reference Oral LD_{50}[2]	3T3 NRU Test Method[3]	NHK NRU Test Method[3]
$LD_{50} \leq 5$	7	6	7
$5 < LD_{50} \leq 50$	12	12	12
$50 < LD_{50} \leq 300$	12	12	12
$300 < LD_{50} \leq 2000$	16	14	16
$2000 < LD_{50} \leq 5000$	11	9	9
$LD_{50} > 5000$	14	11	12

Abbreviations: 3T3=BALB/c 3T3 fibroblasts; NHK=Normal human epidermal keratinocytes; NRU= Neutral red uptake; GHS=Globally Harmonized System of Classification and Labelling of Chemicals (UN 2005).
[1]GHS category for acute oral toxicity.
[2]Number of reference substances tested in each category. Reference acute oral LD_{50} values from rats and mice were generated after evaluating LD_{50} values located through literature searches and references from toxicity databases such as RTECS® (from **Table 4-2**).
[3]Number of reference substances with IC_{50} values from all three laboratories.

7.2 Reproducibility Analyses for the 3T3 and NHK NRU Test Methods

The interlaboratory reproducibility of the 3T3 and NHK NRU IC_{50} values was assessed by comparing the laboratory-specific IC_{50}-LD_{50} linear regressions for each method to a regression calculated using the mean IC_{50} values of the laboratories. The interlaboratory reproducibility of the 3T3 and NHK NRU test methods was also assessed using ANOVA, CV analysis, and analysis of the laboratory mean maximum:minimum IC_{50} ratios, as described in **Section 5.5.2.2.** Intralaboratory reproducibility was assessed using a CV analysis.

7.2.1 Comparison of Laboratory-Specific IC_{50}-LD_{50} Linear Regression Analyses to the Mean Laboratory Regression

The comparisons of laboratory-specific IC_{50}-LD_{50} linear regressions to the mean laboratory regression for each method were made because the 3T3 and NHK NRU test methods are intended for use with IC_{50}-LD_{50} regressions to determine starting doses for acute oral toxicity tests. Laboratory-specific IC_{50}-LD_{50} linear regressions were generated and displayed graphically for each method using the 64 and 68 reference substances for the 3T3 and NHK NRU test methods, respectively, as indicated in **Section 7.1**. The regressions used the geometric mean IC_{50} values for each substance with the rodent acute oral LD_{50} reference value (**Table 4-2**). To determine whether the laboratory-specific regressions were significantly different from one another, they were compared against the mean laboratory regression for each NRU test method that was calculated using the geometric mean of the laboratory mean IC_{50} values and the rodent acute oral LD_{50} reference values. The mean laboratory regression for each NRU test method is in **Figure 7-1** with 95% confidence limits, and shows that the laboratory-specific regressions were all within the 95% confidence limits of the mean laboratory regression.

7.2.2 ANOVA Results for the 3T3 and NHK NRU Test Methods

The ANOVA was performed as discussed in Section **5.5.2.2**. Because the sample sizes from this study were small, usually three observations per laboratory, there may be differences that were statistically significant only because there were too few observations within the laboratories to adequately characterize variability or because the within-laboratory variability was small.

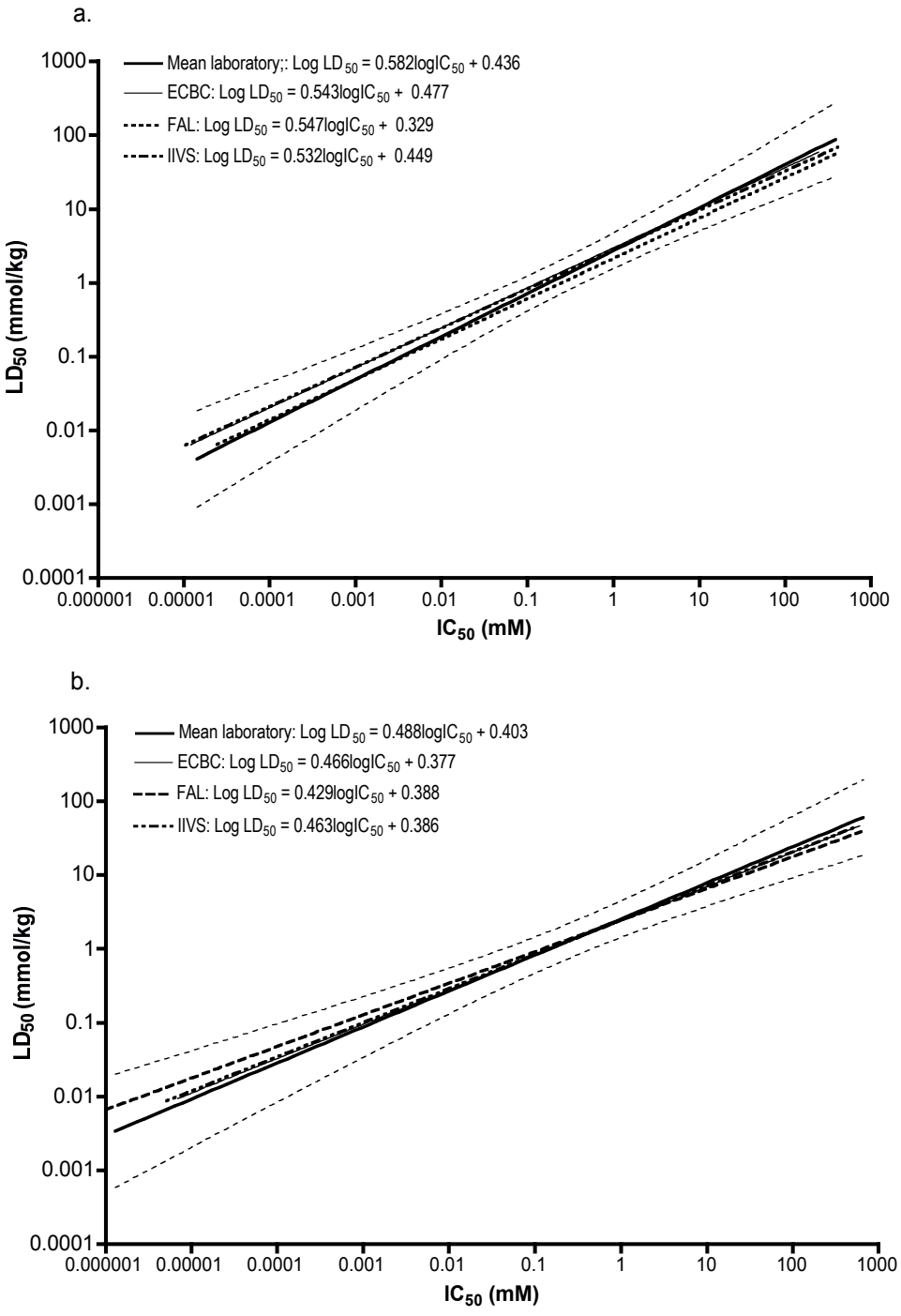
7.2.2.1 *Differences Among the IC_{50} Values in Laboratories Using the 3T3 NRU Test Method*

The ANOVA results in **Table 7-3** show that there were statistically significant ($p < 0.01$) differences among the laboratories for 23 of the 64 (36%) reference substances evaluated. The p values from the contrast analyses, post-hoc tests to determine which laboratory was significantly different from the others at $p < 0.01$ (see **Section 5.5.2.2)**, are also provided in **Table 7-3**. The substances for which statistically significant ANOVA and contrast results were obtained are listed in **Table 7-4** along with columns showing the laboratory with significantly differing values from the other two laboratories. Because significant laboratory differences may have resulted from the insolubility or volatility of the test substance, **Table 7-4** also indicates whether any laboratory reported insolubility or volatility during conduct of the test. Insolubility was suggested by the presence of precipitates in either the stock solutions or in cell culture. Volatility was identified by the need for plate sealers to contain volatile contamination of lower concentration wells by higher concentrations. Insolubility and volatility were reported for only six of the 23 chemicals showing significant

interlaboratory variability. In contrast, 22 of the 41 substances that were classified as generating interlaboratory reproducible data exhibited precipitates and/or volatility.

For the 23 substances that yielded significantly different results among laboratories, contrast analyses indicated that the IC_{50} values produced by ECBC and FAL were frequently different from the other laboratories. ECBC tended to report the lowest IC_{50} values (i.e., highest toxicity) among the laboratories while FAL tended to report the highest values of the three laboratories. ECBC reported significantly different results from the other two laboratories for 15 of the 23 substances; for 13 of the 15, ECBC's mean value IC_{50} was the lowest among the laboratories. FAL reported significantly different results from the other two laboratories for 20 of the 23 substances; for 18 of the 20, FAL's IC_{50} value was the highest among the laboratories. IIVS reported significantly different values for 11 of the 26 substances, with no tendency toward highest or lowest IC_{50} values.

Figure 7-1 Mean Laboratory and Laboratory-Specific 3T3 and NHK NRU Regressions



Abbreviations: 3T3=BALB/c 3T3 fibroblasts; NRU=Neutral red uptake; NHK=Normal human epidermal keratinocytes. Solid lines show the mean laboratory linear regressions for the 3T3 NRU (a) and the NHK NRU (b) test methods with dashed curved lines to show the 95% confidence limits of the regression. The regressions were calculated using 64 and 68 reference substances for the 3T3 and NHK NRU test methods, respectively, as described in **Section 7.1**. Regressions used geometric mean IC_{50} values and reference acute oral LD_{50} values from **Table 4-2**.

Table 7-3 Interlaboratory Reproducibility of the IC_{50} Values from the 3T3 NRU Test Method

Reference Substance/Laboratory	Arithmetic Mean IC_{50} (mg/mL)[1]	Maximum: Minimum Mean IC_{50} Ratio[2]	Arithmetic IntraLab %CV	Arithmetic InterLab %CV	Log Arithmetic Mean IC_{50} (mg/mL)[1]	ANOVA P[3]	Contrast P[4]
Acetaminophen	50.1	1.6		28	1.7	0.171	
ECBC	40.8		22		1.61		NA
FAL	66.2		35		1.82		NA
IIVS	43.4		26		1.64		NA
Acetonitrile	8484	1.5		21	3.93	0.553	
ECBC	6433		2		3.81		NA
FAL	9690		58		3.99		NA
IIVS	9330		13		3.97		NA
Acetylsalicylic acid	760	3.1		56	2.88	<0.001	
ECBC	646		10		2.81		0.581
FAL	1234		24		3.09		<0.001
IIVS	401		16		2.6		<0.001
5-Aminosalicylic acid	1698	1.4		19	3.23	0.054	
ECBC	1467		14		3.17		NA
FAL	2070		16		3.32		NA
IIVS	1557		12		3.19		NA
Aminopterin	0.007	2.4		54	-2.14	0.036	
ECBC	0.005		20		-2.28		NA
FAL	0.012		46		-1.93		NA
IIVS	0.005		23		-2.33		NA
Amitriptyline HCl	7.23	1.3		14	0.86	0.348	
ECBC	6.03		23		0.78		0.163
FAL	7.86		28		0.9		0.469
IIVS	7.81		18		0.89		0.445
Arsenic trioxide	2.51	3.9		61	0.4	0.004	

Table 7-3 Interlaboratory Reproducibility of the IC_{50} Values from the 3T3 NRU Test Method

Reference Substance/Laboratory	Arithmetic Mean IC_{50} (mg/mL)[1]	Maximum: Minimum Mean IC_{50} Ratio[2]	Arithmetic IntraLab %CV	Arithmetic InterLab %CV	Log Arithmetic Mean IC_{50} (mg/mL)[1]	ANOVA P[3]	Contrast P[4]
ECBC	2.41		33		0.38		0.527
FAL	1.04		7		0.02		0.002
IIVS	4.09		52		0.61		0.006
Atropine sulfate	85.6	2.5		49	1.93	0.049	
ECBC	54.1		55		1.73		NA
FAL	133		31		2.12		NA
IIVS	70		8		1.85		NA
Boric acid	2228	3.3		69	3.35	0.01	
ECBC	1497		32		3.18		NA
FAL	3987		17		3.6		NA
IIVS	1202		48		3.08		NA
Busulfan	135	8.0		119	2.13	0.002	
ECBC	40		48		1.6		0.012
FAL	321		56		2.51		<0.001
IIVS	43.7		4		1.64		0.033
Cadmium chloride	0.565	1.4		39	-0.25	0.124	
ECBC	0.48		14		-0.32		NA
FAL	0.4		32		-0.4		NA
IIVS	0.817		53		-0.09		NA
Caffeine	161	1.4		18	2.21	0.481	
ECBC	133		10		2.12		NA
FAL	157		52		2.2		NA
IIVS	191		7.5		2.28		NA
Carbamazepine	109	1.8		35	2.04	0.049	
ECBC	83		14		1.92		NA

Table 7-3 Interlaboratory Reproducibility of the IC_{50} Values from the 3T3 NRU Test Method

Reference Substance/Laboratory	Arithmetic Mean IC_{50} (mg/mL)[1]	Maximum: Minimum Mean IC_{50} Ratio[2]	Arithmetic IntraLab %CV	Arithmetic InterLab %CV	Log Arithmetic Mean IC_{50} (mg/mL)[1]	ANOVA P[3]	Contrast P[4]
FAL	152		37		2.18		NA
IIVS	91.8		12		1.96		NA
Carbon tetrachloride	NA	NA		NA	NA	NA	
ECBC	NA		NA		NA		NA
FAL	NA		NA		NA		NA
IIVS	NA		NA		NA		NA
Chloral hydrate	187	1.6		25	2.27	0.004	
ECBC	151		10		2.18		0.008
FAL	241		10		2.38		0.002
IIVS	170		12		2.23		0.181
Chloramphenicol	161	4.9		67	2.21	<0.001	
ECBC	55.3		22		1.74		<0.001
FAL	273		30		2.44		0.001
IIVS	156		18		2.19		0.165
Citric acid	829	2.4		41	2.92	0.002	
ECBC	473		29		2.68		0.001
FAL	1148		13		3.06		0.003
IIVS	865		19		2.94		0.298
Colchicine	0.047	4.7		85	-1.33	0.001	
ECBC	0.02		11		-1.70		0.0028
FAL	0.093		45		-1.03		0.0005
IIVS	0.028		1		-1.55		0.0914
Cupric sulfate pentahydrate	70.6	21.6		85	1.85	<0.001	
ECBC	82.7		4		1.92		0.001
FAL	123		44		2.09		<0.001

Table 7-3 Interlaboratory Reproducibility of the IC_{50} Values from the 3T3 NRU Test Method

Reference Substance/Laboratory	Arithmetic Mean IC_{50} (mg/mL)[1]	Maximum: Minimum Mean IC_{50} Ratio[2]	Arithmetic IntraLab %CV	Arithmetic InterLab %CV	Log Arithmetic Mean IC_{50} (mg/mL)[1]	ANOVA P[3]	Contrast P[4]
IIVS	5.7		31		0.76		<0.001
Cycloheximide	0.293	5.9		104	-0.53	0.021	
ECBC	0.125		45		-0.9		NA
FAL	0.647		70		-0.19		NA
IIVS	0.109		23		-0.96		NA
Dibutyl phthalate	78.3	9.2		124	1.89	<0.001	
ECBC	23.5		17		1.37		0.012
FAL	191		50		2.28		<0.001
IIVS	20.7		7		1.32		0.005
Dichlorvos	20.3	3.3		57	1.31	0.002	
ECBC	9.8		35		0.99		0.001
FAL	32.8		6		1.52		0.002
IIVS	18.3		11		1.26		0.823
Diethyl phthalate	113	1.7		28	2.05	0.127	
ECBC	85.5		34		1.93		0.092
FAL	147		26		2.17		0.07
IIVS	106		24		2.03		0.846
Digoxin	520	2.8		62	2.72	0.043	
ECBC	351		39		2.54		NA
FAL	892		36		2.95		NA
IIVS	317		21		2.5		NA
Dimethylformamide	5242	1.1		6	3.72	0.296	
ECBC	5343		10		3.73		NA
FAL	5483		9		3.74		NA
IIVS	4900		4		3.69		NA

Table 7-3 Interlaboratory Reproducibility of the IC_{50} Values from the 3T3 NRU Test Method

Reference Substance/Laboratory	Arithmetic Mean IC_{50} (mg/mL)[1]	Maximum: Minimum Mean IC_{50} Ratio[2]	Arithmetic IntraLab %CV	Arithmetic InterLab %CV	Log Arithmetic Mean IC_{50} (mg/mL)[1]	ANOVA P[3]	Contrast P[4]
Diquat dibromide monohydrate	15.1	9.3		120	1.18	0.017	
ECBC	3.9		23		0.59		NA
FAL	36.1		98		1.56		NA
IIVS	5.4		25		0.73		NA
Disulfoton	98.6	2.3		55	1.99	0.003	
ECBC	137		55		2.14		NA
FAL	NA		NA		NA		NA
IIVS	60.4		87		1.78		NA
Endosulfan	8.02	4.2		78	0.9	0.046	
ECBC	5.3		57		0.72		NA
FAL	15.2		78		1.18		NA
IIVS	3.6		42		0.56		NA
Epinephrine bitartrate	59.4	1.2		12	1.77	0.048	
ECBC	51.5		12		1.71		NA
FAL	63.4		11		1.8		NA
IIVS	63.4		3		1.8		NA
Ethanol	6731	1.6		23	3.83	0.075	
ECBC	5360		33		3.73		NA
FAL	8420		14		3.93		NA
IIVS	6413		5		3.81		NA
Ethylene glycol	25292	1.7		26	4.4	0.007	
ECBC	18325		9		4.26		0.004
FAL	31650		24		4.50		0.01
IIVS	25900		12		4.41		0.505
Fenpropathrin	27.2	2.5		49	1.43	0.301	

Table 7-3 Interlaboratory Reproducibility of the IC_{50} Values from the 3T3 NRU Test Method

Reference Substance/Laboratory	Arithmetic Mean IC_{50} (mg/mL)[1]	Maximum: Minimum Mean IC_{50} Ratio[2]	Arithmetic IntraLab %CV	Arithmetic InterLab %CV	Log Arithmetic Mean IC_{50} (mg/mL)[1]	ANOVA P[3]	Contrast P[4]
ECBC	22.6		11		1.35		NA
FAL	42.4		63		1.63		NA
IIVS	16.7		12		1.22		NA
Gibberellic Acid	7842	1.0		3	3.89	0.621	
ECBC	8027		11		3.9		NA
FAL	NA		NA		NA		NA
IIVS	7657		10		3.88		NA
Glutethimide	192	2.3		43	2.28	<0.001	
ECBC	167		4		2.22		0.029
FAL	284.3		7		2.45		<0.001
IIVS	125.3		7		2.1		<0.001
Glycerol	28904	1.9		33	4.46	0.846	
ECBC	20000		15		4.3		NA
FAL	38878		73		4.59		NA
IIVS	27833		39		4.44		NA
Haloperidol	6.26	1.5		24	0.8	0.006	
ECBC	5.3		12		0.72		0.03
FAL	8		8		0.9		0.002
IIVS	5.5		12		0.74		0.061
Hexachlorophene	4.48	1.7		27	0.65	0.174	
ECBC	5		48		0.7		NA
FAL	5.3		33		0.72		NA
IIVS	3.1		9		0.49		NA
Lactic acid	3073	1.2		12	3.49	0.16	
ECBC	2943		11		3.47		NA

Table 7-3 Interlaboratory Reproducibility of the IC_{50} Values from the 3T3 NRU Test Method

Reference Substance/Laboratory	Arithmetic Mean IC_{50} (mg/mL)[1]	Maximum: Minimum Mean IC_{50} Ratio[2]	Arithmetic IntraLab %CV	Arithmetic InterLab %CV	Log Arithmetic Mean IC_{50} (mg/mL)[1]	ANOVA P[3]	Contrast P[4]
FAL	3487		16		3.54		NA
IIVS	2790		9		3.45		NA
Lindane	161	2.9		58	2.21	0.066	
ECBC	125		95		2.1		NA
FAL	266		36		2.43		NA
IIVS	90.4		122		1.96		NA
Lithium carbonate	NA	NA		NA	NA	NA	NA
ECBC	564		12		2.75		NA
FAL	NA		NA		NA		NA
IIVS	NA		NA		NA		NA
Meprobamate	539	2.5		54	2.73	<0.001	
ECBC	353		14		2.55		NA
FAL	877		15		2.94		NA
IIVS	386		2		2.59		NA
Mercury chloride	4.32	1.7		33	0.64	0.021	
ECBC	3.5		5		0.54		NA
FAL	6		31		0.78		NA
IIVS	3.5		3		0.54		NA
Methanol	NA	NA		NA	NA	NA	NA
ECBC	NA		NA		NA		NA
FAL	NA				NA		NA
IIVS	NA				NA		NA
Nicotine	378	1.7		25	2.58	0.128	
ECBC	272		24		2.43		NA
FAL	412		33		2.61		NA

Table 7-3 Interlaboratory Reproducibility of the IC_{50} Values from the 3T3 NRU Test Method

Reference Substance/Laboratory	Arithmetic Mean IC_{50} (mg/mL)[1]	Maximum: Minimum Mean IC_{50} Ratio[2]	Arithmetic IntraLab %CV	Arithmetic InterLab %CV	Log Arithmetic Mean IC_{50} (mg/mL)[1]	ANOVA P[3]	Contrast P[4]
IIVS	450		12		2.65		NA
Paraquat	23.3	1.2		8	1.37	1	
ECBC	21.3		34		1.33		NA
FAL	24.9		67		1.4		NA
IIVS	23.7		64		1.37		NA
Parathion	61.8	6.4		111	1.79	0.014	
ECBC	22.7		53		1.36		NA
FAL	141		70		2.15		NA
IIVS	22		22		1.34		NA
Phenobarbital	612	1.5		21	2.79	0.232	
ECBC	634		21		2.8		NA
FAL	726		35		2.86		NA
IIVS	476		23		2.68		NA
Phenol	70.9	2.1		41		0.011	
ECBC	50.2		22		1.7		NA
FAL	104		24		2.02		NA
IIVS	58.1		12		1.76		NA
Phenylthiourea	119	7.9		90	2.08	0.007	
ECBC	30.1		66		1.48		0.004
FAL	239		28		2.38		0.006
IIVS	89		25		1.95		0.718
Physostigmine	28.8	1.9		30	1.46	0.149	
ECBC	28.2		53		1.45		NA
FAL	37.8		5		1.58		NA
IIVS	20.4		33		1.31		NA

Table 7-3 Interlaboratory Reproducibility of the IC_{50} Values from the 3T3 NRU Test Method

Reference Substance/Laboratory	Arithmetic Mean IC_{50} (mg/mL)[1]	Maximum: Minimum Mean IC_{50} Ratio[2]	Arithmetic IntraLab %CV	Arithmetic InterLab %CV	Log Arithmetic Mean IC_{50} (mg/mL)[1]	ANOVA P[3]	Contrast P[4]
Potassium chloride	3635	1.1		7	3.56	0.846	
ECBC	3352		14		3.53		NA
FAL	3842		31		3.58		NA
IIVS	3710		11		3.57		NA
Potassium cyanide	64.3	10.4		127	1.81	<0.001	
ECBC	15.3		25		1.18		0.001
FAL	159		52		2.2		<0.001
IIVS	18.9		5		1.28		0.006
Procainamide HCl	443	1.2		11	2.65	0.007	
ECBC	400		4		2.6		0.008
FAL	431		1		2.63		0.396
IIVS	497		8		2.7		0.003
2-Propanol	3563	1.6		23	3.55	0.001	
ECBC	2610		9		3.42		<0.001
FAL	3970		4		3.6		0.004
IIVS	4110		4		3.61		0.002
Propranolol HCl	14.9	1.3		16	1.17	0.488	
ECBC	13.6		32		1.13		NA
FAL	13.5		51		1.13		NA
IIVS	17.6		21		1.25		NA
Propylparaben	29.9	3.0		64	1.48	0.001	
ECBC	20.9		16		1.32		0.045
FAL	51.8		29		1.71		<0.001
IIVS	17.1		12		1.23		0.003
Sodium arsenite	0.873	2.8		55	-0.06	0.028	

Table 7-3 Interlaboratory Reproducibility of the IC_{50} Values from the 3T3 NRU Test Method

Reference Substance/Laboratory	Arithmetic Mean IC_{50} (mg/mL)[1]	Maximum: Minimum Mean IC_{50} Ratio[2]	Arithmetic IntraLab %CV	Arithmetic InterLab %CV	Log Arithmetic Mean IC_{50} (mg/mL)[1]	ANOVA P[3]	Contrast P[4]
ECBC	0.5		6		-0.3		NA
FAL	1.4		57		0.15		NA
IIVS	0.7		17		-0.15		NA
Sodium chloride	4764	1.1		3	3.68	0.759	
ECBC	4790		5		3.68		NA
FAL	4625		13		3.67		NA
IIVS	4877		9		3.69		NA
Sodium dichromate dihydrate	0.602	1.2		9	-0.22	0.822	
ECBC	0.603		14		-0.22		NA
FAL	0.657		37		-0.18		NA
IIVS	0.547		17		-0.26		NA
Sodium fluoride	79.8	1.6		22	1.9	0.016	
ECBC	61.3		9		1.79		NA
FAL	96.1		18		1.98		NA
IIVS	82		7		1.91		NA
Sodium hypochlorite	1211	2.5		57	3.08	0.04	
ECBC	823		13		2.92		NA
FAL	805		46		2.91		NA
IIVS	2005		44		3.3		NA
Sodium oxalate	40.8	1.6		23	1.61	0.643	
ECBC	42		41		1.62		NA
FAL	31		28		1.49		NA
IIVS	49.5		53		1.69		NA
Sodium selenate	34.5	4.3		60	1.54	<0.001	
ECBC	12.7		13		1.1		<0.001

Table 7-3 Interlaboratory Reproducibility of the IC_{50} Values from the 3T3 NRU Test Method

Reference Substance/Laboratory	Arithmetic Mean IC_{50} (mg/mL)[1]	Maximum: Minimum Mean IC_{50} Ratio[2]	Arithmetic IntraLab %CV	Arithmetic InterLab %CV	Log Arithmetic Mean IC_{50} (mg/mL)[1]	ANOVA P[3]	Contrast P[4]
FAL	54.2		19		1.73		<0.001
IIVS	36.5		14		1.56		0.026
Strychnine	199	4.7		83	2.3	<0.001	
ECBC	389		21		2.59		<0.001
FAL	124		16		2.09		0.018
IIVS	83.5		6		1.92		<0.001
Thallium Sulfate	7.5	4.9		72	0.88	0.165	
ECBC	2.8		24		0.45		NA
FAL	13.4		78		1.13		NA
IIVS	6.3		28		0.8		NA
Trichloroacetic acid	928	1.6		27	2.97	0.005	
ECBC	762		13		2.88		0.022
FAL	1220		6		3.09		0.002
IIVS	801		14		2.9		0.069
1,1,1-Trichloroethane	15538	2.2		52	4.19	<0.001	
ECBC	NA		NA		NA		NA
FAL	21250		11		4.33		NA
IIVS	9827		2		3.99		NA
Triethylenemelamine	0.568	16.9		135	-0.25	<0.001	
ECBC	0.086		11		-1.07		<0.001
FAL	1.45		18		0.16		<0.001
IIVS	0.169		29		-0.77		0.002
Triphenyltin hydroxide	0.022	1.7		29	-1.66	0.688	
ECBC	0.026		17		-1.59		NA
FAL	0.026		81		-1.59		NA

Table 7-3 Interlaboratory Reproducibility of the IC_{50} Values from the 3T3 NRU Test Method

Reference Substance/Laboratory	Arithmetic Mean IC_{50} (mg/mL)[1]	Maximum: Minimum Mean IC_{50} Ratio[2]	Arithmetic IntraLab %CV	Arithmetic InterLab %CV	Log Arithmetic Mean IC_{50} (mg/mL)[1]	ANOVA P[3]	Contrast P[4]
IIVS	0.015		55		-1.83		NA
Valproic acid	1177	3.3		76	3.07	<0.001	
ECBC	547		12		2.74		NA
FAL	1807		10		3.26		NA
IIVS	NA		NA		NA		NA
Verapamil HCl	35.2	1.2		10	1.55	0.23	
ECBC	32		18		1.51		NA
FAL	34.6		5		1.54		NA
IIVS	38.9		11		1.59		NA
Xylene	NA	NA		NA	NA	NA	NA
ECBC	NA		NA		NA		NA
FAL	NA		NA		NA		NA
IIVS	724		12		2.86		NA

Abbreviations: 3T3=BALB/c 3T3 fibroblasts; NRU=Neutral red uptake; ECBC=Edgewood Chemical Biological Center; FAL= Fund for the Replacement of Animals in Medical Experiments Alternatives Laboratory; IIVS=Institute for *In Vitro* Sciences; NA=No acceptable IC_{50} results reported or calculation was not performed (e.g., for contrast results); CV=Coefficient of variation.

[1]Results reported on the same row with chemical names are the means of all the laboratories. Results reported on the same row as laboratories are the laboratory means.

[2]Maximum laboratory mean IC_{50} divided by minimum laboratory mean IC_{50}.

[3]$p < 0.01$ indicated statistical significance.

[4]Contrasts were performed if ANOVA was significant ($p < 0.01$) to determine which laboratory was different from the other two laboratories. Significant contrasts were denoted by $p < 0.01$. No contrast tests were performed if only two laboratories reported IC_{50} values.

Table 7-4 Reference Substances with Significant ANOVA Differences Among Laboratories for the 3T3 NRU Test Method

Reference Substance	Significant Contrast Results[1]			Insoluble/ Volatile[2]
	ECBC	FAL	IIVS	
Acetylsalicylic acid		H	L	
Arsenic trioxide		L	H	Precipitate
Busulfan		H		
Chloral hydrate	L	H		
Chloramphenicol	L	H		
Citric acid	L	H		
Colchicine	L	H		
Cupric sulfate pentahydrate	M	H	L	
Dibutyl phthalate		H	L	Precipitate
Dichlorvos	L	H		Precipitate
Ethylene glycol	L			
Glutethimide		H	L	
Haloperidol		H		
Meprobamate	L	H	M	
Phenylthiourea	L	H		
Potassium cyanide	L	H	M	Precipitate /Volatile
Procainamide HCl	L		H	
2-Propanol	L	M	H	Volatile
Propylparaben		H	L	
Sodium selenate	L	H		
Strychnine	H		L	Precipitate
Trichloroacetic acid		H		
Triethylenemelamine	L	H		

Abbreviations: ANOVA=Analysis of variance; 3T3=BALB/c 3T3 fibroblasts; NRU=Neutral red uptake; H=Laboratory reported the highest mean IC_{50}; L=Laboratory reported the lowest mean IC_{50}; M=Laboratory reported a mean IC_{50} between the values of the other two laboratories; ECBC=Edgewood Chemical Biological Center; FAL= Fund for the Replacement of Animals in Medical Experiments Alternatives Laboratory; IIVS=Institute for *In Vitro* Sciences.

[1]Laboratories significantly different from the other two at $p < 0.01$.

[2]From **Table 5-11**. Precipitate reported by at least one laboratory is indicated by "Precipitate". Use of plate sealers by at least one laboratory to prevent volatile contamination of control wells indicated by "Volatility".

7.2.2.2 *Differences Among the IC_{50} Values in Laboratories Using the NHK NRU Test Method*
The ANOVA results in **Table 7-5** indicate that there were statistically significant ($p < 0.01$) laboratory differences for six of the 68 (9%) reference substances evaluated. These substances are listed in **Table 7-6** along with columns showing which laboratory's IC_{50} values were statistically significantly different from the other two (as indicated by the contrast results), and indications of insolubility or volatility during conduct of the assay. Insolubility was reported for three of the six substances, but none of the six substances were volatile.

For the six substances that yielded significantly different IC_{50} values among the laboratories, ECBC reported the highest IC_{50} value for four substances and the lowest for one, FAL reported the lowest values for three substances and the highest for two, and IIVS reported the highest IC_{50} value for one substance and the lowest for two.

7.2.3 CV Results for the 3T3 and NHK NRU Test Methods
CV values were calculated as described in Section **5.5.2.2**. **Tables 7-3** and **7-5** provide the intra- and inter-laboratory CV values for the individual reference substances. **Table 7-7** summarizes the CV values for each method and shows that median and mean values were often similar. Median CV values were frequently lower than the corresponding means, which indicated that large individual CV values skewed the CV distributions.

7.2.3.1 *Reproducibility of Intralaboratory CV Values*
Table 7-7 shows that the intralaboratory CV values and mean intralaboratory CV values were the same, 26%, for both NRU test methods. The median intralaboratory CV values were also similar: 23% and 24% for the 3T3 and the NHK NRU test method, respectively. Of the three laboratories, FAL had the highest mean and median CV values and IIVS had the lowest for both methods.

Table 7-5 Reproducibility of the IC_{50} Values from the NHK NRU Test Method

Reference Substance/Laboratory	Arithmetic Mean IC_{50} (mg/mL)[1]	Maximum: Minimum Mean IC_{50} Ratio[2]	Arithmetic IntraLab %CV	Arithmetic InterLab %CV	Log Arithmetic Mean IC_{50} (mg/mL)[1]	ANOVA P[3]	Contrast P[4]
Acetaminophen	526	1.3		13	2.72	0.181	
ECBC	558		15		2.75		NA
FAL	447		19		2.65		NA
IIVS	571		14		2.76		NA
Acetonitrile	10104	1.2		8	4	0.964	
ECBC	10868		72		4.04		NA
FAL	10153		19		4.01		NA
IIVS	9290		4		3.97		NA
Acetylsalicylic acid	613	1.4		15	2.79	0.060	
ECBC	631		3		2.8		NA
FAL	694		14		2.84		NA
IIVS	514		15		2.71		NA
5-Aminosalicylic acid	52.3	2.6		47	1.72	0.044	
ECBC	29.9		22		1.48		NA
FAL	78.2		54		1.89		NA
IIVS	48.8		16		1.69		NA
Aminopterin	682	1.6		27	2.83	0.025	
ECBC	889		20		2.95		NA
FAL	545		8		2.74		NA
IIVS	611		12		2.79		NA
Amitriptyline HCl	9.76	1.4		19	0.99	0.365	
ECBC	10.8		31		1.03		NA
FAL	7.57		72		0.88		NA
IIVS	10.9		10		1.04		NA
Arsenic trioxide	10.4	8.2		91	1.02	<0.001	

Table 7-5 Reproducibility of the IC_{50} Values from the NHK NRU Test Method

Reference Substance/Laboratory	Arithmetic Mean IC_{50} (mg/mL)[1]	Maximum: Minimum Mean IC_{50} Ratio[2]	Arithmetic IntraLab %CV	Arithmetic InterLab %CV	Log Arithmetic Mean IC_{50} (mg/mL)[1]	ANOVA P[3]	Contrast P[4]
ECBC	7.77		33		0.89		0.694
FAL	2.55		75		0.41		<0.001
IIVS	20.9		31		1.32		0.0006
Atropine sulfate	91.9	1.3		13	1.96	0.988	
ECBC	85.4		12		1.93		0.8903
FAL	104		85		2.02		0.9069
IIVS	83.2		25		1.92		0.9832
Boric acid	473	1.2		8	2.67	0.931	
ECBC	440		31		2.64		0.9692
FAL	517		73		2.71		0.7391
IIVS	464		2		2.67		0.768
Busulfan	278	1.2		11	2.44	0.659	
ECBC	253		27		2.4		NA
FAL	268		72		2.43		NA
IIVS	313		12		2.5		NA
Cadmium chloride	1.98	1.2		10	0.3	0.733	
ECBC	2.2		37		0.34		NA
FAL	1.88		65		0.27		NA
IIVS	1.86		8		0.27		NA
Caffeine	661	1.4		21	2.82	0.296	
ECBC	817		31		2.91		NA
FAL	591		32		2.77		NA
IIVS	574		1		2.76		NA
Carbamazepine	128	4.0		85	2.11	0.432	
ECBC	66.1		13		1.82		NA

Table 7-5 Reproducibility of the IC_{50} Values from the NHK NRU Test Method

Reference Substance/Laboratory	Arithmetic Mean IC_{50} (mg/mL)[1]	Maximum: Minimum Mean IC_{50} Ratio[2]	Arithmetic IntraLab %CV	Arithmetic InterLab %CV	Log Arithmetic Mean IC_{50} (mg/mL)[1]	ANOVA P[3]	Contrast P[4]
FAL	253		129		2.4		NA
IIVS	63.9		8		1.81		NA
Carbon tetrachloride	NA	NA		NA	NA	NA	
ECBC	NA		NA		NA		NA
FAL	NA		NA		NA		NA
IIVS	NA		NA		NA		NA
Chloral hydrate	137	1.4		17	2.14	0.302	
ECBC	140		24		2.15		NA
FAL	159		32		2.2		NA
IIVS	112		2		2.05		NA
Chloramphenicol	366	1.3		13	2.56	0.750	
ECBC	318		45		2.5		NA
FAL	414		44		2.62		NA
IIVS	367		22		2.56		NA
Citric acid	424	1.7		25	2.63	0.006	
ECBC	526		16		2.72		0.009
FAL	312		17		2.49		0.002
IIVS	433		5		2.64		0.483
Colchicine	0.007	1.6		22	-2.16	0.174	
ECBC	0.005		46		-2.28		NA
FAL	0.008		10		-2.12		NA
IIVS	0.008		21		-2.09		NA
Cupric sulfate pentahydrate	197	1.1		4	2.29	0.374	
ECBC	190		10		2.28		NA
FAL	195		6		2.29		NA

Table 7-5 Reproducibility of the IC_{50} Values from the NHK NRU Test Method

Reference Substance/Laboratory	Arithmetic Mean IC_{50} (mg/mL)[1]	Maximum: Minimum Mean IC_{50} Ratio[2]	Arithmetic IntraLab %CV	Arithmetic InterLab %CV	Log Arithmetic Mean IC_{50} (mg/mL)[1]	ANOVA P[3]	Contrast P[4]
IIVS	207		3		2.32		NA
Cycloheximide	0.082	2.3		43	-1.09	0.302	
ECBC	0.053		22		-1.28		NA
FAL	0.12		78		-0.92		NA
IIVS	0.071		19		-1.15		NA
Dibutyl phthalate	32.6	2.2		41	1.51	0.408	
ECBC	28.3		27		1.45		NA
FAL	47.4		73		1.68		NA
IIVS	22		6		1.34		NA
Dichlorvos	11.1	1.4		20	1.05	0.181	
ECBC	8.56		27		0.93		NA
FAL	12.4		30		1.09		NA
IIVS	12.2		3		1.09		NA
Diethyl phthalate	145	2.6		44	2.16	0.049	
ECBC	174		8		2.24		NA
FAL	71.5		94		1.85		NA
IIVS	189		18		2.28		NA
Digoxin	0.00314	107.6		88	-2.5	<0.001	
ECBC	0.00538		13		-2.27		<0.001
FAL	0.00005		36		-4.29		<0.001
IIVS	0.00398		7		-2.4		<0.001
Dimethylformamide	7856	1.5		19	3.9	<0.001	
ECBC	9353		2		3.97		<0.001
FAL	7817		1		3.89		0.508
IIVS	6397		3		3.81		<0.001

Table 7-5 Reproducibility of the IC_{50} Values from the NHK NRU Test Method

Reference Substance/Laboratory	Arithmetic Mean IC_{50} (mg/mL)[1]	Maximum: Minimum Mean IC_{50} Ratio[2]	Arithmetic IntraLab %CV	Arithmetic InterLab %CV	Log Arithmetic Mean IC_{50} (mg/mL)[1]	ANOVA P[3]	Contrast P[4]
Diquat dibromide monohydrate	4.73	1.9		37	0.67	0.217	
ECBC	3.59		23		0.56		NA
FAL	6.77		55		0.83		NA
IIVS	3.84		8		0.58		NA
Disulfoton	378	5.8		99	2.58	<0.001	
ECBC	140		19		2.15		0.002
FAL	808		26		2.91		<0.001
IIVS	186		32		2.27		0.018
Endosulfan	2.35	2.4		43	0.37	0.029	
ECBC	3.44		17		0.54		NA
FAL	1.42		50		0.15		NA
IIVS	2.19		20		0.34		NA
Epinephrine bitartrate	90.6	1.5		24	1.96	0.119	
ECBC	115		9		2.06		NA
FAL	81.7		35		1.91		NA
IIVS	75		16		1.88		NA
Ethanol	10184	1.4		18	4.01	0.035	
ECBC	8290		5		3.92		NA
FAL	12013		19		4.08		NA
IIVS	10250		9		4.01		NA
Ethylene glycol	42600	1.3		15	4.63	0.063	
ECBC	38000		12		4.58		NA
FAL	49800		9		4.7		NA
IIVS	40000		13		4.6		NA
Fenpropathrin	2.6	2.0		39	0.41	0.031	

Table 7-5 Reproducibility of the IC_{50} Values from the NHK NRU Test Method

Reference Substance/Laboratory	Arithmetic Mean IC_{50} (mg/mL)[1]	Maximum: Minimum Mean IC_{50} Ratio[2]	Arithmetic IntraLab %CV	Arithmetic InterLab %CV	Log Arithmetic Mean IC_{50} (mg/mL)[1]	ANOVA P[3]	Contrast P[4]
ECBC	3.73		27		0.57		NA
FAL	2.23		28		0.35		NA
IIVS	1.82		17		0.26		NA
Gibberellic Acid	2866	1.0		2	3.46	0.862	
ECBC	2850		14		3.45		NA
FAL	2940		9		3.47		NA
IIVS	2807		4		3.45		NA
Glutethimide	177	1.1		5	2.25	0.968	
ECBC	187		34		2.27		NA
FAL	170		14		2.23		NA
IIVS	176		16		2.24		NA
Glycerol	27108	1.9		31	4.43	0.200	
ECBC	34267		45		4.53		NA
FAL	18023		46		4.26		NA
IIVS	29033		16		4.46		NA
Haloperidol	3.57	1.1		7	0.55	0.935	
ECBC	3.69		27		0.57		NA
FAL	3.72		49		0.57		NA
IIVS	3.29		35		0.52		NA
Hexachlorophene	0.031	2.2		41	-1.5	0.097	
ECBC	0.027		16		-1.57		NA
FAL	0.046		44		-1.34		NA
IIVS	0.021		11		-1.67		NA
Lactic acid	1308	1.0		1	3.12	0.904	
ECBC	1290		4		3.11		NA

Table 7-5 Reproducibility of the IC_{50} Values from the NHK NRU Test Method

Reference Substance/Laboratory	Arithmetic Mean IC_{50} (mg/mL)[1]	Maximum: Minimum Mean IC_{50} Ratio[2]	Arithmetic IntraLab %CV	Arithmetic InterLab %CV	Log Arithmetic Mean IC_{50} (mg/mL)[1]	ANOVA P[3]	Contrast P[4]
FAL	1320		5		3.12		NA
IIVS	1313		11		3.12		NA
Lindane	19.3	1.5		20	1.29	0.203	
ECBC	19.1		17		1.28		NA
FAL	23.2		31		1.37		NA
IIVS	15.6		15		1.19		NA
Lithium carbonate	477	1.3		13	2.68	0.295	
ECBC	411		29		2.61		NA
FAL	486		20		2.69		NA
IIVS	535		6		2.73		NA
Meprobamate	516	4.7		61	2.71	0.027	
ECBC	761		15		2.88		NA
FAL	163		116		2.21		NA
IIVS	624		14		2.8		NA
Mercury chloride	5.87	1.3		15	0.77	0.120	
ECBC	6.87		15		0.84		NA
FAL	5.4		19		0.73		NA
IIVS	5.35		2		0.73		NA
Methanol	1616	1.9		42	3.21	0.007	
ECBC	NA		NA		NA		NA
FAL	1133		19		3.05		NA
IIVS	2100		11		3.32		NA
Nicotine	113	1.4		17	2.05	0.700	
ECBC	94.3		26		1.97		NA
FAL	134		59		2.13		NA

Table 7-5 Reproducibility of the IC_{50} Values from the NHK NRU Test Method

Reference Substance/Laboratory	Arithmetic Mean IC_{50} (mg/mL)[1]	Maximum: Minimum Mean IC_{50} Ratio[2]	Arithmetic IntraLab %CV	Arithmetic InterLab %CV	Log Arithmetic Mean IC_{50} (mg/mL)[1]	ANOVA P[3]	Contrast P[4]
IIVS	112		25		2.05		NA
Paraquat	66.1	2.0		40	1.82	0.047	
ECBC	48.3		13		1.68		NA
FAL	96.6		39		1.98		NA
IIVS	53.4		10		1.73		NA
Parathion	31.4	1.2		8	1.5	0.845	
ECBC	34		30		1.53		NA
FAL	31.2		38		1.49		NA
IIVS	29		29		1.46		NA
Phenobarbital	478	1.9		39	2.68	0.027	
ECBC	693		26		2.84		NA
FAL	360		27		2.56		NA
IIVS	381		18		2.58		NA
Phenol	77.7	1.6		22	1.89	0.094	
ECBC	59.1		36		1.77		NA
FAL	93.2		6		1.97		NA
IIVS	80.8		6		1.91		NA
Phenylthiourea	346	1.5		19	2.54	0.133	
ECBC	363		16		2.56		NA
FAL	401		21		2.6		NA
IIVS	272		26		2.44		NA
Physostigmine	172	1.5		22	2.24	0.623	
ECBC	164		3		2.21		NA
FAL	213		112		2.33		NA
IIVS	139		6		2.14		NA

Table 7-5 Reproducibility of the IC_{50} Values from the NHK NRU Test Method

Reference Substance/Laboratory	Arithmetic Mean IC_{50} (mg/mL)[1]	Maximum: Minimum Mean IC_{50} Ratio[2]	Arithmetic IntraLab %CV	Arithmetic InterLab %CV	Log Arithmetic Mean IC_{50} (mg/mL)[1]	ANOVA P[3]	Contrast P[4]
Potassium chloride	2279	1.3		13	3.36	0.396	
ECBC	2560		17		3.41		NA
FAL	2287		28		3.36		NA
IIVS	1990		8		3.3		NA
Potassium cyanide	45.1	5.3		86	1.65	0.340	
ECBC	29.3		24		1.47		NA
FAL	89		112		1.95		NA
IIVS	16.9		13		1.23		NA
Procainamide HCl	1764	1.4		16	3.25	0.053	
ECBC	1480		14		3.17		NA
FAL	1787		12		3.25		NA
IIVS	2027		11		3.31		NA
2-Propanol	5541	1.7		26	3.74	0.033	
ECBC	5263		11		3.72		NA
FAL	4273		27		3.63		NA
IIVS	7087		7		3.85		NA
Propranolol HCl	36.9	1.5		21	1.57	0.003	
ECBC	38.27		12		1.58		0.325
FAL	43.8		6		1.64		0.006
IIVS	28.6		11		1.46		0.001
Propylparaben	16.8	1.3		16	1.23	0.066	
ECBC	18.1		13		1.26		NA
FAL	18.6		15		1.27		NA
IIVS	13.8		9		1.14		NA
Sodium arsenite	0.532	2.4		44	-0.27	0.061	

Table 7-5 Reproducibility of the IC_{50} Values from the NHK NRU Test Method

Reference Substance/Laboratory	Arithmetic Mean IC_{50} (mg/mL)[1]	Maximum: Minimum Mean IC_{50} Ratio[2]	Arithmetic IntraLab %CV	Arithmetic InterLab %CV	Log Arithmetic Mean IC_{50} (mg/mL)[1]	ANOVA P[3]	Contrast P[4]
ECBC	0.79		32		-0.1		NA
FAL	0.336		56		-0.47		NA
IIVS	0.47		14		-0.33		NA
Sodium chloride	2724	3.2		51	3.44	0.045	
ECBC	3583		7		3.55		NA
FAL	1118		124		3.05		NA
IIVS	3470		9		3.54		NA
Sodium dichromate dihydrate	0.737	1.5		19	-0.13	0.258	
ECBC	0.784		14		-0.11		NA
FAL	0.851		36		-0.07		NA
IIVS	0.576		17		-0.24		NA
Sodium fluoride	47.4	1.4		15	1.68	0.313	
ECBC	48.7		14		1.69		NA
FAL	39.7		24		1.6		NA
IIVS	53.7		13		1.73		NA
Sodium hypochlorite	1580	1.5		20	3.2	0.313	
ECBC	1863		31		3.27		NA
FAL	1243		46		3.09		NA
IIVS	1633		11		3.21		NA
Sodium oxalate	355	1.0		1	2.55	0.926	
ECBC	355		15		2.55		NA
FAL	350		42		2.54		NA
IIVS	360		26		2.56		NA
Sodium selenate	11.2	2.2		40	1.05	0.134	
ECBC	7.47		12		0.87		NA

Table 7-5 Reproducibility of the IC_{50} Values from the NHK NRU Test Method

Reference Substance/Laboratory	Arithmetic Mean IC_{50} (mg/mL)[1]	Maximum: Minimum Mean IC_{50} Ratio[2]	Arithmetic IntraLab %CV	Arithmetic InterLab %CV	Log Arithmetic Mean IC_{50} (mg/mL)[1]	ANOVA P[3]	Contrast P[4]
FAL	16.1		59		1.21		NA
IIVS	10		13		1		NA
Strychnine	69.3	1.9		39	1.84	0.364	
ECBC	100		76		2		NA
FAL	52.5		53		1.72		NA
IIVS	55.1		6		1.74		NA
Thallium Sulfate	0.16	1.6		23	-0.8	0.405	
ECBC	0.198		51		-0.7		NA
FAL	0.153		20		-0.82		NA
IIVS	0.127		16		-0.9		NA
Trichloroacetic acid	427	1.6		24	2.63	0.134	
ECBC	348		18		2.54		NA
FAL	541		28		2.73		NA
IIVS	394		13		2.6		NA
1,1,1-Trichloroethane	NA	NA		NA	NA	NA	
ECBC	8137		7		3.91		NA
FAL	NA		NA		NA		NA
IIVS	NA		NA		NA		NA
Triethylenemelamine	1.95	1.3		12	0.29	0.562	
ECBC	1.69		57		0.23		NA
FAL	2.03		23		0.31		NA
IIVS	2.13		23		0.33		NA
Triphenyltin hydroxide	0.013	3.0		55	-1.89	0.088	
ECBC	0.021		32		-1.68		NA
FAL	0.007		106		-2.15		NA

Table 7-5 Reproducibility of the IC_{50} Values from the NHK NRU Test Method

Reference Substance/Laboratory	Arithmetic Mean IC_{50} (mg/mL)[1]	Maximum: Minimum Mean IC_{50} Ratio[2]	Arithmetic IntraLab %CV	Arithmetic InterLab %CV	Log Arithmetic Mean IC_{50} (mg/mL)[1]	ANOVA P[3]	Contrast P[4]
IIVS	0.011		32		-1.96		NA
Valproic acid	533	1.6		28	2.73	0.081	
ECBC	468		25		2.67		0.331
FAL	702		23		2.85		0.032
IIVS	430		17		2.63		0.135
Verapamil HCl	68.7	1.3		14	1.84	0.624	
ECBC	60.5		22		1.78		NA
FAL	79.4		42		1.9		NA
IIVS	66.2		8		1.82		NA
Xylene	NA	NA		NA	NA	NA	
ECBC	NA		NA		NA		NA
FAL	NA		NA		NA		NA
IIVS	486		38		2.69		NA

Abbreviations: NHK=Normal human epidermal keratinocytes; NRU=Neutral red uptake; ECBC=Edgewood Chemical Biological Center; FAL= Fund for the Replacement of Animals in Medical Experiments Alternatives Laboratory; IIVS=Institute for *In Vitro* Sciences; NA=No acceptable IC_{50} results reported or calculation was not performed (e.g., for contrast results); CV=Coefficient of variation.

[1]Results reported on the same row with chemical names are the means of all the laboratories. Results reported on the same row as laboratories are the laboratory means.

[2]Maximum laboratory mean IC_{50} divided by minimum laboratory mean IC_{50}.

[3]$p < 0.01$ indicated statistical significance.

[4]Contrasts were performed if ANOVA was significant ($p < 0.01$) to determine which laboratory was different from the other two laboratories. Significant contrasts were denoted by $p < 0.01$. No contrast tests were performed if only two laboratories reported IC_{50} values.

Table 7-6 Reference Substances with Significant ANOVA Differences Among Laboratories for the NHK NRU Test Method

Reference Substance	Significant Contrast Results[1]			Solubility/ Volatility[2]
	ECBC	FAL	IIVS	
Arsenic trioxide		L	H	Precipitate
Citric acid	H	L		Precipitate
Digoxin	H	L		
Dimethylformamide	H		L	
Disulfoton	L	H		Precipitate
Propranolol HCl		H	L	

Abbreviations: ANOVA=Analysis of variance; NHK=Normal human epidermal keratinocytes; NRU=Neutral red uptake; H=Laboratory reported the highest mean IC_{50}; L=Laboratory reported the lowest mean IC_{50}; ECBC=Edgewood Chemical Biological Center; FAL= Fund for the Replacement of Animals in Medical Experiments Alternatives Laboratory; IIVS=Institute for *In Vitro* Sciences.
[1]Laboratories significantly different from the other two at $p < 0.01$
[2]From **Table 5-11**. Precipitate reported by at least one laboratory.

7.2.3.2 *Reproducibility of Interlaboratory CV Values*

The mean and median interlaboratory CV for the reference substances were lower in the NHK NRU test method (mean=28%; median=21% vs. mean=47%; median=37% for 3T3 (see **Table 7-7**).

Table 7-7 Summary of CV Results for the 3T3 and NHK NRU Test Methods

CV	3T3 NRU Test Method				NHK NRU Test Method			
	N	Mean	Median	Range	N	Mean	Median	Range
Intralaboratory CV	198	26%	23%	1-122%	204	26%	24%	1-129%
ECBC	64	23%	17%	2-95%	68	23%	20%	2-76%
FAL	64	33%	31%	1-98%	68	43%	34%	1-129%
IIVS	64	21%	14%	1-122%	68	13%	13%	1-35%
Interlaboratory CV	64	47%	37%	3-135%	68	28%	21%	1-91%

Abbreviations: CV=Coefficient of variation; 3T3=BALB/c 3T3 fibroblasts; NHK=Normal human epidermal keratinocytes; NRU=Neutral red uptake; N=number of values; ECBC=Edgewood Chemical Biological Center; FAL= Fund for the Replacement of Animals in Medical Experiments Alternatives Laboratory; IIVS=Institute for *In Vitro* Sciences.
Note: For the 3T3 method, the following substances were excluded because all laboratories did not obtain sufficient IC_{50} data: carbon tetrachloride; disulfoton; gibberellic acid; lithium carbonate; methanol; 1,1,1-trichloroethane; valproic acid; and xylene. For the NHK method, the following substances were excluded because all laboratories did not obtain sufficient IC_{50} data: carbon tetrachloride; methanol; 1,1,1-trichloroethane; and xylene.

7.2.3.3 *Variation of CV with Chemical Property*

To identify chemical characteristics that may be associated with high or low CV values, their associations were assessed for chemical class along with the following chemical attributes: physical state (i.e., solid or liquid), solubility, volatility, molecular weight, log K_{ow}, IC_{50}, and boiling point. The CVs were also examined with respect to their association with the GHS acute oral toxicity class (UN 2005). For categorical characteristics such as physical form, solubility (i.e., precipitate/no precipitate), volatile/not volatile, and chemical class, the mean

CV values and ranges for the groups were compared to one another and to the overall mean CV and CV range for each method. No statistical analyses were performed for these comparisons. Spearman correlation analyses were performed for chemical characteristics measured by continuous variables, such as molecular weight, log K_{ow}, and IC_{50}, and boiling point.

7.2.3.4 *Results of Intralaboratory CV Analysis*

The intralaboratory CV analysis (see **Table 7-8**) uses one mean intralaboratory CV for each reference substance that was calculated from the intralaboratory CV values from each laboratory. There seemed to be little difference in CV values among the categorical physical/chemical/toxicological attributes. The mean intralaboratory CV values for solids and liquids were similar (26 vs. 23% for 3T3; 27 vs. 24% for NHK). The mean intralaboratory CV values for reference substances for which precipitates were observed were similar to values for substances with no precipitates were observed (32 vs. 23% for 3T3; 24 vs. 27% for NHK). The mean intralaboratory CV values for substances that exhibited volatility were similar to those that did not (31 vs. 25% for 3T3; 27 vs. 26% for NHK). Similarly, the substances grouped by GHS acute oral toxicity category (UN 2005) had mean intralaboratory CV values that were similar (20-33% for 3T3; 19-31% for NHK) to the overall mean CV values (26% for both test methods). However, the mean intralaboratory CV values for both NRU test methods tended to increase with decreasing LD_{50}.

Mean intralaboratory CV values were calculated for the chemical classes that contained at least three of the reference substances included in the reproducibility analyses (i.e., 64 substances for 3T3 and 68 substances for NHK). Reference substances in the amide chemical class had unusually low mean intralaboratory CV values for both the 3T3 (13%) and the NHK (10%) NRU test method compared with the overall mean CV (26% for both test methods), but there were only three substances in this chemical class (acetaminophen, dimethylformamide, procainamide HCl). Organic sulfur compounds had a high mean intralaboratory CV for the 3T3 test method (46%), but not for the NHK NRU test method (29%) compared with the overall mean intralaboratory CV for both test methods (26%). The intralaboratory CV values for the remaining chemical classes were unremarkable compared with the overall mean intralaboratory CV values.

For the characteristics amenable to correlation analysis, none of the Spearman correlation coefficients were large (absolute value of $r_s < 0.6$), but several were statistically significantly different from zero ($p < 0.05$). Molecular weight ($p=0.016$), IC_{50} ($p=0.002$), and boiling point ($p=0.009$) exhibited statistically significant correlations to intralaboratory CV for the 3T3 test NRU method. The higher molecular weight substances had higher intralaboratory CV values and the substances with lower IC_{50} values had higher intralaboratory CV values. The finding that substances with higher boiling points had higher CV values was consistent with the categorical analysis of volatility. The substances that exhibited volatile characteristics (i.e., cross contamination of VC wells) in the 3T3 NRU test method had slightly higher mean intralaboratory CV values (31%) than the substances that did not exhibit volatile characteristics (25%).

Table 7-8 Intralaboratory CV Values by Chemical Characteristics for the 3T3 and NHK NRU Test Methods

Class/Attribute	3T3 NRU Test Method			NHK NRU Test Method		
	N[1]	Range	Mean	N[1]	Range	Mean
All chemicals	64	1-122%	26%	68	1-129%	26%
Chemical form						
Solid	51	4-84	26	53	6-57	27
Liquid	13	6-48	23	15	2-40	24
Solubility						
Precipitate[2]	18	11-84	32	19	2-47	24
No precipitate	46	4-55	23	49	7-57	27
Volatility[3]						
Volatile	10	6-84	31	9	11-50	27
Nonvolatile	54	4-55	25	59[2]	2-57	26
Chemical Class						
Alcohol	9	6-42	22	9	10-37	22
Amide	3	4-28	13	3	2-16	10
Amine	3	9-35	18	3	10-24	18
Carboxylic acid	13	4-41	18	14	2-48	23
Heterocyclic	14	6-59	31	14	13-50	32
Organophosphorous	2	NA	NA	3	20-32	26
Organic sulfur	4	36-59	46	5	21-27	29
Phenol	5	14-30	20	5	11-31	19
Polycyclic	4	19-35	27	5	9-38	20
Inorganic	14	9-43	25	15	6-50	29
Inorganic chlorine	5	9-33	19	5	12-50	32
Inorganic sodium	6	9-34	20	6	17-47	30
GHS Acute Oral Toxicity Class						
$LD_{50} \leq 5$ mg/kg	6	9-46	27	7	20-40	30
$5 < LD_{50} \leq 50$	12	13-59	32	12	12-50	31
$50 < LD_{50} \leq 300$	12	11-84	33	12	17-37	25
$300 < LD_{50} \leq 2000$	14	4-51	22	16	6-57	25
$2000 < LD_{50} \leq 5000$	9	9-32	20	9	7-50	30
$LD_{50} > 5000$	11	6-42	20	12	2-40	19
Correlations	**N**	**r_s**	**P value**	**N**	**r_s**	**P value**
Molecular weight	64	0.301	0.016	68	0.181	0.140
Log K_{ow}	45[4]	0.121	0.430	48[4]	0.310	0.032
IC_{50}	64	-0.382	0.002	68	-0.346	0.004
Boiling point	24[5]	0.520	0.009	24[5]	0.226	0.289

Abbreviations: CV=Coefficient of variation; 3T3=BALB/c 3T3 fibroblasts; GHS=Globally Harmonized System of Classification and Labelling of Chemicals (UN 2005); NA=Not applicable because class had less than three observations; NHK=Normal human epidermal keratinocytes; NRU=Neutral red uptake; N=Number of values; r_s=Spearman correlation coefficient; K_{ow}=Octanol:water partition coefficient.

[1]One intralaboratory CV for each chemical was calculated by averaging the CV values for each reference substance.

[2]Identified by laboratory reports of precipitate in the stock reference substance solutions or in cell culture (see **Table 5-11**).

[3]Identified by laboratory reports of using plate sealers to avoid contamination of the VC wells (see **Table 5-11**).

[4]Number of reference substances with CV values and log K_{ow} data.

[5]Number of reference substances with CV values and boiling point data.

Among the IC_{50} values obtained using the NHK NRU test method, two of the characteristics amenable to correlation analysis were statistically significantly different from zero, although the correlation coefficients did not have large magnitudes (absolute value of r_s <0.4). The log

K_{ow} (p=0.032) and IC_{50} (p=0.004) exhibited statistically significant correlations (p <0.05) to the intralaboratory CV. Log K_{ow} was positively correlated (i.e., higher log K_{ow} values were associated with higher mean intralaboratory CV), but the IC_{50} was negatively correlated (i.e., higher log IC_{50} values were associated with lower mean intralaboratory CV) to mean intralaboratory CV.

7.2.3.5 *Results of the Interlaboratory CV Analysis*

Table 7-9 shows the analysis of the interlaboratory CV values. There seemed to be little difference in interlaboratory CV values for most of the categorical physical/chemical characteristics. The mean interlaboratory CV values for solids and liquids were similar (48% for solids vs. 42% for liquids for 3T3, and 28% for solids vs. 21% for liquids for NHK), as were the values for substances for which precipitates were observed versus no precipitates (58% vs. 43% for 3T3, and 24% vs. 28% for NHK), and the values for substances that exhibited volatile characteristics (51% for volatile substances vs. 46% for nonvolatile substances for 3T3, and 32% for volatile substances vs. 26% for nonvolatile substances for NHK).

Mean interlaboratory CV values were calculated for the chemical classes that contained at least three of the reference substances included in the reproducibility analyses (i.e., 64 substances for 3T3 and 68 substances for NHK). Reference substances in the amide chemical class had low mean interlaboratory CV values for both the 3T3 (15%) and the NHK (16%) NRU test methods compared with the overall mean interlaboratory CV (47% and 28%, respectively). Substances in the amine class also had low mean interlaboratory CV values for the 3T3 NRU (13%), but not for the NHK NRU (20%). Organic sulfur compounds had unusually high mean interlaboratory CV values for the 3T3 test method (100%), but not for the NHK NRU (36%) compared with the overall mean interlaboratory CV (47% and 28%, respectively). Because of the low number of reference substances in these classes, these results were deemed to not be significant.

Mean interlaboratory CV values tended to be large for chemicals in the most toxic GHS acute oral toxicity categories, especially with the 3T3 NRU test method. The mean interlaboratory CV for reference substances in the $LD_{50} \leq 5$ mg/kg (72%) and $5 < LD_{50} \leq 50$ mg/kg (78%) classes were larger than the mean overall interlaboratory CV (47%,). For the NHK NRU test method, the mean interlaboratory CV for chemicals in the $5 < LD_{50} \leq 5$ mg/kg (37%) and $5 < LD_{50} \leq 50$ mg/kg (41%) classes were larger than the mean overall interlaboratory CV (28%).

For the characteristics amenable to correlation analysis, none of the correlation coefficients were large (absolute value of r_s <0.6), but IC_{50} (p=0.015) and boiling point (p=0.007) exhibited statistically significant correlations (p <0.05) to interlaboratory CV in the 3T3 test NRU method. There was a negative correlation between interlaboratory CV and IC_{50}, but the correlation between boiling point and interlaboratory CV was positive. The positive correlation of CV with boiling point was largely consistent with the categorical analysis of volatility. The substances that exhibited volatile characteristics in the 3T3 NRU test method had slightly higher mean CV values than substances that did not exhibit volatile characteristics (51% vs. 46%). Only the IC_{50} was significantly correlated (p=0.014) to

interlaboratory CV with a negative correlation (r_s=-0.271) when the NHK NRU test method was used.

Table 7-9 Interlaboratory 3T3 and NHK NRU Test Method CV Values Sorted by Chemical Characteristics

Class/Attribute	3T3 NRU Test Method			NHK NRU Test Method		
	N	Range	Mean	N	Range	Mean
All chemicals	64[1]	3-135%	47%	68[1]	1-91%	28%
Chemical Form						
Solids	51	3-135	48	53	1-91	28
Liquids	13	6-124	42	15	1-44	21
Solubility						
Precipitate[2]	18	7-127	58	19	1-91	24
No precipitate	46	3-135	43	49	1-88	28
Volatility						
Volatile[3]	10	21-127	51	9	8-86	32
Nonvolatile	54	3-135	46	59	1-91	26
Chemical Class						
Alcohol	9	12-119	38	9	11-31	20
Amide	3	6-28	15	3	13-19	16
Amine	3	10-16	13	3	14-24	20
Carboxylic acid	13	6-124	38	14	1-61	26
Heterocyclic	14	8-135	57	14	5-85	32
Organic sulfur	4	78-119	100	5	8-99	36
Organophosphorous	2	NA	NA	3	8-99	42
Phenol	5	19-64	41	5	15-47	28
Polycyclic	4	14-85	44	5	2-88	30
Inorganic	14	3-127	50	15	4-91	30
Inorganic chlorine	5	3-127	45	5	10-86	35
Inorganic sodium	6	3-60	34	6	15-51	32
GHS Acute Oral Toxicity Class						
$LD_{50} \leq 5$ mg/kg	6	12-135	72	7	12-99	37
$5 < LD_{50} \leq 50$	12	33-127	78	12	8-91	41
$50 < LD_{50} \leq 300$	12	8-120	37	12	10-41	26
$300 < LD_{50} \leq 2000$	14	11-85	35	16	1-61	20
$2000 < LD_{50} \leq 5000$	9	3-69	29	9	1-85	27
$LD_{50} > 5000$	11	6-124	41	12	2-44	23
Correlations	**N**	**r_s**	**P value**	**N**	**r_s**	**P value**
Molecular weight	64	0.245	0.051	68	0.169	0.168
Log K_{ow}	45[4]	0.151	0.324	48[4]	0.210	0.151
IC_{50}	64	-0.304	0.015	68	-0.297	0.014
Boiling point	22[5]	0.563	0.007	25[5]	-0.051	0.809

Abbreviations: CV=Coefficient of variation; 3T3=BALB/c 3T3 fibroblasts; GHS=Globally Harmonized System of Classification and Labelling of Chemicals (UN 2005); NA=Not applicable because class had less than three observations; NHK=Normal human epidermal keratinocytes; NRU=Neutral red uptake; N=Number of values; r_s=Spearman correlation coefficient; K_{ow}=Octanol:water partition coefficient.

[1]One intralaboratory CV for each chemical was calculated by averaging the CV values for each reference substance.

[2]Identified by laboratory reports of precipitate in the stock reference substance solutions or in cell culture (see **Table 5-11**).

[3]Identified by laboratory reports of using plate sealers to avoid contamination of the VC wells (see **Table 5-11**).

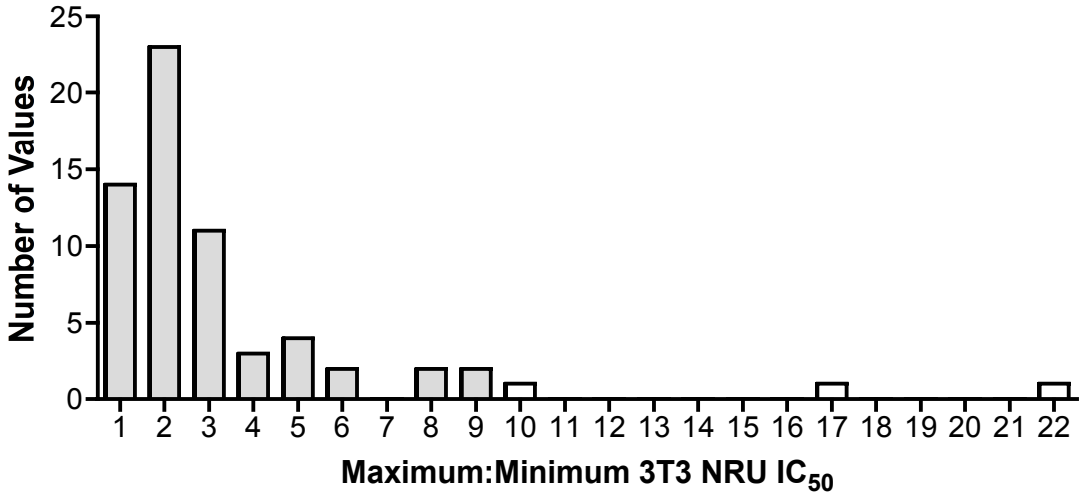
[4]Number of reference substances with CV values and log K_{ow} data.

[5]Number of reference substances with CV values and boiling point data.

7.2.4 Comparison of Maximum to Minimum IC_{50} Values Using Laboratory Means

Interlaboratory reproducibility was also compared by calculating maximum to minimum mean IC_{50} values using the laboratory means from each method, so that the reproducibility of the IC_{50} values could be compared with the reproducibility of the reference LD_{50} values derived in **Section 4.2**. The **Figure 7-2** frequency histogram for the 3T3 NRU test method maximum:minimum mean IC_{50} values shows that approximately half (37) of the 64 reference substances produced ratios less than 2.5-fold of each other, and only nine chemicals had ratios greater than 5.5-fold, including one substance (cupric sulfate pentahydrate) that had a ratio of 22.

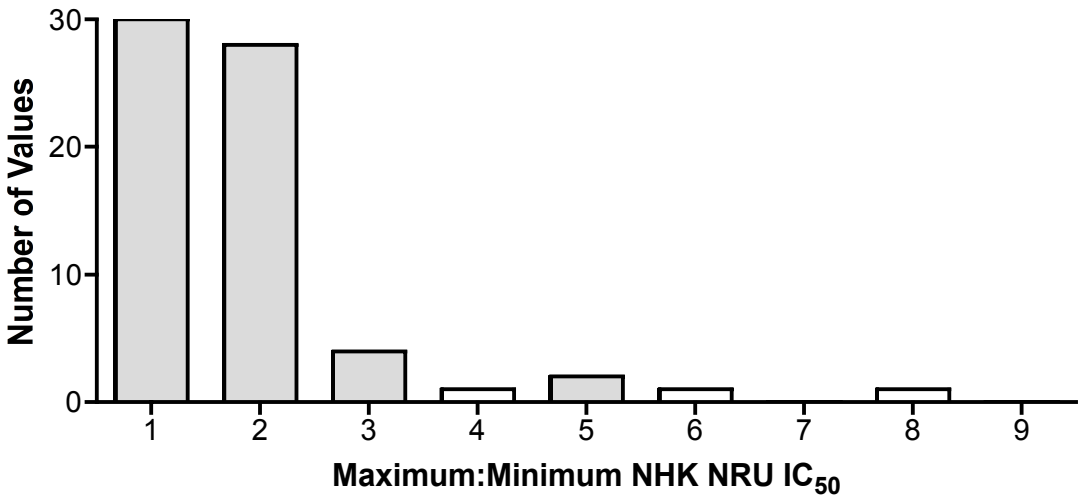
Figure 7-2 Frequency of Maximum:Minimum 3T3 NRU IC_{50} Ratios



Abbreviations: 3T3=BALB/c 3T3 fibroblasts; NRU=Neutral red uptake.
Bars show the number of substances with maximum:minimum 3T3 NRU IC_{50} ratios within ±0.5 units of the bar label (e.g., the first bar indicates that there were 14 reference substances for which the laboratory mean maximum:minimum 3T3 NRU IC_{50} ratios were 0.5 to1.4). The analysis includes 64 reference substances. Carbon tetrachloride, disulfoton, gibberellic acid, lithium carbonate, methanol, 1,1,1-trichloroethane, valproic acid, and xylene were excluded because not all laboratories obtained IC_{50} values.

The **Figure 7-3** frequency histogram for the maximum:minimum mean IC_{50} values for the NHK NRU test method shows that ratios of 58 of the 68 chemicals were less than 2.5-fold of one another. The highest ratio of 108 for digoxin is not shown in the figure. Comparison of **Figures 7-2** and **7-3** shows that the interlaboratory reproducibility of the NHK NRU test method was better than that for the 3T3 NRU test method based on the distribution of the low maximum:minimum IC_{50} ratios.

Figure 7-3 Frequency of Maximum:Minimum NHK NRU IC_{50} Ratios



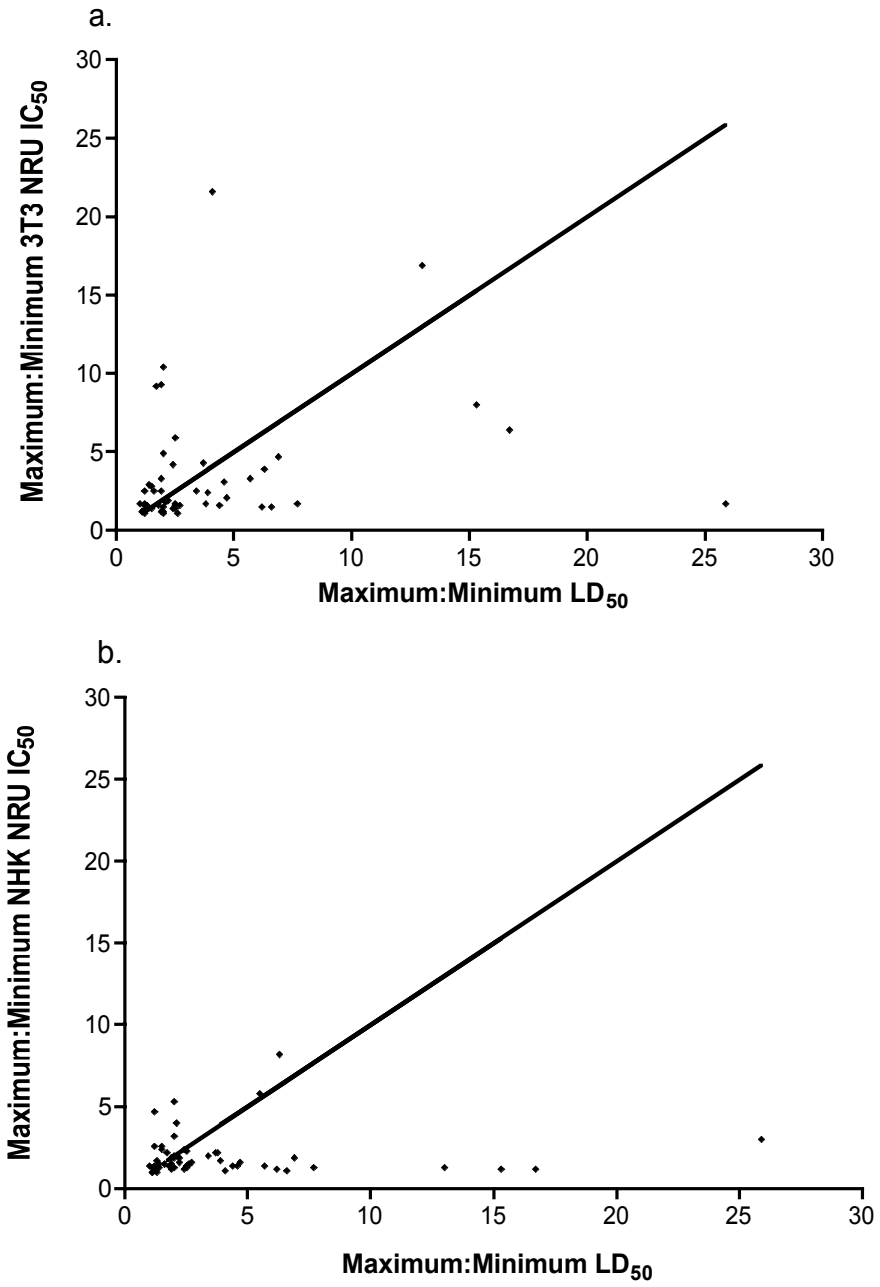
Abbreviations: NHK=Normal human epidermal keratinocytes; NRU=Neutral red uptake.
Bars show the number of substances with maximum:minimum NHK NRU IC_{50} ratios within ±0.5 units of the bar label (e.g., the first bar indicates that there were 30 reference substances for which the laboratory mean maximim:minimum NHK NRU IC_{50} ratios were 0.5 to 1.4). The analysis includes 68 reference substances. Carbon tetrachloride, methanol, 1,1,1-trichloroethane, and xylene were excluded because not all laboratories obtained IC_{50} values. The maximum:minimum IC_{50} for digoxin of 108 was excluded from this figure.

7.2.5 Comparison of the Maximum:Minimum IC_{50} Ratios with the Maximum:Minimum LD_{50} Ratios

To compare the reproducibility of the NRU IC_{50} values with that of the LD_{50} values, the maximum:minimum IC_{50} ratios for each method (shown in **Tables 7-3** and **7-5**) were compared with the maximum:minimum LD_{50} ratios reported in **Table 4-2**. This analysis excluded reference substances for which fewer than three laboratories reported IC_{50} values, and reference substances for which fewer than two acceptable acute oral LD_{50} values were identified. As a result, there were 53 substances analysed for the 3T3 NRU test method and 57 for the NHK NRU test method. The following substances were excluded from both analyses because fewer than two acceptable LD_{50} values could be identified: aminopterin; colchicine; digoxin; epinephrine bitartrate; gluthethimide; phenylthiourea; physostigmine; procainamide HCl, propranolol HCl; propylparaben; and thallium sulfate. Carbon tetrachloride, disulfoton, gibberellic acid, lithium carbonate, methanol, 1,1,1-trichloroethane, valproic acid, and xylene, were excluded from the 3T3 analysis, and carbon tetrachloride, methanol, 1,1,1-trichloroethane, and xylene, were excluded from the NHK analysis, because fewer than three laboratories reported IC_{50} values.

Figure 7-4 shows that the maximum:minimum LD_{50} ratios tend to be larger than either the 3T3 NRU IC_{50} or NHK NRU IC_{50} ratios because there are more points below the theoretical one-to-one correspondence line than above the line. The difference between the LD_{50} maximum:minimum values and the NRU IC_{50} maximum:minimum values is more striking for the NHK since there are fewer points above the line for the NHK graph (**Figure 7-4b**) than for the 3T3 graph (**Figure 7-4a**).

Figure 7-4 Comparison of Maximum:Minimum NRU IC_{50} Ratios to Maximum:Minimum LD_{50} Ratios



Abbreviations: 3T3=BALB/c 3T3 fibroblasts; NRU=Neutral red uptake; NHK=Normal human epidermal keratinocytes. Comparison of maximum:minimum ratios of IC_{50} and LD_{50} for 53 reference substances for the 3T3 NRU test method (a) and 57 reference substances for the NHK NRU test method (b). Solid lines show the theoretical one to one correspondence of maximum:minimum IC_{50} to maximum:minimum LD_{50}.

7.2.6 Normalization of Reference Substance IC_{50} Values Using SLS IC_{50} Values

As an alternate analysis for reproducibility, IC_{50} values for reference substances were normalized to those of the corresponding SLS IC_{50}values. This approach was tested using five reference substances for each test method to determine whether such normalization would reduce the variability, measured using CV values, of the results. The reference substances selected for this evaluation were those for which the ANOVA indicated statistically significant differences among the laboratories. Because there were a number of reference substances that met this criterion for the 3T3 NRU test method, the substances were selected so as to cover a wide range of rodent acute oral toxicity. One reference substance was selected from each GHS category with the exception of the $50 \leq LD_{50} < 300$ mg/kg category. There was no substance represented by this category because there were six acute oral toxicity categories and only five substances were used for this assessment. The reference substances, shown in **Table 7-10**, were busulfan, chloramphenicol, meprobamate, propylparaben, and triethylenemelamine. Because there were only six reference substances with significant ANOVAs in the NHK NRU test method, the last five reference substances in **Table 7-5** (citric acid, digoxin, dimethylformamide, disulfoton, and propranolol HCl) were selected for this analysis.

Millimolar units were used for the IC_{50} values in this analysis since the mole is the most appropriate unit for measuring and comparing biological activity. The IC_{50} value (in mM) for each reference substance was normalized to the corresponding SLS IC_{50} value (in mM) by dividing the SLS IC_{50} by the reference substance IC_{50}. Intra- and inter-laboratory CV values were calculated for both the IC_{50} values and for the SLS IC_{50}:reference substance IC_{50} ratios to determine whether this type of normalization would reduce the interlaboratory CV values.

Table 7-10 shows that the mean intralaboratory CV of the IC_{50} values for the five substances used in the 3T3 evaluation was 22% and the interlaboratory CV was 88%. Normalizing the reference substance IC_{50} values to the SLS IC_{50} yielded a slightly higher intralaboratory CV of 25% and a lower interlaboratory CV of 65%. The mean intralaboratory CV of the IC_{50} values for the five substances used in the NHK evaluation was 14% and the interlaboratory CV was 50%. Normalizing the reference substance IC_{50} values to the SLS IC_{50} yielded a slightly higher intralaboratory CV of 16% and a higher interlaboratory CV of 61%. When the normalization ratios are examined for each chemical-by-laboratory combination (**Table 7-10**), nine CVs increased, five decreased, and one remained the same for the 3T3 NRU test method, and eight increased, six decreased, and one remained the same for the NHK NRU test method. Thus, for the reference substances used in this analysis, normalizing the reference substance IC_{50} to the concurrent SLS IC_{50} did not reduce the overall variability of the measurements, as measured by the CV values.

Table 7-10 CV Values for 3T3 and NHK NRU Test Method IC_{50} Values and Normalized IC_{50} Values

Reference Substance	IC_{50} (mM)[1]	IntraLab CV[2] (%)	InterLab CV[3] (%)	SLS IC_{50}: Substance IC_{50}[4]	IntraLab CV SLS IC_{50}: Substance IC_{50}[5] (%)	InterLab CV SLS IC_{50}: Substance IC_{50}[6] (%)
3T3 NRU Test Method						
Busulfan	0.548		119	0.677		74
ECBC	0.163	48		1.05	70	
FAL	1.30	56		0.109	53	
IIVS	0.177	4		0.877	9	
Chloramphenicol	0.498		67	0.725		29
ECBC	0.171	22		0.847	30	
FAL	0.845	30		0.844	22	
IIVS	0.483	18		0.483	21	
Meprobamate	2.47		54	0.071		39
ECBC	1.62	14		0.085	23	
FAL	4.02	15		0.039	29	
IIVS	1.77	2		0.088	3	
Propylparaben	0.166		64	1.16		49
ECBC	0.116	16		1.29	20	
FAL	0.287	29		0.535	22	
IIVS	0.0949	12		1.65	9	
Triethylene-melamine	0.00278		135	191		87
ECBC	0.000421	11		354	11	
FAL	0.00710	18		21.4	24	
IIVS	0.000827	29		197	23	
Mean		**22**	**88**		**25**	**65**
NHK NRU Test Method						
Citric Acid	2.21		25	0.00587		26
ECBC	2.74	16		0.0053	14	
FAL	1.62	17		0.0076	28	
IIVS	2.25	5		0.0047	16	
Digoxin	4.02E-06		88	62378		168
ECBC	6.89E-06	13		1264	10	
FAL	6.53E-08	36		183479	44	
IIVS	5.10E-06	7		2389	26	
Dimethylform-amide	107		19	0.00011		31
ECBC	128	2		0.00007	7	
FAL	107	1		0.00013	1	
IIVS	87.5	3		0.00013	19	
Disulfoton	1.38		99	0.0140		61
ECBC	0.509	19		0.022	6	
FAL	2.94	26		0.005	5	
IIVS	0.679	32		0.015	20	
Propranolol HCl	0.125		21	0.0947		20
ECBC	0.129	12		0.081	15	

Table 7-10 CV Values for 3T3 and NHK NRU Test Method IC_{50} Values and Normalized IC_{50} Values

Reference Substance	IC_{50} (mM)[1]	IntraLab CV[2] (%)	InterLab CV[3] (%)	SLS IC_{50}: Substance IC_{50}[4]	IntraLab CV SLS IC_{50}: Substance IC_{50}[5] (%)	InterLab CV SLS IC_{50}: Substance IC_{50}[6] (%)
FAL	0.148	6		0.087	25	
IIVS	0.0967	11		0.116	9	
Mean		**14**	**50**		**16**	**61**

Abbreviations: 3T3=BALB/c 3T3 fibroblasts; ECBC=Edgewood Chemical Biological Center; FAL= Fund for the Replacement of Animals in Medical Experiments Alternatives Laboratory; IIVS=Institute for *In Vitro* Sciences; NA=No acceptable IC_{50} results reported or calculation was not performed (e.g., for contrast results); CV=Coefficient of variation.
[1]Results reported on the same row with reference substance names are the arithmetic means of all the laboratories. Results reported on the same row as laboratories are the arithmetic laboratory means.
[2]CV for IC_{50} values from the acceptable tests within each laboratory.
[3]CV calculated using the arithmetic mean IC_{50} values from each laboratory.
[4]Concurrent SLS IC_{50} in mM divided by the reference substance IC_{50}. Results reported on the same row with reference substance names are the arithmetic means of all the laboratories. Results reported on the same row as laboratories are the arithmetic laboratory means.
[5]CV for SLS IC_{50}:reference substance IC_{50} values within each laboratory.
[3]CV calculated using the mean SLS IC_{50}:reference substance IC_{50} values from each laboratory.

7.3 Historical Positive Control (PC) Data

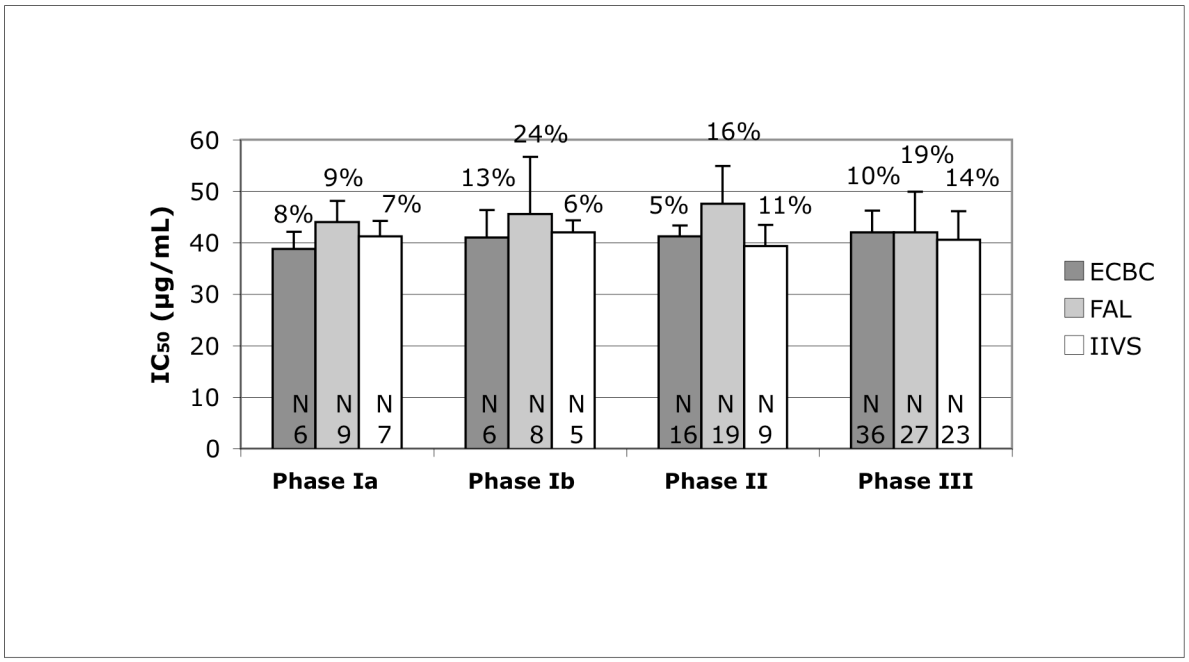
The reproducibility of the PC (SLS) data was assessed by CV analysis, ANOVA, and linear regression over time, as described in **Section 5.5.4.2**. To obtain an assessment of the true variation of SLS IC_{50} values, the reproducibility analyses also included IC_{50} values from SLS tests that failed the test acceptance criterion for the IC_{50} acceptance limits determined for each study phase. Therefore, the values used for this analysis included some that were not included in **Table 5-3**. These additional SLS tests, however, passed all other test acceptance criteria. If more than one SLS test was performed in a single day (for each method and laboratory), the IC_{50} values were averaged to determine a single IC_{50} for the day so that the multiple results from that day would not overly influence the average.

Figure 7-5 shows the average SLS IC_{50} values for each method, laboratory, and study phase. The SLS IC_{50} for the 3T3 test method (**Figure 7-5a**) was relatively consistent over the entire period of the study (approximately 2.5 years). The intralaboratory CV values for the individual study phases ranged from 5% to 24% **(Figure 7-5a)**. With the exception of the Phase Ib CV at FAL, the CV values for each laboratory and phase were less than 20%. The interlaboratory CV values were even smaller, 6% in Phases Ia and Ib, 10% in Phase II, and 2% in Phase III.

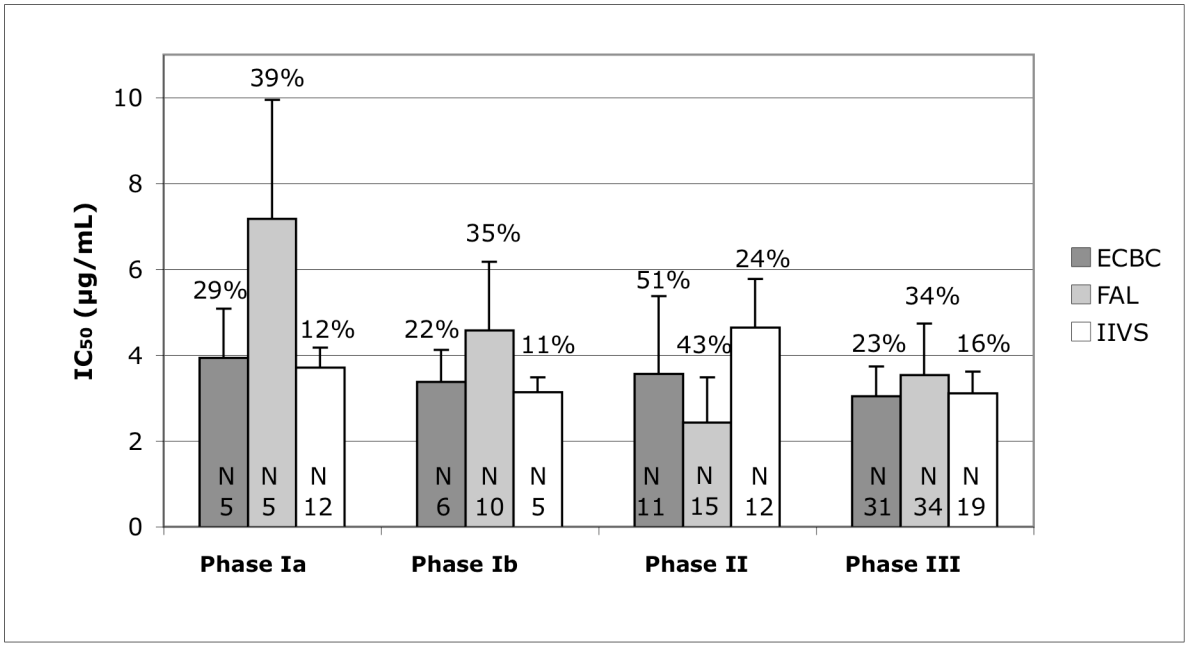
Figure 7-5b shows that the SLS IC_{50} for the NHK NRU test method tended to vary with time, but, with the exception of the values from FAL, there appeared to be no consistent trend. The IC_{50} values from FAL, which changed their cell culture methods after Phase Ib (see **Section 5.3.3.1**), tended to decrease over time. Although the change in cell culture methods reduced the magnitude of the IC_{50}, the variability (as evidenced by the intralaboratory CV values shown in **Figure 7-5b**) remained relatively high (CV $\geq$34% for all FAL study phases).

Figure 7-5 SLS IC_{50} for Each Laboratory and Study Phase

a 3T3 NRU Test Method



b NHK NRU Test Method



Abbreviations: 3T3=BALB/c 3T3 fibroblasts; NHK=Normal human epidermal keratinocytes; NRU=Neutral red uptake; ECBC=Edgewood Chemical Biological Center; FAL= Fund for the Replacement of Animals in Medical Experiments Alternatives Laboratory; IIVS=Institute for *In Vitro* Sciences; N=Number of values.
Note: Bars show mean SLS IC_{50} values. Error bars show standard deviation. Percent values above error bars are intralaboratory CVs.

The CV values for all the laboratories and study phases show that the SLS IC_{50} values in the NHK NRU test method are more variable within laboratories than the corresponding 3T3 SLS IC_{50} values. The CV values for the SLS IC_{50} for the NHK NRU test method ranged from 11 to 51%, with nine of the 12 values greater than 20%. The interlaboratory CV values, which were also greater than those for the 3T3 NRU test method, were 39% in Phase Ia, 21% in Phase Ib, 31% in Phase II, and 8% in Phase III.

7.3.1 ANOVA and Linear Regression Results for the 3T3 NRU Test Method

7.3.1.1 *Variation of SLS IC_{50} Values with Time*

Table 7-11 shows the SLS ANOVA results from the 3T3 test method. When the IC_{50} values in each laboratory were compared, there were no statistically significant differences ($p < 0.01$) among study phases for any laboratory. **Table 7-12** shows that the slopes of the linear regressions of the IC_{50} values over time (expressed as index values) were significantly different from zero for ECBC and FAL ($p=0.001$ and 0.012, respectively), but, because the slopes were so small (0.000204 and -0.000324), and in different directions, these differences were considered to be unimportant, regardless of the statistical conclusions. The slope of the IIVS regression of SLS IC_{50} over time was not significantly different from zero ($p=0.651$; **Table 7-12**), which was consistent with the ANOVA analysis (**Table 7-11**), and showed that SLS IC_{50} from IIVS did not vary with study phase ($p=0.854$). The ANOVA analysis, with study phase as the factor (with laboratories combined), showed that the 3T3 NRU IC_{50} values from all the laboratories were consistent over time ($p=0.304$).

7.3.1.2 *Comparison of SLS IC_{50} Values Among the Laboratories*

When all study phases from each laboratory were combined, ANOVA, with laboratory as the factor, showed that the SLS IC_{50} values in the 3T3 NRU test method differed significantly among the laboratories ($p < 0.006$) **(Table 7-11)**. However, as can be seen in **Figure 7-5a**, the individual laboratory SDs overlap one another.

Table 7-11 ANOVA Results for the SLS IC_{50} Values in the 3T3 NRU Test Method

Study Phase/ Laboratory	ECBC				FAL				IIVS			
	Log Mean IC_{50} (mM)	SD	N	P[1]	Log Mean IC_{50} (mM)	SD	N	P[1]	Log Mean IC_{50} (mM)	SD	N	P[1]
Test for differences between phases within each laboratory												
Phase Ia	-0.876	0.042	6	0.031	-0.811	0.046	9	0.015	-0.850	0.034	7	0.854
Phase Ib	-0.864	0.066	6		-0.846	0.065	8		-0.838	0.025	5	
Phase II	-0.848	0.027	16		-0.796	0.057	19		-0.854	0.025	8	
Phase III	-0.842	0.036	36		-0.851	0.066	27		-0.844	0.041	23	
Test for differences between laboratories (phases combined)												
All Phases	-0.849	0.039	64	0.006	-0.826	0.062	63		-0.847	0.035	44	
Test for differences between phases (laboratories combined)												
Phase Ia	-0.839	0.049	22	0.304								
Phase Ib	-0.850	0.056	19									
Phase II	-0.831	0.047	34									
Phase III	0.845	0.045	86									

Abbreviations: ANOVA=Analysis of variance; SLS=Sodium lauryl sulfate; 3T3=BALB/c 3T3 fibroblasts; NRU=Neutral red uptake; N=Number of values; SD=Standard deviation; ECBC=Edgewood Chemical Biological Center; FAL=Fund for the Replacement of Animals in Medical Experiments Alternatives Laboratory; IIVS=Institute for *In Vitro* Sciences.
[1]Statistically significant at $p < 0.01$.

Table 7-12 Linear Regression Analysis of SLS IC_{50} Values Over Time[1]

Laboratory	Slope	P-value (Slope)[2]	Intercept
3T3 NRU Test Method			
ECBC	0.000204	0.001	-0.874
FAL	-0.000324	0.012	-0.796
IIVS	0.0000304	0.651	-0.850
NHK NRU Test Method			
ECBC	-0.000559	0.002	-1.901
FAL	-0.00112	<0.001	-1.737
IIVS	-0.000445	0.002	-1.885

Abbreviations: SLS=Sodium lauryl sulfate; 3T3=BALB/c 3T3 fibroblasts; NHK=Normal human epidermal keratinocytes; NRU=Neutral red uptake; N=Number of values; ECBC=Edgewood Chemical Biological Center; FAL=Fund for the Replacement of Animals in Medical Experiments Alternatives Laboratory; IIVS=Institute for *In Vitro* Sciences.
[1]Time was expressed as index values. The index value of each test reflected the order of testing without respect to the time lapsing between tests.
[2]Statistically significant from zero at $p < 0.05$.

7.3.2 ANOVA and Linear Regression Results for the NHK NRU Test Method

7.3.2.1 *Variation of SLS IC_{50} Values with Time*

Table 7-13 shows the ANOVA results for the NHK NRU test method. When the IC_{50} values within each laboratory were compared by study phase, the values were statistically different ($p < 0.01$) at each laboratory. The IC_{50} values from the various study phases were also significantly different from one another when the laboratory data were combined ($p < 0.001$). The change in cell culture methods at FAL after Phase Ib (see **Section 5.3.3.1**) contributed to this difference. **Table 7-13** shows that FAL had clearly the lowest log mean SLS IC_{50} for Phases Ia and Ib. Linear regression analyses showed that the IC_{50} slopes over time (expressed as an index values) were statistically significantly less than zero for each laboratory (see **Table 7-12**). Because the slopes were so small (-0.000559, -0.00112, and -0.000445), and negative, their statistical significance was considered to be irrelevant.

7.3.2.2 *Comparison of SLS IC_{50} Values Among the Laboratories*

The ANOVA results, with laboratory as a factor (**Table 7-13),** showed that the SLS IC_{50} was statistically significantly different among the laboratories when the data from the study phases were pooled ($p < 0.001$). **Figure 7-5b** shows that the SLS data from ECBC and IIVS were rather similar to one another for Phases Ia, Ib, and III. The SLS IC_{50} data from FAL are different from the other two laboratories for Phases Ia, Ib, and II, but the SDs for Phase III show that the data from all laboratories produced similar values.

Table 7-13 ANOVA Results for the SLS IC_{50} Values in the NHK NRU Test Method

Study Phase/ Laboratory	ECBC				FAL				IIVS			
	Log Mean IC_{50} (mM)	SD	N	P[1]	Log Mean IC_{50} (mM)	SD	N	P[1]	Log Mean IC_{50} (mM)	SD	N	P[1]
Test for differences between phases within each laboratory												
Phase Ia	-1.867	0.135	5	0.001	-1.656	0.125	5	<0.001	-1.904	0.060	12	<0.001
Phase Ib	-1.936	0.092	6		-1.829	0.141	10		-1.965	0.046	5	
Phase II	-2.007	0.109	11		-1.982	0.173	15		-1.863	0.058	12	
Phase III	-1.990	0.098	31		-1.941	0.113	34		-1.972	0.070	19	
Test for differences between laboratories (phases combined)												
All Phases	-1.971	0.113	53	<0.001	-1.879	0.175	64		-1.924	0.073	48	
Test for differences between phases (laboratories combined)												
Phase Ia	-1.833	0.143	22	<0.001								
Phase Ib	-1.891	0.125	21									
Phase II	-1.964	0.139	38									
Phase III	-1.971	0.100	84									

Abbreviations: ANOVA=Analysis of variance; SLS=Sodium lauryl sulfate; NHK=Normal human epidermal keratinocytes; NRU=Neutral red uptake; N=Number of values; SD=Standard deviation; ECBC=Edgewood Chemical Biological Center; FAL=Fund for the Replacement of Animals in Medical Experiments Alternatives Laboratory; IIVS=Institute for *In Vitro* Sciences.

[1]Statistically significant at $p < 0.01$.

7.4 Laboratory Concordance for Solvent Selection

The solvents used for the reference substances are shown in **Table 7-14**. For Phases Ib and II, the SMT based their selection of solvents on the results provided by BioReliance (see **Table 5-9**) using the solubility protocol in **Appendix G2**. Despite the fact that the solubility of an individual substance might be different in 3T3 and NHK growth media, the SMT selected the same solvent (i.e., medium or DMSO) for both test methods, rather than having different solvents for each method.

BioReliance occasionally achieved higher solubility values for the Phase I and II substances than the three cytotoxicity laboratories (e.g., see the results for arsenic trioxide, aminopterin, and chloramphenicol in **Table 5-10**). The laboratories were using the solubility protocols in **Appendices C3** through **C6** (for Phases Ib and II), which were somewhat different from the protocol used by BioReliance. Although all the laboratories used the same protocols, they did not always obtain similar results with respect to the solvent to be used (e.g., see the results for aminopterin, cadmium chloride, and chloramphenicol in **Table 5-10**). In an attempt to avoid the selection of a solvent for which one or more laboratories could not achieve the desired solubility, the SMT used the solubility data from all the laboratories to determine the solvents to be used for each chemical tested in Phase III. **Table 7-14** shows that cell culture medium was used as the solvent for 38 substances and DMSO was used for 34 substances.

Five of the substances were insoluble in medium and DMSO in at least one testing laboratory. Arsenic trioxide was insoluble at all laboratories. IIVS also found sodium oxalate, strychnine, and triethylenemelamine insoluble in media and DMSO, and FAL found thallium sulfate insoluble in both solvents. Therefore, the SMT used the results from the laboratories that did achieve solubility to select the solvents to be used for testing these substances.

The testing laboratories selected the same solvent for 55 of the 72 reference substances (76%). Excluding the five substances that were found to be insoluble in both solvents by at least one laboratory, there were 12 substances on which the laboratories disagreed: acetaminophen, acetylsalicylic acid, carbamazepine, carbon tetrachloride, chloramphenicol, dichlorvos, meprobamate, methanol, phenobarbital, phenylthiourea, physostigmine, and valproic acid. Each laboratory reported relatively low solubility, $\leq$2 mg/mL, in medium for these substances. Because 2 mg/mL in medium is the departure point for the selection of medium or DMSO, small variations in solubility lead the laboratories to select different solvents. The solubility of acetaminophen, for example was reported as 2 mg/mL in culture media by ECBC and FAL, but <2 mg/mL by IIVS. IIVS found it soluble in 200 mg/mL DMSO and selected DMSO as the solvent. ECBC and FAL selected culture media as the solvent. The SMT selected DMSO as the solvent for acetaminophen to be used by all laboratories so that they would all be assured of obtaining usable test results.

Table 7-14 Solvent Determinations by Laboratory

Reference Substance	Solvent Used for Testing[1]	ECBC	FAL	IIVS
Acetaminophen	DMSO	Medium	Medium	DMSO
Acetonitrile	Medium	Medium	Medium	Medium
Acetylsalicylic acid	DMSO	Medium	DMSO	Medium
Aminopterin	DMSO	DMSO	DMSO	DMSO
5-Aminosalicylic acid	Medium	Medium	Medium	Medium
Amitriptyline HCl	DMSO	DMSO	DMSO	DMSO
Arsenic III trioxide	Medium	ID	ID	ID
Atropine sulfate	Medium	Medium	Medium	Medium
Boric aid	Medium	Medium	Medium	Medium
Busulfan	DMSO	DMSO	DMSO	DMSO
Cadmium II chloride	DMSO	DMSO	DMSO	DMSO
Caffeine	Medium	Medium	Medium	Medium
Carbamazepine	DMSO	Medium	DMSO	DMSO
Carbon tetrachloride	DMSO	Medium	DMSO	Medium
Chloral hydrate	Medium	Medium	Medium	Medium
Chloramphenicol	DMSO	DMSO	DMSO	Medium
Citric acid	Medium	Medium	Medium	Medium
Colchicine	Medium	Medium	Medium	Medium
Cupric sulfate pentahydrate	Medium	Medium	Medium	Medium
Cycloheximide	Medium	Medium	Medium	Medium
Dibutyl phthalate	DMSO	DMSO	DMSO	DMSO
Dichlorvos	DMSO	Medium	DMSO	Medium
Diethyl phthalate	DMSO	DMSO	DMSO	DMSO
Digoxin	DMSO	DMSO	DMSO	DMSO
Dimethylformamide	Medium	Medium	Medium	Medium
Diquat dibromide monohydrate	Medium	Medium	Medium	Medium
Disulfoton	DMSO	DMSO	DMSO	DMSO
Endosulfan	DMSO	DMSO	DMSO	DMSO
Epinephrine bitartrate	Medium	Medium	Medium	Medium
Ethanol	Medium	Medium	Medium	Medium
Ethylene glycol	Medium	Medium	Medium	Medium
Fenpropathrin	DMSO	DMSO	DMSO	DMSO
Gibberellic acid	Medium	Medium	Medium	Medium
Glutethimide	DMSO	DMSO	DMSO	DMSO
Glycerol	Medium	Medium	Medium	Medium
Haloperidol	DMSO	DMSO	DMSO	DMSO
Hexachlorophene	DMSO	DMSO	DMSO	DMSO
Lactic acid	Medium	Medium	Medium	Medium
Lindane	DMSO	DMSO	DMSO	DMSO
Lithium I carbonate	Medium	Medium	Medium	Medium
Meprobamate	DMSO	Medium	Medium	DMSO
Mercury II chloride	DMSO	DMSO	DMSO	DMSO
Methanol	DMSO	Medium	Medium	DMSO
Nicotine	Medium	Medium	Medium	Medium
Paraquat	Medium	Medium	Medium	Medium
Parathion	DMSO	DMSO	DMSO	DMSO
Phenobarbital	DMSO	Medium	DMSO	DMSO
Phenol	Medium	Medium	Medium	Medium
Phenylthiourea	DMSO	DMSO	Medium	DMSO

Table 7-14 Solvent Determinations by Laboratory

Reference Substance	Solvent Used for Testing[1]	ECBC	FAL	IIVS
Physostigmine	DMSO	Medium	DMSO	DMSO
Potassium I chloride	Medium	Medium	Medium	Medium
Potassium cyanide	Medium	Medium	Medium	Medium
Procainamide HCl	Medium	Medium	Medium	Medium
2-Propanol	Medium	Medium	Medium	Medium
Propranolol HCl	DMSO	Medium	Medium	Medium
Propylparaben	DMSO	DMSO	DMSO	DMSO
Sodium arsenite	Medium	Medium	Medium	Medium
Sodium chloride	Medium	Medium	Medium	Medium
Sodium dichromate dihydrate	Medium	Medium	Medium	Medium
Sodium fluoride	Medium	Medium	Medium	Medium
Sodium hypochlorite	Medium	Medium	Medium	Medium
Sodium oxalate	Medium	Medium	Medium	ID
Sodium selenate	Medium	Medium	Medium	Medium
Strychnine	Medium	Medium	Medium	ID
Thallium I sulfate	Medium	Medium	ID	Medium
Trichloroacetic acid	Medium	Medium	Medium	Medium
1,1,1-Trichloroethane	Medium	Medium	Medium	Medium
Triethylenemelamine	DMSO	Medium	DMSO	ID
Triphenyltin hydroxide	DMSO	DMSO	DMSO	DMSO
Valproic acid	DMSO	Medium	DMSO	DMSO
Verapamil HCl	DMSO	DMSO	DMSO	DMSO
Xylene	DMSO	DMSO	DMSO	DMSO
DMSO Total	34	22	29	28
Medium Total	38	49	41	40

Abbreviations: DMSO=Dimethyl sulfoxide; ECBC=Edgewood Chemical Biological Center; FAL= Fund for the Replacement of Animals in Medical Experiments Alternatives Laboratory; IIVS=Institute for *In Vitro* Sciences; ID=Insufficient data to select solvent; Medium=Cell culture medium.
[1]Solvents selected by the SMT for use by all laboratories.

7.5 Summary

Intra- and inter-laboratory reproducibility were assessed by comparing the laboratory-specific IC_{50}-LD_{50} regressions to the mean, across-laboratory regression for each method, ANOVA, CV analysis, and comparison of maximum:minimum mean laboratory IC_{50} values. ANOVA permitted statistical comparisons of laboratories and experimental averages, while controlling for other factors. CV analysis compared the relative magnitudes of variability on a standardized scale. Reproducibility was evaluated using the results from the reference substances that yielded IC_{50} values from all three laboratories: 64 and 68 reference substances in the 3T3 and the NHK NRU test methods, respectively. The analysis of intralaboratory reproducibility, by evaluating the similarity of the laboratory specific IC_{50}-LD_{50} regressions, showed that the laboratory regressions for both NRU test methods were within the 95% confidence limits of the laboratory mean regressions.

The ANOVA showed significant interlaboratory differences for 23 substances in the 3T3 NRU test method and six in the NHK NRU test method. Intralaboratory CV values ranged from 1-122% in the 3T3 test method and 1-129% in the NHK NRU test method. Mean interlaboratory CV values were 26% for both NRU test methods, but NHK had a lower mean

interlaboratory CV (28% vs 47% for 3T3 NRU). Interlaboratory CV values ranged from 3-135% in the 3T3 NRU test method and 1-91% in the NHK NRU test method. FAL had the highest mean intralaboratory CV in both NRU test methods (33% in 3T3, 43% in NHK).

An analysis to determine the relationship between the chemical attributes and interlaboratory CV indicated that chemical structure, physical form, solubility, and volatility had little effect on CV. The CV seemed to be related, however, to GHS acute toxicity category, IC_{50}, and boiling point. Mean interlaboratory CV values were larger for substances in the most toxic GHS categories than for substances in the other toxicity categories, especially with the 3T3 NRU test method. The mean interlaboratory CV for substances in the $LD_{50} \leq 5$ mg/kg (72%) and $5 < LD_{50} \leq 50$ mg/kg (78%) classes were larger than the mean overall interlaboratory CV (47%) with the 3T3 NRU test method. The mean interlaboratory NHK CV was 37% for substances with $LD_{50} \leq 5$ mg/kg, and 41% for substances with $5 < LD_{50} \leq 50$ mg/kg, while the mean overall interlaboratory CV was 28%. A Spearman correlation analysis showed that the IC_{50} was inversely correlated to interlaboratory CV for both the 3T3 ($p=0.015$) and NHK ($p=0.014$) test methods, and that boiling point was positively correlated to interlaboratory CV ($p=0.007$) (i.e., higher boiling points were associated with higher CV values) for the 3T3 but not the NHK NRU test method ($p=0.809$).

The ANOVA results for the PC IC_{50} in the 3T3 NRU test method showed that there were significant differences among laboratories ($p=0.006$) but not among study phases within laboratories ($p > 0.01$). However, interlaboratory CV values, which ranged from 2% to 10% for the different study phases, were small and the intralaboratory CV values ranged from 5% to 24%. The SLS IC_{50} values from the NHK NRU test method were more variable than those from the 3T3 NRU test method. The ANOVA results for SLS in the NHK NRU test method indicated that there were significant differences among laboratories ($p < 0.001$) and among study phases within laboratories ($p \leq 0.001$). A change in cell culture methods at FAL after Phase Ib decreased the SLS IC_{50} in subsequent phases, but FAL's CV values still tended to be higher than in the other laboratories. Intralaboratory CV values for the NHK SLS IC_{50} during the various study phases ranged from 11% to 51% and interlaboratory CV values for SLS in the NHK NRU test method ranged from 8% in Phase III to 39% in Phase Ia.

Cell culture medium was used as the solvent for 38 substances and DMSO was used for 34 substances. Concordance among all three laboratories in selecting the solvent for the reference substances was 76% (55/72).