[This Page Intentionally Left Blank]

8.0   3T3 AND NHK NRU TEST METHOD DATA QUALITY

This section of the BRD presents the extent of adherence to GLP regulations for generation of the validation study data. Data quality is described, along with deviations from the regulations and their effect (if any) on the quality of the data. Statistical analyses are provided to compare the data generation, collection, and reporting by the two GLP compliant laboratories and the one non-GLP compliant laboratory, as well as for the GLP-compliant laboratory that distributed the reference substances and performed solubility studies. Discussions of various quality assurance aspects of the study are included.

8.1   Compliance With Good Laboratory Practice Regulations

8.1.1   Guidelines Followed for Cytotoxicity Testing

8.1.1.1   *Good Laboratory Practices*
The SOW provided the following definition of U.S. Regulatory agency GLPs to each laboratory:
> "Regulations governing the conduct, procedures, and operations of toxicology laboratories; regulations to assure the quality and integrity of the data and to address such matters as organization and personnel, facilities, equipment, facility operations, test and control articles, and validation study protocol, and conduct (U.S. Food and Drug Administration, Title 21 CFR Part 58; U.S. Environmental Protection Agency, Title 40 CFR Part 160)."

IIVS, ECBC, and BioReliance performed testing under GLP guidelines. The details of GLP compliance and training are addressed in **Section 11.2**.

8.1.1.2   *Spirit of GLP*
The SMT determined a definition for "spirit of GLP" and provided the following to the laboratories:
> "Laboratories that are non GLP-compliant shall adhere to GLP principles and other method parameters as put forth in this Statement of Work and the Test Method Protocols (provided by NIEHS/NICEATM); documentation and accountability shall be equal to GLP requirements; laboratories must make assurances that they are equal in performance criteria and that there is parity amongst the laboratories."

FAL performed testing in the "spirit of GLP" (see **Section 11.2.2.1**) by following the international GLP standards referenced in the ECVAM Workshop 37 Report (Cooper-Hannan 1999) and the OECD Principles of GLP (OECD 1998). The laboratory did not have their data and test procedures reviewed by an independent, quality assurance (QA) auditor. The SOW directed FAL to, at a minimum, routinely document their equipment monitoring and record keeping (see **Table 8-1**), and to archive all documents. The FAL already had most of the requested procedures and guidelines in place for routine laboratory procedures before initiation of this study. The various general laboratory-related activities were documented in workbooks and logbooks, and the information was made available to the SMT.

Table 8-1   SMT-Recommended Documentation for FAL

Daily	Per Use	Periodic
Temperatures Laboratory (ambient), incubators, water baths, refrigerators, freezers	Cryogenic Storage Unit Liquid N_2 volume	Laboratory Supplies[1] Lot numbers and expiration dates for stock media formulations and components, NRU reagents, tissue culture plasticware
Humidity/CO_2 Cell culture incubators	Equipment Calibration Balances, pH meters, cell counters	Cells Quantity, and cryogenic storage conditions, for 3T3 and NHK cells
Visual Observations Cell Culture Growth	Reagents Lot numbers and expiration dates of medium/supplements	Equipment Calibration Incubators, laminar flow hoods, autoclaves, micropipettors, spectrophotometer plate readers, computers (software)

Abbreviations: SMT=Study Management Team; FAL=Fund for the Replacement of Animals in Medical Experiments Alternatives Laboratory; NRU=Neutral red uptake; 3T3=BALB/c 3T3 fibroblasts; NHK=Normal human epidermal keratinocytes.
[1]Documentation for laboratory supplies begins when supplies are purchased and received by the laboratory

8.1.1.3   *Good Cell Culture Practices (GCCP)*

The SMT provided guidance in the SOW for implementing GLPs in a cell culture laboratory environment. The initial assumption by the SMT was that each laboratory had the basic cell culture skills and knowledge (e.g., as described in Freshney 2000) to reliably perform the *in vitro* NRU cytotoxicity test methods. Reviews of historical laboratory documents, and scientific and professional exchanges with the laboratory personnel, assured the SMT that each laboratory had demonstrated, through previous validation studies and other experience, that the personnel were capable of providing quality scientific data through the use of good cell culture practices. A comparison of the SOW and the *in vitro* NRU cytotoxicity protocols showed that the guidelines developed for the NICEATM/ECVAM study were harmonious with the guidelines in the ECVAM Good Cell Culture Practices Reports (Hartung 2002; Coecke et al. 2005), and the OECD document on GLPs and *in vitro* studies (OECD 2004a).

8.1.2   Quality Assurance (QA) for NRU Cytotoxicity Test Data

8.1.2.1   *Coded Reference Substances*

BioReliance acquired 73 high purity chemicals (72 reference substances and one positive control substance) from reputable commercial sources. Sixty-four of the reference substances were ≥99% pure, and seven were between 90 and 99% pure. Lactic acid had the lowest purity, 89% (See **Appendix F1**). The substances were coded with unique identification numbers and provided to the testing laboratories in a blinded fashion. Procurement of chemicals and their preparation for distribution was performed under GLP guidelines and the SOW provided by the SMT (see **Appendix G**). **Section 3.4** provides detailed information on the acquisition and distribution of reference substances.

8.1.2.2   *Solubility Testing and Data Review*

All laboratories performed solubility tests on all reference substances using the solvents and procedures specified in the protocols provided by the SMT, and submitted solubility data to the SMT in the form of hard copy printouts and electronic worksheets. The Study Directors reviewed all laboratory procedures and all data produced at their respective laboratories, and the QA designee in each GLP-compliant laboratory reviewed all data in their laboratory. The SMT Project Coordinators served as informal QA reviewers for FAL (i.e., reviewed all the raw data sheets). The errors and omissions detected were reported to FAL, and corrections were requested. The SMT reviewed all solubility data and NRU assay data produced by all of the laboratories.

The SMT reviews of the submitted data in Phases Ia and Ib revealed that, even after data review by the Study Directors, the data files contained an unacceptably high frequency of errors (see **Section 2.6.2.5**). The laboratories were alerted to the problem and personnel from all laboratories attended a weeklong training session at the IIVS laboratories in Gaithersburg, Maryland to enhance harmonization among the laboratories. Errors continued to be found in data files submitted for Phase III after the training, albeit less frequently; however, such errors generally resulted from the rush to rapidly complete the data files for submission to the SMT shortly after the conclusion of each test. The formal QA reviews of the files occurred later in each phase of the study.

The most common errors included typographical mistakes, transcriptional and data entry errors in the Microsoft® EXCEL® and the GraphPad PRISM® 3.0 templates, and incorrect labeling of files. The SMT reviewed every electronic file and hard copy printout throughout the study and alerted the Study Directors of the affected laboratories when errors were found. All data files were checked for consistency within the documents, and for compliance with the protocols. The SMT also documented errors on the hard copy printouts in the form of handwritten notations to the files (at NICEATM) and added these notations to the electronic data summary files compiled for data management. Files that were revised and/or corrected by the Study Director were resubmitted to the SMT and identified as corrected files.

8.1.2.3   *NRU Cytotoxicity Test Tallies*

The Study Directors periodically received individualized test tallies specific to their laboratories from NICEATM that detailed:

- The number of range finder tests performed by the laboratory
- The number of definitive tests performed, and the pass/fail status of each
- The number of PC tests performed, and the pass/fail status of each
- The number of acceptable tests completed
- The test completion status for each chemical (i.e., whether one range finder test had been completed, and the number of acceptable definitive tests had been completed)

The laboratories compared the NICEATM tallies to their own records to verify their consistency and accuracy. Discrepancies were resolved through direct communication between the Study Director and the SMT.

8.1.3   Guidelines Followed for Rodent Acute Oral LD$_{50}$ Data Collection

For the purposes of this validation study, the *in vitro* NRU test methods were proposed for predicting starting doses for rodent acute oral toxicity test methods, rather than as replacement tests for the *in vivo* test method. No *in vivo* tests were performed for this validation study. All *in vivo* data (i.e., rat and mouse LD$_{50}$ values) were collected by NICEATM through reviews of the literature and from publicly available databases. All relevant data and pertinent information were gathered and stored in an Excel$^{®}$ spreadsheet.

8.1.3.1   *Rodent Acute Oral LD$_{50}$ Values Used in the Registry of Cytotoxicity (RC)*

The RC is a database of acute oral LD$_{50}$ values for rats and mice obtained primarily from the 1983/84 RTECS$^{®}$ database compiled by NIOSH, and IC$_{50}$ values from *in vitro* cytotoxicity assays using multiple cell lines and cytotoxicity endpoints for chemicals with known molecular weights (Halle 1998, 2003). Collection and reporting methods used for generating the data in RTECS$^{®}$ were not a part of the data collection hierarchy employed by NIOSH, and the data in this database were not evaluated for quality and accuracy. Many of the values come from secondary sources with no citation to the original report. GLP guidelines were not used to determine acceptable data for the database. The only criterion used by NIOSH for reporting acute oral toxicity data in RTECS$^{®}$ was that the LD$_{50}$ value was the most toxic LD$_{50}$ value for a chemical that could be found in the literature, regardless of the number of other values available, or their distribution.

8.1.3.2   *Rodent Acute Oral LD$_{50}$ Values Collected by NICEATM from Other Sources*

One critical aspect of the validation study design was the establishment of a rat acute oral LD$_{50}$ reference value for each of the 72 reference substances (see **Section 4**). These reference values were used to evaluate the extent to which the two *in vitro* NRU test methods could predict rat acute oral LD$_{50}$ values. Primary rat acute oral LD$_{50}$ studies were located through searching electronic databases, published articles, and secondary references. Rat data were not available for three of the reference substances and mouse acute oral LD$_{50}$ values were used. Only seven of the 455 LD$_{50}$ values collected from the literature were produced under GLP guidelines.

8.2   Results of Data Quality Audits

The QA unit or designee in each GLP laboratory provided a systematic and critical comparison of the data provided in the laboratory's study reports to the raw data in the laboratory records. The SOW provided to each laboratory contained the following guidance regarding QA statements:

> "The Final Reports for all phases of the Validation Study shall be audited by the Quality Assurance unit of the Testing Facility for GLP compliance and a QA Statement shall be provided by the Testing Facility. Each Final Report shall identify: 1) the phases and data inspected, 2) dates of inspection, and 3) dates findings were reported to the Study Director and Testing Facility management. The QA Statement shall identify whether the methods and results described in the Final Report accurately reflect the raw data produced during the Validation Study."

8.2.1   QA Statements

The QA statements from the GLP-compliant laboratories addressed the reviews of:
- Protocols
- Laboratory standard operating procedures (SOPs)

- Laboratory operations, in general
- 3T3 and NHK NRU experiment data
- The report submitted to the SMT

The QA statements from the GLP laboratories affirm that the methods described in the protocols are the methods that the laboratory personnel used, and that the data reported to the SMT accurately reflect the raw data obtained by the laboratory. See **Section 8.2.2** for information about the QA statements for the non-GLP laboratory.

8.2.2   QA Statements from the Laboratories

8.2.2.1   *BioReliance QA Statements*
The Study Director/Laboratory Director provided the following statement in all of the final reports:
  "The solubility studies, acquisition, preparation, and distribution of the test chemicals were conducted in compliance with GLP. Although not audited (per SOW), the work described in this report for Phase X (i.e., Ia, Ib, and II) fully and accurately reflects to the best of my knowledge the raw data generated in the study."

8.2.2.2   *FAL QA Statements*
The Study Director for FAL performed the final review of all data and reports before sending them to the SMT, and provided the following two statements in the final reports provided to the SMT.
- "The laboratory worked under the principles of GLP whilst not being a GLP-compliant laboratory."
- "The report accurately reflects the work undertaken and the results obtained at the FRAME Alternatives Laboratory."

Formal QA statements were not provided to FAL because the SMT performed informal QA reviews.

8.2.2.3   *ECBC QA Statements*
The QA statements reported the particular study phase and laboratory procedures that were examined for GLP compliance. In addition, the laboratory's statement noted that the scope of work, associated protocols, and quality control (QC) acceptance criteria were updated or changed during the study, which made the assessment of the procedures and data for conformance to the SOPs more difficult. However, compliance with the requirements and intent of GLP guidelines was continually assessed during the review of the SOPs and the observance of operations. The QA reviews found the ECBC protocols to be in compliance with the NICEATM/ECVAM study protocols. The aspects of the studies inspected by the QA designee were:
- Review of protocols and laboratory SOPs
- Review of waste handling procedures
- Review of laboratory operations
- Certification of new personnel
- Review of data
- Review of the final report for each testing phase

The QA designee also observed the preparation of reference substances, 96-well plate configuration, application of reference substance, annotation to the workbook, and appropriate sterile technique while performing the testing. The number of inspections of laboratory operations was reduced in the latter phases of the study because the same personnel conducted the testing throughout the entire study.

ECBC Review Dates of the Study Phases
- Phase Ia: July 2002 through May 2003
- Phase Ib: July 2002 through January 2003
- Phase II: May 2003 through February 2004
- Phase III: November 2003 through March 2005

8.2.2.4   *IIVS QA Statements*
Because the IIVS QA unit is small, it carried out reviews of different aspects of the procedures at different times. The IIVS QA Statement reads:

"This study has been divided into a series of in-process phases. Using a random sampling approach, Quality Assurance monitors each of these phases over a series of studies. Procedures, documentation, equipment records, etc., are examined to assure that the study is performed in accordance with the U.S. FDA Good Laboratory Practice regulations (21 CFR 58), the U.S. EPA GLP Standards (40 CFR 792 and 40 CFR 160) and the OECD Principles of Good Laboratory Practice and to assure that the study is conducted according to the protocol and relevant Standard Operating Procedures."

The aspects of the studies inspected by the QA designee were as follows:
- Protocol and initial paperwork
- Reading of the plates (definitive test)
- Dilution of the test articles (definitive test)
- Treatment of the cells
- Termination of treatment and addition of the NR dye (definitive test)
- Cell concentration determination and seeding of the plates (third definitive test)
- Termination of treatment and addition of the NR dye
- Washing the cells
- Draft report and data
- Final report

IIVS Review Dates of Various Aspects of the Test Phases
- Phase Ia: August 2002   Final Report Review: October 2005
- Phase Ib: January 2003   Final Report Review: October 2005
- Phase II: July-August 2003   Final Report Review: October 2005
- Phase III: January-November 2004   Final Report Review: October 2005

8.2.2.5   *Other QA Information*
Data generated by the laboratories and reviewed by their respective Study Directors were submitted to the SMT. Often, the data were provided electronically within days of the end of testing. The SMT was active as a secondary QA reviewer of all information provided by the

Study Directors. If the SMT found discrepancies, the Project Coordinators corresponded with the appropriate Study Director to identify and rectify the error. The Study Director made corrections/adjustments to the discrepancies in data reporting and presented the changes to the SMT. The SMT did not initiate any external data quality audits.

The quality of the reference substances was assured in the form of certificates of analysis provided by the chemical manufacturer to BioReliance at the time of purchase. The SMT and the laboratories obtained certificates of analysis from CAMBREX for Clonetics® NHK culture medium and supplements. In addition, the SMT obtained QC data directly from CAMBREX technical departments concerning the NHK medium's ability to support keratinocyte growth.

8.3   Effect of Deviations or Non-compliance with GLPs

Rates for several types of errors (i.e., documentation, testing methods, and data management) were determined by the SMT. Many of the errors (particularly in Phases Ia and Ib) were the result of minor mistakes (e.g., typographical, mislabeling) and did not affect the quality of the data.

8.3.1   Laboratory Error Rates

The SMT was concerned about the number of errors that were seen in documentation and testing methods during Phases Ia and Ib, and compiled the detected errors from each laboratory. The types of errors found included errors in documentation (e.g., reference substance identification did not match on all associated data sheets; IC_{20} and IC_{80} values were transposed in the EXCEL® template; a test acceptance criterion flag in a data sheet was incorrect) and in testing (e.g., wrong dilution scheme was used for the PC; wrong SLS IC_{50} was used as the PC IC_{50}). Error rates were compiled as the number of tests with errors per total number of tests. As shown in **Table 2-3**, FAL had the highest error rates: 93% for the 3T3 NRU test method and 41% for the NHK NRU test method. The highest error rates in the other laboratories were 10% for the 3T3 NRU test method and 23% for the NHK NRU test method (both ECBC).

There were relatively few errors detected in the Phase III data files. The SMT did not compile the typographical and transcriptional errors found, but reported them directly to the appropriate Study Director so that the data sheets could be immediately corrected. The SMT did not detect errors in the raw optical density data from the 96-well plates provided in each data file. The laboratories and the SMT corrected typographical and transcriptional errors (e.g., incorrect $logIC_{50}$ value entered) in the EXCEL® templates. The EXCEL® template formulas were used for the statistical analyses.

An assessment of error rates was performed specifically for Phase III for one particular clerical error – the transfer of the final results (e.g., ICx values) from the GraphPad PRISM® 3.0 template to the Microsoft® EXCEL® template. It was often necessary for the SMT to revise the EXCEL® data files provided by the laboratories because the incorrect values had been transferred to EXCEL®. **Table 8-2** summarizes the Phase III error rates resulting from the transfer of data from PRISM® to EXCEL®.

Table 8-2   Phase III Error Rates in the Transfer of Data to the EXCEL® Template

Laboratory	Number of Errors Detected	Number of Definitive Tests	Percentage of Tests with Detected Errors
ECBC	49	402	12
FAL	171	513	33
IIVS	25	419	6

Abbreviations: ECBC=Edgewood Chemical Biological Center; FAL=Fund for the Replacement of Animals in Medical Experiments Alternatives Laboratory; IIVS=Institute for *In Vitro* Sciences.

8.3.2   Failure Rates for Definitive and PC Tests

Table 8-3 presents the test failure (i.e., did not meet test acceptance criteria) rates experienced in Phase III. Approximately 25% of all 3T3 definitive tests and 18% of all NHK definitive tests failed. If a definitive test (see **Section 2.3.2.2** for the definition of a definitive test) failed, the laboratory repeated the test and attempted to obtain three acceptable definitive tests for each reference substance in each cell type (see **Section 2.5** for criteria for repeating tests). The PC tests failed 0 to 18% of the time with a combined average failure rate of 8% for both cell types. FAL had the highest individual laboratory test failure rates for 3T3 definitive tests (30%), NHK definitive tests (32%), and NHK PC tests (18%). ECBC had the highest failure rate for 3T3 PC tests (11%). IIVS had no PC test failures.

Table 8-3   Definitive Test and Positive Control (PC) Test Failure Rates in Phase III

Test Type	3T3 NRU Test Method				NHK NRU Test Method				Total
	ECBC	FAL	IIVS	Total	ECBC	FAL	IIVS	Total	
Definitive Tests - Acceptable	169	177	176	522	173	175	174	522	1044
Definitive Tests - Total	215	257	225	697	187	256	194	637	1334
% Failed Definitive Tests	21	30	22	25	8	32	10	18	22
PC Tests - Acceptable	66	40	16	122	58	37	20	115	237
PC Tests - Total	74	42	17	133	59	45	20	124	257
% Failed PC Tests	11	5	6	8	2	18	0	7	8
Definitive Tests Failed Only Because PC Tests Failed	14	6	14	34	0	22	0	22	56
% Definitive Tests Failed Only Because PC Tests Failed	7	2	6	5	0	9	0	4	4

Abbreviations: ECBC=Edgewood Chemical Biological Center; FAL=Fund for the Replacement of Animals in Medical Experiments Alternatives Laboratory; IIVS=Institute for *In Vitro* Sciences; NRU=Neutral red uptake; 3T3=BALB/c 3T3 fibroblasts; NHK=Normal human epidermal keratinocytes.

The Phase III guidelines required each laboratory to provide three acceptable definitive tests for each substance for both cell types (3 x 60 x 2 = 360 definitive tests). PC tests were run concurrently with the definitive tests, and more than one reference substance was usually tested in conjunction with each PC test. Because of test failures, each laboratory performed additional testing to obtain the three acceptable definitive tests required for each substance.

Table 8-4 presents the success rates for each laboratory for Phase III testing and a total for all the laboratories combined.

Table 8-4   Combined Definitive and Positive Control (PC) Test Success Rates for the 3T3 and NHK Methods in Phase III

Test Type	ECBC	FAL	IIVS	Total
Acceptable Definitive Tests/ Total Definitive Tests	342/402	352/513	350/419	1044/1334
% Acceptable Definitive Tests	85%	69%	84%	78%
Acceptable PC Tests/Total PC Tests	124/133	77/87	36/37	237/257
% Acceptable PC Tests	93%	89%	97%	92%

Abbreviations: ECBC=Edgewood Chemical Biological Center; FAL=Fund for the Replacement of Animals in Medical Experiments Alternatives Laboratory; IIVS=Institute for *In Vitro* Sciences; NRU=Neutral red uptake; 3T3=BALB/c 3T3 fibroblasts; NHK=Normal human epidermal keratinocytes.

8.3.3   Intralaboratory Reproducibility

CV values for each method were determined for each reference substance in each laboratory using the IC_{50} values from the acceptable definitive tests, as described in **Section 5.5.2**. **Table 8-5** presents the average CV values for the substances tested in each of the study phases, and for the entire study.

Table 8-5   CV Values for Definitive Tests

Cell Type	Labs	Phases I & II		Phase III		All Phases	
		Number of Reference Substances	Average % CV	Number of Reference Substances	Average % CV	Number of Reference Substances	Average % CV
3T3	ECBC	12	17	57	24	69	23
	FAL	11	28	55	33	66	33
	IIVS	11	20	56	22	68	21
NHK	ECBC	12	24	57	22	69	23
	FAL	12	31	57	45	69	42
	IIVS	12	14	58	14	70	14

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; NHK=Normal human epidermal keratinocytes; CV=Coefficient of variation.

8.3.4   Prediction of GHS Acute Oral Toxicity Categories

Predicted LD_{50} values were determined using the *in vitro* NRU IC_{50} values in the IC_{50}-LD_{50} 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 categories was determined by comparison with categorization based on *in vivo* rat oral LD_{50} values (in mg/kg) in **Table 4-2**. Using the RC rat-only millimole regression, the accuracy of the predictions and the extent of underprediction or overprediction are shown for each laboratory in **Table 8-6**. The laboratories generally agreed with each other in their predictions. Although FAL had the highest error rates and CV values, their predictions of GHS categories were consistent with the other laboratories. The laboratories determined category matches for 25 to 30% of the reference substances for the 3T3 NRU test method and 29 to 31% of the reference substances for the NHK NRU test method. For the 3T3 NRU test method, toxicity was overpredicted for 38% of the reference substances and underpredicted for 33 to 38% of them. For the NHK NRU test method, toxicity was overpredicted for 35 to 38% of the reference substances and underpredicted for 32 to 34% of them. (See **Appendix J** for additional laboratory comparisons for the other *in vitro – in vivo* regressions evaluated in **Section 6**.)

8.4   Availability of Laboratory Notebooks

All laboratories maintained laboratory notebooks using a template provided by IIVS, and provided copies of the notebooks to the SMT (archived at NICEATM) after completion of each testing phase. The notebooks contained information from all aspects of testing including, but not limited to:

- Environmental conditions
- Reagent identification
- Preparation of 96-well plates
- Preparation of reference substances
- Treatment of cell cultures
- Visual observations of cell cultures
- NRU assays
- Data analysis

Table 8-6   GHS Acute Oral Toxicity Category Predictions by Laboratory[1]

	Labs	Total Reference Substances	Category Match	Toxicity Overpredicted	Toxicity Underpredicted
3T3	ECBC	64	30%	38%	33%
	FAL	64	25%	38%	38%
	IIVS	64	27%	38%	36%
NHK	ECBC	68	31%	35%	34%
	FAL	68	29%	38%	32%
	IIVS	68	31%	37%	32%

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; NHK=Normal human epidermal keratinocytes; GHS=Globally Harmonized System for Classification and Labelling of Chemicals (UN 2005).
[1]3T3 and NHK NRU test method IC_{50} data (geometric mean of within laboratory replicates) used with the RC rat-only millimole regression, log LD_{50} (mmol/kg) = 0.439 x log IC_{50} (mM) + 0.621, to assign GHS category. *In vivo* category was based on reference rodent oral LD_{50} values (mg/kg) in **Table 4-2**. For each method, the reference substances evaluated were those for which all three laboratories obtained IC_{50} values.

8.5   Summary

- The determinations of test method and data collection errors showed that FAL consistently had the highest error levels; however, the laboratory's GHS acute oral toxicity category predictions were comparable to the other laboratories' results.
- The laboratories reported no significant deviations from the protocols, and deviations that did occur during the testing phases were generally quickly acknowledged and addressed by the Study Directors. If a deviation occurred that would affect the data (e.g., improper concentration of DMSO solvent), the Study Director would reject the test, notify the SMT, and perform an additional test. Improper transfer of data to either the EXCEL® or PRISM® templates, which would affect the data summaries and analyses, were recognized, documented, and rectified by the Study Director and/or the SMT.
- The SMT reviewed all data sheets to ensure that data were not inadvertently attributed to the incorrect data summary files, and that the correct data were used in all statistical analyses.
- An electronic copy of all data for this validation study can be obtained from NICEATM upon request by mail, fax, or e-mail to Dr. William S. Stokes, NICEATM, NIEHS, P. O. Box 12233, MD EC-17, Research Triangle Park, NC, 27709, (phone) 919-541-2384, (fax) 919-541-0947, (e-mail) niceatm@niehs.nih.gov.

[This Page Intentionally Left Blank]

[This Page Intentionally Left Blank]

9.0   OTHER SCIENTIFIC REPORTS AND REVIEWS OF *IN VITRO* CYTOTOXICITY TEST METHODS AND THEIR ABILITY TO PREDICT *IN VIVO* ACUTE TOXICITY AND OTHER TOXIC EFFECTS

In vitro cytotoxicity methods based on NRU have been evaluated for a number of uses. This section reviews studies that used *in vitro* NRU cytotoxicity methods to:

- Predict acute rodent oral toxicity
- Predict starting doses for acute systemic toxicity tests
- Predict other *in vivo* toxicity endpoints, including phototoxicity and eye irritation.

Section 9.1 describes studies that evaluated *in vitro* cytotoxicity test methods that measured NRU for its ability to predict acute systemic toxicity in rodents, and other *in vivo* endpoints. Also reviewed are studies that evaluated the use of *in vitro* cytotoxicity results to reduce animal use in acute toxicity testing. **Section 9.2** reviews independent evaluations of the use of *in vitro* cytotoxicity methods to predict acute oral toxicity, and to determine starting doses for acute systemic toxicity assays. Also discussed is a 3T3 NRU test method that has been validated and accepted for regulatory use for detecting phototoxic potential using a protocol similar to that used in the NICEATM/ECVAM validation study. The conclusions of these reports will be compared to the conclusions reached in this study, wherever possible. **Section 9.3** reviews published studies that used the *Guidance Document* approach (ICCVAM 2001b), which established the current test method performance standard.

9.1   Relevant Studies

9.1.1   Correlation of NRU Cytotoxicity Values with Rodent Lethality

This section reviews five published *in vitro* cytotoxicity studies that correlated cytotoxicity values (i.e., IC_{20} or IC_{50}) from NRU cytotoxicity test methods that used various cell types, to rat and/or mouse acute LD_{50} values from various exposure routes. In these sections, *italics* are used to identify reference substances tested in the reviewed studies that were also tested in the NICEATM/ECVAM validation study. **Table 9-1** characterizes the substances tested in the reviewed studies by providing the ranges of their rat oral LD_{50} values. Also shown for comparison are the mouse and/or rat oral LD_{50} ranges for the NICEATM/ECVAM validation study and the RC. The table shows that the substances tested by Peloux et al. (1992), Fautrel et al. (1993), and Rasmussen (1999), covered a wide range of rat acute LD_{50} values. The substances used by Roguet et al. (1993) and Creppy et al. (2004) covered a much smaller range. **Table 9-2** characterizes the test substances by chemical class based on NLM Medical Subject Heading (MeSH®) descriptors.

Table 9-1   **Rat Acute Oral LD$_{50}$ Ranges for Test Substances Used in Previous *In Vitro* NRU Cytotoxicity Studies and the NICEATM/ECVAM Study[1]**

Study/Database	N	Rat Acute Oral LD$_{50}$ Range (mg/kg)[2]
Peloux et al. (1992)	30	2 – 14500
Fautrel et al. (1993)	31	2 – 14500
Roguet et al. (1993)	28	0.04 – 176
Rasmussen (1999)	20	1 – 10298
Creppy et al. (2004)	2	48 – 924[5]
NICEATM/ECVAM Validation[3]	72	2 – 19770
RC[4]	347	1 – 31015

Abbreviations: N=Number of substances in the study/database; RC=Registry of Cytotoxicity.
[1]Studies reviewed in **Section 9.1.1**.
[2]Values cited in the studies or from references provided by the studies.
[3]Current study summarized in this BRD.
[4]The RC includes both rat and mouse LD$_{50}$ values.
[5]Upper limit of range is an LD$_{50}$ calculated from the *in vitro* NRU IC$_{50}$ because there was no *in vivo* value available for that substance.

Table 9-2   **Chemical Classes Represented by the Substances Used in Published Studies for Correlation of *In Vitro* NRU Cytotoxicity with Rodent Acute Lethality**

Chemical Class[1]	Study[2]	Chemical Class[1]	Study[2]	Chemical Class[1]	Study[2]
Alcohols	1, 2, 3, 4	Fluorine	3, 4	Nitriles	1, 2
Amides	1, 2, 3	Heterocyclics	1, 2, 3, 4, 5	Nitrogen	3, 4
Amines	1, 2	Hydrocarbons	1, 2, 3, 4, 5	Organophosphates	3, 4
Arsenicals	3, 4	Iron	3	Phenols	3, 4
Carboxylic Acids	1, 2, 3, 4	Lactones	1, 2	Polycyclics	3
Chlorine	3, 4	Lithium	1, 2, 3, 4	Potassium	3, 4
Copper	3, 4	Mercury	3, 4	Sodium	3, 4
Ethers	1, 2	Metals	3, 4	Sulfur	1, 2, 3, 4

Study references: 1=Peloux et al. (1992) (24/25 substances were organic compounds); 2=Fautrel et al. (1993) (30/31 substances were organic compounds); 3=Roguet et al. (1993) (22/30 substances were organic compounds); 4=Rasmussen (1993) (13/20 substances were organic compounds); 5=Creppy et al. (2004) (2/2 substances were organic compounds).
[1]Classification by NLM Medical Subject Heading (MeSH®) descriptors.
[2]Studies reviewed in **Section 9.1.1.**

9.1.1.1   *Peloux et al. (1992)*

The authors used several different *in vitro* cytotoxicity methods with primary rat hepatocytes to determine the correlation with rat/mouse intraperitoneal (i.p.) or intravenous (i.v.) LD$_{50}$ values for the 25 substances tested. The *in vitro* cytotoxicity methods, which used 20-hour test substance exposure durations, assessed the following endpoints: NRU, total protein content, LDH release, MTT reduction. MTT is metabolized by mitochondrial succinate dehydrogenase of viable cells to yield a purple formazan reaction product. The IC$_{50}$ values

obtained using the four endpoints were highly correlated (r = 0.973 to 0.999) to each other. When performing the IC_{50}-LD_{50} regressions, Peloux et al. (1992) used the lowest reported published LD_{50} value for acute rat or mouse studies that administered the test substances using the i.p. or i.v. routes. The IC_{50} values obtained using NRU as the endpoint had the highest correlation coefficient, r = 0.877, to the rat/mouse i.p./i.v. LD_{50} values. The total protein assay yielded r = 0.872, the MTT reduction assay yielded r = 0.808, and the LDH release assay yielded r = 0.789.

Peloux et al. (1992) followed the recommendations of Fry et al. (1988, 1990) and used parenteral LD_{50} values rather than oral LD_{50} values for comparison with *in vitro* values. Fry et al. (1988, 1990) recommended the use of the i.p./i.v. LD_{50} values for comparisons because they proposed that cells *in vivo* receive a more direct test substance exposure via these routes than through the oral route. They had posited that *in vitro* cell cultures would mirror this (direct) toxicity because they also receive direct exposure to test substances via the cell culture medium. The authors also noted that the oral route of exposure presents confounding variables such as, 1) only a fraction of a test substance would be available in the systemic circulation due to limited absorption or pre-systemic metabolism, and 2), the level of the substance in the systemic circulation decreases due to elimination mechanisms (e.g., metabolism, excretion). Fry et al. (1990) had reported a correlation of only r = 0.49 for *in vivo/in vitro* comparisons of oral LD_{50} and IC_{50} values (from a total protein assay) and a correlation of r = 0.68 for i.p. LD_{50} and ID_{50} values[1].

9.1.1.2   *Fautrel et al. (1993)*
Six laboratories tested the cytotoxicity of 31 substances in primary rat hepatocyte cultures using a 24-hour exposure followed by measurement of NRU. The investigators performed linear regression analyses for the prediction of rat i.v., i.p., and oral LD_{50} values from the NRU IC_{50} values. The regressions for the various *in vivo* administration routes did not use the same substances because LD_{50} values were not available for all of the tested substances in all of the routes. Oral, i.v., and i.p. LD_{50} values were available for 27, 24, and 18 substances, respectively, and IC_{50} values were obtained for 15, 14, and 11 of these substances, respectively. The regression for the i.v. data was statistically significant (r = 0.88, n = 11), but the i.p. (r = 0.48, n = 14) and oral regressions (r = 0.17, n = 15) were not. The finding that the i.v. LD_{50} values corresponded more closely with the *in vitro* cytotoxicity data than did the oral LD_{50} was thought to be the result of having fewer pharmacokinetic variables (i.e., absorption, distribution, etc.) to consider following i.v. administration.

9.1.1.3   *Roguet et al. (1993)*
Roguet et al. (1993) tested the cytotoxicity of 28 MEIC substances in primary rat hepatocytes exposed for 21 hours, followed by the measurement of NRU. A correlation of the NRU IC_{50} values to oral LD_{50} values obtained from the unpublished data of B. Ekwall et al. (personal communication) yielded a statistically significant linear correlation (p <0.001) with r = 0.80 when the *in vivo* and *in vitro* data were in molar units. [NOTE: The LD_{50} values subsequently published by Ekwall et al. (1998) were from the 1997 edition of RTECS®.] The authors reported that the toxicities of thioridazine, malathion, and *copper sulfate* were overestimated, and the toxicity of *potassium cyanide* was underestimated by the correlation, but their criteria for over- and under- estimation were not provided.

---

[1] ID_{50}: index of cytotoxicity; concentrations (µg/mL) producing a 50% reduction in protein value.

The *in vivo* toxicity of *potassium cyanide* was also underpredicted in the NICEATM/ECVAM validation study. **Table 6-3** shows that *potassium cyanide* was an outlier for which toxicity was underpredicted when using the IC_{50} values from both the 3T3 and NHK NRU test methods in the RC millimole regression (log LD_{50} mmol/kg = 0.435 log IC_{50} mM + 0.625). The GHS category predictions using both NRU test methods and the RC rat-only millimole regression (log LD_{50} mmol/kg = 0.439 log IC_{50} mM + 0.621), and the RC rat-only weight regression (i.e., log LD_{50} =0.372 log IC_{50} + 2.024), were also higher (i.e., less toxic) than the *in vivo* category (see **Appendix L2**).

9.1.1.4   *Rasmussen (1999)*

Twenty MEIC substances were tested for cytotoxicity using NRU release from 3T3 cells following 24-hr exposure, with and without the addition of a Aroclor-induced rat liver microsomal preparation (S9 mix). Similar to the present validation study, Rasmussen (1999) observed that *xylene* was non-toxic to the cells, even though it was dissolved in ethanol instead of DMSO. In the presence of S9, the cytotoxicities of malathion, 2,4-dichlorophenoxyacetic acid, *propranolol*, thioridazine, *lithium* sulfate, *copper sulfate*, and *thallium sulfate,* were significantly decreased (p <0.05), while the cytotoxicities of *1,1,1-trichloroethane*, *phenol, nicotine*, and *paraquat* were significantly increased (p <0.05).

Because the NICEATM/ECVAM validation study used cells with little or no xenobiotic metabolizing capability, it could be expected that these systems would overpredict the toxicity of substances that would be inactivated by the addition of a metabolizing system, or to underpredict the toxicity of substances that are metabolized to more toxic substances. None of the four substances in common for which toxicity was decreased by the addition of S9 were overpredicted in the NICEATM/ECVAM study. However, the toxicities of two of the four substances in common for which toxicity was increased by the addition of S9, were underpredicted in the NICEATM/ECVAM study. **Table 6-3** shows that *nicotine* was an outlier whose toxicity was underpredicted when using the 3T3 and NHK IC_{50} values in the RC millimole regression (log LD_{50} mmol/kg = 0.435 log IC_{50} mM + 0.625). *Paraquat* was an outlier whose toxicity was underpredicted when using the NHK IC_{50} value in the RC millimole regression. The GHS category predictions for both substances using both NRU test methods with the RC rat-only millimole regression (log LD_{50} mmol/kg = 0.439 log IC_{50} mM + 0.621) and the RC rat-only weight regression (log LD_{50} mg/kg = 0.357 log IC_{50} μg/mL + 2.194) were also higher than the *in vivo* category (see **Appendix L2**).

Although both the IC_{20} and IC_{50} values were determined in the Rasmussen (1999) study, only the IC_{20} values were used for correlations with the rat acute oral LD_{50} values from RTECS[®]. The units of the LD_{50} values were not reported, but the correlations were assumed to be in molar units because the IC_{20} and IC_{50} values were reported in μM units. Significant correlations (p <0.001) between IC_{20} and LD_{50} values were obtained with and without rat liver microsomes. The correlation of IC_{20} with LD_{50} was slightly higher with the S9 mix (r = 0.72 vs. 0.68 for oral LD_{50} values, and 0.82 vs. 0.78 for i.p. LD_{50} values).

Although the presence of S9 increased the cytotoxicity of some substances to the 3T3 cells, it decreased the toxicity of others, and yielded only a small improvement in the correlation to *in vivo* data. Rasmussen (1999) concluded that the toxicity of the S9 mix (0.32 mg protein/mL), itself, was insignificant because it reduced cell survival by less than 10% compared with cells

without S9. However, others have shown that S9 microsomal mixes could produce significant cytotoxic effects. Kohn (1993) showed that an S9 mix containing 0.07 mg protein/mL was cytotoxic to all types of murine neurons in culture when the cells were exposed for four days or longer. Non-neuronal cells tolerated higher concentration exposures of S9, but exhibited cytoplasmic inclusions when exposed to S9 at 0.35 mg protein/mL. Dal Negro et al. (2006) reported 100% cell death of human monocyte-derived U-937 cells when the S9 fraction (1 mg protein/mL) and co-factors were applied to the cells for a 72-hour incubation. Both of these studies used longer exposure durations, and/or higher protein concentrations, than the Rasmussen (1999) study.

9.1.1.5   *Creppy et al. (2004)*

Creppy et al. (2004) used a 48-hour NRU assay to determine the cytotoxicity of ochratoxin A (OTA) and fumonisin B1 (FB1) on cultured C6 glioma (rat brain), Caco-2 (human intestinal), and Vero (green monkey kidney) cells. The IC_{50} determined in the NRU assay was used in the RC millimole regression to predict rodent acute oral LD_{50} values. The predicted LD_{50} for OTA using the C6 glioma cells was similar to mouse LD_{50} values generated from four *in vivo* mouse studies, but the LD_{50} values predicted by the other cell lines were about 50 times greater. The authors found the relative insensitivity of the Vero cells surprising because OTA is a kidney toxin. There were no available *in vivo* rodent oral LD_{50} values with which to compare the predicted LD_{50} of FB1, which ranged from 671 to 924 mg/kg for the three cell types tested.

9.1.2   Use of *In Vitro* Cytotoxicity Data to Reduce the Use of Animals in Acute Oral Toxicity Testing

9.1.2.1   *Halle et al. (1997): Animal Savings with the ATC Method Using Cytotoxicity Data*

This study assessed the animal savings that would be produced by using IC_{50} data in an IC_{50}-LD_{50} regression to determine a starting dose for ATC testing. No cytotoxicity testing was performed for this study. Instead, the authors used the IC_{50} values from the RC database and the RC millimole regression to predict the LD_{50} for 347 RC substances. The predicted LD_{50} values were then used to determine the starting doses for simulated ATC testing.

At the time of the Halle et al. (1997) study, the ATC method (1996 version from OECD) was designed to classify substances using three classes of acute oral toxicity and an unclassified group, as defined by the acute oral toxicity classification system of the EU (see **Table 9-3**). As a result, the fixed doses for the ATC testing were 25, 200, and 2000 mg/kg. The authors used the LD_{50} predicted by the RC IC_{50} and the RC millimole regression for the 347 RC substances as a starting point to estimate the number of ATC dose steps, and number of animals, that would be needed to classify the substances in the EU category associated with the rodent oral LD_{50} (i.e., rat or mouse values from RTECS®). The method required the simulated ATC testing for each substance to start at the fixed ATC dose nearest to the predicted LD_{50}. The outcome of the simulated testing of three animals per fixed dose was determined by the *in vivo* LD_{50}. If the test dose was lower than the *in vivo* LD_{50}, animals were assumed to live and, if the test dose was higher than the LD_{50}, the animals were assumed to die. Testing of the substance would proceed with higher (when the animals lived) or lower fixed doses (when the animals died) until the substance was placed into the EU toxicity category indicated by the *in vivo* rodent oral LD_{50}.

Table 9-3   EU[1] Classes of Acute Oral Toxicity

Category	LD_{50} (mg/kg)
1	$LD_{50} \leq 25$
2	$25 < LD_{50} \leq 200$
3	$200 < LD_{50} \leq 2000$
Unclassified	$LD_{50} > 2000$

Abbreviations: EU=European Union
[1]Anon (1993)

The method of Halle et al. (1997) can be illustrated with digoxin, which has an *in vivo* mouse LD_{50} of 18 mg/kg (from RTECS®). The predicted LD_{50} of 414 mg/kg was calculated using the RC IC_{50} in the RC millimole regression (log LD_{50} [mmol/kg] = 0.435 x log IC_{50} (mM) + 0.625). Simulated ATC testing would start at the nearest fixed dose, 200 mg/kg. The three animals tested were assumed to die, and then three more animals would be tested at 25 mg/kg. The animals tested at 25 mg/kg were assumed to die and digoxin would be classified in category 1 for $LD_{50} \leq 25$ mg/kg. Thus, the classification of digoxin using the 4-category system required six animals.

Using such simulations of ATC testing, Halle et al. (1997) estimated that 2139 animals would be used to test the 347 substances:
- Three hundred twenty-eight would require testing with two doses using three test animals each.
- Nineteen would require testing with three doses using three animals each.

Halle et al. (1997) cited Schlede et al. (1995) in reporting that the average number of animals required to classify substances using the ATC method was 9.11 animals per test. Using this average, ATC testing of the 347 RC substances would require 3161 animals. Thus, Halle et al. (1997) estimated that there would be a 32% reduction ([3161-2139]/3161) in the number of test animals used when the LD_{50} prediction from the RC millimole regression was used with the 1996 version of the ATC method, in lieu of the standardanimal classification procedure (Halle et al. 1997).

The simulated average animal savings for the ATC in the NICEATM/ECVAM validation study at dose-response slopes of 2.0 and 8.3 was 4.8% to 10.2% (0.51 to 1.09 animals) for the 3T3 (67 reference substances) and NHK (68 reference substances) NRU test methods (see **Section 10.3.3.2**), depending on the regression evaluated. This is considerably lower than the average savings of 32% estimated by Halle et al. (1997). However, there are a number of differences between the evaluation performed by Halle et al. (1997) and the NICEATM/ECVAM study that contribute to the difference in calculated animal savings:
- The NICEATM/ECVAM study used six GHS acute toxicity categories for classification whereas Halle et al. (1997) used the EU toxicity classification scheme, which had only four toxicity categories. The accuracy of category prediction by any method would be higher with fewer categories.

- The NICEATM/ECVAM study used experimentally derived *in vitro* cytotoxicity data from a standardized protocol to estimate starting doses (using two regressions based on the RC substances with rat LD_{50} data), whereas Halle et al. (1997) used IC_{50} data from the RC database.
- The reference substances tested in the NICEATM/ECVAM study poorly fit the RC millimole regression. Nearly half of the reference substances evaluated were outliers (28/70 [40%] in the 3T3 NRU test method, and 31/71 [44%] in the NHK NRU test method) (see **Table 6-3**). The RC database had 95/347 (27.4%) substances outside of the prediction intervals.
- The NICEATM/ECVAM study used computer simulations of ATC testing, which incorporated assumptions about mortality distributions, to determine animals used, whereas Halle et al. (1997) used simplified assumptions (i.e., all animals lived when test dose was less than the *in vivo* LD_{50} and all animals died when test dose was greater than the *in vivo* LD_{50}).
- The NICEATM/ECVAM study determined animal savings by comparing animal use with starting doses determined by the *in vitro* data, to animals used at the default starting dose of 300 mg/kg. Halle et al. (1997) used the average animal use for the ATC for comparison to animal use with simulated testing.

9.1.2.2   *Spielmann et al. (1999): Animal Savings Using Cytotoxicity Data with the UDP*
Spielmann et al. (1999) recommended an *in vitro* cytotoxicity procedure as a range finding test for the *in vivo* toxicity test to reduce the number of animals used in acute toxicity tests. The authors identified nine substances in both the RC database and an evaluation of acute toxicity methods by Lipnick et al. (1995). They then compared the LD_{50} values from Lipnick et al. (1995) to LD_{50} predictions calculated when using the RC IC_{50} values in the RC millimole regression formula (log LD_{50} [mmol/kg] = 0.435 x log IC_{50} [mM] + 0.625). For seven of the nine substances, the LD_{50} prediction was within an order of magnitude of the experimental LD_{50} reported by Lipnick et al. (1995). Spielmann et al. (1999) concluded that the RC millimole regression provided an adequate prediction of LD_{50}, and that *in vitro* cytotoxicity data could be used to predict starting doses for the UDP. The authors recommended using the IC_{50}, with the RC millimole regression, to calculate a starting dose (i.e., an estimated LD_{50}) for the UDP, FDP, or ATC method whenever an IC_{50} was available.

If no IC_{50} was available, Spielmann et al. (1997) recommended determining cytotoxicity using a standard cell line and specific cytotoxic endpoint (e.g., NRU, total protein, MTT reduction). They recommended testing 10 to 20 RC substances to demonstrate that the *in vitro* cytotoxicity test methods provide results that are consistent with the RC millimole regression. The resulting IC_{50} values would then be used to calculate a new regression (using the LD_{50} values reported in the RC), which would be compared to the RC millimole regression. If the new regression fit into the acceptance interval (± log 5 of the fitted regression line) of the RC millimole regression, the RC millimole regression would be used to predict starting doses for the UDP. If the new regression is parallel to the RC millimole regression, but outside the ± log 5 acceptance interval, then the new regression would be used for the prediction of the starting dose.

Spielmann et al. (1999) contended that the RC millimole regression provides a sufficient prediction of LD_{50} values from IC_{50} values for substances that do not require metabolic

activation and are not very toxic (i.e., $LD_{50} > 200$ mg/kg). The authors acknowledged that the fit of substances with $LD_{50} < 200$ mg/kg to the RC millimole regression is not good, and attributed the poor fit of these substances to the need for metabolic activation to a more toxic substance. They suggested that the prediction of starting doses using cytotoxicity data can be used with the UDP and ATC methods, but not with the FDP because dosing is not sequential (which contradicted a claim made earlier in the paper that the approach could be used with the FDP). They did not estimate the number of animals that might be saved with this approach, but did recommend that the approach be validated experimentally using several established cell lines with a limited number of representative substances from the RC.

9.1.2.3   *EPA (2004): U.S. EPA HPV Challenge Program Submission*
In response to the EPA HPV Chemical Challenge Program, PPG Industries, Inc., the manufacturer of Propanoic acid, 2-hydroxy-, compound with 3-[2-(dimethylamino)ethyl] 1-(2-ethylhexyl) (4-methyl-1,3-phenylene)bis[carbamate] (1:1) [CASRN 68227-46-3], and the sponsor of this compound, submitted data to the EPA. This is an isolated intermediate used to produce a resin component of paint products. PPG provided the following types of data in their submission to the EPA: physical-chemical, environmental fate and pathway, ecotoxicity, and toxicology. The acute mammalian toxicology data were generated using both *in vitro* and *in vivo* methods.

An *in vitro* NRU cytotoxicity test was conducted with 3T3 cells to estimate a starting dose for the *in vivo* acute UDP oral toxicity test (OECD 2001a) (see **Appendix M1** for the OECD UDP test guideline). The use of this *in vitro* NRU test method was intended to minimize the number of animals used for *in vivo* testing. The estimated LD_{50} of the compound as determined by the NRU assay was 489 mg/kg. Therefore, the starting dose for the UDP study was set at 175 mg/kg, which is the first default dose below the estimated LD_{50} value; this is also the default starting dose for the UDP, and is used when no information on which to base a starting dose is available. A total of fifteen female rats received the compound at 175, 550, or 2000 mg/kg. Five of nine rats treated at 2000 mg/kg died prematurely on Days 2 and 3, and by Day 15, 2/4 surviving animals at this dose had lost up to 25% of their Day 1 body weights. The LD_{50} was estimated to be 2000 mg/kg, with a 95% confidence interval of 1123-5700 mg/kg. Thus, the *in vitro* NRU test method overpredicted the toxicity of the compound by estimating an LD_{50} value that was lower than that determined in the UDP test. The report authors reported that a greater than predicted number of animals was used for the UDP testing because the estimated LD_{50}, 489 mg/kg and, consequently, the starting dose, was much lower than the *in vivo* LD_{50} of 2000 mg/kg. However, because the UDP started with the default starting dose of 175 mg/kg, the claim that more animals were used is incorrect, because animal use with the default starting dose is the baseline against which other animal use should be compared.

9.1.3   Other Evaluations of 3T3 or NHK NRU Test Methods
This section briefly reviews five studies that evaluated NRU test methods for purposes other than the prediction of starting doses for acute oral toxicity assays. NRU test methods using either 3T3 or NHK cells have been evaluated for use as alternatives to the Draize eye irritation test, to measure phototoxicity, and to predict acute lethality in humans. Except for the 3T3 NRU phototoxicity assay, NRU methods have not been scientifically validated by an independent review for any of these purposes or accepted for regulatory use. The use of the

validated 3T3 NRU test method to determine phototoxic potential is addressed in **Section 9.2**.

The *in vitro* NRU protocols evaluated in the five reviewed studies are similar to those used in the NICEATM/ECVAM validation study, all of which were based on the method of Borenfreund and Puerner (1985). The major difference is that most studies used a 24-hour test substance exposure duration for the 3T3 NRU test method, while the NICEATM/ECVAM 3T3 study used a 48-hour exposure duration. The major difference between the NHK protocols used in the reviewed studies and the protocol used in the NICEATM/ECVAM study is that the cell culture medium was changed at the time of test substance application in the NICEATM/ECVAM study.

9.1.3.1   *Draize Eye Irritation*
Triglia et al. (1989)
Four laboratories collaborated in an interlaboratory validation study to test the NHK NRU assay marketed by Clonetics® Corporation[2] for its intra- and inter-laboratory reproducibility and ability to predict *in vivo* ocular irritancy. Each laboratory tested 11 blind-coded surfactant-based substances and compared the IC_{50} values to *in vivo* Draize ocular irritancy scores.

The test exhibited the following performance characteristics for the comparison of *in vitro* and *in vivo* data:

- Specificity (percentage of non-irritants correctly detected) = 93%
- Sensitivity (percentage of true irritants correctly detected) = 80%
- Predictive values (probability that an unknown agent will be properly classified)
    - Positive predictive value = 90%
    - Negative predictive value = 87%

The authors reported that there was excellent correlation among the laboratories, and good correlation between the *in vitro* IC_{50} values and *in vivo* Draize scores (Spearman Rank correlation coefficients between *in vivo* and *in vitro* data for the laboratories ranged from 0.67-0.76). The authors also concluded that the NRU test could not replace the Draize test, but may be an effective screening tool for use in a battery of *in vitro* alternatives

Sina et al. (1995)
Sina et al (1995) evaluated the NHK NRU test method along with six other *in vitro* methods to determine whether they could be used as complimentary tests in a battery approach to estimate ocular irritation. The NRU data correlated poorly with Draize ocular scores for the 33 pharmaceutical intermediates tested. The Spearman correlation coefficient for the IC_{50} and maximum average Draize score (MAS) was -0.10, and the Pearson correlation coefficient was -0.04.

---

[2] Clonetics® Corporation sponsored this study. It was not clear in the publication if Clonetics® Corporation participated as one of the testing laboratories.

Brantom et al. (1997)
This study examined the potential of 10 alternative methods to predict the eye irritation potential of cosmetic ingredients. Four laboratories tested 55 coded substances (23 single ingredients and 32 formulations) using the 3T3 NRU test method, and used the resulting IC_{50} values to predict modified maximum average scores (MMAS) for the Draize test.

An endpoint was generated for each test by interpolation from a plot of percent cell survival versus test substance concentration. A prediction model was developed from data of 30 single ingredients (29 surfactants and one substance not classified by the authors) to equate the IC_{50} value to an MMAS.

The interlaboratory CV for the IC_{50} values was $37.3 \pm 29.8\%$ (7.5 ± 6.8, log transformed). Most of the mean IC_{50} values from a single laboratory differed by plus or minus an order of magnitude from the means of all the laboratories for each substance, which the authors interpreted as "no significant outliers". Correlations of NRU-predicted MMAS scores with *in vivo* MMAS scores yielded Pearson's r values ranging from 0.25 to 0.32 for the four laboratories.

Although the authors concluded the interlaboratory reproducibility was good, the IC_{50} values did not predict the MMAS. The r values for the *in vitro/in vivo* correlations were low (0.246 to 0.316) and the tests all underpredicted irritants and overpredicted non-irritants. Four substances were outside of the 95% confidence intervals and the authors concluded that the 3T3 NRU test method had wide applicability to test the remaining 51 coded substances according to the limitations in the prediction model, but that it was not effective as a stand-alone replacement for the Draize test across the entire irritation scale. The authors did not identify the test substances.

Harbell et al. (1997)
This publication reported the results of the evaluation of 12 *in vitro* cytotoxicity assays to predict ocular irritation. Data were voluntarily submitted to the U.S. Interagency Regulatory Alternatives Group (IRAG), composed of members from CPSC, EPA, and FDA. The NHK NRU test method was one of the tests evaluated by six laboratories testing surfactants and surfactant-containing formulations (the 3T3 NRU test method was not tested). Two laboratories submitted results for the same test substances, but the other four submitted data for various sets of substances and formulations.

The correlation of results from the two laboratories that independently tested the same substances was r=0.99. Correlations between the IC_{50} data and *in vivo* maximum average Draize score (MAS) ranged from -0.92 to -0.54. The IRAG concluded that the assays were suitable as a screening and adjunct assay to assess eye irritation over the range of toxicities found in personal care and household products, and recommended that its use be limited to water-soluble materials. Although the method was also evaluated for surfactants, IRAG recommended that the evaluation continue for its performance in predicting eye irritation for various product classes (e.g., fabric softeners, shampoos). In addition, the substance's physical form should be considered because the *in vitro* toxicity of a solution of the test substance will not necessarily predict toxicity of the parent, solid substance *in vivo*.

9.1.3.2   *Predicting Human Lethal Blood Concentrations (LC)*
Seibert et al. (1992)
This single laboratory study was designed to evaluate various aspects of cellular toxicity in four *in vitro* test systems for their relevance and reliability with respect to acute systemic toxicity, in particular, human LC. The 3T3 NRU test method was one of four methods evaluated with 10 MEIC substances.

The authors stated that final conclusions on the relevance of the *in vitro* systems for *in vivo* data could not be determined because the variations in LC were unknown so that limits for over or underprediction of human *in vivo* toxicity using experimental models could not be defined. In addition, the ability of *in vitro* toxicity to predict *in vivo* toxicity may depend on toxicokinetic factors that were not considered in the *in vitro* systems.

9.2   Independent Scientific Reviews

This section summarizes independent scientific reviews of the use of *in vitro* cytotoxicity methods for the prediction of rodent acute oral toxicity, and for the reduction of animal use in acute toxicity testing. The conclusions of these reviews are compared to the conclusions of the current study. Also discussed is the 3T3 NRU phototoxicity method, because it is similar to the 3T3 NRU test method used in the current validation study and has been validated by ECVAM and is the subject of OECD Test Guideline 432 (OECD 2004).

9.2.1   <u>*In Vitro* Acute Toxicity Testing for the Classification and Labelling of Chemicals</u>

9.2.1.1   *Seibert et al. (1996): ECVAM Workshop 16*
ECVAM sponsored a workshop in 1994 to review the current status of various *in vitro* methods and to determine their potential uses for reducing, refining, and/or replacing the use of laboratory animals for acute systemic toxicity testing. The workshop participants reviewed various types of toxicity, *in vitro* cytotoxicity testing schemes and strategies, inclusion of biokinetic parameters, biotransformation, biodistribution *in vitro* and *in vivo*, and a proposed acute toxicity testing scheme for the classification of substances.

The workshop participants agreed that some studies showed good correlations between *in vitro* cytotoxicity data and LD_{50} values. They also acknowledged that *in vitro* basal cytotoxicity tests could not address all the different of mechanisms of acute systemic toxicity. Additional approaches to replacing animals would have to incorporate the three main types of cellular level toxic effects that can lead to in acute systemic toxicity (i.e., basal cytotoxicity, selective toxicity, and cell-specific function toxicity). The participants determined that it is also important that any alternative method take into account the active concentration and meaningful dose of a test substance in an *in vitro* cell culture system. Quantitative comparisons of test substance concentrations must be made to evaluate the effects of the test substances regarding the three types of cytotoxicity.

The biokinetics of a test substance (determined by its absorption, distribution, metabolism, and elimination) must be considered when making predictions of *in vivo* toxicity using *in vitro* toxicity data. Various methods can be used to convert *in vitro* effective concentrations of a test substance to equivalent body doses. Test substance factors, such as physicochemical characteristics (e.g., pKa, lipophilicity, volatility), estimates of protein binding, and *in vitro*

characteristics (e.g., cell concentration, cell protein concentration, ratio of cell/medium volumes, medium albumin concentration), are needed for such conversions.

An *in vitro* tiered testing scheme was proposed by the workshop participants for using *in vitro* methods to determine the acute oral toxicity of a substance:

- Stage 1: Basal cytotoxicity test
- Stage 2: Hepatocyte-specific cytotoxicity test to assess the role of biotransformation in producing toxicity
- Stage 3: Test system that evaluates non-hepatocyte-specific selective cytotoxicity (i.e., effects on cell-specific functions)

This testing scheme was proposed as an approach to classify substances by their *in vitro* toxicity. The lowest IC_{50} value determined at any of the testing stages would be used to classify a substance (i.e., very toxic, toxic, harmful, and no label). The workshop participants recommended that a feasibility study be conducted to determine the practicability, relevance, and reliability of this tiered testing scheme. As noted in the NICEATM/ECVAM study (see **Section 6.4)**, the *in vitro* basal cytotoxicity tests are not suitable as replacements for rodent acute oral toxicity tests and could only be used as an adjunct test, and not a stand-alone test, for classifying substances for acute oral toxicity. However, *in vitro* tests could be used to identify starting doses for acute toxicity testing to reduce the number of animals used.

9.2.2   Use of *In Vitro* Cytotoxicity Data for Estimation of Starting Doses for Acute Oral Toxicity Testing

9.2.2.1   *ICCVAM (2001a): Estimation of Animal Savings Using Cytotoxicity Data with the ATC Method*

Participants at Workshop 2000 examined the influence of starting dose on animal use in the ATC method (ICCVAM 2001a; Section 2.2.3, pp.12-14; no testing was performed at the Workshop). The participants made inferences from the 1996 version of the ATC method that was based on the EU toxicity classification system (**Table 9-1**). The fixed doses for testing were 25, 200, and 2000 mg/kg. Normally, classification of a substance requires testing three animals in two to four dosing steps (i.e., six to 12 animals). The number of dosing steps increases with increasing distance between the true toxicity class and the starting dose. They estimated that one to three dosing steps could be avoided (i.e., three to nine animals saved) if the optimum starting dose could be predicted by *in vitro* cytotoxicity testing.

The predicted savings of one to three dosing steps was made under ideal conditions. The Workshop 2000 report (ICCVAM 2001a) provides a biometric analysis at a dose-mortality slope of 2.0 that shows that the greatest animal savings would occur for substances with very high and very low toxicity. Three animals are needed to classify a substance in the <25 mg/kg class if the true LD_{50} is 1 mg/kg and 25 mg/kg is the starting dose, but six animals are needed if the test starts from the default starting dose of 200 mg/kg (i.e., an animal savings of 50%). For a substance with a true LD_{50} of 10000 mg/kg, 11.3 animals on average are needed when the default starting dose is used, but only 7.7 animals would be needed at the 2000 mg/kg starting dose (i.e., an animal savings of 31%). For substances with a true LD_{50} of 2000 mg/kg, no animals would be saved by starting at the 2000 mg/kg dose (compared to starting at the default starting dose of 200 mg/kg).

Although these analyses were performed assuming the 1996 ATC method used starting doses of 25, 200, 2000 mg/kg, the Workshop 2000 participants noted that the animal savings that would be produced by improving the starting dose would not be significantly different for the current ATC method that uses GHS doses of 5, 50, 300, and 2000 mg/kg (or up to 5000 mg/kg) (OECD 2001c; see **Appendix M** for the current ATC test guideline). The Workshop 2000 participants did not predict the animal savings when *in vitro* cytotoxicity methods are used to estimate starting doses for the ATC, other than the biometric analysis described above.

The NICEATM/ECVAM study yielded patterns of animal savings with the ATC that were similar to those discussed at the 2000 Workshop (i.e., animal savings were greater for substances with a lower or higher LD_{50} than the default starting dose; see **Section 10.3.3.3**). Depending on the NRU test method and regression evaluated, the average animal savings per test (for the 67 or 68 reference substances evaluated) predicted by the NICEATM/ECVAM 7validation study at a dose-mortality slope of 2.0 were:

- 22.6 to 30.4 % (2.21 to 2.96 animals) for substances in the $LD_{50} \leq 5$ mg/kg category
- 10.2 to 13.0 % (1.17 to 1.51 animals) for substances in the $5 < LD_{50} \leq 50$ mg/kg category
- 3.8 to 4.3 % (0.42 to 0.47 animals) for substances in the $50 < LD_{50} \leq 300$ mg/kg category
- -9.5 to -6.1% (-0.93 to -0.60 animals) for substances in the $300 < LD_{50} \leq 2000$ mg/kg category
- -0.03 to 12.7% (-0.30 to 1.43 animals) for substances in the $2000 < LD_{50} \leq 5000$ mg/kg category
- 17.1 to 25.5% (2.03 to 3.02 animals) for substances with $LD_{50} > 5000$ mg/kg

The major differences between the evaluation reviewed by the Workshop 2000 participants and the NICEATM/ECVAM study were:

- The NICEATM/ECVAM study used the GHS toxicity categories for classification whereas the Workshop participants used the EU classification scheme, which has fewer toxicity categories. The accuracy of category prediction is higher with fewer categories.
- The NICEATM/ECVAM study used *in vitro* cytotoxicity data to estimate starting doses using two regressions based on the RC substances with rat LD_{50} data, whereas the Workshop 2000 participants used the fixed ATC doses as starting doses.
- The NICEATM/ECVAM study used computer simulations of ATC testing for individual substances whereas Workshop 2000 participants used an evaluation that estimated animal use based on fixed *in vivo* LD_{50} values and the fixed ATC doses.

9.2.2.2   *ICCVAM (2001a): Estimation of Animal Savings Using Cytotoxicity Data with the UDP*

Workshop 2000 participants examined the effect of starting dose on animal usage in the UDP assay by making inferences from the computer simulations of animal use shown in the peer-review BRD for the UDP (ICCVAM 2000). When the rule that requires testing to stop when four animals have been tested after the first reversal is used, and no other stopping rules are considered, the animal use is relatively insensitive to the slope of the dose-mortality curve. The number of animals required when the starting dose equals the true LD_{50} is approximately six. However, approximately nine animals are required when the starting dose is 1% of the true LD_{50}. Thus, animal use is 30% less when the starting dose is the true LD_{50} compared to a starting dose that is 1% of the true LD_{50} (ICCVAM 2001a, section 2.2.4, pg. 16). When UDP testing stops based on the likelihood-ratio stopping rule, the animal use depends principally on the slope of the dose-mortality curve. The Workshop 2000 participants estimated that 25 to 40% of the animals would be saved when the starting dose is equal to the true LD_{50}, compared to the savings at a starting dose 1% of the true LD_{50}.

According to the UDP BRD (ICCVAM 2000) used by the Workshop participants, UDP simulations at a mortality-response slope of 2.0 showed that an average of 12.4 animals per test were used when the starting dose was 1% of the true LD_{50}, but an average of 8.7 animals was used when the starting dose was the true LD_{50} (i.e., a 30% reduction). At a slope of 8.3, an average of 11 animals per test were used when the starting dose was 1% of the true LD_{50}, but an average of only six animals were used when the starting dose was the true LD_{50} (i.e., a 46% reduction). The animal savings predicted by Workshop 2000 participants was 25 to 40% based on starting at the true LD_{50} in comparison to starting at a dose that is 1% of the true LD_{50}.

Depending on the regression evaluated, the average animal savings predicted in the NICEATM/ECVAM validation study at dose-response slopes of 2.0 and 8.3 were 5.8 to 7.8% (0.49 to 0.66 animals) using the 3T3 (67 reference substances) and NHK (68 reference substances) NRU test methods (see **Section 10.2.3**). When averaged for the reference substances in each GHS category, the highest mean animal savings at a mortality-response slope of 2.0 was obtained for reference substances in the $2000 < LD_{50} \leq 5000$ mg/kg and $LD_{50} > 5000$ mg/kg categories. Animal savings were 11.3 to 16.7% (1.28 to 1.65 animals) using the 3T3 and NHK NRU test methods for the two regressions evaluated. The average animal savings for the substances in these categories at a dose-mortality slope of 8.3 were 12.1 to 21.0% (1.11 to 1.63 animals) for both methods and regressions. The major differences between the evaluation performed by the Workshop 2000 participants and the NICEATM/ECVAM study were that:

- The default starting dose used for the NICEATM/ECVAM simulations was 175 mg/kg (see **Section 10.2.2**), rather than 1% of the true LD_{50} assumed by the Workshop 2000 participants.
- The NRU IC_{50} was used in two regressions of *in vitro* data against *in vivo* data to estimate starting doses. This estimation was not always close to the true LD_{50}, which was the value used by the Workshop 2000 participants. For example, LD_{50} values predicted by the NICEATM/ECVAM study for phenylthiourea were approximately 540 mg/kg by the 3T3 IC_{50} and

approximately 904 mg/kg by the NHK IC_{50} using the RC rat-only millimole regression. The true *in vivo* LD_{50} for phenylthiourea is 3 mg/kg. Workshop 2000 participants used a best-case scenario when they assumed that *in vitro* cytotoxicity precisely predicted the true LD_{50}.

9.2.3   Validation of the 3T3 NRU Assay for Phototoxicity

An NRU assay using 3T3 cells was validated by ECVAM, and accepted for regulatory use, to detect the phototoxic potential of test substances. The 3T3 NRU test for phototoxicity requires a 60-minute exposure to the test substance, a 50-minute exposure to ultraviolet (UVA, 315-400 nm) light, followed by removal of test substance and incubation for another 24 hours in fresh medium (Spielmann et al. 1998). NR medium is then added, and NRU is measured after a 3-hour incubation. Phototoxic potential is assessed by comparing the differences in cytotoxicity between test plates containing the test substance that have not been exposed to UVA and comparable test plates exposed to UVA.

Two different models, employing the Photoinhibition Factor (PIF) and the Mean Photo Effect (MPE), were validated for the prediction of *in vivo* phototoxic potential. The accuracy of the models for classifying the phototoxic potential of the 30 substances tested in nine laboratories was 88% for the PIF, and 92% for the MPE, when compared with *in vivo* classifications. Interlaboratory variability for classification (i.e., phototoxic vs. non-phototoxic) was assessed using a bootstrapping approach. For each substance, the classification based on a single experiment was compared to the classification based on the mean PIF or mean MPE. The interlaboratory variability for classification was 0 to 18.8% using PIF and 0 to 20% using MPE.

The ECVAM Scientific Advisory Committee confirmed the scientific validity of the method in 1997 (ECVAM 1997) and its regulatory acceptance was noted in Annex V of Council Directive 67/548/EEC part B.41 on phototoxicity, in 2000. An OECD Test Guideline, 432, was finalized in 2004 (OECD 2004). The 3T3 NRU phototoxicity test is used in a tiered testing approach to determine the phototoxic potential of test substances.

The performance of the 3T3 NRU phototoxicity assay could not be compared with the performance of the 3T3 NRU test method used in this validation study because different classification schemes were used (i.e., a two-category classification for the phototoxicity vs. a six-class scheme for acute oral toxicity). The ECVAM measurements of interlaboratory variability also used different techniques and were not comparable to those used for the NICEATM/ECVAM study.

9.2.3.1   *NHK NRU Phototoxicity Assay*

FAL participated in the European Union/European Cosmetic, Toiletry and Perfumery Association (EU/COLIPA) study (30 substances tested using NHK and 3T3 cells) and the ECVAM/COLIPA study (20 substances tested using NHK cells) (Clothier et al. 1999). The studies showed that the NHK NRU test method could be used to predict phototoxic potential. The accuracy for predicting *in vivo* results was similar to that of the 3T3 NRU phototoxicity test (see **Table 9-4**). The NHK NRU phototoxicity test uses the same test substance exposure duration (approximately 2 hours) as the 3T3 NRU test method, but the duration of culture after UV exposure is 72 hours rather than 24 hours. NRU was measured after a 45-minute incubation with NR.

Although the NHK NRU phototoxicity method achieved good concordance with *in vivo* phototoxicity, it has not yet been validated for regulatory use.

Table 9-4   Correct Identification of *In Vivo* Phototoxicants by the NHK NRU Phototoxicity Assay

Study	3T3 NRU Phototoxicity Method	NHK NRU Phototoxicity Method
EU/COLIPA (Spielmann et al. 1998)	29/30 (97%)[1]	28/30 (93%)[1]
ECVAM/COLIPA	NA	18/20 (90%)[1] 19/20 (95%)[2]
Combined Study Data	45/45 (100%)[2]	44/45 (98%)[2]

Abbreviations: 3T3=BALB/c 3T3 fibroblasts; NHK=Normal human epidermal keratinocytes; NRU=Neutral red uptake; EU=European Union; ECVAM=European Centre for the Validation of Alternative Methods; COLIPA=The European Cosmetic Toiletry and Perfumery Association; NA=not available.
[1]Mean Photo Effect (MPF) prediction model.
[2]Photoinhibition Factor (PIF) prediction model.

9.3   Studies Using *In Vitro* Cytotoxicity Methods with Established Performance Standards

The procedure provided in the *Guidance Document* for evaluating basal cytotoxicity assays for use in predicting starting doses for acute oral toxicity assays provides the existing performance standards for the 3T3 and NHK NRU test methods (ICCVAM 2001b).

9.3.1   *Guidance Document* (ICCVAM 2001b)
In addition to guidance for evaluating *in vitro* basal cytotoxicity methods for use in predicting starting doses for rodent acute oral toxicity assays, the *Guidance Document* provided results from testing 11 reference substances using the recommended 3T3 and NHK NRU protocols (ICCVAM 2001b). The 11 substances were chosen from the RC database so as to have a close fit to the RC millimole regression and to cover a wide range of cytotoxicity. The major differences between the *Guidance Document* protocols and the protocols used in this validation study are the reduced NR concentrations (from 50 µg/mL to 25 µg/mL in the 3T3 NRU test method, and from 50 µg/mL to 33 µg/mL in the NHK NRU test method), the increased duration of test substance exposure in the 3T3 NRU test method, from 24 to 48 hours, and the lack of a refeeding step in the NHK NRU test method just prior to substance application (see **Sections 2.6** and **2.7** for further detail). Despite these differences, the *Guidance Document* shows that the test results for the 11 substances in both the 3T3 and NHK NRU test methods were similar to the results in the RC database. The calculated regressions for the 11 *Guidance Document* substances were:
- $\log LD_{50} = 0.506 \log IC_{50} + 0.475$ ($R^2=0.985$) for the 3T3 NRU test method
- $\log LD_{50} = 0.498 \log IC_{50} + 0.551$ ($R^2=0.936$) for the NHK NRU test method
- $\log LD_{50} = 0.435 \log IC_{50} + 0.625$ for the RC millimole regression

The 3T3 and NHK NRU regressions were compared with the RC millimole regression (347 substances) to show that the regression lines, as well as all 11 substance data points, were

within the acceptance interval (± 0.5 log around the regression) of the RC millimole regression (see *Guidance Document* Figures 3 and 4, p.13 [ICCVAM 2001b]).

9.3.2   King and Jones (2003)

This study also tested the 11 substances recommend in the *Guidance Document* using the recommended 3T3 NRU protocol. The IC_{50} - LD_{50} regression obtained was comparable to the RC millimole regression and to the 11 substance regression provided in the *Guidance Document* (ICCVAM 2001b). The regression was log LD_{50} = 0.552 log IC_{50} + 0.503 (R^2=0.929) and the RC millimole regression was log LD_{50} = 0.435 log IC_{50} + 0.625. The 11-substance regression fit within the acceptance interval (± 0.5 log) of the RC millimole regression.

King and Jones (2003) also showed that a 3T3 NRU test method that was adapted for high throughput testing by using three test sample concentrations yielded approximately the same IC_{50} as an eight concentration-response. A regression used to compare the IC_{50} values using the two different concentration-response approaches yielded R^2=0.945.

9.3.3   A-Cute-Tox Project: Optimization and Pre-Validation of an *In Vitro* Test Strategy for Predicting Human Acute Toxicity (Clemedson 2005)

The A-Cute-Tox Project is an Integrated Project under the EU 6th framework program that started in January 2005, with a termination date of January 2010. It was initiated in response to the REACH (Registration, Evaluation, Authorisation and Restriction of Chemicals) Directive and the 7th amendment of the Cosmetics Directive, which calls for the broad replacement of animal experiments for finished products by 2003, and for ingredients by 2009. The project is an extension of the NICEATM/ECVAM validation study and the EDIT program, which is the continuation of the MEIC program. The partnership is made up of the EDIT Consortium, ECVAM, and 35 other European toxicity research group partners.

The aim of the project is to develop a simple and robust *in vitro* testing strategy for prediction of human acute oral toxicity, which could replace the animal acute oral toxicity tests currently used for regulatory purposes. The objectives of A-Cute-Tox are:

- Compilation, critical evaluation, and generation of high quality *in vitro* and *in vivo* data for comparative analysis.
- Identifying factors (e.g., kinetics, metabolism, and organ specificity) that influence the correlation between *in vitro* toxicity (concentration) and *in vivo* toxicity (dosage), and to define an algorithm that accounts for these effects.
- Explore innovative tools and cellular systems to identify new toxicity end-points and strategies to better anticipate animal and human toxicity.
- To design a simple, robust and reliable *in vitro* test strategy associated with the prediction model for acute toxicity that is amenable to high-throughput testing.

The project has been divided into the following workpackages that will be implemented by various configurations of research partners:

- WP1: Generation of a "high quality" *in vivo* database (through literature searches and historical data) and establishment of a depository list of reference substances

- WP2: Generation of a "high quality" *in vitro* database (including data from the NICEATM/ECVAM study, EDIT studies, and MEIC studies)
- WP3: Iterative amendment of the testing strategy
- WP4: New end-points and new cell systems
- WP5: Alerts and correctors in toxicity screening (I): Role of absorption, distribution, and excretion
- WP6: Alerts and correctors in toxicity screening (II): Role of metabolism
- WP7: Alerts and correctors in toxicity screening (III): Role of target organ toxicity (i.e., neuro-, nephro-, hepato-toxicity)
- WP8: Technical optimisation of the amended test strategy
- WP9: Pre-validation of the test strategy

A-Cute-Tox aims to extend the NICEATM/ECVAM and MEIC/EDIT approaches toward a full replacement test strategy by improving the prediction of acute oral toxicity using *in vitro* methods, and then validating the testing procedure.

9.4   Summary

- *In vitro* NRU cytotoxicity test methods using various cell types have been evaluated for their correlation with rodent lethality endpoints (e.g., rat/mouse i.v., i.p., and oral toxicity). Peloux et al. (1992) and Fautrel et al. (1993) showed good correlations (r=0.877 and 0.88, respectively) of *in vitro* cytotoxicity with rodent i.p./i.v. and i.v. toxicity data, respectively.
- 3T3 and NHK NRU test methods have been evaluated for purposes other than the prediction of starting doses for acute toxicity studies (e.g., ocular irritancy; human LC values, *in vivo* phototoxicity).
- A 3T3 NRU test method has been validated by ECVAM for the identification of *in vivo* phototoxic potential.
- No *in vitro* test methods have been validated for the prediction of acute oral toxicity. Estimations of animal savings using *in vitro* cytotoxicity data to estimate starting doses for the UDP did not use actual *in vitro* cytotoxicity data. Instead, animal savings were estimated by assuming that the *in vivo* starting dose equals the true LD_{50}, which is an approach that assumes that cytotoxicity data can perfectly predict *in vivo* lethality. These theoretical predictions of animal savings in the UDP ranged from 25 to 40% (ICCVAM 2001a), as compared with the average animal savings of 5.3 to 7.8% predicted using computer simulation modeling of the UDP for the reference substances tested in the NICEATM/ECVAM study. Halle et al. (1997) used the *in vitro* cytotoxicity data in the RC to determine that an animal savings of 32% can be attained for the ATC method by using the LD_{50} predicted by the RC regression as the starting dose. For the reference substances tested in the NICEATM/ECVAM validation study, most of which were a poor fit to the RC millimole regression, the average animal savings for the ATC, as determined by computer simulation modeling, was 4.8 to 10.2%.

[This Page Intentionally Left Blank]

10.0   ANIMAL WELFARE CONSIDERATIONS (REFINEMENT, REDUCTION, AND REPLACEMENT)

As demonstrated in **Section 6**, *in vitro* basal cytotoxicity methods cannot be used as replacement assays[1] for rodent acute oral toxicity test methods for hazard classification. However, as described in this section, these methods can be used to reduce[2] and refine[3] animal use in the UDP or ATC acute oral toxicity assays, as shown by the computer simulations of such testing. Although the use of *in vitro* cytotoxicity data to determine starting doses for the FDP may reduce the use of animals for the FDP, even though death is not the primary endpoint, such an evaluation will not be provided in this document.

The test guidelines recommend using information on structurally-related substances and the results of any other toxicity tests (EPA 2002b) for the test substance, including *in vitro* cytotoxicity results, to approximate the LD_{50} and the slope of the dose-mortality curve (OECD 2001a; OECD 2001d; EPA 2002a). However, for the purposes of the reduction and refinement evaluation conducted in this section, it was assumed that no information other than 3T3 and NHK NRU IC_{50} data would be available. To determine the extent of animal reduction or refinement that would occur in the UDP and the ATC method when using a starting dose based on 3T3 or NHK NRU IC_{50} values rather than the default starting dose, computer models were used to simulate the *in vivo* testing of the reference substances used in the validation study.

Section 10.1 lists the regressions that were used with IC_{50} data from the 3T3 and NHK NRU test methods to determine starting doses for the UDP and the ATC. **Sections 10.2.1** and **10.3.1** summarize the animal testing procedures in the current test guidelines for the UDP and the ATC, respectively. The procedures for using computer simulation of the animal testing of the reference substances are described in **Sections 10.2.2** and **10.3.2**. The computer simulations were used to determine the numbers of animals used and the numbers of animals that "died" for each test. The modeling was performed using five different dose-mortality slopes[4] (i.e., 8.3, 4.0, 2.0, 0.8, and 0.5) because such slope information was not available for all of the reference substances used. To simplify the presentation of results, the animal use figures provided in **Sections 10.2.3**, **10.2.4**, **10.3.3**, and **10.3.4** include the data for only two of the slopes, 8.3 and 2.0. The slope of 2.0 is the default used for the calculation of LD_{50} by the UDP method (OECD 2001a; EPA 2002a) and the slope of 8.3 is shown to represent substances, such as pesticides, with higher slopes. The results for the other three slopes were calculated, and are provided in **Appendices N** and **Q**. The numbers of animals used are summarized to show the mean number of animals tested when the default starting dose is used and the mean number of animals used when the starting dose was determined from the 3T3 or NHK NRU IC_{50} values. The difference in animal use between the default starting doses and the IC_{50}-based starting doses is referred to as the animal savings. Differences were

---

[1] Replacement alternative: a new or modified test method that replaces animals with nonanimal systems or one animal species with a phylogenetically lower one (e.g., a mammal with an invertebrate).
[2] Reduction alternative: a new or modified test method that reduces the number of animals required.
[3] Refinement alternative: a new or modified test method that refines procedures to lessen or eliminate pain or distress in animals or enhances animal well-being.
[4] The dose-mortality slope is the slope of the dose-response curve for mortality.

tested for statistical significance (at p <0.05) using a one-sided Wilcoxon signed ranked test based on the number of substances evaluated. **Sections 10.2** and **10.3** summarize mean animal use by the total number of substances tested and by the number of substances in each GHS category. **Sections 10.2.4** and **10.3.4** provide the mean number of animal deaths compared to the mean number of animals used for each default and IC_{50}-based starting dose to determine whether the IC_{50}-based starting doses lead to a reduction in the number of animals used and the number that die (i.e., refinement). **Sections 10.2.5** and **10.3.5** discuss concordance for the reference substance outcomes of simulated testing using the IC_{50}-based starting doses, with the outcomes of the default starting doses. Sections 10.4 and 10.5 discuss the impact of accuracy and the impact of prevalence (i.e., the number of substances to be tested in each GHS category) on animal savings.

10.1   Use of the 3T3 and NHK NRU Test Methods to Predict Starting Doses for Rodent Acute Oral Toxicity Assays

The IC_{50} values developed from the 3T3 and NHK NRU test methods were used to predict starting doses for rodent acute oral toxicity tests using the following linear regressions of IC_{50}-LD_{50} values (from **Section 6.3**):

- The RC rat-only millimole regression: $\log LD_{50}$ (mmol/kg) = 0.439 log IC_{50} (mM) + 0.621
- The RC rat-only weight regression: $\log LD_{50}$ (mg/kg) = 0.372 log IC_{50} (μg/mL) + 2.024

The IC_{50} values from each *in vitro* NRU test method were evaluated with each regression and simulated acute oral toxicity test method,. The criteria for the use of a reference substance for this evaluation were that it must have:

- Replicate IC_{50} values from at least one laboratory
- A rat acute oral LD_{50} reference value (from **Table 4-2**)

Sixty-seven and 68 reference substances were evaluated for the 3T3 and the NHK NRU test methods, respectively. Of the 72 reference substances tested, epinephrine bitartrate, colchicine, and propylparaben were excluded because they did not have associated rat oral LD_{50} data. Carbon tetrachloride and methanol were excluded from the 3T3 evaluations, and carbon tetrachloride was excluded from the NHK evaluations, because none of the laboratories achieved sufficient toxicity in any test for the calculation of an $IC_{50.}$

10.2   Reduction and Refinement of Animal Use for the UDP

10.2.1   *In Vivo* Testing Using the UDP
This section describes the general dosing procedure for the UDP (OECD 2001a; EPA 2002a). Although doses, interval between doses, and dose progression, may be adjusted as necessary, the procedures described reflect the default guidance. Guidance on the types of animals that can be used, animal housing, clinical observations, etc., are outside the scope of the current discussion and are provided in the test guidelines (see **Appendices M1** and **M2**).

10.2.1.1   *Main Test*
The UDP is based on a staircase design in which single animals are dosed, in sequence, at 48-hour intervals. The effect on the first animal determines the dose of the next animal. If the first animal dies or is in a moribund state within 48 hours after dosing, the dose administered

to the next animal is lowered by dividing the original dose by one-half log (i.e., 3.2, which is the default dose progression). If the first animal survives, the dose administered to the next animal is increased by one-half log times the original dose. A dose progression of one-half log unit corresponds to a dose-mortality slope of 2.0. The default dose progression can be adjusted if the analyst has prior information upon which to estimate a slope.

The starting dose recommended by the guideline is one dose progression step below the analyst's best estimate of the LD_{50}, because, in the UDP test method, the LD_{50} estimate tends to move toward the starting dose. A default starting dose of 175 mg/kg is used if there is no information on which to base a starting dose. The default dosing scheme, using the dose progression of 3.2, is 1.75, 5.5, 17.5, 55, 175, 550, 1750, and 5000 mg/kg (EPA 2002a) or 1.75, 5.5, 17.5, 55, 175, 550, and 2000 mg/kg (OECD 2001a). The difference between the two reflects the different maximum doses emphasized in the different guidelines. Dosing single animals, upward or downward, in sequence proceeds until the first of three conditions, referred to as stopping rules, is met:

- Three consecutive animals survive at the upper dose limit (2000 or 5000 mg/kg)
- Five reversals[5] occur in any six consecutive animals tested
- Four or more animals have followed the first reversal, and the likelihood-ratios specified by the guideline exceed the critical value. For a wide variety of LD_{50} values and dose-mortality slopes, this rule is satisfied with four to six animals after the first reversal. Three likelihood values are calculated: a likelihood at an LD_{50} point estimate (called the rough estimate or dose-averaging estimate); a likelihood at a value below the point estimate (the point estimate divided by 2.5); and a likelihood at a value above the point estimate (the point estimate multiplied by 2.5). The ratios of the likelihoods are examined to determine whether they exceed a critical value.

If none of these conditions is met, the dosing stops after 15 animals have been used.

10.2.1.2 *Limit Test*
The UDP guidelines include a limit test using three to five animals dosed sequentially at 2000 mg/kg (OECD 2001a) or 5000 mg/kg (EPA 2002a). The EPA guideline for testing at a limit dose calls for proceeding to the main test if the first animal dosed at 5000 mg/kg dies (EPA 2002a). If the first animal lives, two more animals are dosed, in sequence, with 5000 mg/kg. If both animals live, then testing is terminated, and the substance is designated as having an LD_{50} >5000 mg/kg. If one or both animals die, then two more animals are dosed in sequence. As soon as a total of three animals survive, the test is terminated, with the conclusion that LD_{50} >5000 mg/kg. However, the main test is conducted if three animals die.

---

[5] Reversal: a situation where a nonresponse (i.e., animal lives) is observed at some dose, and a response is observed at the next dose tested (i.e. animal dies), or vice versa. Reversal is created by a pair of responses. (See **Appendices M1** and **M2**)

The OECD guideline for testing at a limit dose calls for proceeding to the main test if the first animal dosed at 2000 mg/kg dies (OECD 2001a). If the animal lives, four more animals are sequentially dosed. The main test is performed if three animals die. If three or more animals survive, testing is terminated with the conclusion that the LD_{50} >2000 mg/kg.

10.2.2   Computer Simulation Modeling of the UDP

Ten thousand UDP testing simulations were run for each substance, *in vitro* NRU test method, and dose-mortality slope. Because the analysis assumed there was no information upon which to estimate a dose-mortality slope, the modeling used the default dose progression factor of 3.2, and 5000 mg/kg as the upper limit dose because this upper limit is emphasized in the EPA guideline (EPA 2002a)[6]. If the starting dose estimated from the *in vitro* IC_{50} value was ≥4000 mg/kg, then the limit test, rather than the main test, was performed. If, during the dose progression, the next highest dose to be administered was approximately 4000 mg/kg or greater, then the limit dose of 5000 mg/kg was administered. If a dose one step below the IC_{50}-estimated LD_{50} was used as the starting dose, the other doses administered corresponded to the default doses specified in the guidelines (OECD 2001a; EPA 2002a). The simulation modeling procedures also used a lower limit of 1 mg/kg. Thus, a dose of 1 mg/kg was administered if the dose progression fell below 1 mg/kg. To estimate animal use by the default method, a starting dose of 175 mg/kg was used; the other doses administered after the default starting dose corresponded to the doses specified in the guidelines (OECD 2001a; EPA 2002a).

The simulation was performed using SAS® version 8 (SAS 1999) and implemented the distributional assumptions underlying the dose-mortality relationship. The lowest dose at which an animal dies in response to the administration of a toxic substance varies from animal to animal. For an entire population of animals, mortality is assumed to have a log-normal distribution, with the mean equal to the log of the true LD_{50}. Sigma (σ), the variability of the simulated population, is the inverse of the slope of the dose-mortality curve. Because of a lack of information concerning the actual dose-mortality curves, the simulations assumed several different values of the slope, but no corresponding changes were made in the dose progression. Dose-mortality slopes of 0.5, 0.8, 2.0, 4.0, and 8.3 were used because these were used in the simulation modeling used to evaluate the current version of the UDP guidelines (ICCVAM 2001c).

To model the variability of the IC_{50} values within and among laboratories, the values for each reference substance were log-transformed to normalize their distribution. The mean and variance of these log-transformed values were used to generate a log-normal distribution from which an IC_{50} value was randomly selected. This IC_{50} value was used with the regressions to determine starting doses using two different methods. One method used the LD_{50} estimated from the IC_{50} and the regression as the starting dose, while the other used the closest default dose that was lower than the estimated LD_{50}. The latter method is recommended by the EPA and OECD test guidelines (EPA 2002a; OECD 2001a), and the results from that simulation are presented in **Section 10.2**. The UDP is only usable for regulatory purposes if the starting dose is set below the expected LD_{50}. **Appendix Q** contains

---

[6] The results from UDP simulations for a limit dose of 2000 mg/kg will be presented in a future addendum to this document.

the results obtained when the LD$_{50}$ that was estimated by the IC$_{50}$ and the regression was used as the starting dose.

The simulation procedure used the following steps for each reference substance:
1. The LD$_{50}$ value (in mg/kg) from **Table 4-2** was entered as the true LD$_{50}$ value and the choices of assumed slope were entered as the true slopes for the dose-mortality curve.
2. An IC$_{50}$ value was selected from a distribution identified by the mean and variance of the IC$_{50}$ values for each chemical to reflect the variation in IC$_{50}$ values produced by the different laboratories (see **Tables 5-4** and **5-5** for mean IC$_{50}$ values and standard deviations for the 3T3 and NHK NRU test methods, respectively).
3. The IC$_{50}$ value from Step 2 was used in the regression model being evaluated to predict a LD$_{50}$ value, which was used to determine the starting dose.
4. The dosing simulation was run three times: once with the default starting dose of 175 mg/kg, once at the next default dose below the LD$_{50}$ estimated by the regression being evaluated, and once at a dose equal to that of the LD$_{50}$ estimated by the regression being evaluated.
5. For each simulated trial, the animals are dosed sequentially; therefore for each animal (*i*) there is a corresponding dose (*i*) that is administered to the animal. For the first animal in each trial, it is the starting dose for that trial. For each subsequent animal, the dose is dependent on the previous dose and the previous animal's response, as described in **Section 10.2.1**. For animal (*i*), the probability of a response is computed with the cumulative log-normal distribution at the dose administered. That is,

    $$P(response) = P(x < \log[dose(i)]) \text{ where } x \sim N(\mu, \sigma),$$

    where μ is the log of the true LD$_{50}$ value, and σ is the inverse of the assumed slope of the dose-mortality curve. One observation is then sampled from a binomial distribution with this calculated probability of success to determine whether the animal lives or dies.
6. Dosing simulation is stopped as soon as one of the stopping rules is satisfied.

Steps 2-6 were repeated 10,000 times in order to compute an average animal use for each method evaluated.

10.2.3   Animal Savings in the UDP When Using 3T3- and NHK-Based Starting Doses

10.2.3.1 *The Effect of the Dose-Mortality Slope on Animal Use*
As described in **Section 10.2.2**, the simulation modeling of animal use for the UDP assumed five different dose-mortality slopes in order to assess animal use under various conditions of population variability. **Table 10-1** shows that the number of animals used for the UDP decreases with increasing slope for both the default starting dose and the IC$_{50}$-determined starting dose when based on the RC rat-only millimole regression. The IC$_{50}$-determined starting dose was the next default dose lower than the regression-estimated LD$_{50}$. For example, because the LD$_{50}$ predicted for cadmium chloride by the 3T3 NRU IC$_{50}$ with the RC rat-only millimole regression was 16 mg/kg, the starting dose was 1.75 mg/kg (i.e., the next default dose below the predicted LD$_{50}$). This approach is consistent with the UDP

guidelines (OECD 2001a; EPA 2002a) as a means of reducing the number of animals that might experience pain and suffering from a treatment. This approach also overcomes the nonconservative bias of the UDP, which tends to yield an LD_{50} close to the starting dose.

Table 10-1 shows that, for each dose-mortality slope, the mean number of animals saved was statistically significant (p<0.05) when compared to mean number of animals needed when the default starting dose was used. When expressed as a percentage of the number of animals used when the default starting dose is used, animal savings also generally increased with increasing slope of the dose-response. The animal savings is the same at all slopes tested, but fewer animals are used at the steeper slopes, which increases the relative percentages of animals saved.

Table 10-1   Change in Animal Use[1] with Dose-Mortality Slope for the UDP[2]

Dose-Mortality Slope	With Default Starting Dose[1,3]	With IC_{50}-Based Starting Dose[1,4]	Animals Saved[5]
3T3 NRU Test Method			
0.5	10.01 ±0.10	9.48 ±0.11	0.53* (5.3%)
0.8	9.95 ±0.13	9.34 ±0.14	0.61* (6.1%)
2.0	9.35 ±0.16	8.80 ±0.17	0.54* (5.8%)
4.0	8.68±0.18	8.15 ±0.19	0.52* (6.0%)
8.3	7.95 ±0.18	7.42 ±0.20	0.53* (6.6%)
NHK NRU Test Method			
0.5	10.01 ±0.09	9.53 ±0.12	0.49* (4.9%)
0.8	9.96 ±0.13	9.41 ±0.15	0.55* (5.5%)
2.0	9.36 ±0.16	8.86 ±0.18	0.50* (5.3%)
4.0	8.66 ±0.17	8.18 ±0.20	0.48* (5.6%)
8.3	7.92 ±0.18	7.43 ±0.20	0.49* (6.2%)

Abbreviations: UDP=Up-and-Down Procedure; 3T3=BALB/c 3T3 fibroblasts; NHK=Normal human epidermal keratinocytes; RC=Registry of Cytotoxicity.
*Statistically significant (p<0.05) by a one-sided Wilcoxon signed rank test. Percentage difference is shown in parentheses.
[1]Mean numbers of animals ±standard errors for 10,000 simulations for each of the 67 substances in the 3T3 NRU test method and 68 substances in the NHK NRU test method. Although the simulations used whole animals, averaging the results over a large number of simulations produced fractional numbers. Upper limit dose =5000 mg/kg.
[2]OECD (2001a); EPA (2002a).
[3]Default starting dose = 175 mg/kg.
[4]The starting dose = next lower default dose to the predicted LD_{50}, which was calculated from the IC_{50} value in the RC rat-only millimole regression: $\log LD_{50}$ (mmol/kg) = 0.439 $\log IC_{50}$ (mM) + 0.621. The IC_{50} value for each reference substance was randomly selected from the distribution of values obtained during the testing with each method.
[5]Difference between mean animal use with the default starting dose and mean animal use with the IC_{50}-based starting dose.

To simplify the presentation of animal savings and the comparison of the various regressions and starting doses, the results of subsequent analyses presented in **Section 10.2.3** are limited to the dose-mortality slopes of 2.0 and 8.3. The slope of 2.0 is the default used for the calculation of LD_{50} by the UDP method (OECD 2001a; EPA 2002a) and the slope of 8.3 is shown to represent substances, such as pesticides, with higher slopes. Animal savings results for the other dose-mortality slopes were calculated, and are presented in **Appendices N1-N3.** Although using the next lower default dose to the *in vitro*-determined LD_{50} value overcomes

the bias of the UDP toward the starting dose (OECD 2001a, EPA 2002a) and is the appropriate approach for regulatory use, animal savings results using the estimated LD_{50} as the starting dose were also calculated (see **Appendix Q**).

10.2.3.2   *Mean Animal Use for UDP Simulations – Comparison of Regressions and
            Predictions from the 3T3 and NHK NRU Test Methods*
Table 10-2 shows the mean animal use for the simulated UDP testing of the reference substances described in **Section 10.1**. Mean animal use is shown for the default starting dose and for starting doses that were one default dose lower than the LD_{50} predicted from the *in vitro* NRU methods and the regressions evaluated in **Section 6.4** for the prediction of GHS category. The difference in animal use between the two starting doses is the mean animal savings produced by using the starting dose based on the *in vitro* NRU methods. All differences (i.e., mean animal savings) were statistically significant ($p < 0.05$) by a one-sided Wilcoxon signed rank test. Mean animal savings ranged from 0.49 to 0.66 (6.2% to 7.0%) animals per test depending upon the *in vitro* NRU test method, regression, and dose-mortality slope. The lowest mean animal savings were obtained for the RC rat-only millimole regression (0.49 [6.2%] to 0.54 [5.8%] animals for the different test methods and dose-mortality slopes), and the greatest mean animal savings were obtained with the RC rat-only weight regression (0.54 [6.8%] to 0.66 [7.0%] animals per test).

The animal savings using the *in vitro* NRU test methods with the RC rat-only regressions apply only to the reference substances evaluated in this validation study, and are based on substances pre-selected for their known *in vivo* toxicities and may not be broadly applicable to other substances. **Table 3-4** shows that 22 (38%) of the 58 RC substances selected for testing were known to have a poor fit to the RC millimole regression (i.e., the *in vivo* LD_{50} was outside the RC acceptance interval for the predicted LD_{50}). **Table 6-3** shows that 40% (28/70 for the 3T3) and 44% (31/71 for the NHK) of the reference substances that produced IC_{50} values were outliers. The RC rat-only millimole regression evaluated here is very similar to the RC millimole regression (see **Table 6-5**). Substances with better fits to the regression are more likely to yield greater animal savings.

10.2.3.3   *Animal Savings in the UDP by GHS Acute Oral Toxicity Category Using 3T3- and
            NHK-Based Starting Doses*
Tables 10-3 and **10-4** show mean animal use and mean animal savings for the UDP when the default starting dose and the IC_{50}-predicted starting doses were used, and when the reference substances are grouped by GHS category (UN 2005). The data come from the same analyses as the data provided in **Table 10-2**. The IC_{50}-predicted starting doses were based on the:
- RC rat-only millimole regression (**Table 10-3**)
- RC rat-only weight regression (**Table 10-4**)

Table 10-2   Mean Animal Use[1] in the UDP[2] Using Starting Doses Based on the 3T3 and NHK NRU Test Methods with the Different Regressions

Assay/Regression	With Default Starting Dose[3]	With IC_{50}-Based Starting Dose[4]	Animals Saved[5]	With Default Starting Dose[3]	With IC_{50}-Based Starting Dose[5]	Animals Saved[5]
3T3 NRU Test Method	**Dose-mortality Slope = 2.0**			**Dose-mortality Slope = 8.3**		
RC rat-only millimole[6]	9.35 ±0.16	8.80 ±0.17	0.54* (5.8%)	7.95 ±0.18	7.42 ±0.20	0.53* (6.6%)
RC rat-only weight[7]	9.36 ±0.16	8.70 ±0.16	0.66* (7.0%)	7.94 ±0.18	7.32 ±0.19	0.62* (7.8%)
NHK NRU Test Method	**Dose-mortality Slope = 2.0**			**Dose-mortality Slope = 8.3**		
RC rat-only millimole[6]	9.36 ±0.16	8.86 ±0.18	0.50* (5.3%)	7.92 ±0.18	7.43 ±0.20	0.49* (6.2%)
RC rat-only weight[7]	9.36 ±0.16	8.80 ±0.17	0.56* (6.0%)	7.92 ±0.18	7.38 ±0.20	0.54* (6.8%)

Abbreviations: 3T3=BALB/c 3T3 fibroblasts; NHK=Normal human epidermal keratinocytes; RC=Registry of Cytotoxicity; UDP=Up-and-Down Procedure.
*Statistically significant (p <0.05) by a one-sided Wilcoxon signed rank test. Percentage difference is shown in parentheses.
[1]Mean numbers of animals ±standard errors for 10,000 simulations for each of the 67 substances in the 3T3 NRU test method and 68 substances in the NHK NRU test method. Although the simulations used whole animals, averaging the results over a large number of simulations produced fractional numbers. Upper limit dose =5000 mg/kg.
[2]OECD (2001a); EPA (2002a).
[3]Default starting dose =175 mg/kg.
[4]The starting dose = one default dose lower than the predicted acute oral LD_{50} calculated using the IC_{50} value in the specified regression. The IC_{50} value for each reference substance was randomly selected from the distribution of values obtained during the *in vitro* testing with each test method.
[5]Difference between mean animal use with default starting dose and mean animal use with the IC_{50}-based starting dose.
[6]log LD_{50} (mmol/kg) = 0.439 log IC_{50} (mM) + 0.621.
[7]log LD_{50} (mg/kg) = 0.372 log IC_{50} (μg/mL) + 2.024.

These analyses showed that:
- For each *in vitro* NRU test method and regression, animal savings were statistically significant for substances in the $2000 < LD_{50} \leq 5000$ mg/kg and $LD_{50} > 5000$ mg/kg toxicity categories.
- For substances with $5 < LD_{50} \leq 50$ mg/kg and $50 < LD_{50} \leq 300$ mg/kg, both *in vitro* NRU test methods with each regression used slightly more animals than the default-starting dose, but the differences were not statistically significant.

Animal Savings for the UDP by GHS Acute Oral Toxicity Category Using 3T3- and NHK-Based Starting Doses with the RC Rat-Only Millimole Regression
Table 10-3 shows the animal savings by GHS category when the IC_{50} values are used with the RC rat-only millimole regression. Mean animal savings were statistically significant (p <0.05) by a one-tailed Wilcoxon signed rank test for the following GHS toxicity categories, test methods, and dose-mortality slopes:
- The use of the NHK NRU test method at both dose-mortality slopes for substances with $300 < LD_{50} \leq 2000$ mg/kg that produced savings of 0.49 (6.5%) to 0.52 (6.1%) animals per test.
- The use of the 3T3 NRU test method at the 8.3 dose-mortality slope for substances with $300 < LD_{50} \leq 2000$ mg/kg that produced a saving of 0.31 (4.1%) animals per test.
- The use of both *in vitro* NRU test methods at both dose-mortality slopes for substances with $2000 < LD_{50} \leq 5000$ mg/kg that produced savings of 1.11 (12.1%) to 1.28 (11.9%) animals per test.
- The use of both *in vitro* NRU test methods and both dose-mortality slopes for substances with an $LD_{50} > 5000$ mg/kg that produced savings of 1.47 (14.8%) to 1.58 (20.3%) animals per test.

The mean animal savings for the 3T3 and NHK NRU test methods were similar for most toxicity categories at both dose-mortality slopes, with the mean savings with the 3T3 slightly higher than with the NHK. For the dose-mortality slope of 2.0, the mean animal savings with the 3T3 NRU test method ranged from -0.42 (-5.5%) to 1.58 (16.0%) animals per test for the various toxicity categories, and savings for the NHK NRU test method ranged from -0.34 (-3.5%) to 1.47 (14.8%) animals per test. For the dose-mortality slope of 8.3, animal savings for the 3T3 NRU test method ranged from -0.29 (-4.3%) to 1.58 (20.3%) animals per test and savings for the NHK NRU test method ranged from -0.33 (-3.9%) to 1.47 (19.2%) animals per test. Animal savings were also obtained for highly toxic substances ($LD_{50} \leq 5$ mg/kg) with both the 3T3 (0.96 [9.9%] to 1.14 [10.0%] animals per test) and NHK (0.71 [7.3%] to 0.75 [6.7%] animals per test) NRU test methods, but the savings were not statistically significant.

No mean animal savings (≤ -0.28 animal per test) were observed for substances with $50 < LD_{50} \leq 300$ mg/kg by either the 3T3 or the NHK NRU test method. This category includes the default starting dose of 175 mg/kg. Animal savings were not expected for this category because savings were determined by comparing animal use with the IC_{50}-based starting dose with animal use at the default starting dose. No animal savings (-0.07 to -0.34 animals per test) were observed for substances with $5 < LD_{50} \leq 50$ mg/kg for either NRU test method. None of these differences in animal use was statistically significant.

Table 10-3   Animal Use[1] for the UDP[2] by GHS Acute Oral Toxicity Category[3] Using Starting Doses Based on the 3T3 and NHK NRU Test Methods with the RC Rat-Only Millimole Regression[4]

GHS Acute Oral Toxicity Category[3]	Number of Reference Substances	Dose-mortality Slope = 2.0			Dose-mortality Slope = 8.3		
		With Default Starting Dose[5]	With IC$_{50}$-Based Starting Dose[6]	Animals Saved[7]	With Default Starting Dose[5]	With IC$_{50}$-Based Starting Dose[6]	Animals Saved[7]
		3T3 NRU Test Method					
LD$_{50}$ ≤5 mg/kg	6	11.32 ±0.20	10.19 ±0.70	1.14 (10.0%)	9.70 ±0.28	8.74 ±0.43	0.96 (9.9%)
5 < LD$_{50}$ ≤50 mg/kg	11	9.68 ±0.23	9.74 ±0.45	-0.07 (-0.7%)	8.46 ±0.28	8.54 ±0.47	-0.08 (-1.0%)
50 < LD$_{50}$ ≤300 mg/kg	12	7.76 ±0.10	8.18 ±0.21	-0.42 (-5.5%)	6.61 ±0.19	6.90 ±0.19	-0.29 (-4.3%)
300 < LD$_{50}$ ≤2000 mg/kg	16	8.53 ±0.21	8.14 ±0.21	0.38 (4.5%)	7.46 ±0.24	7.15 ±0.19	0.31* (4.1%)
2000 < LD$_{50}$ ≤5000 mg/kg	10	10.73 ±0.10	9.46 ±0.15	1.28* (11.9%)	9.17 ±0.23	7.96 ±0.31	1.21* (13.2%)
LD$_{50}$ >5000 mg/kg	12	9.87 ±0.34	8.29 ±0.49	1.58* (16.0%)	7.76 ±0.59	6.18 ±0.69	1.58* (20.3%)
		NHK NRU Test Method					
LD$_{50}$ ≤5 mg/kg	6	11.21 ±0.24	10.47 ±0.71	0.75 (6.7%)	9.66 ±0.27	8.95 ±0.52	0.71 (7.3%)
5 < LD$_{50}$ ≤50 mg/kg	11	9.65 ±0.16	9.99 ±0. 45	-0.34 (-3.5%)	8.43 ±0.26	8.77 ±0.49	-0.33 (-3.9%)
50 < LD$_{50}$ ≤300 mg/kg	12	7.78 ±0.11	8.12 ±0.21	-0.34 (-4.4%)	6.57 ±0.19	6.85 ±0.19	-0.28 (-4.2%)
300 < LD$_{50}$ ≤2000 mg/kg	16	8.55 ±0.22	8.03 ±0.23	0.52* (6.1%)	7.49 ±0.25	7.00 ±0.20	0.49* (6.5%)
2000 < LD$_{50}$ ≤5000 mg/kg	10	10.75 ±0.08	9.54 ±0.20	1.21* (11.3%)	9.17 ±0.23	8.06 ±0.29	1.11* (12.1%)
LD$_{50}$ >5000 mg/kg	13	9.87 ±0.32	8.41 ±0.44	1.47* (14.8%)	7.66 ±0.59	6.18 ±0.69	1.47* (19.2%)

Abbreviations: 3T3=BALB/c 3T3 fibroblasts; GHS=Globally Harmonized System of Classification and Labelling of Chemicals (UN 2005); NHK=Normal human epidermal keratinocytes; RC=Registry of Cytotoxicity; UDP=Up-and-Down Procedure.
*Statistically significant (p<0.05) by a one-sided Wilcoxon signed rank test. Percentage difference shown in parentheses.
[1]Mean numbers of animals used ±standard errors for 10,000 simulations for each substance with an upper limit dose of 5000 mg/kg. Although the simulations used whole animals, averaging the results over a large number of simulations produced fractional numbers. Results are provided for 67 substances in the 3T3 NRU test method and 68 substances in the NHK NRU test method. Substances were categorized using the rat acute oral LD$_{50}$ reference values in mg/kg from **Table 4-2**.
[2]OECD (2001a); EPA (2002a).
[3]UN (2005).
[4]The RC rat-only millimole regression is log LD$_{50}$ (mmol/kg) = 0.439 log IC$_{50}$ (mM) + 0.621.
[5]Default starting dose = 175 mg/kg.
[6]The starting dose was one default dose lower than the predicted LD$_{50}$ calculated using the IC$_{50}$ value for each reference substance in the RC rat-only millimole regression. The IC$_{50}$ value for each reference substance was randomly selected from the distribution of values obtained during the testing with each method.
[7]Difference between mean animal use with the default starting dose and mean animal use with the predicted starting dose.

The animal savings from the future use of these *in vitro* NRU test methods with the RC rat-only millimole regression will depend on the proportion of test substances that will fall into each of the GHS categories.

Animal Savings for the UDP by GHS Category Using 3T3- and NHK-Based Starting Doses with the RC Rat-Only Weight Regression

Table 10-4 shows the mean animal savings by GHS acute oral toxicity category when the IC_{50} values are used with the RC rat-only weight regression. A comparison of mean animal savings, by category, with the RC rat-only millimole regression, indicates that, in most cases, animal savings were slightly higher for the RC rat-only weight regression than for the millimole regression. In the RC rat-only weight regression, the mean differences between animal use for the default starting dose and mean animal use with the IC_{50}-determined starting dose were statistically significant ($p < 0.05$) by a one-sided Wilcoxon signed rank test for the following GHS categories, NRU test methods, and dose-mortality slopes:

- The use of the 3T3 NRU test method at the 8.3 mortality-slope for substances with $300 < LD_{50} \leq 2000$ mg/kg that produced a savings of 0.28 (3.8%) animals per test.
- The use of both *in vitro* NRU test methods at both dose mortality slopes for substances with $2000 < LD_{50} \leq 5000$ mg/kg that produced savings of 1.28 (14.0%) to 1.64 (15.2%) animals per test.
- The use of both *in vitro* NRU test methods at both dose-mortality slopes for substances with $LD_{50} > 5000$ mg/kg that produced savings of 1.53 (20.0%) to 1.65 (16.7%) animals per test.

For the dose-mortality slope of 2.0, the mean animal savings (for the various GHS categories) with the 3T3 NRU test method ranged from -0.25 (-3.3%) to 1.65 (16.7%) animals per test, and from -0.24 (-3.1%) to 1.54 (15.6%) animals per test using the NHK NRU test method. At the dose-mortality slope of 8.3, animal savings with the 3T3 NRU test method ranged from -0.18 (-2.7%) to 1.63 (21.0%) animals per test, and savings for the NHK NRU test method ranged from -0.18 (-2.7%) to 1.53 (20.0%) animals per test. Animal savings were also obtained for highly toxic substances ($LD_{50} \leq 5$ mg/kg) with both the 3T3 (0.78 [8.0%] to 0.90 [8.0%] animals per test) and NHK (0.69 [7.1%] to 0.72 [6.4%] animals per test) NRU test methods, but these savings were not statistically significant.

There were no mean animal savings (≤ -0.18 animals per test) for substances with $50 < LD_{50} \leq 300$ mg/kg with either *in vitro* NRU test method. This category includes the default starting dose of 175 mg/kg. Animal savings were not expected for this category because savings were determined by comparing animal use at the IC_{50}-based starting dose with animal use at the default starting dose. For the NHK NRU test method, there were no animal savings (-0.07 to -0.13 animals per test) when used for substances with $5 < LD_{50} \leq 50$ mg/kg. None of these small changes in animal use were statistically significant.

The animal savings from testing new substances with these *in vitro* NRU test methods using the RC rat-only weight regression will depend on the proportion of test substances that fall into each of the GHS categories.

Table 10-4   Animal Use[1] for the UDP[2] by GHS Acute Oral Toxicity Category[3] Using Starting Doses Based on the 3T3 and NHK NRU Test Methods with the RC Rat-Only Weight Regression[4]

GHS Acute Oral Toxicity Category[3]	Number of Reference Substances	Dose-mortality Slope = 2.0			Dose-mortality Slope = 8.3		
		With Default Starting Dose[5]	With IC_{50}-Based Starting Dose	Animals Saved[7]	With Default Starting Dose[5]	With IC_{50}-Based Starting Dose	Animals Saved[7]
		3T3 NRU Test Method					
$LD_{50} \leq 5$ mg/kg	6	11.29 ±0. 20	10.38 ±0.62	0.90 (8.0%)	9.70 ±0.28	8.92 ±0.37	0.78 (8.0%)
$5 < LD_{50} \leq 50$ mg/kg	11	9.71 ±0.22	9.58 ±0.42	0.13 (1.3%)	8.47 ±0.28	8.41 ±0.44	0.06 (0.8%)
$50 < LD_{50} \leq 300$ mg/kg	12	7.74 ±0.10	7.99 ±0.18	-0.25 (-3.3%)	6.58 ±0.19	6.76 ±0.18	-0.18 (-2.7%)
$300 < LD_{50} \leq 2000$ mg/kg	16	8.52 ±0.21	8.16 ±0.19	0.35 (4.1%)	7.46 ±0.24	7.17 ±0.16	0.28* (3.8%)
$2000 < LD_{50} \leq 5000$ mg/kg	10	10.78 ±0.11	9.14 ±0.24	1.64* (15.2%)	9.20 ±0.24	7.61 ±0.37	1.59* (17.3%)
$LD_{50} > 5000$ mg/kg	12	9.87 ±0.34	8.23 ±0.48	1.65* (16.7%)	7.76 ±0.59	6.14 ±0.69	1.63* (21.0%)
		NHK NRU Test Method					
$LD_{50} \leq 5$ mg/kg	6	11.21 ±0.24	10.49 ±0.71	0.72 (6.4%)	9.66 ±0.27	8.97 ±0.52	0.69 (7.1%)
$5 < LD_{50} \leq 50$ mg/kg	11	9.70 ±0.18	9.78 ±0.41	-0.07 (-0.8%)	8.45 ±0.27	8.59 ±0.44	-0.13 (-1.6%)
$50 < LD_{50} \leq 300$ mg/kg	12	7.75 ±0.11	7.99 ±0.21	-0.24 (-3.1%)	6.58 ±0.19	6.76 ±0.18	-0.18 (-2.7%)
$300 < LD_{50} \leq 2000$ mg/kg	16	8.54 ±0.21	8.20 ±0.22	0.34 (3.9%)	7.48 ±0.23	7.17 ±0.16	0.31 (4.1%)
$2000 < LD_{50} \leq 5000$ mg/kg	10	10.77 ±0.08	9.40 ±0.25	1.38*(12.8%)	9.18 ±0.23	7.90 ±0.33	1.28* (14.0%)
$LD_{50} > 5000$ mg/kg	13	9.88 ±0.32	8.34 ±0.44	1.54*(15.6%)	7.66 ±0.56	6.12 ±0.63	1.53* (20.0%)

Abbreviations: 3T3=BALB/c 3T3 fibroblasts; GHS=Globally Harmonized System of Classification and Labelling of Chemicals (UN 2005); NHK=Normal human epidermal keratinocytes; RC=Registry of Cytotoxicity; UDP=Up-and-Down Procedure.
*Statistically significant (p <0.05) by a one-sided Wilcoxon signed rank test. Percent difference is shown in parentheses.
[1]Mean number of animals used ±standard errors for 10,000 simulations for each substance with a limit dose of 5000 mg/kg. Although the simulations used whole animals, averaging the results over a large number of simulations produced fractional numbers. Results are provided for 67 substances for the 3T3 NRU test method and 68 substances for the NHK NRU test method categorized using the rat acute oral LD_{50} reference values in mg/kg from **Table 4-2**.
[2]OECD (2001a); EPA (2002a).
[3]UN (2005).
[4]The RC rat-only weight regression is log LD_{50} (mg/kg) = 0.372 log IC_{50} (µg/mL) + 2.024
[5]Default starting dose = 175 mg/kg.
[6]The starting dose was one default dose lower than the predicted LD_{50} calculated using the IC_{50} values for each reference substance in the RC rat-only weight regression. The IC_{50} value for each reference substance was randomly selected from the distribution of values obtained during the testing with each method.
[7]Difference between mean animal use with the default starting dose and mean animal use with the predicted starting dose.

10.2.4   Refinement of Animal Use for the UDP When Using 3T3- and NHK-Based
          Starting Doses

A procedure refines animal use when it lessens or eliminates pain or distress in animals or
enhances animal well-being (ICCVAM 2003). This section evaluates whether the use of 3T3-
and NHK-based starting doses refines animal use by reducing the number of animals that die
and experience accompanying pain and distress during UDP testing, compared to the number
of animals that die when the default starting dose of 175 mg/kg is used. **Table 10-5** reports
the results for the UDP simulation modeling using the 5000 mg/kg limit dose. For every
regression evaluated, the mean number of deaths when using the IC_{50}-based starting doses
were essentially equal to the mean number of deaths when using the default starting dose.
The percentage of deaths, however, was slightly higher for the IC_{50}-based starting doses than
for the default starting dose because the total number of animals used was lower for the IC_{50}-
based starting doses. Thus, fewer animals were used when using an IC_{50}-based starting dose
compared with use of the default starting dose, but the same numbers of animals died.

**Table 10-5   Animal Deaths[1] in the UDP[2] Using Starting Doses Based on the 3T3 and
                NHK NRU Test Methods**

Assay/Regression	With Default Starting Dose[3]			With IC_{50}-Based Starting Dose[4]		
	Used	Dead	% Deaths	Used	Dead	% Deaths
3T3 NRU Test Method	**Dose-Mortality Slope = 2.0**					
RC rat-only millimole[5]	9.35	4.11	44.0%	8.80	4.09	46.5%
RC rat-only weight[6]	9.36	4.11	43.9%	8.70	4.05	46.6%
	Dose-Mortality Slope = 8.3					
RC rat-only millimole[5]	7.95	3.44	43.3%	7.42	3.43	46.2%
RC rat-only weight[6]	7.94	3.43	43.2%	7.32	3.39	46.3%
NHK NRU Test Method	**Dose-Mortality Slope = 2.0**					
RC rat-only millimole[5]	9.36	4.08	43.6%	8.86	4.07	45.9%
RC rat-only weight[6]	9.36	4.08	43.6%	8.80	4.02	45.7%
	Dose-Mortality Slope = 8.3					
RC rat-only millimole[5]	7.92	3.39	42.8%	7.43	3.39	45.6%
RC rat-only weight[6]	7.92	3.39	42.8%	7.38	3.35	45.4%

Abbreviations: 3T3=BALB/c 3T3 fibroblasts; NHK=Normal human epidermal keratinocytes; RC=Registry of
Cytotoxicity; UDP=Up-and-Down Procedure.
[1]Numbers are mean numbers of animals used for 10,000 simulations for each substance. Although the simulations used
whole animals, averaging the results over a large number of simulations produced fractional numbers. Upper limit dose =
5000 mg/kg. Results are provided for 67 substances in the 3T3 NRU test method and 68 substances in the NHK NRU test
methods.
[2]OECD (2001a); EPA (2002a).
[3]Default starting dose = 175 mg/kg.
[4]The starting dose was one default dose lower than the predicted LD_{50} calculated using the IC_{50} value in the regression
specified. The IC_{50} value for each reference substance was randomly selected from the distribution of values obtained during
the testing with each method.
[5]$\log LD_{50}$ (mmol/kg) = 0.439 $\log IC_{50}$ (mM) + 0.621.
[6]$\log LD_{50}$ (mg/kg) = 0.372 $\log IC_{50}$ (μg/mL) + 2.024.

10.2.5   Accuracy of UDP Outcomes Using the IC_{50}-Based Starting Doses
For each of the reference substances, the outcome of the simulated UDP testing, the simulated LD_{50} was used to classify the substance into a GHS acute oral toxicity category. The accuracy of GHS toxicity category assignments using the IC_{50}-based starting doses was determined by calculating the proportion of reference substances for which the GHS acute oral toxicity category obtained using the IC_{50}-based starting dose matched the categories obtained using the default starting dose.

The concordance between the GHS categories determined using the 3T3 and NHK NRU test methods with the RC rat-only millimole regression, and those determined using the UDP default starting dose, was 96% for 3T3 and 97% for NHK (see **Appendix N1**). The discordant reference substances were acetaminophen and sodium dichromate dihydrate in the 3T3 NRU test method, and acetaminophen, caffeine, and sodium dichromate dihydrate in the NHK NRU test method. The use of the IC_{50}-based starting dose from both *in vitro* NRU test methods resulted in a higher GHS category (i.e., higher simulated LD_{50}) for acetaminophen (simulated LD_{50} = 2047 vs. 1765 mg/kg for 3T3, and LD_{50} = 2174 vs. 1755 mg/kg for NHK), and a lower GHS category for sodium dichromate dihydrate (simulated LD_{50} = 44 vs. 52 mg/kg for 3T3 and LD_{50} = 45 vs. 52 mg/kg for NHK) than when using the default starting dose. The NHK-based starting dose resulted in a lower GHS category for caffeine (simulated LD_{50} = 280 vs. 357 mg/kg).

The concordance of GHS acute toxicity category predictions with those determined using the default starting dose was 97% for the 3T3 and NHK NRU test methods when the RC rat-only weight regression was used (see **Appendix N2**). The discordant reference substances were caffeine and sodium dichromate dihydrate. The simulated LD_{50} outcome for caffeine was lowered from 338 mg/kg for the default starting dose to 272 mg/kg for the 3T3-based starting dose, and from 339 mg/kg to 270 mg/kg for the NHK-based starting dose. The simulated LD_{50} outcome for sodium dichromate dihydrate was lowered from 51 mg/kg for the default starting dose to 48 mg/kg for the 3T3-based starting dose, and from 51 mg/kg to 49 mg/kg for the NHK-based starting dose.

Thus, the use of the IC_{50}-based starting doses did not significantly alter the outcome of the simulated UDP tests compared with the outcome obtained using the default starting doses.

10.3   Reduction and Refinement of Animal Use in the ATC Method

10.3.1   *In Vivo* Testing Using the ATC Method
This section describes the general dosing procedure for the conduct of the ATC procedure (OECD 2001d). The ATC is used to assign a test substance to the appropriate GHS category for classification and labeling. This is done by estimating the range of the LD_{50} values for the test substance, rather than calculating a point estimate of the LD_{50}. The time between administration of test substance doses is determined by the onset, duration, and severity of toxic signs. Guidance on the types of animals to use, animal housing, clinical observations, etc., which are outside the scope of the current discussion, are provided in the test guideline (See **Appendix M3**).

10.3.1.1   *Main Test*
The ATC method uses a stepwise administration of test substances to three animals at a time, at one of a number of fixed doses: 5, 50, 300, and 2000 mg/kg (and 5000 mg/kg, if necessary). The starting dose is selected so that at least some of the animals die at that dose. If no information on which to base a starting dose is available, a default starting dose of 300 mg/kg is used. The next step is determined by the starting dose and the outcome of the three animals tested at the starting dose and may be a decision to stop testing, test additional animals at the same dose, test at the next higher dose, or test at the next lower dose. For example, if two to three animals die or are in a moribund state after receiving the 300 mg/kg starting dose, the next step is to administer 50 mg/kg to three more animals. However, if no, or one, animal dies at 300 mg/kg, three additional animals are tested at that dose. Most substances require two to four dosing steps before they can be classified, and testing can be stopped. See **Appendix M3** for the outcome-based testing sequence for each starting dose.

10.3.1.2   *Limit Test*
For test substances that are likely to be nontoxic, the ATC guideline includes a limit test in which six animals (three animals per step [see **Appendix M3**]) are tested at the limit dose of 2000 mg/kg or three animals are tested at a limit dose of 5000 mg/kg (OECD 2001d).

10.3.2   Computer Simulation Modeling of the ATC Method
The simulation for the ATC method was performed using MATLAB® (The MathWorks, Inc. 1996-2004) computational software, which is functionally comparable with SAS® version 8. Two thousand simulations of ATC testing were run for each substance, *in vitro* NRU test method, and dose-mortality slope, using an upper limit dose of 2000 mg/kg[7]. The simulation implements the distributional assumptions underlying the dose-mortality response. The lowest dose at which an animal dies in response to the administration of a toxic substance varies from animal to animal. For an entire population of animals, mortality is assumed to have a log-normal distribution with the mean equal to the log of the true LD_{50}. Sigma (σ), the variability of the simulated population, is the inverse of the slope of the dose-mortality curve. For any given dose, the probability that an animal will die is computed by the cumulative log-normal distribution:

$$\text{Probability (death)} = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\log dose} e^{\frac{-(t - \log trueLD_{50})^2}{2\sigma^2}} dt$$

Because of a lack of information regarding the real dose-mortality curves, the simulations assumed several different values of the slope (i.e., the inverse of σ). Dose-mortality slopes of 0.5, 0.8, 2.0, 4.0, and 8.3 were chosen, so as to be comparable to the slopes chosen for simulation modeling of the UDP (see **Section 10.2.2**).

To model the variability of the IC_{50} values within and among laboratories, the values for each substance were log-transformed to normalize their distribution. The mean and variance of

---

[7] The results from ATC simulations for a limit dose of 5000 mg/kg will be presented in a future addendum to this document.

these log-transformed values were used to generate a log-normal distribution from which to randomly select an IC_{50} value.

The simulation procedure used the following steps for each substance:

1. The rodent acute oral LD_{50} value (in mg/kg) from **Table 4-2** was entered as the true LD_{50} value and the choices of assumed slope were entered as the true slope for the dose-mortality curve.
2. An IC_{50} value was selected from a distribution identified by the mean and variance of the IC_{50} values computed from the data to reflect that different laboratories produce different IC_{50} values in different situations (see **Tables 5-4** and **5-5** for mean IC_{50} values and standard deviations for the 3T3 and NHK NRU test methods, respectively).
3. The IC_{50} value from Step 2 was used in the regression model being evaluated to compute a predicted LD_{50} value for determining the starting dose.
4. The dosing simulation (of 2000 iterations) was run twice: once with the default starting dose of 300 mg/kg and once with a starting dose equal to the next fixed dose below the predicted LD_{50}, which was estimated by the regression being evaluated (i.e., the IC_{50}-based starting dose). If the IC_{50}-based starting dose was greater than the 2000 mg/kg limit dose, then testing proceeded using the 2000 mg/kg limit test rather than the main test.
5. For every dose group of three animals, one observation was sampled from a binomial distribution with the probability of death calculated by the probability equation for a population of three. The sampled value, referred to as N1, indicates the number of animals, 0, 1, 2, or 3, in the dosing group that die.
6. If N1 ≤1, step 4 is repeated with the same dose. The resulting sampled value from the binomial distribution is referred to as N2.
7. If N2 ≤1 and the dose is the highest dose tested, or the dose has already been decreased, a toxicity category is assigned and testing is terminated. If the dose is not the highest dose tested, or if the dose has not been decreased, the next higher fixed dose is administered and step 4 is repeated.
8. If N1 >1 or N2 >2, and the dose is the lowest dose tested, or if the dose has already been increased, a toxicity category is assigned and testing is terminated. If the dose is not the lowest dose tested, or if the dose has not already been increased, the next lower fixed dose is administered and step 4 is repeated.

10.3.3   Animal Savings for the ATC Method When Using 3T3- and NHK-Based Starting Doses

10.3.3.1 *The Effect of the Dose-Mortality Slope on Animal Use*
As described in **Section 10.3.2**, the simulation modeling of animal use for the ATC used five different dose-mortality slopes to assess animal use under various conditions of population variability. **Table 10-6** shows how mean animal use for the simulated ATC changes with dose-mortality slope for both the default starting dose of 300 mg/kg and a starting dose that was one fixed dose lower than that predicted by the 3T3 and NHK NRU IC_{50} values with the RC rat-only millimole regression. The mean number of animals used for the ATC method

decreased slightly with increasing slope for both the default starting dose and the IC_{50}-based starting dose.

The mean numbers of animals saved at all dose-mortality slopes were statistically significant (p <0.05 by one-sided Wilcoxon signed rank tests) when compared with mean animal use with the default dose, and tended to decrease with increasing slope. To simplify the presentation of animal savings and comparisons of the various regressions and starting doses, subsequent results in **Section 10.3.3** are shown only for dose-mortality slopes of 2.0 and 8.3. As stated earlier, these slopes are shown here because the slope of 2.0 is the default used for the calculation of LD_{50} by the UDP method (OECD 2001a; EPA 2002a) and the slope of 8.3 is shown to represent substances, such as pesticides, with higher slopes. Results for the other dose-mortality slopes were computed, and are presented in **Appendices N3** and **N4**.

Table 10-6   Change in Animal Use[1] with Dose-Mortality Slope in the ATC Method[2]

Dose-Mortality Slope	With Default Starting Dose[1,3]	With IC_{50}- Based Starting Dose[1,4]	Animals Saved[5]
3T3 NRU Test Method			
0.5	11.25 ±0.05	10.56 ±0.17	0.69* (6.1%)
0.8	11.10 ±0.07	10.46 ±0.19	0.64* (5.8%)
2.0	10.89 ±0.12	10.27 ±0.24	0.62* (5.7%)
4.0	10.73 ±0.15	10.15 ±0.26	0.58* (5.4%)
8.3	10.64 ±0.17	10.13 ±0.27	0.51* (4.8%)
NHK NRU Test Method			
0.5	11.25 ±0.05	10.43 ±0.16	0.82* (7.3%)
0.8	11.10 ±0.07	10.31 ±0.18	0.79* (7.1%)
2.0	10.91 ±0.11	10.11 ±0.24	0.80* (7.3%)
4.0	10.75 ±0.15	9.98 ±0.27	0.77* (7.1%)
8.3	10.67 ±0.17	9.96 ±0.29	0.70* (6.6%)

Abbreviations: 3T3=BALB/c 3T3 fibroblasts; ATC=Acute Toxic Class method; NHK=Normal human epidermal keratinocytes; RC=Registry of Cytotoxicity.
*Statistically significant (p <0.05) by a one-sided Wilcoxon rank test. Percent difference is shown in parentheses.
[1]Mean numbers of animals used ± standard errors for 2000 simulations each for 67 substances in the 3T3 NRU test method and 68 substances in the NHK NRU test method. Although the simulations used whole animals, averaging the results over a large number of simulations produced fractional numbers. Upper limit dose = 2000 mg/kg.
[2]OECD (2001d).
[3]Default starting dose = 300 mg/kg.
[4]Next fixed dose lower than the predicted LD_{50} calculated using the IC_{50} value for each reference substance in the RC rat-only millimole regression: log LD_{50} (mmol/kg) = 0.439 log IC_{50} (mM) + 0.621. The IC_{50} value for each reference substance was randomly selected from the distribution of values obtained during the testing with each method.
[5]Difference between mean animal use with the default starting dose and mean animal use with the IC_{50}-based starting dose.

10.3.3.2   *Mean Animal Use for ATC Simulations – Comparison of Regressions and*
            Predictions from the 3T3 and NHK NRU Test Methods

Table 10-7 shows the mean animal use for testing the reference substances using the
simulated ATC method, when the starting dose was the default starting dose and when the
starting dose was one fixed dose lower than that determined by the 3T3 and NHK-predicted
LD_{50}, and the regressions evaluated in **Section 6.4** for prediction of GHS category. The mean
difference in animal use between the two starting doses is the mean animal savings. All mean
animal savings were statistically significant ($p < 0.05$ using one-sided Wilcoxon signed rank
tests), and ranged from 0.51 (4.8%) to 1.09 (10.2%) animals per test depending upon the
NRU test method, regression, and dose-mortality slope. The lowest mean animal savings
were obtained for the RC rat-only millimole regression (0.51 [4.8%] to 0.80 [7.3%] animals
per test), and the highest were obtained with the RC rat-only weight regression (0.91 [8.6%]
to 1.09 [10.2%] animals per test).

The animal savings obtained using the *in vitro* NRU test methods with the RC rat-only
regressions apply only to the reference substances evaluated in this validation study, and are
based on substances pre-selected for their known *in vivo* toxicities and may not be broadly
applicable to other substances. **Table 3-4** shows that 22 (38%) of the 58 RC substances
selected for testing were known to have a poor fit to the RC millimole regression (i.e., the
predicted LD_{50} was outside the RC acceptance interval). **Table 6-3** shows that 40% (28/70 in
the 3T3) and 44% (31/71 in the NHK) of the reference substances that yielded IC_{50} values
were outliers. Substances that better fit the regression are likely to yield greater animal
savings.

Table 10-7   Animal Use[1] for the ATC[2] Method Using Starting Doses Based on the 3T3 and NHK NRU Test Methods with the Different Regressions

Method/Regression	With Default Starting Dose[3]	With IC_{50}- Based Starting Dose[4]	Animals Saved[5]	With Default Starting Dose[3]	With IC_{50}- Based Starting Dose[5]	Animals Saved[5]
3T3 NRU Test Method	**Dose-Mortality Slope = 2.0**			**Dose-Mortality Slope = 8.3**		
RC rat-only millimole[6]	10.89 ±0.12	10.27 ±24	0.62* (5.7%)	10.64 ±0.17	10.13 ±0.27	0.51* (4.8%)
RC rat-only weight[7]	10.89 ±0.12	9.85 ±0.24	1.04* (9.6%)	10.64 ±0.17	9.55 ±0.29	1.09* (10.2%)
NHK NRU Test Method	**Dose-Mortality Slope = 2.0**			**Dose-Mortality Slope = 8.3**		
RC rat-only millimole[6]	10.91 ±0.11	10.11 ±0.24	0.80* (7.3%)	10.67 ±0.17	9.96 ±0.29	0.70* (6.6%)
RC rat-only weight[7]	10.91 ±0.11	9.95 ±0.24	0.96* (8.8%)	10.67 ±0.17	9.75 ±0.30	0.91* (8.6%)

Abbreviations: 3T3=BALB/c 3T3 fibroblasts; ATC=Acute Toxic Class method; NHK=Normal human epidermal keratinocytes; RC=Registry of Cytotoxicity.
*Statistically significant (p<0.05) using a one-sided Wilcoxon signed rank test. Percentage difference is shown in parentheses.
[1]Mean numbers of animals used ±standard errors for 2000 simulations each for 67 substances in the 3T3 NRU test method and 68 substances in the NHK NRU test method. Upper limit dose =2000 mg/kg.
[2]OECD (2001d).
[3]Default starting dose =300 mg/kg.
[4]Starting dose was one fixed dose lower than the predicted LD_{50} calculated using the IC_{50} value for each reference substance in the regression specified. The IC_{50} value for each reference substance was randomly selected from the distribution of values obtained during the testing with each test method.
[5]Difference between mean animal use with the default starting dose and mean animal use with the IC_{50}-based starting dose.
[6]log LD_{50} (mmol/kg) = 0.439 log IC_{50} (mM) + 0.621.
[7]log LD_{50} (mg/kg) = 0.372 log IC_{50} (μg/mL) + 2.024.

10.3.3.3   *Animal Savings in the ATC Method by GHS Acute Oral Toxicity Category Using*
            the 3T3- and NHK -Based Starting Doses

Tables 10-8 and **10-9** show mean animal use and mean animal savings for the ATC when
used with the *in vitro* NRU test methods, organized by GHS category (UN 2005), and when
based on the:

- RC rat-only millimole regression (**Table 10-8**)
- RC rat-only weight regression (**Table 10-9**)

The following data come from the same analyses as the data provided in **Table 10-7**.

The analyses showed that:

- For each *in vitro* NRU test method and regression, the highest mean animal
  savings were generally in the LD_{50} ≤5 mg/kg and LD_{50} >5000 mg/kg toxicity
  categories.
- For each NRU test method and regression, the lowest mean animal savings
  were in the $300 < LD_{50}$ ≤2000 mg/kg toxicity category.

Animal Savings in the ATC Method by GHS Category Using the 3T3- and NHK-Based
Starting Doses with the RC Rat-Only Millimole Regression
Table 10-8 shows the mean animal savings in the ATC method by GHS category for the *in
vitro* NRU test methods used with the RC rat-only millimole regression. Mean differences
between animal use for the default starting dose and with the IC_{50}-determined starting dose
were statistically significant ($p < 0.05$) by a one-sided Wilcoxon signed rank test for the
following GHS toxicity categories, NRU test methods, and dose-mortality slopes:

- The use of both *in vitro* NRU test methods at both dose-mortality slopes for
  substances with $5 < LD_{50}$ ≤50 mg/kg produced savings of 1.15 (9.8%) to 1.33
  (11.4%) animals per test
- The use of the 3T3 NRU test method at both dose-mortality slopes for
  substances with $300 < LD_{50}$ ≤2000 mg/kg used more animals per test (i.e.,
  produced savings of -0.92 [-9.5%] to -1.30 [-14.0%] animals per test)
- The use of both *in vitro* NRU test methods at both dose-mortality slopes for
  substances with LD_{50} >5000 mg/kg produced savings of 2.03 (17.1%) to 2.66
  (22.2%) animals per test

At the dose-mortality slope of 2.0, the mean animal savings with the 3T3 NRU test method
ranged from -0.92 (-9.5%) to 2.68 (27.4%) animals per test, and the animal savings with the
NHK NRU test method ranged from -0.60 (-6.1%) to 2.96 (30.4%) animals per test. At the
dose-mortality slope of 8.3, the mean animal savings with the 3T3 NRU test method ranged
from -1.30 (-14.0%) to 2.70 (29.7%) animals per test, and the animal savings with the NHK
NRU test method ranged from -0.85 (-9.2%) to 2.99 (33.0%) animals per test.

Table 10-8   Animal Savings[1] for the ATC[2] Method by GHS Acute Oral Toxicity Category[3] Using Starting Doses Based on the 3T3 and NHK NRU Test Methods with the RC Rat-Only Millimole Regression[4]

GHS Acute Oral Toxicity Category[3]	Number of Reference Substances	Dose-Mortality Slope = 2.0			Dose-Mortality Slope = 8.3		
		With Default Starting Dose[5]	With IC_{50}-Based Starting Dose[6]	Animals Saved[7]	With Default Starting Dose[5]	With IC_{50}-Based Starting Dose[6]	Animals Saved[7]
		3T3 NRU Test Method					
$LD_{50} \leq 5$ mg/kg	6	9.77 ±0.17	7.09 ±1.09	2.68 (27.4%)	9.08 ±0.08	6.38 ±1.09	2.70 (29.7%)
$5 < LD_{50} \leq 50$ mg/kg	11	11.56 ±0.21	10.39 ±0.52	1.17* (10.2%)	11.75 ±0.16	10.60 ±0.43	1.15* (9.8%)
$50 < LD_{50} \leq 300$ mg/kg	12	10.81 ±0.20	10.39 ±0.17	0.42 (3.9%)	9.42 ±0.26	9.27 ±0.11	0.15 (1.6%)
$300 < LD_{50} \leq 2000$ mg/kg	16	9.75 ±0.07	10.67 ±0.48	-0.92* (-9.5%)	9.26 ±0.10	10.56 ±0.62	-1.30* (-14.0%)
$2000 < LD_{50} \leq 5000$ mg/kg	10	11.22 ±0.08	11.14 ±0.08	0.08 (0.7%)	11.88 ±0.10	11.77 ±0.10	0.11 (0.9%)
$LD_{50} > 5000$ mg/kg	12	11.85 ±0.04	9.82 ±0.78	2.03* (17.1%)	12.00 ±0.000	9.81 ±0.84	2.19* (18.3%)
		NHK NRU Test Method					
$LD_{50} \leq 5$ mg/kg	6	9.74 ±0.16	6.78 ±1.31	2.96 (30.4%)	9.09 ±0.08	6.09 ±1.23	2.99 (33.0%)
$5 < LD_{50} \leq 50$ mg/kg	11	11.56 ±0.21	10.38 ±0.35	1.18* (10.2%)	11.76 ±0.17	10.42 ±0.45	1.33* (11.4%)
$50 < LD_{50} \leq 300$ mg/kg	12	10.83 ±0.21	10.39 ±0.29	0.44 (4.0%)	9.44 ±0.26	9.63 ±0.49	-0.20 (-2.1%)
$300 < LD_{50} \leq 2000$ mg/kg	16	9.77 ±0.06	10.37 ±0.49	-0.60 (-6.1%)	9.26 ±0.10	10.11 ±0.63	-0.85 (-9.2%)
$2000 < LD_{50} \leq 5000$ mg/kg	10	11.22 ±0.08	11.25 ±0.12	-0.03 (-0.3%)	11.87 ±0.10	11.89 ±0.15	-0.02 (-0.2%)
$LD_{50} > 5000$ mg/kg	13	11.86 ±0.03	9.43 ±0.73	2.43* (20.5%)	12.00 ±0.000	9.34 ±0.80	2.66* (22.2%)

Abbreviations: 3T3=BALB/c 3T3 fibroblasts; ATC=Acute Toxic Class method; GHS=Globally Harmonized System of Classification and Labelling of Chemicals (UN 2005); NHK=Normal human epidermal keratinocytes; RC=Registry of Cytotoxicity.
*Statistically significant (p <0.05) by a one-sided Wilcoxon signed rank test. Percentage difference is shown in parentheses.
[1]Mean number of animals used ±standard errors for 2000 simulations for each substance with an upper limit dose of 2000 mg/kg. Results are provided for 67 substances in the 3T3 NRU test method and 68 substances in the NHK NRU test method categorized using the rat acute oral LD_{50} reference values in mg/kg from **Table 4-2**. Although the simulations used whole animals, averaging the results over a large number of simulations produced fractional numbers.
[2]OECD (2001d).
[3]GHS for acute oral toxicity (UN 2005).
[4]The RC rat-only millimole regression is log LD_{50} (mmol/kg) = 0.439 log IC_{50} (mM) + 0.621.
[5]Default starting dose =300 mg/kg.
[6]The starting dose was the next fixed dose lower than the predicted LD_{50} using the IC_{50} for each reference substance in the RC rat-only millimole regression. The IC_{50} value for each reference substance was randomly selected from the distribution of values obtained during the testing with each method.
[7]Difference between mean animal use with the default starting dose and mean animal use with the IC_{50}-based starting dose.

At both the 2.0 and 8.3 dose-mortality slopes, the mean animal savings using the 3T3 NRU test method were lower than the corresponding savings using the NHK NRU test method, for substances in at least four of the six toxicity categories: $LD_{50} \leq 5$ mg/kg; $5 < LD_{50} \leq 50$ mg/kg; $300 < LD_{50} \leq 2000$ mg/kg; and $LD_{50} > 5000$ mg/kg. The mean animal savings per test were higher with the 3T3 NRU test method than the NHK NRU test method for substances in the $2000 < LD_{50} \leq 5000$ mg/kg category at both dose-mortality slopes. For substances in the $50 < LD_{50} \leq 300$ mg/kg category, the mean animal savings using the 3T3 NRU test method was greater than the savings using the NHK NRU test method, when the dose-mortality slope equaled 8.3. When the 3T3 NRU test method was used, the highest mean animal savings occurred when testing substances in the $LD_{50} \leq 5$ mg/kg category (2.68 [27.4%] animals per test at dose-mortality slope = 2.0, and 2.70 [29.7%] at dose-mortality slope = 8.3). When the NHK NRU test method was used, the highest mean animal savings occurred when testing substances in the $LD_{50} \leq 5$ mg/kg category (2.96 [30.4%] animals per test at dose-mortality slope = 2.0, and 2.99 [33.0%] animals per dose at dose-mortality slope = 8.3). However, the animal savings were not statistically significant with either *in vitro* NRU test method.

The smallest mean animal savings (≤ 0.44) in both *in vitro* NRU test methods were observed for substances with LD_{50} values between 50 and 5000 mg/kg. Because the default starting dose was 300 mg/kg, little change in mean animal use was expected for substances in the $50 < LD_{50} \leq 300$ mg/kg and $300 < LD_{50} \leq 2000$ mg/kg categories. The mean animal savings from both *in vitro* NRU test methods and both dose-mortality slopes for the substances in the $50 < LD_{50} \leq 300$ mg/kg category were -0.20 to 0.44 animals per test. There were no animal savings for substances in the $300 < LD_{50} \leq 2000$ mg/kg category using either NRU test method or dose-mortality slope. In fact, significantlyore animals were used when the starting doses were based on the 3T3 NRU IC_{50} than using the default starting dose (-0.92 to -1.30 animals per test). More animals were also used when the starting doses were based on the NHK NRU IC_{50} (-0.85 to -0.60 animals/test), but the difference was not statistically significant.

The animal savings in the various GHS acute oral toxicity categories using the *in vitro* NRU test methods with the RC rat-only millimole regression applies only to the reference substances evaluated in this validation study, and may not be broadly applicable to other substances. The animal savings for future testing using the *in vitro* NRU test methods with the RC rat-only millimole regression will depend on the prevalence of test substances in each of the GHS acute oral toxicity categories.

Animal Savings with the ATC Method by GHS Category Using 3T3- and NHK-Based Starting Doses with the RC Rat-Only Weight Regression
Table 10-9 shows the animal savings for the simulated ATC method by GHS category for the *in vitro* NRU methods used with the RC rat-only weight regression. Mean animal savings were statistically significant ($p < 0.05$) by a one-tailed Wilcoxon signed rank test for the following GHS toxicity categories, NRU test methods, and dose-mortality slopes.

- The use of both *in vitro* NRU test methods at both dose-mortality slopes for substances with $5 < LD_{50} \leq 50$ mg/kg produced savings of 1.25 (10.8%) to 1.51 (13.0%) animals per test.

Table 10-9   Animal Savings[1] for the ATC[2] Method by GHS Acute Oral Toxicity Category[3] Using Starting Doses Based on the 3T3 and NHK NRU Test Methods with the RC Rat-Only Weight Regression[4]

GHS Acute Oral Toxicity Category[3]	Number of Reference Substances	Dose-Mortality Slope = 2.0			Dose-Mortality Slope = 8.3		
		With Default Starting Dose[5]	With IC_{50}-Based Starting Dose[6]	Animals Saved[7]	With Default Starting Dose[5]	With IC_{50}-Based Starting Dose[6]	Animals Saved[7]
		3T3 NRU Test Method					
$LD_{50} \leq 5$ mg/kg	6	9.77 ±0.17	7.56 ±1.03	2.21 (22.6%)	9.08 ±0.08	6.85 ±0.99	2.24 (24.6%)
$5 < LD_{50} \leq 50$ mg/kg	11	11.56 ±0.21	10.06 ±0.38	1.51* (13.0%)	11.75 ±0.16	10.27 ±0.33	1.48* (12.6%)
$50 < LD_{50} \leq 300$ mg/kg	12	10.81 ±0.20	10.35 ±0.18	0.47* (4.3%)	9.42 ±0.26	9.20 ±0.10	0.22 (2.4%)
$300 < LD_{50} \leq 2000$ mg/kg	16	9.75 ±0.07	10.67 ±0.50	-0.93* (-9.5%)	9.26 ±0.10	10.65 ±0.66	-1.39 (-15.0%)
$2000 < LD_{50} \leq 5000$ mg/kg	10	11.22 ±0.08	9.80 ±0.51	1.43* (12.7%)	11.88 ±0.10	9.44 ±0.88	2.43 (20.5%)
$LD_{50} > 5000$ mg/kg	12	11.85 ±0.04	8.83 ±0.83	3.02* (25.5%)	12.00 ±0.00	8.67 ±0.91	3.33* (27.7%)
		NHK NRU Test Method					
$LD_{50} \leq 5$ mg/kg	6	9.74 ±0.16	6.87 ±1.28	2.87 (29.4%)	9.09 ±0.08	6.18 ±1.20	2.91 (32.0%)
$5 < LD_{50} \leq 50$ mg/kg	11	11.56 ±0.21	10.31 ±0.19	1.25* (10.8%)	11.76 ±0.17	10.40 ±0.33	1.36* (11.5%)
$50 < LD_{50} \leq 300$ mg/kg	12	10.83 ±0.21	10.41 ±0.28	0.42 (3.8%)	9.44 ±0.26	9.63 ±0.49	-0.20 (-2.1%)
$300 < LD_{50} \leq 2000$ mg/kg	16	9.77 ±0.62	10.46 ±0.50	-0.69 (-7.1%)	9.26 ±0.10	10.23 ±0.65	-0.97 (-10.4%)
$2000 < LD_{50} \leq 5000$ mg/kg	10	11.22 ±0.09	10.69 ±0.37	0.53 (4.7%)	11.87 ±0.10	11.03 ±0.60	0.84 (7.1%)
$LD_{50} > 5000$ mg/kg	13	11.86 ±0.03	8.91 ±0.78	2.94* (24.8%)	12.00 ±0.00	8.75 ±0.85	3.25* (27.1%)

Abbreviations: 3T3=BALB/c 3T3 fibroblasts; ATC=Acute Toxic Class method; GHS=Globally Harmonized System of Classification and Labelling of Chemicals (UN 2005); NHK=Normal human epidermal keratinocytes; RC=Registry of Cytotoxicity.
*Statistically significant (p <0.05) by a one-sided Wilcoxon signed rank test. Percentage difference is shown in parentheses.
[1]Mean number of animals used ±standard errors for 2000 simulations for each substance with an upper limit dose of 2000 mg/kg. Although the simulations used whole animals, averaging the results over a large number of simulations produced fractional numbers. Results are provided for 67 substances in the 3T3 NRU test method and 68 substances in the NHK NRU test method categorized using the rat acute oral reference LD_{50} values in mg/kg from **Table 4-2**.
[2]OECD (2001d).
[3]GHS for acute oral toxicity (UN 2005).
[4]From **Table 6-2;** log LD_{50} (mg/kg) = 0.372 log IC_{50} (μg/mL) + 2.024
[5]Default starting dose = 300 mg/kg.
[6]The starting dose was one fixed dose lower than the predicted LD_{50} calculated using the IC_{50} for each reference substance in the RC rat-only weight regression. The IC_{50} value for each reference substance was randomly selected from the distribution of values obtained during the testing with each method.
[7]Difference between mean animal use with the default starting dose and mean animal use with the IC_{50}-based starting dose.

- The use of the 3T3 NRU test method at the 2.0 dose-mortality slope for substances with $50 < LD_{50} \leq 300$ mg/kg produced savings of 0.47 (4.3%) animals per test.
- The use of the 3T3 NRU test method at the 2.0 dose-mortality slope for substances with $300 < LD_{50} \leq 2000$ mg/kg produced savings of -0.93 (-9.5%) animals per test (i.e., used more animals per test than the default starting dose).
- The use of the 3T3 NRU test method at the 2.0 dose-mortality slope for substances with $2000 < LD_{50} \leq 5000$ mg/kg produced savings of 1.43 (12.7%) animals per test.
- The use of both *in vitro* NRU test methods at both dose-mortality slopes for substances with $LD_{50} > 5000$ mg/kg produced savings of 2.94 (24.8%) to 3.33 (27.7%) animals per test.

The mean animal savings with the 3T3 and NHK NRU test methods were similar for most acute oral toxicity categories at both dose-mortality slopes; the mean savings for the 3T3 NRU test method was slightly higher than for the NHK NRU test method for most toxicity categories. At the dose-mortality slope of 2.0, the mean animal savings for the 3T3 NRU test method (for the various toxicity categories) ranged from -0.93 (-9.5%) to 3.02 (25.5%) animals per test, and savings for the NHK NRU test method ranged from -0.69 (-7.1%) to 2.94 (24.8%) animals per test. At the dose-mortality slope of 8.3, animal savings with the 3T3 NRU test method ranged from -1.39 (-15.0%) to 3.33 (27.7%) animals per test, and savings with the NHK NRU test method ranged from -0.97 (-10.4%) to 3.25 (27.1%) animals per test.

There were no mean animal savings ($\leq$-0.69 animals) for substances with $300 < LD_{50} \leq 2000$ when either *in vitro* NRU test method was used. The mean animal savings for the substances in the $50 < LD_{50} \leq 300$ mg/kg category using both *in vitro* NRU test methods and dose-mortality slopes were also relatively small (-0.20 to 0.47 animals per test). Because the default starting dose was 300 mg/kg, little change in mean animal use was expected for substances in the $50 < LD_{50} \leq 300$ mg/kg and $300 < LD_{50} \leq 2000$ mg/kg categories. The highest mean animal savings ($\leq$-0.69 animals) occurred for substances with LD50 >5000 mg/kg when either *in vitro* NRU test method was used. For both test methods and dose-mortality slopes, the mean animal savings for substances in this category were 2.94 (24.8%) to 3.33 (27.7%) animals per test and were statistically significant. Mean animal savings were also high (2.21 [22.6%] to 2.91 [32.0%] animals per test) for substances with $LD_{50} \leq 5$ mg/kg, but these savings were not statistically significant.

The animal savings in the various GHS categories using the two *in vitro NRU* test methods with the RC rat-only weight regression applies only to the reference substances evaluated in this validation study, and may not be broadly applicable to other substances.

10.3.4   Refinement of Animal Use in the ATC Method When Using 3T3- and NHK-Based Starting Doses
A procedure refines animal use when it lessens or eliminates pain or distress in animals, or enhances animal well-being (ICCVAM 2003). This section evaluates whether the use of 3T3- and NHK-based starting doses refines animal use by reducing the number of animals that die

when the IC_{50}-predicted starting doses are used, compared to the number of animals that die when using the default ATC starting dose of 300 mg/kg. **Table 10-10** reports the results for the ATC simulation modeling using the 2000 mg/kg limit dose. For every regression evaluated, the mean number of deaths when using the 3T3- and NHK-based starting doses was less than the mean number of deaths when using the default starting dose, by approximately 0.4 to 0.5 deaths per test. For the RC rat-only millimole regression and the RC rat-only weight regression, the percentage of deaths (compared with the numbers of animals used) was also slightly lower with the *in vitro*-based starting dose compared with the default starting dose. In general, fewer animals were used with the *in vitro*-based starting dose, and fewer animals died.

Table 10-10   Animal Deaths[1] for the ATC[2] Method Using Starting Doses Based on the 3T3 and NHK NRU Test Methods

Method/Regression	Default Starting Dose[3]			IC_{50}- Based Starting Dose[4]		
	Used	Dead	% Deaths	Used	Dead	% Deaths
3T3 NRU Test Method	**Dose-Mortality Slope = 2.0**					
RC rat-only millimole[5]	10.89	3.77	34.6%	10.27	3.31	32.2%
RC rat-only weight[6]	10.89	3.77	34.6%	9.85	3.27	33.2%
	Dose-Mortality Slope = 8.3					
RC rat-only millimole[5]	10.64	3.20	30.1%	10.13	2.77	27.3%
RC rat-only weight[6]	10.64	3.20	30.1%	9.55	2.73	28.6%
NHK NRU Test Method	**Dose-Mortality Slope = 2.0**					
RC rat-only millimole[5]	10.91	3.72	34.1%	10.11	3.19	31.6%
RC rat-only weight[6]	10.91	3.72	34.1%	9.95	3.21	32.3%
	Dose-Mortality Slope = 8.3					
RC rat-only millimole[5]	10.67	3.15	29.5%	9.96	2.67	26.8%
RC rat-only weight[6]	10.67	3.15	29.5%	9.75	2.67	27.4%

Abbreviations: 3T3=BALB/c 3T3 fibroblasts; ATC=Acute Toxic Class method; NHK=Normal human epidermal keratinocytes; RC=Registry of Cytotoxicity.
[1]Mean numbers of animals used for 2000 simulations for each of 67 substances in the 3T3 NRU test method and 68 substances in the NHK NRU test method. Although the simulations used whole animals, averaging the results over a large number of simulations produced fractional numbers. Upper limit dose =2000 mg/kg.
[2]OECD (2001d).
[3]Default starting dose =300 mg/kg.
[4]The starting dose was one fixed dose lower than the predicted LD_{50} calculated by using the IC_{50} for each reference substance in the regression evaluated. The IC_{50} value for each reference substance was randomly selected from the distribution of values obtained during the testing with each method.
[5]log LD_{50} (mmol/kg) = 0.439 log IC_{50} (mM) + 0.621.
[6]log LD_{50} (mg/kg) = 0.372 log IC_{50} (μg/mL) + 2.024.

10.3.5   Accuracy of the ATC Method Outcomes Using the IC_{50}-Based Starting Doses
The accuracy of the outcome of the simulated ATC testing (i.e., the simulated GHS acute oral toxicity category) using the IC_{50}-based starting dose was determined by calculating the proportion of reference substances for which the simulated GHS category for the IC_{50}-based starting dose matched the simulated GHS category for the default starting dose.

When the RC rat-only millimole regression with the 3T3 and NHK NRU test methods was used, the concordance of simulated GHS categories for the IC_{50}-based starting doses with those for the default starting dose was 99% for both *in vitro* NRU test methods (see **Appendix N3**). The discordant reference substance in the 3T3 NRU test method was caffeine. The simulated GHS category using the 3T3-based starting dose was $50 < LD_{50} \leq 300$ mg/kg, and the simulated GHS category using the default starting dose was $300 < LD_{50} \leq 2000$ mg/kg.

The discordant reference substance in the NHK NRU test method was sodium dichromate dihydrate. The simulated GHS acute oral toxicity category using the NHK-based starting dose was $5 < LD_{50} \leq 50$ mg/kg and the simulated GHS category using the default starting dose was $50 < LD_{50} \leq 300$ mg/kg. Both discordant substances were predicted to have a starting dose one category below the actual category.

When the RC rat-only weight regression was used with the 3T3 and NHK NRU test methods, the concordance of simulated GHS acute toxicity category predictions with those determined using the default starting dose was 99% and 97% for the 3T3 and the NHK NRU test methods, respectively (see **Appendix N4**). The discordant reference substance in the 3T3 NRU test method was caffeine. The simulated GHS acute oral toxicity category for caffeine using the 3T3-based starting dose was $50 < LD_{50} \leq 300$ mg/kg and that using the default starting dose was $300 < LD_{50} \leq 2000$ mg/kg. The discordant reference substances in the NHK NRU test method were caffeine and sodium dichromate dihydrate. The simulated GHS acute oral toxicity category for caffeine using the NHK-based starting dose was $50 < LD_{50} \leq 300$ mg/kg and the simulated GHS category using the default starting dose was $300 < LD_{50} \leq 2000$ mg/kg. The simulated GHS acute oral toxicity category for sodium dichromate dihydrate using the NHK-based starting dose was $5 < LD_{50} \leq 50$ mg/kg while that for the default starting dose was $50 < LD_{50} \leq 300$ mg/kg. Similar to what was seen with the RC millimole regression, the predicted starting doses for the discordant substances were one GHS category below the actual category.

Thus, the use of the IC_{50}-based starting doses did not significantly alter the outcomes of the simulated ATC tests compared with the outcome based on the default starting dose.

10.4   The Impact of Accuracy on Animal Savings

Two types of accuracy analyses were performed for the NICEATM/ECVAM validation study. The first analyses determined the accuracy of using the NRU IC_{50} values with an IC_{50}-LD_{50} regression to predict LD_{50} values. It calculated the concordance for GHS acute oral toxicity category by comparing the GHS categorization yielded by the NRU-predicted LD_{50} values (using the *in vitro* NRU IC_{50} values in the regressions presented in **Table 6-5)** with the GHS categorization based on rat acute oral LD_{50} data (see **Section 6.4**). The second analysis determined the accuracy of the simulation outcomes using the IC_{50}-based starting doses (see **Sections 10.2.5** and **10.3.5**). It calculated the concordance for the GHS acute oral toxicity category outcomes obtained using the IC_{50}-based starting doses with the GHS category outcomes obtained using the default starting dose. The magnitude of animal savings did not correlate with either determination of accuracy and the accuracy determinations for IC_{50}-based predictions and IC_{50}-based outcomes for GHS category did not correlate with one another.

Animal savings did not correlate with the accuracy of the GHS acute oral toxicity category predictions based on the LD_{50} values calculated using the IC_{50} values in the RC rat-only regressions (see **Sections 6.4.2** and **6.4.3**). Substances in categories with the lowest accuracy produced the highest animal savings. For example, using the RC rat-only millimole regression with the *in vitro* NRU IC_{50} values yielded very low accuracy (0 to 17%) for GHS acute oral toxicity category prediction for substances with LD_{50} >5000 mg/kg (see **Table 6-7**), but the highest animal savings of 14.8 to 20.3% occurred in this category (see **Table 10-3**). Animal savings were small, 4.5 to 6.5%, for substances with $300 \leq LD_{50} \leq 2000$ mg/kg, but the accuracy of 75-81% for GHS acute oral toxicity category prediction was relatively high. The reason that animal savings is unrelated to the accuracy of prediction of GHS acute oral toxicity category based on the LD_{50} values calculated using IC_{50} values in the RC rat-only regressions is because two different standards are used for comparison in the two analyses:

- GHS acute oral toxicity category predictions using IC_{50} values in the RC rat-only regressions are compared with the GHS categories derived from the *in vivo* reference LD_{50}
- The number of animals used (to determine animal savings) was compared with the animal use at the default starting dose of 175 mg/kg for the UDP or 300 mg/kg for the ATC

Despite the relatively poor GHS accuracy for the low toxicity chemicals (the toxicity of almost all were overpredicted by one GHS category), animal savings were greatest due to the fact that testing goes to the limit dose faster.

The accuracy of the simulated GHS toxicity category assignments using the IC_{50}-based starting doses for UDP and ATC test simulations was determined by calculating the proportion of reference substances for which the GHS acute oral toxicity category obtained using the IC_{50}-based starting dose matched the categories obtained using the default starting dose (see **Sections 10.2.5** and **10.3.5**). The accuracy of these GHS toxicity category assignments based on the simulation outcomes does not correlate with animal savings using the IC_{50} values in the RC rat-only regressions (see **Sections 6.4.2** and **6.4.3**). For example, the accuracy of GHS acute oral toxicity category outcomes for the ATC test method when using the RC rat-only millimole regression was 100% for the 3T3 NRU test method for substances with $300 \leq LD_{50} \leq 2000$ mg/kg (see **Appendix N3**). In contrast, the animal savings for those substances was negative at -6.1 to -14.0% (i.e., more animals were used compared with the default starting dose) (see **Table 10-8**). The reason the outcome-based GHS acute oral toxicity category predictions is unrelated to animal savings is that two different parameters are being measured in the two analyses:

- The accuracy of the simulatedGHS acute oral toxicity outcomes using the IC_{50}-based starting doses measured outcome (i.e., simulated GHS category based on the simulated LD_{50} outcome for the UDP and simulated GHS category for the ATC)
- The animal savings analysis measured the number of animals used at the IC_{50}-based starting dose and the default starting dose of 175 mg/kg for the UDP or 300 mg/kg for the ATC

Thus, the measurements for the two analyses are different: outcome (i.e., GHS category) and number of animals used to achieve the outcome.

In addition, accuracy of the GHS toxicity category assignments based on the simulation outcomes does not correlate with the accuracy of the GHS acute oral toxicity category predictions using the IC_{50} values in the RC rat-only regressions (see **Section 6.4.2** and **6.4.3**). For example, the overall accuracy of GHS acute oral toxicity category outcomes for the ATC test method when using the RC rat-only millimole regression was 99% for both *in vitro* NRU test methods (see **Section 10.3.5** and **Appendix N3**). In contrast, the overall accuracy of GHS acute oral toxicity category predictions using the using the IC_{50} values in the RC rat-only millimole regression was 31% for the 3T3 NRU test method and 29% for the NHK NRU test method (see **Table 6-7**). The reason the simulated outcome-based GHS acute oral toxicity category predictions differed from the accuracy of the GHS acute oral toxicity category predictions based on the calculation of LD_{50} using the IC_{50} in the IC_{50}-LD_{50} regression is because two different standards are used for comparison in the two analyses:

- Simulated GHS acute oral toxicity outcomes for the IC_{50}-based starting doses were compared with the simulated GHS category outcomes using the default starting doses
- GHS acute oral toxicity category predictions using the IC_{50} values in the RC rat-only regressions were compared with the GHS category derived from the *in vivo* reference LD_{50}

Thus, despite that the IC_{50} values and IC_{50}-LD_{50} regressions predicted GHS acute oral toxicity categories poorly, the GHS acute oral toxicity category outcomes using the IC_{50}-based starting doses were practically the same as the GHS acute oral toxicity category outcomes using the default starting dose.

10.5   The Impact of Prevalence on Animal Savings

As stated several times in this section, the animal savings for substances tested in the futureusing the 3T3 and NHK NRU test methods to determine the staring dose for rodent acute oral toxicity test methods will depend on the proportion of test substances that fall into each of the GHS acute toxicity hazard categories. Although the prevalence of substances among the different categories will depend, to a large extent, on the mandate of a particular regulatory agency, Spielmann et al. (1999) indicated that 76% (845/1115) of the industrial substances submitted to the Federal Institute for Health Protection of Consumers and Veterinary Medicine in Berlin, Germany, since 1982 had LD_{50} >2000 mg/kg. The extent to which these substances represent the population of substances in commerce is not known. However, if the results of the validation study are broadly applicable to substances to be tested in the future, and if such substances are relatively nontoxic, the selection of starting doses using the *in vitro* NRU test methods may save a considerable number of animals since animal savings for the validation study were highest for the least toxic substances.

10.6   Summary

Computer simulation modeling of UDP testing using the default dose progression shows that, for the subset of reference substances evaluated, the prediction of starting doses using the 3T3 and NHK NRU test methods with the RC rat-only millimole regression resulted in a statistically significant (p <0.05) decrease in the number of animals used by an average of 0.49 (6.2%) to 0.54 (5.8%) animals per test, depending upon the *in vitro* NRU test method and the dose-mortality slope (2.0 or 8.3) used. The mean animal savings improved slightly, to 0.54 (6.8%) to 0.66 (7.0%) animals per test, when the RC rat-only weight regression was used.

When reference substances were grouped by GHS category, there were no mean animal savings by simulated UDP testing for substances with $50 < LD_{50} \leq 300$ mg/kg. The highest, and statistically significant, animal savings were observed with both *in vitro* NRU test methods when testing substances with $2000 < LD_{50} \leq 5000$ mg/kg and $LD_{50} > 5000$ mg/kg. When using the RC rat-only millimole regression, animal savings for these categories ranged from 1.28 (11.9%) to 1.58 (20.3%) animals per test. The use of the RC rat-only weight regression improved animal savings slightly for the substances in these toxicity categories to 1.28 (14.0%) to 1.65 (16.7%) animals per test. Although the use of IC_{50} values to estimate starting doses for the simulated UDP deceased the number of animals used per test, it did not change the number of animals that would have died during the procedures.

Computer simulation modeling of ATC testing showed that, for the reference substances tested in this validation study, the prediction of starting doses using the 3T3 and NHK NRU test methods with the RC rat-only millimole regression resulted in a statistically significant (p <0.05) decrease in the number of animals for ATC testing by an average of 0.51 (4.8%) to 0.80 (7.3%) animals per test, depending upon the *in vitro* NRU test method and the dose-mortality slope (2.0 or 8.3) used. Animal savings improved to a mean of 0.91 (8.6%) to 1.09 (10.2%) animals per test when the RC rat-only weight regression was used.

When test substances were grouped by GHS category, the mean animal savings for ATC testing using the RC rat-only millimole regression were statistically significant with the 3T3 NRU test method at both dose-mortality slopes for substances with $5 < LD_{50} \leq 50$ mg/kg (1.15 [9.8%] to 1.17 [10.2%] animals per test), and for substances with $LD_{50} > 5000$ mg/kg (2.03 [17.1%] to 2.19 [18.3%] animals per test). Significantly more animals were needed when the 3T3-based starting doses were used, than the default starting dose for reference substances with $300 < LD_{50} \leq 2000$ mg/kg (i.e., the animal savings were negative: -0.92 [-9.5%] to -1.30 [-14.0%] animals). The mean animal savings with the NHK NRU test method and the RC rat-only millimole regression were statistically significant at both dose-mortality slopes for substances with $5 < LD_{50} \leq 50$ mg/kg (1.18 [10.2%] to 1.33 [11.4%] animals per test), and for substances with $LD_{50} > 5000$ mg/kg (2.43 [20.5%] to 2.66 [22.2%] animals per test). When the RC rat-only weight regression was used, statistically significant savings in animals used were observed with both *in vitro* NRU test methods and dose-mortality slopes for substances with $5 < LD_{50} \leq 50$ mg/kg (1.25 [10.8%] to 1.51 [13.0%] animals per test), and for substances with $LD_{50} > 5000$ mg/kg (2.94 [24.8%] to 3.33 [27.7%] animals per test). The use of IC_{50} values to estimate starting doses for the ATC refined animal use by producing

approximately 0.5 to 0.6 fewer mean animal deaths per test than when the default starting dose of 300 mg/kg was used.

The use of the IC_{50}-based starting doses did not significantly alter the GHS category outcomes of the simulated UDP or ATC when compared with the outcomes based on the default starting dose. The concordance for GHS acute oral toxicity category for the IC_{50}-based starting dose with the default starting dose was 97 to 99% for both *in vitro* NRU methods and IC_{50}-LD_{50} regressions evaluated.

The magnitude of animal savings did not correlate with the accuracy of GHS categorization yielded by the NRU-predicted LD_{50} values (using the *in vitro* NRU IC_{50} values in the IC_{50}-LD_{50} regressions) or with the accuracy of GHS category outcomes since the accuracy and animals savings analyses used different standards for comparison.

The specific animal savings using the 3T3 and NHK NRU test methods with the RC rat-only regressions apply only to the reference substances evaluated in this validation study, and may not be broadly applicable to other substances. Spielmann et al. (1999) indicated that 76% (845/1115) of the industrial substances submitted to the Federal Institute for Health Protection of Consumers and Veterinary Medicine in Berlin, Germany, since 1982 had LD_{50} >2000 mg/kg. The extent to which these substances represent the population of substances in commerce is not known. However, if the results of the validation study are broadly applicable to substances to be tested in the future, and if such substances are relatively nontoxic, the selection of starting doses using the *in vitro* NRU test methods may save a considerable number of animals since animal savings for the validation study were highest for the least toxic substances.

[This Page Intentionally Left Blank]

11.0   PRACTICAL CONSIDERATIONS

The 3T3 and NHK NRU test methods are proposed as adjuncts, rather than replacements for, *in vivo* acute oral toxicity assays. Data from these *in vitro* basal cytotoxicity test methods are used with a linear regression model to predict the rat acute oral LD_{50} of the test substance, which is then used to determine the starting dose for subsequent rat acute oral toxicity tests, as described in **Sections 10.2.2** and **10.3.2**. This section discusses practical issues involved in using these two *in vitro* NRU test methods for predicting starting doses for rat acute oral toxicity tests. Practical issues that need to be considered with respect to the implementation of these cell culture methods include the need for, and availability of, specialized equipment, personnel training and expertise requirements, cost considerations, and time expenditures.

11.1   Transferability of the 3T3 and NHK NRU Test Methods

Transferability of a test method is defined as the ability of a method or procedure to be accurately and reliably performed in different, competent laboratories (ICCVAM 2003). Accuracy and reliability of these NRU test methods are discussed in **Sections 6** and **7**, respectively.

Protocols for the 3T3 and NHK NRU test methods, including solubility testing, and prequalification of keratinocyte growth medium, have been optimized and are available on the ICCVAM/NICEATM website (http://iccvam.niehs.nih.gov/methods/invitro.htm). The protocols were designed with GLP-compliance in mind and can be easily implemented or adapted by scientists with the appropriate technical experience.

Although the *in vitro* and *in vivo* test methods require some similar, general laboratory skills (e.g., preparation of solutions and test substance doses, record keeping), *in vitro* testing requires skills specific to cell culture systems (e.g., aseptic techniques, microscopic evaluation of cell cultures, propagation of cells in medium) but not to the maintenance, handling, or treatment of rodents.

11.1.1   Facilities and Major Fixed Equipment
The following lists of facility requirements, equipment and supplies, and training and expertise are common to most *in vitro* mammalian cell culture laboratories. Required equipment and supplies are also described in detail in the validation study 3T3 and NHK protocols (**Appendices B** and **C**), the *Guidance Document* (ICCVAM 2001b), and Hartung et al. (2002).

11.1.1.1 *Facility Requirements*
The testing facility should be appropriate for operating a scientific laboratory (e.g., laboratory space, air handling procedures, access to utilities, shipping/receiving department [for appropriate receipt and handling of cell culture materials], etc.). Each facility should provide:

- Adequate facilities, equipment, and supplies
- Proper health and safety guidelines
- Satisfactory quality assurance procedures

Each facility should conform to all appropriate statutes (i.e., local, state, provincial, federal, national, international) concerning safety guidelines (e.g., general workplace safety

guidelines, chemical handling and disposal guidelines, biohazard guidelines). Hartung et al. (2002) provides recommended safety guidelines for working with potentially infectious materials (e.g., HIV, hepatitis B, hepatitis C) and human materials (e.g., cells, tissues, fluids).

11.1.1.2 *Cell Culture Laboratory*
The testing facility should have a designated cell culture laboratory to ensure that *in vitro* cytotoxicity assays are performed under clean and proper aseptic conditions. The dedicated laboratory should be located such that through traffic is minimal to reduce possible disturbances that can lead to contamination which could compromise the cell culture assays. The room temperature of the laboratory should be regulated, monitored, and documented. Access to the laboratory and its supplies and test chemicals should be restricted to appropriate personnel.

11.1.1.3 *Major Equipment*
Each testing facility should have at a minimum the following equipment:
- Incubator (37ºC ± 1ºC, 90% ± 10% humidity, 5.0% ± 1% CO_2/air)
- Laminar flow clean bench/cabinet (standard: "biological hazard")
- Inverted phase contrast microscope
- 96-well plate spectrophotometric plate reader equipped with 540 nm ± 10 nm filter (if testing in 96-well plates)
- Autoclave
- Refrigerator
- Freezer (-70ºC)
- Cryogenic (liquid nitrogen) freezer/storage unit
- Computer

Equipment maintenance and calibration should be routinely performed and documented according to GLP guidelines and testing facility SOPs.

11.1.2 Availability of Other Necessary Equipment and Supplies

11.1.2.1 *General Equipment*
Each testing facility should have at a minimum the following equipment:
- Low speed centrifuge
- Adjustable temperature waterbath
- Pipettors
- Balance
- pH meter
- Cell counting system
- Water bath sonicator
- Magnetic stirrer
- Vortex mixer
- Antistatic bar ionizer (for reduction of static on tissue culture plates)

Equipment maintenance and calibration should be routinely performed and documented as per GLP guidelines and testing facility SOPs. The types of equipment listed in this section are available from scientific and laboratory supply companies (e.g., Thomas Scientific - http://www.thomassci.com/index.jsp; Fisher Scientific - https://www.fishersci.com/).

11.1.2.2 *Cell Culture Materials and Supplies*
The following supplies are needed for the *in vitro* NRU test methods. Specific product and private company names are provided either as an identification of actual materials/brands used in the validation study or as examples. Mention of these names does not imply endorsement of the product or company.

- Tissue culture plasticware (flasks [e.g., 25 cm^2, 75-80 cm^2], 96-well plates, disposable pipettes)
- Laboratory glassware (e.g., flasks, bottles, graduated cylinders)
- Adhesive film plate sealers (e.g., Excel Scientific SealPlate™)
- Sterile filtration systems (e.g., vacuum filtration units with 0.22 μm and 0.45 μm sterile filters)
- Culture medium and supplements (e.g., DMEM; prequalified NHK medium)
- NCS (bovine)
- Balanced salt solutions (e.g., HBSS, D-PBS)

Cell culture supplies are generally available through the major scientific and laboratory supply companies and through specialty companies (e.g., GIBCO, SIGMA-Aldrich, CAMBREX/Biowhittaker, Becton Dickinson). Compositions of culture media, supplements/additives, salt solutions, NRU assay chemicals, and the volumes of each needed for each test method, should be defined. All tissue culture flasks and dishes needed to assure proper cell propagation should be identified.

11.1.2.3 *Cell Cultures*
3T3 Mouse Fibroblasts: BALB/c 3T3 cells, clone 31, can be obtained from national/international cell culture repositories (e.g., American Type Culture Collection [ATCC], Manassas, VA, product # CCL-163).

NHKs: These non-transformed keratinocyte cells from cryopreserved primary or secondary cells can be obtained from national/international cell culture repositories (e.g., CAMBREX Bio Science, 8830 Biggs Ford Road, Walkersville, MD), or isolated from donated tissue using proper collection, preparation, and propagation techniques. It may be difficult, at times, to obtain adequate supplies of keratinocytes; the preparation of a pool of cells depends on the availability of tissue donors. It is recommended that testing laboratories procure of a commercially available stock pool of cells and store them indefinitely in a cryogenic freezer.

All cell stock and cultures used for testing must be certified as free of contamination by mycoplasma and bacteria.

11.1.3   Problems Specific to the NHK NRU Test Method
FAL had difficulty obtaining an adequate supply of NHK medium during the validation study. Communication between the UK distributor and the laboratory was uneven and the SMT attempted to resolve the supply issue on several occasions. The other laboratories periodically had difficulties in obtaining NHK medium and supplements that adequately supported keratinocyte growth. Although the purchased medium and supplements met the manufacturer's QA/QC standards, certain lots of the medium and supplements did not support the growth of NHK cells to the extent needed in the test protocol. To deal with these problems, an NHK medium prequalification protocol was incorporated into the study to

avoid unnecessarily repeating studies because of medium and supplements that did not adequately support cell growth. These experiences illustrate the need for multiple sources of keratinocyte cell culture medium. They also suggest that the NHK results could be more variable than the 3T3 results because of the batch-to-batch differences in NHK growth medium and supplements.

11.2   3T3 and NHK NRU Test Method Training Considerations

The ECVAM Good Cell Culture Practice Task Force Report 1 (Hartung et al. 2002) encouraged the establishment of practices and principles that will reduce uncertainty in the development and application of *in vitro* test methods. Training in good cell culture practices, in conjunction with good laboratory practices, are essential for all *in vitro* cytotoxicity testing and should be employed to ensure that data produced from the 3T3 and NHK NRU test methods are reproducible, credible, and acceptable.

In vitro cytotoxicity test methods require personnel trained specifically in sterile tissue/cell culture techniques and general laboratory procedures. Personnel should have mandatory training in good cell culture practices, in the specialized culture procedures needed for these assays, and in safety and handling practices appropriate to the types of substances that may be tested in the laboratory (Hartung et al. 2002).

The facility management should establish scientific guidelines and procedures, train and supervise professional and technical staff, and evaluate results and performance within their discipline area relative to the testing requirements. Performance of the tests requires a moderate degree of technical capability and a high degree of skill in monitoring and maintaining appropriate cell growth conditions, troubleshooting the potential and real problems in culture systems, and analyzing and interpreting *in vitro* cytotoxicity data. Each individual engaged in the conduct of a study, or responsible for its supervision, shall have education, training, and experience, or combination thereof, to enable that individual to perform the assigned duties. The NRU test methods do not require that personnel be trained to perform *in vivo* testing.

11.2.1   Required Training and Expertise
Personnel performing *in vitro* testing should have training in basic cell culture aspects such as: sterile technique, handling culture media, feeding cultures, cell counting, subculture (trypsinization), detection and elimination of contamination, cell growth and measurement of growth curves, viability assays, and storage and freezing/thawing of cells. Additionally, training is encouraged for special culture procedures such as primary cell and tissue cultures, toxicity testing, and viability assays. Laboratory personnel should be trained in the application of GLP requirements (see **Section 8.1.1**), and in the safe storage, handling, and disposal of toxic substances.

11.2.1.1 *Specific Training and Expertise Needed*
Personnel performing the *in vitro* cytotoxicity test methods should be well experienced in general cell culture techniques and should be able to:
- Work with cryogenic freezing apparatus
- Pipette solutions with large volume pipettors and multi-channel pipettors

- Establish cells in culture vessels under aseptic conditions and monitor growth; recognize normal and abnormal cell growth characteristics; and document observations of cell cultures throughout all aspects of the procedure
- Perform the *in vitro* assays by following the protocols to grow the cells, count, transfer, and feed the cells, treat the cells with test substances, perform application of adhesive plate sealers to culture plates for control of volatile substances, perform the NRU assay, perform optical density measurements, transfer data to electronic templates
- Operate equipment necessary for maintaining cell culture laboratories (e.g., incubators, biohazard hoods, spectrophotometric microtiter plate readers)

11.2.1.2   *General Laboratory Expertise Needed*
Personnel should also be able to understand and perform basic laboratory techniques and laboratory management:

- Prepare cell culture solutions (e.g., culture medium, NRU solutions), measure pH, know proper storage conditions, and maintain proper documentation
- Prepare test substances for application to cell cultures, follow solubility protocols to adequately prepare test chemicals in solution, recognize solubility issues (e.g., insolubility nature of chemical, precipitation), and implement procedures for dissolving the test chemicals
- Monitor and control laboratory environment (e.g., temperature, humidity, lighting, traffic), maintain equipment to support cell cultures (e.g., temperature, humidity, gas flow, calibrations)

11.2.2   <u>Training Requirements to Demonstrate Proficiency</u>
Laboratories establish their own criteria for proficiency but, over the course of training, laboratory personnel should be able to understand the protocol, perform the protocol with guidance from an experienced supervisor/trainer and, eventually, perform the protocol with minimal or no supervision. An experienced supervisor determines when a technician is adequately trained because there are no standardized criteria or tasks that can be used to accurately measure competence. After the technician demonstrates competence in executing all the aspects of the test protocols(s), it is appropriate to perform routine assessments of technical competence using a benchmark, coded control test substance (e.g., SLS). It is essential that the laboratory staff be certified as proficient in using the test methods to test unknowns.

The laboratories in the validation study were selected because of their experience in performing *in vitro* cytotoxicity assays but were required to develop additional skills through Phases I and II (e.g., data collection and transfer to Excel® and PRISM® templates). Inexperienced laboratory personnel were trained by having them perform "training" assays using SLS. In the early phases of the validation study, the laboratories continued training by testing coded reference substances of various toxicities, and performing solubility testing on substances of varying solubilities. These procedures helped improve proficiency among the laboratories for the final phase of the validation study.

11.2.2.1 *Proficiency With GLP-Compliance*

Results from these test methods will be submitted to regulatory agencies that will, for the most part, require GLPs. Laboratories should work toward attaining GLP compliance. GLP compliance in each laboratory is determined by its independent QA unit. ECBC and IIVS conducted this validation study in compliance with GLP (see **Section 8.1.1**). Their respective QA units (as per GLPs) reviewed the various aspects of the study and issued QA statements that addressed whether the test methods and the results described in the Final Report accurately followed the test protocol and reflected the raw data produced during the study, and provided assurance that all testing was done under according to GLP. FAL (which was non-GLP-adherent) followed the GLP standards referenced in **Section 8.1.1** as guidelines for conducting this study. FAL had no QA unit to judge GLP compliance.

11.2.3 Personnel Needed to Perform the *In Vitro* NRU Test Methods

The facility management will be responsible for determining which qualified personel meet the criteria (e.g., scientific knowledge, specialized training) for the following positions needed for adequate performance of the *in vitro* NRU test methods and oversight of the testing.

- Study Director: the individual with the overall responsibility for the technical conduct of the testing (e.g., is familiar with the test procedures, provides SOPs and ensures GLP compliance, analyzes and interprets the data, determines test acceptance, oversees recordkeeping procedures, and produces the test reports.
- Quality Assurance Officer: monitors the testing to assure conformance with GLP requirements; must be independent of the Study Director.
- Laboratory Technician(s): individuals trained in sterile tissue/cell culture techniques and general laboratory procedures and who are capable of performing the test methods according to GLPs.

11.3 Cost Considerations

11.3.1 3T3 and NHK NRU Test Methods

11.3.1.1 *Equipment Costs*

Major instruments and equipment needed to implement the *in vitro* cytotoxicity test methods are described in **Section 11.1.1**. Ranges of costs for some of the equipment were obtained from on-line catalogues for two major scientific equipment and supplies companies (Thomas Scientific - http://www.thomassci.com/index.jsp; Fisher Scientific - https://www.fishersci.com/). These prices are for equipment that will meet the minimum needs of the NRU test methods (see **Table 11-1**). These costs were researched in August 2006.

11.3.1.2 *Costs for Cell Cultures and Supplies*

Supplies such as cell culture chemicals, the reagents used to measure NRU, and cell culture plasticware are available from numerous suppliers, and are not cost prohibitive.

Table 11-1   Costs for Major Laboratory Equipment

Equipment	Range of Costs[1]
Class II Biological Safety Cabinet	$7,300 – $12,200
CO_2 Incubator	$5,100 – $16,400
Spectrophotometer Microplate Reader	$5,000 – $7,500
Freezer (capable of -70°C)	$8,000 – $15,300
Refrigerator	$1,300 – $9,800
Centrifuge (benchtop model)	$2,100 – $8,500
Microscope (inverse phase contrast)	$3,000 – $14,500
Coulter Counter[2, 3]	$3,000 – $9,000
Autoclave (benchtop model)[2]	$3,500 – $15,400
Cryogenic (liquid nitrogen) Storage	$1,000 – $3,700

[1]From on-line scientific equipment catalogues (Thomas Scientific - http://www.thomassci.com/index.jsp; Fisher Scientific - https://www.fishersci.com/). [searched August 2006]
[2]May be useful, but not required for performing the tests.
[3]Other automatic cell counters may be used.

The 3T3 NRU test method is generally less expensive to perform than the NHK NRU test method. One vial of the immortalized 3T3 cells (~$200 [ATCC]) can be propagated indefinitely by passaging cells and periodically cryopreserving batches of cells. The NHK NRU test method requires a fresh sample of primary cells for each test run (~$380 per vial [CAMBREX]). Because primary NHK cells are passaged only once after initiating the culture, there are no cells available to cryopreserve a stock batch of cells. The DMEM medium used for the 3T3 cells is less expensive, more "generic", and more readily available than keratinocyte-specific NHK medium. (See **Table 11-2.**)

11.3.1.3   *Commercial Testing*
The following price quotes are provided as examples of test costs and were acquired from commercial laboratories through Internet contact or through personal communication. Use of information from these specific laboratories does not imply endorsement of them.

A representative of MB Research Laboratories (Spinnerstown, PA, http://www.mbresearch.com/) provided a quote (personal communication, 2005) for an *in vitro* 24-hr cytotoxicity test (but not a 48-hour test period) of $1050 (USP standards[1]) or $1950 (ISO standards[1]) for a set of three test chemicals. The lead laboratory for the NICEATM/ECVAM study, IIVS (Gaithersburg, MD, http://www.iivs.org/) provides

---

[1] USP=United States Pharmacopeia; ISO=International Standards Organization. These organizations provide international standard testing requirements for products that require high quality for public use.

commercial laboratory GLP-compliant testing using this study's protocols (48-hour test period) at a cost of $1120 - $1850 per chemical/sample for one cell type (personal communication 2005) (see **Table 11-2**).

Table 11-2   Costs for Cell Culture Materials and Commercial Laboratory *In Vitro* Cytotoxicity Testing

Item	Cost (approximate)	Number of Tests Possible	Other
3T3 Cells	~$200/vial[1]	indefinite	One vial can produce an indefinite supply of cells by propagating the cells in culture and periodically freezing a pool of cells.
NHK Cells	~$380/vial[2]	~5 (96-well plates)	Since cells are passaged only once beyond cryopreservation, new vials should be thawed as needed to maintain continuous testing.
Dulbeccos' Minimum Essential Medium (D-MEM) with supplements	~$20/500mL[3]	~15 (96-well plates)	Establish cells in culture (~20 mL/vial of cells; 60 mL/3 vials), seed cells in 96-well plates (12 mL/plate; 180 mL/15 plates); prepare stock solution and eight concentration dilutions (~20 mL/chemical; 300 mL/15 plates).
NHK Medium with supplements	~$80/500 mL[2]	~15 (96-well plates)	Same as DMEM (above)
Commercial Laboratory Testing (MB Research Laboratories [GLP-compliant])	$1050/$1950 (USP/ISO) per 3 test materials[4]	1 test/material	*in vitro* NRU cytotoxicity test (24-hour test period)
Commercial Laboratory Testing (Institute for *In Vitro* Sciences [GLP-compliant])	$1120 (GLP) per test material (minimum of 5 materials tested simultaneously)[4]	1 range finder, 2 definitive tests per test material	*in vitro* NRU cytotoxicity test (48-hour test period)
Commercial Laboratory Testing (Institute for *In Vitro* Sciences)	$1850 (GLP) per single test material (tested individually)[4]	1 range finder, 2 definitive tests per test material	*in vitro* NRU cytotoxicity test (48-hour test period)

Abbreviations: 3T3=BALB/c 3T3 fibroblasts; NHK=Normal human epidermal keratinocytes; NRU=Neutral red uptake; USP/ISO= United States Pharmacopeia/International Standards Organization GLP=Good Laboratory Practices
[1]Catalogue price from American Type Culture Collection (ATCC) (http://www.atcc.org/)
[2]Catalogue price from CAMBREX (http://www.cambrex.com/Welcome.asp)
[3]Catalogue price from INVITROGEN (http://www.invitrogen.com/content.cfm?pageid=1)
[4]Personal communication (Raabe 2005)

11.3.2   Rodent Acute Oral Toxicity Testing

As stated in **Section 11.3.1.3**, presentation of price quotes from commercial laboratories provides examples of test costs and does not imply an endorsement of that laboratory. **Table 11-3** provides some commercial prices for acute oral systemic toxicity testing. MB Research Laboratories performs the UDP test at a cost of $750 for three rats and charges $250 for each additional rat needed. In the best-case scenario, the UDP test needs only three rats ($750). In the worst-case scenario, this test would need an additional 12 rats (15 maximum for the test); the total cost of the test would be $3,750. In this costing strategy, $250 is saved for each rat not used by an accurate prediction of the starting dose by the 3T3 or NHK NRU test method. Because the *in vitro* cytotoxicity test costs from $350 to $1850 per chemical, there is no net savings in animal costs if fewer than two to six animals are saved.

Table 11-3   Commercial Prices for Conducting *In Vivo* Acute Rat Toxicity Testing

Test	GLP-Compliant	Non GLP-Compliant	Company
Acute Oral Toxicity UDP: Limit Test - 2000 mg/kg	$1200	$1000	Product Safety Laboratories
Acute Oral Toxicity UDP: Limit Test - 5000 mg/kg	$800	$650	Product Safety Laboratories
Acute Oral Toxicity UDP: LD_{50}	$2700	$2200	Product Safety Laboratories[1]
Acute Oral Rat Toxicity: single dose[2]	$950	NA	Bio Research Laboratories
Acute Oral Rat Toxicity: two doses[2]	$1500	NA	Bio Research Laboratories
Acute Oral Rat Toxicity: LD_{50}	$3000	NA	Bio Research Laboratories
Acute Oral Toxicity – UDP	$730 for the first 3 animals; $250 each additional animal	NA	MB Research Laboratories[1]

Abbreviations: UDP=Up-and-Down Procedure; GLP=Good Laboratory Practices; NA=Not available.
[1]Personal communication (Wnorowski 2005).
[2]Washington State Biological Testing Methods #80-12 For the Designation of Dangerous Waste; Part B: Acute Oral Rat Toxicity Test [http://www.ecy.wa.gov/pubs/80012.pdf]. This test method is an adaptation of the EPA Health Affects Test Guidelines OPPTS 870.110 Acute Oral Toxicity and American Society for Testing and Materials (ASTM) methods E 1163-90 (Standard test method for estimating acute oral toxicity in rats) and E 1372-90 (Standard test method for conducting a 90-day oral toxicity study in rats).

The President of Product Safety Laboratories, Gary Wnorowski, (Dayton, NJ, http://www.productsafetylabs.com/), provided a cost quote of $2700 for determination of a rat LD_{50} value using the UDP test; the cost is independent of the number of rats that are needed. Each test dose is administered ~24-48 hours after the previous dose and each animal test generally does not exceed four days. The time involved in providing the LD_{50} value is approximately three months (initiation of the test to provision of the final report). Having the estimated LD_{50} value would not affect the cost of the *in vivo* test but could reduce the number of animals needed.

Bio Research Laboratories performs the rat acute oral toxicity test using a test method that determines lethality and signs of acute toxicity from a waste sample administered in a single dose, by gavage, to a limited number of rats. The bioassay determines if the test sample

produces an LD$_{50}$ either greater than or less than a regulatory threshold corresponding to a hazardous waste designation (i.e., 5000, 500, 50 mg/kg). A minimum of 10 rats is used at the tested dose for the regulatory threshold value that is relevant to the test sponsor. In this testing scenario, knowledge of the estimated LD$_{50}$ would not reduce animal use or test costs if a single predetermined dose is tested.

11.4   Time Considerations for Performing the 3T3 and NHK NRU Tests

11.4.1   The 3T3 NRU Test Method
Approximately one week is needed to thaw cryopreserved 3T3 cells, propagate them, and passage them at least two times before subculturing them into the 96-well test plates. After subculture into 96-well plates, the cells are incubated another 24 hours to reach the proper confluence, and then exposed to test chemical for 48 hours. The initial 3T3 NRU test (range finder or definitive test) takes approximately 10 days. However, after the cells are established in culture, they can be passaged for approximately two months before having to go back to the cryopreserved cells to start a new culture. A 3T3 NRU test can be completed in less than four consecutive days when started from an established stock culture. Multiple substances can be tested at the same time, and different tests can overlap each other; thus, many substances can be tested in a relatively short time.

11.4.2   The NHK NRU Test Method
Approximately one week is needed to thaw cryopreserved NHK cells, propagate them, and passage them into the 96-well test plates. After subculture into 96-well plates, the cells are incubated another 48-72 hours to reach the proper confluence and then exposed to test chemical for 48 hours. The entire NHK NRU test (range finder or definitive test) requires approximately 11-12 days. Cells can be seeded at different densities from one starter vial in the culture flasks so that passaging the cultures can take place on different days. Once the cells are established in culture, they are passaged once to the 96-well test plates and an NHK NRU test can usually be completed in five to six consecutive days. Multiple substances can be tested at the same time, and different tests can overlap each other; thus, many substances can be tested in a relatively short time.

11.4.3   Prequalification of NHK Medium
The protocol for the prequalification of NHK medium requires nearly identical steps, and similar time-line (i.e., 11-12 days), as required for the NHK rangefinder and definitive tests. **Table 11-2** provides an estimate of how many tests could be performed using one 500 mL bottle of medium with supplements (~15 tests in 96-well plates).

11.4.4   *In Vivo* Testing
According to guidelines for acute oral toxicity testing, single animals or groups of animals are dosed in sequence, usually at 2-4 day intervals, and observations are generally made for up to 14 days (for animals that are not moribund) for the main test and limit dose test (EPA 2002a; OECD 2001a; OECD 2001b, OECD 2001c). The addition of 3T3 or NHK NRU testing to estimate a starting dose prior to the implementation of the UDP main test or limit dose test would take 10-12 days, but could save up to 14 days of observation for every animal not used.

11.4.5   The Limit Test

The *in vitro* NRU test methods can provide a savings of time when used to determine if an *in vivo* acute oral toxicity limit test can be employed as the initial test for a substance with unknown *in vivo* toxicity. If the IC_{50} value from an *in vitro* NRU test could accurately predict an LD_{50} that is greater than, or equal to, the limit dose (i.e., 2000 mg/kg or 5000 mg/kg), the *in vivo* test could start at the limit test dose. This approach has the potential to eliminate the need to do the main test and could result in a net savings of six days for the UDP test method and about one day for the ATC test method. **Table 11-4** illustrates the following:

- Time needed to perform the 3T3 and NHK NRU test
- Time needed to reach the limit test starting dose when initiating the *in vivo* main test using the default starting doses (UDP and ATC)

The times presented in **Table 11-4** use the following assumptions:
- 3T3 cells reach ≤50% confluence in approximately 24 hours
- NHK cells reach >20% confluence in approximately 48 hours
- Animals show no evident toxicity 48 hours post-dosing, and additional animals are dosed at the next higher default dose
- Limit test dose = 5000 mg/kg for the UDP and 2000 mg/kg for the ATC method

Table 11-4   Comparison of Time Needed for *In Vitro* and *In Vivo* Testing

Time	3T3 NRU Test Method	NHK NRU Test Method	UDP (5000 mg/kg upper limit)	ATC (2000 mg/kg upper limit)
Day 1	Seed cells in 96-well plate Incubate for 24 ±2 hr	Seed cells in 96-well plate Incubate for approximately 48 to 72 hr	Dose 1 animal at default dose (175 mg/kg) Observe for 48 hr	Dose 3 animals at default dose (300 mg/kg) Observe for 48 hr
Day 2	Apply test substance Incubate for 48 ±0.5 hr	Incubate	Observe	Observe
Day 3	Incubate	Apply test substance Incubate for 48 ±0.5 hr	No death Dose 1 animal at next default dose (550 mg/kg) Observe 48 hr	0 – 1 animal dies Dose 3 animals at default dose (300 mg/kg) Observe 48 hr
Day 4	NRU: 3 ±0.1 hr Elute NR: 0.33 to 0.75 hr OD_{540} measurement Calculate IC_{50} Estimate LD_{50} and Starting Dose*	Incubate	Observe	Observe
Day 5		NRU: 3 ±0.1 hr Elute NR: 0.33 to 0.75 hr OD_{540} measurement Calculate IC_{50} Estimate LD_{50} and Starting Dose*	No death Dose 1 animal at next default dose (1750 mg/kg) Observe 48 hr	0 – 1 animal dies Dose 3 animals at next default dose (2000 mg/kg) **Starting Point for the Limit Test**
Day 6			Observe	
Day 7			No death Dose 1 animal at next default dose (5000 mg/kg) **Starting Point for the Limit Test**	

Abbreviations: 3T3=BALB/c 3T3 fibroblasts; NHK=Normal human epidermal keratinocytes; NRU=Neutral red uptake; UDP=Up-and-Down Procedure; ATC=Acute Toxic Class method; hr=Hour; NR=Neutral red; OD_{540}=Optical density at 540 nm.

11.5   Summary

- All equipment and supplies should be readily commercially available. During the validation study, direct communication with the NHK medium supplier insured that specific lots of medium were available to the laboratories. The test methods are expected to be transferable to laboratories experienced with mammalian cell culture methods.

- Much of the training and expertise needed to perform the 3T3 and NHK NRU test methods are common to mammalian cell culture procedures. Additional technical training would not be extensive because these test methods are similar to other *in vitro* mammalian cell culture assays, and no extraordinary techniques are necessary. GLP training should be provided to technicians to ensure proper adherence to protocols and documentation procedures.

- Prices for commercial testing for one chemical are $1,120 to $1,850 (**Table 11-2**) for *in vitro* cytotoxicity testing in the 3T3 and NHK test methods, respectively, to determine the IC_{50} (Raabe 2005, personal communication). In contrast, the *in vivo* rat acute oral testing for LD_{50} determination could cost from $750 - $3,750 (**Table 11-3**), depending on the test method used and the toxicity of the test substance. Comparison of costs of *in vitro* testing to *in vivo* testing is difficult because the *in vitro* NRU test methods are not replacements for the animal testing, and animal testing would be performed regardless of the responses of the 3T3 or NHK cells. The use of these *in vitro* NRU test methods may not reduce the overall cost of the *in vivo* rat acute oral toxicity test, but has the potential to reduce the number of animals needed for a study.

[This Page Intentionally Left Blank]

12.0   REFERENCES

Adegunloye BI, Sofola OA. 1997. Differential responses of rat aorta and mesenteric artery to norepinephrine and serotonin *in vitro*. Pharmacology 55:25-31.

Ames A III. 2000. CNS energy metabolism as related to function. Brain Res Rev 34:42-68.

Anon. 1993. Annexes I, II, III, and IV to Commission Directive 93/21/EEC of 27 April 1993 adapting to technical progress for the 18th time Council Directive 67/548/EEC on the approximation of laws, regulations and administrative provisions relating to the classification, packaging and labeling of dangerous substances. Official Journal of the European Communities L110A, 1-86.

ASTM. 1999. American Society of Testing and Materials (ASTM) Standard E691-99, Standard Practice for Conducting an Interlaboratory Study to Determine the Precision of a Test Method. American National Standards Institute. Washington, D.C. Available: http://www.ansi.org/ [accessed 01 November 2006].

ATSDR. 1999a. Toxicological Profile for Cadmium. Update PB/99/166621. U.S. Dept. of Health and Human Services. Atlanta, GA:Agency for Toxic Substances and Disease Registry. Available: http://www.atsdr.cdc.gov/toxpro2.html [accessed 01 November 2006].

ATSDR. 1999b. Toxicological Profile for Diethyl phthalate. PB/95/264214/AS. U.S. Dept. of Health and Human Services. Atlanta, GA:Agency for Toxic Substances and Disease Registry. Available: http://www.atsdr.cdc.gov/toxpro2.html [accessed 01 November 2006].

ATSDR. 2000a. Toxicological Profile for Endosulfan. Update. PB/2000/108023. U.S. Dept. of Health and Human Services. Atlanta, GA:Agency for Toxic Substances and Disease Registry. Available: http://www.atsdr.cdc.gov/toxpro2.html [accessed 01 November 2006].

ATSDR. 2000b. Toxicological Profile for Chromium. Update. PB/2000/108022. U.S. Dept. of Health and Human Services. Atlanta, GA:Agency for Toxic Substances and Disease Registry.Available: http://www.atsdr.cdc.gov/toxpro2.html [accessed 01 November 2006].

ATSDR. 2001. Toxicological Profile for Di-n-butyl phthalate. Update. PB/2001/109104/AS. U.S. Dept. of Health and Human Services. Atlanta, GA:Agency for Toxic Substances and Disease Registry. Available: http://www.atsdr.cdc.gov/toxpro2.html [accessed 01 November 2006].

ATSDR. 2004a. Toxicological Profile for Pyrethrins and Pyrethroids. Update. PB2004-100004. U.S. Dept. of Health and Human Services. Atlanta, GA:Agency for Toxic Substances and Disease Registry. Available: http://www.atsdr.cdc.gov/toxpro2.html [accessed 01 November 2006].

ATSDR. 2004b. Toxicological Profile for Selenium. Update. PB2004-100005. U.S. Dept. of Health and Human Services. Atlanta, GA:Agency for Toxic Substances and Disease Registry. Available: http://www.atsdr.cdc.gov/toxpro2.html [accessed 01 November 2006].

Babich H, Martin-Alguacil N, Raul C, Rosenberg DW, Borenfreund E. 1991. Response of human cell cultures to cytotoxicants requiring metabolic activation. In: Alternative Methods in Toxicology (Goldberg AM, ed.), Vol. 8. New York:Mary Ann Liebert, Inc., 263-276.

Balls M, Botham PA, Bruner LH, Spielmann H. 1995. The EC/HO international validation study on alternatives to the Draize eye irritation test for classification and labelling of chemicals. Toxicol *In Vitro* 9:871-929.

Bernson V, Bondesson I, Ekwall B, Stenberg K, Walum E. 1987. A Multicentre evaluation study of *in vitro* cytotoxicity. Altern Lab Anim 14:144-146.

Bliss CI. 1938. The determination of the dosage-mortality curve from small numbers. Q J Pharm Pharmacol 11:192-214.

Bondesson I, Ekwall B, Hellberg S, Romert L, Stenberg K, Walum E. 1989. MEIC - A new international multicenter project to evaluate the relevance to human toxicity of *in vitro* cytotoxicity tests. Cell Biol Toxicol 5:331-347.

Borenfreund E, Puerner J. 1984. A simple quantitative procedure using monolayer cultures for cytotoxicity assays (HTD/ NR-90). J Tissue Culture Meth 9:7-9.

Borenfreund E, Puerner JA. 1985. Toxicity determination *in vitro* by morphological alterations and neutral red absorption. Toxicol Lett 24:119-124.

Botham PA. 2004. The validation of *in vitro* methods for skin irritation. Toxicol Lett 149:387-390.

Brantom PG, Bruner LH, Chamberlain M, De Silva O, Dupuis J, Earl LK, et al. 1997. A summary report of the COLIPA international validation study on alternatives to the Draize rabbit eye irritation test. Toxicol *In Vitro* 11:141-179.

Calabrese EJ. 2005. Cancer biology and hormesis: human tumor cell lines commonly display hormetic (biphasic) dose responses. Crit Rev Toxicol 35:463-582.

Canadian Centre for Occupational Health and Safety. 2005. IPCS INTOX Database. Available: http://www.intox.org/databank/index.htm [accessed 01 November 2006].

Casarett LJ, Klaassen CD, Doull J, eds. 2001. Casarett and Doull's Toxicology, The Basic Science of Poisons. 6th Edition. New York:McGraw-Hill.

Charles River Laboratories Catalog. 2002. Available: http://www.criver.com [accessed 01 November 2006].

Clemedson C, McFarlane-Abdulla E, Andersson M, Barile FA, Calleja MC, Chesné C, et al. 1996a. MEIC evaluation of acute systemic toxicity. Part I. Methodology of 68 *in vitro* toxicity assays used to test the first 30 reference chemicals. Altern Lab Anim 24 (suppl 1):51-272.

Clemedson C, McFarlane-Abdulla E, Andersson M, Barile FA, Calleja MC, Chesné C, et al. 1996b. MEIC evaluation of acute systemic toxicity. Part II. *In vitro* results from 68 toxicity assays used to test the first 30 reference chemicals and a comparative cytotoxicity analysis. Altern Lab Anim 24 (suppl 1):273-311.

Clemedson C, Barile FA, Ekwall Ba, Gómez-Lechón MJ, Hall T, Imai K, et al. 1998a. MEIC evaluation of acute systemic toxicity. Part III. *In vitro* results from 16 additional methods used to test the first 30 reference chemicals and a comparative cytotoxicity analysis. Altern Lab Anim 26 (suppl 1):93-129.

Clemedson C, Andersson M, Aoki Y, Barile FA, Bassi AM, Calleja MC, et al. 1998b. MEIC evaluation of acute systemic toxicity. Part IV. *In vitro* results from 67 toxicity assays used to test reference chemicals 31-50 and a comparative cytotoxicity analysis. Altern Lab Anim 26 (suppl 1):131-183.

Clemedson C, Nordin-Andersson M, Bjerregaard H.F, Clausen J, Forsby A, Gustafsson H, Hansson U, Isomaa B, Jorgensen C, et al. 2002. Development of an *in vitro* test battery for the estimation of acute human systemic toxicity: an outline of the EDIT project. Altern Lab Anim 30:313-321.

Clemedson C. 2005. A-Cute-Tox project, an integrated project under the EU 6FP with the aim to optimize and pre-validate an *in vitro* test strategy for predicting human acute toxicity. European Society of Toxicology *In Vitro* Newsletter 18:6.

Clothier R, Willshaw A, Cox H, Garle M, Bowler H, Combes R. 1999. The use of human keratinocytes in the EU/COLIPA international *in vitro* phototoxicity test validation study and the ECVAM/COLIPA study on UV filter chemicals. Altern Lab Anim 27:247-259.

Coeck S, Balls M, Bowe G, Davis J, Gstraunthaler G, Hartung T, et al. 2005. Guidance on Good Cell Culture Proactice. A Report of the Second ECVAM Task Force on Good Cell Culture Practice. Altern Lab Anim 33:261-287.

Coldwell BB, Boyd EM. 1966. The acute rectal toxicity of acetylsalicylic acid. Can J Physiol Pharmacol 44: 909-918.

Conolly RB, Lutz WK. 2004. Nonmonotonic dose-response relationships: mechanistic basis, kinetic modeling, and implications for risk assessment. Toxicol Sci 77:151-157.

Cooper-Hannan R, Harbell JW, Coecke S, Balls M, Bowe G, Cervinka, et al. 1999. The Principles of Good Laboratory Practice: application to *in vitro* toxicology studies. Altern Lab Anim 27:539-577.

Cosmetic Ingredient Review Panel. 1983. Final report on the safety assessment of sodium borate and boric acid. J Am Coll Toxicol 2:87-125.

CPSC. 2003. Commercial Practices. Hazardous Substances and Articles; Administration and Enforcement Regulations. 16CFR1500.42.

Creppy EE, Chiarappa P, Baudrimont I, Borracci P, Moukha S, Carratu MR. 2004. Synergistic effects of fumonisin B1 and ochratoxin A: are *in vitro* cytotoxicity data predictive of in vivo acute toxicity? Toxicology 201:115-23.

Curren R, Bruner L, Goldberg A, Walum E. 1998. 13th meeting of the Scientific Group on Methodologies for the Safety Evaluation of Chemicals (SGOMSEC): validation and acute toxicity testing. Environ Health Perspect 106 (suppl 2): 419-425.

Curren RD, Moyer GO, Wilt N, Clear ML, Sizemore AM, Mun G. 2003. Assessment of protocol variables in cytotoxicity assays utilizing BALB/c 3T3 cells and normal human keratinocytes [Abstract]. Toxicologist 72:157. Available: http://www.iivs.org/pages/publication_view.php?doc_id=43 [accessed 01 November 2006].

Dal Negro G, Vandin L, Bonato M, Sciuscio D. 2006. Toward refinement of the colony-forming unit-granulocyte/macrophage clonogenic assay: inclusion of a metabolic system. Toxicol *In Vitro* 20:743-9.

DOT. 2003. Shippers--General Requirements for Shipments and Packagings. Class 6, Division 6.1 − Definitions. 49CFR173.132. Available: http://www.gpoaccess.gov/cfr/index.html [accessed 01 November 2006].

DOT. 2003. Shippers--General Requirements for Shipments and Packagings. Assignment of Packing Group and Hazard Zones for Division 6.1 Materials. 49CFR173.133. Available: http://www.gpoaccess.gov/cfr/index.html [accessed 01 November 2006].

DOT. 1999. Assignment of packing group and hazard zones for Division 6.1 materials. 49CFR. SubChapter C. Available: http://www.gpoaccess.gov/cfr/index.html [accessed 01 November 2006].

Deichmann WB. 1969. Toxicology of Drugs and Chemicals. New York:Academic Press, Inc., 67.

Dixon WJ, Massey FJ. 1981. Introduction to Statistical Analysis, 4th ed. Milwaukee:Quality Press.

Ekwall B. 1983. Screening of toxic compounds in mammalian cell cultures. Ann New York Acad Sci 407:64-77.

Ekwall B, Clemedson C, Craaford B, Ekwall B, Hallander S, Walum E, Bondesson I. 1998a. MEIC evaluation of acute systemic toxicity. Part V. Rodent and human toxicity data for the 50 reference chemicals. Altern Lab Anim 26 (suppl 2):571-616.

Ekwall B, Clemedson C, Craaford B, Ekwall B, Hallander S, Walum E, Bondesson I. 1998b. MEIC evaluation of acute systemic toxicity. Part VI. The prediction of human toxicity by rodent LD50 values and results from 61 *in vitro* methods. Altern Lab Anim 26 (suppl 2):617-658.

Ekwall B, Clemedson C, Ekwall B, Ring P, Romert L. 1999. EDIT: A new international multicentre programme to develop and evaluate batteries on *in vitro* tests for acute and chronic systemic toxicity. Altern Lab Anim 27:339-349.

Ekwall B, Ekwall B, Sjostrom M. 2000. MEIC evaluation of acute systemic toxicity: Part VIII. Multivariate partial least squares evaluation, including the selection of a battery cell line tests with a good prediction of human acute lethal peak blood concentrations for 50 chemicals. Altern Lab Anim 28 (suppl 1):201-234.

EPA. 1992. Reregistration Eligibility Document Citric Acid. List D, Case 4204. Office of Pesticide Programs. Washington, DC:U.S. Environmental Protection Agency. Available: http://www.epa.gov/oppsrrd1/REDs/factsheets/4024fact.pdf [accessed 01 November 2006].

EPA. 1995. Reregistration Eligibility Decision (RED) Diquat Dibromide. EPA 738-R-95-016. Office of Prevention, Pesticides and Toxic Substances. Washington, DC:U.S. Environmental Protection Agency. Available: http://www.epa.gov/oppsrrd1/REDs/0288.pdf [accessed 01 November 2006].

EPA. 1996. Trichloroacetic acid (CASRN 76-03-9) file in the Integrated Risk Information System. Office of Research and Development. Washington, DC:U.S. Environmental Protection Agency. Available: http://www.epa.gov/iris [accessed 01 November 2006].

EPA. 1998. Product Properties Test Guidelines OPPTS 830.7840 Water Solubility: Column Elution Method; Shake Flask Method. EPA 712–C–98–041. Washington, DC:U.S. Environmental Protection Agency.

EPA. 2000a. High Production Challenge Program. Office of Pollution Prevention and Toxics (OPPT). Washington, DC:U.S. Environmental Protection Agency. Available: http://www.epa.gov/chemrtk/ [accessed 01 November 2006].

EPA. 2000b. Toxicological Review of Chloral Hydrate. EPA/635/R-00/006. Washington, DC:U.S. Environmental Protection Agency. Available: www.epa.gov/iris [accessed 01 November 2006].

EPA. 2002a. Health Effects Test Guidelines OPPTS 870.1100 Acute Oral Toxicity. EPA 712–C–02–190. Washington, DC: U.S. Environmental Protection Agency. Available: http://www.epa.gov/opptsfrs/home/testmeth.htm [accessed 01 November 2006].

EPA. 2002b. Statements of Policies and Interpretations. Toxicological and Ecological Studies. 40CFR159.165. Available: http://www.gpoaccess.gov/cfr/index.html [accessed 01 November 2006].

EPA. 2002c. Health Effects Test Guidelines OPPTS 870.1000 Acute Toxicity Testing - Background. EPA 712–C–02–189. Washington, DC: U.S. Environmental Protection Agency. Available: http://www.epa.gov/opptsfrs/home/testmeth.htm/ [accessed 01 November 2006].

EPA. 2002d. Emergency Planning and Notification. 40CFR355. Available: http://www.access.gpo.gov/nara/cfr/waisidx_00/40cfr355_00.html [accessed 01 November 2006].

EPA. 2003a. Good Laboratory Practice Standards. Toxic Substances Control Act. 40CFR792. Available: http://www.access.gpo.gov/nara/cfr/waisidx_03/40cfr792_03.html [accessed 01 November 2006].

EPA. 2003b. Good Laboratory Practice Standards. Federal Insecticide, Fungicide, and Rodenticide Act. 40CFR160. Available: http://www.access.gpo.gov/nara/cfr/waisidx_03/40cfr160_03.html [accessed 01 November 2006].

EPA. 2004. U.S. EPA HPV Challenge Program Submission. Propanoic acid, 2-hydroxy-, compd. with 3-[2-(dimethylamino)ethyl] 1-(2-ethylhexyl) (4-methyl-1,3-phenylene)bis[carbamate] (1:1) [CASRN 68227-46-3]. Submitted by PPG Industries, Inc. Revised December 17, 2004. Available: http://www.epa.gov/chemrtk/pubs/hpvrstp.htm [accessed 01 November 2006].

European Centre for the Validation of Alternative Methods. (ECVAM). 1997. Statement on the scientific validity of the 3T3 NRU PT test (an *in vitro* test for phototoxic potential). Available: http://ecvam.jrc.it/index.htm [accessed 01 November 2006].

EU. 2003. Report from the Commission to the Council and the European Parliament. Third Report from the commission to the Council and the European Parliament of the statistics on the number of animals used for experimental and other scientific purposes in the member states of the European Union. Brussels, 22.01.2003, COM (2003) 19 final.

Fautrel A, Chesné C, Guillouzo A, De Sousa G, Placidi M, Rahmani R, et al. 1993. A multicentre study of acute *in vitro* cytotoxicity in rat hepatocytes: tentative correlation between *in vitro* toxicity and *in vivo* data. Altern Lab Anim 21:281-284.

FDA. 1993. Toxicological Principles for the Safety Assessment of Direct Food Additives and Color Additives Used in Food. Center for Food Safety and Applied Nutrition. Washington, DC: U.S. Food and Drug Administration.

FDA. 2003. Good Laboratory Practice for Nonclinical Laboratory Studies. 21CFR Part 58. Available: http://www.access.gpo.gov/nara/cfr/waisidx_03/21cfr58_03.html [accessed 01 November 2006].

Fentem J, Fry J, Garle M, Gülden M, Seibert H, Voss J-U, et al. 1993. An International Evaluation of Selected *In Vitro* Toxicity Test Systems for Predicting Acute Systemic Toxicity. A report prepared for DGXI, CEC; Contract Numbers B92/B4-3063/14086 & B92/B4-3040/14087. Nottingham, UK:Fund for Replacement of Animals in Medical Experiments.

Fentem JH, Archer GEB, Balls M, Botham PA, Curren RD, Earl LK, et al. 1998. The ECVAM international validation study on *in vitro* tests for skin corrosivity. 2. Results and evaluation by the Management Team. Toxicol *In Vitro* 12:483-524.

Fentem JH, Briggs D, Chesne C, Elliott GR, Harbell JW, Heylings JR, et al. 2001. A prevalidation study on *in vitro* tests for acute skin irritation: results and evaluation by the Management Team. Toxicol *In Vitro* 15:57-93.

Freshney RI. 2000. Culture of Animal Cells: A Manual of Basic Technique, 4th edition. New York:Wiley-Liss.

Garle M., Fentem JH, Fry JR. 1994. *In vitro* cytotoxicity tests for the prediction of acute toxicity *in vivo*. Toxicol *In Vitro* 8:1303-1312.

Gennari A, van den Berghe C, Casati S, Castell J, Clemedson C, Coecke S, et al. 2004. Strategies to replace *in vivo* acute systemic toxicity testing. The report and recommendations of ECVAM Workshop 50. Altern Lab Anim 32:437-459.

Gettings SD, Bagley DM, Demetrulias JL, Dipasquale LC, Hintze KL, Rozen MG, et al. 1991. The CTFA evaluation alternatives program: An evaluation of *in vitro* alternatives to the Draize primary eye irritation test. (Phase I) Hydro-alcoholic formulations; (Part 2) Data analysis and biological significance. *In Vitro* Toxicol 4:247-288.

Gettings SD, Bagley DM, Chudkowski M, Demetrulias JL, Dipasquale LC, Galli CL et al. 1992. Development of potential alternatives to the Draize eye test - The CTFA evaluation of alternatives program - Phase II - Review of materials and methods. Altern Lab Anim 20:164-171.

Gettings SD, Dipasquale LC, Bagley DM, Casterton PL, Chudkowski M, Curren RD, et al. 1994a. The CTFA evaluation of alternatives program: An evaluation of *in vitro* alternatives to the Draize primary eye irritation test. (Phase II) Oil/water emulsions. Food Chem Toxicol 32:943-976.

Gettings SD, Hintze KL, Bagley DM, Casterton PL, Chudkowski M, Curren RD, et al. 1994b. The CTFA evaluation of alternatives program: Phase III (surfactant-based formulations). *In Vitro* Toxicol 7:166.

Gettings SD, Lordo RA, Hintze KL, Bagley DM, Casterton PL, Chudkowski M, et al. 1996. The CFTA evaluation of alternatives program: An evaluation of *in vitro* alternatives to the Draize primary eye irritation test. (Phase III) Surfactant-based Formulations. Food Chem Toxicol 34:79-117.

Gilman AG, Goodman LS, Rall TW, Murad F (eds.). 1985. Goodman and Gilman's The Pharmacological Basis of Therapeutics, 7th Edition. New York:Macmillan Publishing Co.

Glaxo Wellcome. 2000. Myleran (busulfan) Product Information. Research Triangle Park, NC:Glaxo Wellcome, Inc.

GraphPad Software. 1994-2000. Prism® version 3.0a for Macintosh. GraphPad Software, San Diego, California. Available: www.graphpad.com [accessed 01 November 2006].

Grisham JW, Smith GJ. 1984. Predictive and Mechanistic Evaluation of Toxic Responses in Mammalian Cell Culture Systems. Pharmacol Rev 36(suppl):151S-171S.

Gülden M, Dierickx P, Seibert H. 2006. Validation of a prediction model for estimating serum concentrations of chemicals which are equivalent to toxic concentrations *in vitro*. Toxicol *In Vitro* 20:1114-1124.

Guzzie PJ. 1994. Lethality testing. In: *In Vitro* Toxicology (Gad SC, ed). New York:Raven Press, 57-86.

Hackenberg U, Bartling H. 1959. Messen und Rechnen im pharmakologischen Laboratorium mit einem speziellen Zahlensystem (WL24-System). Arch Exp Pathol Pharmakol 235:437-463.

Halle W, Liebsch M, Traue D, Spielmann H. 1997. Reduktion der Tierzahlen bei der Einstufung von Stoffen in die EU-Toxizitätsklassen fur akute orale Toxizität mit Hilfe von Daten aus dem Register der Zytotoxizität (RC). ALTEX 14:8-15.

Halle W. 1998. Toxizitätsprüfungen in Zellkulturen für eine Vorhersage der akuten Toxizität (LD$_{50}$) zur Einsparung von Tierversuchen. Life Sciences/ Lebenswissenschaften, Volume 1, Jülich: Forschungszentrum Jülich.

Halle W. 2003. The Registry of Cytotoxicity: Toxicity testing in cell cultures to predict acute toxicity (LD50) and to reduce testing in animals. Altern Lab Anim 31:89-198. (*English translation of Halle 1998*)

Halle W, Spielmann H. 1992. Two procedures for the prediction of acute toxicity (LD50) from cytotoxicity data. Altern Lab Anim 20:40-49.

Hardman JG, Limbird LE, Molinoff PB, Ruddon RW, Gilman AG, eds. 1996. Goodman and Gilman's The Pharmacological Basis of Therapeutics, 9th Edition. New York:McGraw-Hill.

Harlan Company. 2002. Available: http://www.harlan.com/us/index.htm [accessed 01 November 2006].

Harbell JW, Koontz SW, Lewis RW, Lovell D, Acosta D. 1997. IRAG working group 4. Cell cytotoxicity assays. Interagency Regulatory Alternatives Group. Food Chem Toxicol 35:79-126.

Hartung T, Balls M, Bardouille C, Blanck O, Coecke S, Gstrauthaler G, et al. 2002. Good cell culture practice: ECVAM Good Cell Culture Practice Task Force report 1. Altern Lab Anim 30:407-414. Available: http://ecvam.jrc.it/publication/index5007.html [accessed 01 November 2006].

Heimann R, Rice RH. 1983. Rat esophageal and epidermal-keratinocytes - intrinsic differences in culture and derivation of continuous lines. J Cell Physiol 117:362-367.

Hintze KL, Janus J, Marenus KD, Muscatiello MJ, Pape WJW, Renskers KJ, et al. 1992. Development of potential alternatives to the Draize eye test - The CTFA evaluation of alternatives program - Phase II - Review of materials and methods. Altern Lab Anim 20:164-171.

Hunter WJ, Lingk W, Recht JP. 1979. Intercomparison study on the determination of single administration toxicity in rats. J Assoc Off Anal Chem 62:864-873.

Hyclone®. 1996. Heat inactivation – are you wasting your time? Art to Science 15:1-5. Available: http://www.hyclone.com/library/arttoscience.htm [accessed 27 July 2006].

ICCVAM. 2000. The Revised Up-and-Down Procedure: A Test Method for Determining the Acute Oral Toxicity of Chemicals and Products. Proposed Test Method and Background Review Document. Research Triangle Park, NC:National Institute for Environmental Health Sciences. Available: http://iccvam.niehs.nih.gov/ [01 November 2006].

ICCVAM. 2001a. Report Of The International Workshop On *In Vitro* Methods For Assessing Acute Systemic Toxicity. NIH Publication No. 01-4499. Research Triangle Park, NC:National Institute for Environmental Health Sciences. Available: http://iccvam.niehs.nih.gov/ [accessed 01 November 2006].

ICCVAM. 2001b. Guidance Document On Using *In Vitro* Data To Estimate In Vivo Starting Doses For Acute Toxicity. NIH Publication No. 01-4500. Research Triangle Park, NC:National Institute for Environmental Health Sciences. Available: http://iccvam.niehs.nih.gov/ [accessed 01 November 2006].

ICCVAM. 2001c. The Revised Up-and-Down Procedure: A Test Method for Determining the Acute Oral Toxicity of Chemicals. NIH Publication No. 02-4501. Research Triangle Park, NC:National Institute for Environmental Health Sciences. Available: http://iccvam.niehs.nih.gov/ [accessed 01 November 2006].

ICCVAM. 2003. ICCVAM Guidelines for the Nomination and Submission of New, Revised, and Alternative Test Methods. NIH Publication No. 03-4508. Research Triangle Park, NC:National Institute for Environmental Health Sciences. Available: http://iccvam.niehs.nih.gov/ [accessed 01 November 2006].

ICCVAM. 2006. Peer Review Panel Report: The Use of *In Vitro* Basal Cytotoxicity Test Methods for Estimating Starting Doses for Acute Oral Systemic Toxicity Testing. Research Triangle Park, NC: National Institute for Environmental Health Sciences. Available: http://iccvam.niehs.nih.gov/methods/invitro.htm [accessed 02 August 2006].

IOM. 2004. Forging a Poison Prevention and Control System. Washington: National Academies Press.

INVITTOX. 1991. The FRAME Modified Neutral Red Uptake Cytotoxicity Test. Protocol Number 3 (IP-3A). Available: http://embryo.ib.amwaw.edu.pl/invittox/ [accessed 01 November 2006].

Karczmar A. 1998. Invited review: Anticholinesterases: dramatic aspects of their use and misuse. Neurochem Int 32:401-411.

King AV, Jones PA. 2003. In-house assessment of a modified *in vitro* cytotoxicity assay for higher throughput estimation of acute toxicity. Toxicol *In Vitro* 17:717-22.

Knox P, Uphill PF, Fry JR, Benford J, Balls M. 1986. The FRAME multicentre project on *in vitro* cytotoxicity. Food Chem Toxicol 24:457-463.

Kohn J, Durham HD. 1993. S9 liver fraction is cytotoxic to neurons in dissociated culture. Neurotoxicology 14:381-386.

Life Technologies, Inc. 1997. Culture of human keratinocytes in defined serum-free medium. Focus 19:1-5. Available: http://www.invitrogen.co.jp/focus/191002.pdf [accessed 27 July 2006].

Lipnick RL, Cotruvo JA, Hill RN, Bruce RD, Stitzel KA, Walker AP, et al. 1995. Comparison of the up-and-down, conventional LD50, and fixed-dose acute toxicity procedures. Food Chem Toxicol 33:223-231.

Litchfield JT, Wilcoxon F. 1949. A simplified method of evaluating dose-effect experiments. J Pharmacol Exp Therap 96:99-113.

Litovitz TL, Klein-Schwartz W, White S, Cobaugh DJ, Youniss J, Drab A, et al. 2000. 1999 Annual report of the american association of poison control centers toxic exposure surveillance system. Am J Emerg Med 18:517-574.

The MathWorks. 1996-2004. MATLAB[®] software. Natick, MA:The MathWorks, Inc.

Matsuno T, Masuda, Shimizu. 1971. Pipobroman: acute toxicity and general toxicological reaction. Kiso to Rinsho Clinical Report 5:1894-1917.

MDL Information Systems 2001, 2002. Registry of Toxic Effects of Chemical Substances (RTECS) accessed online by subscription.

Mediatech, Inc. 2006. Heat Inactivation of Serum. Technical Information #S055.0303.001. Available: http://cellgro.bizatomic.net/shop/customer/home.php [accessed 27 July 2006].

Microsoft Corporation. 1998-2001. Microsoft® Excel® 2000 software for Mac®. Redmond, WA:Microsoft Corporation.

Miller LC, Tainter ML. 1944. Estimation of the LD50 and its error by means of logarithmic-probit graph paper. Proc Soc Exp Biol Med 57:261-264.

Miyazaki H, Sekine T, Endou H. 2004. The multispecific organic anion transport family: properties and pharmacological significance. Trends Pharmacol Sci 25:654-662.

National Toxicology Program. 2002. Annual Plan For Fiscal Year 2002. Public Health Service, Department of Health and Human Services. NIH Publication No. 03-5309. Research Triangle Park, NC:National Institute for Environmental Health Sciences. Available: http://ntp.niehs.nih.gov/index.cfm?objectid=03611496-026C-BB0F-B2A1527014B2D4B1 [accessed 01 November 2006].

NLM. 2005. Haz-Map, Occupational Exposure to Hazardous Agents, Specialized Information Services, National Institutes of Health, Department of Health & Human Services. Washington, DC:National Library of Medicine. Available: http://hazmap.nlm.nih.gov/ [accessed 01 November 2006].

NLM. 2000, 2001, 2002, 2005. Hazardous Substances Data Bank. Accessed online via TOXNET. National Institutes of Health, Department of Health & Human Services. Washington, DC:National Library of Medicine. Available: http://toxnet.nlm.nih.gov/cgi-bin/sis/htmlgen?HSDB [accessed 01 November 2006].

Occupational Safety and Health Act. 2003. Toxic and Hazardous Substances. Hazard Communication 29CFR1910.1200. Available: http://www.gpoaccess.gov/cfr/index.html [accessed 01 November 2006].

Occupational Safety and Health Act. 2000. Labeling Requirements for Pesticides and Devices. Labeling Requirements. 40CFR156.10. Available: http://www.gpoaccess.gov/cfr/index.html [accessed 01 November 2006].

OECD. 1998. OECD Series on Principles of Good Laboratory Practice and Compliance Monitoring Number 1: OECD principles on Good Laboratory Practice. (as revised in 1997). ENV/MC/CHEM(98)17. Paris:Organisation for Economic Co-operation and Development. Available: http://www.oecd.org [accessed 01 November 2006].

OECD. 2000. Guidance Document on the Recognition, Assessment and Use of Clinical Signs as Humane Endpoints for Experimental Animals Used in Safety Evaluation. Environmental Health and Safety Monograph Series on Testing and Assessment No. 19. Paris, France:Organisation for Economic Co-operation and Development. Available: http://www.oecd.org [accessed 01 November 2006].

OECD. 2001a. Guideline for Testing of Chemicals, 425, Acute Oral Toxicity – Up-and-Down Procedure. Paris, France:Organisation for Economic Co-operation and Development. Available: http://www.oecd.org [accessed 01 November 2006].

OECD. 2001b. OECD Series on Testing and Assessment, 33, Harmonized Integrated Classification System for Human Health and Environmental Hazards of Chemical Substances and Mixtures. ENV/JM/MONO(2001)6. Paris, France:Organisation for Economic Co-operation and Development. Available: http://www.oecd.org [accessed 01 November 2006].

OECD. 2001c. Guideline for Testing of Chemicals, 420, Acute Oral Toxicity – Fixed Dose Method. Paris, France:Organisation for Economic Co-operation and Development. Available: http://www.oecd.org [accessed 01 November 2006].

OECD. 2001d. Guideline For Testing of Chemicals, 423, Acute Oral Toxicity – Acute Toxic Class Method. Paris, France:Organisation for Economic Co-operation and Development. Available: http://www.oecd.org [accessed 01 November 2006].

OECD. 2001e. Guidance Document on Acute Oral Toxicity Testing. 24. Paris, France:Organisation for Economic Co-operation and Development. Available: http://www.oecd.org [accessed 01 November 2006].

OECD. 2004a. OECD Series on Principles of Good Laboratory Practice and Compliance Monitoring Number 14: Advisory Document of the Working Group on Good Laboratory Practice: The Application of the Principles of GLP to *In Vitro* Studies. ENV/JM/MONO (2004)26. Paris, France:Organisation for Economic Co-operation and Development. Available: http://www.oecd.org [accessed 01 November 2006].

OECD. 2004b. Guideline for Testing of Chemicals, 432, *In Vitro* 3T3 NRU Phototoxicity Test. Paris, France:Organisation for Economic Co-operation and Development. Available: http://www.oecd.org [accessed 01 November 2006].

Orphan Medical, Inc. 1999. NDA 20-954 Busulfex™ (busulfan) Injection. NDA 20-954. Available: www.fda.gov/cder/foi/label/1999/20954lbl.pdf [accessed 01 November 2006].

Pesticide Action Network North America. 2005. Pesticide Action Network [PAN] Pesticides Database. Available: http://www.pesticideinfo.org/Index.html [accessed 01 November 2006].

Peloux AF, Federici C, Bichet N, Gouy D, Cano JP. 1992. Hepatocytes in primary culture: an alternative to LD_{50} testing? Validation of a predictive model by multivariate analysis. Altern Lab Anim 20:8-26.

Phillips JC, Gibson WB, Yam J, Alden CL, Hard GC. 1990. Survey of the QSAR and *in vitro* approaches for developing non-animal methods to supersede the *in vivo* LD50 test. Food Chem Toxicol 28:375-394.

Raabe HA, Moyer GO, Mun GC, Sizemore AM, Harbell JW, Merrill JC. 2004. Induction of a zone of cell death in multi-well plates by refeeding [Abstract]. Toxicologist 78:50. Available: http://www.iivs.org/pages/publication_view.php?doc_id=45 [accessed 28 August 2006].

Rasmussen ES. 1999 Cytotoxicity of MEIC chemicals nos. 11-30 in 3T3 mouse fibroblasts with and without microsomal activation. In Vitr Mol Toxicol 12:125-132.

Riddell RJ, Panacer DS, Wilde SM, Clothier RH, Balls M. 1986. The importance of exposure period and cell type in *in vitro* cytotoxicity tests. Altern Lab Anim 14:86-92.

Roguet R. Colovio J, Gaetani Q., Sossou KG, Rogier A. 1993. Cytotoxicity of 28 MEIC chemicals to rat hepatocytes using two viability endpoints: correlation with acute toxicity data in rat and man. Altern Lab Anim 21:216-224.

SAS Institute, Inc. 1999. SAS/STAT User's Guide Version 8. Cary, NC:SAS Institute, Inc.

Schlede E, Mischke U, Roll R, Kayser D. 1992. A national validation study of the acute-toxic-class method--an alternative to the LD50 test. Arch Toxicol 66:455-470.

Schlede E, Mischke U, Diener W, Kayser D. 1995. The international study of the acute toxic class method (oral). Arch Toxicol 69:659-670.

Schmahl D, Osswald H. 1970. Experimental studies on the carcinogenic effects of anticancer chemotherapeutics and immunosuppressive agents. Arzneimittelforschung 20:1461-1467.

Seibert H, Gulden M, Kolossa M, Schepers G. 1992. Evaluation of the relevance of selected *in vitro* toxicity test systems for acute systemic toxicity. Altern Lab Anim 20:240-245.

Seibert H, Balls M, Fentem JH, Bianchi V, Clothier RH, Dierickx PJ, et al. 1996. Acute toxicity testing *in vitro* and the classification and labelling of chemicals. The report and recommendations of ECVAM Workshop 16. Altern Lab Anim 24:499-510.

Sina JF, Galer DM, Sussman RG, Gautheron PD, Sargent EV, Leong B, et al. 1995. A collaborative evaluation of seven alternatives to the Draize eye irritation test using pharmaceutical intermediates. Fundam Appl Toxicol 26:20-31.

Sekine T, Cha SH, Endou H. 2000. The multispecific organic anion transporter (OAT) family. Pflugers Arch 440:337-350.

Spielmann H, Genschow E, Liebsch M, Halle W. 1999. Determination of the starting dose for acute oral toxicity (LD$_{50}$) testing in the up and down procedure (UDP) from cytotoxicity data. Altern Lab Anim 27:957-966.

Spielmann H, Balls M, Dupuis J, Pape WJW, Pechovitch G, de Silva O, et al. 1998. EU/COLIPA "*In Vitro P*hototoxicity" Validation Study, Results of Phase II (blind trial), Part 1: the 3T3 NRU phototoxicity test. Toxicol *In Vitro* 12:305-327.

Spielmann H, Balls M, Dupuis J, Pape WJW, De Silva O, Holzhütter HG, et al. 1998b. A study on the phototoxic potential of UV filter chemicals from Annex VII of EU Directive 76/768/EEC in the 3T3 NRU *in vitro* phototoxicity test. Altern Lab Anim 26:679-705.

Spielmann H, Liebsch M, Kalweit S, Moldenhauer F, Wirnsberger T, Holzhütter HG, et al. 1996. Results of a validation study in Germany on two *in vitro* alternatives to the Draize eye irritation test, the HET-CAM test and the 3T3 NRU cytotoxicity test. Altern Lab Anim 24**:**741-858.

Spielmann H, Kalweit S, Liebsch M, Wirnsberger T, Gerner I, Bertram-Neis E, et al. 1993. Validation study of alternatives to the Draize eye irritation test in Germany: Cytotoxicity testing and HET-CAM test with 136 industrial chemicals. Toxicol *In Vitro* 7:505-510.

Spielmann H, Gerner S, Kalweit S, Moog R, Wirnserberger T, Krauser K, et al. 1991. Interlaboratory assessment of alternatives to the Draize eye irritation test in Germany. Toxicol *In Vitro* 5:539-542.

Stallard N, Whitehead A. 2004. A statistical evaluation of the fixed dose procedure. Altern Lab Anim 32(suppl 2):13-21.

Taconic Farms, Inc. 2002. Taconic Farms Animal Models, Sprague Dawley® Outbred Rats. Available: http://www.taconic.com/anmodels/spragued.htm [accessed 01 November 2006].

Thompson WR. 1947. Use of moving averages and interpolation to estimate median-effective dose. Bacteriol Rev 11:115-145.

Thomson PDR®. 2004. Physicians' Desk Reference, 58th edition. Montvale, NJ:Thomson PDR.

Triglia D, Wegener PT, Harbell J, Wallace K, Mathesen D, Shopsis C. 1989. Interlaboratory validation study of the keratinocyte neutral red bioassay from Clonetics Corporation. In: *In Vitro* Toxicology, New Directions (Goldberg A, ed). Alternative Methods in Toxicology, Volume 7. New York:Mary Ann Liebert, 357-365.

Tsao MC, Walthall BJ, Ham RG. 1982. Clonal growth of normal human epidermal-keratinocytes in a defined medium. J Cell Physiol 110: 219-229.

UN. 1977. United Nations Economic and Social Council. Joint meeting of the RID safety committee and the group of experts on the transportation of dangerous goods. Trans/GE 15/R 274, 2. New York and Geneva:United Nations.

UN. 2005. Globally Harmonized System of Classification and Labelling of Chemicals (GHS), First Revised Edition. [ST/SG/AC.10/30/Rev.1]. New York and Geneva:United Nations. Available: http://www.unece.org/trans/danger/publi/ghs/ghs_rev01/01files_e.html [accessed 01 November 2006].

Vernot EH, MacEwen JD, Haun CC, Kinkead ER. 1977. Acute toxicity and skin corrosion data for some organic and inorganic compounds and aqueous solutions. Toxicol Appl Pharmacol 42:417-423.

Wallum E. 1998. Acute oral toxicity. Environ Health Perspect 106:497-503.

Weil CS. 1952. Tables for convenient calculation of median effective dose (LD50 or ED50) and instructions in their use. Biometrics 8:249-263.

Weil CS, Wright GJ. 1967. Intra- and interlaboratory comparative evaluation of single oral test. Toxicol Appl Pharmacol 11:378-388.

WHO. 1998. The WHO Recommended Classification of Pesticides by Hazard and Guidelines to Classification 1998-1999. International Programme of Chemical Safety. WHO/PCS/98.21. Geneva:World Health Organization.

Willshaw A, Moore LJ, Balls M. 1994. *In vitro* alternatives for the detection of photoirritant chemicals - the EEC COLIPA trial. Toxicol *In Vitro* 8:723-725.

Zbinden G, Flury-Roversi M. 1981. Significance of the LD50-test for the toxicological evaluation of chemical substances. Arch Toxicol 47:77-99.

Zuang V, Balls M, Botham PA, Coquett A, Corsini E, Curren RD, et al. 2002. Follow-up to the ECVAM prevalidation study on *in vitro* tests for acute skin irritation. Altern Lab Anim 30:109-129.

[This Page Intentionally Left Blank]

13.0 GLOSSARY[1]

Accuracy[2]: (a) The closeness of agreement between a test method result and an accepted reference value. (b) The proportion of correct outcomes of a test method. It is a measure of test method performance and one aspect of "relevance". Accuracy is highly dependent on the prevalence of positives in the population being examined.

Acute Toxic Class (ATC) method: An acute oral systemic toxicity test method based on testing groups of animals at fixed doses in a sequential manner. The lethality outcomes are used to classify a test substance into the appropriate GHS acute oral toxicity category.

ANOVA: One-way (and two-way) analysis of variance. ANOVA compares the measurements (continuous variables) of three or more groups when the data are categorized in one way (one-way) or two ways (two-way). ANOVA assumes that the populations compared are normally distributed and that the variances for the groups to be compared are approximately equal.

Assay[2]: The experimental system used. Often used interchangeably with "test" and "test method."

Biphasic dose-response: Dose-response in which cytotoxicity increases (as dose increases), plateaus, and then increases again. See **Section 2.6.3**.

Category prediction: The acute oral GHS hazard category that includes the predicted LD_{50} value for a test chemical.

Coded substances: Substances labeled by code rather than name so that they can be tested and evaluated without knowledge of their identity or anticipation of test results. Coded substances are used to avoid intentional or unintentional bias when evaluating laboratory or test method performance.

Coefficient of determination: In linear regression, it denotes the proportion of the variance in Y and X that is shared. Its value ranges between zero and one and it is commonly called called "R^2." For example, $R^2 = 0.45$, indicates that 45% of the variance in Y can be explained by the variation in X and that 45% of the variance in X can be explained by the variation in Y.

---

[1] The definitions in this Glossary are restricted to their uses with respect to *in vitro* cytotoxicity testing and the NRU test methods.
[2] Definition used by the Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM 2003).

Coefficient of variation: A statistical representation of the precision of a test. It is expressed as a percentage and is calculated as follows:

$$\left(\frac{standard\ deviation}{mean} \right) \times 100\%$$

Concordance[2]: The proportion of all substances tested that are correctly classified as positive or negative. It is a measure of test method performance and one aspect of "relevance." The term is often used interchangeably with "accuracy." Concordance is highly dependent on the prevalence of positives in the population being examined. In the NICEATM/ECVAM study, concordance was used to describe the proportion of test substances that were correctly classified into GHS acute oral toxicity hazard categories, or to describe the proportion of test substances for which the laboratories obtained the same classification result.

Confluence: A state in which cells in culture come into contact with other cells in the same culture to form a complete sheet of cells (monolayer). For this study, confluence is determined as a percentage of cell coverage of the tissue culture vessel growth surface (e.g., cell monolayer has 80% confluency).

Cytotoxicity: The adverse effects resulting from interference with structures and/or processes essential for cell survival, proliferation, and/or function. For most chemicals, toxicity is a consequence of non-specific alterations in "basal cell functions" (i.e., via mitochondria, plasma membrane integrity, etc.), which may then lead to effects on organ-specific functions and/or death of the organism. These effects may involve the integrity of membranes and the cytoskeleton, cellular metabolism, the synthesis and degradation or release of cellular constituents or products, ion regulation, and cell division.

Definitive test: The main test of the cytotoxicity assay for determining the IC_{50}. The concentration closest to the range finder test IC_{50} serves as the midpoint of the concentrations tested in a definitive test. Compared to the range finder test, the definitive test uses a smaller dilution factor for the concentrations tested.

Discordant chemicals: Chemicals for which the LD_{50} is not accurately predicted by the IC_{50} (and the associated regression formula) or the GHS toxicity category is not accurately predicted by the IC_{50} (and the associated regression formula). Also referred to as "outliers."

EDIT: Evaluation-guided Development of New *In vitro* Test Batteries. An international project initiated by Björn Ekwall in 1998 and continued by the Scandinavian Society for Cell Toxicology to develop new *in vitro* tests for toxicity and toxicokinetics to be incorporated into test batteries for predicting acute and chronic systemic toxicity.

Endpoint[2]: The biological process, response, or effect assessed by a test method.

Fixed Dose Procedure (FDP): An acute oral systemic toxicity test method based on testing groups of animals at fixed doses. Evident toxicity outcomes are used to classify a test substance into the appropriate GHS acute oral toxicity category.

Geometric mean: The antilog of the mean of the logarithm of the values. It is less affected by extreme values than the arithmetic mean.

Globally Harmonized System (GHS): A classification system presented by the United Nations that provides (a) a harmonized criteria for classifying substances and mixtures according to their health, environmental and physical hazards, and (b) a harmonized hazard communication elements, including requirements for labeling and safety data sheets.

Good Laboratory Practices (GLP)[2]: Regulations promulgated by the U.S. Food and Drug Administration and the U.S. Environmental Protection Agency, and principles and procedures adopted by the Organization for Economic Cooperation and Development and Japanese authorities that describe record keeping and quality assurance procedures for laboratory records that will be the basis for data submissions to national regulatory agencies.

Guidance Document: *Guidance Document on Using In Vitro Data to Estimate In Vivo Starting Doses for Acute Toxicity* (ICCVAM 2001b).

Hazard[2]: The potential for an adverse health or ecological effect. A hazard potential results only if an exposure occurs that leads to the possibility of an adverse effect being manifested.

Hill function: The IC_{50} values are determined from the concentration-response using a Hill function which is a four parameter logistic mathematical model relating the concentration of the test chemical to the response (typically following a sigmoidal shape).

$$Y = \text{Bottom} + \frac{\text{Top} - \text{Bottom}}{1 + 10^{(\log EC50 - \log X)\text{HillSlope}}}$$

where Y=response (i.e., % viability), X is the substance concentration producing the response, Bottom is the minimum response (0% viability, maximum toxicity), Top is the maximum response (maximum viability), EC_{50} is the substance concentration at the response midway between Top and Bottom, and HillSlope describes the slope of the curve. When Top=100% viability and Bottom=0% viability, the EC_{50} is the equal to the IC_{50}.

Hill function (rearranged): Some unusual dose-responses did not fit the Hill function well. To obtain a better model fit, the Bottom parameter was estimated without constraints (the previous practice was to use Bottom=0). However, when Bottom≠0, the EC_{50} reported by the Hill function was not the same as the IC_{50} since the Hill function defines EC_{50} as the point midway between Top and Bottom. Thus, the Hill function calculation using the Prism[®] software was rearranged to calculate the concentration corresponding to the IC_{50} as follows:

$$\log IC_{50} = \log EC_{50} - \frac{\log\left(\dfrac{Top - Bottom}{Y - Bottom} - 1\right)}{HillSlope}$$

where IC_{50} is the concentration producing 50% toxicity, EC_{50} is the concentration producing a response midway between the Top and Bottom responses; Top is the maximum response (maximum survival), Bottom is the minimum response (0% viability, maximum toxicity), Y=50 (i.e., 50% response), and HillSlope describes the slope of the response. The X from the standard Hill function equation is replaced, in the rearranged Hill function equation, by the IC_{50}.

Hormesis: a dose-response characterized by a compound's ability to cause an opposite effect at low doses than it causes at high doses. A stimulatory effect at low doses and an inhibitory effect in high doses is often the observed manifestation of hormesis.

IC_{50}: test chemical concentration producing 50% inhibition of the endpoint measured (i.e., cell viability).

Interlaboratory reproducibility[2]: A measure of whether different qualified laboratories using the same protocol and test substances can produce qualitatively and quantitatively similar results. Interlaboratory reproducibility is determined during the prevalidation and validation processes and indicates the extent to which a test method can be transferred successfully among laboratories.

Intralaboratory repeatability[2]: The closeness of agreement between test results obtained within a single laboratory when the procedure is performed on the same substance under identical conditions within a given time period.

Intralaboratory reproducibility[2]: The first stage of validation; a determination of whether qualified people within the same laboratory can successfully replicate results using a specific test protocol at different times.

In vitro: In glass. Refers to assays that are carried out in an artificial system (e.g., in a test tube or petri dish) and typically use single-cell organisms, cultured cells, cell-free extracts, or purified cellular components.

In vivo: In the living organism. Refers to assays performed in multicellular organisms.

K_{ow}: Octanol:water partition coefficient.

LC_{50}: Acute lethal serum or blood concentrations.

LD_{50}: The calculated value of the oral dose that produces lethality in 50% of test animals (rats and mice). The LD_{50} values serve as reference values for the *in vitro* tests.

LD_{50} (initial): Acute oral rat and mouse LD_{50} values used during the chemical selection process. For RC chemicals, LD_{50} values were those used in the RC database, which were largely from the 1983/84 RTECS®. For chemicals that were not included in the RC, the initial LD_{50} values came from HSDB or 2002 RTECS®.

LD$_{50}$ (reference): Acute oral rodent LD$_{50}$ values from rats and mice were located through literature searches and references from major toxicity databases such as RTECS$^{\textregistered}$. Studies were reviewed to identify the most appropriate LD$_{50}$ values for each chemical. Values obtained using feral animals, preanesthetized animals, or animals less than 4 weeks of age were not used. Values reported as inequalities were not used. Reference LD$_{50}$ values were determined by calculating the geometric mean of the acceptable LD$_{50}$ values. Data were used in generation of the laboratory-specific and combined-laboratory 3T3 and NHK NRU regressions.

Maximum:minimum value: Ratio of minimum acceptable LD$_{50}$ (or IC$_{50}$) to maximum acceptable LD$_{50}$ (or IC$_{50}$).

MEIC: Multicentre Evaluation of *In Vitro* Cytotoxicity. An international effort established by the Scandinavian Society for Cell Toxicology and initiated in 1983 to evaluate the relationship and relevance of *in vitro* cytotoxicity for predicting the acute toxicity of chemicals in humans.

Millimolar regressions: Linear regressions with IC$_{50}$ values in mmol/L and LD$_{50}$ values in mmol/kg.

Negative control: An untreated sample containing all components of a test system, except the test substance solvent, which is replaced with a known non-reactive material, such as water. This sample is processed with test substance-treated samples and other control samples to determine whether the solvent interacts with the test system.

Neutral red (NR): A weakly cationic water-soluble dye that stains living cells by readily diffusing through the plasma membrane and concentrating in lysosomes where it electrostatically binds to the anionic lysosomal matrix.

Neutral red uptake (NRU): Concentration of neutral red dye in the lysosomes of living cells. Altering the cell surface or the lysosomal membrane by a toxicological agent causes lysosomal fragility and other adverse changes that gradually become irreversible. The NRU test method makes it possible to distinguish between viable, damaged, or dead cells because these changes result in decreased uptake and binding of NR measurable by optical density absorption readings in a spectrophotometer.

NHK: Normal **H**uman epidermal **K**eratinocytes (from neonatal foreskin).

Optical density (OD): The absorption (i.e., OD measurement) of the resulting colored solution (colorimetric endpoint) in the NRU assay measured at 540 nm ± 10 nm in a spectrophotometric microtiter plate reader using blanks as a reference

Outlier: For any measurement, an extreme value in the NICEATM/ECVAM study was referred to as an "outlier" if it passes a statistical test for outliers at the 99% level. With respect to chemicals, it refers to chemicals that do not fit (using the specified criteria) an IC$_{50}$-LD$_{50}$ linear regression model. It may also refer to chemicals for which the predicted

acute oral GHS toxicity category does not match the reference *in vivo* GHS acute oral toxicity category.

Performance[2]: The accuracy and reliability characteristics of a test method (see "accuracy", "reliability").

pH: A measure of the acidity or alkalinity of a solution. pH 7.0 is neutral; higher pHs are alkaline, lower pHs are acidic.

Plate reader: A spectrophotometric device for measuring light intensity as a function of color/wavelength (i.e., optical density/absorption at 540 nm ± 10 nm for NRU) in 96-well microtiter tissue culture plates.

Positive control: A sample containing all components of a test system and treated with a substance known to induce a positive response, which is processed with the test substance-treated and other control samples to demonstrate the sensitivity of each experiment and to allow for an assessment of variability in the conduct of the assay over time.

Predictivity[2]: Proportion of *in vivo* category matches for all substances with *in vitro* predictions for a particular category. Predictivity is an indicator of test accuracy.

Protocol[2]: The precise, step-by-step description of a test, including the listing of all necessary reagents, criteria and procedures for the evaluation of the test data.

Quality assurance (QA)[2]: A management process by which adherence to laboratory testing standards, requirements, and record keeping procedures is assessed independently by individuals other than those performing the testing.

Quality control (QC): A management process for ensuring products or services are designed and produced to adhere to a defined set of quality criteria to meet or exceed customer requirements; similar to quality assurance.

Range finder: Initial test performed to determine starting doses for the main (definitive) test. The NRU assays test eight concentrations of the test chemical or the PC by diluting the stock solution in log dilutions to cover a large concentration range.

RC millimole regression: $\log (LD_{50}) = 0.435 \times \log (IC_{50}) + 0.625$; for estimating an LD_{50} value in mmol/kg (body weight) from an IC_{50} value in mM. Developed using the 347 IC_{50} and oral LD_{50} (282 rat and 65 mouse) values from the RC.

RC rat-only millimole regression: $\log (LD_{50}) = 0.439 \times \log (IC_{50}) + 0.621$; for estimating an LD_{50} value in mmol/kg (body weight) from an IC_{50} value in mM. Developed from the IC_{50} values (in mM) and acute oral LD_{50} values (in mmol/kg) for the 282 substances with rat LD_{50} values in the RC database (Halle 1998, 2003).

RC rat-only weight regression: $\log (LD_{50}) = 0.372 \times \log (IC_{50}) + 2.024$; for estimating an LD_{50} value in mg/kg (body weight) from an IC_{50} value in µg/mL. Developed from the IC_{50} values (in µg/mL) and acute oral LD_{50} values (in mg/kg) for the 282 substances with rat LD_{50} values in the RC database (Halle 1998, 2003).

Reduction alternative[2]: A new or modified test method that reduces the number of animals required.

Reference substances: Substances selected for use during the research, development, prevalidation, and validation of a proposed test method because their response in the *in vivo* reference test method or the species of interest is known (see "reference test"). Reference substances should represent the classes of chemicals for which the proposed test method is expected to be used and cover the range of expected responses (negative, weak to strong positive).

Reference test method[2]: The accepted *in vivo* test method used for regulatory purposes to evaluate the potential of a test substance to be hazardous to the species of interest.

Refinement alternative[2]: A new or modified test method that refines procedures to lessen or eliminate pain or distress in animals or enhances animal well-being.

Registry of Cytotoxicity (RC): Database that consists of *in vivo* acute oral toxicity data (i.e., LD_{50} values) from rats and mice and *in vitro* cytotoxicity data (i.e., IC_{50} values) from multiple cell lines and cytotoxicity endpoints for 347 chemicals with known molecular weights (Halle 1998, 2003). A regression model constructed from these data was proposed by ZEBET, as a method to reduce animal use by identifying the most appropriate starting doses for acute oral systemic toxicity tests.

Relevance[2]: The extent to which a test method correctly predicts or measures the biological effect of interest in humans or another species of interest. Relevance incorporates consideration of the "accuracy" or "concordance" of a test method.

Reliability[2]: A measure of the degree to which a test method can be performed reproducibly within and among laboratories over time. It is assessed by calculating intra- and inter-laboratory reproducibility and intralaboratory repeatability.

Replacement alternative[2]: A new or modified test method that replaces animals with nonanimal systems or one animal species with a phylogenetically lower one (e.g., a mammal with an invertebrate).

Reproducibility[2]: The consistency of individual test results obtained in a single laboratory (intralaboratory reproducibility) or in different laboratories (interlaboratory reproducibility) using the same protocol and test substances (see intra- and inter-laboratory reproducibility).

RTECS®: Registry of Toxic Effects for Chemical Substances. Compendium of data extracted from the open scientific literature. The database includes toxicity data (e.g., acute

toxicity) and specific numeric toxicity values (e.g., LD_{50}). Compiled by the U.S. National Institute for Occupational Safety and Health (NIOSH) and now licensed to MDL Information Systems, Inc.

Sensitivity[2]: The proportion of all positive substances that are classified correctly as positive in a test method. It is a measure of test method accuracy.

Simulation modeling: Computer simulation modeling of the acute systemic toxicity assays to determine animal use. The simulation process uses a simulated population of animals for testing, a reference endpoint (i.e., "true" LD_{50} value), and its assumed log-normal distribution. Mortality is assumed to have a mean equal to the log of the true LD_{50}. The SD, which reflects the variability of the simulated population, is the inverse of the slope of the dose-mortality curve. Due to a lack of information for the real dose-mortality curve, the simulations assumed slopes of 0.5, 0.8, 2, 4, and 8.3.

Solubility: The amount of a test substance that can be dissolved (or thoroughly mixed with) culture medium or solvent. The solubility protocol was based on a U.S. EPA guideline (EPA 1998) that involves testing for solubility in a particular solvent, beginning at a relatively high concentration and proceeding to successively lower concentrations by adding more solvent as necessary for dissolution. Testing stops when, upon visual observation, the procedure produces a clear solution with no cloudiness or precipitate.

Solvent control: An untreated sample containing all components of a test system, including the solvent that is processed with the test substance-treated and other control samples to establish the baseline response for the samples treated with the test substance dissolved in the same solvent. When tested with a concurrent negative control, this sample also demonstrates whether the solvent interacts with the test system.

Specificity[2]: The proportion of all negative substances that are classified correctly as negative in a test method. It is a measure of test method accuracy.

Spirit of GLP: Guidance provided in the Statement of Work specifically for the non GLP-compliant laboratory that participated in the validation study. Based on the GLP standards referenced in the ECVAM Workshop 37 Report (Cooper-Hannan et al. 1999) and the OECD Principles of GLP (OECD 1998). "*Laboratories that are non GLP-compliant shall adhere to GLP principles and other method parameters. Documentation and accountability shall be equal to GLP requirements. Laboratories must make assurances that they are equal in performance criteria and that there is parity amongst the laboratories.*"

TESS: Toxic Exposure Surveillance System. A comprehensive poisoning surveillance database maintained by the American Association of Poison Control Centers (AAPCC).

Test[2]: The experimental system used; used interchangeably with "test method" and "assay".

Test method[2]: A process or procedure used to obtain information on the characteristics of a substance or agent. Toxicological test methods generate information regarding the ability of a

substance or agent to produce a specified biological effect under specified conditions. Used interchangeably with "test" and "assay". See also "validated test method" and "reference test".

Test method component: Structural, functional, and procedural elements of a test method that are used to develop the test method protocol. These components include unique characteristics of the test method, critical procedural details, and quality control measures.

3T3: BALB/c 3T3 clone A31 mouse fibroblasts developed in 1968 from disaggregated 14- to 17-day-old BALB/c mouse embryos (American Type Culture Collection [ATCC]; # CCL-163).

Tiered testing: A testing strategy where all existing information on a test substance is reviewed, in a specified order, before *in vivo* testing.

Toxicity underpredicted: Measured l LD_{50} value of a test substance is lower than the predicted LD_{50} value.

Toxicity overpredicted: Measured LD_{50} value of a test substance is higher than the predicted LD_{50} value.

Transferability[2]: The ability of a test method or procedure to be accurately and reliably performed in different, competent laboratories.

Up-and-Down Procedure (UDP): An acute oral systemic toxicity test method used to minimize the number of animals required to estimate the acute oral toxicity of a chemical, estimate the LD_{50} and confidence interval (CI), and observe signs of toxicity. Single animals are tested sequentially. Subsequent doses are based on the outcome of the previous animal.

Validated test method[2]: An accepted test method for which validation studies have been completed to determine the accuracy and reliability of this method for a specific proposed use.

Validation[2]: The process by which the reliability and accuracy of a procedure are established for a specific purpose.

Vehicle control (VC): The VC consists of appropriate cell culture medium for the cells in the test (i.e., DMEM for 3T3 cells and keratinocyte growth medium for the NHK cells). For chemicals dissolved in DMSO, the VC consists of medium with the same amount of solvent as that used in the test chemical concentrations that are applied to the 96-well test plate. The final DMSO concentration is ≤0.5% (v/v) in the VCs.

Volatility: Ability of a test chemical to evaporate. A general indicator of volatility issues in the NRU test methods is the percent difference in the mean OD values for the two VC columns on the test plate. If the difference is greater than 15%, then chemical volatility can be suspected, especially if the VC adjacent to the highest test concentration had a

significantly reduced OD value. Volatility may be an issue for compounds with a specific gravity of less than 1.

Weight of evidence (process): The strengths and weaknesses of a collection of information are used as the basis for a conclusion that may not be evident from the individual data.

Weight regressions: Linear regressions with IC_{50} values in µg/mL and LD_{50} values in mg/kg.

ZEBET: The German National Center for the Documentation and Evaluation of Alternative Methods to Animal Experiments.