ICCVAM TEST METHOD EVALUATION REPORT

IN VITRO CYTOTOXICITY TEST METHODS FOR ESTIMATING STARTING DOSES FOR ACUTE ORAL SYSTEMIC TOXICITY TESTING

Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM)

National Toxicology Program (NTP) Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM)

National Toxicology Program
P.O. Box 12233
Research Triangle Park, NC 27709

November 2006
NIH Publication No. 07-4519

National Institute of Environmental Health Sciences
National Institutes of Health
U.S. Public Health Service
Department of Health and Human Services
This document is available electronically at:
http://iccvam.niehs.nih.gov/methods/invitro.htm
TABLE OF CONTENTS

APPENDICES ........................................................................................................................................ iii
LIST OF TABLES ...................................................................................................................................... iv
LIST OF ABBREVIATIONS AND ACRONYMS ................................................................................ v
INTERAGENCY COORDINATING COMMITTEE ON THE VALIDATION
OF ALTERNATIVE METHODS - DESIGNATED AGENCY
REPRESENTATIVES ........................................................................................................................ vii
ACKNOWLEDGMENTS .................................................................................................................. viii
PREFACE ............................................................................................................................................ xiii
EXECUTIVE SUMMARY .................................................................................................................... xvii

1.0 INTRODUCTION .................................................................................................................................. 1
  1.1 Evaluation of the Use of *In Vitro* Cytotoxicity Test Methods to
     Estimate Acute Oral Toxicity ........................................................................................................ 1
  1.2 Evaluation of the Use of *In Vitro* Cytotoxicity Test Methods to
     Estimate Starting Doses for Acute Oral Toxicity Tests ........................................................... 2
  1.3 Peer Review of the NICEATM/ECVAM Validation Study .......................................................... 3
  1.4 Report Organization ...................................................................................................................... 5

2.0 ICCVAM RECOMMENDATIONS FOR *IN VITRO* NEUTRAL RED
UPTAKE (NRU) BASAL CYTOTOXICITY TEST METHODS .............................................................. 7
  2.1 Test Method Description .............................................................................................................. 7
    2.1.1 General Test Method Procedures ........................................................................................ 7
    2.1.2 Protocol Similarities and Differences for the 3T3 and NHK
         NRU Test Methods .................................................................................................................... 7
  2.2 Reference Substances .................................................................................................................. 8
  2.3 Test Method Accuracy ................................................................................................................. 9
  2.4 Test Method Reliability (Inter- and Intra-Laboratory
     Reproducibility) .......................................................................................................................... 14
  2.5 Animal Welfare Considerations: Reduction, Refinement, and
     Replacement .................................................................................................................................... 15
  2.6 ICCVAM Recommendations for Test Method Uses ...................................................................... 22
  2.7 ICCVAM Recommendations for Future Studies .......................................................................... 23

3.0 ICCVAM RECOMMENDED PERFORMANCE STANDARDS ....................................................... 25
3.1 Essential Test Method Components for *In Vitro* Basal Cytotoxicity Assays to Predict Starting Dose for Acute Oral Toxicity Tests ..........25

3.1.1 *In Vitro* Cell Culture Conditions......................................................26
3.1.2 Application of the Test Substances..........................................................27
3.1.3 Control Substances...........................................................................28
3.1.4 Viability Measurements........................................................................29
3.1.5 Interpretation of Results......................................................................29
3.1.6 Test Report.........................................................................................30

3.2 Reference Substances for *In Vitro* Basal Cytotoxicity Assays to Predict Starting Doses for Acute Oral Toxicity Tests .................31

3.3 Accuracy and Reliability Standards..........................................................34

3.3.1 Accuracy and Reliability for the NRU Test Methods.........................34
3.3.2 Accuracy and Reliability for Me-Too Assays.......................................34

4.0 REFERENCES.............................................................................................39
APPENDICES

APPENDIX A PEER REVIEW PANEL REPORTS................................................. A-1
   A1 Peer Review Panel Report: The Use of In Vitro Basal Cytotoxicity
       Test Methods for Estimating Starting Doses for Acute Oral Systemic
       Toxicity Testing.................................................................................... A-3
   A2 Summary Minutes from Peer Review Panel Meeting on May 23, 2006 .... A-65

APPENDIX B RELEVANT FEDERAL ACUTE ORAL TOXICITY
REGULATIONS AND TESTING GUIDELINES .......................B-1
   B1 Table of Relevant Acute Oral Toxicity Regulations.................................B-3
   B3 OECD Guideline 423: Acute Oral Toxicity – Acute Toxic Class
       Method ....................................................................................................B-35
   B4 OECD Guideline 420: Acute Oral Toxicity – Fixed Dose Procedure.........B-51
   B5 Health Effects Test Guidelines OPPTS 870.1100: Acute Oral Toxicity ....B-67
   B6 OECD Series on Testing and Assessment Number 24: Guidance
       Document on Acute Oral Toxicity Testing................................................B-107

APPENDIX C RECOMMENDED TEST METHOD PROTOCOLS ................... C-1
   C1 Test Method Protocol for the BALB/c 3T3 NRU Cytotoxicity Test
       Method ....................................................................................................C-3
   C2 Test Method Protocol for the NHK NRU Cytotoxicity Test Method........C-37

APPENDIX D FEDERAL REGISTER NOTICES AND PUBLIC
COMMENTS.......................................................................................... D-1
   D1 Federal Register Notices ........................................................................ D-3
   D2 ICCVAM Consideration of Public Comments Received in Response to
       Federal Register Notices ......................................................................... D-27

APPENDIX E ICCVAM RECOMMENDATIONS FROM THE 2000
INTERNATIONAL WORKSHOP ON IN VITRO
METHODS FOR ASSESSING ACUTE SYSTEMIC
TOXICITY ............................................................................................E-1
**LIST OF TABLES**

Table 2-1  Prediction of GHS Acute Oral Toxicity Category by the 3T3 and NHK NRU Test Methods and the RC Rat-Only Millimole Regression ........................................11

Table 2-2  Prediction of GHS Acute Oral Toxicity Category by the 3T3 and NHK NRU Test Methods and the RC Rat-Only Weight Regression ..............................13

Table 2-3  Animal Use for the UDP by GHS Acute Oral Toxicity Category Using Starting Doses Based on the 3T3 and NHK NRU Test Methods with the RC Rat-Only Millimole Regression ..................................................17

Table 2-4  Animal Use for the UDP by GHS Acute Oral Toxicity Category Using Starting Doses Based on the 3T3 and NHK NRU Test Methods with the RC Rat-Only Weight Regression .........................................................18

Table 2-5  Animal Use for the ATC Method by GHS Acute Oral Toxicity Category Using Starting Doses Based on the 3T3 and NHK NRU Test Methods with the RC Rat-Only Millimole Regression .........................................20

Table 2-6  Animal Use for the ATC Method by GHS Acute Oral Toxicity Category Using Starting Doses Based on the 3T3 and NHK NRU Test Methods with the RC Rat-Only Weight Regression ..................................................21

Table 3-1  Recommended Reference Substances for Evaluation of In Vitro Basal Cytotoxicity Methods for Predicting the Starting Dose for Rodent Acute Oral Toxicity Tests ..................................................................................32

Table 3-2  Prediction of GHS Acute Oral Toxicity Category by 3T3 NRU Test Method Using the Recommended Reference Substances and the RC Rat-Only Millimole Regression .........................................................36

Table 3-3  Prediction of GHS Acute Oral Toxicity Category by 3T3 NRU Test Method Using the Recommended Reference Substances and the RC Rat-Only Weight Regression ........................................................................37
# LIST OF ABBREVIATIONS AND ACRONYMS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>3T3</td>
<td>BALB/c mouse fibroblasts, clone A31</td>
</tr>
<tr>
<td>ADME</td>
<td>Absorption, distribution, metabolism, excretion</td>
</tr>
<tr>
<td>ANOVA</td>
<td>Analysis of variance</td>
</tr>
<tr>
<td>ATC</td>
<td>Acute Toxic Class method</td>
</tr>
<tr>
<td>ATWG</td>
<td>Acute Toxicity Working Group</td>
</tr>
<tr>
<td>BRD</td>
<td>Background review document</td>
</tr>
<tr>
<td>CASRN</td>
<td>Chemical Abstracts Service Registry Number</td>
</tr>
<tr>
<td>CPSC</td>
<td>U.S. Consumer Product Safety Commission</td>
</tr>
<tr>
<td>CS</td>
<td>Calf serum</td>
</tr>
<tr>
<td>CV</td>
<td>Coefficient of variation</td>
</tr>
<tr>
<td>°C</td>
<td>Degrees Celsius</td>
</tr>
<tr>
<td>DMSO</td>
<td>Dimethyl sulfoxide</td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic acid</td>
</tr>
<tr>
<td>DPBS</td>
<td>Dulbecco’s Phosphate Buffered Saline</td>
</tr>
<tr>
<td>ECBC</td>
<td>U.S. Army Edgewood Chemical Biological Center</td>
</tr>
<tr>
<td>ECVAM</td>
<td>European Centre for the Validation of Alternative Methods</td>
</tr>
<tr>
<td>EDIT</td>
<td>Evaluation-guided Development of New <em>In Vitro</em> Tests</td>
</tr>
<tr>
<td>EPA</td>
<td>U.S. Environmental Protection Agency</td>
</tr>
<tr>
<td>ETOH</td>
<td>Ethanol</td>
</tr>
<tr>
<td>FAL</td>
<td>FRAME Alternatives Laboratory</td>
</tr>
<tr>
<td>FL</td>
<td>Fluorescein leakage</td>
</tr>
<tr>
<td>FR</td>
<td>Federal Register</td>
</tr>
<tr>
<td>FRAME</td>
<td>Fund for Replacement of Animals in Medical Experiments</td>
</tr>
<tr>
<td>GHS</td>
<td>Globally Harmonized System of Classification and Labelling of Chemicals (UN 2005)</td>
</tr>
<tr>
<td>HPV</td>
<td>High Production Volume</td>
</tr>
<tr>
<td>IC₅₀</td>
<td>Test substance concentration producing 50% inhibition of the endpoint measured</td>
</tr>
<tr>
<td>ICCVAM</td>
<td>Interagency Coordinating Committee on the Validation of Alternative Methods</td>
</tr>
<tr>
<td>IIVS</td>
<td>Institute for <em>In Vitro</em> Sciences</td>
</tr>
<tr>
<td>ILS</td>
<td>Integrated Laboratory Systems</td>
</tr>
<tr>
<td>LD₅₀</td>
<td>Dose that produces lethality in 50% of test animals</td>
</tr>
<tr>
<td>LDH</td>
<td>Lactate dehydrogenase</td>
</tr>
<tr>
<td>MEIC</td>
<td>Multicentre Evaluation of <em>In Vitro</em> Cytotoxicity</td>
</tr>
<tr>
<td>MTT</td>
<td>[3-(4,5-dimethylthiazol-2-yl)2,5-diphenyl tetrazolium bromide]</td>
</tr>
<tr>
<td>NCS</td>
<td>Newborn calf serum</td>
</tr>
<tr>
<td>NHK</td>
<td>Normal human epidermal keratinocytes</td>
</tr>
<tr>
<td>NICEATM</td>
<td>National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods</td>
</tr>
<tr>
<td>NIEHS</td>
<td>National Institute of Environmental Health Sciences</td>
</tr>
<tr>
<td>NR</td>
<td>Neutral red</td>
</tr>
<tr>
<td>NRR</td>
<td>Neutral red release</td>
</tr>
<tr>
<td>NRU</td>
<td>Neutral red uptake</td>
</tr>
<tr>
<td>Acronym</td>
<td>Definition</td>
</tr>
<tr>
<td>---------</td>
<td>------------</td>
</tr>
<tr>
<td>NTP</td>
<td>U.S. National Toxicology Program</td>
</tr>
<tr>
<td>OD</td>
<td>Optical density</td>
</tr>
<tr>
<td>OECD</td>
<td>Organisation for Economic Cooperation and Development</td>
</tr>
<tr>
<td>PC</td>
<td>Positive control</td>
</tr>
<tr>
<td>QSAR</td>
<td>Quantitative structure-activity relationship</td>
</tr>
<tr>
<td>RC</td>
<td>Registry of Cytotoxicity</td>
</tr>
<tr>
<td>RTECS</td>
<td>Registry of Toxic Effects of Chemical Substances</td>
</tr>
<tr>
<td>SACATM</td>
<td>Scientific Advisory Committee on Alternative Toxicological Methods</td>
</tr>
<tr>
<td>SLS</td>
<td>Sodium lauryl sulfate</td>
</tr>
<tr>
<td>SMT</td>
<td>Study Management Team</td>
</tr>
<tr>
<td>TESS</td>
<td>Toxic Exposure Surveillance System</td>
</tr>
<tr>
<td>UDP</td>
<td>Up-and-Down Procedure</td>
</tr>
<tr>
<td>UN</td>
<td>United Nations</td>
</tr>
<tr>
<td>VC</td>
<td>Vehicle control</td>
</tr>
<tr>
<td>XTT</td>
<td>[Sodium 3,3,3-[(Phenylamino)carbonyl]-3,4-Tetrazolium-Bis(4-methoxy-6-nitro)benzenesulfonic acid hydrate]</td>
</tr>
<tr>
<td>ZEBET</td>
<td>German Centre for Documentation and Evaluation of Alternative Methods to Animal Experiments</td>
</tr>
</tbody>
</table>
INTERAGENCY COORDINATING COMMITTEE ON THE VALIDATION OF ALTERNATIVE METHODS - DESIGNATED AGENCY REPRESENTATIVES

Agency for Toxic Substances and Disease Registry
• Moiz Mumtaz, Ph.D.

Consumer Product Safety Commission
• Marilyn L. Wind, Ph.D. (Vice-Chair)
* Patricia Bittner, M.S.
* Kailash C. Gupta, D.V.M., Ph.D.
* Kristina Hatlelid, Ph.D.

Department of Agriculture
• Jodie Kulpa-Eddy, D.V.M.
◊ Elizabeth Goldentyer, D.V.M.

Department of Defense
• Robert E. Foster, Ph.D.
◊ Patty Decot
* Harry Salem, Ph.D.

Department of Energy
• Michael Kuperberg, Ph.D.
◊ Marvin Stodolsky, Ph.D.

Department of the Interior
• Barnett A. Rattner, Ph.D.
◊ Sarah Gerould, Ph.D.

Department of Transportation
• George Cushmac, Ph.D.
◊ Steve Hwang, Ph.D.

Environmental Protection Agency
Office of Science Coordination and Policy
• Karen Hamernik, Ph.D.

Office of Research and Development
◊ Julian Preston, Ph.D.
* Suzanne McMasters, Ph.D.

OECD Test Guidelines Program
* Jerry Smrchek, Ph.D.

Office of Pesticides Programs
* Amy Rispin, Ph.D.
* Deborah McCall

Food and Drug Administration
• Leonard M. Schectman, Ph.D. (Chair)

Office of Science and Health Coordination
◊ Susanne Fitzpatrick, Ph.D., D.A.B.T.

Center for Drug Evaluation and Research
* Abigail C. Jacobs, Ph.D.

Center for Devices and Radiological Health
* Raju Kamnula, D.V.M., Ph.D., D.A.B.T.
* Melvin E. Stratmeyer, Ph.D.

Center for Biologics Evaluation and Research
* Richard McFarland, Ph.D., M.D.
* Ying Huang, Ph.D.

Center for Food Safety and Nutrition
* David G. Hattan, Ph.D.
* Robert L. Bronaugh, Ph.D.

Center for Veterinary Medicine
* Devaraya Jagannath, Ph.D.
* M. Cecilia Aguila, D.V.M.

National Center for Toxicological Research
* William T. Allaben, Ph.D.

Office of Regulatory Affairs
* Lawrence A. D’Hoostelaere, Ph.D.

National Cancer Institute
• Alan Poland, M.D.
◊ T. Kevin Howcroft, Ph.D.

National Institute of Environmental Health Sciences
• William S. Stokes, D.V.M., D.A.C.L.A.M.
◊ John R. Bucher, Ph.D., D.A.B.T.
* Rajendra S. Chhabra, Ph.D., D.A.B.T
* Jerrold J. Heindel, Ph.D.

National Institute for Occupational Safety and Health
• Paul Nicolaysen, V.M.D.
◊ K. Murali Rao, M.D., Ph.D.

National Institutes of Health
• Margaret D. Snyder, Ph.D.

National Library of Medicine
• Vera Hudson, M.S.
◊ Jeanne Goshorn, M.S.

Occupational Safety and Health Administration
• Surender Ahir, Ph.D.
ACKNOWLEDGMENTS

The following individuals are acknowledged for their contributions and oversight of the in vitro acute toxicity test method evaluation process.

Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) Acute Toxicity Working Group (ATWG)

Consumer Product Safety Commission (CPSC)
Kailash Gupta, D.V.M, Ph.D.
Cassandra Prioleau, Ph.D.
Marilyn Wind, Ph.D. (ATWG Chair, ICCVAM Vice Chair)

Department of Energy (DOE)
Po-Yung Lu, Ph.D.

Environmental Protection Agency (EPA)
Karen Hamernik, Ph.D.
Masih Hashim, Ph.D.
Marianne Lewis
Elizabeth Margosches, Ph.D.
Deborah McCall
John Redden, Ph.D.
Amy Rispin, Ph.D.

Food and Drug Administration (FDA)
Leonard Schechtman, Ph.D. (ICCVAM Chair)
Kenneth Hastings, Ph.D.
Abigail Jacobs, Ph.D.
Suzanne Morris, Ph.D.
David Morse, Ph.D.
Thomas Umbreit, Ph.D.

National Institute for Occupational Safety & Health (NIOSH)
Steven Reynolds, Ph.D.

National Institute of Environmental Health Sciences (NIEHS)
Rajendra Chhabra, Ph.D., D.A.B.T.
William Stokes, D.V.M., D.A.C.L.A.M.
(ICCVAM Executive Director; NICEATM Director)
Raymond Tice, Ph.D. (NICEATM Deputy Director)

European Centre for the Validation of Alternative Methods (ECVAM)
Liaisons
Silvia Casati, Ph.D.
Pilar Prieto, Ph.D.
In Vitro Acute Toxicity Peer Review Panel

David H. Blakey, D.Phil., Health Canada, Ottawa, Ontario, Canada
June Bradlaw, Ph.D., International Foundation for Ethical Research (IFER), Rockville, Maryland
Robert Copeland, Ph.D., Howard University College of Medicine, Washington, DC
Gianni Dal Negro, D.V.M., Ph.D., GlaxoSmithKline Medicine Research Centre, Verona, Italy
Marion Ehrich, Ph.D., RPh., D.A.B.T. Virginia-Maryland Regional College of Veterinary Medicine, Blacksburg, Virginia
Eugene Elmore, Ph.D., University of California, Irvine, Irvine, California
Benjamin Gerson, M.D., Thomas Jefferson University School of Medicine, Philadelphia, Pennsylvania
Michael Greene, Ph.D., U.S. Consumer Product Safety Commission, Bethesda, MD
Janice Kuhn, Ph.D., D.A.B.T., Stillmeadow Inc., Sugar Land, Texas
Daniel Marsman, D.V.M., Ph.D., D.A.B.T., Procter & Gamble Company, Cincinnati, Ohio
Andrew Rowan, Ph.D., Humane Society of the United States, Washington, DC
Hasso Seibert, Ph.D., University Medical School Schleswig-Holstein, Kiel, Germany
Nigel Stallard, Ph.D., The University of Warwick, Coventry, United Kingdom
Katherine Stitzel, D.V.M., (Panel Chair), Consultant, West Chester, Ohio
Shinobu Wakuri, MSc., Hatano Research Institute, Japan
Daniel Wilson, Ph.D., D.A.B.T., The Dow Chemical Company, Midland, Michigan
National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods

David Allen, Ph.D.
ILS, Inc.

Bradley Blackard, M.S.P.H.
ILS, Inc.

Sue Brenzel
ILS, Inc.

Thomas Burns, M.S.
ILS, Inc.

Patricia Ceger, M.S.
ILS, Inc.

Jeff Charles, Ph.D., M.B.A., D.A.B.T.
ILS, Inc.

Neepa Choksi, Ph.D.
ILS, Inc.

Frank Deal, M.S.
ILS, Inc.

Linda Litchfield
ILS, Inc.

Deborah McCarley
NIEHS

Michael Paris
ILS, Inc.

William Stokes, D.V.M., D.A.C.L.A.M. (Director)
NIEHS

Judy Strickland, Ph.D., D.A.B.T.
ILS, Inc.

Raymond Tice, Ph.D. (Deputy Director)
NIEHS

James Truax, M.A.
ILS, Inc.

Doug Winters, M.S.
ILS, Inc.
Participants in the *In Vitro* Cytotoxicity Validation Study

**BioReliance Corp.**  
(Chemical Distribution)  
Martin Wenk, Ph.D. – Principal Investigator

**Institute for In Vitro Sciences (IIVS)**  
(Lead Laboratory – Protocols)  
Hans Raabe, M.S. – Study Director  
Greg Mun – Laboratory Manager  
Angela Sizemore – Research Technician  
Gregory O. Moyer – Research Technician  
John Harbell, Ph.D. – Scientific Director

**U.S. Army Edgewood Chemical Biological Center (ECBC)**  
(Testing Laboratory)  
Cheng Cao, Ph.D. – Study Director  
Janna Madren-Whalley – Research Technician  
Chundakkadu Krishna, Ph.D. – Research Technician  
James J. Valdes, Ph.D. – Scientific Advisor

**FRAME Alternatives Laboratory (FAL)**  
(University of Nottingham, UK)  
(Lead Laboratory – Software)  
Richard Clothier, Ph.D. – Study Director  
Nicola Bourne – Research Technician  
Monika Owen – Research Technician  
Rachel Budworth – Research Technician

**Constella Group (Statistical Analyses)**  
Patrick Crockett, Ph.D.  
Eric Harvey, Ph.D.  
Wendell Jones, Ph.D.  
Robert Lee, M.S.  
Jessica L. Matthews, M.S.  
Michael Riggs, Ph.D.  
Janine Wilcox  
Nicole Williams

**Statistical Consultant**  
Joseph Haseman, Ph.D.

**NIEHS**  
Grace Kissling, Ph.D. – Contract Project Officer  
Molly Vallant - Contract Project Officer

**Study Management Team (SMT)**  
NICEATM  
William Stokes, D.V.M., D.A.C.L.A.M. (NIEHS) – Director, NICEATM  
Raymond Tice, Ph.D. (NIEHS) – Deputy Director, NICEATM  
Judy Strickland, Ph.D., D.A.B.T. (ILS, Inc.) – Project Coordinator  
Michael Paris (ILS, Inc.) – Assistant Project Coordinator  
Jeffrey Charles, Ph.D., D.A.B.T. (ILS, Inc.) – Advisor

**ECVAM**  
Thomas Hartung, Ph.D., M.D. – Head of Unit from 2002  
Silvia Casati, Ph.D. – Task Leader  
Michael Balls, D. Phil. – Head of Unit until June 2002
PREFACE

The Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) is charged by the ICCVAM Authorization Act of 20001 with evaluating the scientific validity of new, revised, and alternative toxicological test methods applicable to U.S. Federal agency safety testing requirements. ICCVAM is required to also provide recommendations to U.S. Federal agencies regarding the usefulness and limitations of test methods following their scientific evaluation. This report provides the ICCVAM’s recommendations for using two in vitro methods for estimating the acute oral toxicity potential of chemicals and other substances. These recommendations are based on a thorough ICCVAM evaluation of the scientific validation status of the test methods.

ICCVAM initiated a review of the validation status of in vitro methods for estimating acute oral toxicity in 1999 in response to a request from the U.S. Environmental Protection Agency (EPA) Office of Pesticides, Prevention, and Toxic Substances. The request was based on recently published studies that showed a correlation between in vitro and in vivo acute toxicity. An International Workshop on In Vitro Methods for Assessing Acute Systemic Toxicity was subsequently convened by ICCVAM and the National Toxicology Program (NTP) Center for the Evaluation of Alternative Toxicological Methods (NICEATM) in October 2000. Workshop participants concluded that the proposed in vitro methods had not yet undergone adequate studies to determine if they could meet regulatory requirements for acute toxicity testing. However, an in vitro approach previously proposed by the German Centre for Documentation and Evaluation of Alternative Methods to Animal Experiments (ZEBET) was recommended by workshop participants as a high priority for further evaluation (ICCVAM 2001a). In vitro cytotoxicity data was proposed as an approach for estimating starting doses for in vivo acute toxicity studies based on a correlation between in vitro IC50 and in vivo LD50 values2. Such a strategy might reduce the number of animals required for an acute oral toxicity test by identifying a starting dose closer to the actual LD50. A Guidance Document on Using In Vitro Data to Estimate In Vivo Starting Doses for Acute Toxicity (ICCVAM 2001b) was subsequently prepared by some of the workshop participants with the assistance of ICCVAM and NICEATM to provide interim in vitro cytotoxicity protocols and instructions for implementing the approach.

ICCVAM agreed with the workshop participants that in vitro basal cytotoxicity test methods should have a high priority for validation studies. NICEATM and the European Centre for the Validation of Alternative Methods (ECVAM) subsequently developed a collaboration 1) to further to characterize the usefulness and limitations of in vitro cytotoxicity assays as predictors of starting doses for rodent acute oral toxicity test methods, and 2) to develop a high quality database of in vitro cytotoxicity data that could be used to determine what other in vitro tests would be needed to accurately estimate acute toxicity hazard classification categories. NICEATM and ECVAM designed an international, multi-laboratory validation study to evaluate the performance of two standardized in vitro neutral red uptake (NRU) test

---

2 The IC50 is the test substance concentration that produces 50% inhibition of the endpoint measured. The LD50 is the dose that produces lethality in 50% of the test animals.
methods, using the ZEBET approach based on the Registry of Cytotoxicity (RC)\(^3\) regression model. One test method used BALB/c 3T3 mouse fibroblasts (3T3) while the other used normal human epidermal keratinocytes (NHK).

The validation study, which used 72 reference substances in a phased validation study design, was initiated in August 2002 and completed in January 2005. Upon completion, NICEATM, in coordination with the ICCVAM Acute Toxicity Working Group (ATWG) and ICCVAM, prepared a comprehensive draft background review document (BRD) containing the study results and analyses. ICCVAM subsequently convened an international independent Peer Review Panel (hereafter, Panel) meeting on May 23, 2006, to review the BRD, to evaluate the extent to which established validation and acceptance criteria had been addressed for the two methods, and to provide comments on draft ICCVAM recommendations on test method uses, future studies, draft test method protocols, and draft performance standards. The Panel meeting was open to the public and members of the public were provided an opportunity to submit written comments in advance of the meeting or verbally at the meeting. Public comments were also solicited on the Panel report via a Federal Register (FR) notice\(^4\) announcing the availability of the Panel report. The draft BRD, the Panel report, and all public comments were then made available to the Scientific Advisory Committee on Alternative Toxicological Methods (SACATM)\(^5\), for its consideration during a public teleconference meeting. The SACATM agreed with the consensus conclusions of the Panel\(^6\).

ICCVAM and the ATWG considered the Panel report, public comments, and SACATM comments in preparing the final test method recommendations provided in this report. Briefly, ICCVAM recommends that, while the two standardized \textit{in vitro} test methods (3T3 and NHK NRU test methods) are not sufficiently accurate to predict acute oral toxicity for the purposes of hazard classification, they can be used in a weight-of-evidence\(^7\) approach to determine the starting dose for the current acute oral \textit{in vivo} toxicity protocols. ICCVAM recommends that these test methods should be considered and used where determined appropriate before testing is conducted using animals. This approach should reduce the number of animals needed for acute oral toxicity testing studies, and for highly toxic substances, it should reduce the numbers of animals that die or need to be humanely killed.

In accordance with the ICCVAM Authorization Act of 2000, this report will be made available to the public and provided to U.S. Federal agencies for consideration. Agencies

---

\(^3\) The RC is a database of acute oral LD\(_{50}\) values for rats and mice obtained from RTECS\(^\circ\) and IC\(_{50}\) values from \textit{in vitro} cytotoxicity assays using multiple cell lines and cytotoxicity endpoints for 347 chemicals with known molecular weights (Halle 1998, 2003).

\(^4\) Vol. 71, No. 132, pp. 39122-39123.

\(^5\) The SACATM advises the ICCVAM, NICEATM, and the Director of the NIEHS on priorities and activities related to the development, validation, scientific review, regulatory acceptance, implementation, and national and international harmonization of new, revised, and alternative toxicological test methods.

\(^6\) SACATM (2006).

\(^7\) A weight-of-evidence approach is the use of the strengths and weaknesses of a collection of information as the basis for a conclusion that may not be evident from the individual data. For estimating starting doses, \textit{in vitro} data is considered, or “weighed” along with all other data and information (“evidence”), such as the LD\(_{50}\) of related substances, quantitative structure-activity relationship (QSAR) predictions, and other existing data, to estimate a dose that is likely to be close to the actual LD\(_{50}\) value.
with applicable testing regulations and/or guidelines are required by law to respond to ICCVAM within 180 days after receiving the recommendations. These responses will be made available to the public on the ICCVAM website (http://iccvam.niehs.nih.gov) in accordance with the ICCVAM Authorization Act requirements.

**Acknowledgments**

The efforts of many individuals who contributed to the preparation, review, and revision of this report are gratefully acknowledged. We especially recognize the members of the Panel for their thoughtful evaluations and generous contributions of time and effort. Special thanks are extended to Dr. Katherine Stitzel for serving as the Peer Review Panel Chair. The efforts of the ATWG were invaluable for assuring a meaningful and comprehensive review. We thank the chair of the ATWG, Dr. Marilyn Wind (Consumer Products Safety Commission [CPSC]) for her effective leadership of this group. The efforts of the NICEATM staff in preparing the BRD, organizing the Panel meeting, and preparing this final report are greatly appreciated. Finally, we want to acknowledge the invaluable contributions of the laboratory staff and study directors for the validation study, the international validation study management team, and the project coordinators for the independent validation study, Dr. Judy Strickland and Mr. Michael Paris. This was the first joint validation study by NICEATM and ECVAM, and we want to thank all of the team for ensuring excellent international coordination and communication. The experiences gained from this international cooperation are already facilitating a recently initiated second collaborative validation study involving NICEATM, ECVAM, and the newly established Japanese Center for the Validation of Alternative Methods (JaCVAM). International collaboration by these three centers of validation excellence will ensure high quality validation studies and take advantage of broad international expertise and experience with scientific validation.

William S. Stokes, D.V.M. D.A.C.L.A.M.
RADM, U.S. Public Health Service
Director, NICEATM
Executive Director, ICCVAM

Leonard S. Schechtman, Ph.D.
Deputy Director for Washington Operations
National Center for Toxicological Research
U.S. Food and Drug Administration
Chairman, ICCVAM
EXECUTIVE SUMMARY

This Test Method Evaluation Report (TMER) describes an evaluation by the Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) of the use of \textit{in vitro} basal cytotoxicity test methods for estimating starting doses for acute oral toxicity tests. This evaluation provides validation information that should be helpful to various stakeholders (e.g., applicable U.S. Federal regulatory agencies, the international regulatory community, the pesticide and other commercial chemical industries) in determining when these test methods might be useful for specific testing situations. Appropriate use of these \textit{in vitro} test methods is expected to further reduce and refine\textsuperscript{8} animal use for acute oral toxicity testing.

An international, multi-laboratory validation study to evaluate the usefulness and limitations of two \textit{in vitro} neutral red uptake (NRU) test methods was organized by the National Toxicology Program (NTP) Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM) and the European Centre for the Validation of Alternative Methods (ECVAM). In the validation study, three laboratories tested 72 reference substances for cytotoxicity in BALB/c 3T3 mouse fibroblasts (3T3) and normal human epidermal keratinocytes (NHK). The resulting data were used to estimate starting doses for rodent acute oral toxicity testing, based on linear regressions developed from the Registry of Cytotoxicity (RC\textsuperscript{9}) database.

NICEATM, in coordination with the ICCVAM Acute Toxicity Working Group (ATWG) and ICCVAM, prepared a comprehensive draft background review document (BRD) to describe results and analyses generated from the study. On March 21, 2006, public availability of the draft BRD was announced in a \textit{Federal Register (FR)} notice\textsuperscript{10}. An international independent Peer Review Panel (hereafter, Panel) convened in a public session by ICCVAM on May 23, 2006, reviewed the BRD, evaluated the extent that the BRD addressed established validation and acceptance criteria, and provided comment on the draft ICCVAM recommendations on the use of these test methods, future studies, draft test method protocols, and draft performance standards.

On July 11, 2006, an \textit{FR} notice\textsuperscript{11} announced the public availability of and requested public comments on the Peer Review Panel Report: The Use of \textit{In Vitro} Basal Cytotoxicity Test Methods for Estimating Starting Doses for Acute Oral Systemic Toxicity Testing. The Panel Report indicated that the information presented in the draft BRD was generally sufficient for its purpose. The Panel concluded that the applicable validation criteria were adequately

\textsuperscript{8} A reduction alternative is a new or modified test method that reduces the number of animals required. A refinement alternative is a new or modified test method that refines procedures to lessen or eliminate pain or distress in animals or enhances animal well-being (ICCVAM 2003).

\textsuperscript{9} The RC is a database of acute oral LD\textsubscript{50} values for rats and mice obtained from RTECS\textsuperscript{®} and IC\textsubscript{50} values from \textit{in vitro} cytotoxicity assays using multiple cell lines and cytotoxicity endpoints for 347 chemicals with known molecular weights (Halle 1998, 2003).

\textsuperscript{10} Vol. 71, No. 54, pp. 14229-14231; available at \url{http://iccvam.niehs.nih.gov/methods/invitro.htm}.

\textsuperscript{11} Vol. 71, No. 132, pp. 39122-39123; available at \url{http://iccvam.niehs.nih.gov/methods/invitro.htm}. 
addressed for use of these in vitro test methods in a weight-of-evidence\textsuperscript{12} approach to determine starting doses for acute oral toxicity tests.

The accomplishments of the validation study included standardization and optimization of the two NRU protocols that were evaluated and improvement of the LD\textsubscript{50}\textsuperscript{13} database for the 72 reference substances after review of the literature values. The IC\textsubscript{50} results obtained using the protocols showed that the IC\textsubscript{50} values in the RC could generally be reproduced with a single cell type and in vitro cytotoxicity endpoint\textsuperscript{14}. Although the validation study improved the in vivo LD\textsubscript{50} data for the reference chemicals by evaluating LD\textsubscript{50} values from the scientific literature, IC\textsubscript{50} - LD\textsubscript{50} regressions calculated using the validation study data were not different from those calculated using RC data. The validation study also characterized the reproducibility of the NRU test methods and estimated the animal savings that would occur when they are used to determine starting doses for the Up-and-Down Procedure (UDP) (OECD 2001a; EPA 2002a) and the Acute Toxic Class (ATC) method (OECD 2001b).

**Accuracy and Reliability**
The NICEATM/ECVAM validation study standardized the 3T3 and NHK NRU test methods and improved the LD\textsubscript{50} database for 72 substances. IC\textsubscript{50} - LD\textsubscript{50} regressions were performed for each in vitro NRU test method. The resulting IC\textsubscript{50} - LD\textsubscript{50} regressions are consistent with and support continued use of the RC database. The RC rat-only millimole regression, which is applicable to substances with known molecular weight, was based on 282 (of 347) RC substances with rat oral LD\textsubscript{50} data. The RC rat-only data were converted to a weight basis (i.e., mg/kg) to develop the RC rat-only weight regression, which is applicable to mixtures or other substances without a known molecular weight. The accuracy of the in vitro NRU test methods when used with each of the regressions was characterized by determining the proportion of reference substances for which their Globally Harmonized System of Classification and Labelling of Chemicals (GHS; UN 2005) categories (based on rat acute oral LD\textsubscript{50} data) were correctly predicted.

Using the RC rat-only millimole regression, the 3T3 NRU test method correctly predicted the GHS hazard category of 31% (21/67) of the reference substances successfully tested, while the NHK NRU test method predicted correctly 29% (20/68 reference substances). The accuracy of the 3T3 NRU test method was 69% (46/67 reference substances) for correct category prediction ±1 category. The corresponding accuracy of the NHK NRU test method was 75% (51/68 reference substances) for correct category prediction ±1 category.

Using the RC rat-only weight regression, both NRU test methods correctly predicted the GHS hazard category of 31% (21/67 - 3T3; 21/68 - NHK) of the reference substances successfully tested. The accuracy for the 3T3 NRU test method was 75% (50/67 reference substances) for correct category prediction ±1 category. The corresponding accuracy for the

\textsuperscript{12} A weight-of-evidence approach is the use of the strengths and weaknesses of a collection of information as the basis for a conclusion that may not be evident from the individual data. For estimating starting doses, in vitro data is considered, or “weighed” along with all other data and information (“evidence”), such as the LD\textsubscript{50} of related substances, quantitative structure-activity relationships (QSAR) predictions, and other existing data, to estimate a dose that is likely to be the closest to the actual LD\textsubscript{50} value.

\textsuperscript{13} The LD\textsubscript{50} is the dose that produces lethality in 50% of the test animals.

\textsuperscript{14} The IC\textsubscript{50} is the test substance concentration that produces 50% inhibition of the endpoint measured.
NHK NRU test method was 75% (51/68 reference substances) for correct category prediction ±1 category.

Reproducibility was evaluated using the results from the 64 reference substances tested in 3T3 cells and the 68 substances tested in NHK cells that yielded IC\textsubscript{50} values in all three laboratories. Intra- and inter-laboratory reproducibility of the 3T3 and NHK NRU IC\textsubscript{50} data were assessed using analysis of variance (ANOVA), coefficient of variation (CV) analysis, comparison of the laboratory-specific IC\textsubscript{50} - LD\textsubscript{50} regressions, and comparison of maximum:minimum mean laboratory IC\textsubscript{50} values.

Results for the positive control (sodium lauryl sulfate [SLS]) IC\textsubscript{50} values from the 3T3 NRU test method indicated that there were significant differences among laboratories (p = 0.006, ANOVA), but not between study phases within laboratories (p > 0.01). In addition, interlaboratory CV values were relatively low (2 to 16%). Results for the SLS IC\textsubscript{50} from the NHK NRU test method showed significant differences among laboratories (p < 0.001) and among study phases within laboratories (p ≤ 0.001). The use of a different cell culture method at the Fund for the Replacement of Animals in Medical Experiments Alternatives Laboratory (FAL) was considered to be responsible for SLS IC\textsubscript{50} differences among the laboratories in test Phases Ia and Ib. Interlaboratory CV values were 39% and 21%, respectively, for Phases Ia and Ib, and 31% and 8%, respectively, for Phases II and III. The linear regression analyses of the SLS IC\textsubscript{50} over time (within each laboratory) for both test methods indicated that IC\textsubscript{50} values generated over the duration of the validation study were stable.

ANOVA for the reference substances showed significant laboratory differences for 23 substances with the 3T3 NRU test method, and six substances with the NHK NRU test method. Mean intralaboratory CV values were 26% for both test methods, but the NHK NRU test method had a lower mean interlaboratory CV (28% vs. 47%). The maximum:minimum mean laboratory IC\textsubscript{50} ratios for the 3T3 NRU test method ranged from 1.1 to 21.6, with 58% (37/64) of the reference substances having ratios of less than 2.5. The maximum:minimum mean laboratory IC\textsubscript{50} ratios for the NHK NRU test method ranged from 1.0 to 107.6, with 85% (58/68) reference substances having ratios of less than 2.5. Thus, overall, reproducibility was generally better with the NHK NRU test method.

**Animal Reduction and Refinement**

The NICEATM/ECVAM validation study used computer models to simulate the in vivo testing of the reference substances in the UDP (OECD 2001a; EPA 2002a) and the ATC method (OECD 2001b), using either the default starting dose (175 mg/kg for the UDP, 300 mg/kg for the ATC) or the starting dose determined using the 3T3 and NHK NRU test methods. The simulations were used to estimate, per substance, the number of animals that would be used and their associated survival rate. The modeling was performed using five different dose-mortality slopes\textsuperscript{15} (i.e., 8.3, 4.0, 2.0, 0.8, and 0.5) because slope information was not available for many of the reference substances. Both RC rat-only regressions were used to determine starting doses from IC\textsubscript{50} data obtained using either the 3T3 or NHK NRU test methods. In principle, animal savings with the Fixed Dose Procedure (FDP; OECD

\textsuperscript{15} The dose-mortality slope is the slope of the dose-response curve for mortality.
2001c) could be estimated even though death is not the primary endpoint, but the validation study did not include this analysis.

Computer simulation of the UDP testing showed that, for the substances with rat acute oral LD$_{50}$ reference data tested in the validation study (67 substances for 3T3, 68 substances for NHK) an average of 0.49 animals (6.2%) to 0.66 animals (7.0%) would be saved. No animal savings were predicted for reference substances with 50 &lt; LD$_{50}$ ≤300 mg/kg, which is the range where the default starting dose of 175 mg/kg occurs. The highest animal savings were predicted for substances with 2000 &lt; LD$_{50}$ ≤5000 mg/kg and LD$_{50}$ &gt;5000 mg/kg for both NRU test methods (1.28 [11.9%] to 1.65 animals [16.7%] per test). The greatest animal savings were observed for substances in these categories because the limit test, which would be used for such substances, uses fewer animals than the main test. Although using the 3T3 and NHK NRU IC$_{50}$ values to estimate starting doses for the simulated UDP deceased the number of animals used, it did not change the number of animals that would be expected to be euthanized or die.

Computer simulation of ATC method testing showed that, for the substances tested in the validation study, NRU test methods resulted in an average savings of 0.51 animals (4.8%) to 1.09 animals (10.2%) per test. No animal savings were predicted for substances with 300 &lt; LD$_{50}$ ≤2000 mg/kg, which is where the default dose of 300 mg/kg would have been used. Mean animal savings for substances with 2000 &lt; LD$_{50}$ ≤5000 mg/kg ranged from -0.03 animals (-0.03%) to 0.11 animals (0.9%) for the RC rat-only millimole regression and from 0.53 animals (4.7%) to 2.43 animals (20.5%) for the RC rat-only weight regression. For both regressions evaluated, mean animal savings for substances with LD$_{50}$ &gt;5000 mg/kg ranged from 2.03 animals (17.1%) to 3.33 animals (27.7%). The greatest reduction in animal use occurs for substances in this category because the limit test uses fewer animals than the main test.

The use of the IC$_{50}$-based starting doses did not significantly alter the GHS category outcomes of the simulated UDP (based on LD$_{50}$ outcome) 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 simulated GHS category outcomes because the accuracy and animals savings analyses used different standards for comparison.

**ICCVAM Test Method Recommendations: Current Uses**

ICCVAM’s recommendations for use of the in vitro NRU test methods are as follows:

1. The 3T3 and NHK NRU test methods are not sufficiently accurate to predict acute oral toxicity for the purpose of regulatory hazard classification.
2. For the purposes of acute oral toxicity testing, the 3T3 and NHK NRU test methods may be used in a weight-of-evidence approach to determine the
starting dose for the current acute oral toxicity protocols (i.e., the UDP, the ATC method).

3. Consistent with the U.S. Government Principles on the Use of Animals in Research, Testing, and Education\(^{16}\), and the U.S. Public Health Service Policy on Humane Care and Use of Laboratory Animals (PHS 2002), \textit{in vitro} basal cytotoxicity test methods as part of a weight-of-evidence approach to estimate the starting dose for acute oral \textit{in vivo} toxicity test methods should be considered and used where appropriate before testing is conducted using animals. For some types of substances, this approach will reduce the number of animals needed. In some testing situations, the approach may also reduce the numbers of animals that die or need to be humanely killed.

4. The starting doses for substances with certain toxic mechanisms that are not expected to be active in 3T3 or NHK cells (e.g., those that are neurotoxic or cardiotoxic) will likely be underpredicted by these \textit{in vitro} basal cytotoxicity test methods. Therefore, the results from basal cytotoxicity testing with such substances may not be appropriate for estimating starting doses.

5. The regression formula used to determine starting doses for test substances with known molecular weights and high purity should be the revised RC millimole regression line, based on substances with rat LD\(_{50}\) data, with IC\(_{50}\) values in mmol/L and LD\(_{50}\) values in mmol/kg. The regression formula used to determine starting doses for mixtures, test substances with low or unknown purity, or test substances with unknown molecular weights should be the revised RC regression line, based on substances with rat LD\(_{50}\) data, with IC\(_{50}\) values in µg/mL and LD\(_{50}\) values in mg/kg.

6. The performance of other \textit{in vitro} basal cytotoxicity test methods that are based on similar scientific principles and that measure or predict the same biological response (i.e., basal cytotoxicity and the rat acute oral LD\(_{50}\) value, respectively) should be demonstrated to meet or exceed the accuracy and reliability of the 3T3 and NHK NRU test methods.

7. Compared to the NHK NRU test method, the 3T3 NRU test method appears to be less labor intensive and less expensive to conduct; therefore, the 3T3 NRU test method is recommended for general use. Although the 3T3 NRU test method was less reproducible than the NHK NRU test method, it produced slightly higher animal savings and accuracy for prediction of GHS acute oral toxicity category using the IC\(_{50}\) and the revised RC regressions evaluated for the prediction of LD\(_{50}\).

\textbf{ICCVAM Recommended Test Method Protocols}

ICCVAM recommends the use of \textit{in vitro} NRU protocols that are compliant with Good Laboratory Practice guidelines (OECD 1998; EPA 2003a, 2003b; FDA 2003). The recommended protocols, provided in Appendix C, use 3T3 or NHK cells with a 48-hour exposure duration for test substances. After test substance exposure, cells are incubated with neutral red (NR) dye. NRU is determined by the comparison with the optical density

---

measurements of untreated vehicle controls. The IC$_{50}$ is calculated by applying a Hill function to the concentration-response data.

**ICCVAM Recommendations: Performance Standards**

The purpose of performance standards is to communicate the basis by which adequately validated new proprietary (e.g., copyrighted, trademarked, registered) and nonproprietary test methods have been determined to have sufficient accuracy and reliability for specific testing purposes (see **Section 3**). The three elements of performance standards are:

- Essential test method components (i.e., structural, functional, and procedural elements of a validated test method that a proposed, mechanistically and functionally similar test method should adhere to)
- A minimum list of reference chemicals for assessing the accuracy and reliability of the proposed test method
- The accuracy and reliability values that should be achieved by the proposed test method using the minimum list of reference chemicals.

The test method performance standards provided in this report can be used to evaluate the acceptability of other *in vitro* basal cytotoxicity test methods that are based on similar scientific principles and that measure or predict the same biological response (i.e., basal cytotoxicity and the rat acute oral LD$_{50}$, respectively). Such methods should adhere to the essential test method components recommended in this report (see **Section 3.1**). Similar test methods can be evaluated by testing 30 reference substances (see **Table 3-1**) that cover all six hazard classification categories (i.e., the entire range of acute oral toxicity). The performance of the test method should be comparable to or better than the accuracy and reliability of the 3T3 and NHK NRU test methods in order to be considered acceptable for determining starting doses for acute oral toxicity tests or for use in a battery of *in vitro* test methods for estimating acute oral toxicity (see **Section 3.3**).

**ICCVAM Recommendations: Future Studies**

ICCVAM recommends the following future studies in order to advance the use of *in vitro* methods for assessing acute oral toxicity for regulatory hazard classification purposes:

1. Additional data should be collected using the 3T3 NRU basal cytotoxicity test method to evaluate its usefulness for predicting the rodent acute oral toxicity of chemical mixtures.

2. To supplement the high quality validation database started by this study, additional high quality comparative *in vitro* basal cytotoxicity data should be collected when rat acute oral toxicity testing is conducted. However, *in vivo* testing should not be conducted solely to collect data to assess the usefulness of the NRU test method. Periodic evaluations of the expanded database should be conducted to further characterize the usefulness and limitations of using *in vitro* cytotoxicity data as part of a weight-of-evidence approach to estimate starting doses.

3. Additional efforts should be conducted to identify *in vitro* tests and other methods necessary to achieve accurate acute oral hazard classification; studies should be conducted to investigate the potential use of *in vitro* cell-based test methods that incorporate mechanisms of action and evaluations of ADME
(absorption, distribution, metabolism, excretion) to provide improved estimates of acute toxicity hazard categories. Methods should be developed to extrapolate from in vitro toxic concentrations to equivalent doses in vivo.

4. The in vivo database of reference substances used in this validation study should be used to evaluate the utility of other non-animal approaches to estimate starting doses for acute oral toxicity tests (e.g., widely available software that uses quantitative structure-activity relationships [QSAR]).

5. Standardized procedures to collect in vivo measurements and observations pertinent to an understanding of the mechanisms of lethality should be included in future rat acute oral toxicity studies. Such information will likely be necessary to support the further development of predictive mechanism-based in vitro methods.

6. An expanded list of reference substances with rat acute oral LD50 values substantiated by high quality in vivo data (including data currently held by industry) should be developed for use in future in vitro test method development and validation studies.
1.0 INTRODUCTION

The Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) is charged by the ICCVAM Authorization Act of 2000\(^{17}\) with evaluating the scientific validity of new, revised, and alternative toxicological test methods applicable to U.S. Federal agency safety testing requirements. Following such evaluations, ICCVAM is required to provide recommendations to U.S. Federal agencies regarding the usefulness and limitations of such methods.

1.1 Evaluation of the Use of *In Vitro* Cytotoxicity Test Methods to Estimate Acute Oral Toxicity

ICCVAM initiated a review of the validation status of *in vitro* methods for estimating acute oral toxicity in 1999, in response to a request from the U.S. Environmental Protection Agency (EPA) Office of Pesticides, Prevention, and Toxic Substances. This request was based on recently published studies that showed a correlation between *in vitro* cytotoxicity and *in vivo* acute toxicity. In October of 2000, the National Toxicology Program (NTP), the National Institute of Environmental Health Sciences (NIEHS), and the EPA sponsored the International Workshop on *In Vitro* Methods for Assessing Acute Systemic Toxicity, which was announced in the *Federal Register* (FR) (Vol. 65, No. 184, pp. 57203-57205; available at http://iccvam.niehs.nih.gov/methods/invidocs/6557203.htm). Invited scientific experts and ICCVAM agency scientists were assigned to one of four Breakout Groups and prepared recommendations on the following:

- *In Vitro* Screening Methods for Assessing Acute Toxicity
- *In Vitro* Methods for Toxicokinetic Determinations
- *In Vitro* Methods for Predicting Organ Specific Toxicity
- Chemical Data Sets for Validation of *In Vitro* Acute Toxicity Test Methods

Workshop participants concluded that none of the proposed *in vitro* methods reviewed had been formally evaluated for reliability and relevance, and that their usefulness and limitations for generating information to meet regulatory requirements for acute toxicity testing had not been adequately assessed. However, an *in vitro* approach previously proposed by ZEBET (the German Centre for Documentation and Evaluation of Alternative Methods to Animal Experiments) was recommended by workshop participants as a high priority for rapid adoption so that data could be generated to establish its usefulness with a large number of chemicals (ICCVAM 2001a). The proposal was to use *in vitro* cytotoxicity data to estimate starting doses for *in vivo* acute toxicity studies. Since a correlation between IC\(_{50}\)\(^{18}\) and LD\(_{50}\)\(^{19}\) values had been determined based on retrospective literature reviews, such a strategy might reduce the use of animals for acute oral toxicity tests by identifying a starting dose closer to the LD\(_{50}\). To provide sample *in vitro* cytotoxicity protocols and instructions for using *in vitro* data to predict starting doses for acute rodent oral toxicity tests, the *Guidance Document on Using In Vitro Data to Estimate In Vivo Starting Doses for Acute Toxicity* (ICCVAM 2001b) was prepared by workshop participants with the assistance of ICCVAM and NICEATM.


\(^{18}\) The IC\(_{50}\) is the test substance concentration that produces 50% inhibition of the endpoint measured.

\(^{19}\) The LD\(_{50}\) is the dose that produces lethality in 50% of the test animals.
1.2 Evaluation of the Use of *In Vitro* Cytotoxicity Test Methods to Estimate Starting Doses for Acute Oral Toxicity Tests

ICCVAM agreed with workshop participants that *in vitro* basal cytotoxicity test methods should have a high priority for validation studies. Therefore, the NTP Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM) collaborated with the European Centre for the Validation of Alternative Methods (ECVAM), a component of the European Commission’s Joint Research Centre, to further characterize the usefulness of *in vitro* cytotoxicity assays as predictors of starting doses for acute oral systemic toxicity assays. NICEATM and ECVAM designed a multi-laboratory validation study using 72 reference substances to evaluate the performance of two standardized *in vitro* neutral red uptake (NRU) test methods, based on the ZEBET approach using the Registry of Cytotoxicity (RC)\(^{20}\) millimole regression model. The objectives for the validation study were to:

- Further standardize and optimize the *in vitro* NRU cytotoxicity protocols using BALB/c 3T3 mouse fibroblasts (3T3) and normal human epidermal keratinocytes (NHK) to maximize test method reliability (intralaboratory repeatability, intra- and inter-laboratory reproducibility)
- Assess the accuracy of the two standardized *in vitro* 3T3 and NHK NRU basal cytotoxicity test methods for estimating rodent oral LD\(_{50}\) values across the five United Nations (UN) Globally Harmonized System of Classification and Labelling of Chemicals (GHS; UN 2005) categories of acute oral toxicity, as well as unclassified toxicities
- Estimate the reduction and refinement in animal use achievable from using the *in vitro* 3T3 and NHK NRU test methods to identify starting doses for *in vivo* acute oral toxicity tests, assuming that no other information were available
- Develop high quality *in vivo* acute oral lethality and *in vitro* NRU cytotoxicity databases that can be used to support the investigation of other *in vitro* test methods necessary to improve the prediction of *in vivo* acute oral lethality

The validation study proceeded in four phases so that the Study Management Team (SMT) could evaluate the reproducibility of results after each phase and refine the protocols, if necessary, before proceeding to the next phase. Three laboratories participated in testing the 72 reference substances using the 3T3 and NHK NRU test methods, beginning in August 2002 and ending in January 2005:

- The U.S. Army Edgewood Chemical Biological Center, Edgewood, MD (ECBC)
- Fund for the Replacement of Animals in Medical Experiments Alternatives Laboratory, Nottingham, UK (FAL)
- The Institute for *In Vitro* Sciences, Gaithersburg, MD (IIVS)

\(^{20}\) The RC is a database of acute oral LD\(_{50}\) values for rats and mice obtained from RTECS\(^{\circledR}\) and IC\(_{50}\) values from *in vitro* cytotoxicity assays using multiple cell lines and cytotoxicity endpoints for 347 chemicals with known molecular weights (Halle 1998, 2003).
BioReliance Corporation (Rockville, MD) procured and distributed the coded reference substances and performed solubility tests prior to their distribution to the testing laboratories, but did not perform any of the \textit{in vitro} tests.

NICEATM, in coordination with the ICCVAM Acute Toxicity Working Group (ATWG) and ICCVAM, prepared a comprehensive draft background review document (BRD) to summarize the procedures and results generated from the validation study. On March 21, 2006, the availability of the draft BRD was announced in an \textit{FR} notice\textsuperscript{21}. The BRD was made available to the public in electronic format on the ICCVAM/NICEATM website (available at \url{http://iccvam.niehs.gov}) and in print upon request to NICEATM.

1.3 \textbf{Peer Review of the NICEATM/ECVAM Validation Study}

An international independent Peer Review Panel (hereafter, Panel) convened by ICCVAM on May 23, 2006, reviewed the BRD, evaluated the extent that the BRD addressed established validation and acceptance criteria, and provided comment on the draft ICCVAM recommendations on the use of these test methods, future studies, draft test method protocols, and draft performance standards. Comments from the public and scientific community were provided to the Panel and made available on the ICCVAM/NICEATM website (\url{http://iccvam.niehs.nih.gov/methods/invitro.htm}). On July 11, 2006, an \textit{FR} notice\textsuperscript{22} announced the availability of the Peer Review Panel Report: The Use of In Vitro Basal Cytotoxicity Test Methods for Estimating Starting Doses for Acute Oral Systemic Toxicity Testing. The Panel report (see Appendix A) indicated that the information presented in the draft BRD was generally sufficient for its purpose. The Panel concluded that the objectives of the validation study were appropriate, and agreed that the applicable validation criteria were adequately addressed for use of these \textit{in vitro} test methods in a weight-of-evidence\textsuperscript{23} approach to determine starting doses for acute oral toxicity tests.

Regarding the draft ICCVAM recommendations for test method uses, the Panel agreed that:

- Neither of the NRU test methods can be used as alternatives for the \textit{in vivo} acute oral toxicity test for the purposes of hazard classification.
- The \textit{in vitro} NRU test methods may be useful in a weight-of-evidence approach to determine the starting dose for acute oral \textit{in vivo} toxicity protocols.
- The NRU test methods should be considered before animals are used.
- The RC rat-only regression should be used to estimate the LD\textsubscript{50} from IC\textsubscript{50} data. When the molecular weight of a test substance is known, the molar regression should be used; however, a regression based on weight rather than molar units should be used when the molecular weight of the test substance is unknown.

\textsuperscript{21} Vol. 71, No. 54, pp. 14229-14231; available at \url{http://iccvam.niehs.nih.gov/methods/invitro.htm}.

\textsuperscript{22} Vol. 71, No. 132, pp. 39122-39123; available at \url{http://iccvam.niehs.nih.gov/methods/invitro.htm}.

\textsuperscript{23} A weight-of-evidence approach is the use of the strengths and weaknesses of a collection of information as the basis for a conclusion that may not be evident from the individual data. For estimating starting doses, \textit{in vitro} data is considered, or “weighed” along with all other data and information (“evidence”), such as the LD\textsubscript{50} of related substances, quantitative structure-activity relationships (QSAR) predictions, and other existing data, to estimate a dose that is likely to be the closest to the actual LD\textsubscript{50} value.
• Other *in vitro* basal cytotoxicity test methods that are based on similar scientific principles and that measure or predict the same biological response (i.e., basal cytotoxicity and the rat acute oral LD$_{50}$ value, respectively) should meet or exceed the accuracy and reliability of the 3T3 and NHK NRU test methods.

• The 3T3 NRU test method, based on relative ease of performance and cost, should be recommended for general use, but cautioned that one test method should not be preferred over the other.

• The NRU test methods are appropriate for substances that interfere with energy utilization or alkylation of proteins and other macromolecules.

Regarding the draft ICCVAM recommendations for future studies, the Panel agreed that:

• Additional data for the 3T3 NRU test method should be collected to evaluate its usefulness for predicting starting doses with chemical mixtures.

• High quality comparative *in vitro* basal cytotoxicity data should be collected in tandem with *in vivo* rat acute oral toxicity test results to further evaluate the use of these test methods for predicting the starting doses for acute oral toxicity tests.

• Additional *in vitro* tests and other methods necessary to achieve accurate acute oral hazard classification should be investigated.

• The *in vivo* database of reference substances used in the validation study should be used to evaluate the utility of other non-animal approaches to estimate starting doses for rat acute oral toxicity tests.

• Standardized procedures to collect information pertinent to an understanding of the mechanisms of lethality should be included, to the extent possible, in future rat acute oral toxicity studies.

• An expanded list of reference substances with estimated rat LD$_{50}$ values substantiated by high quality *in vivo* data should be developed for use in future *in vitro* test development and validation.

The draft BRD, the Panel report, and all associated public comments were made available to the Scientific Advisory Committee on Alternative Toxicological Methods (SACATM$^{24}$) for their consideration. The SACATM endorsed the Panel Report. ICCVAM and the ATWG then considered the Panel Report, all public comments, and the comments of SACATM in preparing the final test method recommendations that are provided in this report. This report will be made available to the public and provided to U.S. Federal agencies for consideration, in accordance with the ICCVAM Authorization Act of 2000. The final BRD, *In Vitro Cytotoxicity Test Methods for Estimating Acute Oral Systemic Toxicity* (ICCVAM 2006), revised in response to the Panel and ATWG comments, will also be provided as background information and technical support for this report. Agencies with applicable testing regulations and guidelines (Appendix B) are required by law to respond to ICCVAM within 180 days of

$^{24}$ The SACATM advises the ICCVAM, NICEATM, and the Director of the NIEHS on priorities and activities related to the development, validation, scientific review, regulatory acceptance, implementation, and national and international harmonization of new, revised, and alternative toxicological test methods.
receiving the ICCVAM recommendations. These responses will be made available to the public on the ICCVAM website (http://iccvam.niehs.nih.gov) as they are received.

1.4 Report Organization

Section 1.0 of this report provides the background of the NICEATM/ECVAM validation study for the use of in vitro cytotoxicity test methods to predict starting doses for acute oral toxicity test methods and this resulting ICCVAM test method evaluation report.

Section 2.0 describes the NRU protocols evaluated in the validation study, the reference substances tested, and the accuracy and reliability results from the validation study. Also included are ICCVAM’s recommendations for test method uses and future studies, which were finalized after consideration of the Panel Report, public comments, and the comments of SACATM, and were based on the results of the validation study. The recommendations for future studies are intended to advance the use of alternative methods for the prediction of acute toxicity.

Section 3.0 provides recommended performance standards for application to future test methods that are based on similar scientific principles and that measure or predict the same biological or toxic effect. The three elements of performance standards are essential test method components (i.e., structural, functional, and procedural elements of a validated test method that a proposed, mechanistically and functionally similar test method should adhere to), a minimum list of reference substances for assessing the accuracy and reliability of the proposed test method, and the accuracy and reliability values that should be achieved by the proposed test method using the minimum list of reference substances.
2.0 ICCVAM RECOMMENDATIONS FOR IN VITRO NEUTRAL RED UPTAKE (NRU) BASAL CYTOTOXICITY TEST METHODS

The following technical summary provides a synopsis of the performance analysis described in the BRD (ICCVAM 2006) which indicates the current validation status of the in vitro 3T3 and NHK NRU basal cytotoxicity test methods, including what is known about their reliability and accuracy, the scope of the substances tested, and standardized protocols. These results form the basis for the ICCVAM Recommendations for test method uses and future studies that are presented at the end of this section.

2.1 Test Method Description

The NRU cytotoxicity assay procedure is based on the ability of viable cells to incorporate and bind neutral red (NR), a supravital dye. NR is a weakly cationic dye that readily diffuses through the plasma membrane and concentrates in lysosomes where it electrostatically binds to the anionic lysosomal matrix. Toxicants can alter the cell surface or the lysosomal membrane to cause lysosomal fragility and other adverse changes that gradually become irreversible. Thus, cell death and/or inhibition of cell growth decreases the amount of NR retained by the culture. Healthy proliferating mammalian cells, when properly maintained in culture, continuously divide and multiply over time. A toxic substance, regardless of site or mechanism of action, will interfere with this process and result in a reduction of the growth rate as reflected by cell number. Cytotoxicity is expressed as a concentration dependent reduction of the uptake of NR after substance exposure to the cells, thus providing a sensitive, integrated signal of both cell integrity and growth inhibition.

2.1.1 General Test Method Procedures

3T3 and NHK cell cultures are grown in 96-well microtiter plates and exposed to a reference substance and/or positive control (PC). After the predetermined incubation time, the reference substance and PC are removed and NR solution is applied to the cells. The cells are incubated again, the excess NR solution is removed, and NR is eluted from the cells. The NRU is determined by using a microtiter plate reader/spectrophotometer to measure the optical density (OD; at a wavelength of 540 ±10 nm) of the eluted NR dye in the 96-well plate. A calculation of cell viability expressed as NRU is made for each concentration of a reference substance and PC by using the mean NRU OD of six replicate values (minimum of four acceptable replicate wells) per test concentration. The cell viability OD value is compared with the mean NRU OD of all vehicle control (VC) values (provided VC values have met the VC acceptance criteria). Relative cell viability is then expressed as percentage of untreated VC.

2.1.2 Protocol Similarities and Differences for the 3T3 and NHK NRU Test Methods

A number of protocol procedures and conditions are common to both the 3T3 and NHK NRU test methods (see Appendices C1 and C2 for specific protocols for the test methods). Both NRU test methods use the same solvents to dissolve reference substances and the PC, the same culture conditions, the same 96-well plate format, and the same duration of exposure, and both employ the use of a range finder test before performing the definitive (main) test. In addition, both NRU test methods follow identical NRU procedures and calculate cell viability and the IC$_{50}$ using the same procedures.
There are three differences between the protocols for the 3T3 and NHK NRU test methods. The first is the use of newborn calf serum (NCS) in the 3T3 cell culture medium. The NHK cells require a keratinocyte-specific serum-free medium. The second is that the 3T3 cells require less time (approximately 24 hours) to reach appropriate the confluence for testing than the NHK cells (approximately 24 to 72 hours). The third difference is the application and volume of test substance. For the 3T3 NRU test method, all culture medium is removed from the 3T3 cells and 50 µL/well of medium with substance is added immediately. For the NHK cells, 125 µL/well of medium with test substance is added to the 125 µL/well of medium already on the cells.

2.2 Reference Substances

Seventy-two reference substances were selected for testing in the NICEATM/ECVAM validation study. These substances were selected to represent: (1) the complete range of in vivo acute oral LD50 values; (2) the types of substances regulated by the various regulatory authorities; and (3) those with human toxicity data and/or human exposure potential. To insure that the complete range of toxicity was covered, the GHS (UN 2005) was used to select 12 substances for each of five acute oral toxicity categories and 12 unclassified substances. The set of selected reference substances had the following characteristics:

- Thirty-five percent (27/77) were pharmaceuticals, 22% (17/77) were pesticides, 10% (8/77) were solvents, and 6% (5/77) were food additives. The remaining substances were used for a variety of manufacturing and consumer products. The number of assigned uses (77) is greater than the number of selected substances because some of the substances have more than one use.
- Relevance of the substances to human exposure is indicated by the fact that 58% (42/72) were included in the Multicentre Evaluation of In Vitro Cytotoxicity (MEIC) study, 24% (17/72) of which were included also in the Evaluation-guided Development of New In Vitro Tests (EDIT) program; 64% (46/72) had human exposures reported by the Toxic Exposure Surveillance System (TESS); 71% (51/72) had been evaluated by NTP; and 25% (18/72) were on the EPA High Production Volume (HPV) list.
- Eighty-one percent (58/72) of the substances were in the RC database; 38% (22/58) of which were outliers with respect to the RC millimole regression (log LD50 [mmol/kg] = 0.435 x log IC50 [mM] + 0.625). The RC millimole regression underpredicted the toxicity of 77% (17/22) of the outliers and overpredicted the toxicity of 23% (5/22).
- Seventy-nine percent (57/72) were organic compounds and 21% (15/72) were inorganic. The most commonly represented classes of organic compounds were heterocyclics (25%, 14/57), carboxylic acids (25%, 14/57), and alcohols (18%, 10/57).
- Twenty-six percent (19/72) were known to have active metabolites and three others were expected to have active metabolites based on their chemical structures.
- Many of the substances produced toxicity in more than one organ system. The most common target organs were liver (17 substances) and kidney (15 substances). Other target organs included the nervous system (40 substances).
2.3 Test Method Accuracy

The ability of the 3T3 and NHK NRU test methods to correctly predict rodent acute oral systemic toxicity is based on the validity of the in vivo – in vitro (i.e., IC₅₀ - LD₅₀) regression model. It is the IC₅₀ - LD₅₀ regression that establishes the relationship between the 3T3 and NHK NRU IC₅₀ values and the predicted LD₅₀ values that were used to set the starting doses for the computer simulated acute oral toxicity assays performed for the NICEATM/ECVAM validation study.

The validation study tested two regressions for their use with the NRU IC₅₀ values to predict LD₅₀ values. The first regression – the RC rat-only millimole regression – was calculated using the 282 substances in the RC dataset of 347 substances that had a reported rat oral LD₅₀ value (65 substances had mouse-only LD₅₀ values). The LD₅₀ data for the regression were limited to one species to decrease the variability in LD₅₀ values produced by combining the data of two species. Rats were selected because they are the preferred species for most acute oral toxicity testing (i.e., the Up-and-Down Procedure [UDP; EPA 2002b; OECD 2001a], the Acute Toxic Class method [ATC; OECD 2001b], and the Fixed Dose Procedure [FDP; OECD 2001d]). The second regression – the RC rat-only weight regression – was a transformation of the RC rat-only millimole regression to weight units (mg/kg for LD₅₀ and µg/mL for IC₅₀) in order to make the regression applicable to the testing of mixtures and substances without a known molecular weight.

The ability of the 3T3 and NHK NRU IC₅₀ data to correctly predict rat acute oral LD₅₀ values, based on using the RC rat-only millimole regression and the RC rat-only weight regression, was evaluated by determining the extent to which the appropriate GHS acute oral toxicity category was identified for each reference substance. This approach permits an assessment of accuracy specific to each GHS hazard classification category.

Tables 2-1 and 2-2, which are divided into upper and lower sub-tables, provide accuracy data for the 3T3 and NHK NRU test methods, respectively. For each part, the toxicity categories corresponding to the reference rat acute oral LD₅₀ data are provided in rows that are labeled on the far left side of the table. The toxicity categories predicted by the in vitro NRU assays and the associated regressions are provided in columns that are labeled across the top of each part (i.e., 3T3 or NHK NRU-predicted toxicity category) of the table. The numbers at the intersections of the reference rat oral LD₅₀ rows and 3T3 or NHK NRU-predicted toxicity category columns are the numbers of substances with in vitro category predictions that correspond to the various in vivo categories. The right sides of the tables also provide summaries containing, for each rat acute oral toxicity category and for the total number of substances evaluated:

- The number of substances
- The accuracy of the 3T3 or NHK NRU prediction
- The percentage of substances for which toxicity has been overpredicted and underpredicted by the in vitro NRU test methods
In each of the 3T3 and NHK sections of the table, a summary of predictivity is also provided for each predicted toxicity category along with the percentage of substances with category (i.e., toxicity) underpredicted and overpredicted.

**Table 2-1** shows the concordance of the observed (i.e., the rat acute oral LD₅₀) and predicted GHS acute oral toxicity categories (UN 2005) for each *in vitro* NRU cytotoxicity test method using the geometric mean IC₅₀ values (of the three validation study laboratories) in the RC rat-only millimole regression, log LD₅₀ (mmol/kg) = 0.439 log IC₅₀ (mM) + 0.621. Accuracy is the agreement of the *in vitro* NRU category predictions with those based on the rat acute oral LD₅₀ reference values.

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

- Zero (0%) of six substances with LD₅₀ < 5 mg/kg was correctly predicted
- One (9%) of 11 substances in the 5 < LD₅₀ ≤ 50 mg/kg category was correctly predicted
- Five (42%) of 12 substances in the 50 < LD₅₀ ≤ 300 mg/kg category were correctly predicted
- Thirteen (81%) of 16 substances in the 300 < LD₅₀ ≤ 2000 mg/kg category were correctly predicted; however, this toxicity category was also predicted for 32 other substances (71%; 32/45) that did not match this category *in vivo*. Thus, the predictivity for this category was 29% (13/45 substances predicted for this category matched the *in vivo* category).
- None (0%) of the 10 substances in the 2000 < LD₅₀ ≤ 5000 mg/kg category were correctly predicted
- Two (17%) of the 12 substances with LD₅₀ > 5000 mg/kg were correctly predicted
Table 2-1  Prediction of GHS Acute Oral Toxicity Category by the 3T3 and NHK NRU Test Methods and the RC Rat-Only Millimole Regression

<table>
<thead>
<tr>
<th>Reference Rat Oral LD50 (mg/kg)</th>
<th>3T3 NRU-Predicted GHS Category (mg/kg)</th>
<th>Total</th>
<th>Accuracy</th>
<th>Toxicity Overpredicted</th>
<th>Toxicity Underpredicted</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>LD50 &lt;5</td>
<td>5 &lt; LD50 ≤50</td>
<td>50 &lt; LD50 ≤300</td>
<td>300 &lt; LD50 ≤2000</td>
<td>2000 &lt; LD50 ≤5000</td>
</tr>
<tr>
<td>LD50 &lt;5</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>5 &lt; LD50 ≤50</td>
<td>0</td>
<td>1</td>
<td>6</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>50 &lt; LD50 ≤300</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>7</td>
<td>0</td>
</tr>
<tr>
<td>300 &lt; LD50 ≤2000</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>13</td>
<td>0</td>
</tr>
<tr>
<td>2000 &lt; LD50 ≤5000</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>10</td>
<td>0</td>
</tr>
<tr>
<td>LD50 &gt;5000</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>8</td>
<td>2</td>
</tr>
<tr>
<td>Total</td>
<td>0</td>
<td>4</td>
<td>13</td>
<td>45</td>
<td>3</td>
</tr>
<tr>
<td>Predictivity</td>
<td>0%</td>
<td>25%</td>
<td>38%</td>
<td>29%</td>
<td>0%</td>
</tr>
<tr>
<td>Category Overpredicted</td>
<td>0%</td>
<td>25%</td>
<td>15%</td>
<td>40%</td>
<td>67%</td>
</tr>
<tr>
<td>Category Underpredicted</td>
<td>0%</td>
<td>50%</td>
<td>46%</td>
<td>31%</td>
<td>33%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Reference Rat Oral LD50 (mg/kg)</th>
<th>NHK NRU-Predicted Toxicity Category (mg/kg)</th>
<th>Total</th>
<th>Accuracy</th>
<th>Toxicity Overpredicted</th>
<th>Toxicity Underpredicted</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>LD50 &lt;5</td>
<td>5 &lt; LD50 ≤50</td>
<td>50 &lt; LD50 ≤300</td>
<td>300 &lt; LD50 ≤2000</td>
<td>2000 &lt; LD50 ≤5000</td>
</tr>
<tr>
<td>LD50 &lt;5</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>5 &lt; LD50 ≤50</td>
<td>0</td>
<td>2</td>
<td>5</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>50 &lt; LD50 ≤300</td>
<td>0</td>
<td>1</td>
<td>6</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>300 &lt; LD50 ≤2000</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>12</td>
<td>1</td>
</tr>
<tr>
<td>2000 &lt; LD50 ≤5000</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>10</td>
<td>0</td>
</tr>
<tr>
<td>LD50 &gt;5000</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>7</td>
<td>6</td>
</tr>
<tr>
<td>Total</td>
<td>0</td>
<td>5</td>
<td>15</td>
<td>40</td>
<td>8</td>
</tr>
<tr>
<td>Predictivity</td>
<td>0%</td>
<td>40%</td>
<td>40%</td>
<td>30%</td>
<td>0%</td>
</tr>
<tr>
<td>Category Overpredicted</td>
<td>0%</td>
<td>40%</td>
<td>13%</td>
<td>43%</td>
<td>75%</td>
</tr>
<tr>
<td>Category Underpredicted</td>
<td>0%</td>
<td>20%</td>
<td>47%</td>
<td>28%</td>
<td>25%</td>
</tr>
</tbody>
</table>

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

The RC rat-only millimole regression is log LD50 (mmol/kg) = log IC50 (mM) x 0.439 + 0.621. Numbers in table represent numbers of substances.

Reference rat oral LD50 values in mg/kg (see BRD Table 4-2) (ICCVAM 2006)

1Carbon tetrachloride excluded because no laboratory attained sufficient toxicity for the calculation of an IC50.

2Methanol excluded because no laboratory attained sufficient toxicity for the calculation of an IC50.

3Propylparaben excluded because no rat LD50 was identified (see BRD Table 4-2) (ICCVAM 2006).
The overall accuracy of the NHK NRU test method for correctly predicting the GHS acute oral toxicity classification, when the prediction was based on the RC rat-only millimole regression, was 29% (20/68 substances). Toxicity was overpredicted for 40% (27) and underpredicted for 31% (21) of the 68 substances. The pattern of concordance between in vitro and in vivo results for the NHK NRU test method with the RC rat-only millimole regression was similar to that for the 3T3 NRU test method with the exception that the toxicity of all substances with LD$_{50}$ >50000 mg/kg were not correctly predicted. For this analysis, in terms of each GHS acute oral toxicity classification category:

- Zero (0%) of six substances with LD$_{50}$ <5 mg/kg were correctly predicted
- Two (18%) of 11 substances in the 5 < LD$_{50}$ ≤50 mg/kg category were correctly predicted
- Six (50%) of 12 substances in the 50 < LD$_{50}$ ≤300 mg/kg categories were correctly predicted
- 12 (75%) of 16 substances in the 300 < LD$_{50}$ ≤2000 mg/kg category were correctly predicted; however, this toxicity category was also predicted for 28 (70%; 28/40) other substances with in vivo data that did not match the category. Thus, the predictivity for this category was 30% (12/40).
- Zero (0%) of 10 substances in the 2000 < LD$_{50}$ ≤5000 mg/kg category were correctly predicted
- None (0%) of 13 substances with LD$_{50}$ >5000 mg/kg were correctly predicted

Table 2-2 shows the concordance of the observed and predicted GHS acute oral toxicity categories for each in vitro NRU test method using the geometric mean IC$_{50}$ values (of the three validation study laboratories) and the RC rat-only weight regression. The regression formula for the RC rat-only weight regression is log LD$_{50}$ (mg/kg) = 0.372 log IC$_{50}$ (µg/mL) + 2.024.

The overall accuracy of the 3T3 NRU test method with the RC rat-only weight regression was 31% (21) for the results from 67 substances. The toxicity was overpredicted for 33% (24) and underpredicted for 36% (22) of the 67 substances. For this analysis, in terms of each GHS acute oral toxicity classification category:

- Zero (0%) of six substances with LD$_{50}$ <5 mg/kg were correctly predicted
- One (9%) of 11 substances in the 5 < LD$_{50}$ ≤50 mg/kg GHS acute oral toxicity category was correctly predicted
- Four (33%) of 12 substances in the 50 < LD$_{50}$ ≤300 mg/kg GHS acute oral toxicity category were correctly predicted; however, since 10 other substances were also predicted for this category, the predictivity was 29% (4/14)
- Twelve (75%) of 16 substances in the 300 < LD$_{50}$ ≤2000 mg/kg GHS acute oral toxicity category were predicted correctly. Since a total of 40 substances were predicted for this category, the predictivity was 30% (12/40)
- Four (40%) of 10 substances in the 2000 < LD$_{50}$ ≤5000 mg/kg GHS acute oral toxicity category were correctly predicted; however, since a total of 11 substances were predicted for this category, the predictivity was 36% (4/11).
- Zero (0%) of 12 substances with LD$_{50}$ >5000 mg/kg were correctly predicted
Table 2-2  Prediction of GHS Acute Oral Toxicity Category by the 3T3 and NHK NRU Test Methods and the RC Rat-Only Weight Regression

<table>
<thead>
<tr>
<th>Reference Rat Oral LD50 (mg/kg)</th>
<th>3T3 NRU-Predicted Toxicity Category (mg/kg)</th>
<th>NHK NRU-Predicted Toxicity Category (mg/kg)</th>
<th>Total</th>
<th>Accuracy</th>
<th>Toxicity Over-predicted</th>
<th>Toxicity Under-predicted</th>
</tr>
</thead>
<tbody>
<tr>
<td>LD50 &lt;5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0%</td>
<td>0%</td>
<td>100%</td>
</tr>
<tr>
<td>5 &lt; LD50 ≤50</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0%</td>
<td>0%</td>
<td>100%</td>
</tr>
<tr>
<td>50 &lt; LD50 ≤300</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0%</td>
<td>0%</td>
<td>100%</td>
</tr>
<tr>
<td>300 &lt; LD50 ≤2000</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0%</td>
<td>0%</td>
<td>100%</td>
</tr>
<tr>
<td>2000 &lt; LD50 ≤5000</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0%</td>
<td>0%</td>
<td>100%</td>
</tr>
<tr>
<td>LD50 &gt;5000</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0%</td>
<td>0%</td>
<td>100%</td>
</tr>
<tr>
<td>Total</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0%</td>
<td>0%</td>
<td>100%</td>
</tr>
<tr>
<td>Predicitivity</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>100%</td>
</tr>
<tr>
<td>Category Overpredicted</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>100%</td>
</tr>
<tr>
<td>Category Underpredicted</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>100%</td>
</tr>
</tbody>
</table>

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

1The RC rat-only weight regression is log LD50 (mg/kg) = log IC50 (µg/mL) x 0.372 + 2.024.
2Reference rat oral LD50 values in mg/kg (BRD Table 4-2) (ICCVAM 2006).
3Epinephrine bitartrate excluded because no rat LD50 was identified (see BRD Table 4-2) (ICCVAM 2006).
4Colchine excluded because no rat LD50 was identified (see BRD Table 4-2) (ICCVAM 2006).
5Carbon tetrachloride excluded because no laboratory attained sufficient toxicity for the calculation of an IC50.
6Methanol excluded because no laboratory attained sufficient toxicity for the calculation of an IC50.
7Propylparaben excluded because no rat LD50 was identified (see BRD Table 4-2) (ICCVAM 2006).
The overall accuracy of the NHK NRU test method with the RC rat-only weight regression was 31% (21/68). Toxicity was overpredicted for 37% (22) and underpredicted for 32% (25) of the 68 substances. For this analysis, in terms of each GHS acute oral toxicity classification category:

- Zero (0%) of six substances with LD$_{50} < 5$ mg/kg were correctly predicted
- One (9%) of 11 substances in the $5 <$ LD$_{50} \leq 50$ mg/kg GHS acute oral toxicity category was correctly predicted
- Five (42%) of 12 substances in the $50 <$ LD$_{50} \leq 300$ mg/kg GHS acute oral toxicity category were correctly predicted; however, since six other substances were also predicted for this category, the predictivity was 33% (3/9)
- Thirteen (81%) of 16 substances in the $300 <$ LD$_{50} \leq 2000$ mg/kg GHS acute oral toxicity category were predicted correctly; however, since 29 other substances were also predicted for this category, the predictivity was 31% (13/42)
- One (10%) of 10 substances in the $2000 <$ LD$_{50} \leq 5000$ mg/kg GHS acute oral toxicity category was correctly predicted
- One (8%) of 13 substances with LD$_{50} > 5000$ mg/kg was correctly predicted

### 2.4 Test Method Reliability (Inter- and Intra-Laboratory Reproducibility)

Reproducibility is the consistency of individual test results obtained within a single laboratory (intralaboratory reproducibility) or among different laboratories (interlaboratory reproducibility) using the same protocol and test samples. Reproducibility was evaluated using the results from the reference substances that yielded IC$_{50}$ values from all three validation study laboratories (i.e., 64 and 68 reference substances for the 3T3 and the NHK NRU test methods, respectively). Intra- and inter-laboratory reproducibility of the 3T3 and NHK NRU IC$_{50}$ data were assessed using analysis of variance (ANOVA), coefficient of variation (CV) analysis, comparison of the laboratory-specific IC$_{50}$ - LD$_{50}$ regressions to one another, and comparison of maximum:minimum mean laboratory IC$_{50}$ values (see BRD [ICCVAM 2006] Section 7 for reliability and reproducibility analyses for the NICEATM/ECVAM validation study). As indicated below, reproducibility was generally better for the NHK NRU test method.

Although ANOVA results for the PC, sodium lauryl sulfate (SLS), IC$_{50}$ values for the 3T3 NRU test method indicated there were significant differences among laboratories ($p = 0.006$) but not between study phases within laboratories ($p > 0.01$), the data show (see BRD Figure 7-5 [ICCVAM 2006]) that laboratory means and standard deviations from each testing phase overlap which indicated that the IC$_{50}$ was stable between testing phases. Interlaboratory CV values for SLS with the 3T3 NRU test method were relatively low and ranged from 2 to 16% for the various study phases. ANOVA results for the SLS IC$_{50}$ for the NHK NRU test method also showed significant differences between laboratories ($p < 0.001$) but also between study phases within laboratories ($p \leq 0.001$). A modified cell culturing method at FAL was likely responsible for SLS IC$_{50}$ differences among the laboratories in Phases Ia and Ib. Interlaboratory CV values were 39% and 21%, respectively, for Phases Ia and Ib and 31% and 8%, respectively, for Phases II and III. Very small but significantly different slopes ($p < 0.05$; slope ranges from -0.00032 to 0.00020 for 3T3 and -0.0011 to -0.0004 for NHK) for linear regression analyses of the SLS IC$_{50}$ over time (within each laboratory) for both NRU
test methods indicated that SLS IC$_{50}$ was relatively stable over the 2.5 year duration of the study.

The assessment of reproducibility for reference substances by the comparisons of laboratory-specific IC$_{50}$ - LD$_{50}$ regressions indicated that the regressions were not significantly different from one another because the regressions for each laboratory were within the 95% confidence limits of the mean laboratory regressions. The similarity of the laboratories in LD$_{50}$ predictions (via regression) for the reference substances is relevant with respect to the reproducibility analyses since the NRU methods are proposed for use with the regressions in determining starting doses for rodent acute oral toxicity tests.

ANOVA results for the reference substances showed significant laboratory differences for 23 substances for the 3T3 NRU test method, but only for six substances for the NHK NRU test method. Mean intralaboratory CV values were 26% for both methods, but the NHK NRU test method had a lower mean interlaboratory CV (28% vs 47% for 3T3). An analysis to determine the relationship, if any, between substance attributes and interlaboratory CV values indicated that physical form, solubility, and volatility had little effect on CV values. However, the magnitude of the CV seemed to be related to chemical class, GHS acute toxicity category, IC$_{50}$, and boiling point, although the usefulness of these relationships has not been established.

Mean interlaboratory CV values were larger for substances in the most toxic GHS categories than for substances in the other toxicity categories, especially with the 3T3 NRU test method. The mean interlaboratory CV for substances in the LD$_{50}$ ≤ 5 mg/kg (72%) and 5 < LD$_{50}$ ≤ 50 mg/kg (78%) classes were larger than the mean overall interlaboratory CV (47%) with the 3T3 NRU test method. The mean interlaboratory NHK CV was 37% for substances with LD$_{50}$ ≤ 5 mg/kg, and 41% for substances with 5 < LD$_{50}$ ≤ 50 mg/kg, while the mean overall interlaboratory CV was 28%. A Spearman correlation analysis showed that the IC$_{50}$ was inversely correlated to interlaboratory CV for both the 3T3 (p = 0.015) and NHK (p = 0.014) test methods, and that boiling point was positively correlated to interlaboratory CV (p = 0.007) (i.e., higher boiling points were associated with higher CV values) for the 3T3 but not the NHK NRU test method (p = 0.809).

The maximum:minimum mean laboratory IC$_{50}$ values for the 3T3 NRU test method ranged from 1.1 to 21.6, with 37 (58%) of the 64 reference substances having values less than 2.5. In contrast, the maximum:minimum mean laboratory IC$_{50}$ values for the NHK NRU test method ranged from 1.0 to 107.6, with 58 (85%) of the 68 reference substances having values less than 2.5.

2.5 Animal Welfare Considerations: Reduction, Refinement, and Replacement

Computer models were used to simulate the testing of the reference substances in two currently accepted sequential rodent acute oral toxicity test methods, the UDP (OECD 2001a; EPA 2002a) and the ATC method (OECD 2001b) using either the default starting dose (175 mg/kg for the UDP, 300 mg/kg for the ATC), or the starting dose determined by the 3T3 and NHK NRU test methods (see BRD [ICCVAM 2006] Section 10 for simulation modeling and analyses for the study). The simulations (10,000 per run for the UDP and 2000 per run for the ATC) were used to estimate, per substance, the number of animals that would be used and
their associated survival rate. The modeling was performed using five different dose-mortality slopes\(^{25}\) (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, determination of animal use included the data for only two of the slopes, 2.0 and 8.3. The slope of 2.0 is the default used for the calculation of LD\(_{50}\) by the UDP method and the slope of 8.3 represents substances, such as pesticides, with higher slopes. Starting doses determined by either 3T3 or NHK NRU were tested as were the two RC rat-only regressions, one based on molar units, the other on mg/kg \((\textit{in vivo})\) and µg/mL \((\textit{in vitro})\).

Computer simulation of the UDP testing showed that, for the substances with rat acute oral LD\(_{50}\) reference data tested in the validation study (67 for 3T3, 68 for NHK), the NRU-based starting doses resulted in the use of fewer animals for UDP testing (compared with using the default starting dose of 175 mg/kg). An average of 0.49 animals (6.2%, slope=8.3; NHK NRU test method) to 0.54 animals (5.8%, slope=2.0; 3T3 NRU test method) would be saved with the RC rat-only millimole regression (\textbf{Table 2-3}). The RC rat-only weight regression predicted mean animal savings of 0.54 animals (6.8%, slope=8.3; NHK NRU test method) to 0.66 animals (7.0%, slope=2.0; 3T3 NRU test method) (\textbf{Table 2-4}). When substances were grouped by GHS acute oral toxicity category, no animal savings were predicted for substances with 50 <LD\(_{50}\) ≤300 mg/kg; this category includes the default starting dose of 175 mg/kg. The highest statistically significant animal savings were predicted for substances with 2000 < LD\(_{50}\) ≤5000 mg/kg and LD\(_{50}\) >5000 mg/kg for both NRU test methods. The greatest animal savings were observed for substances in these categories because the limit test, which would be used for such substances, uses fewer animals than the main test. 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. Using the RC rat-only weight regression produced animal savings of 1.28 (14.0%) to 1.65 animals (16.7%) for the substances in these toxicity categories. Although using the 3T3 and NHK NRU IC\(_{50}\) values to estimate starting doses for the simulated UDP deceased the number of animals used, it did not change the number of animals that died.

\(^{25}\) The dose-mortality slope is the slope of the dose-response curve for mortality.
Table 2-3 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\(^4\)

<table>
<thead>
<tr>
<th>GHS Acute Oral Toxicity Category(^3)</th>
<th>Number of Reference Substances</th>
<th>With Default Starting Dose(^5)</th>
<th>With IC(_{50}) Based Starting Dose(^6)</th>
<th>Animals Saved(^7)</th>
<th>With Default Starting Dose(^5)</th>
<th>With IC(_{50}) Based Starting Dose(^6)</th>
<th>Animals Saved(^7)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Dose-mortality Slope = 2.0</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LD(_{50}) ≤ 5 mg/kg</td>
<td>6</td>
<td>11.32 ± 0.20</td>
<td>10.19 ± 0.70</td>
<td>1.14 (10.0%)</td>
<td>9.70 ± 0.28</td>
<td>8.74 ± 0.43</td>
<td>0.96 (9.9%)</td>
</tr>
<tr>
<td>5 &lt; LD(_{50}) ≤ 50 mg/kg</td>
<td>11</td>
<td>9.68 ± 0.23</td>
<td>9.74 ± 0.45</td>
<td>-0.07 (-0.7%)</td>
<td>8.46 ± 0.28</td>
<td>8.54 ± 0.47</td>
<td>-0.08 (-1.0%)</td>
</tr>
<tr>
<td>50 &lt; LD(_{50}) ≤ 300 mg/kg</td>
<td>12</td>
<td>7.76 ± 0.10</td>
<td>8.18 ± 0.21</td>
<td>-0.42 (-5.5%)</td>
<td>6.61 ± 0.19</td>
<td>6.90 ± 0.19</td>
<td>-0.29 (-4.3%)</td>
</tr>
<tr>
<td>300 &lt; LD(_{50}) ≤ 2000 mg/kg</td>
<td>16</td>
<td>8.53 ± 0.21</td>
<td>8.14 ± 0.21</td>
<td>0.38 (4.5%)</td>
<td>7.46 ± 0.24</td>
<td>7.15 ± 0.19</td>
<td>0.31* (4.1%)</td>
</tr>
<tr>
<td>2000 &lt; LD(_{50}) ≤ 5000 mg/kg</td>
<td>10</td>
<td>10.73 ± 0.10</td>
<td>9.46 ± 0.15</td>
<td>1.28* (11.9%)</td>
<td>9.17 ± 0.23</td>
<td>7.96 ± 0.31</td>
<td>1.21* (13.2%)</td>
</tr>
<tr>
<td>LD(_{50}) &gt; 5000 mg/kg</td>
<td>12</td>
<td>9.87 ± 0.34</td>
<td>8.29 ± 0.49</td>
<td>1.58* (16.0%)</td>
<td>7.76 ± 0.59</td>
<td>6.18 ± 0.69</td>
<td>1.58* (20.3%)</td>
</tr>
<tr>
<td>Overall Mean</td>
<td>67</td>
<td>9.35 ± 0.16</td>
<td>8.80 ± 0.17</td>
<td>0.54* (5.8%)</td>
<td>7.95 ± 0.18</td>
<td>7.42 ± 0.20</td>
<td>0.53* (6.6%)</td>
</tr>
<tr>
<td><strong>Dose-mortality Slope = 8.3</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LD(_{50}) ≤ 5 mg/kg</td>
<td>6</td>
<td>11.21 ± 0.24</td>
<td>10.47 ± 0.71</td>
<td>0.75 (6.7%)</td>
<td>9.66 ± 0.27</td>
<td>8.95 ± 0.52</td>
<td>0.71 (7.3%)</td>
</tr>
<tr>
<td>5 &lt; LD(_{50}) ≤ 50 mg/kg</td>
<td>11</td>
<td>9.65 ± 0.16</td>
<td>9.99 ± 0.45</td>
<td>-0.34 (-3.5%)</td>
<td>8.43 ± 0.26</td>
<td>8.77 ± 0.49</td>
<td>-0.33 (-3.9%)</td>
</tr>
<tr>
<td>50 &lt; LD(_{50}) ≤ 300 mg/kg</td>
<td>12</td>
<td>7.78 ± 0.11</td>
<td>8.12 ± 0.21</td>
<td>-0.34 (-4.4%)</td>
<td>6.57 ± 0.19</td>
<td>6.85 ± 0.19</td>
<td>-0.28 (-4.2%)</td>
</tr>
<tr>
<td>300 &lt; LD(_{50}) ≤ 2000 mg/kg</td>
<td>16</td>
<td>8.55 ± 0.22</td>
<td>8.03 ± 0.23</td>
<td>0.52* (6.1%)</td>
<td>7.49 ± 0.25</td>
<td>7.00 ± 0.20</td>
<td>0.49* (6.5%)</td>
</tr>
<tr>
<td>2000 &lt; LD(_{50}) ≤ 5000 mg/kg</td>
<td>10</td>
<td>10.75 ± 0.08</td>
<td>9.54 ± 0.20</td>
<td>1.21* (11.3%)</td>
<td>9.17 ± 0.23</td>
<td>8.06 ± 0.29</td>
<td>1.11* (12.1%)</td>
</tr>
<tr>
<td>LD(_{50}) &gt; 5000 mg/kg</td>
<td>13</td>
<td>9.87 ± 0.32</td>
<td>8.41 ± 0.44</td>
<td>1.47* (14.8%)</td>
<td>7.66 ± 0.59</td>
<td>6.18 ± 0.69</td>
<td>1.47* (19.2%)</td>
</tr>
<tr>
<td>Overall Mean</td>
<td>68</td>
<td>7.95 ± 0.18</td>
<td>7.42 ± 0.20</td>
<td>0.50* (5.3%)</td>
<td>7.92 ± 0.18</td>
<td>7.43 ± 0.20</td>
<td>0.49* (6.2%)</td>
</tr>
</tbody>
</table>

**3T3 NRU Test Method**

**NHK NRU Test Method**

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

*Statistically significant (p <0.05) by a one-sided Wilcoxon signed rank test. Percentage difference shown in parentheses.

1Mean 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 reference LD\(_{50}\) values in mg/kg from BRD Table 4-2 (ICCVAM 2006).

2OECD (2001a); EPA (2002a).

3UN (2005).

4The RC rat-only millimole regression is log LD\(_{50}\) (mmol/kg) = 0.439 log IC\(_{50}\) (mM) + 0.621.

5Default starting dose = 175 mg/kg.

6The 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.

7Difference between mean animal use with the default starting dose and mean animal use with the predicted starting dose.
### Table 2-4 Animal Use for the UDP by GHS Acute Oral Toxicity Category Using Starting Doses Based on the 3T3 and NHK NRU Test Methods with the RC Rat-Only Weight Regression

<table>
<thead>
<tr>
<th>GHS Acute Oral Toxicity Category</th>
<th>Number of Reference Substances</th>
<th>Dose-mortality Slope = 2.0</th>
<th>Dose-mortality Slope = 8.3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>With Default Starting Dose</td>
<td>With IC₅₀-Based Starting Dose</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LD₅₀ ≤ 5 mg/kg</td>
<td>6</td>
<td>11.29 ±0.20</td>
<td>10.38 ±0.62</td>
</tr>
<tr>
<td>5 &lt; LD₅₀ ≤ 50 mg/kg</td>
<td>11</td>
<td>9.71 ±0.22</td>
<td>9.58 ±0.42</td>
</tr>
<tr>
<td>50 &lt; LD₅₀ ≤ 300 mg/kg</td>
<td>12</td>
<td>7.74 ±0.10</td>
<td>7.99 ±0.18</td>
</tr>
<tr>
<td>300 &lt; LD₅₀ ≤ 2000 mg/kg</td>
<td>16</td>
<td>8.52 ±0.21</td>
<td>8.16 ±0.19</td>
</tr>
<tr>
<td>2000 &lt; LD₅₀ ≤ 5000 mg/kg</td>
<td>10</td>
<td>10.78 ±0.11</td>
<td>9.15 ±0.24</td>
</tr>
<tr>
<td>LD₅₀ &gt; 5000 mg/kg</td>
<td>12</td>
<td>9.87 ±0.34</td>
<td>8.23 ±0.48</td>
</tr>
<tr>
<td>Overall Mean</td>
<td>67</td>
<td>9.36 ±0.16</td>
<td>8.70 ±0.16</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LD₅₀ ≤ 5 mg/kg</td>
<td>6</td>
<td>11.21 ±0.24</td>
<td>10.49 ±0.71</td>
</tr>
<tr>
<td>5 &lt; LD₅₀ ≤ 50 mg/kg</td>
<td>11</td>
<td>9.70 ±0.18</td>
<td>9.78 ±0.41</td>
</tr>
<tr>
<td>50 &lt; LD₅₀ ≤ 300 mg/kg</td>
<td>12</td>
<td>7.75 ±0.11</td>
<td>7.99 ±0.21</td>
</tr>
<tr>
<td>300 &lt; LD₅₀ ≤ 2000 mg/kg</td>
<td>16</td>
<td>8.54 ±0.21</td>
<td>8.20 ±0.22</td>
</tr>
<tr>
<td>2000 &lt; LD₅₀ ≤ 5000 mg/kg</td>
<td>10</td>
<td>10.77 ±0.08</td>
<td>9.40 ±0.25</td>
</tr>
<tr>
<td>LD₅₀ &gt; 5000 mg/kg</td>
<td>13</td>
<td>9.88 ±0.32</td>
<td>8.34 ±0.44</td>
</tr>
<tr>
<td>Overall Mean</td>
<td>68</td>
<td>9.36 ±0.16</td>
<td>8.80 ±0.17</td>
</tr>
</tbody>
</table>

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

*Statistically significant (p <0.05) by a one-sided Wilcoxon signed rank test. Percent difference is shown in parentheses.

1Mean 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 reference LD₅₀ values in mg/kg from BRD Table 4-2 (ICCVAM 2006).

2OECD (2001a); EPA (2002a).

3UN (2005).

4The RC rat-only weight regression is log LD₅₀ (mg/kg) = 0.372 log IC₅₀ (µg/mL) + 2.024

5Default starting dose = 175 mg/kg.

6The starting dose was one default dose lower than the predicted LD₅₀ calculated using the IC₅₀ values for each reference substance in the RC rat-only weight regression. The IC₅₀ value for each reference substance was randomly selected from the distribution of values obtained during the testing with each method.

7Difference between mean animal use with the default starting dose and mean animal use with the predicted starting dose.
Computer simulation of ATC method testing showed that, for the substances tested in the validation study, the prediction of starting doses using the NRU test methods resulted in a savings of 0.51 animals (4.8%, slope=8.3 [3T3]) to 0.80 animals (7.3%, slope=2.0 [NHK]) per test when using the RC rat-only millimole regression (Table 2-5). The RC rat-only weight regression produced animal savings of 0.91 animals (8.6%, slope=8.3) to 1.09 animals (10.2%, slope=8.3) (Table 2-6). No animal savings were predicted for substances with $300 < \text{LD}_{50} \leq 2000$ mg/kg when reference substances were grouped by GHS acute oral toxicity category; this category includes the default starting dose of 300 mg/kg. Statistically significant mean animal savings for ATC testing were highest for substances with $5 < \text{LD}_{50} \leq 50$ mg/kg and for substances with $\text{LD}_{50} > 5000$ mg/kg. Mean animal savings using the RC rat-only millimole regression for both test methods for substances with $5 < \text{LD}_{50} \leq 50$ mg/kg ranged from 1.15 animals (9.8%, slope=8.3) to 1.33 animals (11.4%, slope=8.3). Mean animal savings for substances with $\text{LD}_{50} > 5000$ mg/kg ranged from 2.03 animals (17.1%, slope=2) to 2.66 animals (22.2%, slope=8.3). Using the RC rat-only weight regression, mean animal savings for both test methods for substances with $5 < \text{LD}_{50} \leq 50$ mg/kg ranged from 1.25 animals (10.8%, slope=2) to 1.51 animals (13.0%, slope=2.0). Mean animal savings for both test methods for substances with $\text{LD}_{50} > 5000$ mg/kg ranged from 2.94 animals, (24.8%, slope=2.0) to 3.33 animals (27.7%; slope=8.3).

Animal savings did not correlate with the accuracy of the GHS acute oral toxicity category predictions based on the $\text{LD}_{50}$ values calculated using the $\text{IC}_{50}$ values in the RC rat-only regressions. The reason that animal savings is unrelated to the accuracy of prediction of GHS acute oral toxicity category based on the $\text{LD}_{50}$ values calculated using $\text{IC}_{50}$ values in the RC rat-only regressions is because two different standards were used for comparison in the two analyses:

- GHS acute oral toxicity category predictions were compared with the GHS categories derived from the in vivo reference rat oral $\text{LD}_{50}$
- 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 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 use of the $\text{IC}_{50}$-based starting doses did not significantly alter the GHS category outcomes of the simulated UDP (based on $\text{LD}_{50}$ outcome) or ATC when compared with the outcomes based on the default starting dose. The concordance for GHS acute oral toxicity category for the $\text{IC}_{50}$-based starting dose with the default starting dose was 97 to 99% for both in vitro NRU methods and $\text{IC}_{50}$-$\text{LD}_{50}$ regressions evaluated.
Table 2-5  Animal Use\(^1\) 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\(^4\)

<table>
<thead>
<tr>
<th>GHS Acute Oral Toxicity Category(^3)</th>
<th>Number of Reference Substances</th>
<th>Dose-Mortality Slope = 2.0</th>
<th>Dose-Mortality Slope = 8.3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>With Default Starting Dose(^5)</td>
<td>With IC(_{50})-Based Starting Dose(^6)</td>
</tr>
<tr>
<td>LD(_{50}) ≤ 5 mg/kg</td>
<td>6</td>
<td>9.77 ±0.17</td>
<td>7.09 ±1.09</td>
</tr>
<tr>
<td>5 &lt; LD(_{50}) ≤ 50 mg/kg</td>
<td>11</td>
<td>11.56 ±0.21</td>
<td>10.39 ±0.52</td>
</tr>
<tr>
<td>50 &lt; LD(_{50}) ≤ 300 mg/kg</td>
<td>12</td>
<td>10.81 ±0.20</td>
<td>10.39 ±0.17</td>
</tr>
<tr>
<td>300 &lt; LD(_{50}) ≤ 2000 mg/kg</td>
<td>16</td>
<td>9.75 ±0.07</td>
<td>10.67 ±0.48</td>
</tr>
<tr>
<td>2000 &lt; LD(_{50}) ≤ 5000 mg/kg</td>
<td>10</td>
<td>11.22 ±0.08</td>
<td>11.14 ±0.08</td>
</tr>
<tr>
<td>LD(_{50}) &gt; 5000 mg/kg</td>
<td>12</td>
<td>11.85 ±0.04</td>
<td>9.82 ±0.78</td>
</tr>
<tr>
<td>Overall Mean</td>
<td>67</td>
<td>10.89 ±0.12</td>
<td>10.27 ±0.24</td>
</tr>
</tbody>
</table>

| 3T3 NRU Test Method                  |                               |                           |                           |
|--------------------------------------|                               |                           |                           |
| LD\(_{50}\) ≤ 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\(_{50}\) ≤ 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\(_{50}\) ≤ 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\(_{50}\) ≤ 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\(_{50}\) ≤ 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%)      |
| Overall Mean                        | 68                            | 10.91 ±0.11               | 10.11 ±0.24               | 0.80* (7.3%)       | 10.67 ±0.17               | 9.96 ±0.29                | 0.70* (6.6%)       |

| NHK NRU Test Method                  |                               |                           |                           |
|--------------------------------------|                               |                           |                           |
| LD\(_{50}\) ≤ 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\(_{50}\) ≤ 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\(_{50}\) ≤ 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\(_{50}\) ≤ 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\(_{50}\) ≤ 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%)      |
| Overall Mean                        | 68                            | 10.91 ±0.11               | 10.11 ±0.24               | 0.80* (7.3%)       | 10.67 ±0.17               | 9.96 ±0.29                | 0.70* (6.6%)       |

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

*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 reference LD\(_{50}\) values in mg/kg from BRD Table 4-2 (ICCVAM 2006). 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.
**Table 2-6** Animal Use\(^1\) 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\(^4\)

<table>
<thead>
<tr>
<th>GHS Acute Oral Toxicity Category(^3)</th>
<th>Number of Reference Substances</th>
<th>Dose-Mortality Slope = 2.0</th>
<th></th>
<th>Dose-Mortality Slope = 8.3</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>With Default Starting Dose(^5)</td>
<td>With IC(_{50}) Based Starting Dose(^6)</td>
<td>Animals Saved(^7)</td>
<td>With Default Starting Dose(^5)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3T3 NRU Test Method</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LD(_{50}) ≤5 mg/kg</td>
<td>6</td>
<td>9.77 ±0.17</td>
<td>7.56 ±1.03</td>
<td>2.21 (22.6%)</td>
<td>9.08 ±0.08</td>
</tr>
<tr>
<td>5 &lt; LD(_{50}) ≤50 mg/kg</td>
<td>11</td>
<td>11.56 ±0.21</td>
<td>10.06 ±0.38</td>
<td>1.51* (13.0%)</td>
<td>11.75 ±0.16</td>
</tr>
<tr>
<td>50 &lt; LD(_{50}) ≤300 mg/kg</td>
<td>12</td>
<td>10.81 ±0.20</td>
<td>10.35 ±0.18</td>
<td>0.47* (4.3%)</td>
<td>9.42 ±0.26</td>
</tr>
<tr>
<td>300 &lt; LD(_{50}) ≤2000 mg/kg</td>
<td>16</td>
<td>9.75 ±0.07</td>
<td>10.67 ±0.50</td>
<td>-0.93* (-9.5%)</td>
<td>9.26 ±0.10</td>
</tr>
<tr>
<td>2000 &lt; LD(_{50}) ≤5000 mg/kg</td>
<td>10</td>
<td>11.22 ±0.08</td>
<td>9.80 ±0.51</td>
<td>1.43* (12.7%)</td>
<td>11.88 ±0.10</td>
</tr>
<tr>
<td>LD(_{50}) &gt;5000 mg/kg</td>
<td>12</td>
<td>11.85 ±0.04</td>
<td>8.83 ±0.83</td>
<td>3.02* (25.5%)</td>
<td>12.00 ±0.00</td>
</tr>
<tr>
<td>Overall</td>
<td>67</td>
<td>10.89 ±0.12</td>
<td>9.85 ±0.24</td>
<td>1.04* (9.6%)</td>
<td>10.64 ±0.17</td>
</tr>
<tr>
<td>NHK NRU Test Method</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LD(_{50}) ≤5 mg/kg</td>
<td>6</td>
<td>9.74 ±0.16</td>
<td>6.87 ±1.28</td>
<td>2.87 (29.4%)</td>
<td>9.09 ±0.08</td>
</tr>
<tr>
<td>5 &lt; LD(_{50}) ≤50 mg/kg</td>
<td>11</td>
<td>11.56 ±0.21</td>
<td>10.31 ±0.19</td>
<td>1.25* (10.8%)</td>
<td>11.76 ±0.17</td>
</tr>
<tr>
<td>50 &lt; LD(_{50}) ≤300 mg/kg</td>
<td>12</td>
<td>10.83 ±0.21</td>
<td>10.41 ±0.28</td>
<td>0.42 (3.8%)</td>
<td>9.44 ±0.26</td>
</tr>
<tr>
<td>300 &lt; LD(_{50}) ≤2000 mg/kg</td>
<td>16</td>
<td>9.77 ±0.62</td>
<td>10.46 ±0.50</td>
<td>-0.69 (-7.1%)</td>
<td>9.26 ±0.10</td>
</tr>
<tr>
<td>2000 &lt; LD(_{50}) ≤5000 mg/kg</td>
<td>10</td>
<td>11.22 ±0.09</td>
<td>10.69 ±0.37</td>
<td>0.53 (4.7%)</td>
<td>11.87 ±0.10</td>
</tr>
<tr>
<td>LD(_{50}) &gt;5000 mg/kg</td>
<td>13</td>
<td>11.86 ±0.03</td>
<td>8.91 ±0.78</td>
<td>2.94* (24.8%)</td>
<td>12.00 ±0.00</td>
</tr>
<tr>
<td>Overall Mean</td>
<td>68</td>
<td>10.91 ±0.11</td>
<td>9.95 ±0.24</td>
<td>0.96* (8.8%)</td>
<td>10.67 ±0.17</td>
</tr>
</tbody>
</table>

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

*Statistically significant (p <0.05) by a one-sided Wilcoxon signed rank test. Percentage difference is shown in parentheses.

1Mean 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 reference LD\(_{50}\) values in mg/kg from BRD Table 4-2 (ICCVAM 2006).

2OECD (2001d).

3GHS for acute oral toxicity (UN 2005).

4\( \log LD_{50} \) (mg/kg) = 0.372 \( \log IC_{50} \) (µg/mL) + 2.024

5Default starting dose = 300 mg/kg.

6The 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.

7Difference between mean animal use with the default starting dose and mean animal use with the IC\(_{50}\)-based starting dose.
2.6 ICCVAM Recommendations for Test Method Uses

ICCVAM’s recommendations for use of these test methods are as follows:

1. The 3T3 and NHK NRU test methods are not sufficiently accurate to predict acute oral toxicity for the purpose of regulatory hazard classification (see Section 2.3 above and Section 6 of the BRD [ICCVAM 2006]).

2. For the purposes of acute oral toxicity testing, the 3T3 and NHK NRU test methods may be used in a weight-of-evidence approach to determine the starting dose for the current acute oral toxicity protocols (i.e., the UDP, the ATC method).

3. Consistent with the U.S. Government Principles on the Use of Animals in Research, Testing, and Education\(^{26}\), and the U.S. Public Health Service Policy on Humane Care and Use of Laboratory Animals (PHS 2002), \textit{in vitro} basal cytotoxicity test methods as part of a weight-of-evidence approach to estimate the starting dose for acute oral \textit{in vivo} toxicity test methods should be considered and used where appropriate before testing is conducted using animals. For some types of substances, this approach will reduce the number of animals needed. In some testing situations, the approach may also reduce the numbers of animals that die or need to be humanely killed.

4. The starting doses for substances with certain toxic mechanisms that are not expected to be active in 3T3 or NHK cells (e.g., those that are neurotoxic or cardiotoxic) will likely be underpredicted by these \textit{in vitro} basal cytotoxicity test methods. Therefore, the results from basal cytotoxicity testing with such substances may not be appropriate for estimating starting doses.

5. The regression formula used to determine starting doses for test substances with known molecular weights and high purity should be the revised RC millimole regression line, based on substances with rat LD\(_{50}\) data, with IC\(_{50}\) values in mmol/L and LD\(_{50}\) values in mmol/kg. The regression formula used to determine starting doses for mixtures, test substances with low or unknown purity, or test substances with unknown molecular weights should be the revised RC regression line, based on substances with rat LD\(_{50}\) data, with IC\(_{50}\) values in \(\mu g/mL\) and LD\(_{50}\) values in mg/kg.

6. The performance of other \textit{in vitro} basal cytotoxicity test methods that are based on similar scientific principles and that measure or predict the same biological response (i.e., basal cytotoxicity and the rat acute oral LD\(_{50}\) value, respectively) should be demonstrated to meet or exceed the accuracy and reliability of the 3T3 and NHK NRU test methods (see Section 3.0 for ICCVAM Recommended Performance Standards).

7. Compared to the NHK NRU test method, the 3T3 NRU test method appears to be less labor intensive and less expensive to conduct; therefore, the 3T3 NRU test method is recommended for general use. Although the 3T3 NRU test method was less reproducible than the NHK NRU test method, it produced slightly higher animal savings and accuracy for prediction of GHS acute oral

toxicity category using the IC$_{50}$ and the revised RC regressions evaluated for the prediction of LD$_{50}$.

### 2.7 ICCVAM Recommendations for Future Studies

ICCVAM recommends the following future studies in order to advance the use of *in vitro* methods for assessing acute oral toxicity for regulatory hazard classification purposes:

1. Additional data should be collected using the 3T3 NRU basal cytotoxicity test method to evaluate its usefulness for predicting the rodent acute oral toxicity of chemical mixtures.

2. To supplement the high quality validation database started by this study, additional high quality comparative *in vitro* basal cytotoxicity data should be collected when rat acute oral toxicity testing is conducted. However, *in vivo* testing should not be conducted solely to collect data to assess the usefulness of the NRU test method. Periodic evaluations of the expanded database should be conducted to further characterize the usefulness and limitations of using *in vitro* cytotoxicity data as part of a weight-of-evidence approach to estimate starting doses.

3. Additional efforts should be conducted to identify *in vitro* tests and other methods necessary to achieve accurate acute oral hazard classification; studies should be conducted to investigate the potential use of *in vitro* cell-based test methods that incorporate mechanisms of action and evaluations of ADME (absorption, distribution, metabolism, excretion) to provide improved estimates of acute toxicity hazard categories. Methods should be developed to extrapolate from *in vitro* toxic concentrations to equivalent doses *in vivo*.

4. The *in vivo* database of reference substances used in this validation study should be used to evaluate the utility of other non-animal approaches to estimate starting doses for acute oral toxicity tests (e.g., widely available software that uses quantitative structure-activity relationships [QSAR]).

5. Standardized procedures to collect *in vivo* measurements and observations pertinent to an understanding of the mechanisms of lethality should be included in future rat acute oral toxicity studies. Such information will likely be necessary to support the further development of predictive mechanism-based *in vitro* methods.

6. An expanded list of reference substances with rat acute oral LD$_{50}$ values substantiated by high quality *in vivo* data (including data currently held by industry) should be developed for use in future *in vitro* test method development and validation studies.
3.0 ICCVAM RECOMMENDED PERFORMANCE STANDARDS

The purpose of performance standards is to communicate the basis by which validated new proprietary (e.g., copyrighted, trademarked, registered) and nonproprietary test methods have been determined to have sufficient accuracy and reliability for specific testing purposes. Performance standards can then be used to evaluate the accuracy and reliability of other test methods that are based on similar scientific principles and that measure or predict the same biological or toxic effect. The three elements of performance standards are essential test method components (see Section 3.1), a minimum list of reference substances for assessing the accuracy and reliability of the proposed test method (see Section 3.2), and the accuracy and reliability values that should be achieved by the proposed test method using the minimum list of reference substances (see Section 3.3).

The 3T3 and NHK NRU test methods are not sufficiently accurate to predict the acute oral toxicity of substances for the purposes of regulatory hazard classification and labeling. However, these test methods may be used in a weight-of-evidence approach to determine the starting dose for the UDP (OECD 2001a; EPA 2002a) and the ATC (OECD 2001b) rodent acute oral toxicity test methods. The performance of other in vitro basal cytotoxicity test methods that are based on similar scientific principles and that measure or predict the same biological response (i.e., basal cytotoxicity and the rat acute oral LD<sub>50</sub>, respectively) should meet or exceed the accuracy and reliability of the 3T3 and NHK NRU test methods.

The extent to which proposed in vitro basal cytotoxicity test methods should demonstrate comparable performance to these two in vitro NRU cytotoxicity test methods should be considered on a case-by-case basis.

3.1 Essential Test Method Components for In Vitro Basal Cytotoxicity Assays to Predict Starting Doses for Acute Oral Toxicity Tests

These test method components consist of essential structural, functional, and procedural elements of a validated test method that should be included in the protocol of a proposed, mechanistically and functionally similar test method. Essential test method components include unique characteristics of the test method, critical procedural details, and quality control measures. Adherence to essential test method components will help to assure that a proposed test method is structurally and functionally similar to the corresponding validated test method.

The basic steps of an in vitro basal cytotoxicity assay are as follows:

- The test substance is dissolved in an appropriate solvent and applied as a solution to cells that, under control conditions, would be expected to be growing exponentially throughout the exposure period.
- The test substance is incubated with the cells for a specified period of time.
- The test substance is removed and an endpoint indicative of cell viability or cytotoxicity is measured.
- The IC<sub>50</sub> value is calculated (i.e., the concentration at which cell viability or growth is inhibited by 50% compared to control values).
Many different \textit{in vitro} basal cytotoxicity methods might be used to estimate rat acute oral LD$_{50}$ values and, thus, to predict the starting dose for a rodent acute oral lethality assay. \textit{In vitro} basal cytotoxicity data determined using various primary cells and permanent non-differentiated finite or transformed cell lines, generally exhibits the same concentration-response cytotoxicity relationship when exposed to the same xenobiotic, regardless of the toxic endpoints investigated. The following endpoints are sufficiently characteristic of basal cytotoxicity (Spielmann et al. 1999; Halle 1998, 2003):

\begin{itemize}
  \item \textit{Inhibition of cell proliferation}: cell number, cell protein, deoxyribonucleic acid (DNA) content, DNA synthesis, colony formation
  \item \textit{Cell viability - metabolic markers}: metabolic inhibition test, mitochondrial reduction of tetrazolium salts into soluble dye
  \item \textit{Decreased cell viability - membrane markers}: NRU into cell lysosomes, Trypan Blue exclusion, cell attachment/cell detachment for monolayer cultures
  \item \textit{Differentiation markers}: functional or morphological differentiation within cell clusters, intracellular morphology
\end{itemize}

Markers of the release of intracellular components, such as the enzyme lactate dehydrogenase (i.e., LDH release test) or of dye introduced into the cells previous to chemical exposure as occurs, for example, in the fluorescein leakage (FL) test or the Neutral Red Release (NRR) test, are not considered to be characteristic for basal cytotoxicity because they specifically detect damage of the outer cell membrane and generally are associated with short-term chemical exposure (ICCVAM 2001b). A chemical that specifically damages only cell membranes, however, will be detected correctly in one of the tests for basal cytotoxicity listed above.

Investigators using an \textit{in vitro} basal cytotoxicity system for prediction of the \textit{in vivo} starting dose for acute oral toxicity studies must be able to demonstrate that the assay is valid for its intended use. This includes demonstrating that any modification to the existing validated reference test method does not adversely affect its performance characteristics. \textit{In vitro} systems may be used to test solids, liquids, and emulsions of any chemical or product class. The liquids can be aqueous or nonaqueous; solids can be soluble or insoluble in water. The samples may be pure chemicals, dilutions, formulations, or waste. Test substances must be soluble in cell culture medium, dimethyl sulfoxide (DMSO), or ethanol (ETOH). The test method endpoint (i.e., percent of control values) is used to generate an IC$_{50}$ value in mM (if the substance’s molecular weight is known, and, if not, in $\mu$g/mL) and the IC$_{50}$ value is used in the regressions developed to estimate the LD$_{50}$ value in mmol/kg (or mg/kg).

The following is a description of the essential test method components for \textit{in vitro} basal cytotoxicity assays to predict starting doses for acute oral toxicity/lethality tests.

3.1.1 \textit{In Vitro} Cell Culture Conditions

\begin{itemize}
  \item A mammalian cell line (or primary cells) is used that divides rapidly with doubling times of less than 30 hours under standard culture conditions, preferably with calf serum (CS), NCS, or serum-free medium (ICCVAM 2001b).
\end{itemize}
Cells are allowed to propagate in sterile tissue culture vessels (e.g., flasks) and then are subcultured to other sterile tissue culture vessels (e.g., 96 well-plates) for use in testing. Initial cell seeding should be done at a density that allows for exponential growth throughout the exposure period.

Appropriate cell culture growth conditions are maintained throughout the testing period (e.g., 37 °C ±1 °C, 90% ±10% humidity, 5.0% ±1% CO2/air). The cell cultures should be free of contamination with bacteria, mycoplasma, or fungi.

Cell culture media should be prequalified by the testing laboratory via a standardized protocol before initiating the test to guarantee that the media provide cells with appropriate nutrients to meet the growth criteria required for the test method.

### 3.1.2 Application of the Test Substances

#### Test Substance Preparation

- Test substance solutions should be prepared in cell culture medium within an hour before application to the cell cultures (unless the stability of the test substance in the solvent used requires shorter times or allows longer times).
- Standard protocol methods for solubility procedures can include mixing the test substance by vortexing, sonication, warming, and stirring. Test substances should be fully solubilized (i.e., no visual observation of test substance in the dosing solution) before application.
- An inherent limitation to in vitro cytotoxicity is the testing of volatile substances since the material may evaporate before application to the cells or may not remain in the test vessel when incubated. If volatility is predicted or identified for a test substance (e.g., by detection of cross-contamination of the high concentrations of test substance in culture with lower concentrations or controls in the test vessel), measures can be employed to test moderately volatile substances (e.g., cover the test plate with a CO2 permeable plastic film cover/sealer).

#### Cytotoxicity Test

- Each cytotoxicity test should contain a range of test substance concentrations such that the IC50 value can be determined with at least one cytotoxic point between 0 – 50% viability and at least one cytotoxic point between 50 – 100% viability.
- A minimum of three adequate data points should be collected for each test substance concentration. (Note: The NICEATM/ECVAM validation study required the testing of six replicates for each test substance concentration with at least four successful replicates.)
- Blanks (i.e., culture vessels without cells) should be available for assessing background interference when measuring the endpoint.
- Cell monolayers in tissue culture vessels should be adequately covered (e.g., a minimum of 100 μL of test substance solution per well in a 96-well test plate).
- The substance exposure period should be at least the duration of one cell cycle (i.e., approximately 24 to 72 hours) (Riddell et al. 1986). [Note: The
NICEATM/ECVAM validation study required an exposure period of 48 hours for 3T3 and NHK cells; the cell cycle duration (i.e., doubling time) for these cells ranged from 17 to 19 (3T3) and 10 to 22 (NHK) hours in log phase.

- At the end of the exposure period, most endpoints require washing the test substance from the cells with an appropriate buffering solution (e.g., Dulbecco’s Phosphate Buffered Saline [DPBS]) before applying the endpoint material (e.g., neutral red dye). Washing cells to remove the test substance is the default recommendation unless it is known that washing would interfere with measurement of the endpoint.

3.1.3 Control Substances

Vehicle Controls (VC): The VCs provide the reference for 100% cell growth in the test vessel and, thus, the vehicle (or solvent) must be compatible with the cell culture system (i.e., not cause cytotoxicity or reduce cell growth through other mechanisms) and should not alter the properties of the test substance. The VCs should contain the solvent at the concentration applied to the cells. For example, DMSO and ETOH at a final concentration \( \leq 0.5\% \) [v/v] were demonstrated to be compatible with cell growth for 3T3 and NHK cells in the NICEATM/ECVAM validation study. If the compatibility of the solvent with the cell culture system is unknown, cultures with and without the solvent should be included in each experiment.

Positive Controls (PC): The purpose of a PC substance is to demonstrate that the cell culture system is responding with adequate sensitivity to a cytotoxic agent for which the magnitude of the cytotoxic response is well characterized. The PC substance should be tested concurrently with (and independent of) the test substance. The PC should be well characterized for its cytotoxicity potential and each test should generate a response that is comparable to the historic IC\(_{50}\) range generated by the laboratory. A laboratory should perform a minimum of 10 cytotoxicity tests using the PC over a number of days to develop a minimum historical database of IC\(_{50}\) data. Typically, for biologically based test methods, suggested acceptable ranges for the PC response are within two to three standard deviations of the historical mean response, but developers of proprietary test methods may establish tighter ranges. Sodium lauryl sulfate (SLS) is an effective PC substance for use in in vitro basal cytotoxicity test methods. [Note: The NICEATM/ECVAM validation study used SLS as the PC and required 2.5 standard deviations of the historical mean response as the acceptable range.]

Benchmark Controls: Benchmark controls may be useful to demonstrate that the test method is functioning properly for detecting the cytotoxic potential of substances of a specific chemical class or a specific range of responses, or for evaluating the relative cytotoxic potential of a cytotoxic test substance. Appropriate benchmark controls should have the following properties:

- Consistent and reliable source(s) for the substance
- Structural and functional similarity to the class of the substance being tested
- Known physical/chemical characteristics
- Supporting data on known effects in animal models
- Known potency in the range of response (including moderate response)
3.1.4 Viability Measurements

- Only standardized, quantitative methods should be used to measure cell viability. The protocol should be compatible with laboratory apparatus such as spectrophotometers that allow a quick and precise measurement of the endpoint.
- Non-specific dye binding must not interfere with the viability measurement. A measurement endpoint that is well established and that has good interlaboratory reproducibility should be used (ICCVAM 2001b).
- A detailed concentration-response experiment should be conducted using a progression factor that yields graded effects between no effect and total cytotoxicity. Any desired toxicity measure can be derived from a well-designed concentration-response experiment.
- Preference should be given to endpoints that determine either cell proliferation or cell viability (e.g., NRU, MTT [3-(4,5,diimethylthiazol-2yl)2,5-diphenyl tetrazolium bromide], XTT [Sodium 3,3,-(phenylamino)carbonyl]-3,4-tetrazolium-bis(4-methoxy-6-nitro)benzenesulfonic acid hydrate]) (ICCVAM 2001b).
- Simple endpoints such as total protein content are not recommended, as they may underpredict the toxicity of certain test substances by including protein from dead cells.
- A lack of information and a low level of accuracy characterize experiments that seek only to identify the highest tolerated dose or the lowest cytotoxic dose.

Colorimetric endpoints (e.g., NRU) should have the optical density (OD) spectrascopically-measured at the appropriate wavelength (e.g., 540 nm ±10 nm for NRU) and OD values for blanks should be subtracted from the vehicle control and test substance ODs.

3.1.5 Interpretation of Results

IC50 Determination: The endpoint values obtained at each concentration of the test substance can be used to calculate the percentage of cell viability or growth relative to the negative (vehicle) control, which is arbitrarily set at 100%. The cell viability criteria used to determine an IC50 value must be clearly defined and documented, and be shown to be appropriate. In general, such criteria are established during test optimization, tested during a prevalidation phase, and confirmed in a validation study.

Regression Formula: The recommended regression formulas to predict LD50 values from IC50 values are

- The RC rat-only millimole regression for substances with known molecular weight: \( \log LD_{50} (\text{mmol/kg}) = 0.439 \log IC_{50} (\text{mM}) + 0.621 \)
- The RC rat-only weight regression for mixtures and substances with no known molecular weight: \( \log LD_{50} (\text{mg/kg}) = 0.372 \log IC_{50} (\mu g/mL) + 2.024 \)
3.1.6 Test Report
The test report should include the following information, if relevant to the conduct of the study:

Test Substances and Control Substances
- Chemical name(s) such as Chemical Abstracts Service Registry Number (CASRN) and molecular weight (if known), followed by other names, if known
- Formulation (if available) of the test substance if the material is a mixture
- Purity and composition of the substance or preparation (in percentage[s] by weight)
- Physicochemical properties such as physical state, volatility, pH, stability, chemical class, water solubility relevant to the conduct of the study
- Treatment of the test/control substances prior to testing, if applicable (e.g., vortexing, sonication, warming; solvent used)
- Stability, if known

Justification of the In Vitro Test Method and Protocol Used

Test Method Integrity
- The procedure used to insure the integrity (i.e., accuracy and reliability) of the test method over time
- If the test method employs proprietary components, documentation on the procedure used to ensure their integrity from “lot-to-lot” and over time
- The procedures that the user may employ to verify the integrity of the proprietary components

Criteria for an Acceptable Test
- Acceptable concurrent PC ranges based on historical data
- Acceptable negative and solvent/VC data

Test Conditions
- Cell system used
- Calibration information for measuring device used for measuring cell viability (e.g., spectrophotometer)
- Details of test procedure used
- Test doses used
- Description of any modifications of the test procedure
- Reference to historical data of the model
- Description of evaluation criteria used

Results
- Tabulation of data from individual test samples (e.g., OD values and calculated percentage cell viability data for the test substance and the PC and negative and benchmark controls, reported in tabular form, including data from replicate repeat experiments as appropriate, and means ± the standard deviation for each trial)
- Calculated IC$_{50}$ value
- Calculated starting dose (i.e., LD$_{50}$ value) using IC$_{50}$ value in regression formula
- Regression formula (prediction model) used
Description of Other Effects Observed
Discussion of the Results
Conclusion

3.2 Reference Substances for In Vitro Basal Cytotoxicity Assays to Predict Starting Doses for Acute Oral Toxicity Tests

Reference substances are used to assess the accuracy and reliability of a proposed, mechanistically and functionally similar test method and are a representative subset of those used to demonstrate the reliability and the accuracy of the validated test method. These substances:

- Are representative of the range of responses that the validated test method is capable of measuring or predicting
- Have produced consistent results in the validated test method
- Will reflect the accuracy of the validated test method
- Have well-defined chemical structures
- Are readily available
- Are not associated with excessive hazard or prohibitive disposal costs

The subset of 30 reference substances in Table 3-1 was chosen from the 72 reference substances used in the NICEATM/ECVAM validation study. Reference substances that exhibited solubility difficulties or were volatile in culture during this study are included as a secondary subset and are recommended for investigational purposes only.

The substances in this list represent the following types of chemical classes: acyclic hydrocarbons, alcohols, amides, amines, arsenical compounds, boron compounds, cadmium compounds, carboxylic acids, chlorine compounds, cyclic hydrocarbons, fluorine compounds, heterocyclics, mercury compounds, nitro compounds, organometallics, phenols, organophosphorous compounds, polycyclics, potassium compounds, sodium compounds, sulfur compounds, and ureas.
Table 3-1  Recommended Reference Substances for Evaluation of In Vitro Basal Cytotoxicity Methods for Predicting the Starting Dose for Rodent Acute Oral Toxicity Tests

<table>
<thead>
<tr>
<th>Reference Substance</th>
<th>CASRN</th>
<th>Rodent Oral LD₅₀¹</th>
<th>3T3 IC₅₀²</th>
<th>NHK IC₅₀²</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>mg/kg</td>
<td>mmole/kg</td>
<td>µg/mL</td>
</tr>
<tr>
<td>LD₅₀ ≤5 mg/kg</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mercury II chloride</td>
<td>7487-94-7</td>
<td>1</td>
<td>0.0037</td>
<td>4.122</td>
</tr>
<tr>
<td>Triethylenemelamine</td>
<td>51-18-3</td>
<td>1</td>
<td>0.0049</td>
<td>0.2722</td>
</tr>
<tr>
<td>Cycloheximide</td>
<td>66-81-9</td>
<td>2</td>
<td>0.0071</td>
<td>0.1874</td>
</tr>
<tr>
<td>Busulfan</td>
<td>55-98-1</td>
<td>2</td>
<td>0.0081</td>
<td>77.68</td>
</tr>
<tr>
<td>Phenylthiourea</td>
<td>103-85-5</td>
<td>3</td>
<td>0.0197</td>
<td>78.98</td>
</tr>
<tr>
<td>5 &lt; LD₅₀ ≤50 mg/kg</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dichlorvos</td>
<td>62-73-7</td>
<td>17</td>
<td>0.0769</td>
<td>17.74</td>
</tr>
<tr>
<td>Digoxin</td>
<td>20830-75-5</td>
<td>18</td>
<td>0.0230</td>
<td>445.5</td>
</tr>
<tr>
<td>Sodium arsenite</td>
<td>7784-46-5</td>
<td>41</td>
<td>0.3156</td>
<td>0.7587</td>
</tr>
<tr>
<td>Triphenyltin hydroxide</td>
<td>76-87-9</td>
<td>44</td>
<td>0.1199</td>
<td>0.0172</td>
</tr>
<tr>
<td>Sodium dichromate dihydrate</td>
<td>7789-12-0</td>
<td>50</td>
<td>0.1908</td>
<td>0.5867</td>
</tr>
<tr>
<td>50 &lt; LD₅₀ ≤300 mg/kg</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hexachlorophene</td>
<td>70-30-4</td>
<td>61</td>
<td>0.1499</td>
<td>4.195</td>
</tr>
<tr>
<td>Cadmium II chloride</td>
<td>10108-64-2</td>
<td>88</td>
<td>0.4801</td>
<td>0.5177</td>
</tr>
<tr>
<td>Sodium oxalate</td>
<td>62-76-0</td>
<td>155</td>
<td>1.160</td>
<td>37.14</td>
</tr>
<tr>
<td>Sodium fluoride</td>
<td>7681-49-4</td>
<td>180</td>
<td>4.290</td>
<td>78.02</td>
</tr>
<tr>
<td>Diquat dibromide monohydrate</td>
<td>6385-62-2</td>
<td>231</td>
<td>0.6714</td>
<td>8.040</td>
</tr>
<tr>
<td>300 &lt; LD₅₀ ≤2000 mg/kg</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Amitriptyline HCl</td>
<td>549-18-8</td>
<td>361</td>
<td>1.150</td>
<td>7.054</td>
</tr>
<tr>
<td>Propranolol HCl</td>
<td>3506-09-0</td>
<td>470</td>
<td>1.589</td>
<td>14.11</td>
</tr>
<tr>
<td>Atropine sulfate monohydrate</td>
<td>5908-99-6</td>
<td>639</td>
<td>0.9204</td>
<td>76.03</td>
</tr>
<tr>
<td>Acetylsalicylic acid</td>
<td>50-78-2</td>
<td>1000</td>
<td>5.549</td>
<td>676.4</td>
</tr>
<tr>
<td>Carbamazepine</td>
<td>298-46-4</td>
<td>1957</td>
<td>8.282</td>
<td>103.2</td>
</tr>
<tr>
<td>Reference Substance</td>
<td>CASRN</td>
<td>Rodent Oral LD_{50}^1</td>
<td>3T3 IC_{50}^2</td>
<td>NHK IC_{50}^2</td>
</tr>
<tr>
<td>-----------------------------</td>
<td>---------</td>
<td>-----------------------</td>
<td>--------------</td>
<td>--------------</td>
</tr>
<tr>
<td></td>
<td></td>
<td>mg/kg</td>
<td>mmole/kg</td>
<td>µg/mL</td>
</tr>
<tr>
<td></td>
<td>2000 &lt; LD_{50} ≤ 5000 mg/kg</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Acetaminophen</td>
<td>103-90-2</td>
<td>2404</td>
<td>15.90</td>
<td>47.66</td>
</tr>
<tr>
<td>Potassium chloride</td>
<td>7447-40-7</td>
<td>2602</td>
<td>34.90</td>
<td>3555</td>
</tr>
<tr>
<td>Chloramphenicol</td>
<td>56-75-7</td>
<td>3393</td>
<td>10.50</td>
<td>130.2</td>
</tr>
<tr>
<td>Lactic acid</td>
<td>50-21-5</td>
<td>3730</td>
<td>41.41</td>
<td>3044</td>
</tr>
<tr>
<td>Trichloroacetic acid</td>
<td>76-03-9</td>
<td>4999</td>
<td>30.59</td>
<td>901.8</td>
</tr>
<tr>
<td></td>
<td>LD_{50} &gt; 5000 mg/kg</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ethylene glycol</td>
<td>107-21-1</td>
<td>8567</td>
<td>138.0</td>
<td>24435</td>
</tr>
<tr>
<td>Gibberellic acid</td>
<td>77-06-5</td>
<td>6305</td>
<td>18.20</td>
<td>7810</td>
</tr>
<tr>
<td>Sodium hypochlorite</td>
<td>7681-52-9</td>
<td>10328³</td>
<td>138.7³</td>
<td>1040</td>
</tr>
<tr>
<td>Dibutyl phthalate</td>
<td>84-74-2</td>
<td>11998</td>
<td>43.11</td>
<td>43.37</td>
</tr>
<tr>
<td>Glycerol</td>
<td>56-81-5</td>
<td>12691</td>
<td>137.8</td>
<td>24345</td>
</tr>
</tbody>
</table>

**Secondary Subset**

<table>
<thead>
<tr>
<th>Precipitating Substances⁴</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>LD_{50} ≤ 5 mg/kg</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arsenic trioxide</td>
<td>1327-53-3</td>
<td>20</td>
<td>0.1000</td>
<td>2.072</td>
<td>0.0105</td>
<td>6.840</td>
</tr>
<tr>
<td>Parathion</td>
<td>56-38-2</td>
<td>2</td>
<td>0.0069</td>
<td>37.42</td>
<td>0.1285</td>
<td>30.26</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Volatile Substances⁵</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>300 &lt; LD_{50} ≤ 2000 mg/kg</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phenol</td>
<td>108-95-2</td>
<td>414</td>
<td>4.400</td>
<td>66.32</td>
<td>0.7047</td>
<td>75.03</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>LD_{50} &gt; 5000 mg/kg</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Ethanol</td>
<td>64-17-5</td>
<td>14008</td>
<td>304.15</td>
<td>6523</td>
<td>141.6</td>
<td>10018</td>
</tr>
<tr>
<td>2-Propanol</td>
<td>67-63-0</td>
<td>5843</td>
<td>97.21</td>
<td>3489</td>
<td>58.04</td>
<td>5364</td>
</tr>
</tbody>
</table>

Abbreviations: CASRN=Chemical Abstracts Service Registry Number; 3T3=Neutral red uptake assay using BALB/c 3T3 fibroblasts; NHK=Neutral red uptake assay using normal human epidermal keratinocytes.

¹The dose that produces lethality in 50% of test animals (rats or mice). Values used in the RC (Halle 1998, 2003) unless otherwise noted.
²Reference substance concentration (geometric mean of laboratory means) producing 50% inhibition of the endpoint measured (i.e., cell viability).
³LD_{50} values were calculated as the geometric mean of values obtained in the literature (see BRD Section 4) (ICCVAM 2006).
⁴Reference substances expected to precipitate at cytotoxic concentrations.
⁵Reference substances expected to contaminate neighboring wells at high concentrations.
3.3 Accuracy and Reliability Standards

The third element of the performance standards is the determination of accuracy (also known as relevance) and reliability values.

3.3.1 Accuracy and Reliability for the NRU Test Methods

To demonstrate technical proficiency with the validated 3T3 or NHK NRU test method, ICCVAM recommends that the user evaluate his/her ability to calculate IC\textsubscript{50} values for a minimum of two unclassified substances and two substances from each of the five GHS hazard categories (i.e., at least 12 of the 30 reference substances) listed in Table 3-1. The resulting IC\textsubscript{50} values should be within 2.5 standard deviations of the IC\textsubscript{50} values reported in the table.\textsuperscript{27} A linear regression calculated using the LD\textsubscript{50} values provided in Table 3-1 and the resulting IC\textsubscript{50} values should not differ from a linear regression calculated using the LD\textsubscript{50} and the IC\textsubscript{50} values provided in Table 3-1. Also, the intralaboratory CV values for the IC\textsubscript{50} of the reference substances selected should not exceed 129% for the NHK NRU test method or 98% for the 3T3 NRU test method and the mean CV should not exceed 27% for either test method.

3.3.2 Accuracy and Reliability for Me-Too Assays

A proposed test method that is functionally and mechanistically similar to the 3T3 NRU test method should use the selected reference substances to assess accuracy and reliability. The ICCVAM Recommendations (see Section 2.6) propose the general use of the 3T3 NRU test method because it appears to be less labor intensive and less expensive to conduct compared to the NHK NRU test method. Thus, the accuracy and reliability standards presented below focus on the 3T3 NRU test method.

Before using a candidate \textit{in vitro} basal cytotoxicity test to predict starting doses, the correlation between the \textit{in vitro} and the \textit{in vivo} test methods must be established quantitatively by using the new test method to test 12 of the 30 reference substances. After testing, the IC\textsubscript{50} data are used to calculate a linear regression formula (least square method) for the selected reference substances using the corresponding LD\textsubscript{50} values provided in Table 3-1. The resulting regression is compared against a regression using the 3T3 NRU IC\textsubscript{50} and the LD\textsubscript{50} values provided in this table. If the regressions are not statistically significantly different based on a comparison of slope and intercept (at p <0.05), then the test is considered suitable to generate IC\textsubscript{50} data to use with the recommended regression formula for estimating starting doses for acute oral toxicity/lethality tests.

The overall accuracy of the 3T3 NRU test method for correctly predicting GHS acute oral toxicity classification of the 30 reference substances using the RC rat-only millimole regression was 33%. \textit{In vivo} toxicity was overpredicted for 33% and underpredicted for 34%. Seventy-seven percent of the reference substances were classified in the correct category, or within one category above or below the correct category (see Table 3-2). For this analysis, in terms of each GHS acute oral toxicity category:

\textsuperscript{27} Replicate IC\textsubscript{50} values must be determined for each reference substance in order to calculate the standard deviation.
• Zero (0%) of 5 substances with LD$_{50}$ <5 mg/kg was correctly predicted
• One (20%) of 5 substances in the 5 < LD$_{50}$ ≤50 mg/kg category was correctly predicted
• Four (80%) of 5 substances in the 50 < LD$_{50}$ ≤300 mg/kg category were correctly predicted
• Four (80%) of 5 substances in the 300 < LD$_{50}$ ≤2000 mg/kg category were correctly predicted; however, this toxicity category was also predicted for 11 other substances that did not match this category in vivo. Thus, the predictivity for this category was 27%.
• Zero (0%) of the 5 substances in the 2000 < LD$_{50}$ ≤5000 mg/kg category were correctly predicted
• One (20%) of the 5 substances with LD$_{50}$ >5000 mg/kg were correctly predicted. The predictivity for this category was 27%.

The overall accuracy of the 3T3 NRU test method for correctly predicting GHS acute oral toxicity classification of the 30 reference substances using the RC rat-only weight regression was 30% (see Table 3-3). In vivo toxicity was overpredicted for 33% and underpredicted for 37%. For this analysis, in terms of each GHS acute oral toxicity category:

• Zero (0%) of 5 substances with LD$_{50}$ <5 mg/kg was correctly predicted
• One (20%) of 5 substances in the 5 < LD$_{50}$ ≤50 mg/kg category was correctly predicted
• Three (60%) of 5 substances in the 50 < LD$_{50}$ ≤300 mg/kg category were correctly predicted
• Three (60%) of 5 substances in the 300 < LD$_{50}$ ≤2000 mg/kg category were correctly predicted.
• Two (40%) of the 5 substances in the 2000 < LD$_{50}$ ≤5000 mg/kg category were correctly predicted.
• Zero (0%) of the 5 substances with LD$_{50}$ >5000 mg/kg were correctly predicted.
Table 3-2  Prediction of GHS Acute Oral Toxicity Category by the 3T3 NRU Test Method Using the Recommended Reference Substances and the RC Rat-Only Millimole Regression

<table>
<thead>
<tr>
<th>Reference Rodent Oral LD₅₀² (mg/kg)</th>
<th>NRU-Predicted GHS Category (mg/kg)</th>
<th>Total</th>
<th>Accuracy</th>
<th>Toxicity Over-predicted</th>
<th>Toxicity Under-predicted</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>LD₅₀ &lt; 5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>5 &lt; LD₅₀ ≤ 50</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>20%</td>
<td>0%</td>
</tr>
<tr>
<td>50 &lt; LD₅₀ ≤ 300</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>80%</td>
<td>0%</td>
</tr>
<tr>
<td>300 &lt; LD₅₀ ≤ 2000</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>80%</td>
<td>0%</td>
</tr>
<tr>
<td>2000 &lt; LD₅₀ ≤ 5000</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>0%</td>
<td>100%</td>
</tr>
<tr>
<td>LD₅₀ &gt; 5000</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>20%</td>
<td>80%</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>0</td>
<td>8</td>
<td>27%</td>
<td>0%</td>
</tr>
</tbody>
</table>

Predictivity: 0% 33% 50% 27% 0% 100%

Category Overpredicted: 0% 0% 13% 47% 67% 0%

Category Underpredicted: 0% 67% 38% 27% 33% 0%

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

¹The RC rat-only millimole regression is log LD₅₀ (mmol/kg) = log IC₅₀ (mM) x 0.439 + 0.621. Numbers in table represent numbers of substances.
²From Table 3-1.
### Table 3-3  Prediction of GHS Acute Oral Toxicity Category by the 3T3 NRU Test Method Using the Recommended Reference Substances and the RC Rat-Only Weight Regression

<table>
<thead>
<tr>
<th>Reference Rodent Oral LD&lt;sub&gt;50&lt;/sub&gt; (mg/kg)</th>
<th>NRU- Predicted GHS Category (mg/kg)</th>
<th>Total</th>
<th>Accuracy</th>
<th>Toxicity Over-predicted</th>
<th>Toxicity Under-predicted</th>
</tr>
</thead>
<tbody>
<tr>
<td>LD&lt;sub&gt;50&lt;/sub&gt; &lt; 5</td>
<td>0</td>
<td>0</td>
<td>0%</td>
<td>0%</td>
<td>100%</td>
</tr>
<tr>
<td>5 &lt; LD&lt;sub&gt;50&lt;/sub&gt; ≤ 50</td>
<td>0</td>
<td>1</td>
<td>20%</td>
<td>0%</td>
<td>80%</td>
</tr>
<tr>
<td>50 &lt; LD&lt;sub&gt;50&lt;/sub&gt; ≤ 300</td>
<td>0</td>
<td>3</td>
<td>100%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>300 &lt; LD&lt;sub&gt;50&lt;/sub&gt; ≤ 2000</td>
<td>0</td>
<td>2</td>
<td>58%</td>
<td>80%</td>
<td>20%</td>
</tr>
<tr>
<td>2000 &lt; LD&lt;sub&gt;50&lt;/sub&gt; ≤ 5000</td>
<td>0</td>
<td>3</td>
<td>0%</td>
<td>0%</td>
<td>100%</td>
</tr>
<tr>
<td>LD&lt;sub&gt;50&lt;/sub&gt; &gt; 5000</td>
<td>0</td>
<td>0</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
</tbody>
</table>

**Predictivity**  
0% 100% 33% 20% 40% 0%  
**Category Overpredicted**  
0% 0% 22% 33% 60% 0%  
**Category Underpredicted**  
0% 0% 44% 47% 0% 0%  

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

1The RC rat-only weight regression is log LD<sub>50</sub> (mg/kg) = log IC<sub>50</sub> (ug/mL) x 0.372 + 2.024. Numbers in table represent numbers of substances.

2From Table 3-1.
4.0 REFERENCES


