TABLE OF CONTENTS

APPENDICES

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS AND ABBREVIATIONS

A-CUTE-TOX	A-Cute-Tox Project (EU Research & Development Integrated Project)
ADME	Absorption, distribution, metabolism, and elimination
ANOVA	Analysis of variance
ASTDR	Agency for Toxic Substances and Disease Registry
ASTM	American Society for Testing and Materials
ATC	Acute Toxic Class method
ATCC	American Type Culture Collection
ATWG	Acute Toxicity Working Group
BBB	Blood:brain barrier
BPE	Bovine pituitary extract
BRD	Background Review Document
^{o}C	Degrees Celsius
CAS	Chemical Abstracts Service
CASRN	Chemical Abstracts Service Registry Number
CCOHS	Canadian Centre for Occupational Health and Safety (CCOHS)
CDER	U.S. FDA Center for Drug Evaluation and Research
CESARS	Chemical Evaluation Search and Retrieval System
CFU	Colony forming units
CHRIS	Chemical Hazard Response
CI	Confidence interval
CICADS	Concise International Chemical Assessment Documents
CIS	ILO Occupational Safety and Health Information Centre
CNS	Central nervous system
COLIPA	The European Cosmetic Toiletry and Perfumery Association
CPSC	U.S. Consumer Product Safety Commission
CSF	Colony stimulating factor
CTFA	Cosmetic, Toiletries and Fragrance Association
CV	Coefficient of variation
DART®/ETIC	Developmental and Reproductive Toxicology/Environmental Teratology Information Center
DEA	U.S. Drug Enforcement Administration
DHHS	U.S. Department of Health and Human Services
DIMDI	Deutsches Institut fur Medizinische Dokumentation und Information (The German Institute for Medical Documentation and Information)
DNA	Deoxyribose nucleic acid
DMEM	Dulbecco's Modification of Eagle's Medium
DMSO	Dimethyl sulfoxide
D-PBS	Dulbecco's phosphate buffered saline
DOD	U.S. Department of Defense
DOT	U.S. Department of Transportation
EC	European Commission
EC_{50}	Concentration of a substance that produces 50% of the maximum possible response for that substance

ECBC	U.S. Army Edgewood Chemical Biological Center
ECETOC	European Centre for Ecotoxicology and Toxicology of Chemicals
EC/HO	European Commission/British Home Office
ECVAM	European Centre for the Validation of Alternative Methods
EDIT	Evaluation-guided development of new *in vitro* tests
EHC	Environmental Health Criteria
EHS	EPA's Extremely Hazardous Substance list
EPA	U.S. Environmental Protection Agency
ERG	Emergency Response Guidebook
ETOH	Ethanol (Ethyl alcohol)
EU	European Union
EXTONET	The Extension Toxicology Network
FAL	FRAME Alternatives Laboratory
FAO	UN Food and Agriculture Organization
FB1	Fumonisin B1
FDA	U.S. Food and Drug Administration
FDP	Fixed Dose Procedure
FIFRA	U.S. Federal Insecticide, Fungicide, and Rodenticide Act
FR	Federal Register
FRAME	Fund for the Replacement of Animals in Medical Experiments
GABA	Gamma amino butyric acid
GCCP	Good cell culture practices
GHS	Globally Harmonized System (of Classification and Labeling of Chemicals)
GLP	Good Laboratory Practices
gm	Grams
HBSS	Hanks' balanced salt solution
HPV	High Production Volume
hr	Hour(s)
HSDB	Hazardous Substances Data Bank
HSG	Health and Safety Guides
HTD	Highest tolerated dose
IARC	International Agency for Research on Cancer
IC_{20}	Concentration producing 20% inhibition of the endpoint measured
IC_{50}	Concentration producing 50% inhibition of the endpoint measured
IC_{80}	Concentration producing 80% inhibition of the endpoint measured
ICCVAM	Interagency Coordinating Committee for the Validation of Alternative Methods
ICSC	International Chemical Safety Cards
ID	Insufficient data
ID_{50}	Index of cytotoxicity; dose producing a 50% reduction in protein value
IIVS	Institute for *In Vitro* Sciences
ILO	International Labour Organisation
i.m.	Intramuscular
INVITOXX	*In Vitro* Techniques in Toxicology (ERGATT FRAME ECVAM Data bank)

IOM	Institute of Medicine
i.p.	Intraperitoneal
IPCS	International Programme on Chemical Safety
IRAG	Interagency Regulatory Alternatives Group
IRPTC	International Register of Potentially Toxic Chemicals
ISO	International Standards Organization
IUCLID	International Uniform Chemical Information Database
i.v.	Intravenous
JECFA	Joint Expert Committee on Food Additives
JMPR	Joint Meeting on Pesticide Residues
KBM®	Keratinocyte basal medium
kg	Kilogram
K_{ow}	Octanol-water partition coefficient
L	Liter
LC	Lethal blood concentration
LD_{50}	Dose that produces lethality in 50% of test animals
LDH	Lactate dehydrogenase
MAS	Maximum average Draize score
MEIC	Multicentre Evaluation of *In Vitro* Cytotoxicity
MeSH®	Medical Subject Heading
µL	Microliters
µm	Micrometers
µM	Micromoles
mg	Milligram
MIT	Metabolic inhibition test
mL	Milliliter
mM	Millimolar
MMAS	Modified maximum average score
mmol	Millimoles
MPE	Mean photo effect
MSDS	Material Safety Data Sheets
MTT	3-(4,5-Dimethylthiazol-2-yl)-2,5-diphenyl tetrazolium bromide
N	Number (of substances)
NA	Not applicable
NADH	Nicotine adenine dinucleotide (reduced)
NC	Not calculated
NCS	Newborn calf serum
NCTR	U.S. FDA National Center for Toxicological Research
n.d.	Not detectable
NHK	Normal human epidermal keratinocytes
NICEATM	National Toxicology Program Center for the Evaluation of Alternative Toxicological Methods
NIEHS	U.S. National Institute of Environmental Health Sciences
NIH	U.S National Institutes of Health
NIOSH	U.S. National Institute for Occupational Safety and Health
NLM	National Library of Medicine

NR	Neutral red
NRU	Neutral red uptake
NTP	U.S. National Toxicology Program
OAT	Organic anionic transporters
OD	Optical density
OD_{540}	Optical density (absorbance) at a wavelength of 540 nm
OECD	Organisation for Economic Co-operation and Development
OHM/TADS	EPA Oil and Hazardous Materials/Technical Assistance Data System
OPP	U.S. EPA Office of Pesticide Programs
OPPTS	EPA Office of Prevention, Pesticides and Toxic Substances
ORD	U.S. EPA Office of Research and Development
OSHA	U.S. Occupational Safety and Health Administration
OTA	Ochratoxin A
PBS	Phosphate buffered saline
PC	Positive control
PDS	Pesticide Data Sheets
pg	Picogram
PG	Packing group
PIF	Photoinhibition factor
PIMS	Poisons Information Monographs
pK	Acid/base dissociation constant
PLS	Partial Least Squares (analysis)
PPIS	EPA Pesticide Product Information System
PPT	Precipitate
QA	Quality assurance
QC	Quality control
R^2	Coefficient of determination
r_s	Spearman correlation coefficient
RC	Registry of Cytotoxicity
REACH	Registration, evaluation, authorisation and restriction of chemicals
RTECS®	Registry of Toxic Effects of Chemical Substances
RTK NET	The Right-to-Know Network
SD	Standard deviation
SIDS	OECD Screening Information Data Sets
SIS	Scientific Information Service
SLS	Sodium lauryl sulfate
SMT	Study management team
SOP	Standard operating procedure
3T3	BALB/c mouse fibroblasts, clone A31 (ATCC # CCL-163)
TESS	Toxic Exposure Surveillance System
TG	Test guideline
TRI	U.S. EPA Toxics Release Inventory
TSCA	Toxic Substances Control Act
UDP	Up-and-Down Procedure
UN	United Nations

UNEP	United Nations Environment Programme
USP	U.S. Pharmacopoeia
UV	Ultraviolet (light)
VC	Vehicle control
WHO	World Health Organization
XTT	2,3-Bis(2-methoxy-4-nitro-5-sulfophenyl)-2H-tetrazolium-5-carboxanilide
ZEBET	Zentralstelle zur Erfassung und Bewertung von Ersatz- und Ergänzungsmethoden zum Tierversuch (German Center for Documentation and Evaluation of Alternative Methods to Animal Experiments)

ACKNOWLEDGMENTS

The following individuals are acknowledged for their contributions to the in vitro acute toxicity test method review 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.
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.

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

Dave 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.

Jeffery 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

Jim Truax, M.A.
ILS, Inc.

Participants in the *In Vitro* Cytotoxicity Validation Study

NICEATM gratefully acknowledges the generous contributions of the individuals who directly participated in the NICEATM/ECVAM Validation Study. Their time and efforts are greatly appreciated.

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 – Advisor
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

[This Page Intentionally Left Blank]

PREFACE

The Institute of Medicine estimates that more than 4 million poisonings occur annually in the United States (Institute of Medicine [IOM] 2004). In 2001, 30,800 deaths placed poisoning as the second leading cause of injury-related death behind automobile accidents (42,433 deaths) (IOM 2004). In order to ensure that all potentially hazardous substances have proper warning labels, regulatory agencies require determination of acute toxicity hazard potential of substances and products. This determination for oral acute toxicity hazard is currently made using a test that requires laboratory rats. Historically, lethality estimated by the LD_{50} (i.e., the dose of a test substance that produces death in 50% of the animals tested) has been a primary toxicological endpoint in acute toxicity tests.

The conventional LD_{50} acute oral toxicity *in vivo* test method has been modified in various ways to reduce and refine[1] animal use in toxicity testing (OECD 2001a, c, d, e; EPA 2002a). Most recently, the LD_{50} was replaced, for hazard classification testing purposes, with the UDP, based on an Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) technical evaluation and formal ICCVAM recommendations (ICCVAM 2000, 2001c). This method now reduces animal use by over 70% compared to the previous method.

In 1999, at the request of the U.S. Environmental Protection Agency (EPA) Office of Pesticides, Prevention, and Toxic Substances, ICCVAM reviewed the validation status of *in vitro* methods for estimating acute oral toxicity. This request was based on studies published in recent years that showed a correlation between *in vitro* and *in vivo* acute toxicity. *In vitro* cytotoxicity methods have been evaluated as another means to reduce and refine the use of animals and these methods may be helpful in predicting *in vivo* acute toxicity. Since moving the starting dose closer to the LD_{50} reduces the number of animals necessary for the acute oral systemic toxicity test, the use of *in vitro* cytotoxicity assays to predict a starting dose close to the LD_{50} may reduce animal use.

In October of 2000, the International Workshop on *In Vitro* Methods for Assessing Acute Systemic Toxicity sponsored by the National Toxicology Program (NTP), the National Institute of Environmental Health Sciences (NIEHS) and the EPA was convened in Arlington, VA. The Organizing Committee invited 33 expert scientists from academia, industry, and government agencies to participate in the Workshop. 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

[1] 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).

Workshop participants concluded that none of the proposed *in vitro* methods 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 proposed by the German Center for Documentation and Evaluation of Alternative Methods to Animal Experiments (ZEBET) was recommended for rapid adoption so that data could be generated to establish its usefulness with a large number of chemicals (ICCVAM 2001a). In addition, a separate *Guidance Document on Using In Vitro Data to Estimate In Vivo Starting Doses for Acute Toxicity* (ICCVAM 2001b) was prepared to provide sample cytotoxicity protocols and instructions for using *in vitro* data to predict starting doses for acute *in vivo* systemic toxicity tests.

ICCVAM, which is charged with coordinating the technical evaluations of new, revised, and alternative test methods with regulatory applicability (ICCVAM Authorization Act of 2000, Public Law 106-545; available: http://iccvam.niehs.nih.gov/about/PL106545.pdf), agreed that *in vitro* basal cytotoxicity test methods should have a high priority for evaluation. The NTP 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 lethality assays. NICEATM and ECVAM designed a multi-laboratory validation study to evaluate the performance of two standardized *in vitro* basal cytotoxicity test methods using 72 reference substances with the ZEBET approach of using the Registry of Cytotoxicity (RC) regression model. Based on the procedures described in the *Guidance Document* (ICCVAM 2001b), the validation study used two mammalian cell types (i.e., BALB/c 3T3 mouse fibroblasts [3T3] and primary normal human epidermal keratinocytes [NHK]) for *in vitro* basal cytotoxicity test methods with a neutral red uptake (NRU) cell viability endpoint to predict starting doses for acute oral systemic toxicity test methods. The inclusion of human cells in the validation study also implements another workshop recommendation, that of evaluating whether cytotoxicity in human or rodent cells can be used to predict human acute toxicity.

The objectives identified for the validation study were to:

- Further standardize and optimize the *in vitro* NRU basal cytotoxicity protocols using 3T3 and NHK cells 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 basal cytotoxicity 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

Scientists assembled for the ICCVAM-sponsored scientific peer review panel meeting ("Panel") on May 23, 2006 independently assessed the usefulness and limitations of the *in vitro* basal cytotoxicity test methods to predict starting doses for acute oral systemic toxicity test methods. The Background Review Document (BRD) on the two *in vitro* NRU test methods prepared by NICEATM and provided to the peer review panel and the public contains:

1. Comprehensive summaries of the data generated in the validation study
2. An analysis of the accuracy and reliability of the test method protocols
3. Related information characterizing the potential animal savings produced by using the *in vitro* basal cytotoxicity test methods as adjuncts to specific acute systemic toxicity test methods

The Panel also evaluated draft test method performance standards, protocols, and draft ICCVAM recommendations for test method uses and future studies. The public was invited to provide comments on the BRD and other documents and to attend the Panel meeting. Prior to the Panel meeting, public comments provided about the documents were provided to the Panel for their consideration. The BRD can be obtained from the ICCVAM/NICEATM Web site (http://iccvam.niehs.nih.gov) or by contacting NICEATM.

Following the conclusion of the Panel meeting, the ICCVAM and its Acute Toxicity Working Group (ATWG) considered the Panel report, the performance standards for the use of *in vitro* basal cytotoxicity test methods to predict starting doses for acute systemic toxicity test methods, and any public comments in preparation of its final test method recommendations for these *in vitro* basal cytotoxicity test methods. These recommendations will be made available to the public and provided to the U.S. Federal agencies for consideration, in accordance with the ICCVAM Authorization Act of 2000 (Public Law 106-545).

On behalf of the ICCVAM, we gratefully acknowledge the many contributions of all who participated in the *in vitro* cytotoxicity validation study and those who assisted in the preparation of the documents evaluated at the peer review meeting. We extend a special thanks to the participating laboratory Study Directors and scientists who worked diligently to provided critical data and information. We also thank the ECVAM scientists who participated in the management of the validation study and who provided valuable information, comments, and opinions throughout the study. The efforts of the ATWG members were instrumental in assuring a complete and informative BRD. The efforts of the NICEATM staff in coordinating the validation study, providing timely distribution of

information, and preparing the various documents are acknowledged and appreciated. We especially acknowledge Dr. Judy Strickland and Mr. Michael Paris for their coordination of the validation study and preparation of the BRD and other documents.

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

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

EXECUTIVE SUMMARY

This Background Review Document (BRD) reports the results of a validation study, organized and managed 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), conducted to characterize two *in vitro* basal cytotoxicity tests for determining starting doses for rodent acute oral toxicity assays. In conducting this validation study, the protocols for two *in vitro* neutral red uptake (NRU) assays using BALB/c mouse fibroblast 3T3 cells (3T3) and normal human epidermal keratinocytes (NHK) were standardized and optimized, and the LD_{50} values for the reference substances were refined. The accuracy and reliability of the two *in vitro* NRU test methods were determined using 72 reference substances of various toxicities. Computer simulations were used to estimate the potential reduction in animal usage that could be accomplished by the use of either of these *in vitro* test systems. One outcome of this effort has been the generation of high quality *in vivo* lethality and *in vitro* cytotoxicity reference databases that will be useful in the development of other *in vitro* toxicity tests.

The validation study showed that the 3T3 and NHK NRU test methods are not sufficiently accurate as stand-alone methods to correctly predict rodent acute oral toxicity. However, based on computer simulations for the reference substances tested in this study, the use of either of these two *in vitro* basal cytotoxicity test methods for the selection of starting doses for rodent acute oral toxicity testing has the potential to reduce the number of animals used per test and, in some cases, the number of substance-induced animal deaths.

Introduction and Rationale

Although *in vitro* basal cytotoxicity test methods are not currently regarded as suitable replacements for rodent acute oral toxicity tests (Spielmann et al. 1999; ICCVAM 2001a), such methods have been examined as a possible approach to reduce and refine[2] the use of animals for such testing. An international Multicentre Evaluation of *In Vitro* Cytotoxicity (MEIC) was initiated in 1983 to evaluate the relationship between *in vitro* cytotoxicity and acute human toxicity. Tests of 50 substances in 61 *in vitro* assays by multiple laboratories led to the identification of a battery of three human cell line assays whose cytotoxicity responses were highly correlated to human lethal blood concentrations (Bondesson et al. 1989; Clemedson et al 1996, 1996a; Ekwall et al. 1998a, 1998b, 2000). The Registry of Cytotoxicity (RC), initially published in 1998, is a database of 347 substances that currently consists of acute oral toxicity data from rats and mice and *in vitro* cytotoxicity data from studies using various mammalian cell types with a number of different toxic endpoints (Halle 1998, 2003). A regression formula, the RC millimole regression, constructed from these data was proposed by ZEBET, the German National Centre for the Documentation and Evaluation of Alternative Methods to Animal Experiments, as a method to reduce animal use by identifying the most appropriate starting doses for acute oral toxicity tests (Halle 1998, 2003; Spielmann et al. 1999).

[2] 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).

These, and other, initiatives to use *in vitro* cytotoxicity test methods to reduce animal use in acute toxicity testing were evaluated at the International Workshop on *In Vitro* Methods for Assessing Acute Systemic Toxicity, in October 2000 ("Workshop 2000"; ICCVAM 2001a). This workshop was organized by the U.S. Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) and NICEATM. Pursuant to this workshop, ICCVAM recommended (ICCVAM 2001a) further evaluation of the use of *in vitro* cytotoxicity data as one of the approaches that could be used to estimate the starting doses for rodent acute oral toxicity studies. The recommendations are based on preliminary information suggesting that this approach could reduce the number of animals used in such studies (i.e., reduction), minimize the number of animals that receive lethal doses (i.e., refinement), and avoid underestimating hazard. To assist in the adoption and implementation of the ZEBET approach, the *Guidance Document on Using In Vitro Data to Estimate In Vivo Starting Doses for Acute Toxicity* (hereafter referred to as *Guidance Document*; ICCVAM 2001b) was prepared by ICCVAM with the assistance of the workshop participants.

In its recommendations for further evaluations, ICCVAM concurred with the Workshop 2000 recommendation that near-term validation studies should focus on two standard basal cytotoxicity assays: one using a human cell NHK system and one using a rodent cell (3T3) system. Historical data for *in vitro* cytotoxicity testing using mouse 3T3 cells are available (e.g., Balls et al. 1995; Brantom et al. 1997; Gettings et al. 1991, 1994a, 1994b; Spielmann et al. 1991, 1993, 1996), as are historical data for *in vitro* basal cytotoxicity testing using NHK cells (e.g., Gettings et al. 1996; Harbell et al. 1997; Sina et al. 1995; Willshaw et al. 1994).

NICEATM, in partnership ECVAM, designed an international, multi-laboratory validation study to evaluate the reduction or refinement in animal use that could result from using cytotoxicity data from the 3T3 and NHK NRU test methods to estimate starting doses for two rodent acute oral toxicity test methods, the Up-and-Down Procedure (UDP; OECD 2001a; EPA 2002a) and the Acute Toxic Class (ATC) method (OECD 2001d). The NRU protocols, as presented in the *Guidance Document*, were the initial basis of the NICEATM/ECVAM validation study protocols. These protocols were originally derived from the BALB/c 3T3 Cytotoxicity Test, INVITTOX Protocol No. 46 (available at the FRAME-sponsored INVITTOX database [http://embryo.ib.amwaw.edu.pl/invittox/]), the 3T3 cell studies by Borenfreund and Puerner (1984, 1985) and the rat epidermal keratinocyte study of Heimann and Rice (1983). A detailed description of the 3T3 and NHK NRU test method protocols used in the NICEATM/ECVAM validation study is provided in **Section 2**.

Protocol Components

Many protocol components used in the validation study are similar for the 3T3 and NHK cells. The following procedures are common to both cell types:

- Testing was performed in four phases (Ia, Ib, II, and III)
- Preparation of reference substances and positive control
- Cell culture environment conditions
- Determination of test substance solubility
- Configuration of 96-well plates for testing samples
- 48-hour exposure to test substance
- Range finder and definitive testing

- Microscopic evaluation of cell cultures for toxicity
- Measurement of NRU
- Data analysis

The main differences in the test methods for the two cell types are:

- The conditions of propagation of the cells in culture
- The cell growth medium components
- The volumes of reference substance added to the 96-well plate

Three laboratories participated in testing the 72 reference substances in both cell types:

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

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 *in vitro* tests.

Validation Reference Substances

The 72 reference substances were selected to represent: (1) the complete range of *in vivo* acute oral toxicity (encompassing all five GHS acute oral toxicity categories as well as lower toxicities [GHS; UN 2005]); (2) the types of substances regulated by various regulatory authorities; and (3) substances with human toxicity data and/or human exposure potential. To ensure that the complete range of toxicity was covered, 12 substances were selected for each of the five acute oral toxicity categories, with an additional 12 substances with lower toxicities (i.e., LD_{50} >5000 mg/kg). A discussion of the characteristics and sources of the reference substances can be found in **Section 3**. The selected reference substances had the following characteristics:

- 58 (81%) of the 72 substances were also included in the RC, and 38% (22/58) of these were outliers with respect to the RC millimole regression.
- 27 (35%) of the substances were pharmaceuticals, 17 (22%) were pesticides, 8 (10%) were solvents, and 5 (6%) 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.
- 57 (79%) were organic compounds and 15 (21%) were inorganic; well-represented classes of organic compounds included heterocyclics, carboxylic acids, and alcohols.
- 22 (31%) substances were known, or expected to have, toxicologically active metabolites.
- Many of the selected substances had multiple target organs/effects; including neurological, liver, kidney, and cardiovascular effects.

Table ES-1 reports the number of substances that were tested and the number of substances used for the various analyses performed.

Table ES-1 Datasets Used for Validation Study Analyses[1]

Use	3T3 NRU[1]	NHK NRU[1]	Characteristics of Dataset
Testing	72	72	Substances tested
Comparison of laboratory IC_{50}-LD_{50} regressions to one another	47	51	RC substances with IC_{50} values from all laboratories and reference rat oral LD_{50} values
Comparison of combined-laboratory IC_{50}-LD_{50} regressions to a regression calculated with RC data	47	47	RC substances with IC_{50} values for both test methods from all laboratories and rat oral reference LD_{50} values
Prediction of GHS accuracy using IC_{50} values in IC_{50}-LD_{50} regressions; prediction of starting doses for acute oral toxicity test (UDP and ATC) simulations	67	68	Substances with IC_{50} values from at least one laboratory
Reproducibility of acceptable rat oral LD_{50} values	NA	NA	62 substances with more than one acceptable rat oral LD_{50} value
Reproducibility of IC_{50} values	64	68	Substances with IC_{50} values from all laboratories

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

Rodent Acute Oral Toxicity Reference Data

Because the 3T3 and NHK NRU test methods are intended to be used as adjuncts to rodent acute oral toxicity test methods, the LD_{50} values from rodent acute oral toxicity tests are the most appropriate reference data for evaluating the *in vitro* IC_{50} values (i.e., the test chemical concentration that reduces cell viability by 50%). Rodent acute oral LD_{50} reference data for the 72 reference substances were obtained from the literature. It was not possible to limit the data to studies conducted under Good Laboratory Practice (GLP) guidelines (OECD 1998; EPA 2003a, 2003b; FDA 2003) because only 2% of the published data retrieved were from such studies. Although mouse toxicity data were initially considered for inclusion in the database, the accuracy analyses were restricted to rat data. A total of 459 acute rodent oral LD_{50} values were identified for the reference substances. Reference LD_{50} values for each substance were identified by excluding studies with the following characteristics:

- Feral rats
- Rats <4 weeks of age
- Anesthetized rats
- Test substance administered in food or capsule
- LD_{50} reported as a range or an inequality

For substances with multiple LD_{50} values (i.e., from different sources), the rodent reference LD_{50} values for use in the validation study were determined by calculating a geometric mean of the available values for each reference substance. The reference LD_{50} values for 19 (26%) of the 72 substances varied sufficiently from the initial LD_{50} values that came from the RC

database and other summary sources, that the substances were reclassified into different GHS categories.

The reliability of the calculated rat acute oral LD_{50} reference values was assessed by comparison to other evaluations of the performance of rodent acute oral toxicity tests. For the 62 reference substances that had more than one LD_{50} value, the maximum:minimum ratios ranged from 1.1 to 25.9, with most below an order of magnitude.

Test Method Accuracy

Although the 3T3 and NHK NRU test methods are not intended to be used as replacements for rodent acute oral toxicity tests, they were evaluated for their ability to correctly predict the reference LD_{50} values (i.e., accuracy[3]). The rationale for evaluating the accuracy of LD_{50} predictions is that the current acute oral toxicity test methods (i.e., UDP, ATC, and Fixed Dose Procedure [FDP; OECD 2001c]) call for starting doses to be placed as close as possible and just below the true LD_{50}. When the starting dose is close to the true LD_{50} for a test substance, fewer animals are needed. When the starting dose is below the true LD_{50}, there is reduced pain and suffering because doses tend to be lower, and the test outcome bias is more conservative (i.e., higher toxicity). Regression models developed using IC_{50} and LD_{50} values were used to derive estimated LD_{50} values from 3T3 or NHK NRU IC_{50} values.

A number of different analyses were performed in an attempt to improve the estimation of the rat acute oral LD_{50}. IC_{50}-LD_{50} regressions (in millimole units) were calculated for each *in vitro* cytotoxicity test method and participating laboratory using the 3T3 and NHK IC_{50} values. Because the regressions for each NRU test method among laboratories were not significantly different from one another (for each NRU test method, $p > 0.5$), the regression for each NRU test method was based on data pooled across the laboratories. This combined-laboratory regression was then compared to the RC data using a regression based on RC IC_{50} and LD_{50} data for the 47 substances common to the validation study and the RC, with rat acute oral LD_{50} reference values, and with both 3T3 and NHK IC_{50} values produced by all three participating laboratories. The statistical comparison of slope and intercept (simultaneously) using an F test showed that neither the 3T3 regression nor the NHK regression was significantly different from the RC regression for the 47 substances ($p = 0.61$ and 0.76 respectively). These outcomes support use of the RC millimole regression.

Reference substances that fit the RC millimole regression poorly (i.e., outliers) were evaluated to determine whether there were relationships between their outlier status and their physical or chemical characteristics. Because the IC_{50}-LD_{50} regressions for the 3T3 and NHK NRU test methods yielded results that were not different from the RC regression for 47 substances, the RC millimole regression was preferred for analysis of outliers because it was based on a much larger data set and because it had established acceptance limits (Halle 1998, 2003). Certain chemical structural classes, boiling points, molecular weights, and log K_{OW} values were related with outliers, but solubility in the 3T3 or NHK medium and the cells' lack of xenobiotic metabolic capability did not correlate with outlier status. Because these *in vitro* NRU test methods are based upon basal cytotoxicity, the mechanism of toxicity was

[3] Accuracy is the agreement between a test method result and an accepted reference value (ICCVAM 2003).

also considered as a characteristic to explain the presence of outliers. Twenty-two reference substances were neurotoxic or cardiotoxic and were not expected to be active in the 3T3 and NHK cell cultures. Of these 22 substances, 13 (59%) were outliers (i.e., they fit the RC millimole regression poorly) using the 3T3 NRU and 12 (55%) were outliers using the NHK NRU. These substances represented 13/28 (46%) and 12/31 (39%) of the outliers for the 3T3 and the NHK NRU test methods, respectively. More information on the outlier analysis is presented in **Section 6.2**.

The potential variation produced by combining the LD_{50} values of two rodent species in the RC millimole regression was eliminated by developing a regression based solely on RC substances with rat LD_{50} data (i.e., the RC rat-only millimole regression). The RC rat-only data were also converted to a weight basis for an additional regression, the RC rat-only weight regression, for applicability to mixtures or to substances for which molecular weight is unknown.

The accuracy of the *in vitro* NRU test methods when used with each of the IC_{50}-LD_{50} regressions was characterized by determining the proportion of reference substances for which their GHS categories (based on rat acute oral LD_{50} data) were correctly predicted. The accuracy of the RC rat-only millimole regression was 31% (21/67 reference substances) and 29% (20/68 reference substances) with the 3T3 and the NHK NRU test methods, respectively. The accuracy of the RC rat-only weight regression was similar, 31% with the 3T3 NRU test method (21/67 reference substances) and 31% with the NHK NRU test method (21/68 reference substances). The poor accuracy is due, in part, to the skewness of the reference substance set with respect to the fit of the reference substances to the regressions and to the differences between cell cultures and whole animal exposures. Each regression showed a general trend to underpredict the toxicity of the most toxic chemicals, and to overpredict the toxicity of the least toxic chemicals. A detailed discussion of the accuracy analyses is presented in **Section 6.4.**

Test Method Reliability

Reproducibility is the consistency of individual test results obtained in a single laboratory (intralaboratory reproducibility) or among different laboratories (interlaboratory reproducibility) using the same protocol and test samples. Reproducibility was evaluated using results from the 64 reference substances tested in 3T3 cells and the 68 substances tested in NHK cells that yielded replicate IC_{50} values in all three laboratories. Intra- and inter-laboratory reproducibility of the 3T3 and NHK NRU IC_{50} data was assessed using analysis of variance (ANOVA), coefficient of variation (CV) analysis, comparison of the laboratory-specific IC_{50}-LD_{50} regressions, and comparison of maximum:minimum mean laboratory IC_{50} values. Reproducibility was generally better with the NHK NRU test method.

Although ANOVA results for the positive control (sodium lauryl sulfate [SLS]) IC_{50} values from the 3T3 NRU test method indicated that there were significant differences among laboratories ($p = 0.006$) but not between study phases within laboratories ($p > 0.01$), the data show (see **Figure 7-5**) that laboratory means and standard deviations from each testing phase overlap , and that the IC_{50} was stable between testing phases. The interlaboratory CV values for the various study phases ranged from 2 to 16%. ANOVA results for the SLS IC_{50} from

the NHK NRU test method showed significant differences among laboratories ($p < 0.001$) and among study phases within laboratories ($p \leq 0.001$). The use of a different cell culture method at FAL was responsible for SLS IC_{50} differences among the laboratories in Phases Ia and Ib. After harmonization of culture methods across laboratories, the laboratory means and standard deviations were similar for Phases II and III (see **Figure 7-5**). Interlaboratory CV values for the NHK NRU for Phases Ia and Ib, were 39% and 21%, respectively. Interlaboratory CV values for Phases II and III were 31% and 8%, respectively. The linear regression analyses of the SLS IC_{50} over time (within each laboratory) for both NRU test methods indicated that IC_{50} values generated over the 2.5-year duration of the study were stable.

For the reference substances, the similarity among the laboratories' LD_{50} predictions (via regression) from IC_{50} values (see **Figure 7-1**) was considered significant with respect to the reproducibility analyses because these *in vitro* NRU test methods are proposed for use in determining starting doses for acute oral toxicity tests using the predicted LD_{50}. ANOVA showed significant laboratory differences for 23 substances with the 3T3 NRU test method (see **Table 7-4**) and six substances with the NHK NRU test method (see **Table 7-6**). Mean intralaboratory CV values were 26% for both NRU test methods, but the NHK NRU test method had a lower mean interlaboratory CV (28% vs. 47%). An analysis to determine the relationship, if any, between reference substance attributes and interlaboratory CV indicated that chemical class, physical form, solubility, and volatility had little effect. The CV seemed to be related instead to the GHS hazard category, the IC_{50}, and boiling point (see **Section 7.2.3**). However, the usefulness of these relationships is not known. Mean interlaboratory CV values were larger for substances in the most toxic GHS hazard categories than for substances in the other toxicity categories, especially with the 3T3 NRU test method. The mean interlaboratory CV for substances in the $LD_{50} \leq 5$ mg/kg (72%) and $5 < LD_{50} \leq 50$ mg/kg (78%) categories were larger than the mean overall interlaboratory CV (47%) with the 3T3 NRU test method. When the NHK NRU test method was used, the mean interlaboratory CV was 37% for substances with $LD_{50} \leq 5$ mg/kg, and 41% for substances with $5 < LD_{50} \leq 50$ mg/kg, and the mean overall interlaboratory CV was 28%. A Spearman correlation analysis indicated that the IC_{50} was inversely correlated to interlaboratory CV for both the 3T3 ($p = 0.015$) and NHK ($p = 0.014$) NRU test methods, and that boiling point was positively correlated to interlaboratory CV ($p = 0.007$) for the 3T3 but not the NHK ($p = 0.809$) NRU test method.

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

Data Quality

The laboratories reported no significant deviations from the protocols, and deviations that did occur were acknowledged and addressed by the Study Directors. Tests that had deviations affecting the data were rejected by the Study Directors and repeated. The computation of test method and data collection errors showed that the non-GLP laboratory consistently had the highest error rate and the lowest intralaboratory reproducibility for IC_{50} results; however, the

laboratory's GHS acute oral toxicity category predictions were comparable to that for the other laboratories.

An electronic copy of all data for the validation study can be obtained from NICEATM upon request by mail, fax, or e-mail to Dr. William S. Stokes, NICEATM, NIEHS, P. O. Box 12233, MD EC-17, Research Triangle Park, NC, 27709, (phone) 919-541-2384, (fax) 919-541-0947, (e-mail) niceatm@niehs.nih.gov.

Other Scientific Reports and Reviews

3T3 and NHK NRU methods have been evaluated for purposes other than the prediction of starting doses for acute toxicity studies (e.g., ocular irritancy; human lethal blood concentrations, *in vivo* phototoxicity). *In vitro* NRU cytotoxicity test methods using various cell types have been evaluated for their correlation with rodent lethality endpoints (e.g., rat/mouse intravenous[i.v.], intraperitoneal [i.p.], and oral toxicity). Peloux et al. (1992) and Fautrel et al. (1993) showed good correlations (r =0.88) of *in vitro* cytotoxicity with rodent i.p./i.v. and i.v. toxicity data, respectively. A 3T3 NRU test method has been validated by ECVAM for the identification of *in vivo* phototoxic potential.

No *in vitro* test methods have been validated for the prediction of acute oral toxicity. Estimations of animal savings using *in vitro* cytotoxicity data to estimate starting doses for the UDP did not use actual *in vitro* cytotoxicity data. Instead, animal savings were estimated by assuming that the *in vivo* starting dose equals the true LD_{50}, which is an approach that assumes that cytotoxicity data can perfectly predict *in vivo* lethality. These theoretical predictions of animal savings in the UDP ranged from 25-40% (ICCVAM 2001a), as compared with the average animal savings of 5.3-7.8% predicted using computer simulation modeling of the UDP for the reference substances tested in the NICEATM/ECVAM study. Halle et al. (1997) used the *in vitro* cytotoxicity data in the RC to determine that an animal savings of 32% can be attained for the ATC method by using the LD_{50} predicted by the RC millimole regression as the starting dose. For the reference substances tested in the NICEATM/ECVAM validation study, most of which were a poor fit to the RC millimole regression, the average animal savings for the ATC, as determined by computer simulation modeling, was 4.8-10.2%.

Animal Welfare Considerations: Reduction, Refinement, and Replacement

Computer models were used to simulate testing of the reference substances using to the UDP and ATC test methods. In principle, animal savings with the FDP could be estimated even though death is not the primary endpoint, but the validation study did not include this analysis. The number of animals that would be used, and the number of animals that would survive or die during the UDP or ATC procedure, were determined for the default starting doses and compared with those when starting dose was based on LD_{50} values determined from IC_{50} values for each reference substance using the RC rat-only regressions.

Computer simulation of UDP testing showed that, for the reference substances used in this validation study, using the 3T3 or NHK NRU test methods and the RC rat-only millimole regression to identify the starting dose resulted in the use of fewer animals per test by an average of 5.3% (0.50 animals) to 6.6% (0.53 animals), depending upon the assumed

mortality-response slope and *in vitro* NRU test method used. The RC rat-only weight regression predicted mean animal savings of 6.0% (0.56 animals) to 7.8% (0.62 animals). When substances were grouped by GHS acute oral toxicity category, there were no animal savings for substances in the 50 <LD_{50} ≤300 mg/kg category because the default starting dose is in this range. The greatest animal savings were observed for substances with 2000 < LD_{50} ≤5000 mg/kg and LD_{50} >5000 mg/kg because the limit test, which would be used for such substances, uses fewer animals that the main test. Animal savings for these toxicity categories using the RC rat-only millimole regression ranged from 11.3% (1.21 animals) to 20.3% (1.58 animals) per test. Use of the RC rat-only weight regression produced animal savings of 12.8% (1.38 animals) to 21.0% (1.63 animals) per test. Although the use of the 3T3 and NHK NRU test methods to estimate starting doses for the simulated UDP decreased the numbers of animals used per test, it did not change the numbers of animals that died.

Computer simulation of ATC testing showed that, for the reference substances used in this validation study, using the 3T3 or NHK NRU test methods and the RC rat-only millimole regression to identify the starting dose resulted in a savings of 4.8% (0.51 animals) to 7.3% (0.80 animals) per test, depending upon the assumed mortality-response slope and the *in vitro* NRU test method used. The use of the RC rat-only weight regression produced animal savings of 8.6% (0.91 animals) to 10.2% (1.09 animals). When substances were grouped by GHS acute oral toxicity category, there were no animal savings for substances in the 300 < LD_{50} ≤2000 mg/kg category because this category contains the default starting dose for the ATC method. Animal savings were highest for substances with 5 < LD_{50} ≤50 mg/kg and LD_{50} >5000 mg/kg. The mean animal savings for both *in vitro* NRU test methods for substances with 5 < LD_{50} ≤50 mg/kg ranged from 9.8% (1.15 animals) to 11.4% (1.33 animals) per test for the RC rat-only millimole regression. The greatest reduction in animal use would be for substances with LD_{50} >5000 mg/kg because the limit test used fewer animals than the main test. Animal savings for these substances ranged from 17.1%, (2.03 animals) to 22.2% (2.66 animals) per test for the RC rat-only millimole regression. When the RC rat-only weight regression was used, the mean animal savings with both *in vitro* NRU test methods for substances with 5 < LD_{50} ≤50 mg/kg ranged from 10.8% (1.25 animals) to 13.0% (1.51 animals) per test. Mean animal savings for substances with LD_{50} >5000 mg/kg ranged from 24.8% (2.94 animals) to 27.7% (3.33 animals) per test. The use of IC_{50} values to estimate starting doses for the ATC tests refined animal use by producing fewer animal deaths by approximately 0.5 to 0.6 animals per test.

Simulations for the UDP and ATC method showed that the use of cytotoxicity results to estimate starting doses did not significantly alter the GHS categorizations compared with the categories determined using default starting doses. This concordance was 97 to 99% for the 3T3 and NHK NRU test methods.

Practical Considerations

Practical issues with respect to the implementation of these *in vitro* NRU test methods include the need for, and availability of, appropriate cell culture equipment, training and expertise, cost, and time expenditure. The ECVAM Good Cell Culture Practice Task Force Report 1 (Hartung et al. 2002) encourages the establishment of laboratory practices and

principles that will reduce uncertainty in the development and application of *in vitro* test methods.

All equipment and supplies are readily available, and the *in vitro* NRU test methods are easily transferable to laboratories experienced with mammalian cell culture techniques. Much of the training and expertise needed to perform the 3T3 and NHK NRU test methods are common to people with mammalian cell culture experience. Additional technical training would not be intensive because these methods are similar in general performance to other *in vitro* mammalian cell culture assays. GLP training should be provided to laboratory personnel (including study directors and principal investigators) to ensure proper adherence to test protocols and data documentation and verification procedures.

Prices for commercial *in vitro* NRU cytotoxicity testing to determine the IC_{50} for one substance ranged from $1120 to $1850. It is not clear if the price of an *in vivo* test would be reduced if it were preceded by an *in vitro* cytotoxicity test to set the starting dose. Thus, use of these test methods may not reduce the overall cost of rodent acute oral toxicity testing and may increase the cost, but their use has the potential to reduce the number of animals and the time needed for a study. The greatest savings in time and animals will occur if the IC_{50} data determine that the rodent acute oral toxicity limit test should be performed, rather than the main test. Based on the cost and technical procedures associated with cell culture maintenance, the 3T3 NRU test method is less expensive and less complicated to conduct than the NHK NRU test method.