Reproducibility Analyses for *In Vitro* Neutral Red Uptake Methods From a Validation Study to Evaluate *In Vitro* Cytotoxicity Assays for Estimating Rodent Acute Systemic Toxicity

MW Paris[1, 2], JA Strickland[1,2], S Casati[3], RR Tice[1], WS Stokes[1], H Raabe[4], C Cao[5], R Clothier[6], J Harbell[4], G Mun[4], A Sizemore[4], GO Moyer[4], J Madren-Whalley[5], C Krishna [5], M Owen[6], N Bourne[6], E Harvey[8], R Lee[8], JH Haseman[7], P Crockett[8], ML Wenk[9], and MK Vallant[10]

[1]National Toxicology Program (NTP) Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM), NIEHS/NIH/DHHS, Research Triangle Park, North Carolina, USA; [2]Integrated Laboratory Systems, Inc., Research Triangle Park, North Carolina, USA; [3]ECVAM, JRC, Ispra, Italy; [4]IIVS, Gaithersburg, MD; [5]US Army ECBC, Aberdeen Proving Ground, MD; [6]Univ. of Nottingham, Nottingham, UK; [7]Consultant, Raleigh, NC; [8]Constella Group, Durham, NC, [9]BioReliance Corp, Rockville, MD; [10]NIEHS, RTP, NC

Introduction

Accidental poisoning is a more serious public health problem than is generally recognized. The Institute of Medicine estimates that more than 4 million poisoning episodes occur annually in the United States (Institute of Medicine 2004). In 2001, 30,800 deaths placed poisoning as the second leading cause of injury-related death behind automobile accidents (42,433 deaths) (Institute of Medicine 2004). The hazard potential for poisoning in humans is assessed by acute oral toxicity testing in rodents, which is a regulatory requirement for many substances and products. However, ethical and societal demands call for decreasing the numbers of animals used for such studies.

In October, 2000, the International Workshop on *In Vitro* Methods for Assessing Acute Systemic Toxicity reviewed the validation status of *in vitro* methods directed toward reducing and refining the use of laboratory animals for acute oral systemic toxicity (i.e., lethality) testing (ICCVAM 2001a). Workshop participants reviewed data demonstrating that animal use could be reduced, for sequential *in vivo* testing procedures such as the Up-and-Down Procedure (UDP; EPA 2002; OECD 2001a) and Acute Toxic Class (ATC; OECD 2001b) methods, if the starting dose was close to the oral LD_{50}. Spielmann et al. (1999) showed that *in vitro* basal cytotoxicity assays could be used to predict starting doses for *in vivo* acute systemic toxicity assays. Thus, one of the workshop recommendations for reducing and refining the use of animals for acute systemic toxicity assays was the publication of guidance for using *in vitro* basal cytotoxicity assays to estimate the starting doses for acute oral lethality assays (ICCVAM 2001b). The guidance publication provides details and examples on how to execute such an approach (ICCVAM 2001b).

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) subsequently designed and initiated a multi-laboratory validation study using the approach described in the *Guidance Document* (ICCVAM 2001b). The objectives of the validation study were:

- to further standardize and optimize two *in vitro* basal cytotoxicity protocols for neutral red uptake (NRU) assays using BALB/c 3T3 mouse fibroblasts (3T3) and normal human keratinocytes (NHK) to maximize intra- and inter-laboratory reproducibility.
- to assess the accuracy of these standardized *in vitro* cytotoxicity assays for estimating rodent oral LD_{50} values (**see poster 1969**) and human lethal concentrations across the Globally Harmonised System (GHS; OECD 2001) categories of acute oral toxicity.
- to estimate the reduction and refinement (i.e., reduced pain, suffering, and deaths) in animal use that would result from using these *in vitro* cytotoxicity assays to estimate the starting dose for *in vivo* acute toxicity tests (**see poster 1968**).
- to generate a high quality *in vitro* database that can be used to support investigation of other methods necessary to improve the accuracy of *in vitro* assessments of acute systemic toxicity.

The following posters also illustrate facets of the NICEATM/ECVAM Validation Study.

- **Poster 1968:** Reducing Animal Use In Acute Systemic Toxicity Testing By Using In Vitro Cytotoxicity Assays For Estimating Starting Doses
- **Poster 1969:** The Use Of Mouse Fibroblast (3T3) And Normal Human Epidermal Keratinocyes (NHK) Cytotoxicity Assays For Estimating Acute Oral Toxicity Of Formulations And Mixtures.

More information on ICCVAM and NICEATM can be accessed at: http://iccvam.niehs.nih.gov/

ICCVAM NICEATM

U.S. Department of Health and Human Services
National Institutes of Health
National Institute of Environmental Health Sciences

In Vitro Basal Cytotoxicity Test Method

The Neutral Red Uptake (NRU) Cytotoxicity Assay[1]

This cytotoxicity assay is based on the ability of viable cells to incorporate and bind neutral red (NR), a weak cationic supravital dye. NR readily penetrates cell membranes by non-ionic diffusion and accumulates intracellularly in lysosomes. Cytotoxicity is expressed as a concentration dependent reduction of the uptake of NR after chemical exposure.

General NRU Procedures

- BALB/c 3T3 cells or NHK cells are seeded into 96-well plates to form a sub-confluent monolayer (24 h for 3T3 cells; 48-72 h for NHK cells)
- Cells are exposed for 48 h to the reference substance in culture medium
- Culture medium is removed and NR dye medium is added for a 3 h incubation
- NR medium is discarded and NR desorbing fixative is added to the cells
- NR absorption is measured at optical density (OD) 540nm ± 10nm (uptake test method endpoint)
- Percent of control values at which cell viability or growth is inhibited is calculated to generate an IC_{50} value in g/mL
- The IC_{50} value is used in the regression formula to estimate the rodent acute oral LD_{50} value in mg/kg

[1]see Borenfreund and Puerner (1984); protocol of Riddell et al. (1986).

Study Phases

Seventy-two coded substances (12 per GHS class) covering a wide range of toxicity were tested in several phases using NRU assays with 3T3 cells and human NHK cells.

Phase Ia: Laboratory Evaluation Phase

At least 10 replicate tests of the positive control (PC) chemical (sodium laurel sulfate [SLS]) were performed with each cell type. The mean IC_{50} ± 2 standard deviations was calculated for each cell type for each laboratory and these values were used as acceptance criteria for PC performance in future assays. The protocols were revised as necessary to achieve intra-/inter-laboratory reproducibility.

Phase Ib: Laboratory Evaluation Phase

Each laboratory tested the same three coded substances of varying toxicities and generated three replicate acceptable tests/substance with each cell type. The protocols were refined and testing repeated, if necessary, to increase intra-/inter-laboratory reproducibility.

Phase II: Laboratory Qualification Phase

Each laboratory tested the same nine coded substances covering the full range of GHS toxicity categories and generated three replicate acceptable tests/substance with each cell type. Each laboratory assured that corrective actions taken in Phase I achieved the desired results. The protocols were further refined and testing repeated, if necessary, to increase test method reproducibility. Protocols were finalized for Phase III.

Phase III: Laboratory Testing Phase

Each laboratory tested the same 60 coded substances covering the full range of GHS toxicity categories and generated three replicate acceptable tests/substance with each cell type using the final protocols.

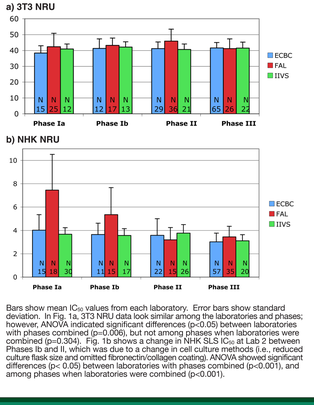
Reproducibility

Intralaboratory reproducibility is the agreement of results produced when qualified people within the same laboratory perform the test method using the same test protocol at different times (ICCVAM 2003). Interlaboratory reproducibility is the agreement of results from different qualified laboratories using the same protocol and reference substances. Interlaboratory reproducibility indicates the extent to which a test method can be transferred successfully among laboratories.

Intra- and inter-laboratory reproducibility of the 3T3 and NHK NRU test methods were determined using:

- analysis of variance (ANOVA). IC_{50} values were first converted to mM units and then log transformed to obtain normal distributions. The results for each substance were analyzed using one-way ANOVA with SAS PROC GLM (SAS Institute 1999). When significant differences among the laboratories were detected ($p<0.05$), contrast analyses were performed to compare the results of each laboratory with that of the other two laboratories. Due to the multiple comparisons performed for the contrasts, a significant difference among the laboratories was defined as $p<0.01$.
- coefficient of variation (CV) analysis. The coefficient of variation (CV) was calculated by dividing the SD by the mean IC_{50} value and then multiplying by 100.
- comparison of laboratory-specific IC_{50}-LD_{50} regressions to one another (for each test method) and by evaluating laboratory concordance for the GHS acute oral toxicity category predictions.

Figure 1. Variability of the positive control, SLS, IC_{50}



Bars show mean IC_{50} values from each laboratory. Error bars show standard deviation. In Fig. 1a, 3T3 NRU data look similar among the laboratories and phases; however, ANOVA indicated significant differences ($p<0.05$) between laboratories with phases combined ($p=0.006$), but not among phases when laboratories were combined ($p=0.304$). Fig. 1b shows a change in NHK SLS IC_{50} at Lab 2 between Phases Ib and II, which was due to a change in cell culture methods (i.e., reduced culture flask size and omitted fibronectin/collagen coating). ANOVA showed significant differences ($p< 0.05$) between laboratories with phases combined ($p<0.001$), and among phases when laboratories were combined ($p<0.001$).

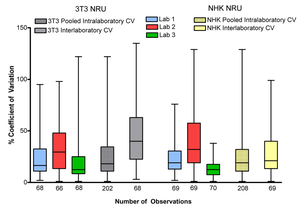
Table 1. Linear Regression Analysis of SLS IC_{50} Over Time[1]

Test Method/ Laboratory	Slope	P-value (Slope)[2]	Intercept
3T3 NRU			
Lab 1	0.0002040	0.001	-0.874
Lab 2	-0.0003240	0.012	-0.796
Lab 3	0.0000304	0.651	-0.850
NHK NRU			
Lab 1	-0.000559	0.002	-1.901
Lab 2	-0.001120	0.000	-1.737
Lab 3	-0.000445	0.002	-1.885

[1]Time was expressed as index values.
[2]Statistically significant from zero at $p<0.05$.

Table 1 shows that the slopes of the linear regressions of the IC_{50} values over time (expressed as index values) were statistically different from zero for Lab 1 and Lab 2 ($p<0.001$ and 0.012, respectively). Since the slopes were so small, they were considered to be unimportant.

Figure 2. Reproducibility of 3T3 and NHK NRU by Coefficient of Variation Analysis



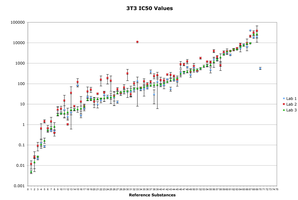
% Coefficient of variation (CV) = standard deviation/mean X 100. Boxes show median, and lower (Q1) and upper (Q3) quartiles. Error bars show range. Total number of reference substances is less than 72 because of insufficient toxicity to produce an IC_{50} for some substances in some or all laboratories. Intralaboratory CV values are shown for each laboratory and each assay (pooled laboratory data). Mean intralaboratory CV = 26% for both the 3T3 and NHK NRU test methods. Mean interlaboratory CV = 46% for 3T3 and 28% for NHK. Although no criterion for acceptable CV was determined for this study, other studies have used CV < 30% as an acceptable CV range for both intra- and inter-laboratory reproducibility to reflect an acceptable maximum for normal biological variability (ECVAM: Zuang et al. 2002; Fentem et al. 2001).

Table 2. NICEATM/ECVAM Validation Study Reference Substances

1	Aminopterin	37	Strychnine[1]
2	Triphenyltin hydroxide	38	Phenylthiourea[1]
3	Colchicine[1]	39	Lindane
4	Cycloheximide	40	Carbamazepine
5	Triethylenemelamine[1]	41	Diethyl phthalate
6	Sodium dichromate dihydrate	42	Glutethimide[1]
7	Sodium arsenite	43	Chloramphenicol[1]
8	Cadmium II chloride	44	Chloral hydrate[1]
9	Hexachlorophene	45	Caffeine
10	Mercury II chloride	46	Digoxin[2]
11	Endosulfan	47	Meprobamate[1]
12	Arsenic III trioxide[1,2]	48	Acetylsalicylic acid[1]
13	Diquat dibromide monohydrate	49	Nicotine
14	Haloperidol[1]	50	Phenobarbital
15	Cupric sulfate pentahydrate[1]	51	Procainamide HCl[1]
16	Thallium I sulfate	52	Valproic acid[1]
17	Amitriptyline HCl	53	Xylene
18	Fenpropathrin	54	Trichloroacetic acid[1]
19	Propylparaben[1]	55	Citric acid[1,2]
20	Propranolol HCl[2]	56	Boric acid
21	Dichlorvos[1]	57	5-Aminosalicylic acid
22	Potassium cyanide[1]	58	Sodium hypochlorite
23	Physostigmine	59	Lactic acid
24	Dibutyl phthalate[1]	60	Potassium I chloride
25	Parathion	61	2-Propanol[1]
26	Paraquat	62	Sodium chloride
27	Sodium selenate[1]	63	Dimethylformamide[2]
28	Verapamil HCl	64	Ethanol
29	Acetaminophen	65	Gibberellic acid
30	Busulfan[1]	66	Acetonitrile
31	Sodium oxalate	67	1,1,1-Trichloroethane[1]
32	Phenol	68	Ethylene glycol[1]
33	Disulfoton[1,2]	69	Glycerol
34	Epinephrine bitartrate	70	Lithium I carbonate
35	Atropine sulfate	71	Carbon tetrachloride
36	Sodium I fluoride	72	Methanol[2]

[1]The ANOVA results indicate that there were statistically significant ($p<0.05$) laboratory differences for the 3T3 NRU test method for this reference substance.
[2]The ANOVA results indicate that there were statistically significant ($p<0.05$) laboratory differences for the NHK NRU test method for this reference substance.

Figure 3. Comparison of the 3T3 NRU assay IC_{50} values for the 72 reference substances from the validation study laboratories



Data points are the mean IC_{50} values (µg/mL) for each reference substance for each laboratory ± 1 standard deviation (SD). Data were sorted by 3T3 NRU IC_{50} values (ascending order) from Lab 3 (lead laboratory in the validation study). Reference substance identification is found in Table 2. The ANOVA results indicate that there were statistically significant ($p<0.05$) laboratory differences for the 3T3 NRU test method for the reference substances identified by the superscript in Table 2.

For the 3T3 NRU test method, the following laboratories/substances did not obtain sufficient IC_{50} data for the calculation of an intralaboratory CV: carbon tetrachloride at any laboratory; disulfoton at Lab 2; gibberellic acid at Lab 2; lithium carbonate at Lab 2 and Lab 3; methanol at any laboratory; 1,1,1-trichloroethane at Lab 1; valproic acid at Lab 3; and xylene at Lab 1 and Lab 2.

For the 3T3 NRU test method, the following substances did not obtain sufficient IC_{50} data for the calculation of an interlaboratory CV: carbon tetrachloride, lithium carbonate; methanol; and xylene.

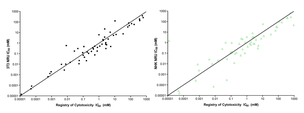
Figure 4. Comparison of the NHK NRU assay IC_{50} values for the 72 reference substances

Data points are the mean IC_{50} values (µg/mL) for each reference substance for each laboratory ± 1 standard deviation. Data were sorted using the same parameters for Fig. 3. Reference substance identification is found in Table 2. The ANOVA results indicate that there were statistically significant ($p<0.05$) laboratory differences for the 3T3 NRU test method for the reference substances identified by the superscript in Table 2.

For the NHK assay, the following laboratories/substances did not obtain sufficient IC_{50} data for the calculation of an intralaboratory CV: carbon tetrachloride at any laboratory; methanol at Lab 1; 1,1,1-trichloroethane at Lab 2 and Lab 3; and xylene at Lab 1 and Lab 2.

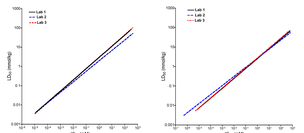
For the NHK assay, the following substances did not yield sufficient IC_{50} data for the calculation of an interlaboratory CV: carbon tetrachloride; 1,1,1-trichloroethane; and xylene.

Figure 5. Comparison of 3T3 and NHK NRU IC_{50} Values to RC IC_{50} Values



3T3 and NHK NRU values were geometric means of the mean values (for each reference substance tested) from each of three laboratories. The Registry of Cytotoxicity (RC) database contains acute oral LD_{50} values for rats and mice obtained from RTECS® and IC_{50} values from *in vitro* cytotoxicity assays using multiple cell lines and cytotoxicity endpoints for substances with known molecular weights. RC IC_{50} values were geometric means of various assays and cytotoxicity endpoints collected from the scientific literature (Halle 2003). Comparison of the mean IC_{50} values for the 58 (of 72) study reference substances common to the Registry of Cytotoxicity (RC) yielded a high correlation (r=0.955 for 3T3; r=0.824 for NHK).

Figure 6. The Laboratory-specific IC_{50}-LD_{50} Regressions for each NRU Test Method



A goodness of fit F-test indicated that the laboratory-specific regressions for both test methods were not significantly different ($p<0.05$) from one another ($p=0.796$ for the 3T3 NRU and $p=0.985$ for the NHK NRU). Thus, the prediction of LD_{50} by the laboratories would not be significantly different.

Table 3. Accuracy of GHS Toxicity Category Predictions by Laboratory[1]

	Laboratories	Total Reference Substances	Category Match	Toxicity Overpredicted	Toxicity Underpredicted
3T3	Lab 1	69	29%	41%	30%
	Lab 2	67	28%	43%	28%
	Lab 3	69	28%	41%	32%
NHK	Lab 1	69	28%	42%	30%
	Lab 2	69	28%	41%	32%
	Lab 3	70	29%	40%	31%

[1]GHS-Globally Harmonized System categories of acute oral toxicity with LD_{50} in mg/kg (UN 2003). 3T3 and NHK NRU test method IC_{50} data (geometric mean of within laboratory replicates) used with the RC regression: log(LD_{50} mmol/kg) = 0.425 x log(IC_{50} mM) +0.625.

GHS Category	LD_{50} range
1	$LD_{50} \leq 5$ mg/kg
2	$5 < LD_{50} \leq 50$ mg/kg
3	$50 < LD_{50} \leq 300$ mg/kg
4	$300 < LD_{50} \leq 2000$ mg/kg
5	$2000 < LD_{50} \leq 5000$ mg/kg
unclassified	$LD_{50} > 5000$ mg/kg

Conclusions

- For both the 3T3 and NHK NRU test methods, the IC_{50} geometric means calculated for the 58 (of 72) study reference substances common to the Registry of Cytotoxicity (RC) were strongly correlated with the IC_{50} values reported in the RC database. Comparison of the mean IC_{50} values yielded a high correlation (r=0.955 for 3T3; r=0.824 for NHK). (Figure 5).
- Positive Control (SLS) Reproducibility
  - The ANOVA results for SLS in the 3T3 NRU test method indicated that there were significant differences among laboratories ($p=0.006$) and among study phases within laboratories ($p=0.031$ for Lab 1 and $p=0.015$ for Lab 2) with the exception of Lab 3 ($p=0.854$).
  - The ANOVA results for SLS in the NHK NRU test method indicated also that there were significant differences among laboratories ($p<0.001$) and among study phases within laboratories ($p=0.031$ for Lab 1 and $p<0.001$ for Lab 2 and Lab 3).
  - Although the ANOVAs indicated significant differences between laboratories for SLS, Figures 1a and 1b show that the IC_{50} for SLS was consistent throughout the study except for changes in the NHK SLS IC_{50} at laboratory 1 due to changes in cell culture methods (i.e., reducing culture flask size and omitting fibronectin/collagen coating had a positive impact on reproducibility). Linear regression analysis of the IC_{50} over time indicates that, although the slopes were statistically significant, they were very small (Table 1).
  - A positive control should be tested throughout a study as a method for determining reproducibility.
- Intra- and Inter-laboratory Reproducibility for the Reference Substances
  - The CV analysis indicated that the mean CV for the reference substances was the same for both assays (mean CV = 26%), but interlaboratory reproducibility was better for the NHK assay (mean CV = 28% vs 46% for 3T3) (Figure 2).
  - For the 72 coded chemicals, 26 showed significant IC_{50} differences among laboratories for the 3T3 assay and 7 showed significant differences for the NHK assay.
  - Reproducibility among the laboratories was also demonstrated by the similarity of the laboratory specific regressions (Figure 6) and by the similar accuracy of the GHS toxicity category predictions (Table 3). The consistency of responses across laboratories (Figures 3 and 4) is well confirmed.

The results of these analyses and their relevance and impact are still under investigation.

References

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

EPA. 2002b. Health Effects Test Guidelines OPPTS 870.1100 Acute Oral Toxicity. EPA 712–C–02–190. Washington, DC: U.S. Environmental Protection Agency.

Fentem et al. 2001. A prevalidation study on in vitro tests for acute skin irritation: results and evaluation by the Management Team. Toxicology In Vitro. 15:57-93.

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

ICCVAM. 2001a. Report Of The International Workshop On In Vitro Methods For Assessing Acute Systemic Toxicity. NIH Publication No. 01-4499. NIEHS, Research Triangle Park, NC. http://iccvam.niehs.nih.gov/.

ICCVAM. 2001b. Guidance Document On Using In Vitro Data To Estimate In Vivo Starting Doses For Acute Toxicity. NIH Publication No. 01-4500. NIEHS, Research Triangle Park, NC. http://iccvam.niehs.nih.gov/.

ICCVAM. 2003. ICCVAM Guidelines for the Nomination and Submission of New, Revised, and Alternative Test Methods. NIH Publication No. 03-4508. NIEHS, Research Triangle Park, NC. http://iccvam.niehs.nih.gov/.

Institute of Medicine. 2004. Forging a Poison Prevention and Control System. Washington: National Academies Press.

OECD. 2001a. Guideline for Testing of Chemicals, 425, Acute Oral Toxicity – Up-and-Down Procedure. Paris France: OECD. http://www.oecd.org.

OECD. 2001b. Guideline For Testing of Chemicals, 423, Acute Oral Toxicity – Acute Toxic Class Method. Paris France: OECD. http://www.oecd.org.

Riddell et al. 1986. An evaluation of three *in vitro* cytotoxicity assays. Food Chem Toxicol 24: 469–471.

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

Spielmann et al. 1999. Determination of the starting dose for acute oral toxicity (LD50) testing in the up and down procedure (UDP) from cytotoxicity data. Altern Lab Anim **27**:957-966.

UN. 2005. Globally Harmonized System of Classification and Labelling of Chemicals (GHS). First Revised Edition. [ST/SG/AC.10/30/Rev.1]. United Nations, New York and Geneva.

Zuang et al. 2002. Follow-up to the ECVAM prevalidation study on in vitro tests for acute skin irritation. Altern Lab Anim 30:109-129.

Acknowledgments

This poster was supported by the Intramural Research Program of the NIH, National Institute of Environmental Health Sciences. ILS staff supported by NIEHS contract N01-ES 35504. The views expressed above do not necessarily represent the official positions of any federal agency.

Phase III Neutral Red Uptake Protocols for 3T3 and NHK cells are available at: http://iccvam.niehs.nih.gov/methods/invitro.htm