Results of the Final Phase of a Validation Study to Evaluate *In Vitro* Cytotoxicity Assays for Estimating Rodent Acute Systemic Toxicity

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

[1]NICEATM, RTP, NC, USA; [2]ILS, Inc, RTP, NC, USA; [3]ECVAM, JRC, Ispra, Italy; [4]IIVS, Gaithersburg, MD, USA; [5]US Army ECBC, APG, MD, USA; [6]Univ. of Nottingham, UK; [7]Consultant, Raleigh, NC, USA; [8]Constella Group, Durham, NC, USA; [9]BioReliance Corp, Rockville, MD, USA; [10]NIEHS/NIH/DHHS, RTP, NC, USA

Introduction

Acute oral toxicity testing in laboratory animals is used to characterize the hazard associated with human exposure to a substance. 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 toxicity testing (ICCVAM 2001a). Participants considered the use of *in vitro* cytotoxicity assays to predict acute *in vivo* lethality (Spielmann et al. 1999) to be sufficiently reliable that guidance should be published for using these assays to estimate starting doses for acute oral lethality tests. *Guidance Document on Using In Vitro Data to Estimate In Vivo Starting Doses for Acute Toxicity* (ICCVAM 2001b) provides details and examples on how to execute this approach.

The validation study organized by the National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM) and the European Centre for the Validation of Alternative Methods (ECVAM) implements the *Guidance Document* approach as well as another workshop recommendation to evaluate the ability of *in vitro* data from rodent and human cells to predict rodent *in vivo* LD50 data and human *in vivo* mortality data. This poster

- compares the results of the validation study to those of the Registry of Cytotoxicity (RC) (Halle 2003), on which the *Guidance Document* (ICCVAM 2001b) is based,
- reports the reliability and accuracy of the *in vitro* test methods, and
- reports the reduction of animal use for acute oral toxicity testing when using these *in vitro* methods to determine starting doses.

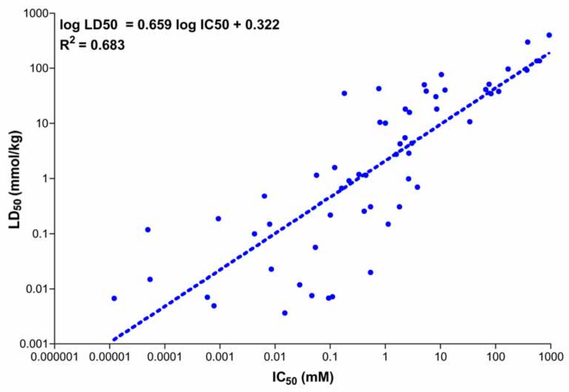

Figure 1. Regression between cytotoxicity (IC_{50x}) and rodent acute oral LD_{50} values for the 58 Registry of Cytotoxicity chemicals tested in this study.
The *Guidance Document* approach (ICCVAM 2001b) uses the RC regression, which is based on *in vitro – in vivo* data from 347 chemicals, to predict starting doses for acute oral toxicity assays from *in vitro* cytotoxicity data. The RC IC_{50x} values are geometric means of multiple endpoints and cell types while the LD_{50} values come largely from the 1983/84 Registry of Toxic Effects for Chemical Substances (Halle 2003). The RC regression is: log (LD_{50}) = 0.435 log (IC_{50}) + 0.625; R^2=0.452. The regression of the 58 chemicals (using RC data) in common with the RC was significantly different from the RC regression for 347 chemicals when slopes and intercepts were simultaneously compared ($p < 0.0001$).

Study Objectives

- To further standardize and optimize two *in vitro* cytotoxicity protocols in order to maximize intra- and inter-laboratory reproducibility.
- To assess the accuracy of two standardized *in vitro* cytotoxicity assays for estimating rodent oral LD_{50} values and human lethal concentrations across the five Globally Harmonized System (GHS; UN 2003) categories of acute oral toxicity as well as unclassified toxicities.
- To estimate the reduction and refinement (i.e., reduced deaths) in animal use that would result from using *in vitro* cytotoxicity assays to estimate starting doses for *in vivo* acute toxicity testing.
- To generate high quality *in vivo* and *in vitro* databases that can be used to support investigation of other test methods necessary to improve the accuracy of *in vitro* assessments of acute systemic toxicity.

Study Design

Test 72 reference chemicals in neutral red uptake (NRU) assays using mouse fibroblast (BALB/c) 3T3 cells and normal human keratinocytes (NHK).

Study Phases

Phase Ia: Laboratory Evaluation – Completed Nov 2002
Development of positive control database for each laboratory (N=3)

- Test positive control chemical (sodium laurel sulfate [SLS]) with each test method in at least 10 replicate NRU tests.
- Calculate mean $IC_{50} \pm 2$ standard deviations for each test method/lab for acceptance criteria for positive control performance in subsequent testing.

Phase Ib: Laboratory Evaluation – Completed May 2003
Limited chemical testing for possible protocol refinement

- Each lab tests the same three coded chemicals of varying toxicities three times with each assay.
- Refine protocols and repeat, if necessary, until acceptable reproducibility is achieved.

Phase II: Laboratory Qualification – Completed Nov 2003

- Each lab tests nine coded chemicals covering the range of GHS toxicity categories. Three replicate tests/chemical for each assay.
- Assure that corrective actions taken in Phase I have achieved the desired results.
- Further refine protocols and re-test if necessary to achieve acceptable results.
- Finalize protocols for Phase III.

Phase III: Laboratory Testing Phase – Completed Jan 2005

- Each lab tests 60 coded chemicals three times using the standardized and optimized protocol for each assay.

Results

Comparison to the Registry of Cytotoxicity

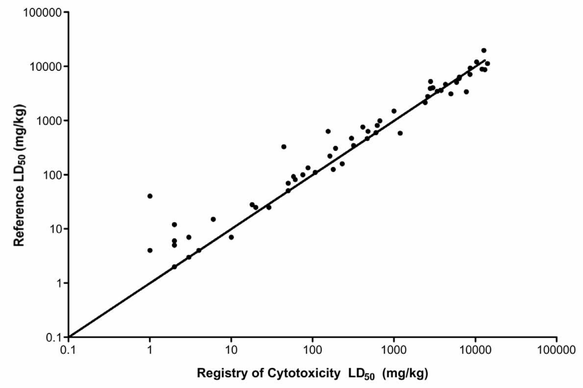

Figure 2. Comparison of reference LD_{50} values to RC LD_{50} values
Reference LD_{50} values were geometric means of acceptable LD_{50} values identified during a literature review (Paris et al. 2003). RC LD_{50} values were based largely on the 1983/84 Registry of Toxic Effects of Chemical Substances®.

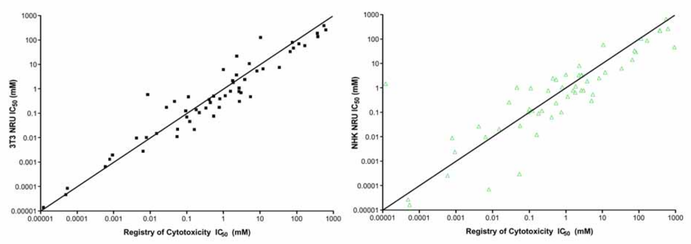

Figure 3. Comparison of 3T3 and NHK NRU IC_{50} values to RC IC_{50} values
3T3 and NHK NRU values were geometric means of the mean values (for each chemical tested) from each of three laboratories. RC IC_{50} values were geometric means of various assays and cytotoxicity endpoints collected from the scientific literature (Halle 2003).

Reliability

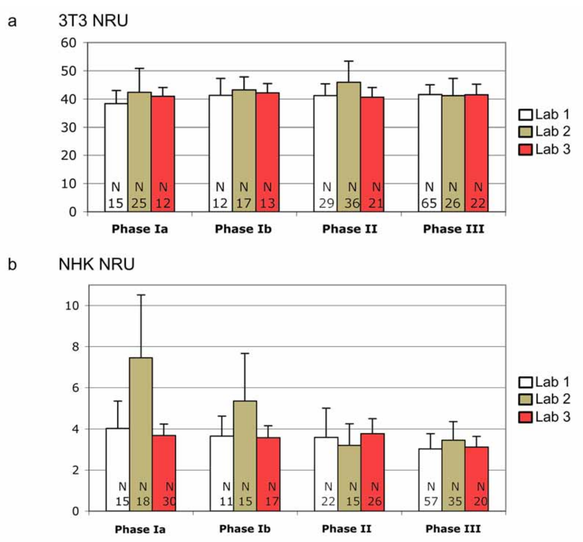

Figure 4. Variability of the positive control, SLS, IC_{50}
Bars show mean IC_{50} values from each lab. Error bars show standard deviation. **Fig. 4a** shows 3T3 NRU data look similar among the labs and phases, however, ANOVA indicated significant differences ($p < 0.05$) among labs with phases combined ($p=0.006$), but not among phases when labs were combined ($p=0.304$). **Fig. 4b** shows a change in NHK SLS IC_{50} at FAL between Phases Ib and II, which was due to a change in cell culture methods (i.e., decrease in culture flask size and omission of fibronectin/collagen coating). ANOVA showed significant differences ($p < 0.05$) among labs with phases combined ($p<0.001$), and among phases when labs were combined ($p<0.001$).

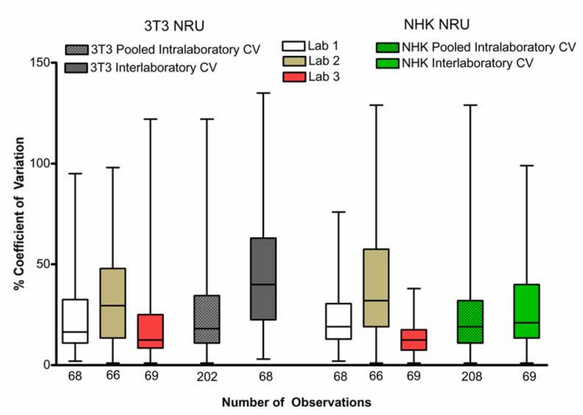

Figure 5. Reproducibility of 3T3 and NHK NRU by coefficient of variation analysis
%Coefficient of variation (CV) = standard deviation/mean X 100. Boxes show median, first quartile, and third quartile. Error bars show range. Total number of chemicals is less than 72 because of insufficient toxicity to produce an IC_{50} for some chemicals in some or all laboratories. Intralaboratory CV values are shown for each laboratory and each assay (pooled laboratory data). Mean intralaboratory CV = 26% for 3T3 and for NHK. Mean interlaboratory CV = 46% for 3T3 and 28% for NHK.

Accuracy

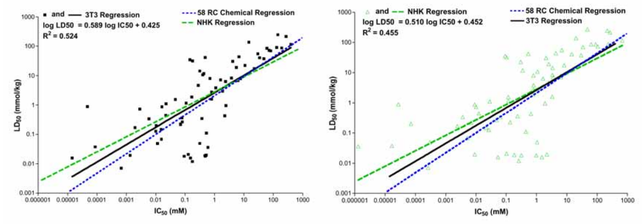

Figure 6. Linear regression analyses for 3T3 and NHK NRU test methods on regression of 58 RC chemicals. 3T3 and NHK NRU linear regressions use log IC_{50} results (geometric mean of 3 labs) with log LD_{50} reference values (Paris et al. 2003). Left panel shows 3T3 NRU results (70 chemicals) with NHK regression and 58 RC chemical regression (Fig. 1) while right panel shows NHK NRU results (71 chemicals) with 3T3 regression and 58 RC chemical regression. No laboratory achieved sufficient toxicity for calculation of an IC_{50} for carbon tetrachloride (both assays) or methanol (3T3 assay). Simultaneous comparison of slope and intercept showed that neither the 3T3 regression ($p = 0.929$) nor the NHK regression ($p = 0.144$) was different from the 58 RC chemical regression.

Table 1. Predictionof GHS Toxicity Category[1] by 3T3 and NHK NRU Regressions

Reference *In Vivo* LD_{50}[2]	3T3 NRU Predicted GHS Toxicity Category						Total	Accuracy[2]	Hazard Underpredicted	Hazard Overpredicted
	< 5	5 - 50	50 - 300	300 - 2000	2000 - 5000	> 5000				
< 5	0	2	3	2	0	0	7	0%	100%	0%
5 - 50	2	2	5	3	0	0	12	17%	67%	17%
50 - 300	0	2	7	3	0	0	12	58%	25%	17%
300 - 2000	1	0	6	9	0	0	16	56%	0%	44%
2000 - 5000	0	0	1	8	2	0	11[4]	18%	0%	82%
> 5000	0	0	2	4	2	4	12[5]	33%	0%	67%
Total	3	6	24	29	4	4	70	34%	26%	40%
Predictivity	0%	33%	29%	31%	50%	100%				
Category Underpredicted	0%	33%	33%	28%	0%	0%				
Category Overpredicted	100%	33%	38%	41%	50%	0%				
Reference *In Vivo* LD_{50}[2]	**NHK NRU Predicted GHS Toxicity Category**						**Total**	**Accuracy[2]**	**Hazard Underpredicted**	**Hazard Overpredicted**
	< 5	5 - 50	50 - 300	300 - 2000	2000 - 5000	> 5000				
< 5	0	1	2	4	0	0	7	0%	100%	0%
5 - 50	2	2	5	3	0	0	12	17%	67%	17%
50 - 300	0	3	5	4	0	0	12	42%	33%	25%
300 - 2000	1	0	4	11	0	0	16	69%	0%	31%
2000 - 5000	0	0	1	10	0	0	11[4]	0%	0%	100%
> 5000	0	0	2	5	6	0	13	0%	0%	100%
Total	3	6	19	37	6	0	71	25%	27%	48%
Predictivity	0%	33%	26%	30%	0%	0				
Category Underpredicted	0%	17%	37%	30%	0%	0				
Category Overpredicted	100%	50%	37%	41%	100%	0%				

[1]GHS-Globally Harmonized System categories for acute oral toxicity; LD_{50} in mg/kg (UN 2003). 3T3 and NHK regressions are shown in Fig. 6. Numbers in table represent number of chemicals.
[2]Accuracy is the proportion of correct outcomes for the NRU predictions of *in vivo* GHS toxicity category.
[3]*In vivo* reference LD_{50} values determined in this study through literature searches (Paris et al. 2003).
[4]Carbon tetrachloride excluded because no laboratory attained sufficient toxicity for the calculation of an IC_{50}.
[5]Methanol excluded because no laboratory attained sufficient toxicity in 3T3 assay for the calculation of an IC_{50}.

Table 2. Estimated Reduction of Animal Use for the Up-and Down Procedure[1] Using 3T3 and NHK NRU Assays to Predict Starting Doses

Assay/ Toxicity Category	Animals Used: Default Starting Dose	Standard Deviation	Cytotoxicity Predicted Starting Dose[2]	Standard Deviation	Animals Saved[3]
3T3					
< 5 mg/kg	11.2	2.3	10.1	2.9	1.1 (10%)
5-50 mg/kg	9.8	2.4	9.6	2.7	0.2 (2%)
50-300 mg/kg	7.8	2.0	8.3	2.3	-0.6 (-8%)
300-2000 mg/kg	8.6	2.4	8.3	2.3	0.4 (4%)
2000-5000 mg/kg	10.7	2.9	8.7	2.9	2.0 (19%)
> 5000 mg/kg	9.8	3.4	8.0	3.5	1.8 (19%)
NHK					
< 5 mg/kg	11.2	2.3	10.6	2.7	0.6 (5%)
5-50 mg/kg	9.7	2.4	9.9	2.8	-0.2 (-2%)
50-300 mg/kg	7.8	2.0	8.2	2.2	-0.5 (-6%)
300-2000 mg/kg	8.6	2.4	8.1	2.3	0.6 (6%)
2000-5000 mg/kg	9.1	2.9	10.7	2.9	1.7 (15%)
> 5000 mg/kg	9.8	3.3	8.2	3.4	1.6 (16%)

[1]Results produced by simulation modelling using reference LD_{50} values determined in this study as the true LD_{50}. Up-and-Down Procedure from OECD (2001) and EPA (2002).
[2]Using the 3T3 and NHK NRU regressions in Figure 6.
[3]Comparing default starting dose (175 mg/kg) with cytotoxicity predicted starting dose. Negative savings indicate more animals were used (compared with default).

Conclusions

- The IC_{50} and LD_{50} data collected for this study were similar to those used in the RC (Figs 2 and 3).
- The IC_{50} for the positive control, SLS, was consistent throughout the study except for changes in the NHK SLS IC_{50} at one lab due to change in cell culture methods (i.e., decrease in culture flask size and omission of fibronectin/collagen coating).
- Intralaboratory reproducibility for reference chemicals 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) (see Fig. 5).
- Judging by goodness of fit (i.e., R^2 values), 3T3 IC_{50} values had a better correlation with acute oral LD_{50} values (see Fig. 6).
- The 3T3 regression had higher accuracy than the NHK regression for GHS acute oral toxicity category predictions (34% vs. 25%) (see Table 1).
- For the chemicals tested, animal savings using the NRU methods for estimating starting doses for the Up–and–Down procedure for acute systemic toxicity were slightly greater for the 3T3 assay than for the NHK assay (see Table 2).
- A reliable database of *in vitro* and *in vivo* toxicity values has been established for 72 chemicals. These data can be used to investigate other test methods necessary to improve the accuracy of *in vitro* assessments of acute systemic toxicity for chemicals that poorly fit the *in vitro-in vivo* regressions.

References

Ekwall B, et al. 1998. ATLA 26:571-616
EPA. 2002. EPA 712–C–02–190.
Halle W. 2003. ATLA 31:89-198.
ICCVAM. 2001a. NIH Publication 01-4499. http://iccvam.niehs.nih.gov/
ICCVAM. 2001b. NIH publication 01-4500. http://iccvam.niehs.nih.gov/
Paris et al. 2003. Toxicologist 72:156.
OECD. 2001. TG 425.
Spielmann et al. 1999. ATLA 27: 957-966.
UN. 2003. http://www.unece.org/trans/danger/publi/ghs/officialtext.html

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

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