[This Page Intentionally Left Blank]

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

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

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

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

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

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

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

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

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

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

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

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

10.2 Reduction and Refinement of Animal Use for the UDP

10.2.1 *In Vivo* Testing Using the UDP

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

10.2.1.1 *Main Test*

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

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

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

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

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

10.2.1.2 *Limit Test*

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

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

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

10.2.2 Computer Simulation Modeling of the UDP

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

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

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

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

the results obtained when the LD_{50} that was estimated by the IC_{50} and the regression was used as the starting dose.

The simulation procedure used the following steps for each reference substance:

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

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

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

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

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

10.2.3.1 *The Effect of the Dose-Mortality Slope on Animal Use*

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

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

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

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

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

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

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

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

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

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

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

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

Tables 10-3 and **10-4** show mean animal use and mean animal savings for the UDP when the default starting dose and the IC_{50}-predicted starting doses were used, and when the reference substances are grouped by GHS category (UN 2005). The data come from the same analyses as the data provided in **Table 10-2**. The IC_{50}-predicted starting doses were based on the:

- RC rat-only millimole regression (**Table 10-3**)
- RC rat-only weight regression (**Table 10-4**)

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

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

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

These analyses showed that:

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

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

Table 10-3 shows the animal savings by GHS category when the IC_{50} values are used with the RC rat-only millimole regression. Mean animal savings were statistically significant ($p < 0.05$) by a one-tailed Wilcoxon signed rank test for the following GHS toxicity categories, test methods, and dose-mortality slopes:

- The use of the NHK NRU test method at both dose-mortality slopes for substances with $300 < LD_{50} \leq 2000$ mg/kg that produced savings of 0.49 (6.5%) to 0.52 (6.1%) animals per test.
- The use of the 3T3 NRU test method at the 8.3 dose-mortality slope for substances with $300 < LD_{50} \leq 2000$ mg/kg that produced a saving of 0.31 (4.1%) animals per test.
- The use of both *in vitro* NRU test methods at both dose-mortality slopes for substances with $2000 < LD_{50} \leq 5000$ mg/kg that produced savings of 1.11 (12.1%) to 1.28 (11.9%) animals per test.
- The use of both *in vitro* NRU test methods and both dose-mortality slopes for substances with an $LD_{50} > 5000$ mg/kg that produced savings of 1.47 (14.8%) to 1.58 (20.3%) animals per test.

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

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

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

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

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

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

[1]Mean numbers of animals used ±standard errors for 10,000 simulations for each substance with an upper limit dose of 5000 mg/kg. Although the simulations used whole animals, averaging the results over a large number of simulations produced fractional numbers. Results are provided for 67 substances in the 3T3 NRU test method and 68 substances in the NHK NRU test method. Substances were categorized using the rat acute oral LD_{50} reference values in mg/kg from **Table 4-2**.

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

[3]UN (2005).

[4]The RC rat-only millimole regression is $\log LD_{50}$ (mmol/kg) = 0.439 $\log IC_{50}$ (mM) + 0.621.

[5]Default starting dose = 175 mg/kg.

[6]The starting dose was one default dose lower than the predicted LD_{50} calculated using the IC_{50} value for each reference substance in the RC rat-only millimole regression. The IC_{50} value for each reference substance was randomly selected from the distribution of values obtained during the testing with each method.

[7]Difference between mean animal use with the default starting dose and mean animal use with the predicted starting dose.

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

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

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

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

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

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

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

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

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

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

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

[1]Mean number of animals used ±standard errors for 10,000 simulations for each substance with a limit dose of 5000 mg/kg. Although the simulations used whole animals, averaging the results over a large number of simulations produced fractional numbers. Results are provided for 67 substances for the 3T3 NRU test method and 68 substances for the NHK NRU test method categorized using the rat acute oral LD_{50} reference values in mg/kg from **Table 4-2**.

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

[3]UN (2005).

[4]The RC rat-only weight regression is $\log LD_{50}$ (mg/kg) = 0.372 $\log IC_{50}$ (µg/mL) + 2.024

[5]Default starting dose = 175 mg/kg.

[6]The starting dose was one default dose lower than the predicted LD_{50} calculated using the IC_{50} values for each reference substance in the RC rat-only weight regression. The IC_{50} value for each reference substance was randomly selected from the distribution of values obtained during the testing with each method.

[7]Difference between mean animal use with the default starting dose and mean animal use with the predicted starting dose.

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

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

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

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

Abbreviations: 3T3=BALB/c 3T3 fibroblasts; NHK=Normal human epidermal keratinocytes; RC=Registry of Cytotoxicity; UDP=Up-and-Down Procedure.

[1]Numbers are mean numbers of animals used for 10,000 simulations for each substance. Although the simulations used whole animals, averaging the results over a large number of simulations produced fractional numbers. Upper limit dose = 5000 mg/kg. Results are provided for 67 substances in the 3T3 NRU test method and 68 substances in the NHK NRU test methods.

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

[3]Default starting dose = 175 mg/kg.

[4]The starting dose was one default dose lower than the predicted LD_{50} calculated using the IC_{50} value in the regression specified. The IC_{50} value for each reference substance was randomly selected from the distribution of values obtained during the testing with each method.

[5]$\log LD_{50}$ (mmol/kg) = 0.439 $\log IC_{50}$ (mM) + 0.621.

[6]$\log LD_{50}$ (mg/kg) = 0.372 $\log IC_{50}$ (μg/mL) + 2.024.

10.2.5 Accuracy of UDP Outcomes Using the IC_{50}-Based Starting Doses

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

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

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

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

10.3 Reduction and Refinement of Animal Use in the ATC Method

10.3.1 *In Vivo* Testing Using the ATC Method

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

10.3.1.1 *Main Test*

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

10.3.1.2 *Limit Test*

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

10.3.2 Computer Simulation Modeling of the ATC Method

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

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

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

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

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

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

The simulation procedure used the following steps for each substance:

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

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

10.3.3.1 *The Effect of the Dose-Mortality Slope on Animal Use*

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

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

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

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

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

Abbreviations: 3T3=BALB/c 3T3 fibroblasts; ATC=Acute Toxic Class method; NHK=Normal human epidermal keratinocytes; RC=Registry of Cytotoxicity.

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

[1]Mean numbers of animals used ± standard errors for 2000 simulations each for 67 substances in the 3T3 NRU test method and 68 substances in the NHK NRU test method. Although the simulations used whole animals, averaging the results over a large number of simulations produced fractional numbers. Upper limit dose = 2000 mg/kg.

[2]OECD (2001d).

[3]Default starting dose = 300 mg/kg.

[4]Next fixed dose lower than the predicted LD_{50} calculated using the IC_{50} value for each reference substance in the RC rat-only millimole regression: $\log LD_{50}$ (mmol/kg) = 0.439 $\log IC_{50}$ (mM) + 0.621. The IC_{50} value for each reference substance was randomly selected from the distribution of values obtained during the testing with each method.

[5]Difference between mean animal use with the default starting dose and mean animal use with the IC_{50}-based starting dose.

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

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

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

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

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

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

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

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

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

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

The analyses showed that:

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

Animal Savings in the ATC Method by GHS Category Using the 3T3- and NHK-Based Starting Doses with the RC Rat-Only Millimole Regression

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

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

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

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

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

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

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

[1]Mean number of animals used ±standard errors for 2000 simulations for each substance with an upper limit dose of 2000 mg/kg. Results are provided for 67 substances in the 3T3 NRU test method and 68 substances in the NHK NRU test method categorized using the rat acute oral LD_{50} reference values in mg/kg from **Table 4-2**. Although the simulations used whole animals, averaging the results over a large number of simulations produced fractional numbers.

[2]OECD (2001d).

[3]GHS for acute oral toxicity (UN 2005).

[4]The RC rat-only millimole regression is $\log LD_{50}$ (mmol/kg) = 0.439 $\log IC_{50}$ (mM) + 0.621.

[5]Default starting dose =300 mg/kg.

[6]The starting dose was the next fixed dose lower than the predicted LD_{50} using the IC_{50} for each reference substance in the RC rat-only millimole regression. The IC_{50} value for each reference substance was randomly selected from the distribution of values obtained during the testing with each method.

[7]Difference between mean animal use with the default starting dose and mean animal use with the IC_{50}-based starting dose.

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

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

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

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

Table 10-9 shows the animal savings for the simulated ATC method by GHS category for the *in vitro* NRU methods used with the RC rat-only weight regression. Mean animal savings were statistically significant ($p < 0.05$) by a one-tailed Wilcoxon signed rank test for the following GHS toxicity categories, NRU test methods, and dose-mortality slopes.

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

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

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

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

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

[1]Mean number of animals used ±standard errors for 2000 simulations for each substance with an upper limit dose of 2000 mg/kg. Although the simulations used whole animals, averaging the results over a large number of simulations produced fractional numbers. Results are provided for 67 substances in the 3T3 NRU test method and 68 substances in the NHK NRU test method categorized using the rat acute oral reference LD_{50} values in mg/kg from **Table 4-2**.

[2]OECD (2001d).

[3]GHS for acute oral toxicity (UN 2005).

[4]From **Table 6-2;** $\log LD_{50}$ (mg/kg) = 0.372 $\log IC_{50}$ (μg/mL) + 2.024

[5]Default starting dose = 300 mg/kg.

[6]The starting dose was one fixed dose lower than the predicted LD_{50} calculated using the IC_{50} for each reference substance in the RC rat-only weight regression. The IC_{50} value for each reference substance was randomly selected from the distribution of values obtained during the testing with each method.

[7]Difference between mean animal use with the default starting dose and mean animal use with the IC_{50}-based starting dose.

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

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

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

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

10.3.4 Refinement of Animal Use in the ATC Method When Using 3T3- and NHK-Based Starting Doses

A procedure refines animal use when it lessens or eliminates pain or distress in animals, or enhances animal well-being (ICCVAM 2003). This section evaluates whether the use of 3T3- and NHK-based starting doses refines animal use by reducing the number of animals that die

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

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

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

Abbreviations: 3T3=BALB/c 3T3 fibroblasts; ATC=Acute Toxic Class method; NHK=Normal human epidermal keratinocytes; RC=Registry of Cytotoxicity.

[1]Mean numbers of animals used for 2000 simulations for each of 67 substances in the 3T3 NRU test method and 68 substances in the NHK NRU test method. Although the simulations used whole animals, averaging the results over a large number of simulations produced fractional numbers. Upper limit dose =2000 mg/kg.

[2]OECD (2001d).

[3]Default starting dose =300 mg/kg.

[4]The starting dose was one fixed dose lower than the predicted LD_{50} calculated by using the IC_{50} for each reference substance in the regression evaluated. The IC_{50} value for each reference substance was randomly selected from the distribution of values obtained during the testing with each method.

[5]$\log LD_{50}$ (mmol/kg) = 0.439 $\log IC_{50}$ (mM) + 0.621.

[6]$\log LD_{50}$ (mg/kg) = 0.372 $\log IC_{50}$ (μg/mL) + 2.024.

10.3.5 Accuracy of the ATC Method Outcomes Using the IC_{50}-Based Starting Doses

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

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

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

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

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

10.4 The Impact of Accuracy on Animal Savings

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

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

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

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

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

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

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

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

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

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

10.5 The Impact of Prevalence on Animal Savings

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

10.6 Summary

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

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

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

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

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

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

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

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