Model Validation in Acute Aquatic Toxicity Testing[1]

L.S. McCarty

L.S. McCarty Scientific Research & Consulting, Newmarket, Ontario, Canada.

Financial Support from the European Chemical Industry Council (CEFIC) Long-Range Initiative Project ECO-16 is gratefully acknowledged

Abstract

A quality control evaluation of the 96 h acute toxicity testing protocol focused on 3 key assumptions: 1, it must be possible to estimate steady-state LC50s; 2, LC50s must occur at an equivalent exposure duration; 3, all substantive toxicity modifying factors should be adequately controlled.

About 8% of the tests failed assumption 1. Examination of remaining data indicated remaining variance from unquantified effects of toxicity modifying factors, thereby failing assumption 3. Such flaws in results generated via recommended LC50 testing protocols means test data do not represent consistent, comparable measures of relative toxicity.

Current regulations are acceptable due to the use of semiquantitative, policy-driven development guidance that considers such data uncertainty. Quantitative applications such as QSARs, mixture toxicity, and regulatory chemical grouping are compromised.

Failures to validate key assumptions justify a formal QC review of the LC50 and related toxicity testing protocols. Interim improvements in design, execution, interpretation, and regulatory applications of testing using exposure-based dose surrogates are warranted.

Background

The key assumption underlying LC50 toxicity and related exposure-based toxicity tests is well-known:

"The incipient LC50 (lethal concentration for 50 percent of individuals on long exposure) is recommended as the most useful single criterion of toxicity."(Wuhrmann, 1952, from Sprague, 1969)

The LC50 conceptual model is: several constant exposure concentrations compensate for influences of various physical, chemical, and biological modifying factors, such that levels lie in a range where 50% mortality can be interpolated. At steady-state the LC50 concentration is a surrogate dose for unknown organism concentrations. In turn it is a surrogate for concentrations at the unknown site(s) of toxic action in the organism. If all toxicity modifying factors are accounted for LC50 differences are due to toxicodynamic differences.

The technical model for toxicokinetics is the one-compartment, first-order kinetics model widely used in bioconcentration. No formal toxicodynamic technical model is defined. Despite various limitations and caveats, this approach has been used, largely unchanged, for more than 50 years.

Materials and Methods

The widely used fathead minnow acute toxicity database (e.g., *Russom et al.,* 1997) was employed as results are reported in detail. A subset of 370 tests on 325 chemicals was selected for organic chemicals from diverse chemical groups.

Steady-state LC50 estimates and elimination rate constants (k_2) were derived using non-linear curve fitting of time and inverse toxicity data (i.e., 1/LC50) to a 1CFOK model (McCarty *et al.,* 1992). Where there were curve-fitting difficulties, but steady-state was evident by inspection, only LC50s were reported.

These results were used to test the validity of three assumptions key to determining whether conceptual and technical toxicity testing models are appropriate and sufficient for test results to be considered consistent, comparable measures of relative toxicity:

1. The minimum basis for relative toxicity comparisons is that an effective surrogate relationship among water, whole organism, and internal target doses is achieved for all tests. The validation test is establishing a steady-state LC50 estimate.

2. Optimal comparability between tests is ensured by an equivalent exposure duration for all tests. The validation test is that steady-state LC50s are reached at an equivalent exposure duration. Herein it is assumed the toxicokinetic phase dominates and the toxicodynamic response after the effective organism dose is achieved is rapid.

3. A catchall assumption is that the LC50 test protocol addresses and adequately controls all toxicity modifying factors. The validation test is that modifying factor influences are not detected in testing results.

Results and Discussion – 1

The results LC50 analyses are summarized in Table 1. For Group 1 (No SS-No K2) 61 tests on 58 chemicals (61 of 736 tests, 8%) failed validation of assumption 1. Of these chemicals 7% (43 of 604) have no replicate tests and no valid LC50 is present in the FHM database. See Table 2.

LC50s that fail validation of assumption 1 should be treated the same as results that fail statistical model validation: they are invalid and unusable.

Table 1. Steady-State LC50 and Time to Steady-State Status for 370 Tests with 325 Chemicals

Estimation Group	Number of Tests	Comments
1. No SS-No K2 Steady-state: No Kinetics: No	61 tests 58 chemicals	Insufficient data for SS LC50 determination k2: cannot be estimated Tss: cannot be estimated
2. SS-K2 Steady-state: Yes Kinetics: Yes	209 tests 172 chemicals	Steady-state LC50: 0.000011-918 mmol/kg k2: 0.0034 to 6.5 h^{-1} Tss: 0.424 to 815 h
3. SS-No K2 Steady-state: Yes Kinetics: No	100 tests 95 chemicals	Steady-state LC50: 0.00025-709 mmol/kg k2: cannot be estimated Tss: cannot be estimated

Table 2. 96h LC50 Tests Where Steady-State Assumption Validation Failed (Group 1)

*(+-)-sec-Butylamine	*4-(tert-butyl)Benzamide (para)	*Malononitrile
*1-(Carboxymethyl) Pyridinium Chloride	*4-Acethpyridine	*Methyl methacrylate
*1,4-bis(3-aminopropyl) piperazine	*4-Actamidophenol	*N-Vinylcarbazole
*1,4-Dioxane #1 & 2	*4-Bromoaniline (para)	*Tert-Butylstyrene
*2-(Ethylamino) Ethanol	*4-Chloroaniline	1,1,1-Trichloroethane #1
*2-Acetyl-1-Methylpyrrole	*4-Decylaniline	2-Chloroaniline #2
*2-Chloro-6-Methylbenzonitrile	*4-Methoxyphenol	2-Chlorophenol #1
*2,2-Dichloroacetamide	*4-Nitrobenzaldehyde	2-Fluorotoluene #1
*2,3-Dimethyl-1,3-Butadiene	*5-Bromovanillin	2-Hydroxypropyl Acrylate #2
*2,3-Dimethylvaleraldehyde	*Acetone oxime	2-Propyn-1-ol #2
*2,4-Dimethylphenol	*Acrylamide	2,4,6-Trichlorophenol #1
*2,5-Dinitrophenol	*Allyl isothiocyanate	4-Chloro-3-Methyl Phenol #2
*2,6-Dichlorobenzamide	*Benzoic Acid, Sodium Salt	4-Fluorophenyl Ether #1
*3-(4-tert-Butylphenoxy) Benzaldehyde	*Dicofol	a,a,a-Trifluoro-M-Tolualdehyde #3
*3-Chloro-1-Propanol	*Di-N-Butylterephthalate	Chlorpyrifos #2
*3-Methoxyphenol	*Ethyl Hexanoate	N,N-Diphenylformamide #1
*3-Picoline	*Hexachloro-1,3-Butadiene	Neoabietic Acid #1
*3,3-Dimethylglutaric acid	*Isophorone	Phenol #1, 3, 4
*3'-Aminoacetophenone (meta)	*Isopimaric Acid	

*** indicates no replicate**

Figure 1: T_{ss} & Steady-State LC50s for FHM Groups 2 & 3

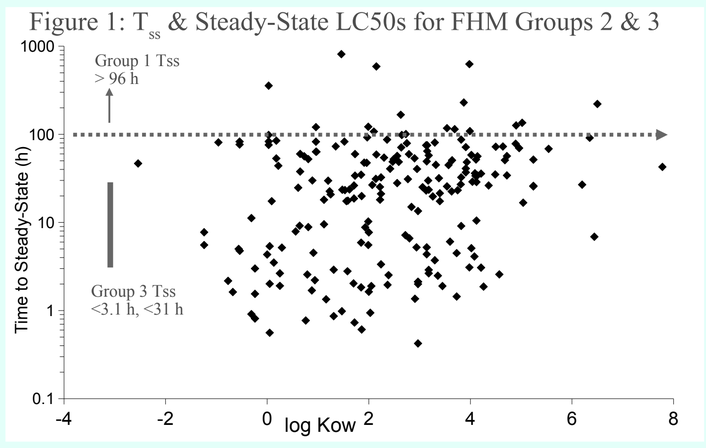
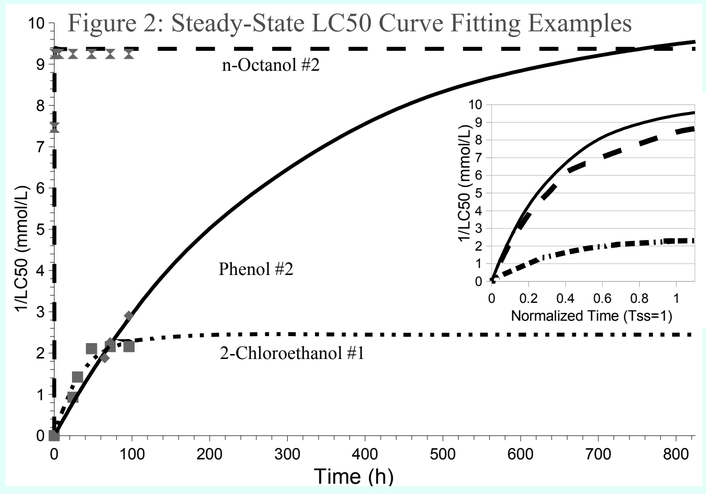


Figure 2: Steady-State LC50 Curve Fitting Examples



Groups 2 & 3 appear to meet Assumption 2 as Tss varies ~1900x, but steady-state is reached. Apparent differences in exposure duration are related to use of clock-based time rather than normalized time (i.e., Tss= 4 * T½). Only at steady-state are assumptions validated and LC50s based on comparable exposures.

Results and Discussion – 2

Despite meeting assumptions 1 & 2, additional analysis indicated residual, uncontrolled/uncorrected influences of some toxicity modifying factors:

- exposure durations less/more than steady-state
- body size variations
- metabolic biotransformation
- different modes of toxic action

Influences were ~ 3.5x, 4x, 3x for toxicokinetic (total is ~10x) and at least 10x for toxicodynamic factors. This means that LC50 results are not consistent comparable measures of relative toxicity; assumption 3 validation fails for the entire fathead minnow database.

Noisiness in inter/intra laboratory testing has been attributed to unexplained sources; this work indicates design and implementation deficiencies in approved testing protocols contribute significantly. Other modifying factors are likely involved but evaluation is hampered by minimalistic reporting requirements.

Most aquatic toxicity tests use the same conceptual model, differing in exposure character/duration and response endpoints. If data that fail or cannot undergo assumption validation are excluded as potentially flawed, little toxicity testing data will be available for theoretical or regulatory applications.

Regulatory Consequences

Data Quality Assessment principles indicate that even poor data can be useful where inadequacies are considered in decision-making policy (U.S. EPA, 2006). Thus, existing aquatic regulations developed by case-by-case approach are adequate as protocols are semi-quantitive, considering both data quality and quantity.

However, QSARs, mixture toxicity, and grouping by chemical similarity or use, as well as targeted risk reduction within groupings, are quantitative in nature, requiring precise, accurate toxicity metrics. Such approaches, components of new regulatory approaches in REACH, CEPA, and HPV Challenge, are compromised by such data flaws. Caution is warranted as, without correction/consideration of deficiencies, outcomes may be erroneous.

Advances in toxicological test modelling and interpretation (e.g., Jager *et al.,* 2011) should be employed to revise standard testing protocols. Perhaps the most important change that can be immediately instituted is that explicit model description, routine assumption validation, and detailed data reporting be mandatory for the conduct of new aquatic toxicity testing and considered in evaluation of existing data.

Conclusions and Recommendations

About 8% of the toxicity tests failed assumption 1 and are invalid. The remainder failed assumption 3 due to variance of ~10-20x from unquantified effects of toxicity modifying factors embedded in LC50 estimates. **Thus, standard LC50 results are not consistent, comparable measures of relative toxicity**.

Most currently available aquatic toxicity data were generated with a similar conceptual model and are also flawed, but inadequate reporting precludes data quality validation.

Current regulations developed with available testing data are acceptable as policy-driven development guidance addresses such data uncertainty. However, quantitative applications such as QSARs, mixture toxicity, and regulatory grouping by chemical characteristics or uses will be compromised.

Assumption validation failures justify a formal QC examination of LC50 and related testing protocols. Interim improvements in design, execution, interpretation, and regulatory applications of LC50 and related protocols are warranted.

References

Jager, T, Albert, C, Preuss, TG, Ashauer, R, 2011. General Unified Threshold Model of Survival - a toxicokinetic-toxicodynamic framework for ecotoxicology. Environ. Sci. Technol. 45:2529-2540.

McCarty, LS, Mackay, D, Smith, AD, Ozburn, GW, Dixon, DG, 1992. Residue-based interpretation of toxicity and bioconcentration QSARs from aquatic bioassays: neutral narcotic organics. Environ. Toxicol. Chem. 11:917-930.

Russom, CL, Bradbury, SP, Broderius, SJ, Hammermeister, DE, Drummond, RA, 1997. Predicting modes of toxic action from chemical structure: Acute toxicity in the fathead minnow (*Pimephales promelas*). Environ. Toxicol. Chem. 16:948-967.

Sprague, JB, 1969. Measurement of pollutant toxicity to fish I. Bioassay methods for acute toxicity. Water Res. 3:793-821.

US EPA, 2006. Data Quality Assessment: A Reviewer's Guide. EPA QA/G9R. U.S. Environmental Protection Agency, Office of Environmental Information, Washington DC.

[1] McCarty, L.S., 2012, in press. Model Validation in Aquatic Toxicity Testing: Implications for Regulatory Practice. Regul. Toxicol. Pharmacol. DOI: 10.1016/j.yrtph.2012.04.009