2

Multiple Endpoints and Integrated Test Strategies

Neil Krivanek and Rick Corley

ABSTRACT: *The basic questions of what toxic endpoints test for and the appropriate time to conduct these tests has remained an issue for the regulatory and toxicology community since the creation of TSCA. In order to maximize the return of multiple endpoint screening tests, we have examined several important components of these tests. Initial selection of compounds for testing should consider potential human exposure and production volumes.*

Simple acute studies in rodents should precede multi-dose studies. For multiple endpoint screening properly designed 28-day exposure studies can generate data for most endpoints. We have examined some of the boundaries which impact quality and interpretation of test data.

This paper will discuss some key interrelated items in toxicology screening testing for multiple endpoints that are pertinent to this OTA workshop on the evaluation of the existing chemicals review program which is administered by the EPA.

The scope of this paper is first to describe the current state for evaluating multiple toxicity endpoint in mammalian systems and secondly describe the important issues we see that need to be addressed in screening tests for existing chemicals under TSCA.

Screens, by their nature, involve a series of trade-offs or compromises. They need to be sensitive enough to identify subtle hazards, selective enough to minimize false positives and negatives yet be manageable enough (cost and time) to evaluate large numbers of materials. For the purposes of this paper, a screen can consist of either a battery of studies or a single study which incorporates numerous endpoints that can be completed within a reasonable period of time (i.e. less than one year or budget cycle) at a reasonable cost per chemical. Such an exploratory screen should provide sufficient information to identify potential systems affected (e.g. respiratory, cardiac, digestive, reproductive, neurological, immunological, etc.), its severity and dose-responsiveness to make a preliminary assessment of potential risk (e.g. margins of safety). Appropriately designed screens may also be used to direct further research to either more fully characterize an effect or to establish the relevancy of an effect to exposed populations, human or otherwise.

The focus on this manuscript is on the initial assessment (screen) for potential toxicity to a variety of organ systems as the first in a possible tiered evaluation.

The main question which needs to be consistently asked is "what" endpoints to test for and "when" to conduct these tests. For industrial chemicals, including those which come under TSCA, the goals are to safely manufacture, use, and dispose of these materials. These toxicity data are used to develop programs for safe handling and use, for occupational exposure limits, appropriate warnings for use, and appropriate information on Material Safety Data Sheets (MSDS's).

TESTING TRIGGERS

What we would like to discuss is what toxicity information triggers certain types of testing. To date most testing has been done on a case-by-case basis. The types of tests have been largely triggered by the toxicity of the chemical and by structure-activity relationships. When to test has generally been driven by a combination of potential exposures and production volume, with em-

phasis on exposure. However within the last few years, strategies have been emerging in several parts of the world, notably Europe, which link a certain amount of testing with a certain volume of production of the chemical without an assessment of exposure. This particular issue of what to test for and when to test has also been controversial since the implementation of TSCA.

To illustrate, I will list two examples of priority setting for testing. This is shown in table 2-1. This is a comparison of the Screening Information Data Set (SIDS) from Organization for Economic Cooperation and Development (OECD) and the development of a criteria document by National Institute of Occupational Safety and Health (NIOSH). This side by side comparison shows that each contains the same basic elements of the initial exposure assessment. This is part of the first step to decide what is important for prioritization. Both efforts emphasize chemical use patterns, type and extent of workplace exposure, and number of workers exposed.

The concept of a production volume trigger for defining testing strategies is shown in table 2-2 and 2-3. In table 2-2, for 1 ton/year with and without a 5 ton cumulative trigger the information required focuses on physical chemical properties, acute animal and aquatic testing, and limited repeated dose studies in animals for up to 28 days. In table 2-3, for volumes of greater than 100 tons/year to less than 1000 tons/year with 500 tons cumulative, testing is more comprehensive.

Table 2-1: Comparison between OECD SIDS and U.S. NIOSH Data Requirements for Chemicals-in-Commerce

SIDS **Initial exposure assessment**	**NIOSH** ***Criteria documents &*** **preliminary prioritizations**
Identification of occupational or professional uses of the chemical or products in which it is contained, and of use in consumer products	Frequency of use (occupations, processes, number of industries)
Uses in consumer products	
Function of chemical (for each consumer product identified)	
Weight fraction of chemical (actual or recommended)	
Form or product (as marketed, e.g., aerosol, powder, liquid)	
Workplace exposure, frequency and duration of such exposure	Extent of exposure and background exposures
Number of workers (in range of situations including manufacture, maintenance, and use)	Number of workers exposed
Quantities per media (time dimension of release, type of release, and uncertainties in estimates)	Substitution; interactions
An indication of measured exposure levels (expressed in an appropriate statistical form, e.g., geometric mean and standard deviation)	Technical feasibility of controls, quality of available data, severity and type of adverse health effects
An overview of monitoring data in the environment (with specifications of conditions)	Availability of sampling and analytical methods
Any additional information that will help to focus the exposure assessment	Other considerations

Table 2-2: Overview of Base Set Testing Requirements: 6th Amendment vs. 7th Amendment to the Dangerous Substances Directive

6th Amendment Annex VII (trigger: 1 ton/year)	7th Amendment Annex VII, Part A (trigger: 1 ton/year or 5 tons cumulative)
UV, IR, NMR spectra	UV, IR, NMR spectra
Methods of detection	Methods of detection
Melting point	Melting point
Boiling point	Boiling point
Relative density	Relative density
Vapor pressure	Vapor pressure
Surface tension	Surface tension
Water volubility	Water volubility
Fat volubility	—
$P K_{ow}$	**PK_{ow}**
Flash point	Flash point
Flammability	Flammability
Explosive properties	Explosive properties
Auto-flammability	Self-ignition temperature
Oxidizing properties	Oxidizing properties
—	*Granulometry (particle size distribution)*
Acute oral LD_{50}	Acute oral LD_{50}
Acute inhalation LC_{50} or acute cutaneous LD_{50}	Acute inhalation LC_{50} or acute cutaneous LD_{50}
Skin irritation	Skin irritation
Eye irritation	Eye irritation
Skin sensitization	Skin sensitization
28-day sub acute study	Repeated dose toxicity (28-days)
Ames assay	Mutagenicity, bacterial (reverse mutation) test
Non-bacterial mutagenicity	Chromosomal aberration or damage
	Toxicokinetic behavior assessment
—	*Reproductive screening test*
Fish acute toxicity	Fish acute toxicity
Daphnia acute toxicity	Daphnia acute toxicity
—	*Algal growth inhibition*
	Bacterial inhibition
Biodegradation	Biodegradation
Abiotic degradation	Abiotic degradation
—	*Adsorption/desorption screening test*

Developmental/reproductive endpoints, extra mutagenicity, toxicokinetics, and environmental repeated dose and bioaccumulation are added.

For volumes greater than 1000 tons/year 5000 cumulative, Level 2 of testing is activated (table 2-4). This includes chronic effects (including carcinogenicity, second species developmental, further toxicokinetics, organ specificity, and other species in the ecological sphere such as birds and other fishes. Besides production volumes, it is important to consider the potential for human exposure when deciding on testing strategy. Although specific quantitation of this aspect has not been incorporated into a specific regulation, it

Table 2-3: Additional Requirements with 100 to 1,000 Tons Annual Production and 500 Tons Cumulative Production

Level 1 Attachment 8 (Complementary tests)
• Further physical & chemical testing
• Fertility study (one or two generation)
• Teratogenesis (one species)
• Subchronic and or chronic toxicity study
• Additional mutagenesis
• Screening for carcinogenesis
• Toxicokinetics
• Prolonged toxicity to Daphnia
• Phytotoxicity (higher plant)
• Bioaccumulation (preferably fish)
• Further degradation tests (if poor degradation found)
• Further absomtion/resorption

Table 2-4: Additional Requirements with Greater than 1,000 Tons Annual Production or 5000 Tons Cumulative Production

Level 2 Attachment 8 (Complementary tests)
• Fertility (3 generation)
• Chronic toxicity
• Carcinogenicity
• Teratogenesis (2nd species)
• Further toxicokinetics
• Complementary studies to determine organ toxicity
• Complementary tests on accumulation, degradation mobility and absorption/desorption
• Complementary tests on fish
• Toxicity to birds
• Complementary tests on other organisms

Table 2-5: Corollaries of Descriptive Toxicity Testing

- Tests are designed to *characterize* toxic effects
- Tests are not designed to *demonstrate safety*
- There are no set of tests that have to be conducted on *every* chemical-in-commerce
- Tests are dependent upon:
  - Use
  - Potential exposure
  - Chemical/physical properties
  - Structural analogs

Table 2-6: Components of Descriptive Toxicity Study

- Relevant route of administration
- Primarily conducted with rats or mice
- Typically three or more dose levels, plus controls
- Includes a batter of clinical observations
- Clinical pathology
- Necropsy
- Microscopic evaluation of tissues
- May include other special tests

should not be disregarded. For example, a large volume chemical used only as a site limited intermediate has limited potential for human exposure, while a low volume chemical which becomes a component of a consumer product has a significant potential for human exposure and, therefore, a likely different testing strategy.

ANIMAL TESTING

Before addressing some of the issues pertaining to testing strategy, I would like to describe the simple, acute toxicity testing in mammals (table 2-5).

For the simple descriptive tests, there are basic components (table 2-6). Over the years there have been clear improvements. The first one is the characterization of the test material. For pure compounds purity can be in the 99+ % range. For technical grade material, the amount and content of each component can be determined to very low levels such as ppm. In general, impurities at >0.1% must be identified. Another important feature is stability of the material, knowing that composition doesn't change from time of manufacturer to completion of tests is critical.

The test species that are used: rats, mice, guinea pigs, rabbits and dogs have become more consistent; one can select a particular strain of animal and generally they remain healthy during the course of the study. Similar upgrades with emphasis on consistency and quality are found in animal food, water, housing, and lack of disease.

Table 2-7: Subacute Tests

- Range finding for subchronic studies
  - Potential target organs
  - Palatability/application limits
  - Oral, dermal, and inhalation
- Primarily in the rat, mouse, (rabbit, dog)
- 5/sex/dose, 3 dose levels plus control
- Limited clinical pathology
- Gross and partial histopathology
  - Liver, kidney, lung, skin, and gross lesions

Table 2-8: Subchronic Tests

- Most complete "short-term" studies
- Primarily in the rat, mouse, rabbit, and dog
- Range-finding for chronic studies
  - Identifies target organs
  - Oral, dermal, and inhalation
- Rodents: 10/sex/dose, 3 dose levels plus controls
  - Recovery groups
- Full in-life clinical observations
- Full clinical pathology
- Gross and full histopathology (top dose versus control)

For acute toxicity testing data development there have been improvements as well. Two examples are worth mentioning. In the classic evaluation for lethal effects, the LD50 and LC50 are considered as the basic values. However, this exact value is being superseded by the Approximate Lethal Dose (ALD) and Approximate Lethal Concentration (ALC). These latter two values provide enough information for most determinations and the fine tuning to an LD50 has become less important (2). Also ALDs and ALCs use fewer animals thus addressing some animal welfare concerns. Further development in this area has been the use of a limit dose test, which is aimed at getting basic information. OECD recommends a dose of at least 2,000 mg/kg; if mortality is observed a full study could be done (5). Lethality data remains important, but this starting point may not need to be so well defined.

The second example is the determination of irritation. In the classic paradigm, there was dermal testing in rabbits and eye testing in rabbits. Dermal testing remains generally unchanged, but significant changes in eye testing strategy have been made. The dermal response in rabbit is used to determine whether or not an eye test will be conducted. Even when eye tests are conducted, less material is used, 0.1 ml vs. 1 ml amounts. Furthermore, if the pH is less than 2 or greater than 11.5 eye testing is not performed, because it would be painful to the animal and results, based upon past experiences, would most likely show the material to be at least a severe irritant.

Beyond acute studies other endpoints are only addressed with multi-dose and exposure studies.

The subacute and subchronic tests have not changed format significantly (tables 2-7 and 2-8) Their main use is to determine multiple exposure effects, i.e. cumulative toxicity and begin to focus on identifying target organs. The major evaluation improvements over acute studies are whole animal observations while under test and histopathological evaluation of tissues for microscopic effects. The data collected are shown in table 2-9, along with the data analysis, and data interpretation.

An example of detailed organ evaluation is shown for two commonly examined tissues - liver and kidney (table 2-10). Besides microscopic evaluation, cellular enzymes and other cellular components are analyzed. A specific for these organs can be evaluated. The additional end points shown in this table are not usually done routinely, but often become part of a research

Table 2-9: Data Collection and Analysis

Data Collection

- Measured endpoints
- Clinical signs
- Body and organ weight
- Hematology
- Pathology
  - Gross lesions
  - Histopathology

Data Analysis

- Statistics and multiple comparisons
- Good Laboratory Practices

Table 2-10: Multiple Endpoints for a 90-day Study

Target Organ	Core endpoints	Additional endpoints
Liver	Histopathology Clinical chemistry ALT, AST, AP, glue, chol., bili, prot., alb., GGT, triglycerides Organ weight Clinical signs	Additional stains Electron microscopy Cell proliferation Enzyme levels
Kidney	Histopathology Clinical chemistry BUN, protein, electrolytes Urinalysis protein, pH, specific gravity Organ weight	Additional stains Electron microscopy Cell proliferation Enzyme levels

investigation examining mode or mechanism of action.

The overall above discussion briefly describes the basic mammalian evaluation of an existing chemical. The chronic exposure component for which all practical purposes is aimed at determining whether or not the chemical has carcinogenic potential is beyond the scope of this discussion.

ISSUES

The second portion of this document addresses several issues which have emerged with the implementation of TSCA and have an impact on how testing is carried out.

The following tables were prepared to give a view of costs for conducting the various studies (table 2-1 1). For comparison purposes, a previous publication in 1973 (3) shows that costs of toxicity tests were about one-tenth of what they are today. This averages out to over a 10% increase on an annual basis. Table 2-12 shows how long a study needs to be run, ie. exposure duration in order to develop adequate information on a particular endpoint and table 2-13 addresses the various non-cancer endpoint of general interest. From these two tables, duration of exposure of at least 28 day provides appropriate data in the rat model. The 14 day study is likely insufficient in duration to reach a steady state for metabolism and lesion development and the 90 day study may not provide that much more information. A 90 day study covers the male rat sperm cycle of about 60 days, whereas 28 days might have limitations. These tables are useful in helping decide what duration of testing should be considered for a chemical and raises a fundamental issue in experimental design which focuses on length of exosure. Thus, we find that a 28 day study would

Table 2-11: Cost of Various Laboratory Studies for Acute and Chronic Toxicity

Test	cost ($1000)
Acute battery • Oral LD50 • Inhalation LC50 • Dermal LD50 • Eye Irritation • Skin Sensitivity	30-40
Mutagenicity battery • Ames • CHO/HGPRT • Mouse micronucleus	40-60
Repeated exposure • oral, dermal, or inhalation	35- 90*
Subchronic • oral, dermal, or inhalation	120- 200'
Metabolism	50-250
Developmental	120-160
Reproduction	350-500
Chronic/Oncogenicity	600-1200

*Costs for each route of exposure

Table 2-12: Duration of Test Exposures Needed to Generate Adequate Information on Various Endpoints

	Exposure duration (with rats)			
Endpoint	**Acute**	**14 Day**	**28 Day**	**90 Day**
Lethality	+	+	+	+
Clinical signs	+	+	+	+
Toxic signs time course	-	+	+	+
Body weight and food data	-	+	+	+
Hematology		+	+	+
Gross necropsy	*	+	+	+
Clinical chemistry	-	+	+	+
Histopathology	*	+	+	+
Target organ		*	+	+
Dose-response	-	+	+	+

+ = Test needed; - = Test not needed; ± = Test maybe needed

be an appropriate screen for longer term, dose level selections for special studies and perhaps more importantly for describing systemic toxicity.

A previously publication comparing 6 month studies to longer term ones, suggests that 6 months is adequate for identifying non-cancer endpoints vs. 12 months or longer studies (1,4). An up-to-date comparison of 14 vs 28 vs 90 day duration studies is needed so that a data base is established to make case for shorter duration studies, with the caveat that a comprehensive 28 day study can be sufficient for determining repeated exposure effects.

The implementation of the Good Laboratory Practice (GLP) standards under TSCA published in FR November 19, 1983 and revised in 1989, FR August 17, 1989, added several layers of details, especially documentation of procedures and protocols to the conduct of studies under TSCA with GLPs. However, these efforts add significantly to the testing costs, measured in time, required to complete a study, amount of documentation, and quality assurance compliance activities. An unintended side effect of this burden has been a tendency to raise the threshold for deciding whether to conduct a study which, for screening tests, could be counter-productive.

Table 2-13: Test Exposures Needed to Generate Information on Various Non-Cancer Endpoints

	Exposure duration (with rats)			
Endpoint	**Acute**	**14 Day**	**28 Day**	**90 Day**
Neurotoxicity	±	+	+	+
Immunological	-	±	+	+
Reproductive		±	+	+
Pharmacokinetics	-	+	+	+
Mutagenicity	-	+	+	+

+ = Test needed; - = Test not needed; ± = Test maybe needed

However, with GLP in place the overall quality of studies is increased in a way measured by having sufficient detail to available to describe all aspects of the study.

An issue which we raise are the costs (defined both in absolute dollars, as well as utilization of finite resources) of doing testing under TSCA. This is tied in with a second issue of data gaps versus data needs.

As we have heard from several discussions during this conference, there are approximately 60,000 chemicals on the TSCA inventory and about 10,000 of those materials are out in commerce. Of that 10,000 approximately 1,000 have sufficient toxicity data developed to make judgment for risk assessments, although there is considerable variability in the amount and type of information.

The number of additional endpoints which can be added on to any study will have limitations. Two items that begin to put boundaries on add-ons are the ability to properly manage the logistics of the study and secondly, the interpretation of the data. The first item will be affected by GLP's and the second brings into light whether the experimental design was appropriate for addressing the particular toxicity endpoint.

We would also like to share some experiences with you concerning TSCA Section 4 test programs. These programs require significant amounts of time to develop and once finalized their implementation is attached to a timeline for completion and submittal of reports to the appro-

priate TSCA office. Once received by the agency internal review takes place, but the timeline for that is not known to the parties responsible for conducting the studies. Generally, there is no conclusionary meeting between the agency and parties affected about the study results and any associated issues which may have been raised by those studies. Examples of this are also found with TSCA Section 4 test rules, such as triethylene glycol ethers, cyclohexanone and the phenylene diamines. We would propose that data interpretation and recommendations should be discussed by both parties so that there is a value beyond than just completing the test requirements. The opportunity for this has been consistently missed.

CONCLUSION

Only when properly designed and conducted toxicity studies are carried out, can effective strategies for predicting potential hazards of chemicals be realized. We believe that simple, acute studies, followed by 28 days multi-dose rodent studies by appropriate exposure routes can address the majority of toxicity endpoints for further prioritization of testing. As part of the initial strategy, the selection of existing chemicals under TSCA should include evaluation of potential human exposure and production volume triggers.

REFERENCES

1. Betton, G. et al., *Human and Experimental Toxicology* 13:221-232, *1994.*
2. Brown, V.K.H., "Acute Toxicity Testing - a Critique," Chapter 3 in *Testing for Toxicity,* J.W. Garrod, ed. (Taylor and Francis Ltd., 1981).
3. Ghering, P.J., Rowe, V.K. and McCollister, S.B. *Food and Cosmetic Toxicology* 11:1097-1110, 1973.
4. Lumley, C.E. and Walker, S.R. Reg. *Toxicol. Pharmacology* 6:66-72, *1986.*
5. *Offical Journal of the European Communities,* L383QA, V35, Dec 29, 1992