March 20, 2013

DATA QUALITY IN TOXICOLOGY STUDIES:
A Key Element in Systematic Review for Evaluating Chemical Risks

Richard A. Becker Ph.D. DABT
American Chemistry Council

American Chemistry Council

Data Quality: A Key Element

Systematic Review for Evaluating Chemical Risks

- Goal of systematic review for evaluating chemical hazards/risks
- Stages/Phases of a systematic review
- Types of toxicology studies
- Understanding regulatory toxicology
- Approaches for evaluating data quality and study reliability
- Conclusions

Goal of systematic review for evaluating chemical hazards/risks

"Consistent with established best practices of systematic evidence-based reviews, we support use of transparent, objective criteria for determining data quality and study reliability.

Such criteria allow data from laboratory experiments, epidemiological investigations, and cutting-edge mechanistic research from all relevant studies, GLP and non-GLP, and from all investigators, regardless of affiliation or funding source, to be comprehensively and systematically reviewed, given appropriate weight, and integrated in a manner that provides a robust understanding of the mode of action and the potential hazards and risks that exposures to a substance could pose."

Conrad JW, Jr, Becker RA. 2011b. Environ Health Perspect. 119: a508-a509.

Stages/Phases of a Systematic Review

Phase 1. Define Causal Question and Develop Criteria for Study Selection

Phase 2. Develop and Apply Criteria for Review of Individual Studies

Phase 3. Integrate and Evaluate Evidence (Weight of Evidence Framework)

Phase 4. Draw Conclusions Based on Inferences

Types of Toxicology Studies

Human Epidemiology Studies

- Ecological
- Cross-Sectional
- Cohort Study
- Case Control
- Occupational
- Case Reports

In Vivo Lab Animal Studies

- Test Guideline (TG) and non-TG
- Acute, Subchronic, Chronic/Carcinogenicity, Repro, Neuro, Immuno, Developmental,
- Mechanistic

In Vitro Studies

- Test Guideline (TG) and non-TG
- Genetox, Cell Transformation, Cytoxicity,
- Mechanistic

In Silico (Computer) Studies

- Guidance compliant (e.g. OECD QSAR Principles) and those that aren't
- Phys/Chem, (Q)SAR, read-across, Fate & Transport, etc.

Understanding Regulatory Toxicology

Regulatory Science is the science of developing and applying tools, standards, and approaches to assess the safety of regulated products.

(certain statutes and regulations may also require evaluating efficacy, quality and performance)

Why Are Study Reliability and Data Quality So Important In Regulatory Science?

Scientific Confidence is needed - to be assured that products are used safely:

- **What are the hazards and risks and how can health and the environment be protected?**
- **Establish regulatory actions to limit uses / exposures (transport modalities, define uses, require controls, define personal protective equipment, etc.)**

Regulatory Toxicology Studies

- Validated test systems: validation guidance from ICCVAM, ECVAM, OECD
- Standardized Test Guidelines: primarily EPA and OECD
- Quality Assurance and Quality Control: Good Laboratory Practices (EPA, FDA, NIEHS/NTP in US, EU (REACH, PPP, etc.) and OECD globally,

Guidance for Validating Test Methods

- Establish the relevance and reliability of a method for its proposed / intended use
- Interagency Coordinating Committee for Validation of Alternative Methods (ICCVAM) http://iccvam.niehs.nih.gov/
- European Centre for the Validation of Alternative Methods (ECVAM) – has evolved into European Union Reference Laboratory on Alternatives to Animal Testing http://ihcp.jrc.ec.europa.eu/our_labs/eurl-ecvam
- Japanese Center for the Validation of Alternative Methods (JaVAM) http://jacvam.jp/en/
- Organization for Economic Cooperation and Development (OECD): Guidance Document 34 "Validation and International Acceptance of New or Updated Test Methods for Hazard Assessment" http://search.oecd.org/officialdocuments/displaydocumentpdf/?cote=env/jm/mono(2005)14&doclanguage=en

Test Method Validity (1): from ICCVAM

1. The scientific and regulatory rationale for the test method, including a clear statement of its proposed use, should be available.

2. The relationship of the test method's endpoint(s) to the biologic effect of interest must be described. Although the relationship may be mechanistic or correlative, tests with biologic relevance to the toxic process being evaluated are preferred.

3. A detailed protocol for the test method must be available and should include a description of the materials needed, a description of what is measured and how it is measured, acceptable test performance criteria (e.g., positive and negative control responses), a description of how data will be analyzed, a list of the species for which the test results are applicable, and a description of the known limitations of the test including a description of the classes of materials that the test can and cannot accurately assess.

Test Method Validity (2): from ICCVAM

4. The extent of within-test variability, and the reproducibility of the test within and among laboratories must have been demonstrated. Data must be provided describing the level of intra- and interlaboratory reproducibility and how it varies over time. The degree to which biological variability affects test reproducibility should be addressed.

5. The test method's performance must have been demonstrated using reference chemicals or test agents representative of the types of substances to which the test method will be applied, and should include both known positive and known negative agents. Unless it is hazardous to do so, chemicals or test agents should be tested under code (blinded) to exclude bias.

6. Sufficient data should be provided to permit a comparison of the performance of a proposed substitute test with that of the test it is designed to replace. Performance should be evaluated in relation to existing relevant toxicity testing data, and relevant toxicity information from the species of concern. Reference data from the comparable traditional test method should be available and of acceptable quality.

Test Method Validity (3):from ICCVAM

7. The limitations of the method must be described; for example, in vitro or other non-animal test methods may not replicate all of the metabolic processes relevant to chemical toxicity that occur in vivo.

8. Ideally, all data supporting the validity of a test method should be obtained and reported in accordance with Good Laboratory Practices (GLPs). Aspects of data collection not performed according to GLPs must be fully described, along with their potential impact.

9. All data supporting the assessment of the validity of the test method must be available for review. Detailed protocols should be readily available and in the public domain. The method(s) and results should be published or submitted for publication in an independent, peer-reviewed publication. The methodology and results should have been subjected to independent scientific review.

Standardized Test Guidelines: primarily EPA and OECD

OCSPP Home
Basic Information
Pesticides
Chemicals
Science Policy
Pollution Prevention
Laws & Regulations
Information Sources
Test Methods & Guidelines
A to Z Subject Index
En español
For KIDS

OCSPP Harmonized Test Guidelines

Series 870 - Health Effects Test Guidelines

The FINAL guidelines on this page are part of a series of test guidelines that have been developed by the Office of Chemical Safety ar development of test data for submission to the Agency.

A Master List (PDF) (28 pp, 80K, About PDF) of the OCSPP Harmonized Test Guidelines is available | Microsoft Excel Version (84K) (Excel vi

More information about OCSPP Harmonized Test Guidelines.

You will need Adobe Reader to view some of the files on this page. S

Group A – Acute Toxicity Test Guidelines

- 870.1000 - Acute Toxicity Testing--Background (December 2002)
- 870.1100 - Acute Oral Toxicity (December 2002)
- 870.1200 - Acute Dermal Toxicity (August 1998)
- 870.1300 - Acute Inhalation Toxicity (August 1998)
- 870.2400 - Acute Eye Irritation (August 1998)
- 870.2500 - Acute Dermal Irritation (August 1998)
- 870.2600 - Skin Sensitization (March 2003) (PDF)

Group B – Subchronic Toxicity Test Guidelines

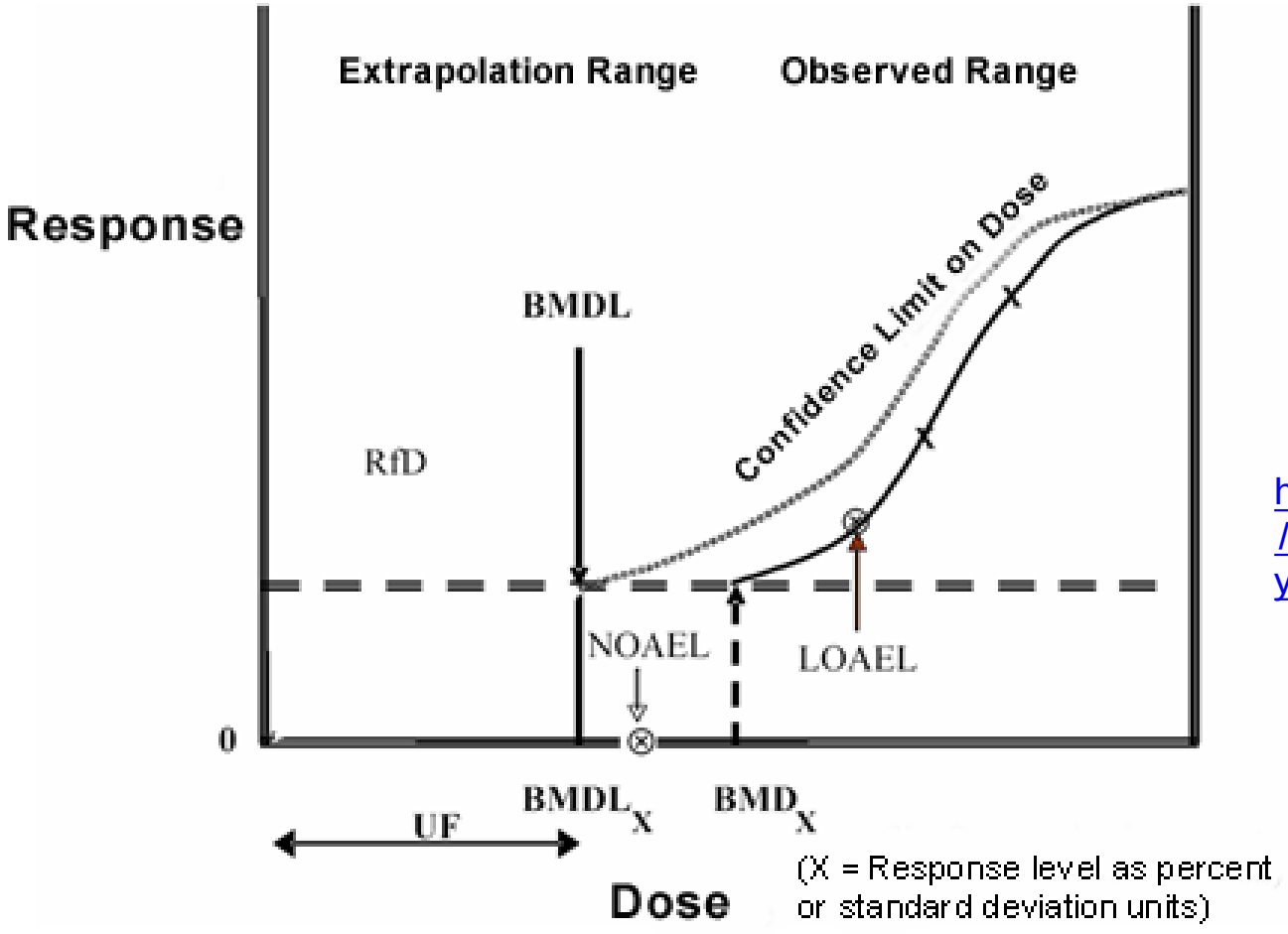
- 870.3050 - Repeated Dose 28-Day Oral Toxicity Study in Rodents (July 2000)
- 870.3100 - 90-Day Oral Toxicity in Rodents (August 1998)
- 870.3150 - 90-Day Oral Toxicity in Nonrodents (August 1998)
- 870.3200 - 21/28-Day Dermal Toxicity (August 1998)
- 870.3250 - 90-Day Dermal Toxicity (August 1998
- 870.3465 - 90-Day Inhalation Toxicity (August 1998)
- 870.3550 - Reproduction/Developmental Toxicity Screening Test (July 2000)
- 870.3650 - Combined Repeated Dose Toxicity Study with the Reproduction/Developmental Toxicity Screening Test (July 2000)
- 870.3700 - Prenatal Developmental Toxicity Study (August 1998)
- 870.3800 - Reproduction and Fertility Effects (August 1998)

Group C – Chronic Toxicity Test Guidelines

Dose Selection: What bias?

(2) **Administration of test and control substances**—(i) **Dose levels and dose selection.** (A) At least three-dose levels and a concurrent control should be used. Healthy animals should be randomly assigned to the control and treatment groups, in a manner which results in comparable mean body weight values among all groups. The dose levels should be spaced to produce a gradation of toxic effects. Unless limited by the physical/chemical nature or biological properties of the test substance, the highest dose should be chosen with the aim to induce some reproductive and/or systemic toxicity but not death or severe suffering. In the case of parental mortality, this should not be more than approximately 10 percent. The intermediate dose levels should produce minimal observable toxic effects. The lowest dose level should not produce any evidence of either systemic or reproductive toxicity (i.e., the no-observed-adverse-effect level, NOAEL) or should be at or near the limit of detection for the most sensitive endpoint. Two- or four-fold intervals are frequently optimal for spacing the dose levels, and the addition of a fourth test group is often preferable to using very large intervals (e.g., more than a factor of 10) between dosages.

Bench Mark Dose (BMD): Uses Entire Curve



The 95% confidence limit on the BMD is determined (BMDL) and that value is used as the Point of Departure (POD) for the analysis

http://water.epa.gov/learn/training/standardsacademy/health_page12.cfm

Test Guidelines + Good Laboratory Practices

- **To ensure that studies use sufficient and relevant dosing protocols, adequate replicates of animals for meaningful statistical analysis, interim analysis when applicable, and analysis of endpoints (organ weights, clinical chemistry, histopathology, etc.) which are considered validated by regulatory organizations for use in safety assessment.**
- **A typical GLP study submitted to a regulatory agency (including those funded by NIEHS/NTP) contains all raw data collected during the course of the study.**
- **Thereby allowing an independent review and audit of the study and independent analysis of the findings.**

Good Laboratory Practices

FDA:
http://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/cfrsearch.cfm?cfrpart=58

EPA:
http://www.epa.gov/compliance/monitoring/programs/fifra/glp.html

OECD:
http://www.oecd.org/chemicalsafety/testing/oecdseriesonprinciplesofgoodlaboratorypracticeglpandcompliancemonitoring.htm

Klimisch Method (1): Study Quality & Reliability

Reliability - evaluating the inherent quality of a test report or publication relating to preferably standardized methodology and the way the experimental procedure and results are described to give evidence of the clarity and plausibility of the findings;

Relevance - covering the extent to which data and tests are appropriate for a particular hazard identification or risk characterization; and

Adequacy - defining the usefulness of data for hazard/risk assessment purposes. When there is more than one study for each endpoint, the greatest weight is attached to the study that is the most reliable and relevant.

Klimisch HJ, Andreae E and Tillmann U (1997). A systematic approach for evaluating the quality of experimental and ecotoxicological data. Reg.Tox. and Pharm. 25:1-5

Klimisch Method (2): Scoring

1 = reliable without restrictions: "studies or data…generated according to generally valid and/or internationally accepted testing guidelines (preferably performed according to GLP) or in which the test parameters documented are based on a specific (national) testing guideline…or in which all parameters described are closely related/comparable to a guideline method."

2 = reliable with restrictions: "studies or data…(mostly not performed according to GLP), in which the test parameters documented do not totally comply with the specific testing guideline, but are sufficient to accept the data or in which investigations are described which cannot be subsumed under a testing guideline, but which are nevertheless well documented and scientifically acceptable."

3 = not reliable: "studies or data…in which there were interferences between the measuring system and the test substance or in which organisms/test systems were used which are not relevant in relation to the exposure (e.g., unphysiologic pathways of application) or which were carried out or generated according to a method which is not acceptable, the documentation of which is not sufficient for assessment and which is not convincing for an expert judgment."

4 = not assignable: "studies or data….which do not give sufficient experimental details and which are only listed in short abstracts or secondary literature (books, reviews, etc.).

Klimisch (3): Refinement

Klimisch Criteria for Reliability Categories

Code	Justification
1	Guideline study (OECD, *etc.*)
	Comparable to guideline study
	Test procedure according to national standards (DIN, *etc.*)
2	Acceptable, well-documented publication/study report which meets scientific principles
	Basic data given; comparable to guidelines/standards
	Comparable to guideline study with acceptable restrictions
3	Method not validated
	Documentation insufficient for assessment
	Does not meet important criteria of today standard methods
	Relevant methodological deficiencies
	Unsuitable test system
4	Only short abstract available
	Only secondary literature (review, tables, books, *etc.*)

Criteria for Reliability Categories (modified by ECETOC)

Code	Category of reliability
1	Reliable without restriction
1a	'Good laboratory practice' guideline study (OECD, EC, EPA, FDA, *etc.*)
1b	Comparable to guideline study
1c	Test procedure in accordance with national standard methods (AFNOR, DIN, *etc.*)
1d	Test procedure in accordance with generally accepted scientific standards and described in sufficient detail
2	Reliable with restrictions
2a	Guideline study without detailed documentation
2b	Guideline study with acceptable restrictions
2c	Comparable to guideline study with acceptable restrictions
2d	Test procedure in accordance with national standard methods with acceptable restrictions
2e	Study well documented, meets generally accepted scientific principles, acceptable for assessment
2f	Accepted calculation method
2g	Data from handbook or collection of data
3	Not reliable
3a	Documentation insufficient for assessment
3b	Significant methodological deficiencies
3c	Unsuitable test system
4	Not assignable
4a	Abstract
4b	Secondary literature
4c	Original reference not yet available
4d	Original reference not translated
4e	Documentation insufficient for assessment

ToxRTool (1): Improved Approach for Study Quality and Reliability

"Evaluation of the reliability of toxicological data is of key importance for regulatory decision-making." European Commission's Joint Research Centre

ToxRTool: a tool to assess the reliability of toxicological data
http://ihcp.jrc.ec.europa.eu/our_labs/eurl-ecvam/archive-publications/toxrtool

Toxicology Letters 189 (2009) 138–144

Contents lists available at ScienceDirect

Toxicology Letters

journal homepage: www.elsevier.com/locate/toxlet

"ToxRTool", a new tool to assess the reliability of toxicological data

Klaus Schneider[a,*], Markus Schwarz[a], Iris Burkholder[b], Annette Kopp-Schneider[b], Lutz Edler[b], Agnieszka Kinsner-Ovaskainen[c], Thomas Hartung[d], Sebastian Hoffmann[e]

ToxRTool (2): Improved Approach

Criteria No.	Criteria Group I: Test Substance Identification	Score	Evaluator's explanations, comments on criteria, *etc.*
1	**WAS THE TEST SUBSTANCE IDENTIFIED?**		
2	Is the purity of the substance given?		
3	Is information on the source/origin of the substance given?		
4	Is all information on the nature and/or physico-chemical properties of the test item given, which you deem indispensable for judging the data (see explanations for examples)?		
		[Total]	
	Criteria Group II: Test organism characterisation		
5	**IS THE SPECIES GIVEN?**		
6	Is the sex of the test organism given?		
7	Is information given on the strain of test animals plus, if considered necessary to judge the study, other specifications (see explanation for examples)?		
8	Is age or body weight of the test organisms at the start of the study given?		
9	For repeated dose toxicity studies only (give point for other study types): Is information given on the housing or feeding conditions?		
		[Total]	

ToxRTool (3): Improved Approach

	Criteria Group III: Study design description		
10	**IS THE ADMINISTRATION ROUTE GIVEN?**		
11	**ARE DOSES ADMINISTERED OR CONCENTRATIONS IN APPLICATION MEDIA GIVEN?**		
12	**ARE FREQUENCY AND DURATION OF EXPOSURE AS WELL AS TIME-POINTS OF OBSERVATION EXPLAINED?**		
13	**WERE NEGATIVE (WHERE REQUIRED) AND POSITIVE CONTROLS (WHERE REQUIRED) INCLUDED (GIVE POINT ALSO, WHEN ASBSENT BUT NOT REQUIRED, SEE EXPLANATIONS FOR STUDY TYPES AND THEIR RESPECTIVE REQUIREMENTS ON CONTROLS)?**		
14	**IS THE NUMBER OF ANIMALS (IN CASE OF EXPERIMENTAL HUMAN STUDIES: NUMBER OF TEST PERSONS) PER GROUP GIVEN?**		
15	Are sufficient details of the administration scheme given to judge the study (see explanation for examples)?		
16	For inhalation studies and repeated dose toxicity studies only (given point for other study types): Were achieved concentrations analytically verified or was stability of the test substance otherwise ensured or made plausible?		
		[Total]	
	Criteria Group IV: Study results documentation		
17	Are the study endpoint(s) and their method(s) of determination clearly described?		
18	Is the description of the study results for all endpoints investigated transparent and complete?		
19	Are the statistical methods applied for data analysis given and applied in a transparent manner (give also point, if not necessary/applicable, see explanations)?		
		[Total]	
	Criteria Group V: Plausibility of study design and results		
20	**IS THE STUDY DESIGN CHOSEN APPROPRIATE FOR OBTAINING THE SUBSTANCE-SPECIFIC DATA AIMED AT (SEE EXPLANATIONS FOR DETAILS)?**		
21	Are the quantitative study results reliable (see explanations for arguments)?		
		[Total]	
	A Numerical result leads to initial Category:	[#]	
	B Checking UPPER-CASE BOLD scores leads to revised Category:	[#]	
	C Evaluator's proposal: Category:		
	D Justification in case evaluator deviates from B:		

ToxRTool (4): Improved Approach

Reliability Categorization (defined by Klimisch *et al.*, 1997)				(Proposed) Consequence
	In vivo	*In vitro*		
1	18-21	15-18	Reliable without restriction	Useful, check relevance for intended purpose
2	11-14	11-14	Reliable with restrictions	Potentially useful, check relevance for intended purpose
3	<11 or not all key criteria met*	<11 or not all key criteria met*	Not reliable	Generally not to be used as key study, but depending on the short-comings of the study, may still be useful in a weight-of evidence (WoE) approaches or as supportive information
4			Not assignable	Generally not to be used as key study, but depending on the short-comings of the study, may still be useful in WoE approaches or as supportive information (This category is not an outcome of this evaluation tool)

EPA's Five General Assessment Factors for Evaluating Quality of Scientific Information

1. **Soundness – The extent to which the scientific and technical procedures, measures, methods or models employed to generate the information are reasonable for, and consistent with, the intended purpose.**

2. **Applicability and Utility – The extent to which the information is relevant for the Agency's intended use.**

3. **Clarity and Completeness – The degree of clarity and completeness with which the data, assumptions, methods, quality assurance, sponsoring organizations and analyses employed to generate the information are documented.**

4. **Uncertainty and Variability - The extent to which the variability and uncertainty (quantitative and qualitative) in the information or the procedures, measures, methods or models are evaluated and characterized.**

5. **Evaluation and Review - The extent of independent verification, validation and peer review of the information or the procedures, measures, methods or models.**

USEPA (2003). A summary of general assessment factors for evaluating the quality of scientific and technical information. Science Policy Council, Washington DC. EPA/100/B-03/001. *http://www.epa.gov/osa/spc/pdfs/assess2.pdf*

Conclusions

- **Use of transparent, objective criteria for determining data quality and study reliability of toxicity studies are best practices.**
- **There are existing approaches, endorsed and used by regulatory agencies globally, for determining data quality and study reliability for toxicity studies: both tests guideline studies and academic, non-guideline studies.**
- **Such criteria allow data from laboratory experiments, epidemiological investigations, and cutting-edge mechanistic research from all relevant studies, GLP and non-GLP, and from all investigators, regardless of affiliation or funding source, to be:**
    - **comprehensively and systematically reviewed**
    - **given appropriate weight, and**
    - **integrated in a manner that provides a robust understanding of the mode of action and the potential hazards and risks that exposures to a substance could pose.**

Some Further Reading

Bevan and Strother. Best Practices for Evaluating Method Validity, Data Quality and Study Reliability of Toxicity Studies for Chemical Hazard and Risk Assessments. 2012. http://arasp.americanchemistry.com/Data-Quality-Evaluation

Conrad JW, Jr, Becker RA. 2011. Enhancing credibility of chemical safety studies: emerging consensus on key assessment criteria. Environ Health Perspect. 119:757–764. and Conrad JW, Jr, Becker RA. 2011b. Chemical Safety Studies: Conrad and Becker Respond. Environ Health Perspect. 119: a508–a509.

Becker RA, Janus ER, White RD, Kruszewski FH, Brackett RE. Good Laboratory Practices and safety assessments. Environ Health Perspect. 2009;117:A482–A483. and Becker RA, Janus ER, White RD, Kruszewski FH, Brackett RE. 2010. Good Laboratory Practices: Becker et al. 118A194–A195.A195.

Tyl RW. Basic exploratory research versus guideline-compliant studies used for hazard evaluation and risk assessment: bisphenol A as a case study. Environ Health Perspect. 2009;117:1644–1651.

Lutter R et al., Data Disclosure for Chemical Evaluations Environ Health Perspect 121:145–148 (2013).

Goldman LR and Silbergeld EK. Assuring Access to Data for Chemical Evaluations. Environ Health Perspect 121:149–152 (2013).