April 23, 2014

DATA QUALITY AND STUDY RELIABILITY:
Key Elements in a Systematic Review for Evaluating Chemical Risks

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

American® Chemistry Council

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.

Recommendations

To assure data of sufficient quality are used, transparent criteria must be established upfront and then consistently applied throughout the assessment to identify studies and to evaluate their quality, relevance, and reliability.

Develop protocols for reviewing and evaluating study quality for each major type of study: epi, animal, in vitro

- **For epi: consider implementing the procedures described in Money et al, 2013**
- **For animal studies: consider implementing the refined Klimisch approach (delineated by ECETOC) or the ToxR Tool**
- **For in vitro studies: consider implementing the ToxR Tool**

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.).

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]

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

Conclusions re: In Vivo & In Vitro Studies

- **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.**

Establishing Quality and Reliability of Human Epidemiological Studies

Regulatory Toxicology and Pharmacology 66 (2013) 241–247

Contents lists available at SciVerse ScienceDirect

Regulatory Toxicology and Pharmacology

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

A systematic approach for evaluating and scoring human data

Chris D. Money [a,*], John A. Tomenson [b], Michael G. Penman [c], Peter J. Boogaard [d], R. Jeffrey Lewis [e]

Data type/ category description for human data quality criteria (chronic outcomes)
Type 1: reliable without restriction
Type 2: reliable with restriction
Type 3: not reliable
Type 4: not assignable

Human Epidemiological Studies:
Type 1 Reliable Without Restriction

- subjects represent appropriate exposure distributions of persons at risk
- emphasis on measuring & reporting response parameters
- adequate recruitment & and follow up to maximize participation and reduce loss
- exposure assessment made independent of outcome, with as little measurement error as possible, using well-established methods, quantitative, validated, individual-level data
- Outcome data collected independent of exposure status & rigorously ascertained for both cases and non-cases (or controls in a case control study)
- serious biases have been reduced by design, controlled through statistical adjustment and/or quantified through sensitivity analyses
- sample size/exposure range was sufficient to study the question under investigation
- data were analyzed comprehensively, using appropriate statistical techniques
- methodology and results comprehensively & transparently reported according the STROBE guidelines

Human Epidemiological Studies: Type 2 Reliable With Restriction

- Study & data possess most of the elements of a ‘‘Type 1’’ quality study, but overall quality compromised due to minor, but obvious, methodological limitations.
  - Examples of limitations:
    - limited measurement data available to validate estimated individual-level exposure data; imprecision because of a small sample size or low exposure range.
    - Study design may not be optimal e.g. a cross sectional design which does not allow inferences to be made about the time order of events,
    - Study in which subject selection procedures and/or post entry loss to follow up introduces the possibility of selection bias

Human Epidemiological Studies:
Type 3 Not Reliable

- Studies have serious methodological flaws that make the results uninterpretable regarding a causal association
- Fail to meet one or more of the most basic standards necessary to interpret epidemiologic research, such as:
  - appropriate study design
  - adequate selection of study subjects.

- Ecological studies and linkage studies: many such studies will be well conducted investigations and be useful for hypothesis generation, but their design renders them uninformative for hazard identification or risk assessment

Type 4 Not Assignable

- Studies or data from the literature which do not give sufficient details about the methodology used to assess their quality or which are only listed in short abstracts or secondary literature

Quality and Reliability: Use in Selecting the Critical (Key) Study

C.D. Money et al. / Regulatory Toxicology and Pharmacology 66 (2013) 241–247

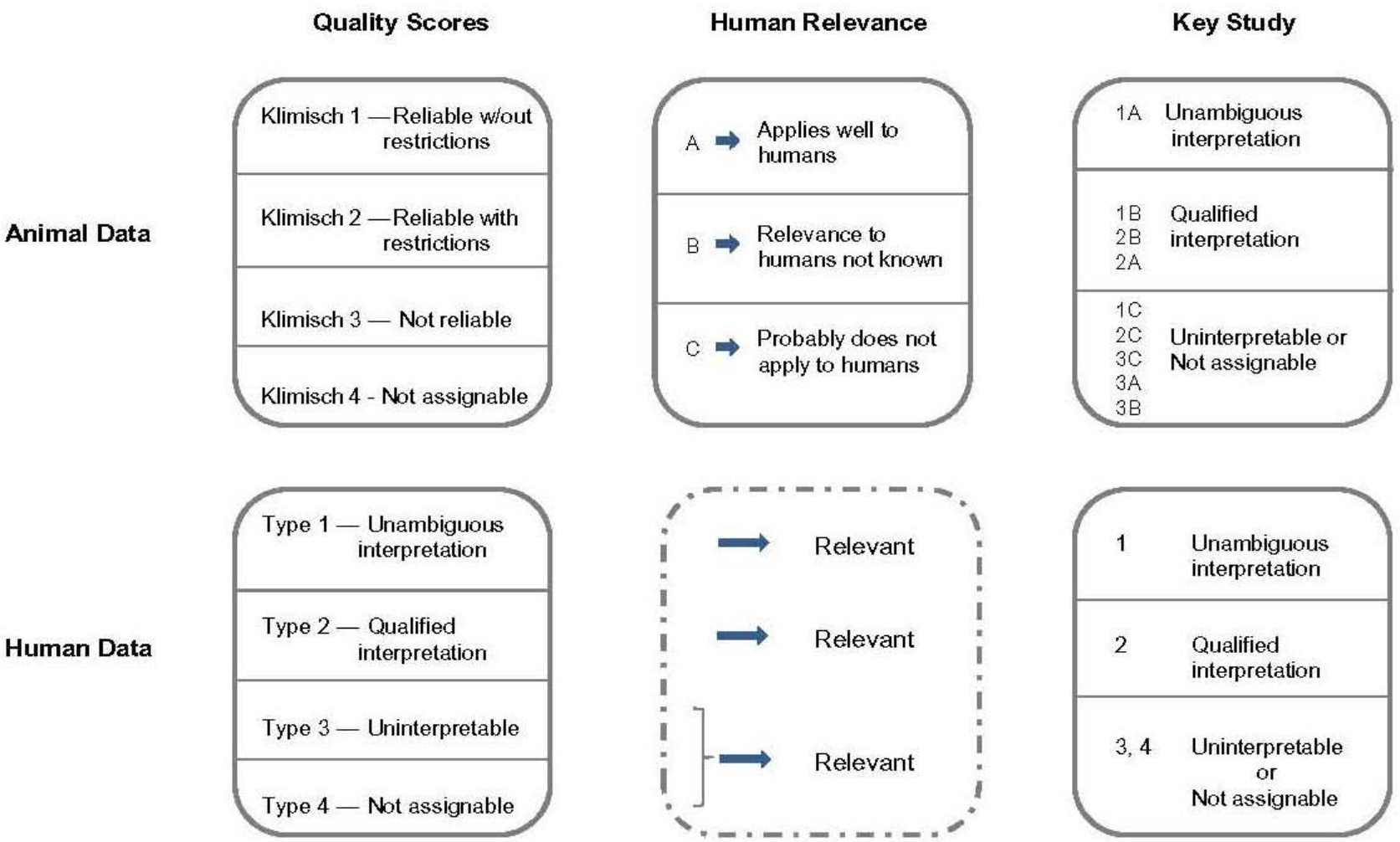

Fig. 1. Considerations when determining key studies under REACH.

Cross Sectional Epi Studies

- **Not a useful type of study for establishing causal relationships but can be useful for hypothesis generation**
- **Use of spot sample biomonitoring data in observational epi studies can be problematic for substances with short half-lives -- particularly in non-representative sample sizes**

STROBE Statement

Strengthening the reporting of observational studies in epidemiology

von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP; STROBE Initiative. **The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE)statement: guidelines for reporting observational studies**. PLoS Med. 2007 Oct 16;4(10):e296. PMID: 17941714

Hypothesized Mode of Action

Toxicologic Pathology, 34:209–219, 2006

ISSN: 0192-6233 print / 1533-1601 online
DOI: 10.1080/01926230600642625

Proposed Mode of Action for In Utero Effects of Some Phthalate Esters on the Developing Male Reproductive Tract

RAYMOND M. DAVID

Data shows that DEP:

1) does not alter testosterone synthesis in the testes

2) does not alter gene expression for steroidogenesis

3) does not produce genital system developmental malformations in rodents

Hypothesized MOA and Recent Epi Studies

- **Anogenital Index in humans: un-validated and its physiological significance in humans is unknown (McEwen and Renner, 2006; Romano-Riquer et al., 2007); whether AGI relates to clinically meaningful outcomes awaits further study**
- **Suzuki and colleagues (2012) found no correlation between maternal DEP exposure and AGI in 111 Japanese mother-infant pairs**
- **In 65 mother-infant pairs, Huang and colleagues (2009), no relationship was observed between AGI and MEP concentrations in maternal urine or amniotic fluid**
- **a prospective, case-control study of 5200 pregnant women (Chevrier et al., 2012) reported no evidence of increased risk of male genital anomalies with prenatal exposure to phthalates, as inferred from maternal urinary phthalate metabolites**

Improving Use of MOA in IRIS

1. Read review articles & formulate initial hypotheses. Ask "what are the are possible MOAs that could be operating wrt human risk, etc." Output - a set of initial alternative hypothesized MOAs

2. Use these initial hypotheses in designing the literature search. Focus on key events. Look at studies wrt the specific chemical itself and also the hypothesized MOAs (general biological knowledge).

3. Collect literature & evaluate for data quality & reliability

4. Match up the literature with each hypothesized MOA by Key Event. Read lit and if necessary refine the hypotheses (If add'l lit search is needed, then do this)

5. For each hypothesis, line up evidence with KEs and integrate to arrive at an overall WoE for each hypothesis.

6. Comparison of alternative hypotheses. E. G., "Overall WoE for hypothesis A is strongly supported. The exception is…." Overall WoE for B is not strongly supported, this hypothesis cannot explain/account for xxxxx." Ask: "is there info that could improve/inform understanding of the leading hypothesis?" "Could slight modification of the leading hypothesis increase the overall consistency of more of the data with this hypothesis?"

Recommendations

To assure data of sufficient quality are used, transparent criteria must be established upfront and then consistently applied throughout the assessment to identify studies and to evaluate their quality, relevance, and reliability.

Develop protocols for reviewing and evaluating study quality for each major type of study: epi, animal, in vitro

- **For epi: consider implementing the procedures described in Money et al, 2013**
- **For animal studies: consider implementing the refined Klimisch approach (delineated by ECETOC) or the ToxR Tool**
- **For in vitro studies: consider implementing the ToxR Tool**