Presented at
2014 SRA Annual Meeting
December 8, 2014
Denver, CO

Evaluation of Study Quality Criteria Frameworks

Julie E. Goodman, Ph.D., DABT, ACE, ATS
Heather N. Lynch, M.P.H.
Robyn L. Prueitt, Ph.D., DABT
Jade A. Tabony, M.S.
Lorenz R. Rhomberg, Ph.D., FATS
Nancy B. Beck, Ph.D., American Chemistry Council

Poster 176

ABSTRACT

The prevailing approaches for the systematic review and evaluation of chemical toxicity are currently being reconsidered, with specific focus on the evaluation of individual studies and their integration into the overall body of evidence. This renewed interest has arisen, in part, as a result of several prominent reviews of these approaches by special committees of the National Research Council (NRC), among others. We conducted a critical evaluation of several available study quality criteria frameworks. We assessed the criteria separately for human, animal, and *in vitro* studies, as well as systematic reviews, then evaluated commonalities across disciplines. We also considered the potential implications of applying criteria frameworks and how they bear on fundamental risk assessment questions. We found that the available frameworks within each discipline differed in terms of their intended purpose and level of guidance in decision-making. All frameworks across disciplines shared common themes, including the adequate reporting of specific details of study conditions and design/protocol, selection and randomization (where applicable) of study groups, outcome assessment methods and applicability, reporting the results of unadjusted and adjusted analyses (*i.e.*, avoiding selective reporting), and consideration of potential confounders or bias. We identified the most informative study quality considerations, which will enable researchers to implement more objective and standardized methods for evaluating studies and, ultimately, improve risk assessment methods.

BACKGROUND

A key element of any systematic review of chemical toxicity is an objective and standardized method for reporting and evaluating the quality of individual studies. Several frameworks provide guidance regarding criteria that should be considered when assessing individual study results and systematic reviews. We reviewed some commonly cited frameworks to determine similarities and differences among criteria and their bearing on the interpretation of study results.

OBJECTIVE

To identify the most common and useful criteria for evaluating the quality of studies used to assess potential causal relationships between chemicals and health effects.

METHODS

- Evaluated 10 publicly available study quality criteria frameworks commonly used in evaluations of chemical toxicity (Table 1)
- Note that some frameworks apply to multiple study types (animal, human, *in vitro*, and systematic reviews); we reviewed the criteria for all of the study types included in each of the 10 frameworks
- Tabulated criteria specified in each framework for human, animal, and *in vitro* studies, as well as systematic reviews
- Reported similarities and differences among frameworks and their implications on interpretation of results

Table 1 Frameworks Reviewed for Evaluating Study Quality Criteria

Study Quality Criteria System	Human Studies	Animal Studies	*In vitro* Studies	Systematic Reviews
Integrated Risk Information System (IRIS) Risk of Bias (RoB) Evaluation	√	√		√
The National Toxicology Program's (NTP's) Office of Health Assessment and Translation (OHAT) Approach	√	√		√
Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) System	√			
Money *et al.* (2013) Approach	√			
Navigation Guide	√	√		√
Animal Research: Reporting of *In vivo* Experiments (ARRIVE) Guidelines		√	√	
Klimisch System		√	√	
Organisation for Economic Co-operation and Development (OECD) Guidance Document (GD) 34		√	√	
Toxicological Data Reliability Assessment Tool (ToxRTool)		√	√	
Assessment of Multiple Systematic Reviews (AMSTAR) System				√

Table 2 Frameworks Reviewed for Evaluating Study Quality Criteria

The number of frameworks addressing each criterion is shown in each column. Frameworks specify whether each criterion should be a reporting requirement [Report]; a score, based on the degree to which a criterion is fulfilled [Score], or the fulfillment of a specific requirement outlined by authors [Y/N]. Criteria in blue are those that are common among at least half of frameworks for each study type.

Criteria	Human (5 Frameworks)			Animal (7 Frameworks)			*In vitro* (4 Frameworks)		
	Report	Score	Y/N[a]	Report	Score	Y/N[a]	Report	Score	Y/N[a]
Study Objectives	1			3			2		
Study Design/Setting	3			5			2		
Participant/Animal Characteristics	3			6	1		--	--	--
Study Size	3			5	1	1	1		1
Study Power Analysis	1								
Blinding and Randomization		3		2	3		2		
Comparison/Control Groups	1	2		6	1		3	1	
Husbandry	--	--	--	3	3	1	--	--	--
Inclusion/Exclusion Criteria	1			2	2		--	--	--
Experimental Procedure	--	--	--	4	1		2	1	1
Participation Rate/Attrition	1	2		1	2		--	--	--
Statistical Methods	1	2	1	2	2	1	2		1
Exposure Measurement Methods/Dose Admin.	1	3	1	5	2		3	1	
Confounding and Bias	1	3	1		2		--	--	--
Outcome Assessment	1	2	1	3	2	1	2	1	1
Result Reporting	2	3		3	3	1	2		1
Adherence to Protocol, Deviations		2			2				
Limitations	2			2			2		
Interpretation and Implications	1		1	2			2		
Generalizability	2			3			3		
Funding Source/Conflict of Interest	3	1		3	1		1		

Notes:

-- = Not applicable.

(a) Y/N criteria include specific requirements defined by authors, e.g., "Attrition minimized" or "Statistical analyses appropriate for endpoint."

Table 3 Summary of Systematic Review Study Quality Criteria Addressed in Frameworks

The number of frameworks addressing each criterion is shown in each column. Frameworks specify whether each criterion should be a reporting requirement [Report]; a score, based on the degree to which a criterion is fulfilled [Score]; or the fulfillment of a specific requirement outlined by authors [Y/N]. Criteria in blue are those that are common among at least half of the frameworks.

Criteria	Number of Frameworks (Out of Total of 4)		
	Report	Score	Y/N[a]
Review Objective Identified			3
A priori Design/Protocol for the Review			4
Comprehensive Literature Search of More than One Database			4
Details of the Search Strategy	4		
Two Independent Reviewers of Data			4
Procedure for Disagreements between Study Reviewers			4
List of Excluded and Included Studies			3
Study Characteristics Reported (*e.g.*, in a table)			2
Assessment and Documentation of the Scientific Quality of Each Study		3	1
Appropriate Methods to Combine Findings Across Studies[b]			2
Overall Confidence Rating for Body of Evidence		3	
Qualitative Assessment of Publication Bias		1	2
Overall Conclusions for Hazard Identification[c]		3	
Statement of Possible Conflict of Interest in Both Systematic Review and Included Studies		1	2
Discussion of Deviation from Review Protocol			1

Notes:

(a) Y/N criteria include specific requirements defined by authors, e.g., "Attrition minimized" or "Statistical analyses appropriate for endpoint."

(b) Tests were done to ensure studies were suitable to combine, such as Chi-square test for homogeneity.

(c) For example, studies are pooled to arrive at a final conclusion for the body of evidence, e.g., "sufficient," "suggestive," or "inadequate" epidemiologic evidence of an association.

RESULTS

- For some systems (*e.g.*, OHAT, IRIS RoB), there is no single document with detailed guidance on how to apply the criteria; several documents must be consulted for implementation
- Purposes of frameworks varied; *e.g.*, some consisted entirely of study reporting requirements, while others provided guidance for evaluating study design and methods
- Some common themes among all 10 frameworks, which can be interpreted as commonly recognized quality considerations that are the most informative for evaluating causal relationships, are:
  - Reporting of study conditions and design/protocol
  - Representative population selection and randomization of study groups
  - Accurate exposure measurement and outcome assessment methods
  - Appropriate statistical methods
  - Complete results reporting (*i.e.*, presenting results of all analyses)
  - Consideration of potential confounders and other biases

DISCUSSION

- The term "risk of bias" is defined differently by various stakeholders and has been applied to some, but not all, of the study quality frameworks
  - NRC defines "study quality" as the extent to which the research was conducted to the highest possible standards and "risk of bias" as issues that impact internal validity and characteristics that could introduce systematic error
  - US EPA defines "risk of bias" as an evaluation of both internal validity as well as the quality of methods and adequate reporting of results
  - Sorting out the distinctions between and the exact meanings of the terms should be discussed among stakeholders
- A study of higher quality is not necessarily correct, nor is one of lower quality necessarily wrong; if study outcomes are inconsistent, extraneous factors in lower quality studies are more likely to have affected the results
- Consideration of the relevant criteria for causal inference that we identified will enable researchers to implement more objective and standardized methods for evaluating study quality and, ultimately, improve risk assessment methods

ACKNOWLEDGMENT

This work was funded by the American Chemistry Council (ACC). The authors retain sole responsibility for the content.

COPIES OF THE FULL TABLES FROM THE WHITE PAPER AND FULL CITATIONS AVAILABLE UPON REQUEST. IF YOU WOULD LIKE A REPRINT OF THIS POSTER, PLEASE LEAVE A BUSINESS CARD IN THE ENVELOPE.