

# Embracing Rigor and Reproducibility in Scientific Research

Thursday, January 23, 2020  
12:30 PM EST

Webinar sponsored by



Graduate Student  
Leadership Committee  
(GSLC)

## Speakers

Robert Chapin, PhD  
G. Jean Harry, PhD  
Daniele Wikoff, PhD

## Moderators

Tamara Young, Catheryne (Katie) Chiang and Sumira Phatak

# Webinar Moderators



**Tamara L. Young**

*GSLC Professional Development Committee  
Chair*

Doctoral Candidate  
University of New Mexico



**Catheryne (Katie) Chiang**

*GSLC Professional Development Committee  
Secretary*

Recent PhD Recipient  
University of Illinois at Urbana-Champaign



**Sumira Phatak**

*GSLC Professional Development Committee  
Volunteer*

Predocctoral Research Fellow at USDA  
Utah State University

# Webinar Speakers



**Robert E. Chapin, PhD**

Senior Research Fellow, Retired

Pfizer

NIEHS



**G. Jean Harry, PhD**

Group Leader, Neurotoxicology  
Laboratory, National Toxicology  
Program

NIEHS



**Daniele Wikoff, PhD**

Senior Managing Scientist,  
Practice Director, Health  
Sciences

Toxstrategies, Inc.



# **Part 1:**

# **The Importance of Rigor and Reproducibility in Research**

**Robert E Chapin, PhD**  
NIEHS, then Pfizer, now retired

# What is the replication crisis and what does it mean to you?

In 2010, Florian Prinz and some others performed an internal review of projects in Discovery at the pharma company Bayer Healthcare in Germany. They asked: How often can we effectively replicate the major findings of a key scientific paper?

The answer: only about 1/3<sup>rd</sup> of the time. For fully 2/3<sup>rd</sup>s of all projects, there were inconsistencies which either prolonged or delayed the project, or killed it entirely. Put another way: they can trust published results less than half the time.

# What is the replication crisis and what does it mean to you?

Many authors and papers have since confirmed similar experiences as Prinz. I had several similar experiences at NIEHS, and then many at Pfizer, where the results from academic papers, especially from high-impact journals, could not be replicated. (and sometimes our own work could not be replicated in our own lab 😞🧐)

This is an existential threat to the scientific enterprise. If we cannot believe published results, then 1) why bother? and 2) why fund it?

# So we're all on the same page, let's define some terms

Oh, wait...the literature doesn't even agree on what these mean. I like the definitions provided by Hans Plesser, although others will switch the definitions for Replicability and Reproducibility.

**Repeatability** (Same team, same experimental setup): The measurement can be obtained with stated precision by the same team using the same measurement procedure, the same measuring system, under the same operating conditions, in the same location on multiple trials. For computational experiments, this means that a researcher can reliably repeat her own computation.

**Replicability** (Different team, same experimental setup): The measurement can be obtained with stated precision by a different team using the same measurement procedure, the same measuring system, under the same operating conditions, in the same or a different location on multiple trials. For computational experiments, this means that an independent group can obtain the same result using the author's own artifacts

## Definitions, cont.

**Reproducibility** (Different team, different experimental setup):  
The measurement can be obtained with stated precision by a different team, a different measuring system, in a different location on multiple trials. For computational experiments, this means that an independent group can obtain the same result using artifacts which they develop completely independently.

From Plesser, H.E. (2018) Reproducibility vs. Replicability: A Brief History of a Confused Terminology. *Front. Neuroinform.* 11:76. doi: 10.3389/fninf.2017.00076

But also see: S.N. Goodman, D., Fanelli, and **John P.A. Ioannidis** (2016) What does research reproducibility mean? *Science Translational Medicine* 8(341) ps12

# The Decline Effect

First described in the 1930's and most recently re-emphasized by Jonathan Schooler at UCSB, this effect says that many published scientific effects seem to diminish with time and number of replicates. (Nature **470**:437, 2011). The differences between treated and control groups diminish with time.

Schooler believes that this would be reduced if **all** data were published, including negative findings.

John Ioannidis has claimed that "Most research findings are false" (PLoS Medicine 2(8): e124, 2005), and then "proved" that very assertion in that article.

# Probably many causes/contributors:

1. Most importantly: A refusal to acknowledge the scope of the problem by most established scientists. (~ climate change denial)
2. → Small n
3. → Insufficient experimental replication
4. The desperate drive to get papers in the “right” journals for institutional and reputational retention.
5. Poor experimental design for the question(s) being asked
6. Poor or inappropriate statistics
7. Lack of blinding during analysis
8. Small effect sizes
9. More tested relationships (hypothesis fishing) and
10. Selective outcome reporting
11. No pre-specification of a relationship to test
12. Reduced replicability is more likely when there is greater flexibility in designs or outcomes

# Incentives

We would be remiss if we did not mention the skewed incentive system in science, which, despite many Op-Eds and much public hand-wringing, remains heavily tied to publication of surprising findings in prestigious journals. Funding review panels expect applicants to correctly predict what will happen in their experiments, but even more than that, they love high-impact good surprises. Surprises are more likely to happen when one employs several characteristics from the preceding slide.

This is also closely tied to the idea of a “beefy” CV proving your value as a scientist.

And need we mention the drive of the institutional “publish or perish” mindset?

# Where Hope Lies:

The need for a strong CV is greatest early in your career, so trainees are most susceptible to all these pressures, and are least in a position to resist them. Trainees have little ability to impact the behavior or strategy of the more established scientists in the lab. Still, the most durable change often comes from the bottom, and if trainees demand methods to combat this problem, their mentors and deans will (eventually....we hope...maybe....) follow.

Additionally, trainees could demand more “air time” at meetings about this, and help pressure SOT to detail how the Society will address the issue, or → could volunteer to lead sessions on this problem and how other Societies are addressing it.

# Where Hope Lies:

One way to guard against this is to solicit a lab working in the same or similar field to replicate your experiment, with the proviso that they become co-authors on the resulting paper. Note that this assumes effective replication, which is a win-lose experiment, and if it doesn't replicate, you know there will likely be some ticked-off PI's.

This is an imperfect solution, and I'm hoping you can collectively come up with something better.



# **Part 2: Experimental Design to Ensure Rigor and Reproducibility**

**G. Jean Harry, PhD**  
National Institute of Environmental Health Sciences

# Why perform experiments?

- To evaluate hypothesis
- To indulge investigator's curiosity
- To try out a new method or technique
- To establish the existence of a new phenomenon
- To explore the conditions under which a phenomenon occurs

All are valid reasons to perform experiments but all should be designed in such a way to provide information to the field of study and be interpreted within the specific framework.

# Before you begin your research

- Establish a clear understanding of the research questions and objectives
- Ensure that you have a very good understanding of relevant literature
- Establish a conviction of what kinds of data are most needed
- What is the scientific rationale for initiating the study?
- What is the hypothesis that you are testing?
- Do the experiments conducted address this hypothesis?
- Will your findings be relevant and advance the field?
- If you have null findings would they also be meaningful?
- Get early feedback on concepts, approach, experimental design, statistical analysis, data interpretation
- Ensure you are using valid methods

# The science (data) does not stand alone

The overall study needs to be set within a framework that identifies

- the importance of the question being asked
- the appropriateness of the experiments
- how the data advances our understanding
- how to move forward

# Before you start your first experiment

Consult with a biostatistician

Statistics cannot save a bad experimental design!

- Clearly establish what comparisons you want to be able to make
- Outline what you consider to be “experiment controls”
- Identify assay controls that would not be included in the statistical analysis of the experimental effect
- Have pilot data or information in hand to show control variability and robustness of the endpoint

# Hypothesis development

- Do your homework and know the status of the research in the area
- As you identify the critical publications that set the basis for your research ideas – do not totally rely on those data
  - Outline pilot studies to determine if you can replicate those findings critical to your study. Remember pilot studies should be designed with the same care as any study.
  - In designing pilot studies, you will need to critically review the methods of the published study and likely find it lacking in important details. Do not hesitate to contact the authors if this is important to your study.

# Hypothesis development

- Clearly identify the overall hypothesis  
This can be rather general but realistic as to what can be accomplished
- Develop the specific aims to address the hypothesis  
These should represent the goals that you hope to achieve with the experiments thus, they are more focused and specific than the overall hypothesis and set the framework for the experiments
- Outline the experiments – ensure that they address the specific aims

As your experiments progress, this will become a reiterative process with modifications and adjustments along the way

# Model selection

Once you have your hypothesis and specific aims you now need to work out the model system you will employ. This requires careful consideration

1. What is the scientific basis for selection of model system?
2. Is this an established model system in the literature?
3. How difficult is the model system?
4. What is the reproducibility of the model system?  
e.g., what is the variability of the endpoint of interest
5. How robust is the endpoint under study?
6. Is it relevant to your overall hypothesis?

# What is the point of controls

For any experiment, the goal of the experimenter is to obtain data that can be interpreted in a manner to address the specific aims of the research. To achieve, this as many variables as possible that could undermine the interpretation of an experimental effect should be placed under experiment control.

# What is the point of controls

## **The chief antagonist of statistical reliability is “Chance”**

“Chance” may refer to the combined effects of uncontrolled variables. If such variables are controllable then chance is a simple excuse for sloppy experimentation.

If the uncontrolled variables are actually unknown then “chance” relates to “ignorance” and science attempts to unmask ignorance. Statistics are used to evaluate the degree of that ignorance.

# Controls to consider

1. Model Controls
2. Assay Controls
3. Experimental Controls

## Examples:

Background controls for genetically modified animals or cells

Identical sham controls that recapitulate exactly the experimental conditions

Identical vehicle controls for any dosing conditions

Age and sex matched controls

Assay positive and negative controls

Always concurrent controls and any experimental conditions being compared

# How to “control” for things you cannot control

1. Randomization where appropriate (random assign, random select)
2. Maintain “experimental design” across all assays
3. Replicate findings
4. Value of experimental biological replicates not technical replicates

# Data analysis

The overall goal of any data analysis is not only to determine an effect but also to determine the generality of the data.

How representative are the data?

To how many samples/subjects do they apply?

In this regard, examining only the mean  $\pm$  SD (SEM), regardless of the sample size, will not provide this information. Rather, identifying the distribution and the % of the sample that falls around the mean will provide information on the generality of the data.

# Data analysis

The Experimental Design defines:

How you can analyze the data

What questions you can ask with your data

Helps to determine the n size needed and the approach to obtaining that n size and replicates

If you do not establish a good experimental design you will waste your efforts as you will not be able to address your questions and in the worst case – no questions. Experimental design does not mean elaborate or costly study – but a study for which you can analyze your data.

# What to analyze

Discuss your project with a biostatistician prior to experiments

Multiple endpoints may be collected but not all may be relevant to the hypothesis

While assay controls are essential for QA they are not appropriate to include in the experimental analysis

Report null as well as positive effects as they apply to the hypothesis

# Reproduce your findings

Your scientific reputation is on the line

Remember that all authors are responsible for the accuracy of the data reported.

How to reproduce –

Direct reproduction of critical data

Systematic replication – rather than simply repeating the experiment, use the collected data as a basis for performing new experiments for additional related data e.g. if this then that; if this then not that

# Report your findings

Report methods in sufficient details to allow someone else to repeat your experiments.

Report all data – direct experimental data with supplementary material for any supporting data

Report statistical details for each analysis showing statistical significance (e.g. ANOVA – F value, degree of freedom, p value)

**Be Proud of Your Work!**



## **Part 3:**

# **Challenges Toxicologists Face in the Quest for Scientific Rigor**

*Alternatively, "How Systematic Review is a Platform for Rigor and Reproducibility"*

**Daniele Wikoff, PhD**

**ToxStrategies**

***Evidence Based Toxicology Collaboration***

***Toxicological Sciences, Regulatory Toxicology and Pharmacology***

# Assessing Rigor and Reliability In Reviews (or Risk Assessment, etc.)

*Challenge: how do we accommodate for rigor and reliability when assessing all available scientific evidence to inform decisions?*

***(One) Solution = Systematic Review***

## By definition, a Systematic Revue is:

- ◆ *A systematic review attempts to identify, appraise and synthesize all the empirical evidence that meets pre-specified eligibility criteria to answer a specific research question. Researchers conducting systematic reviews use explicit, systematic methods that are selected with a view aimed at minimizing bias, to produce more reliable findings to inform decision making. – Cochrane Colloquium, 2014*
- ◆ *A scientific investigation that focuses on a specific question and uses explicit, pre-specified scientific methods to identify, select, assess, and summarize the findings of similar but separate studies. – IOM, 2011*

# Benefits of Systematic Review

- ◆ Transparency (and reproducibility)
- ◆ Rigor
- ◆ Reliable and credible (reduce bias)
- ◆ *Inherently assess reproducibility*

*Does not fully eliminate subjectivity and scientific judgements*

# The Rigor... Example: IOM - Finding What Works in Health Care – Standards for Systematic Reviews

= 21 Standards and  
82 elements of  
performance

STANDARDS MARCH 2011


INSTITUTE OF  
OF THE N  
Advising the nation • Im

For more information visit [www.iom.edu](http://www.iom.edu)

## Finding What Works in Health Care

### Standards for Systematic Reviews

These standards are for systematic reviews of comparative effectiveness research of therapeutic medical or surgical interventions



#### Standards for Initiating a Systematic Review

**STANDARD 2.1**  
**Establish a team with appropriate expertise and experience to conduct the systematic review**

- 2.1.1 Include expertise in the pertinent clinical content areas
- 2.1.2 Include expertise in systematic review methods
- 2.1.3 Include expertise in searching for relevant evidence
- 2.1.4 Include expertise in quantitative methods
- 2.1.5 Include other expertise as appropriate

**STANDARD 2.2**  
**Manage bias and conflict of interest (COI) of the team conducting the systematic review**

- 2.2.1 Require each team member to disclose potential COI and professional or intellectual bias
- 2.2.2 Exclude individuals with a clear financial conflict
- 2.2.3 Exclude individuals whose professional or intellectual bias would diminish the credibility of the review in the eyes of the intended users

**STANDARD 2.3**  
**Ensure user and stakeholder input as the review is designed and conducted**

- 2.3.1 Protect the independence of the review team to make the final decisions about the design, analysis, and reporting of the review

**STANDARD 2.4**  
**Manage bias and COI for individuals providing input to the systematic review**

- 2.4.1 Require individuals to disclose potential professional or intellectual bias
- 2.4.2 Exclude input from individuals who would diminish the credibility of the review in the eyes of the intended users

**STANDARD 2.5**  
**Formulate the topic for the systematic review**

- 2.5.1 Confirm the need for a new review
- 2.5.2 Develop an analytic framework that links the health outcomes of interest and the clinical questions to be addressed in the review
- 2.5.3 Use a standard format to articulate the question of interest
- 2.5.4 State the rationale for each clinical question
- 2.5.5 Refine each question based on user input

**STANDARD 2.6**  
**Develop a systematic review protocol**

- 2.6.1 Describe the context from both a decision-maker and a patient perspective
- 2.6.2 Describe the study selection criteria (inclusion/exclusion)
- 2.6.3 Describe precisely when time points, interventions will be addressed
- 2.6.4 Describe the search strategy and relevant evidence
- 2.6.5 Describe the procedures for resolving disagreements
- 2.6.6 Describe the data extraction process
- 2.6.7 Describe the process of resolving disagreements about study selection and data extraction
- 2.6.8 Describe the approach to synthesizing individual studies

**STANDARD 2.6.9**  
Describe the method for evaluating the body of evidence, including the quantitative and qualitative synthesis strategies

**STANDARD 3.1**  
**Conduct a comprehensive search**

- 3.1.1 Work with a librarian or other trained professional to perform the search strategy
- 3.1.2 Design the search strategy to address the research question
- 3.1.3 Use an independent specialist to peer review the search strategy
- 3.1.4 Search bibliographic databases
- 3.1.5 Search citation indexes
- 3.1.6 Search literature cited in references
- 3.1.7 Update the search at a pace of generation of new research question
- 3.1.8 Search subject-specific databases are unlikely to be identified
- 3.1.9 Search regional bibliographic databases are unlikely to be identified

**STANDARD 3.2**  
**Take action to address potential reporting bias**

- 3.2.1 Search grey literature, trial registries, and other sources about studies
- 3.2.2 Invite researchers to register their studies and report results

**STANDARD 3.3**  
**Use an independent specialist to peer review the search strategy**

- 3.3.1 Use an independent specialist to peer review the search strategy

**STANDARD 3.4**  
**Document the search**

- 3.4.1 Provide a line-by-line description of the search strategy, including the date of every search for each database, web browser, etc.
- 3.4.2 Document the disposition of excluded studies, including reasons for their exclusion

**STANDARD 3.5**  
**Manage data collection**

- 3.5.1 At a minimum, use two or more reviewers, working independently, to extract critical data from each study. For each study, one individual could extract the data, and the second individual independently verify the accuracy and completeness. Establish a procedure for resolving discrepancies that give final decision-making power to the reviewer

**STANDARD 3.6**  
**Link publications from the same study to avoid including data from the same study more than once**

- 3.6.1 Link publications from the same study to avoid including data from the same study more than once

**STANDARD 3.7**  
**Use standard data extraction forms developed for the specific systematic review**

- 3.7.1 Use standard data extraction forms developed for the specific systematic review

**STANDARD 4.1**  
**Use a prespecified method to evaluate the body of evidence**

**NOTE:** The order of the standards does not indicate in which they are carried out.

- 4.1.1 For each outcome, systematically evaluate the following characteristics of the body of evidence:
  - Risk of bias
  - Consistency
  - Precision
  - Directness
  - Reporting bias
- 4.1.2 For bodies of evidence that include individual studies, also systematically evaluate the following characteristics for each outcome:
  - Dose-response association
  - Plausible confounding that would bias the observed effect
  - Strength of association
- 4.1.3 For each outcome specified in the protocol, use consistent language to characterize the confidence in the estimates of the effect of the intervention

**STANDARD 4.2**  
**Conduct a qualitative synthesis**

- 4.2.1 Describe the clinical and methodological characteristics of the included studies, including their size, inclusion or exclusion criteria, subgroups, timeliness, and other characteristics
- 4.2.2 Describe the strengths and limitations of the included individual studies and patterns across studies

### Standards for Reporting Systematic Reviews

#### STANDARD 5.1 Prepare final report using a structured format

- 5.1.1 Include a report title
- 5.1.2 Include an abstract
- 5.1.3 Include an executive summary
- 5.1.4 Include a summary written for the lay public
- 5.1.5 Include an introduction (rationale and objectives)
- 5.1.6 Include a methods section. Describe the following:
  - Research protocol
  - Eligibility criteria (criteria for including and excluding studies in the systematic review)
  - Analytic framework and key questions
  - Databases and other information sources used to identify relevant studies
  - Search strategy
  - Study selection process
  - Data extraction process
  - Methods for handling missing information
  - Information to be extracted from included studies
  - Methods to appraise the quality of individual studies
  - Summary measures of effect size (e.g., risk ratio, difference in means)
  - Rationale for pooling (or not pooling) results of included studies
  - Methods of synthesizing the evidence (qualitative and meta-analysis)
  - Additional analyses, if done, indicating which were prespecified

#### STANDARD 5.7 Include a results section. Organize the presentation of results around key questions. Describe the following (repeat for each key question):

- Study selection process
- List of excluded studies and reasons for their exclusion
- Appraisal of individual studies' quality
- Qualitative synthesis
- Meta-analysis of results, if performed (explain rationale for doing one)
- Additional analyses, if done, indicating which were prespecified
- Tables and figures

#### STANDARD 5.8 Include a discussion section. Include the following:

- Summary of the evidence
- Strengths and limitations of the systematic review
- Conclusions for each key question
- Gaps in evidence
- Future research needs

#### STANDARD 5.9 Include a section describing funding sources and COI

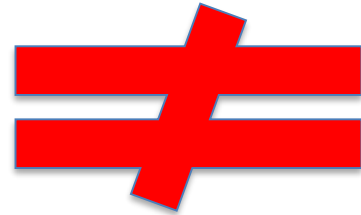
#### STANDARD 5.2 Peer review the draft report

- 5.2.1 Use a third party to manage the peer review process
- 5.2.2 Provide a public comment period for the report and publicly report on disposition of comments

#### STANDARD 5.3 Publish the final report in a manner that ensures free public access

# Common Misperception

**Systematic Search**



**Systematic Review**

# Common Components of SR

## Problem Formulation

- Scoping, scientific needs/objectives, feasibility
- Develop PECO question/statement and context

## Protocol Development

- Determine methods for selecting, appraising, and evaluating evidence
- Document methods (*a priori*)

## Identify Evidence Base

- Implement search strategy (syntax, databases, etc.)
- Screen and select studies via inclusion/exclusion criteria

## Individual Study Assessment

- Extract data
- Conduct critical appraisal for risk of bias (internal validity) and possibly other elements of study quality/relevance

## Body of Evidence Assessment

- Structured synthesis and integration –quantitative or qualitative
- *Considers reproducibility*

## Reporting

- Comprehensive documentation of approach, findings, and conclusions in a public forum

# Global Use of Systematic Review (By Authoritative Bodies)

## Review of EPA's Integrated Risk Information System (IRIS) Process



EFSA Journal 2010; 8(6):1637

### GUIDANCE OF EFSA

#### Application of systematic review methodology to food and feed safety assessments to support decision making<sup>1</sup>

EFSA Guidance for those carrying out systematic reviews

European Food Safety Authority

European Food Safety Authority

#### ABSTRACT

Systematic reviews are commonly used to synthesize evidence on a specific question. This document provides guidance on how to conduct a systematic review of food and feed safety. It covers the process from identifying the question to reporting the results. The document is intended for those who are responsible for food and feed safety assessments. It provides a framework for conducting systematic reviews and includes examples of key steps. The document is intended for those who are responsible for food and feed safety assessments. It provides a framework for conducting systematic reviews and includes examples of key steps.

#### KEY WORDS

Systematic review, food safety, feed safety

<sup>1</sup> On request from EFSA, Question No EFSA-Q-2007-0012  
<sup>2</sup> Correspondence: [ama@efsa.europa.eu](mailto:ama@efsa.europa.eu)  
Acknowledgement: EFSA wishes to thank methodology to food and feed safety assessments: Jon Deeks, Geoff Tranter, Rappo M O'Connor, Andrew Pullin, Andriana Rajafono, Jean-Lou Dorne, and Karin M. Ni Sirpa Kärenlampi.

Suggested citation: European Food Safety Authority (EFSA) (2010) Application of systematic review methodology to food and feed safety assessments to support decision making. EFSA Journal 2010; 8(6):1637. Online: [www.efsa.europa.eu](http://www.efsa.europa.eu)

© European Food Safety Authority, 2010

#### EVENT REPORT

APPROVED: 8 March 2018  
doi:10.2903/journal.efsa.2018.EN-1396

### EFSA Scientific Colloquium 23 – Joint European Food Safety Authority and Evidence-Based Toxicology Collaboration Colloquium

#### Evidence integration in risk assessment: the science of combining apples and oranges

25–26 October 2017  
Lisbon, Portugal

European Food Safety Authority

#### Abstract

In evidence-based scientific assessments, evidence synthesis is the step that occurs after collecting the data relevant to a clearly formulated research question and appraising the validity of the studies selected for the assessment, according to structured and pre-defined approaches. When studies are readily comparable, evidence synthesis is usually carried out through meta-analysis. In hazard assessment in chemical risk assessment (CRA), the process for combining evidence, 'evidence integration', is a recognised challenge as the underlying evidence bases are very diverse and not readily comparable (owing e.g. to varying degrees of validity and precision, diverse data types, different populations and species, models, end-points, routes of exposure, and evidence streams - human observational studies, experimental animal studies, in vitro and computational models data). The European Food Safety Authority (EFSA) and the Evidence-Based Toxicology Collaboration (EBTC) organised a Colloquium to develop a multistakeholder understanding of the best practices, challenges and research needs for evidence integration in CRA, with a focus on hazard identification and on combining multiple studies and end-points for dose-response modelling. The methods discussed included: qualitative methods for integrating evidence within- and across evidence streams; bias-adjusted meta-analysis; quantitative approaches to combine evidence across evidence streams; and quantitative approaches for combining multiple end-points and multiple studies for dose-response modelling. All these methods showed advantages and needs for further development, testing, validation and effective implementation. Support to this could be provided by: more published primary toxicological and epidemiological data; optimisation of study design; a shared primary data repository; the establishment of a community of knowledge of toxicologists, epidemiologists and statisticians. Equally, to be conducted soundly, evidence integration in CRA should be undertaken by multidisciplinary groups (toxicologists and methodologists knowledgeable of the various integration techniques). EFSA and EBTC will continue the collaboration towards the development, testing and validation of best practices for evidence-based CRA.

© European Food Safety Authority and Evidence-Based Toxicology Collaboration, 2018

**Key words:** Evidence synthesis, evidence integration, evidence-based, chemical risk assessment, toxicology, environmental health, hazard assessment, dose-response modelling

**Question number:** EFSA-Q-2017-00712

**Correspondence:** any enquires related to this output should be addressed to [scientific.colloquia@efsa.europa.eu](mailto:scientific.colloquia@efsa.europa.eu)

[www.efsa.europa.eu/publications](http://www.efsa.europa.eu/publications)

EFSA Supporting publication 2018:EN-1396



National Toxicology Program  
U.S. Department of Health and Human Services

### Handbook for Conducting a Literature-Based Health Assessment Using OHAT Approach for Systematic Review and Evidence Integration

#### OHAT Risk of Bias Rating Tool for Human and Animal Studies

#### INTRODUCTION

This document is written to outline a tool for evaluating individual study risk of bias or internal validity – the assessment of whether the design and conduct of a study compromised the credibility of the link between exposure and outcome (Higgins and Green 2011, IOM 2011, Viswanathan *et al.* 2012). The risk-of-bias rating tool presents a parallel approach to evaluating risk of bias in human and non-human animal studies to facilitate consideration of risk of bias across elements and across evidence streams with common terms and categories.

This tool was developed based on the most recent guidance from the Agency for Healthcare Research and Quality (Viswanathan *et al.* 2012, 2013), the Cochrane risk-of-bias tool for non-randomized studies of interventions (Sterne *et al.* 2014), Cochrane Handbook (Higgins and Green 2011), CLARITY Group at McMaster University (2013), SYRCLE's risk-of-bias tool for animal studies (Hooijmans *et al.* 2014), the Navigation Guide (Johnson *et al.* 2013, Koustas *et al.* 2013, Johnson *et al.* 2014, Koustas *et al.* 2014, Woodruff and Sutton 2014), comments from the public and technical advisors on draft methods and risk-of-bias instructions (NTP 2013d, c, b, a), staff at other federal agencies, and other sources (Downs and Black 1998, Genaidy *et al.* 2007, Dwan *et al.* 2010, Shamlivan *et al.* 2010, Shamlivan *et al.* 2011, Krauth *et al.* 2013, Wells *et al.* 2014).

For each study, risk of bias is assessed at the outcome level because certain aspects of study design and conduct may increase risk of bias for some outcomes and not others within the same study.



EPA Document# 760-P1-8001  
Office of Chemical Safety and  
Pollution Prevention

### APPLICATION OF SYSTEMATIC REVIEW IN TSCA RISK EVALUATIONS

MAY 2018

the IRIS Process  
Studies and Toxicology  
and Life Studies

RESEARCH COUNCIL  
NATIONAL ACADEMIES

NATIONAL ACADEMIES PRESS  
D.C.



#### TCEQ Guidelines for Systematic Review and Evidence Integration

Prepared by the  
Toxicology Division

White Paper  
December 20, 2017

TEXAS COMMISSION ON ENVIRONMENTAL QUALITY

# Challenge - Transitioning to Evidence-Based Toxicology (from evidence-based medicine)

- ◆ Different questions (e.g., comparative effectiveness of an intervention vs. characterization of hazard and risk)
  - *PECO vs. PICO*
    - **Population, Exposure/Intervention, Comparator, Outcome**
  - *Need flexibility; in toxicology risk assessment, often need more than yes/no*
- ◆ Different evidence base (i.e., RCTs vs. laboratory animal/observational epi studies)
  - *Need tools tailored to these types of evidence*
  - *Need methods that accommodate highly heterogeneous data*
  - *Need approaches to answer different questions*

# Example: Study Quality

*Not just relevance - multiple aspects of study validity are important*

## OHAT Risk of Bias Rating Tool for Human and Animal Studies

### INTRODUCTION

This document is written to outline a tool for the assessment of whether the design of an exposure and outcome (Higgins) of-bias rating tool presents a parallel animal studies to facilitate consideration with common terms and categories.

This tool was developed based on the Risk of Bias and Quality (Viswanathan *et al.* 2012, 2013) of interventions (Sterne *et al.* 2014), Cochrane Handbook for Systematic Reviews of Interventions (McMaster University (2013), SYRCL's Navigation Guide (Johnson *et al.* 2013), Woodruff and Sutton 2014), comment risk-of-bias instructions (NTP 2013d, c, and Black 1998, Genaidy *et al.* 2007, D Krauth *et al.* 2013, Wells *et al.* 2014).

For each study, risk of bias is assessed and the overall score may increase risk of bias for some studies.



United States  
Environmental Protection Agency

EPA Document# 740-P1-8001  
Office of Chemical Safety and  
Pollution Prevention

## APPLICATION OF SYSTEMATIC REVIEW IN TSCA RISK EVALUATIONS

MAY 2018

Table G-8. Metric Weighting Factors and Range of Weighted Metric Scores for *In Vitro* Toxicity Studies

Domain Number / Description	Metric Number/Description	Metric Score	Metric Weighting Factor	Weighted Score	Rationale (Table G - 14)
1. Test Substance	1. Test Substance Identity		2	0.0	
	2. Test Substance Source		1	0.0	
	3. Test Substance Purity		1	0.0	
2. Test Design	4. Negative and Vehicle Controls		2	0.0	
	5. Positive Controls		2	0.0	
	6. Assay Procedures		1	0.0	
	7. Standards for Test		1	0.0	
3. Exposure Characterization	8. Preparation and Storage of Test Substance		1	0.0	
	9. Consistency of Exposure Administration		1	0.0	
	10. Reporting of Concentrations		2	0.0	
	11. Exposure Duration		2	0.0	
	12. Number of Exposure Groups and Dose Spacing		1	0.0	
	13. Metabolic Activation		1	0.0	
4. Test model	14. Test Model		2	0.0	
	15. Number per Group		1	0.0	
5. Outcome Assessment	16. Outcome Assessment Methodology		2	0.0	
	17. Consistency of Outcome Assessment		1	0.0	
	18. Sampling Adequacy		1	0.0	
	19. Blinding of Assessors		1	0.0	
6. Confounding/ Variable Control	20. Confounding Variables in Test design and Procedures		1	0.0	
	21. Outcomes Unrelated to Exposure		1	0.0	
	22. Data Analysis		1	0.0	
	23. Data Interpretation		1	0.0	
	24. Cytotoxicity Data		1	0.0	
7. Data Presentation and Analysis	25. Reporting of Data		1	0.0	
	Sum (if all metrics scored) =		0		
Range of Overall Scores, where					
Overall Score = Sum of Weighted Scores/Sum of Metric Weighting Factor					
High		Medium		Low	
≥1 and <1.7		≥1.7 and <2.3		≥2.3 and ≤3	

Is my study a high-quality study?

# Systematic Review in Risk Assessment

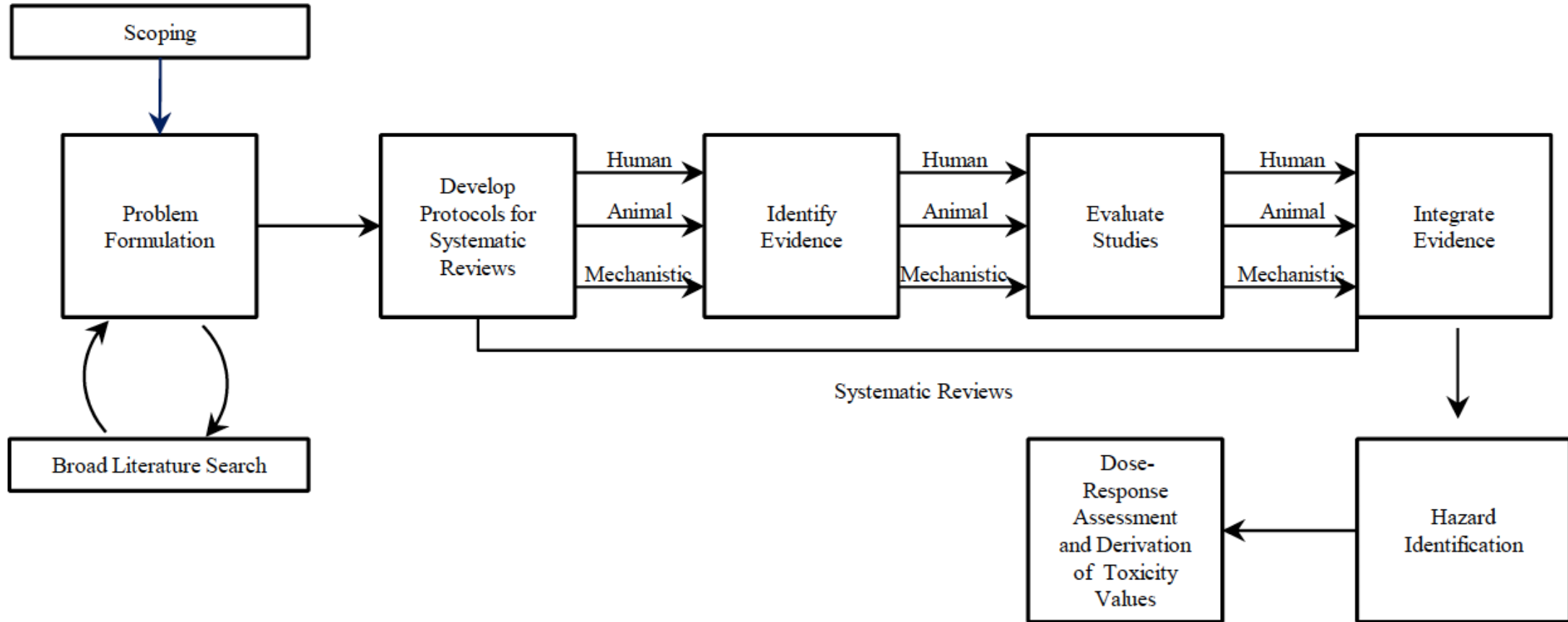


Figure S-1, NAS, 2014 ; <https://www.nap.edu/catalog/18764/review-of-epas-integrated-risk-information-system-iris-process>

# How might systematic review impact you?

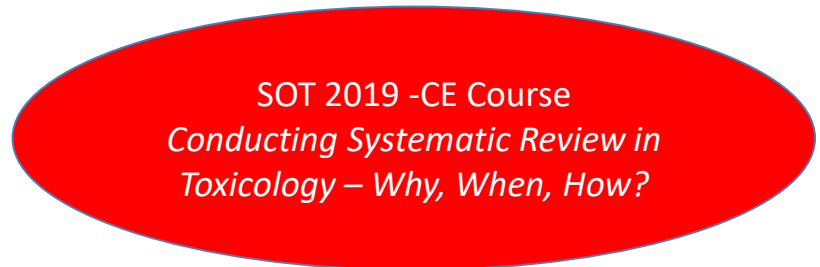
- ◆ Important to understand the global movement towards use of systematic review (and other evidence-based methods)
  - ❖ Consider these approaches when designing research questions
  - ❖ Peer-review journal standards, etc., are evolving
- ◆ Important to understand that studies will be critically appraised for reliability (*and reproducibility*)
  - ❖ Highlights the importance of complete reporting and consideration of internal validity aspects (bias) during study design and conduct
- ◆ Peer-review of systematic reviews requires understanding of the method (as well as the topic of the underlying evidence base)

# Actions You Can Take

*From the perspective of a journal editor...*

- ◆ Don't underestimate the value of planning an experiment
- ◆ Consider structured abstracts
- ◆ Ensure complete reporting (allow for reproducibility)
  - ❖ Heavy emphasis on completeness of methods
  - ❖ Report all findings (including null findings)
  - ❖ Utilize repositories (e.g., Open Science Framework, Zenodo, Dryad) and supplemental material for raw data
- ◆ Cover all aspects of validity
  - ❖ (see, for example, tools from SciRAP <http://www.scirap.org/>)
- ◆ Stick to the science

# Selected Reading: SR in Toxicology



SOT 2019 -CE Course  
Conducting Systematic Review in  
Toxicology – Why, When, How?

REVIEW ARTICLE

## A primer on systematic reviews in toxicology

Sebastian Hoffmann<sup>1,13</sup> · Rob B. M. de Vries<sup>2</sup> · Martin L. Stephens<sup>1</sup> · Nancy B. Beck<sup>3</sup> · Hubert A. A. M. Dirven<sup>4</sup> · John R. Fowle III<sup>5</sup> · Julie E. Goodman<sup>6</sup> · Thomas Hartung<sup>7</sup> · Ian Kimber<sup>8</sup> · Manoj M. Lahu<sup>9</sup> · Kristina Thayer<sup>10</sup> · Paul Whaley<sup>11</sup> · Daniele Wikoff<sup>2</sup> · Katya Tsaouli<sup>10</sup>

Received: 13 April 2017 / Accepted: 4 May 2017  
© The Author(s) 2017. This article is an open access publication

**Abstract** Systematic reviews, pioneered in medicine, provide a transparent, methodologically reproducible means of summarizing the evidence on a precisely framed research question. This approach is now being adopted in toxicology as a well-established approach in many other fields. Systematic reviews are receiving increasing attention as a potential tool for answering toxicology research questions. A larger framework of evidence-based toxicology is needed to address the challenges and obstacles of, as well as the advantages of, using this approach. In this paper, we describe the emerging and adopting systematic reviews in toxicology, and the steps being explored. To provide the toxicologist with a starting point for conducting or undertaking a systematic review, we herein summarize available resources from various fields of application. We discuss the systematic review process in ten steps, starting with planning the review, identifying the question, and writing and publishing the review, concluding with interpretation and recommendations. We have identified the specific methods and tools used in the conduct of toxicological questions and have summarized the available resources.

✉ Katya Tsaouli

FORUM ARTICLE

## The Emergence of Systematic Review in Toxicology

Martin L. Stephens<sup>a,1</sup> Kellyn Betts<sup>b</sup> Nancy B. Beck<sup>c</sup> Vincent Cogliano<sup>d</sup>

Kay Dickersin<sup>e</sup> Susan M. Gorman<sup>f</sup> Thomas Hartung<sup>g</sup> Roberta W. Scherer<sup>h</sup>

<sup>a</sup>Johns Hopkins Center for Communications Programs, Technology Writer, Takoma Park, MD  
<sup>d</sup>US Environmental Protection Agency, Office of Research and Development, Public Health, Baltimore, MD  
<sup>e</sup>Washington University in St. Louis, St. Louis, MO  
<sup>f</sup>CAAT-Europe, Germany  
<sup>g</sup>Assessment and Translation Research Center, Environmental Health Sciences Division, European Food Safety Authority, Parma, Italy



### Implementing systematic review techniques in chemical risk assessment: Challenges, opportunities and recommendations

Paul Whaley<sup>a</sup>, Crispin Halsall<sup>a\*</sup>, Marlene Agerstrand<sup>b</sup>, Elisa Aiassa<sup>d</sup>, Diane Benford<sup>c</sup>, Gary Bilotta<sup>e</sup>, David Coggon<sup>f</sup>, Chris Collins<sup>g</sup>, Ciara Dempsey<sup>h</sup>, Raquel Duarte-Davidson<sup>g</sup>, Rex Fitzgerald<sup>h</sup>, Malika Galay-Burgos<sup>g</sup>, David Gee<sup>i</sup>, Sebastian Hoffmann<sup>1</sup>, Juleen Lam<sup>1</sup>, Toby Lasserson<sup>1</sup>, Len Levy<sup>10</sup>, Steven Lipworth<sup>11</sup>, Sarah Mackenzie Ross<sup>12</sup>, Olwenn Martin<sup>1</sup>, Catherine Meads<sup>13</sup>, Monika Meyer-Baron<sup>14</sup>, James Miller<sup>1</sup>, Camilla Pease<sup>15</sup>, Andrew Rooney<sup>1</sup>, Alison Sapiets<sup>16</sup>, Gavin Stewart<sup>17</sup>, David Taylor<sup>18</sup>

<sup>a</sup> Lancaster Environment Centre, Lancaster University, Lancaster LA1 4YW, UK  
<sup>b</sup> Department of Environmental Science and Analytical Chemistry, Stockholm University, SE-106 91, Stockholm, Sweden  
<sup>c</sup> Food Standards Agency, Aviston House, 125 Kingsway, London WC2B 6NH, UK  
<sup>d</sup> Assessment and Methodological Support Unit, European Food Safety Authority, Via Carlo Magno 1/a 43126, Parma, Italy  
<sup>e</sup> Aquatic Research Centre, University of Brighton, Lewes Road, Brighton BN2 4GJ, UK  
<sup>f</sup> MRC Lifecourse Epidemiology Unit, University of Southampton, MRC Lifecourse Epidemiology Unit, Southampton General Hospital, Southampton SO16 0YD, UK  
<sup>g</sup> Centre for Radiation, Chemical and Environmental Hazards, Public Health England, Harwell Science and Innovation Campus, Didcot, Oxfordshire OX11 0BQ, UK  
<sup>h</sup> Swiss Centre for Applied Human Toxicology, University of Basel, Muesmattenstrasse 64, 4055 Basel, Switzerland  
<sup>i</sup> Institute for the Environment, Health and Society, Brunel University London, Kingston Lane, Uxbridge UB8 3PH, UK  
<sup>1</sup> Evidence-Based Toxicology Collaboration (EBTC), Sternbergstr. 15, 33106 Paderborn, Germany  
<sup>10</sup> University of California San Francisco, Program on Reproductive Health and the Environment, San Francisco, CA, USA  
<sup>11</sup> Cochran Editorial Unit, Cochran Central Executive, St Albans House, 57-59 Highgate Road, London NW11 4JX, UK  
<sup>12</sup> Institute of Environment, Health, Risks and Futures, School of Energy, Environment and Agrifood, Cranfield University, Cranfield, Bedfordshire MK43 0AL, UK

EDITORIAL

## Systematic Reviews in Toxicology

Daniele S. Wikoff<sup>a</sup> and Gary W. Miller<sup>1,†,‡</sup>

<sup>a</sup>ToxStrategies, Inc., Asheville, North Carolina 28804; <sup>1</sup>Editor-in-Chief, Toxicological Sciences; and <sup>2</sup>Department of Environmental Health, Rollins School of Public Health, Emory University, Atlanta, Georgia 30322

<sup>†</sup>To whom correspondence should be addressed at: Department of Environmental Health, Rollins School of Public Health, Emory University, Atlanta, GA 30322, USA. E-mail: gary.miller@toxology.org

As a field, toxicology generates a massive amount of complex data. The increased amount of –omic scale data, combined with the myriad of models, from organs on a chip to cell culture and whole animals, has created a dizzying challenge. It has become exceedingly difficult for a scientist to survey the literature and develop a sense of what the data are saying. At Toxicological Sciences we believe the careful review of the literature is important to help direct future research. Contemporary systematic reviews, instituted (Johnson and Miller, 2014) to help identify trends and emerging topic areas. These articles do not focus on recent findings to capture the upward spiral of new areas. Yet, there is often a need to objectively synthesize data already out there to determine overall findings and data gaps; this is especially important for regulatory and for topics for which there have been conflicting findings.

The concepts presented here draw upon the collective expertise of the field.

By definition, systematic review is a method for answering specific research questions. It uses a predefined, multistep process to identify, select, critically assess, and synthesize evidence from scientific studies to reach a conclusion (JOM, 2011; Rooney et al., 2014). The term ‘systematic review’ implies a weight of evidence that is more robust than a narrative review. Systematic reviews, often referred to as meta-analyses, can offer advantages in providing a synthesis of experts to provide experiential and topic-specific information. In many fields, however, systematic reviews have been preferred platform for providing comprehensive and transparent reviews (JOM, 2011). Thus, it is not surprising that a systematized approach that subjects the literature to the same rigor that is applied to individual experiments. Long used in the fields of medicine and other disciplines, the systematic review is gaining significance in the field of toxicology (Stephens et al., 2016). This is highlighted by integration of this tool by a number of agencies worldwide, including the European Food Safety Authority (EFSA, 2010), National Institute of Environmental Health Sciences (Rooney et al., 2014), U.S. Environmental Protection Agency (Bahaduri and Thayer, 2018) and Drug Administration (USFDA, 2009), and the World Health Organization (WHO, 2014). Systematic reviews have been the subject of a variety of symposia and workshops at the Society of Toxicology annual meeting, the Toxicology Forum, the Society for Risk Analysis, and the International Colloquium (EFSA, 2017; NTP, 2017).

Preface

### Assuring high-quality evidence reviews for chemical risk assessment: Five lessons from guest editing the first environmental health journal special issue dedicated to systematic review

While systematic review (SR), the rigorous methodology for selecting, appraising and synthesising existing evidence in order to answer a research question, may not yet be mainstream among environmental scientists and toxicologists, interest in the methods and what they may bring to chemical risk research is growing rapidly and is evident in an exponential increase in publications over the last 20 years (Fig. 1). Mirroring the rapid growth of a nascent literature is the proliferation of initiatives, many of which are collaborative, seeking to extend the conduct of systematic reviews to pre-clinical research and laboratory animal experimentation. These include the Systematic Review Centre for Laboratory Animal Experimentation<sup>1</sup> (SYRACLE) and the Collaborative Approach to Meta-Analysis and Review of Animal Data from Experimental Studies<sup>2</sup> (CAMARADES), while efforts to apply SR methods to

tools for evidence-based decision-making and undermining the case for using SR methods to synthesise evidence in CRA. With the issue of quality assurance in mind we have drawn up a number of lessons which, while perhaps common knowledge in other fields, have been reinforced for us while editing this Special Issue. The lessons are aimed at SR authors, reviewers and, importantly, journal editors who are being faced with an increasing number of manuscripts that purport to be systematic reviews. We believe this is the first Special Issue dedicated to systematic review published by an environmental health journal. In spite of the inevitable imperfections this entails, we hope the reader agrees this Special Issue has been a success. We would like to thank all the authors, peer reviewers and funders who contributed to this Special Issue and

© The Author(s) 2018. Published by Oxford University Press  
All rights reserved. For permissions, please e-mail: journals.permissions@oup.com

**Thank you!**

*See you in Anaheim...!*

