The Navigation Guide: The Theory and Practice of Systematic Reviews in Environmental Health Sciences

Patrice Sutton, MPH  
Academic Coordinator  
University of California, San Francisco  
Program on Reproductive Health and the Environment  
Patrice.Sutton@ucsf.edu

Paula Johnson, PhD, MPH  
Research Scientist  
Lead, California Safe Cosmetics Program  
California Department of Public Health  
Paula.Johnson@cdph.ca.gov

California Department of Toxic Substances Control  
ELC Seminar Series  
April 22, 2015
NO CONFLICT TO DECLARE
Exposure Is Everywhere Everyday
Federal reserve data on chemical production is only offered as relative production, which is unit-less. A specific reference year is chosen and values are calculated relative to that year’s production. In this particular data set 2007 is the reference year and is assigned a value of 100.

Data from: U.S. Federal Reserve Board, Division of Research and Statistics
U.S. Chemical Production Volume Compared to Population

9,500,000,000,000 pounds of chemicals

- 30,000 pounds of chemicals per person

U.S. POPULATION: 313,000,000

US EPA CDR Fact Sheet: Chemical Snapshot, June 2014. The total reported (domestically manufactured and imported) for 2012.
Industrial Chemicals in Virtually Every U.S. Pregnant Woman
State of the Science of Endocrine Disrupting Chemicals - 2012

Edited by Åke Bergman, Jerrold J. Heindel, Susan Jobling, Karen A. Kidd and R. Thomas Zoeller
The status quo

... “to a disturbing extent, babies are born pre-polluted.”
Sir Austin Bradford Hill

incompleteness of science …

“does not confer upon us a freedom to ignore the knowledge we already have, or to postpone the action that it appears to demand at a given time”


http://www.bmj.com/content/bmj/305/6868/1521.full.pdf
Acting on the science to prevent harm
Reflection on Theoretical Approaches

Psychoanalysis:

Psychoanalytic theory maintains a foundational belief in the unconscious, which consists of thoughts, feelings, and past events that are kept out of awareness. The unconscious experience significantly impacts daily functioning, especially in regard to a person’s character and personality. Psychoanalytic therapy aims to reconstruct members’ character by making unconscious thoughts conscious through analyzing the transference and defenses.

In contrast to psychoanalytic theory, solution-oriented thought integrates insight into behavior development, and so non-conscious decisions can be corrected. Based on this, I will always relate Freud’s foundational concepts such as the unconscious mind and the importance of dreams to contemporary perspectives.

In addition, the emphasis on comprehension in my mind, however, psychoanalysis is also

of great importance, one of the key elements in psychoanalysis is the analysis of the unconscious mind, and it is not practical to think that all thoughts and feelings are kept out of awareness.
He Said

She Said
Clinical sciences have faced and addressed these same challenges

Thank you Lisa Bero
A Comparison of Results of Meta-analyses of Randomized Control Trials and Recommendations of Clinical Experts

Treatments for Myocardial Infarction

Elliott M. Antman, MD; Joseph Lau, MD; Bruce Kupelnick; Frederick Mosteller, PhD; Thomas C. Chalmers, MD

Objective.—To examine the temporal relationship between accumulating data from randomized control trials of treatments for myocardial infarction and the recommendations of clinical experts writing review articles and textbook chapters.

Data Sources.—(1) MEDLINE search from 1966 to present; search terms used were myocardial infarction, clinical trials, multicenter studies, double-blind method, meta-analysis, and the text word “random;”; (2) references from pertinent articles and books; and (3) all editions of English-language general medical texts and manuals and review articles on treatment of myocardial infarction.

Study Selection.—Randomized control trials of therapies for reducing the risk of dying from myocardial infarction.

SHORTENING the time between medical research discoveries and clinical implementation of new technologies by practicing physicians has been a concern of the American public since Congress established the Heart, Cancer, and Stroke Program over 25 years ago. An undesirable lag still exists, and overcoming it is one of the goals of the newly formed Agency for Health Care Policy
Evidence Based Medicine

Aims to apply the best available evidence gained from the scientific method to clinical decision making.

THE COCHRANE COLLABORATION®
The Convergence of SCIENCE AND GOVERNANCE

Research, Health Policy, and American States

Daniel M. Fox
But Evidence Based Medicine Methodologies Are Not Directly Transferable to Environmental Science!
EXHIBIT 1

Streams Of Evidence For Chemical Toxicity Assessment In Clinical And Environmental Health Sciences

Clinical

- New pharmaceutical developed
- In vitro and in vivo toxicity testing
- Human experimental studies (randomized controlled trials)
- Enters marketplace and clinic
  - Post-exposure observation studies

Environmental

- Chemicals introduced prior to 1976 (N = 62,000)
- New synthetic chemical developed
- Limited assessment by EPA
- Enters marketplace and homes, schools, workplaces, communities, and consumer goods
  - Ad hoc post-exposure observational studies
  - Ad hoc in vitro and in vivo toxicity testing


Program on Reproductive Health and the Environment
Program on Reproductive Health and the Environment

Navigation Guide Work Group
Overview of Navigation Guide Systematic Review Methodology

1. Specify Study Question
   Is human environmental exposure to a chemical a reproductive health risk?

2. Select Evidence

3. Rate Quality & Strength of the Evidence:
   Strength of Evidence in Non-Human Systems
   - Sufficient
   - Limited
   - Inadequate
   - Evidence of lack of toxicity
     - Known to be Toxic to Human Reproduction
     - Probably Toxic
     - Possibly Toxic
     - Not Classifiable
   (Strength of human & non-human evidence are combined into 1 of 5 possible strength of evidence summary statements)

4. Grade Strength of Recommendation:
   Strength of Evidence (from Step 3 above)
   Exposure
   - High
     - Known to be Toxic: S
     - Probably Toxic: S
     - Possibly Toxic: S
     - Not Classifiable: S
   - Medium
     - Known to be Toxic: S
     - Probably Toxic: S
     - Possibly Toxic: D
     - Not Classifiable: D
   - Lower
     - Known to be Toxic: D
     - Probably Toxic: D
     - Possibly Toxic: D
     - Not Classifiable: D

   Is a Less Toxic Alternative Available?
   Patient Values and Preferences
   Strong or Discretionary Recommendation

No Recommendation needed
Systematic review approach for each evidence stream

Human Data

“PECO” Statement  Systematic search  Select Studies  Extract Data & Data Analysis  Rate Quality of Evidence  Rate the Strength of Evidence

Overall Conclusion

Non Human Data

“PECO” Statement  Systematic search  Select Studies  Extract Data & Data Analysis  Rate Quality of Evidence  Rate Strength of Evidence
The Navigation Guide Systematic Review Methodology: A Rigorous and Transparent Method for Translating Environmental Health Science into Better Health Outcomes

Tracey J. Woodruff and Patrice Sutton

Program on Reproductive Health and the Environment, University of California, San Francisco, Oakland, California, USA

BACKGROUND: Synthesizing what is known about the environmental drivers of health is instrumental to taking prevention-oriented action. Methods of research synthesis commonly used in environmental health lag behind systematic review methods developed in the clinical sciences over the past 20 years.

OBJECTIVES: We sought to develop a proof of concept of the “Navigation Guide,” a systematic and transparent method of research synthesis in environmental health.

DISCUSSION: The Navigation Guide methodology builds on best practices in research synthesis in evidence-based medicine and environmental health. Key points of departure from current methods of expert-based narrative review prevalent in environmental health include a prespecified protocol, standardized and transparent documentation including expert judgment, a comprehensive search strategy, assessment of “risk of bias,” and separation of the science from values and preferences. Key points of departure from evidence-based medicine include assigning a “moderate” quality rating to human observational studies and combining diverse evidence streams.

CONCLUSIONS: The Navigation Guide methodology is a systematic and rigorous approach to research synthesis that has been developed to reduce bias and maximize transparency in the evaluation of environmental health information. Although novel aspects of the method will require further development and validation, our findings demonstrated that improved methods of research synthesis under development at the National Toxicology Program and under consideration by the U.S. Environmental Protection Agency are fully achievable. The institutionalization of robust methods of systematic and transparent review would provide a concrete mechanism for linking science to timely action to prevent harm.


Address correspondence to T.J. Woodruff, UCSF Program on Reproductive Health and the Environment, 1330 Broadway, Suite 1135, Oakland, CA 94612 USA. Telephone: (510) 350-1241. E-mail: woodruff@obgyn.ucsf.edu

We are indebted to D. Atchley, D. Axelrad, L. Bero, P. Johnson, E. Koustas, and J. Lam for providing invaluable comments and suggestions on this com-
Protocol

UCSF Program on Reproductive Health and the Environment

Navigation Guide Protocol for Rating the Quality and Strength of Human and Non-Human Evidence
December 5, 2012
Photo credit: http://water.usgs.gov/nasqan/sample_collection_methods.html
Risk of Bias

Characteristics of a study that can introduce systematic errors in the magnitude or direction of the results

Higgins and Green 2011
Rate Risk of Bias for Each Individual Study

<table>
<thead>
<tr>
<th>Domains</th>
<th>Determinations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recruitment strategy</td>
<td>Low risk</td>
</tr>
<tr>
<td>Blinding</td>
<td>Probably low risk</td>
</tr>
<tr>
<td>Exposure assessment</td>
<td>Probably high risk</td>
</tr>
<tr>
<td>Confounding</td>
<td>High risk</td>
</tr>
<tr>
<td>Incomplete outcome data</td>
<td></td>
</tr>
<tr>
<td>Selective reporting</td>
<td></td>
</tr>
<tr>
<td>Conflict of interest</td>
<td></td>
</tr>
<tr>
<td>Other bias</td>
<td></td>
</tr>
</tbody>
</table>

(for each risk of bias domain)
Pharmaceutical industry sponsorship and research outcome and quality: systematic review
Joel Lexchin, Lisa A Bero, Benjamin Djulbegovic, Otavio Clark

Abstract

Objective To investigate whether funding of drug studies by the pharmaceutical industry is associated with outcomes that are favourable to the funder and whether the methods of trials funded by pharmaceutical companies differ from the methods in trials with other sources of support.

Methods Medline (January 1966 to December 2002) and Embase (January 1980 to December 2002) searches were supplemented with material identified in the references and in the authors’ personal files. Data were independently abstracted by three of the authors and disagreements were resolved by consensus.

Results 30 studies were included. Research funded by drug companies was less likely to be published than research funded by other sources. Studies sponsored by pharmaceutical companies were more likely to have outcomes favouring the sponsor than were studies with other sponsors (odds ratio 4.05; 95% confidence interval 2.98 to 5.51; 18 comparisons). None of the 13 studies that analysed methods reported that studies funded by industry was of

Conclusion Systematic bias favours products which are made by the company funding the research. Explanations include the selection of an inappropriate comparator to the product being investigated and publication bias.
“... the biggest threat to [scientific] integrity [is] financial conflict of interest”

Drummund Rennie
Deputy Editor (West), JAMA
Rate the Quality of the Body of Evidence

- High: Animal evidence
- Moderate: Human evidence
- Low

Program on Reproductive Health and the Environment

University of California
San Francisco
Rate the Quality of the Body of Evidence

Factors for downgrading/upgrading evidence were derived directly from factors used in GRADE and Cochrane

Quality of Evidence

across all studies.

Downgrade Criteria
- Risk of bias across studies
- Indirectness
- Inconsistency
- Imprecision
- Publication bias

Upgrade Criteria
- Large magnitude of effect
- Dose response
- All possible confounding accounted for

Rating
(based on all quality criteria)
- High quality
- Moderate quality
- Low quality
Rating Strength of Evidence

CONSIDERATIONS
1. What is the quality of the data?
2. What is the direction of the effect?
3. What is our confidence in the effect?
4. Are there other compelling attributes of the data that may influence certainty?

Rating Strength of Evidence
- Sufficient
- Limited
- Inadequate
- Evidence of lack of toxicity
Integration of each evidence stream

- **Sufficient**
  - Known to be toxic
  - Probably toxic
  - Possibly toxic
  - Not classifiable

- **Limited**
  - Possibly toxic

- **Inadequate**
  - Possibly toxic

- **Sufficient**
  - Known to be toxic

- **Limited**
  - Possibly toxic

- **Inadequate**
  - Possibly toxic
Implementing Systematic Review at the National Toxicology Program: Status and Next Steps

Linda S. Birnbaum, Kristina A. Thayer, John R. Bucher, Mary S. Wolfe

National Institute of Environmental Health Sciences, National Institutes of Health, Department of Health and Human Services, Research Triangle Park, North Carolina, E-mail: bucher@niehs.nih.gov

http://dx.doi.org/10.1289/ehp.1306711 [online 01 April 2013]

Systematic Review and Evidence Integration for Literature-Based Environmental Health Science Assessments

Andrew A. Rooney, Abe L. Boyles, Mary S. Wolfe, John R. Bucher, and Kristina A. Thayer

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

January 9, 2015
Office of Health Assessment and Translation (OHAT)
Division of the National Toxicology Program
National Institute of Environmental Health Sciences

## Comparison of Narrative reviews and Navigation Guide/OHAT

<table>
<thead>
<tr>
<th>Reference</th>
<th>Specify study question</th>
<th>Specify inclusion/exclusion criteria</th>
<th>Conduct reproducible search</th>
<th>Assess Risk of Bias</th>
<th>Data analysis and/or meta-analyses</th>
<th>Summary of findings table</th>
<th>Assess quality of body of evidence</th>
<th>Integrate evidence streams</th>
</tr>
</thead>
<tbody>
<tr>
<td>Post et al 2012</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lindstrom et al 2011</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stahl et al 2011</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White et al 2011</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Steenland et al 2010</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DeWitt et al 2009</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Olsen et al 2009</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jensen and Leffers 2008</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lau et al 2007</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Butenhoff et al 2004</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kennedy et al 2004</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lau et al 2004</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hekster et al 2003</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kudo and Kawashima 2003</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Navigation, OHAT/NTP</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
“…systematic-review standards provide an approach that would substantially strengthen the IRIS process…” NAS 2014

“EPA should consistently use a more systematic approach to evaluating the literature ……….” NAS 2014
UCSF: Proof of Concept Documented in 5 Case Studies of Applying the Navigation Guide

1. PFOA and fetal growth (published)
2. Maternal glomerular filtration rate (GFR) and fetal growth (published)
3. Triclosan and reproductive & developmental health outcomes (in preparation)
4. Air pollution and Autism Spectrum Disorder (in progress)
5. PBDEs and Attention Deficit Disorder and IQ (in progress)
### Strengths

- Permits action on available data
- Systematic and transparent
- Based on empirically-proven methods
- Can identify evidence gaps for future work
- Can support identification of safer alternatives
- Separates science from values and preferences

### Limitations

- Analysis limited to available data
- Novel parts of methodology need validation
- Further definition of moving from quality of evidence to strength of evidence
- Does not address non-scientific barriers to prevention-oriented action
Case Study: Triclosan and DART

Source: fda.gov
Navigation Guide systematic review methodology

“PECO” Statement
- Draft protocol with PECO (Population, Exposure, Comparator, Outcome)
- Review protocol
- Define inclusion/exclusion criteria
- Define quality and strength criteria

Systematic search
- Formulate search terms
- Conduct literature search

Select Studies
- Set up forms for screening
- Conduct title/abstract, then full text screening according to criteria
- Reconcile any differences between screeners
- Conduct search of reference and citation lists

Extract Data & Data Analysis
- Extract data from included studies
- Contact study authors for any additional data or info.
- Rate risk of bias of individual studies
- Summarize data for group
- Devise data analysis plan
- Conduct data analysis

Rate Quality of Evidence
- Review data summaries
- Rate overall quality across studies according to criteria
- Reconcile any differences between coauthors
- Summarize responses of coauthors

Rate the Strength of Evidence
- Review data summaries
- Rate overall strength of evidence according to criteria
- Reconcile any differences between coauthors
- Summarize responses of coauthors

Conclusion
Integration of each evidence stream

Human Data

Non Human Data

Overall Conclusion
Specify the Study Question

*Does exposure to triclosan have adverse effects on human development or reproduction?*
PECO

Participants
Exposure
Comparator
Outcome
Participants:

Humans or animals (whole organism studied during the reproductive or developmental time period, tissue, organ, cell line or components), or computer models of humans or animals.
Participants:
Humans or animals (whole organism studied during the reproductive or developmental time period, tissue, organ, cell line or components), or computer models of humans or animals.

Exposure:
Developmental - Pre-conception (exposure of either or both parents or, if relevant, preceding generations), prenatal (exposure of pregnant female and/or directly of fetus), or postnatal (until the time of sexual maturation) exposure, by any route, to triclosan.

Reproductive - Exposure to triclosan at any time preceding assessment of reproductive outcome.
Participants:
Humans or animals (whole organism studied during the reproductive or developmental time period, tissue, organ, cell line or components), or computer models of humans or animals.

Exposure:
Developmental - Pre-conception (exposure of either or both parents or, if relevant, preceding generations), prenatal (exposure of pregnant female and/or directly of fetus), or postnatal (until the time of sexual maturation) exposure, by any route, to triclosan.
Reproductive - Exposure to triclosan at any time preceding assessment of reproductive outcome.

Comparator:
Comparable populations or subjects (humans, non-human, tissues, organs, cell lines or components) exposed to vehicle-only treatment or lower levels of triclosan than the more highly exposed subjects.
Outcome:
Reproductive effects: alterations in hormone levels; effects on male or female gametes (production, maturation, or transport), fertility, fecundity, estrous cycles, menstrual cycles, endocrine function, sexual behavior, gestation, parturition, lactation, age at puberty or reproductive senescence or menopause; pregnancy complications; increased pregnancy wastage; or alterations in size, morphology, or function of reproductive organs.

Developmental effects: fetal loss or resorption, stillbirth, neonatal or subsequent mortality, alterations in sex ratio, altered fetal or postnatal growth, structural malformations and variations, altered gestation length, functional deficits such as alterations in behavior, and morbidity. In addition to effects of prenatal exposure during all or any part of gestation, developmental toxicity can result from:

- Pre-conception exposure of parental or previous generations causing genetic mutation or epigenetic changes, which in turn affect development of unexposed offspring.
- Postnatal exposure when the developing offspring is more susceptible to adverse effects of the toxic agent than is the mature animal:
  - Qualitatively: Effect not seen in similarly-exposed adults
  - Quantitatively: Effect seen at lower doses, or to a greater extent, in immature organisms than in adults
Select the Evidence

9,485 records identified through database searching

5,203 duplicates removed

4,282 titles/abstracts screened

4,034 studies excluded as met one or more exclusion criteria

Excluded as rodent studies were available:
1. 24 whole animal invertebrate studies
2. 13 in vitro studies
3. 13 fish studies
4. 7 amphibian studies

248 full-text articles assessed for

159 studies excluded as met one or more exclusion criteria

248 full-text articles assessed for

24 studies excluded after full-text assessment

6 human studies

19 relevant human and animal studies

13 rodent studies

13 rodent studies

8 rodent studies reporting hormone concentration outcomes included in qualitative analysis

6 studies (10 separate datasets) included in quantitative analysis (meta-analysis)
Reference Screening

4282 title/abstracts retrieved

4034 refs excluded

~18 hours reviewer time

248 full-texts screened

197 refs excluded

~6 hours reviewer time

51

Title/abstract exclusion criteria:
1. No original data
2. No triclosan exposure
3. No triclosan-related toxicity (adverse effects) data
4. Other reason (explanation required)

Full-text exclusion criteria:
1. No original data
2. Developmental and/or reproductive toxicity (DART) not reported
3. Triclosan exposure doesn't precede DART outcomes
4. No comparator group
5. Other reason (explanation required)

Note: DART outcomes of interest include, but are not limited to, outcomes described in U.S. EPA guidelines on risk assessment of developmental and reproductive toxicity
Refid: 46, Association of exposure to phenols and idiopathic male infertility

Attachments
46 Chen et al.pdf

Widespread human exposure to phenols has been documented recently, and some phenols which are potential endocrine disruptors have demonstrated adverse effects on male reproduction in animal and in vitro studies. However, implications about exposure to phenols and male infertility are scarce in humans. Case-control study of 677 idiopathic infertile men and 713 fertile controls was conducted. Urinary levels of bisphenol A, benzophenone-3, pentachlorophenol, triclosan, 4-tol-octylphenol (4-t-OP), 4-n-octylphenol (4-n-OP) and 4-nonylphenol (4-n-NP) and semen parameters were measured. After multivariate adjustment, we found 4-t-OP, 4-n-OP and 4-n-NP exposure was associated with idiopathic male infertility (p-value for trend: <0.0001, 0.014 and 0.001, respectively). Aside from these associations, 4-t-OP and 4-n-NP exposure was also associated with idiopathic male infertility with abnormal semen parameters. Moreover, we observed significant associations between sum alkyphenols (APs) exposure and idiopathic male infertility. There were no relationships between exposure to other phenols and idiopathic male infertility in the present study. Our study provides the first evidence that exposure to APs (4-t-OP, 4-n-OP and 4-n-NP) is associated with idiopathic male infertility. 2013 Elsevier B V.
Endpoints from 12 rodent studies

144 unique outcomes reported...

Hormone concentrations
Growth
Viability
Organ weight
Gestational length
Sex ratio
Feminization
Developmental landmarks
Birth defects
Histology and morphology
Postnatal Administration of Triclosan and Thyroxine (T4)

**Figure 5.** Postnatal Triclosan administration and thyroxine concentration.
A. Postnatal triclosan administration and thyroxine concentration as a percentage of the control group for doses up to 300mg/kg/day
B. Postnatal beta-estimates for dose response and the random effects meta-analysis estimate
Rating Quality
Risk of Bias for non-human experimental studies

1. Sequence generation
   Was it appropriately randomized?

2. Allocation concealment
   Was allocation adequately concealed?

3. Blinding
   Were study personnel and outcome assessors blinded?

4. Incomplete outcome data
   Did study authors report all incomplete outcome data?

5. Selective outcome reporting
   Were outcomes reported selectively?

6. Conflict of interest
   Was study supported by entity with financial interest?

7. Other potential threats to validity
   Any other problems that could bias?
Rating Quality
Risk of Bias for non-human experimental studies

<table>
<thead>
<tr>
<th>Study</th>
<th>Sequence generation</th>
<th>Allocation concealment</th>
<th>Blinding</th>
<th>Incomplete outcome data</th>
<th>Selective outcome reporting</th>
<th>Conflict of interest</th>
<th>Other bias</th>
</tr>
</thead>
<tbody>
<tr>
<td>Paul et al. 2012</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stoker et al. 2010(^a)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stoker et al. 2010(^b)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Paul et al. 2010(^a)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rodriguez and Sanchez 2010</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Paul et al. 2010(^b)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zorilla et al. 2009</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kumar et al. 2009</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Axelstad et al. 2013</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Legend:
- Low risk
- Probably low risk
- Probably high risk
- High risk
- Unclear
Rating Quality
Risk of Bias for human studies

1. Are the study groups free from baseline differences?
2. Was knowledge of the exposure groups adequately prevented during the study?
3. Were exposure assessment methods robust?
4. Were outcome assessment methods robust?
5. Were confounding and effect modification adequately addressed?
6. Were incomplete outcome data adequately addressed?
7. Are reports of the study free of suggestion of selective outcome reporting?
8. Was the study free of support from a company, study author, or other entity having a financial interest in any of the exposures studied?
Rating Quality
Risk of Bias for human studies

Koepe et al. 2013
Cullinan et al. 2012
Allmyr et al. 2010
Rating Quality across all studies

Non-human evidence started as “high” quality rating (comparable to RCTs)

Human evidence started as “moderate” quality rating
Factors that **DECREASE** quality level

Possible ratings: 0=no change; -1 or -2 downgrade 1 or 2 levels.

1. **Risk of bias (study limitations)**
   - Rated down if most relevant evidence were from studies that suffered from high risk of bias
   - Conservative approach: confident that substantial risk of bias across most of body of evidence to downgrade

2. **Indirectness**
   - Rated down if evidence was not directly comparable to the question of interest
   - Based evaluation on PECO statement: population, exposure, comparator, outcome

3. **Inconsistency**
   - Rated down for widely different estimates of effect (heterogeneity or variability in results)
   - Considered variance in point estimates, confidence intervals overlap, $I^2$, tests for heterogeneity
Factors that **DECREASE** quality level

Possible ratings: 0=no change; -1 or -2 downgrade 1 or 2 levels.

4. **Imprecision**
   - Rated down if studies had few participants and few events (wide confidence intervals)

5. **Publication bias**
   - Rated down if studies were thought to be missing from body of evidence, resulting in an underestimate of true effects from exposure
   - Considered if there were early negative studies that were small in size, studies were small and sponsored by industry, unpublished studies showed different results from published studies, search not comprehensive
Factors that INCREASE quality level

Only applicable for human evidence
Possible ratings: 0=no change; +1 or +2 upgrade 1 or 2 levels.

1. **Large magnitude of effect**
   - Upgraded if modeling suggested confounding alone unlikely to explain associations with relative risk greater than 2 or very unlikely to explain relative risk greater than 5
   - Circumstances for upgrading occur infrequently

2. **Dose response**
   - Upgraded if consistent dose response gradient in one or multiple studies, and/or dose response across studies

3. **Confounding minimizes effect**
   - Upgraded if all possible residual confounders or biases reduced demonstrated effect, or suggested a spurious effect when results show no effect
   - Example is autism and vaccination - observational studies showed no association even though empirically confirmed bias that parents more likely to remember vaccination after because of publicity about possible association
# Summary of rating quality of the human hormonal evidence

<table>
<thead>
<tr>
<th>Category</th>
<th>Downgrades</th>
<th>Rationale</th>
</tr>
</thead>
<tbody>
<tr>
<td>Risk of bias</td>
<td>Seven (0); Four (-1)</td>
<td>Two of the three studies, one large and one small, have “low” or “probably low” risk of bias for all domains.</td>
</tr>
<tr>
<td>Indirectness</td>
<td>Five (0): Six (-1)</td>
<td>One study (Cullinan et al.) is of an older age group not representative of reproductive age where thyroid is a developmental or reproductive concern; Cullinan et al. exposure assessment by toothpaste use only is indirect.</td>
</tr>
<tr>
<td>Inconsistency</td>
<td>Eleven (0)</td>
<td>Although there are few studies on which to base this rating, the results are not inconsistent.</td>
</tr>
<tr>
<td>Imprecision</td>
<td>Eleven (0)</td>
<td>Considered downgrade to -1 here, based on Koeppe et al. wide confidence intervals, which is the majority of the data for this outcome.</td>
</tr>
<tr>
<td>Publication bias</td>
<td>Eleven (0)</td>
<td>These studies are not uniformly small and there is a larger study (Koeppe et al.) showing no effect for some outcomes. A comprehensive literature search did not identify studies with conflicting results.</td>
</tr>
</tbody>
</table>

## Upgrades

<table>
<thead>
<tr>
<th>Category</th>
<th>Downgrades</th>
<th>Rationale</th>
</tr>
</thead>
<tbody>
<tr>
<td>Large magnitude of effect</td>
<td>Eleven (0)</td>
<td>The studies found null or minimal effects only.</td>
</tr>
<tr>
<td>Dose-response</td>
<td>Ten (0); One (+1)</td>
<td>There is no or minimal evidence of a dose-response gradient.</td>
</tr>
<tr>
<td>Confounding minimizes effect</td>
<td>Eleven (0)</td>
<td>There is no evidence that residual confounding may be influencing results.</td>
</tr>
</tbody>
</table>

## Overall Quality of Evidence

(Initial rating is “Moderate”)

| Overall Quality of Evidence | Six (Moderate); Five (Low) |                                                                                                                                            |
Summary of rating quality of the non-human mammalian hormonal evidence

<table>
<thead>
<tr>
<th>Category</th>
<th>Downgrades</th>
<th>Rationale</th>
</tr>
</thead>
<tbody>
<tr>
<td>Risk of bias</td>
<td>Nine (-1); Two (0); One (0/-1)</td>
<td>(-1): There is “probably high” risk of bias across several domains; (0): Concern about overall risk of bias does not rise to the level of a downgrade; (0/-1): Most of these studies have “probably high” risk, rather than “high risk,” and this is mostly due to unknown information about the studies.</td>
</tr>
<tr>
<td>Indirectness</td>
<td>Twelve (0)</td>
<td>Animal changes (in rodents) are reflective of what is seen in humans and the outcomes are directly relevant to humans.</td>
</tr>
<tr>
<td>Inconsistency</td>
<td>Twelve (0)</td>
<td>There is not substantial heterogeneity in studies across postnatal dosing for thyroxine; lack of consistency between post and prenatal dosing has a biological explanation.</td>
</tr>
<tr>
<td>Imprecision</td>
<td>Twelve (0)</td>
<td>We judged that the confidence intervals are not wide for the T4 studies or the meta-analysis.</td>
</tr>
<tr>
<td>Publication bias</td>
<td>Twelve (0)</td>
<td>Studies include null findings as well as positive findings from studies with high risk for conflict of interest.</td>
</tr>
<tr>
<td>Overall Quality of Evidence (Initial rating is “High”)</td>
<td>Moderate</td>
<td>We downgraded one level based on concerns about risk of bias.</td>
</tr>
</tbody>
</table>
CONSIDERATIONS
1. What is the quality of the data?
2. What is the direction of the effect?
3. What is our confidence in the effect?
4. Are there other compelling attributes of the data that may influence certainty?
**Sufficient evidence of toxicity**

The available evidence includes results from one or more well-designed, well-conducted studies, and the conclusion is unlikely to be strongly affected by the results of future studies.

**Human:** A positive relationship is observed between exposure and outcome where chance, bias, and confounding can be ruled out with reasonable confidence.

**Non-human:** A positive relationship is observed between exposure and adverse outcome in multiple studies or a single appropriate study in a single species.

**Limited evidence of toxicity**

Confidence in the relationship is constrained by such factors as: the number, size, or quality of individual studies, or inconsistency of findings across individual studies. As more information becomes available, the observed effect could change, and this change may be large enough to alter the conclusion.

**Human:** A positive relationship is observed between exposure and outcome where chance, bias, and confounding cannot be ruled out with reasonable confidence.

**Non-human:** The data suggest a positive relationship between exposure and adverse outcome, but there are important limitations in the quality of the body of evidence.

**Inadequate evidence of toxicity**

The available evidence is insufficient to assess effects of the exposure. Evidence is insufficient because of: the limited number or size of studies, low quality of individual studies, or inconsistency of findings across individual studies. More information may allow an assessment of effects.

**Evidence of lack of toxicity**

**Human:** No relationship is observed between exposure and outcome, and chance, bias, and confounding can be ruled out with reasonable confidence. The available evidence includes consistent results from more than one well-designed, well-conducted study at the full range of exposure levels that humans are known to encounter, and the conclusion is unlikely to be strongly affected by the results of future studies. The conclusion is limited to the age at exposure and/or other conditions and levels of exposure studied.

**Non-human:** Data on an adequate array of endpoints from more than one study with at least two species showed no adverse effects at doses that were minimally toxic in terms of inducing an adverse effect. Information on pharmacokinetics, mechanisms, or known properties of the chemical class may also strengthen the evidence. The conclusion is limited to the species, age at exposure, and/or other conditions and levels of exposure studied, and is unlikely to be strongly affected by the results of future studies.
Definitions of “Strength of Evidence”

Derived from:

### Strength of Evidence Results

<table>
<thead>
<tr>
<th>Rating</th>
<th>Rationale</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inadequate</td>
<td>The available evidence is insufficient to assess effects of the exposure. Evidence is insufficient because of: the limited number or size of studies, low quality of individual studies, or inconsistency of findings across individual studies. More information may allow an assessment of effects.</td>
</tr>
<tr>
<td>Sufficient</td>
<td>We found sufficient evidence that exposure to triclosan alters hormone levels in rats, based on reduced thyroxine levels.</td>
</tr>
</tbody>
</table>
Integration of each evidence stream for thyroxine

### Sufficient
- Known to be toxic

### Limited
- Probably toxic
- Possibly toxic

### Inadequate
- Possibly toxic
- Not classifiable
Conclusion for Triclosan case study

Based on our evaluation using the Navigation Guide criteria, we concluded that there was “sufficient” non-human evidence and “inadequate” human evidence of an association between triclosan exposure and thyroxine concentrations, and consequently, triclosan is “possibly toxic” to reproductive and developmental health.
Integrating the streams of evidence for PFOA (1st case study)

**Strength of Evidence in Non-Human Systems**

- **Sufficient**: Known to be Toxic to Human Reproduction
- **Limited**: Possibly Toxic
- **Inadequate**: Not Classifiable
- **Evidence of Lack of Toxicity**: Probably Not Toxic

**Conclusion:** Human exposure to PFOA is known to be toxic to human reproduction and development based on sufficient evidence of decreased fetal growth in both human and non-human mammalian species.
Meta-analysis is a very useful tool

Grams change in birth weight per 1 ng/mL PFOA increase

Mean change in body weight (g) per 1 mg/kg BW/day PFOA dose
Acknowledgements

Navigation Guide Funders (since 2009)

- US Environmental Protection Agency
- US Environmental Protection Agency STAR (RD83467801)
- National Institute for Environmental Health Sciences (ES018135)
- California Environmental Protection Agency
- Fred Gellert Foundation
- Clarence Heller Foundation
- New York Community Trust
- Forsythia Foundation
- Passport Foundation
- Johnson Family Foundation
- Heinz Endowments
- Rose Foundation
- Kaiser Permanente
- Planned Parenthood Federation of America
- UCSF Phillip R Lee Institute for Health Policy Studies
Program on Reproductive Health and the Environment

Navigation Guide Systematic Review Method

References


Navigation Guide Systematic Review Method

References

• Fox D., Bero L. Systematic Reviews: Perhaps “The Answer to Policy Makers’ Prayers”? Environ Health Perspect DOI:10.1289/ehp.1408599 http://ehp.niehs.nih.gov/1408599/


