

**Applying the Navigation Guide
Case Study #2**

**Reproductive and Developmental Effects of Exposure
to
Triclosan**

**A Systematic Review of the Evidence
Protocol**

PROTOCOL INFORMATION.....	3
PROTOCOL.....	5
Background.....	5
The Navigation Guide	5
Properties of Triclosan (CAS# 3380-34-5)	6
Exposure to Triclosan and Possible Adverse Effects	6
Objectives	7
Methods.....	8
Criteria for Selecting Studies	8
Characteristics Outlined Using the PECO Structure.....	9
Search Methods	9
Study Selection Criteria.....	11
Data Collection	13
Risk of Bias Determination	13
Analysis	13
Quality and Strength of Evidence Ratings	15
SUPPLEMENTARY INFORMATION	17
Appendix I. Search Terms.....	17
Appendix II. List of Toxicological Websites to Search	19
Appendix III: Exclusion Criteria Screening Forms	20
Title and Abstract Screening Form.....	20
Full-Text Screening Form	20
Appendix IV: Data Collection Forms.....	21
Appendix V: Instructions for Making Risk of Bias Determinations	35
Non-human Experimental Studies.....	35
Human Studies.....	44
Appendix VI: Instructions for Grading the Quality and Strength of Evidence	53
Figure 1. Combining Streams of Human and Non-Human Evidence.....	66
Table 1. Strength of Evidence Definitions for Human Evidence.....	67
Table 2. Strength of Evidence Definitions for Non-human Evidence.....	68
ABOUT THE ARTICLE.....	69
ACKNOWLEDGEMENTS	70
References.....	71

PROTOCOL INFORMATION

Authors

Paula I. Johnson, PhD, MPH
Postdoctoral Scholar
Program on Reproductive Health and the Environment
University of California, San Francisco

Erica Koustas, PhD
ORISE Postdoctoral Fellow
U.S. EPA Office of Policy, National Center for Environmental Economics

Hanna M. Vesterinen, PhD
Postdoctoral Fellow
Program on Reproductive Health and the Environment
University of California, San Francisco

Patrice Sutton, MPH
Research Scientist
Program on Reproductive Health and the Environment
University of California, San Francisco

Dylan Atchley, BS
Research Assistant
Program on Reproductive Health and the Environment
University of California, San Francisco

Allegra N. Kim, PhD, MS
Research Scientist
California Environmental Protection Agency, Office of Environmental Health Hazard Assessment

Marlissa Campbell, PhD
Staff Toxicologist
California Environmental Protection Agency, Office of Environmental Health Hazard Assessment

James M. Donald, PhD
Chief, Reproductive Toxicology and Epidemiology Section
California Environmental Protection Agency, Office of Environmental Health Hazard Assessment

Lisa Bero, PhD
Co-Director, San Francisco Branch United States Cochrane Center
Professor, Department of Clinical Pharmacy, School of Pharmacy, University of California, San Francisco

Saunak Sen, PhD
Associate Professor, Epidemiology and Biostatistics
University of California, San Francisco

Daniel Axelrad, MPP
Environmental Scientist

U.S. EPA, Office of Policy National Center for Environmental Economics

Lauren Zeise

Deputy Director, Scientific Affairs

California Environmental Protection Agency, Office of Environmental Health Hazard Assessment

Tracey J. Woodruff, PhD, MPH

Professor and Director

Program on Reproductive Health and the Environment

University of California, San Francisco

Contact: Paula Johnson, UCSF Program on Reproductive Health and the Environment,
1330 Broadway, Suite 1135, Oakland, CA 94612 johnsonp@obgyn.ucsf.edu (510) 350-1245

PROTOCOL

Background

The Navigation Guide

In health care, weight of evidence reviews of scientific research have played a transformative role in speeding the incorporation of science into therapeutic and preventative action at the individual and societal level. These systems are not fully transferable to environmental health science, primarily because of differences between clinical and environmental health science in: (1) the types and completeness of evidence; and (2) how decisions to expose patients or populations are made. In the clinical arena, decisions about exposure to an exogenous substance are made based on weighing risks and benefits to patients' health. Evidence concerning the safety, efficacy, and dosing of new therapeutic drugs includes information from randomized controlled trials (RCTs), in which treatments are administered in controlled experiments designed to answer specific questions. Systematic approaches and methodologies have been developed to incorporate experimental evidence into recommendations for treatment, such as the Grading of Recommendations, Assessment, Development and Evaluation (GRADE) approach, adopted by the Cochrane Collaboration [1, 2].

In contrast to evidence available for clinical decision-making, environmental health risk assessments typically deal with chemicals of no therapeutic value, so RCTs are seldom available. Environmental health risk assessments, therefore, rely heavily on animal toxicology studies and human observational studies. Emerging methods in environmental toxicology promote basing assessments on data from high-throughput, rapid screening experiments performed *in vitro* or even *in silico*.

In an attempt to harmonize the approaches for assessing evidence in the clinical and environmental health sciences, the Navigation Guide was developed to provide a reproducible framework to evaluate the quality and strength of evidence about the relationship between environmental exposures and reproductive and developmental health [3]. The Navigation Guide is a systematic and transparent methodology that proceeds from best practices in the clinical arena but accounts for the differences in evidence and decision context described above. The Navigation Guide methodology outlines four key steps: 1) specify the study question; 2) select the evidence; 3) rate the evidence; and 4) grade the strength of the recommendation.

Previously we applied the Navigation Guide methodology in a proof-of-concept study to evaluate the human and non-human evidence of perfluorooctanoic acid (PFOA) on fetal growth [4, 5]. As part of our ongoing proof-of-concept for the Navigation Guide methodology, this systematic review evaluates the evidence for the effects of exposure to the widely-used biocide triclosan on endpoints of developmental and/or reproductive toxicity. The human health rationale for this review relates to the widespread human exposure to triclosan and evidence of reproductive and developmental health effects, as described below.

This systematic review employs the Navigation Guide methodology for steps 1-3 for both human and non-human evidence streams. The results will be compared to the criteria in the Navigation Guide for separately rating the strength of the human and non-human evidence according to one

of the following four statements: 1. Sufficient; 2. Limited; 3. Inadequate; or 4. Evidence of lack of toxicity (see Appendix VI, Tables 1 and 2). Subsequently, in step 4, we will combine the ratings of the strength of the non-human and human evidence into one of the following five summary statements about the overall strength of the evidence according to the criteria in the Navigation Guide: 1. Known to be toxic; 2. Probably toxic; 3. Possibly toxic; 4. Probably not toxic; or 5. Not classifiable.

Properties of Triclosan (CAS# 3380-34-5)

Triclosan, or 2,4,4'-trichloro-2'-hydroxydiphenyl ether, is a synthetic, broad spectrum anti-microbial agent that possesses mostly antibacterial properties but also has effects on viruses and fungi. It is used to prevent bacterial growth on the skin and preserve products against microbial deterioration. Triclosan is manufactured in either the ethyl or phenol form. Its mechanism of action as an anti-microbial agent is by blocking the active site of an essential enzyme (enoyl-acyl carrier protein reductase inhibitor; ENR) in fatty acid synthesis, which is necessary for bacterial growth. Triclosan is regulated by U.S. EPA when used as a bacteriostat, fungistat, mildewstat, or deodorizer (<http://www.epa.gov/oppsrrd1/REDs/234ored.pdf>).

Exposure to triclosan and possible adverse effects

Products that contain triclosan are marketed under numerous trade names, including Irgasan[®], Irgacare[®], Biofresh[®], and Microban[®]. Triclosan was developed over 40 years ago and introduced as a surgical scrub [6] and was first brought into the commercial market in 1972 as an antibacterial agent in soaps [7]. It has since been introduced to thousands of consumer products, including: dental products, cosmetics, deodorants, personal care products, first aid items, kitchenware, computer equipment, clothes, and children's toys [8]. A 2001 study reported 76% of national brand liquid soaps contained triclosan [9]. Triclosan is also used as a material preservative in adhesives, fabrics, vinyl, plastics (toys, toothbrushes), polyethylene, polyurethane, polypropylene, floor wax, textiles (footwear, clothing, awnings, upholstery fabrics, tents), caulking compounds, sealants, rubber, concrete mixtures, grouts, latex paints, mattresses, mulch, brooms, shower curtains, toilet bowls, urinals, garbage cans, refuse container liners, and insulation. Commercial, institutional and industrial uses include conveyor belts, fire hoses, dye bath vats and ice making equipment. Triclosan is also applied to HVAC coils [10].

Through such widespread use triclosan has become broadly disseminated into the environment; it was found in over half of streams sampled in 30 states in the U.S. [11], in sewage sludge applied to agriculture [11], as well as in human breast milk and urine [12, 13]. The U.S. Centers for Disease Control and Prevention (CDC) detected triclosan in 74.6% of a representative sample of the general U.S. population at least six years of age in the 2003-2004 National Health and Nutrition Examination Survey (NHANES) [13]. Other smaller studies also indicate widespread exposure [14, 15].

Triclosan was previously thought harmless to humans and animals because they lack the ENR enzyme targeted by triclosan. However, there is growing concern over triclosan's possible effects on public health, both direct, e.g., skin irritation [16, 17], and indirect, e.g., antibiotic resistance [18]. Triclosan has been reported to have endocrine effects in amphibians and fish [19-21] and reduces sperm production in rats [22]. One study that examined pubertal stages of girls in relation to biomarkers of phytoestrogens, phthalates, and phenols found a "suggestive inverse association" between triclosan and pubic hair development, though the trend was not monotonic [23]. Thus,

based on its presence in numerous consumer products, the environment, and human breast milk, and the evidence of possible endocrine effects, we decided to review the evidence for reproductive and developmental effects of triclosan.

Aim

- To answer the question: “Does exposure to triclosan have adverse effects on human development or reproduction?”

Objectives

Our specific objectives are:

- To identify all of the literature concerning the effects of triclosan exposure on developmental or reproductive outcomes, including *in vivo*, *in vitro*, and *in silico* studies, as well as observational studies in humans;
- To conduct a meta-analysis of the effect of triclosan on these outcomes and assess for potential sources of heterogeneity;
- To assess the risk of bias of individual studies and individual outcomes and, where appropriate, assess their impact on measures of effect size;
- To rate the strength of the experimental and observational evidence on the developmental and reproductive toxicity of triclosan according to one of the following four statements: 1. Sufficient; 2. Limited; 3. Inadequate; or 4. Evidence of lack of toxicity;
- To combine the strength of evidence ratings for each evidence stream into one of five overall statements of toxicity: 1. Known to be toxic; 2. Probably toxic; 3. Possibly toxic; 4. Not classifiable; or 5. Probably not toxic.

Methods

Criteria for Selecting Studies

We will select studies in which exposure to triclosan was documented, measured, or estimated, and potential effects on development and reproduction or the reproductive system were evaluated. Identification of developmental and reproductive effects is guided by standard definitions and guidance for assessment of these endpoints such as the U.S. EPA guidelines for risk assessment of developmental and reproductive toxicity [24, 25].

Broad categories of adverse developmental effects are death, structural abnormalities, growth alterations, and functional deficits [24]. Within these broad categories, developmental effects include, but are not limited to:

- fetal loss or resorption, spontaneous abortion, and neonatal and subsequent mortality;
- structural malformations and variations;
- alterations in fetal or postnatal growth including body weight, crown-rump length, cranial circumference and delayed ossification
- morbidity;
- altered sex ratio;
- altered gestation length; and
- neurobehavioral effects such as altered functional behaviors or cognitive deficits.

Adverse reproductive outcomes include, but are not limited to:

- reduced fertility;
- Impotence;
- menstrual disorders;
- spontaneous abortion;
- low birth weight and other developmental (including heritable) defects;
- premature reproductive senescence;
- genetic damage to the ovum, spermatozoon, and precursors;
- alterations in ovulation;
- increased pregnancy wastage (including stillbirth);
- adverse effects on sexual behavior;
- adverse effects on lactation;
- impaired sperm or seminal fluid production (including alterations in number, morphology, motility, and ability to fertilize);
- altered endocrine function; and
- alterations in size, morphology, or function of reproductive organs.

Studies that are eligible for review will address the study question and the characteristics as outlined in the following “PECO” aid.

Characteristics outlined using the PECO structure

“PECO” is an aid used to formulate an answerable question in a systematic review of health interventions. The acronym stands for “Population/Participants”, “Exposure,” “Comparator” and “Outcomes.”

Population:

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

Search Methods

We will employ a variety of methods to identify relevant data, as outlined below. Our search will not be limited by language or publication date.

Electronic Searches

We will perform electronic searches of online databases (PubMed, ISI Web of Science, Biosis Previews, Embase and Toxline) using the search terms outlined in Appendix I.

To assist in the development of a list of terms relevant to our search strategy we will use the following databases to compile synonyms for triclosan: Medical Subject Headings (MeSH), PubChem, Sigma-Adlrich, and ChemSpider (<http://pubchem.ncbi.nlm.nih.gov/summary/summary.cgi?q=nama&cid=5564>; <http://www.sigmaaldrich.com/catalog/product/sigma/72779?lang=en®ion=US>; <http://www.chemspider.com/Chemical-Structure.5363.html>).

In addition we will identify further synonyms from the following known reviews or original research articles on triclosan: Dann et al, 2011; James et al, 2010; Ciba et al, 2004; Rodricks et al, 2010; Fang et al, 2010; and APUA, 2011 [26-31].

Substance (triclosan): We will combine “triclosan” and its synonyms in a Boolean search using the “OR” statement. We will search for terms in titles and abstracts (using the [tiab] function in Pubmed, topic search in Web of Science and Biosis Previews; “ti,ab.” function in Embase) or MeSH headings (using the [mh] function).

Web of Science and Biosis Previews

To develop a Web of Science and Biosis Previews search filter, we will modify the PubMed search filter. This will consist of removing all MeSH terms and formatting the search terms for the Web of Science database. Specifically, we will perform a topic search in Web of Science, which will search all keywords, titles, and abstracts for search terms.

Embase

We will develop our Embase search filter using the same method as described above for Web of Science and Biosis Previews. We will use the “ti,ab.” function to limit the search to titles and abstracts.

Toxline

We will develop our Toxline search filter using the same method as described above for Web of Science and Biosis Previews. Specifically we will limit our search to include chemical synonyms and CAS numbers and to exclude Pubmed records.

Toxic Substance Control Act Docket

We will search for unpublished data from primary manufacturers of triclosan by searching the Toxic Substances Control Act (TSCA) docket (<http://www.ntis.gov/products/ots.aspx>) using the following CAS number and each of the synonymous terms: 335-67-1, perfluorooctanoic acid, perfluorooctanoic acid, perfluoro-n-octanoic acid, pentadecafluorooctanoic acid, APFO, perfluorinated octanoic acid, perfluorooctanoate. TSCA requires that manufacturers report positive toxicological findings to U.S. EPA and these reports are maintained in the TSCA docket [32].

Searching Other Resources

We will use other methods to find additional studies that are not identified through electronic sampling and may be in the grey literature i.e., technical reports from government agencies or scientific research groups, working papers from research groups or committees, white papers, preprints, conference proceedings, personal communications, etc.

These methods include:

- Searching the websites listed in Appendix II.
- Hand searching the reference list of all studies that are included after full text review (prior to study author contact, if applicable) and use Web of Science to search for articles that cite the included studies.
- Personal communication with authors to request unpublished data or if they have knowledge of additional data from other authors.

Study Selection Criteria

All search results will be imported or manually entered into EndNote reference management software. We will eliminate duplicate references before we begin evaluating the eligibility of the studies identified.

Title and abstract screening

Each reference will be screened in duplicate. Two of five possible reviewers (HV, MC, AK, DSA, and RB) will independently conduct a title and abstract review of each reference from the literature search results to determine whether a reference meets the selection criteria for inclusion. Each author will be assigned a non-random subset of references to screen.

Furthermore, references will be assigned in such a manner to ensure that the same two authors do not always screen the same references (i.e. reviewer one will be assigned the 1st half of the references; reviewer two, second half; reviewer three, 1st and 3rd quarter; reviewer four 2nd quarter; reviewer 5; 4th quarter).

References which are included at the title/abstract screening level will be subject to a full text review by two of the same five authors.

One author (PJ) will be brought in to settle any discrepancies between the reviewers resulting from each step of the review process. To ensure quality control, PJ will also perform title and abstract screening of five percent of the search results and five percent or 5 papers, whichever is greater, of search results eligible for full text review. The review of selection criteria for inclusion of articles against the pre-specified inclusion and exclusion criteria will be performed using a structured form in DistillerSR (Evidence Partners; available at: <http://www.systematic-review.net>), an online application designed specifically for the screening and data extraction phases of a systematic review (see Appendix III for title and abstract inclusion/exclusion form).

Reports in any language, from any year, will be eligible for inclusion. All reports that compare humans or other animals exposed to triclosan to appropriate comparators and evaluate them for reproductive or developmental effects, as described in the PECO statements above, will be eligible for inclusion.

The title/abstract screening form will be used to screen and exclude references if one or more of the following criteria are met:

1. Report does not contain original data;
2. There was no triclosan exposure;
3. There was no triclosan-related toxicity (effects) data; and/or
4. Other reason (explanation required).

These criteria are expected to exclude studies of triclosan efficacy (unless adverse effects are reported) and studies of bacterial resistance to triclosan. Some reports will also be categorized as “of interest/supplemental” if they appear to be inappropriate for inclusion in a meta-analysis, but may include relevant information.

The following types of records will be included at the title/abstract level:

- Any study on comparative triclosan exposures and health effects
- Mechanistic studies, i.e., studies that investigate potential biological or physiological mechanisms that underlie toxicity, to be reviewed at full-text level for further decision/categorization;
- Studies with “exposed” versus “unexposed” or “less exposed” comparisons, even if triclosan exposure is not quantified; and
- Studies with triclosan exposure at any life stage, if subsequent effects are reported.

For citations where the database contains no abstract, authors will attempt to obtain the abstracts from an internet search. Articles for which the abstract remains unavailable will be screened based on titles and PubMed MeSH headings. Any study not excluded based on above criteria will be included for full-text review.

Full-Text Screening

Citations eligible for full text review will be screened and excluded if one or more of the following criteria are met:

1. Article does not contain original data;
2. Study subjects were not exposed to triclosan prior to assessment of reproductive or developmental outcomes (this will not exclude cross-sectional epidemiological studies);
3. Reproductive or developmental outcomes were not reported;
4. There was no comparator (control group or exposure range comparison); and/or
5. Other reasons (explanation required).

Citations will be included if they meet the PECO statement criteria, that is, their subjects represent humans or other animals (including *in vitro* and *in silico*), they include exposure comparisons for relevant periods, and they report developmental or reproductive health outcomes. Other citations that include useful information, such as exposure or mechanistic data, will be noted but not included. Information from these references may be used to aid interpretation of data or inform risk of bias determinations.

For articles (including non-English articles) which are not available in the database, we will attempt to obtain articles from a broad internet search. If resources permit, potentially relevant non-English articles will be translated into English to determine eligibility.

Data Collection

Two authors will independently extract the study characteristics and data from all of the included articles into a custom Microsoft Access database (see Appendix IV for the study characteristics data collection form). The extracted characteristics will be used to evaluate reporting quality, risk of bias and/or to conduct statistical analyses, and were compiled by combining aspects of a variety of checklists and criteria [33-37]. Study authors will be contacted when additional information is required. If study authors do not respond to requests after being contacted through 2 email messages over the course of 1 month or more, review authors will note that attempts to contact study researchers were unsuccessful.

Risk of Bias Determination

Risk of bias will be assessed for both human and animal studies using domains from the Cochrane Collaboration's "Risk of Bias" tool and the Agency for Healthcare Research and Quality's (AHRQ) criteria that address selection bias, confounding, performance bias, attrition bias, detection bias, and reporting bias [34, 38]. We have modified terminology and concepts in these domains to make them appropriate for animal toxicological studies and human observational studies where appropriate (Appendix V).

Informed by empirical data from meta-analyses conducted on pharmacological treatments and studies of risk of bias and sponsorship [39-41], we will additionally assess funding source and declared conflicts of interest as sources of bias. We will also search for each study in PubMed and note if there has been a retraction of the published article in order to determine if the study may be fraudulent. Two review authors will independently make risk of bias determinations for each study across all domains and then compare their results. Any remaining discrepancies will be reviewed by 4 co-authors. If, upon further discussion the 4 co-authors cannot reach agreement on an appropriate risk of bias determination for a particular domain, the more conservative judgment will be selected (e.g. if one reviewer makes a judgment of 'yes' and the other makes a judgment of 'probably yes', the 'probably yes' judgment will be used). If additional data or information are acquired from study authors, risk of bias judgments will be modified to reflect the updated study information.

We will attempt to minimize the impact of publication bias by: (1) implementing a comprehensive search of the literature using multiple sources and methods in order to identify all published and unpublished studies that meet the eligibility criteria; and (2) if possible, using funnel plot analysis and/or other statistical analyses (e.g. Egger regression [42] and "trim and fill" [43] of the studies included in the systematic review), as appropriate.

Analysis

Where appropriate we will perform a meta-analysis to summarize the effects of triclosan on developmental and reproductive outcomes and to assess the impact of study design characteristics on findings. Characteristics from each study will be compiled and reviewed to establish comparability between studies or to identify data transformations necessary to ensure such comparability. A few key characteristics include:

- Study design
- Type (species, strain, and genetic background) of animal model used

- Dose levels, exposure route, and dosing preparation of triclosan for experimental studies
- Health outcome assessed
- Type of data/statistic available

Summaries of these characteristics for each included study will be assessed by two reviewers in pairs to determine comparability between studies and to identify any heterogeneity concerns.

After gathering data for the meta-analysis, we will consult with a statistician (SS) to identify appropriate statistical methods to analyze the data, and to determine whether further modifications are required prior to performing the meta-analysis. Our proposed approach is to fit linear dose-response models (with the dose variable log-transformed) to each set of study data. From each study, the estimated slope of the linear model and its associated standard error will be collected. These estimates will then be combined across comparable studies, using a random effects model to account for potential heterogeneity across studies. The final quantitative result will be the combined estimate of the slope of the linear dose-response model with an associated confidence interval. However, consultation with SS will guide the identification of a statistical approach that is appropriate for the study data available.

To test statistical heterogeneity across the study estimates, we will estimate the variance component corresponding to between-study variability, and use a likelihood ratio test for the null hypothesis that between-study variability is absent. A p-value of 0.05 or less will be considered statistically significant. Furthermore, to assess the impact of between-study heterogeneity on the meta-analysis, the I^2 test statistic will be calculated and evaluated by considering the magnitude/direction of effects, strength of evidence for heterogeneity (e.g., p-value from a chi squared test or a confidence interval for I^2), and the Cochrane’s guide to interpretation as follows:

- 0%-40%: might not be important;
- 30%-60%: may represent moderate heterogeneity;
- 50%-90%: may represent substantial heterogeneity;
- 75%-100%: considerable heterogeneity.

Sensitivity analyses will be performed by examining the effects of excluding studies with particularly heterogeneous results as well as performing subgroup analyses based on excluding subsets of studies with shared characteristics that might be influential.

If possible, i.e. there are enough studies, we will assess for the presence of publication bias by funnel plotting and Egger regression on the estimates of effect size [42]. In addition, if these methods suggest that publication bias is present we will use “trim and fill” to predict the impact of hypothetical “missing” studies [43].

Quality and Strength of Evidence Ratings

Upon completion of the data collection, risk of bias determination, and data analysis, each of the co-authors will compare the results of the systematic review to the criteria in the Navigation Guide for rating the quality and strength of the evidence. All co-authors will be given explicit directions before rating (see Appendix VI, “Instructions for Rating the Quality and Strength of Evidence”).

The initial quality level of experimental animal data will be considered as high, comparable to human randomized controlled trials. This assumption is supported by the level of study control exercised in experimental animal studies and the limited heterogeneity within the study population. This is also consistent with GRADE's consideration of randomization as a key determinant of the default "high" [35]. Furthermore, studies suggest that humans are more sensitive to chemical exposures than animals, strengthening the applicability of findings from experimental animal studies to human health outcomes [44]. The initial quality level of human observational data will be considered moderate.

Factors that may decrease the quality level of the body of evidence include:

1. Risk of Bias Across Studies: Study limitations – a substantial risk of bias across body of evidence;
2. Indirectness: Evidence was not directly comparable to the question of interest (i.e., population, exposure, comparator, and outcome).
3. Inconsistency: Widely different estimates of effect (heterogeneity or variability in results);
4. Imprecision: Studies had few participants and few events (wide confidence intervals); and
5. Publication Bias: Studies missing from body of evidence, resulting in an overestimate or underestimate of true effects from exposure.

Because experimental animal studies begin as "high" quality, there is no opportunity for upgrading the body of animal evidence. Upgrading the "moderate" quality of human observational data is possible according to the following factors. Factors that may increase the quality level of the body of evidence include:

1. Large magnitude of effect: Upgraded if modeling suggested confounding alone unlikely to explain associations with large magnitude of effect.
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 consideration of all plausible residual confounders or biases would underestimate the effect or suggest a spurious effect when results show no effect.

Possible ratings for quality of evidence are "high," "moderate," or "low." Possible downgrades or upgrades are: 0 (no change), -1 (1 level downgrade), -2 (2 level downgrade), +1 (1 level upgrade) or +2 (2 level upgrade). The ratings of the separate factors are not necessarily added together into a score, e.g. a -1 downgrade for inconsistency and a -1 downgrade for imprecision does not automatically dictate an overall -2 downgrade for the body of evidence. Judgment is exercised to determine if the rationale behind each downgrade warrants an overall downgrade of 1 or 2 levels. The same applies to upgrading the overall body of evidence. Likewise, a -1 downgrade for one factor and a +1 upgrade for another factor do not automatically cancel out and determine 0 downgrades or upgrades for the overall body of evidence.

Authors who decide to rate quality down or up need to specify the 1 or 2 criteria most responsible for their decision while documenting all factors that contributed to the final decision. After independently evaluating the quality of the evidence, co-authors will compare their evaluations and any discrepancies between the reviewers' decisions will be resolved through discussion until consensus is reached. The rationale for each decision on each of the five factors will be recorded.

Subsequent to consensus on the quality of the evidence, the co-authors will rate the strength of evidence. The overall strength of the body of human evidence is based on a combination of four criteria: (1) Quality of body of evidence (i.e., the rating from the previous step); (2) Direction of effect; (3) Confidence in effect; and (4) Other compelling attributes of the data that may influence certainty. The results of rating of the strength of the non-human evidence will then be compared to the criteria specified in the Navigation Guide and described according to one of the following four statements: 1. Sufficient; 2. Limited; 3. Inadequate; or 4. Evidence of lack of toxicity (Appendix VI, Tables 1 and 2) [3]. Any discrepancies between the reviewers' decisions will be resolved through discussion. The senior author (TW) will be the ultimate arbiter of the discrepancies that cannot be resolved through consensus among the co-authors.

At the conclusion of the systematic review of the non-human and human evidence, we will combine these ratings as specified in the Navigation Guide, a process adapted from IARC's method which allows for combining the results of human and non-human evidence into a single concise statement of health hazard [3, 45]. The end result will be one of five possible statements on the impact of triclosan on reproductive or developmental health: 1. known to be toxic; 2. probably toxic; 3. possibly toxic; 4. unclassifiable; or 5. probably not toxic to reproductive or developmental health.

The ultimate goal of these efforts will be to develop transparent science-based estimates of health impacts of environmental toxicants to support the development of recommendations for prevention of adverse effects caused by environmental toxicants.

SUPPLEMENTARY INFORMATION

Appendix I. Search Terms

A literature search was conducted on June 5th 2013 using the following search terms:

Search	PubMed
#1 Substance terms	triclosan [mh] OR triclosan* [tiab] OR 3380-34-5 [rn] OR Irgasan [tiab] OR "Colgate Total" [tiab] OR (enoyl acyl carrier protein reductase [tiab] AND inhibit*[tiab]) OR pHisoHex [tiab] OR methyltriclosan [tiab] OR "methyl triclosan" [tiab] OR "methyl-triclosan" [tiab] OR "methyl-triclosan" [tiab] OR "Colgate Palmolive" [tiab] OR TCCP [tiab] OR trichloro-2'-hydroxydiphenyl ether [tiab] OR "5-chloro-2-(2,4-dichlorophenoxy)phenol" [tiab] OR "2,4,4'-trichloro-2'-hydroxydiphenyl ether" [tiab] OR "DP-300 " [tiab] OR mentadent [tiab] OR polychlorobiphenylol* [tiab] OR Microshield [tiab] OR pHisoderm [tiab] OR Irgacare [tiab] OR Microban [tiab] OR "2,4,4'-trichloro-2'-hydroxy-diphenyl ether" [tiab]) OR "2-(2,4-dichlorophenoxy)-5-chlorophenol" [tiab] OR Aquasept [tiab] OR "Ster-Zac" [tiab] OR Playskool [tiab] OR "5-chloro-(2,4-dichlorophenoxy)phenol" [tiab]) OR "Ultra Fresh" [tiab] OR Gamophen [tiab] OR C ₁₂ H ₇ Cl ₃ O ₂ [tiab] OR "Bacti-Stat" [tiab] OR Tinosan [tiab] OR Irgaguard [tiab] OR Cloxifenol [tiab] OR Aveeno [tiab] OR Ch3565 [tiab] OR GP41-353 [tiab] OR logamel [tiab] OR "Colgate Total" [tiab] OR "phenol, 5-chloro-2-(2,4-dichlorophenoxy)" [tiab] OR "Araldite hardener" [tiab] OR "J-Cloth" [tiab] OR "Ultra Fresh" [tiab] OR Trisan [tiab] OR "Bauer 5000" [tiab] OR Biofresh [tiab] OR Amicor [tiab] OR "CGP 433" [tiab] OR Aquasept [tiab] OR "California Paints" [tiab] OR "reach toothbrush" [tiab] OR "Clean & Clear" [tiab] OR "ether, 2'-hydroxy-2,4,4'-trichlorodiphenyl" [tiab] OR "phenyl ether, 2'-hydroxy-2,4,4'-trichloro-" [tiab] OR "HSDB 7194" [tiab] OR "2,4,4'-trichloro-2'-hydroxy-diphenyl ether" [tiab] OR "2-Hydroxy-2',4,4'-trichloro diphenyl ether" [tiab] OR "Jason Natural Cosmetics" [tiab]

Search	Web of Science & Biosis Previews
#1 Substance terms	TS=(Triclosan* OR 3380-34-5 OR Irgasan OR "Colgate Total" OR (enoyl acyl carrier protein reductase AND inhibit*) OR pHisoHex OR methyltriclosan OR "methyl triclosan" OR "methyl-triclosan" OR "methyl-triclosan" OR "Colgate Palmolive" OR TCCP OR trichloro-2'-hydroxydiphenyl ether OR "5-chloro-2-(2,4-dichlorophenoxy)phenol" OR "2,4,4'-trichloro-2'-hydroxydiphenyl ether" OR "DP-300 " OR mentadent OR polychlorobiphenylol* OR Microshield OR pHisoderm OR Irgacare OR Microban OR "2,4,4'-trichloro-2'-hydroxy-diphenyl ether") OR "2-(2,4-dichlorophenoxy)-5-chlorophenol" OR Aquasept OR "Ster-Zac" OR Playskool OR "5-chloro-(2,4-dichlorophenoxy)phenol") OR "Ultra Fresh" OR Gamophen OR C ₁₂ H ₇ Cl ₃ O ₂ OR "Bacti-Stat" OR Tinosan OR Irgaguard OR Cloxifenol OR Aveeno OR Ch3565 OR GP41-353 OR logamel OR "Colgate Total" OR "phenol, 5-chloro-2-(2,4-dichlorophenoxy)" OR "Araldite hardener" OR "J-Cloth" OR "Ultra Fresh" OR Trisan OR "Bauer 5000" OR Biofresh OR Amicor OR "CGP 433" OR Aquasept OR "California Paints" OR "reach toothbrush" OR "Clean & Clear" OR "ether, 2'-hydroxy-2,4,4'-trichlorodiphenyl" OR "phenyl ether, 2'-hydroxy-2,4,4'-trichloro-" OR "HSDB 7194" OR "2,4,4'-trichloro-2'-hydroxy-diphenyl ether" OR "2-Hydroxy-2',4,4'-trichloro diphenyl ether" OR "Jason Natural Cosmetics"

Search	Embase
#1 Substance terms	(Triclosan* OR 3380-34-5 OR Irgasan OR "Colgate Total" OR (enoyl acyl carrier protein reductase AND inhibit*) OR pHisoHex OR methyltriclosan OR "methyl triclosan" OR "methyl-triclosan" OR "methyl-triclosan" OR "Colgate Palmolive" OR TCCP OR trichloro-2'-hydroxydiphenyl ether OR "5-chloro-2-(2,4-dichlorophenoxy)phenol" OR "2,4,4'-trichloro-2'-hydroxydiphenyl ether" OR "DP-300 " OR mentadent OR polychlorobiphenylol* OR Microshield OR pHisoderm OR Irgacare OR Microban OR "2,4,4'-trichloro-2'-hydroxy-diphenyl ether") OR "2-(2,4-dichlorophenoxy)-5-chlorophenol" OR Aquasept OR "Ster-Zac" OR Playskool OR "5-chloro-(2,4-dichlorophenoxy)phenol") OR "Ultra Fresh" OR Gamophen OR C ₁₂ H ₇ Cl ₃ O ₂ OR "Bacti-Stat" OR Tinosan OR Irguard OR Cloxifenol OR Aveeno OR Ch3565 OR GP41-353 OR logamel OR-"Colgate Total" OR "phenol, 5-chloro-2-(2,4-dichlorophenoxy)" OR "Araldite hardener" OR "J-Cloth" OR "Ultra Fresh" OR Trisan OR "Bauer 5000" OR Biofresh OR Amicor OR "CGP 433" OR Aquasept OR "California Paints" OR "reach toothbrush" OR "Clean & Clear" OR "ether, 2'-hydroxy-2,4,4'-trichlorodiphenyl" OR "phenyl ether, 2'-hydroxy-2,4,4'-trichloro-" OR "HSDB 7194" OR "2,4,4'-trichloro-2'-hydroxy-diphenyl ether" OR "2-Hydroxy-2',4,4'-trichloro diphenyl ether" OR "Jason Natural Cosmetics").ti,ab.

Search	Toxline
#1 Substance terms	Triclosan* OR 3380-34-5 OR Irgasan OR "Colgate Total" OR (enoyl acyl carrier protein reductase AND inhibit*) OR pHisoHex OR methyltriclosan OR "methyl triclosan" OR "methyl-triclosan" OR "methyl-triclosan" OR "Colgate Palmolive" OR TCCP OR trichloro-2'-hydroxydiphenyl ether OR "5-chloro-2-(2,4-dichlorophenoxy)phenol" OR "2,4,4'-trichloro-2'-hydroxydiphenyl ether" OR "DP-300 " OR mentadent OR polychlorobiphenylol* OR Microshield OR pHisoderm OR Irgacare OR Microban OR "2,4,4'-trichloro-2'-hydroxy-diphenyl ether") OR "2-(2,4-dichlorophenoxy)-5-chlorophenol" OR Aquasept OR "Ster-Zac" OR Playskool OR "5-chloro-(2,4-dichlorophenoxy)phenol") OR "Ultra Fresh" OR Gamophen OR C ₁₂ H ₇ Cl ₃ O ₂ OR "Bacti-Stat" OR Tinosan OR Irguard OR Cloxifenol OR Aveeno OR Ch3565 OR GP41-353 OR logamel OR-"Colgate Total" OR "phenol, 5-chloro-2-(2,4-dichlorophenoxy)" OR "Araldite hardener" OR "J-Cloth" OR "Ultra Fresh" OR Trisan OR "Bauer 5000" OR Biofresh OR Amicor OR "CGP 433" OR Aquasept OR "California Paints" OR "reach toothbrush" OR "Clean & Clear" OR "ether, 2'-hydroxy-2,4,4'-trichlorodiphenyl" OR "phenyl ether, 2'-hydroxy-2,4,4'-trichloro-" OR "HSDB 7194" OR "2,4,4'-trichloro-2'-hydroxy-diphenyl ether" OR "2-Hydroxy-2',4,4'-trichloro diphenyl ether" OR "Jason Natural Cosmetics"

Appendix II. List of Toxicological Websites to Search

- ATSDR Interaction Profiles <http://www.atsdr.cdc.gov/interactionprofiles/>
- ATSDR Toxicological Profiles <http://www.atsdr.cdc.gov/toxproz.html>
- CalEPA Office of Environmental Health Hazard Assessment <http://www.oehha.ca.gov/risk.html>
- Chem ID <http://chem.sis.nlm.nih.gov/chemidplus/>
- Chemfinder www.chemfinder.com/
- Chemspider <http://www.chemspider.com>
- Chemical Carcinogenesis Research Information System <http://www.nlm.nih.gov/pubs/factsheets/ccrisfs.html>
- DART <http://toxnet.nlm.nih.gov/cgi-bin/sis/htmlgen?DARTETIC>
- EPA Acute Exposure Guideline Levels <http://www.epa.gov/oppt/aegl/chemlist.htm>
- EPA IRIS edocket and official records; IRIS Hotline 202-566-1676
- EPA IRIS internet www.epa.gov/iris
- EPA NEPIS <http://www.epa.gov/nepis/>
- EPA NSCEP <http://www.epa.gov/ncepihom/>
- EPA Science Inventory <http://www.epa.gov/gateway/science/>
- EPA Substance Registry System <http://www.epa.gov/srs/>
- Environmental Mutagen Information Center <http://library.wlu.ca/resource/emic>
- European Chemicals Agency http://echa.europa.eu/home_en.asp
- GENETOX <http://toxnet.nlm.nih.gov/cgi-bin/sis/htmlgen?GENETOX>
- Health Canada First Priority List Assessments http://www.hcsc.gc.ca/hecs_sesc/exsd/psli.htm
- Health Canada Second Priority List Assessments http://www.hcsc.gc.ca/hecs_sesc/exsd/psl2.htm
- Hazardous Substances Data Bank <http://toxnet.nlm.nih.gov/cgi-bin/sis/htmlgen?HSDB>
- IARC <http://monographs.iarc.fr/htdig/search.html>
- ILSI <http://www.ilsil.org/>
- IPCS INCHEM <http://www.inchem.org/>
- ITER http://iter.ctcnet.net/publicurl/pub_search_list.cfm
- NIOSHTIC 2 <http://www2.cdc.gov/nioshtic2/Nioshtic2.htm>
- US National Toxicology Program Management Status Report http://ntpserver.niehs.nih.gov/main_pages/NTP_ALL_STDY_PG.html
- US National Toxicology Program Results and Status Search http://ntpserver.niehs.nih.gov/main_pages/NTP_ALL_STDY_PG.html
- US National Toxicology Program Report on Carcinogens <http://ntpserver.niehs.nih.gov/NewHomeRoc/AboutRoC.html>
- TERA <http://www.tera.org/>
- Toxicology Data Network <http://toxnet.nlm.nih.gov/>
- Toxline <http://toxnet.nlm.nih.gov/cgi-bin/sis/htmlgen?TOXLINE>
- RTECS Toxcenter <http://www.cdc.gov/niosh/rtecs/default.html>
- WHO assessments – CICADS, EHC <http://www.who.int/ipcs/assessment/en/>
- USEPA Health and Environmental Studies Online <http://hero.epa.gov/>
- TSCA Test Submissions: <http://www.ntis.gov/products/ots.aspx>
- FIFRA docket: <http://www.regulations.gov>

Appendix III: Exclusion Criteria Screening Forms

Title and Abstract Screening Form

1. Check to exclude
 - Exclude (no triclosan exposure, no toxicity/adverse effects, no original data)
Definition: Article does not appear to contain original data (e.g. review article, systematic review, editorial, etc.)
2. Of interest/supplemental
 - Useful reviews, relevant information
Definition: Article will not be used in meta-analysis, but appears to provide useful information
3. Check to include
 - Include
Definition: Include; retrieve full article
4. Comments: (explain here if reason for exclusion is other than reasons provided in #1 above)

Full-Text Screening Form

1. Include this study? (check one):
 - Yes – study is relevant
Definition: Study meets inclusion criteria
 - Whole animal invertebrate study
Definition: Study meets inclusion criteria, but is categorized separately
 - Possibly (unclear)
Definition: Reviewer is uncertain as to whether study meets inclusion criteria
 - Useful information
Definition: Study does not meet inclusion criteria, but provides useful information
 - No - exclude (Possible reasons for exclusion: no original data, developmental and/or reproductive toxicity (DART) not reported, triclosan exposure doesn't precede DART outcomes, no comparator group, other; Note: DART outcomes of interest include but are not limited to, outcomes described in U.S. EPA guidelines for risk assessment of developmental and reproductive toxicity)
Definition: Study does not meet inclusion criteria
2. If you think this might be a duplicate study, explain why:
3. Language, if not English:

4. Comments: (explain here if reason for exclusion is other than reasons provided in #1 above)

Appendix IV: Data Collection Forms

The source criteria checklists for extraction terms include: gold standard publication checklist (GSPC) [36]; ARRIVE guidelines (ARRIVE) [37]; Cochrane Handbook for Systematic Reviews of Interventions data collection checklist (Cochrane) [2]; GRADE criteria for randomized control trials (GRADE) [1].

A. Data Collection for Non-human studies

Fields are free-form except where choices are shown in italics

Refid:

Reviewer:

- *Hanna*
- *Erica*
- *Other (identify yourself)*

Publication year:

Authors' declared conflicts of interest:

- *None declared*
- *Declared*

If declared, provide details:

Study funding source:

- *Government grant*
- *Industry funded*
- *Nonprofit organization grant*
- *Other (details below)*

Study funding source details:

What are the study objectives?:

Animal transportation mode to facility:

- *Truck*
- *Plane*
- *Other (details below)*
- *Not stated*

Animal transportation mode details:

Animal transportation shipping time:

Length of quarantine period after arrival at facility:

Origin and source of animals (commercial breeder, stock #, etc):

Site(s) of experiment (city, state, country):

Are animals immune compromised?:

- *Yes*
- *No*
- *Not stated*
- *Other (details below)*

Immune compromised details:

Microbiological status of facility:

- *SPF*
- *Germ-free*
- *Gnotobiotic*
- *Conventional*
- *Other (details below)*
- *Not stated*

Microbiological status of facility details:

Cage ventilation/setup:

- *Static, filter-top cages*

- *Ventilated caging system*
- *Other (details below)*
- *Not stated*

Cage ventilation/setup details:

Animal room temperature:

Animal room humidity:

Animal room lighting cycle:

- *12-hour light cycle*
- *Other (details below)*
- *Not stated*

Animal room lighting cycle details:

Number of animals per cage during experiment:

Was cage enrichment provided?:

- *Yes*
- *No*
- *Not stated*
- *Other (details below)*

Number of animals and cage enrichment details:

Nutrition (feed brand, ingredients, batch #, etc):

Feeding regimen:

- *Ad libitum*
- *Timed feedings*
- *Other (details below)*
- *Not stated*

Feeding regimen details:

Water provisions:

- *Ad libitum*
- *Timed administration*
- *Limited quantity offered*
- *Other (details below)*
- *Not stated*

Water provision details:

Water system type:

- *Water bottle*
- *Continuous automated watering system*
- *Other (details below)*
- *Not stated*

Water system type details:

Was drinking water purified or pre-treated?:

- *Yes*
- *No*
- *Not stated*
- *Other (details below)*

Water purification/pre-treated details:

Description of breeding methods for dams in experiment:

Have animals undergone previous procedures?:

- *Yes*
- *No*
- *Not stated*
- *Other (details below)*

Previous procedures details:

Euthanasia method for dams, if measuring fetal outcomes:

- *Carbon dioxide*
- *Cervical dislocation or decapitation*
- *Pharmaceutical intervention*
- *Other (details below)*
- *Not stated*

Euthanasia method details:

Description of animal wellbeing throughout study:

Miscellaneous comments by reviewer regarding study info:

Type of study:

- *Preconception*
- *Pregnancy*
- *Both preconception and pregnancy*
- *Multigenerational*
- *Fetal exposure only*
- *Other (details below)*

Type of study details:

Study design:

- *Randomized animal study*
- *Other (details below)*

Study design details:

Animal species :

- *Mouse*
- *Rat*

- *Chicken*
- *Other (details below)*

Animal species details:

Animal strain:

- *CD-1*
- *Sprague-Dawley*
- *Other (details below)*
- *Not stated*

Animal strain details (including exact genetic code, if available):

Animal genetic background:

- *Inbred*
- *Genetically modified (transgenic/knockout)*
- *Other (details below)*
- *Not stated*

Animal genetic background details:

Experimental generation for outcome measurement, if multigenerational study:

Age (at arrival, treatment, mating, pregnancy, etc.):

Developmental stage (at arrival):

- *Virgin*
- *Pregnant*
- *Other (details below)*

Developmental stage (at arrival) details:

Developmental stage (at treatment):

- *Virgin*
- *Pregnant*

- *Other (details below)*

Developmental stage (at treatment) details:

Total number of intervention groups:

Total number of control groups:

Number of animals allocated to each group:

If a power calculation was done, was the sample size of the experiment sufficient?:

- *Yes*
- *No*

Doses of triclosan administered:

Frequency of triclosan administration:

- *Daily*
- *Weekly*
- *Other (details below)*

Frequency of triclosan administration details:

How frequently animals weighed to determine dose:

- *Daily*
- *Weekly*
- *More than weekly*
- *Other (details below)*
- *Not stated*

How frequently animals weighed to determine dose details:

Duration of triclosan treatment:

- *Time window prior to pregnancy*
- *Time window prior to and during pregnancy*
- *Time window during pregnancy*

- *Other (details below)*

Duration of triclosan treatment details:

Administration route:

- *Gavage*
- *Other (details below)*

Administration route details:

Chemical name:

- *Triclosan (CAS# 3380-34-5)*
- *Other (details below)*

Chemical name details:

Details on chemical mixture preparation:

Other chemical information:

Outcomes measured:

Method of fetal growth measurement:

- *Weight*
- *Length*
- *Other (details below)*

Method of fetal growth measurement details:

Developmental stage for progeny at outcome measurement:

- *Near term (E18 for mice or E21 for rat)*
- *At birth*
- *Other (details below)*

Developmental stage for progeny outcome measurement details:

Unit of measurement (for weight, etc.):

- *Grams*

- *Millimeters*
- *Other (details below)*

Unit of measurement (for weight, etc.) details:

If applicable, how progeny weighed (individual or litter, etc.):

Sex (for animals where outcome measured):

- *Males only*
- *Females only*
- *Males and females*
- *Other (details below)*

Sex (for animals where outcome measured) details:

Were dead animals included in outcome assessment?:

- *Yes*
- *No*

How many animals/litters alive vs. dead?:

How was parturition monitored?:

- *Continuously*
- *AM check*
- *Other (details below)*
- *Not stated*

Parturition monitoring details:

Number animals analyzed (for intervention and outcome):

How were data reported (mean, median, raw data, etc.)?:

Growth measurement data for each group (ie, outcome):

How growth measurement data were reported (table, figure, etc):

How was precision reported (standard error, CI, etc.)?:

- *Standard error*
- *Standard deviation*
- *Confidence intervals*
- *Other (details below)*
- *Not stated*

How precision reported details:

Precision estimates:

How precision estimates were reported (table, figure, etc):

Miscellaneous comments by reviewer regarding data analysis:

B. Data Collection for Human studies

Fields are free-form except where choices (in italics) are shown

SOURCE

Refid:

Reviewer:

- (identify yourself)

Publication year:

Authors' declared conflicts of interest:

- *None declared*
- *Declared*

If declared, provide details:

Study funding source:

- *Government grant*
- *Industry funded*
- *Nonprofit organization grant*

- *Other*

Study funding source details:

What are the study objectives?:

Site(s) of data collection (city, state, country):

METHODS

Study duration:

Study design:

- *Cross-sectional*
- *Cohort, prospective*
- *Cohort, retrospective*
- *Case-control*
- *Ecological*
- *Other (list details below)*

Study design details:

Characteristics of study population:

- *Cohort (give description, e.g. NHANES 2004-2006)*
- *Sample size of total cohort*
- *Sample size (each exposure group)*
- *Age (each exposure group)*
- *Co-morbidities*
- *Other relevant details (list below)*

Study subject details:

Exposure measurement:

- *Preconception*

- *Pregnancy*
- *Both preconception and pregnancy*
- *Multigenerational*
- *Fetal/infant exposure only*
- *Other (details below)*
- *Record when exposure occurred or was measured, in relation to outcome measurement*

Exposure measurement details:

Source of exposure data:

- *Biomonitoring (list specific matrix)*
- *Environmental monitoring (list specific matrix)*
- *Questionnaire (list specific determinant of exposure)*
- *Other (specify)*

Total number of exposed groups:

Total number of non-exposed groups:

Number of subjects in each group:

Concentrations of triclosan measured, and units:

Frequency of triclosan measurements if more than once:

Number of replicate measurements taken:

Chemical name:

- *Triclosan*
- *Other (details below)*

Chemical name details:

Other chemical information:

Outcomes measured:

Method of fetal growth measurement:

- *Weight*
- *Length*
- *Other (details below)*

Method of fetal growth measurement details:

Gestational age at outcome measurement:

- *At birth*
- *Other (details below)*

Birth outcome measurement details:

Unit of measurement (for weight, etc.):

- *Grams*
- *Millimeters*
- *Other (details below)*

Unit of measurement (for weight, etc.) details:

Sex (where outcome measured):

- *Males only*
- *Females only*
- *Males and females*
- *Other (details below)*

Number subjects analyzed (for exposure and outcome):

Number of missing participants:

RESULTS

Statistical methods:

- *Statistical tests employed*

- Statistic (odds ratio, adjusted odds ratio, beta estimate, etc.)
- p-values given
- Confidence intervals given
- Confounding adjustments in statistical tests

Were known confounders accounted for by study design?

Were known confounders accounted for by analysis?

How were data reported (mean, median, raw data, etc.)?:

Growth measurement data for each group (ie, outcome):

How growth measurement data were reported (table, figure, etc):

Summary data for each exposure group

Estimate of effect with confidence interval and p-value

How was precision reported (standard error, CI, etc.)?:

- *Standard error*
- *Standard deviation*
- *Confidence intervals*
- *Other (details below)*
- *Not stated*

How precision reported details:

Precision estimates:

How precision estimates were reported (table, figure, etc):

Miscellaneous comments by reviewer regarding data analysis:

Appendix V: Instructions for Making Risk of Bias Determinations

Please answer LOW RISK, PROBABLY LOW RISK, PROBABLY HIGH RISK, HIGH RISK or NOT APPLICABLE and provide details/justification.

A. Non-human Experimental Studies

1. SEQUENCE GENERATION

Adequate sequence generation minimizes bias by using a random component to ensure the sequence is unpredictable.

Was the allocation sequence adequately generated?

Criteria for a judgment of LOW risk of bias:

The investigators describe a random component in the sequence generation process such as:

- Referring to a random number table;
- Using a computer random number generator;
- Coin tossing;
- Shuffling cards or envelopes;
- Throwing dice;
- Drawing of lots.

Note that use of minimization (e.g., ensuring similar animal weights for all groups) does not put the study at risk of bias if combined with a random component.

Criteria for the judgment of PROBABLY LOW risk of bias:

There is insufficient information about the sequence generation process to permit a judgment of low risk of bias, but there is indirect evidence that suggests the sequence generation process was random, as described by the criteria for a judgment of low risk of bias, such as:

- Study authors make a simple statement such as ‘we randomly allocated’, but do not provide details regarding specific random component used in the sequence generation process; or
- Study authors describe randomization for one experiment, and the methods for a second experiment are similar but do not specifically mention randomization.

Criteria for the judgment of PROBABLY HIGH risk of bias:

There is insufficient information about the sequence generation process to permit a judgment of high risk of bias, but there is indirect evidence that suggests a non-random component in the sequence generation process or that a random component, as described by the criteria for a judgment of low risk of bias, was not used such as:

- Study authors do not make any statement about sequence generation and the review author does not find indirect evidence suggesting random sequence generation.

Criteria for the judgment of HIGH risk of bias:

The investigators state clearly that a random component was not used or describe a non-random component in the sequence generation process. Usually, the description would involve some systematic, non-random approach, for example:

- Sequence generated by date of birth;
- Sequence generated by some rule based on date (or day) of arrival at facility;
- Sequence generated by some rule based on record number.

Other non-random approaches happen much less frequently than the systematic approaches mentioned above and tend to be obvious. They usually involve judgment or some method of non-random categorization of animals, for example:

- Allocation by judgment of the investigator;
- Allocation by availability of the intervention;
- Alternate allocation.

2. ALLOCATION CONCEALMENT

Adequate allocation concealment minimizes bias by protecting the allocation sequence before and until assignment.

Was allocation adequately concealed?

Criteria for a judgment of LOW risk of bias:

Investigators could not foresee assignment because one of the following, or an equivalent method, was used to conceal allocation:

- Sequentially numbered treatment containers of identical appearance to control; or
- Sequentially numbered prepared route of administration (e.g., pre-prepared water dosed with chemical) of identical appearance; or
- Study personnel assigned allocation, and separate study personnel administered treatment without knowledge of assignments; or
- Sequentially numbered, opaque, sealed envelopes.

Criteria for the judgment of PROBABLY LOW risk of bias:

There is insufficient information about allocation concealment to permit a judgment of low risk of bias, but there is indirect evidence that suggests the allocation was adequately concealed, as described by the criteria for a judgment of low risk of bias such as:

- Review author finds indirect evidence suggesting allocation concealment, but study authors do not provide details about how investigators were prevented from foreseeing assignment; or

- Study authors state that animals were given identification numbers prior to treatment, or authors describe allocation concealment for one experiment, and the methods for a second experiment are similar but do not specifically mention allocation concealment.

Criteria for the judgment of PROBABLY HIGH risk of bias:

There is insufficient information about allocation concealment to permit a judgment of high risk of bias, but there is indirect evidence that suggests the allocation was not adequately concealed, as described by the criteria for a judgment of high risk of bias such as:

- Study authors do not make any statement about allocation concealment and the review author does not find indirect evidence suggesting allocation concealment.

Criteria for the judgment of HIGH risk of bias:

Investigators handling experimental animals could possibly foresee assignments and thus introduce bias, such as allocation based on:

- Using an open random allocation schedule (e.g. a list of random numbers); or
- Alternation or rotation; or
- Non-random and known criteria, such as date of birth; or
- Same study personnel performed sequence generation and administered initial treatment; or
- Any other explicitly unconcealed procedure.

3. BLINDING OF PERSONNEL AND OUTCOME ASSESSORS

Adequate blinding minimizes bias by protecting the sequence after assignment.

Was knowledge of the allocated interventions adequately prevented during the study?

Criteria for a judgment of LOW risk of bias:

Any one of the following:

- No blinding, but the review author judges that the outcome and the outcome measurement are not likely to be influenced by lack of blinding (e.g., lab test performed by a source not connected with the study); or
- Investigators report blinding of key study personnel; or
- Some key study personnel were not blinded, but outcome assessment was blinded and the non-blinding of others unlikely to introduce bias.

Criteria for the judgment of PROBABLY LOW risk of bias:

There is insufficient information about blinding to permit a judgment of low risk of bias, but there is indirect evidence that suggests the study was adequately blinded, as described by the criteria for a judgment of low risk of bias such as:

- Study authors state that some study personnel were blinded, but it is unclear if all important personnel were blinded; or
- Study authors state that animals were given identification numbers prior to outcome assessment; or
- Study authors describe blinding for one experiment, and the methods for a second experiment are similar but do not specifically mention blinding; or
- The review author judges certain aspects of the outcome or outcome measurement are unlikely to be influenced by lack of blinding, but the review author does not feel confident enough to permit a low risk of bias judgment.

Criteria for the judgment of PROBABLY HIGH risk of bias:

There is insufficient information about blinding to permit a judgment of high risk of bias, but there is indirect evidence that suggests the study was not adequately blinded, as described by the criteria for a judgment of high risk of bias such as:

- Study authors do not make any statement about blinding and the review author does not find indirect evidence suggesting blinding.

Criteria for the judgment of HIGH risk of bias:

Any one of the following:

- No blinding or incomplete blinding, and the outcome or outcome measurement is likely to be influenced by lack of blinding; or
- Some key study personnel were not blinded, and the non-blinding of others likely to introduce bias; or
- Study authors state the study is “open label” (i.e., study was conducted such that investigators were aware of assignments to treatment groups).

4. INCOMPLETE OUTCOME DATA

Missing outcome data, due to exclusion during the study or the analysis, raise the possibility that the observed treatment effect is biased; addressing incomplete outcome data minimizes this potential bias.

Were incomplete outcome data adequately addressed?

Criteria for a judgment of LOW risk of bias:

Review author is confident that the animals included in the analysis are exactly those who were randomized into the experiment. The number of animals allocated to treatment groups is reported for outcomes of interest and data are provided indicating adequate follow up of all animals from the beginning of the study (including for all offspring, if applicable), or any one of the following:

- The number of animals allocated is reported and matches the number of animals reported for each outcome (i.e., no missing outcome data); or

- Reasons for missing outcome data unlikely to be related to true outcome (for survival data, censoring unlikely to be introducing bias); or
- Missing outcome data is provided and is balanced in numbers across treatment groups, with similar reasons for missing data across groups; or
- For dichotomous outcome data, the proportion of missing outcomes compared with the observed frequency of the outcome is not enough to have a biologically relevant impact on the outcome results; or
- For continuous outcome data, plausible change in outcome (difference in means or standardized difference in means) among missing outcomes not enough to have a biologically relevant impact on the outcome results.

Criteria for the judgment of PROBABLY LOW risk of bias:

There is insufficient information about incomplete outcome data to permit a judgment of low risk of bias, but there is indirect evidence that suggests incomplete outcome data were adequately addressed, as described by the criteria for a judgment of low risk of bias such as:

- Study authors do not report numbers of animals allocated to treatment groups or only provide a range of numbers, but provide data indicating adequate follow up of all animals from the beginning of the study (including offspring, if applicable); or
- Study authors report number of animals allocated to treatment groups, but do not provide data indicating adequate follow up for a subset of animals or only provide a qualitative statement about missing outcome data (e.g., authors state equal losses for all treatment and control groups).

Criteria for the judgment of PROBABLY HIGH risk of bias:

There is insufficient information about incomplete outcome data to permit a judgment of high risk of bias, but there is indirect evidence that suggests incomplete outcome data was not adequately addressed, as described by the criteria for a judgment of high risk of bias such as:

- Study authors do not report numbers of animals allocated to treatment groups, but do provide data indicating adequate follow up for a subset of animals or provide a qualitative statement about missing outcome data (e.g., authors state equal losses for all treatment and control groups); or
- Study authors provide a range for numbers of animals allocated to treatment groups, but do not provide data indicating adequate follow up of all animals from beginning of study (including offspring, if applicable) or only provide a qualitative statement about missing outcome data (e.g., authors state equal losses for all treatment and control groups); or
- Study authors analyze a *randomly* selected subset of animals for outcomes of interest (e.g., weighed a subset of dams or a subset of pups per litter).

Criteria for the judgment of HIGH risk of bias:

Review author is not confident that the animals included in the analysis are exactly those who were randomized into the experiment. The number of animals allocated to treatment groups is not reported for outcomes of interest and either one of the following:

- Data are not provided to indicate that there was adequate follow up of all animals from the beginning of the experiment (including offspring, if applicable); or
- Only a subset of animals were examined for outcome of interest (e.g., weighed a subset of dams or a subset of pups per litter), and study authors did not specify that selection of the subset was random or the selection included a non-random component.

Additionally, any one of the following:

- Reason for missing outcome data likely to be related to true outcome, with either imbalance in numbers or reasons for missing data across treatment groups; or
- For dichotomous outcome data, the proportion of missing outcomes compared with the observed frequency of the outcome is enough to have a biologically relevant impact on the outcome results; or
- For continuous outcome data, change in outcome (difference in means or standardized difference in means) among missing is enough to have a biologically relevant impact on the outcome results.

5. SELECTIVE OUTCOME REPORTING

Selective outcome reporting may introduce a risk of bias if study authors exclude a subset of the original variables recorded, on the basis of the results, from the report or publication.

Are reports of the study free of suggestion of selective outcome reporting?

Criteria for a judgment of LOW risk of bias:

All of the study's pre-specified (primary and secondary) outcomes outlined in the protocol, methods, abstract, and/or introduction that are of interest in the review have been reported in the pre-specified way (i.e., the outcomes outlined in the methods section match what is reported in the results section and vice versa), and the number of animals analyzed for outcomes of interest is provided.

Criteria for the judgment of PROBABLY LOW risk of bias:

There is insufficient information about selective outcome reporting to permit a judgment of low risk of bias, but there is indirect evidence that suggests the study was free of selective reporting, as described by the criteria for a judgment of low risk of bias such as:

- All of the study's pre-specified (primary and secondary) outcomes outlined in the protocol, methods, abstract, and/or introduction that are of interest in the review have been reported in the pre-specified way, but study authors report the number of animals analyzed for outcomes of interest as a range or report values for which numbers of animals analyzed need to be calculated by the review author; or
- Not all of the study's pre-specified primary outcomes (as outlined in the protocol, methods, abstract, and/or introduction) that are of interest in the review have

been reported in the pre-specified way, but study authors provided detailed raw data for outcomes of interest.

Criteria for the judgment of PROBABLY HIGH risk of bias:

There is insufficient information about selective outcome reporting to permit a judgment of high risk of bias, but there is indirect evidence that suggests the study was not free of selective reporting, as described by the criteria for a judgment of high risk of bias such as:

- All of the study's pre-specified (primary and secondary) outcomes outlined in the protocol, methods, abstract, and/or introduction that are of interest in the review have been reported in the pre-specified way, but study authors do not report the number of animals analyzed for outcomes of interest; or
- Not all of the study's pre-specified primary outcomes (as outlined in the protocol, methods, abstract, and/or introduction) that are of interest in the review have been reported, but study authors report the number of animals analyzed for outcomes of interest, or report the numbers as a range, or report values for which numbers of animals analyzed need to be calculated by the review author.

Criteria for the judgment of HIGH risk of bias:

One of more of the following:

- Not all of the study's pre-specified primary outcomes (as outlined in the protocol, methods, abstract, and/or introduction) that are of interest in the review have been reported in the pre-specified way (i.e., the outcomes outlined in the methods section do not match what is reported in the results section or vice versa), and the number of animals analyzed for outcomes of interest is not provided; or
- The study report fails to include results for a key outcome that would be expected to have been reported for such a study.

6. CONFLICT OF INTEREST

Conflicts of interest may introduce risk of bias when outside financial interests compromise, or have the appearance of compromising, the design, conduct, or outcome of the study.

Was the study free of support from a company, study author, or other party having a financial interest in any of the treatments studied?

Criteria for a judgment of LOW risk of bias:

The study did not receive support from a company, study author, or other party having a financial interest in the outcome of the study. A conflict of interest statement is provided to indicate the study authors have no financial interests and there is evidence of the parties not having a financial interest. Examples of this evidence include the following:

- Funding source is limited to government, non-profit organizations, or academic grants funded by government, foundations and/or non-profit organizations without financial interest in the treatments studied;
- Chemicals or other treatments used in study were purchased from a supplier or donated by a party without financial interest in the treatments studied;
- Staff affiliated with financially interested parties are not mentioned in the acknowledgements section;
- Parties with a financial interest in the outcome of the study were not involved in the design, conduct, analysis, or reporting of the study and authors had complete access to the data;
- Study authors make a claim denying conflicts of interest;
- Study authors are not affiliated with companies with financial interest, and there is no reason to believe a conflict of interest exists;
- All study authors are affiliated with a government agency (are prohibited from involvement in projects for which there is a conflict of interest or an appearance of conflict of interest).

Criteria for the judgment of PROBABLY LOW risk of bias:

There is insufficient information to permit a judgment of low risk of bias, but there is indirect evidence that suggests the study is free of conflicts of interest, as described by the criteria for a judgment of low risk of bias, such as:

- A conflict of interest statement denying financial interests is not provided, but associated funds and/or persons appear to be free of financial interests in study outcome and are unaffiliated with parties with a financial interest.

Criteria for the judgment of PROBABLY HIGH risk of bias:

There is insufficient information to permit a judgment of high risk of bias, but there is indirect evidence that suggests the study is not free of conflicts of interest, as described by the criteria for a judgment of high risk of bias, such as:

- A conflict of interest statement denying financial interests is provided, but the study received support from a company, study author, or other party having financial interests in the study outcome.

Criteria for the judgment of HIGH risk of bias:

The study received support from a company, study author, or other party having a financial interest in the outcome of the study. Examples of support include:

- Research funds;
- Writing services;
- Author/staff from study was an employee of or otherwise affiliated with a company or other party having a financial interest;
- Company or other party with financial interest limited author access to the data;
- Party with financial interest was involved in the design, conduct, analysis, or reporting of the study;
- Study authors claim a conflict of interest.

7. OTHER POTENTIAL THREATS TO VALIDITY

Other potential threats to validity can include any potential risk of bias identified by the review author that is not otherwise addressed in the other domains.

Was the study apparently free of other problems that could put it at a risk of bias?

Criteria for a judgment of LOW risk of bias:

The study appears to be free of other sources of bias.

Criteria for the judgment of PROBABLY LOW risk of bias:

There is insufficient information to permit a judgment of low risk of bias, but there is indirect evidence that suggests the study was free of other threats to validity, as described by the criteria for a judgment of low risk of bias.

Criteria for the judgment of PROBABLY HIGH risk of bias:

There is insufficient information to permit a judgment of high risk of bias, but there is indirect evidence that suggests the study was not free of other threats to validity, as described by the criteria for a judgment of high risk of bias.

Criteria for the judgment of HIGH risk of bias:

There is at least one important risk of bias. For example, the study:

- Stopped early due to some data-dependent process;
- Had extreme baseline imbalance (improper control group);
- Has been claimed to have been fraudulent;
- The conduct of the study is affected by interim results (e.g. using additional animals from a subgroup showing a greater effect);
- There is deviation from the study methods in a way that does not reflect typical practice;
- There is pre-randomization administration of a treatment that could enhance or diminish the effect of a subsequent, randomized, intervention; inappropriate administration of an intervention (or co-intervention);
- Occurrence of “null bias” due to treatments being insufficiently well delivered or overly wide inclusion criteria;
- An insensitive instrument is used to measure outcomes (which can lead to underestimation of both beneficial and harmful effects);
- Selective reporting of subgroups;
- Had some other problem.

B. Human Studies

1. Are the study groups free from baseline differences?

Criteria for a judgment of LOW risk of bias:

There were no baseline differences among study groups or adjustment techniques were used to correct for the differences.

Examples of baseline differences:

- Protocols for recruitment or inclusion/exclusion criteria were applied differently across study groups
- Study participants were recruited at different times
- Study participants were recruited from different populations and proportions of participants from each population in each study group are not uniform
- Participation rates were inadequate or not comparable across study groups

Criteria for the judgment of PROBABLY LOW risk of bias:

There is insufficient information about participant selection to permit a judgment of low risk of bias, but there is indirect evidence that suggests that participant recruitment and inclusion/exclusion criteria was consistent, as described by the criteria for a judgment of low risk of bias.

Criteria for the judgment of PROBABLY HIGH risk of bias:

There is insufficient information about participant selection to permit a judgment of high risk of bias, but there is indirect evidence that suggests that participant recruitment or inclusion/exclusion criteria was inconsistent, as described by the criteria for a judgment of high risk of bias.

Criteria for the judgment of HIGH risk of bias:

There were baseline differences among study groups and no adjustment was used to correct for the differences, such as:

- Protocols for recruitment or inclusion/exclusion criteria were applied differently across study groups
- Study participants were recruited at different times
- Study participants were recruited from different populations and proportions of participants from each population in each study group are not uniform
- Participation rates were inadequate or not comparable across study groups

Criteria for the judgment of NOT APPLICABLE (risk of bias domain is not applicable to study):

There is evidence that participant selection is not an element of study design capable of introducing risk of bias in the study.

2. Was knowledge of the exposure groups adequately prevented during the study?

Criteria for a judgment of LOW risk of bias:

Any one of the following:

- No blinding, but the review authors judge that the outcome and the outcome measurement are not likely to be influenced by lack of blinding; or
- Blinding of key study personnel was ensured, and it is unlikely that the blinding could have been broken; or
- Some key study personnel were not blinded, but outcome assessment was blinded and the non-blinding of others is unlikely to introduce bias.

Criteria for the judgment of PROBABLY LOW risk of bias:

There is insufficient information about blinding to permit a judgment of low risk of bias, but there is indirect evidence that suggests the study was adequately blinded, as described by the criteria for a judgment of low risk of bias. For example, investigators were effectively blinded to the exposure and outcome groups, as the exposure was measured separately and the outcome was obtained from a hospital record.

Criteria for the judgment of PROBABLY HIGH risk of bias:

There is insufficient information about blinding to permit a judgment of high risk of bias, but there is indirect evidence that suggests the study was not adequately blinded, as described by the criteria for a judgment of high risk of bias.

Criteria for the judgment of HIGH risk of bias:

Any one of the following:

- No blinding or incomplete blinding, and the outcome or outcome measurement is likely to be influenced by lack of blinding; or
- Blinding of key study personnel attempted, but likely that the blinding could have been broken; or
- Some key study personnel were not blinded, and the non-blinding of others likely to introduce bias.

Criteria for the judgment of NOT APPLICABLE (risk of bias domain is not applicable to study):

There is evidence that blinding is not an element of study design capable of introducing risk of bias in the study.

3. Were exposure assessment methods robust?

Criteria for a judgment of LOW risk of bias:

The reviewers judge that there is low risk of exposure misclassification and:

- There is high confidence in the accuracy of the exposure assessment methods, such as methods that have been tested for validity and reliability in measuring the targeted exposure; or
- Less-established or less direct exposure measurements are validated against well-established or direct methods

AND if applicable, appropriate QA/QC for methods are described and are satisfactory, with at least three of the following items reported, or at least two of the following items reported plus evidence of satisfactory performance in a high quality inter-laboratory comparison:

- Limit of detection or quantification;
- standards recovery;
- measure of repeatability;
- investigation and prevention of blanks contamination.

Criteria for the judgment of PROBABLY LOW risk of bias:

There is insufficient information about the exposure assessment methods to permit a judgment of low risk of bias, but there is indirect evidence that suggests that methods were robust, as described by the criteria for a judgment of low risk of bias. Studies only reporting that the QA/QC items above were satisfactory but not reporting all of the actual numbers may receive a judgment of “probably low risk of bias.”

Criteria for the judgment of PROBABLY HIGH risk of bias):

There is insufficient information about the exposure assessment methods to permit a judgment of high risk of bias, but there is indirect evidence that suggests that methods were not robust, as described by the criteria for a judgment of high risk of bias.

Criteria for the judgment of ‘NO’ (i.e. high risk of bias):

The reviewers judge that there is high risk of exposure misclassification and any one of the following:

- There is low confidence in the accuracy of the exposure assessment methods; or
- Less-established or less direct exposure measurements are not validated and are suspected to introduce bias that impacts the outcome assessment (example: participants are asked to report exposure status retrospectively, subject to recall bias)
- Uncertain how exposure information was obtained

Criteria for the judgment of NOT APPLICABLE (risk of bias domain is not applicable to study):

There is evidence that exposure assessment methods are not capable of introducing risk of bias in the study.

4. Were outcome assessment methods robust?

Criteria for a judgment of LOW risk of bias:

The reviewers judge that there is low risk of outcome misclassification and:

- Outcomes were assessed and defined consistently across all study participants, using valid and reliable measures; or
- Less-established or less direct outcome measurements are validated against well-established or direct methods
- AND, if applicable, appropriate QA/QC for methods are described and are satisfactory.

Criteria for the judgment of PROBABLY LOW risk of bias:

There is insufficient information about the outcome assessment methods to permit a judgment of low risk of bias, but there is indirect evidence that suggests that methods were robust, as described by the criteria for a judgment of low risk of bias. Appropriate QA/QC for methods are not described but the review authors judge that the outcome and the outcome assessment are objective and uniform across study groups.

Criteria for the judgment of PROBABLY HIGH risk of bias:

There is insufficient information about the outcome assessment methods to permit a judgment of high risk of bias, but there is indirect evidence that suggests that methods were not robust, as described by the criteria for a judgment of high risk of bias.

Criteria for the judgment of HIGH risk of bias:

The reviewers judge that there is high risk of outcome misclassification and any one of the following:

- There is low confidence in the accuracy of the outcome assessment methods; or
- Less-established or less direct outcome measurements are not validated and are suspected to introduce bias that impacts the outcome assessment
- Uncertain how outcome information was obtained

Criteria for the judgment of NOT APPLICABLE (risk of bias domain is not applicable to study):

There is evidence that outcome assessment methods are not capable of introducing risk of bias in the study.

5. Were confounding and effect modification adequately addressed?

Criteria for a judgment of LOW risk of bias:

The study appropriately assessed and accounted for (i.e., matched, stratified, multivariate analysis or otherwise statistically controlled for) important potential confounders, or reported that potential confounders were evaluated and omitted because inclusion did not substantially affect the results. The determination of specific confounders may be informed by, but not limited to, the studies included in the review. Potential interaction or effect modification was evaluated and adequately addressed.
AND the important potential confounders and effect modifiers were measured consistently across study groups using valid and reliable methods.

Criteria for the judgment of PROBABLY LOW risk of bias:

The study accounted for most but not all of the important potential confounders and effect modifiers
AND this lack of accounting is not expected to introduce substantial bias.

Criteria for the judgment of PROBABLY HIGH risk of bias:

The study accounted for some but not all of the important potential confounders and effect modifiers
AND this lack of accounting may have introduced substantial bias.

Criteria for the judgment of HIGH risk of bias:

The study did not account for or evaluate important potential confounders or effect modifiers.
OR the important potential confounders and effect modifiers were not measured consistently across study groups using valid and reliable methods.

6. Were incomplete outcome data adequately addressed?

Criteria for a judgment of LOW risk of bias:

Participants were followed long enough to obtain outcome measurements and:

- No missing outcome data; or
- Reasons for missing outcome data unlikely to be related to true outcome (for survival data, censoring unlikely to introduce bias); or
- Attrition or missing outcome data balanced in numbers across exposure groups, with similar reasons for missing data across groups; or
- For dichotomous outcome data, the proportion of missing outcomes compared with observed event risk not enough to have a relevant impact on the intervention effect estimate; or
- For continuous outcome data, plausible effect size (difference in means or standardized difference in means) among missing outcomes not enough to have a relevant impact on the observed effect size; or
- Missing data have been imputed using appropriate methods

Criteria for the judgment of PROBABLY LOW risk of bias:

There is insufficient information about incomplete outcome data to permit a judgment of low risk of bias, but there is indirect evidence that suggests incomplete outcome data was adequately addressed, as described by the criteria for a judgment of low risk of bias.

Criteria for the judgment of PROBABLY HIGH risk of bias:

There is insufficient information about incomplete outcome data to permit a judgment of high risk of bias, but there is indirect evidence that suggests incomplete outcome data was not adequately addressed, as described by the criteria for a judgment of high risk of bias.

Criteria for the judgment of HIGH risk of bias:

Participants were not followed long enough to obtain outcome measurements OR
Any one of the following:

- Reason for missing outcome data likely to be related to true outcome, with either imbalance in numbers or reasons for missing data across exposure groups; or
- For dichotomous outcome data, the proportion of missing outcomes compared with observed event risk enough to induce biologically relevant bias in intervention effect estimate; or
- For continuous outcome data, plausible effect size (difference in means or standardized difference in means) among missing outcomes enough to induce biologically relevant bias in observed effect size; or
- Potentially inappropriate application of imputation.

Criteria for the judgment of NOT APPLICABLE (risk of bias domain is not applicable to study):

There is evidence that incomplete outcome data is not capable of introducing risk of bias in the study.

7. Are reports of the study free of suggestion of selective outcome reporting?

Criteria for a judgment of LOW risk of bias:

All of the study's pre-specified (primary and secondary) outcomes outlined in the protocol, methods, abstract, and/or introduction that are of interest in the review have been reported in the pre-specified way.

Criteria for the judgment of PROBABLY LOW risk of bias:

There is insufficient information about selective outcome reporting to permit a judgment of low risk of bias, but there is indirect evidence that suggests the study was free of selective reporting, as described by the criteria for a judgment of low risk of bias.

Criteria for the judgment of PROBABLY HIGH risk of bias:

There is insufficient information about selective outcome reporting to permit a judgment of high risk of bias, but there is indirect evidence that suggests the study was not free of selective reporting, as described by the criteria for a judgment of high risk of bias.

Criteria for the judgment of HIGH risk of bias:

Any one of the following:

- Not all of the study's pre-specified primary outcomes (as outlined in the protocol, methods, abstract, and/or introduction) have been reported; or
- One or more primary outcomes is reported using measurements, analysis methods or subsets of the data (e.g. subscales) that were not pre-specified; or
- One or more reported primary outcomes were not pre-specified (unless clear justification for their reporting is provided, such as an unexpected effect); or
- One or more outcomes of interest are reported incompletely

Criteria for the judgment of NOT APPLICABLE (risk of bias domain is not applicable to study):

There is evidence that selective outcome reporting is not capable of introducing risk of bias in the study.

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?

Criteria for a judgment of LOW risk of bias:

The study did not receive support from a company, study author, or other entity having a financial interest in the outcome of the study. Examples include the following:

- Funding source is limited to government, non-profit organizations, or academic grants funded by government, foundations and/or non-profit organizations;
- Chemicals or other treatment used in study were purchased from a supplier;
- Company affiliated staff are not mentioned in the acknowledgements section;
- Authors were not employees of a company with a financial interest in the outcome of the study;
- Company with a financial interest in the outcome of the study was not involved in the design, conduct, analysis, or reporting of the study and authors had complete access to the data;
- Study authors make a claim denying conflicts of interest;
- Study authors are unaffiliated with companies with financial interest, and there is no reason to believe a conflict of interest exists;
- All study authors are affiliated with a government agency (are prohibited from involvement in projects for which there is a conflict of interest or an appearance of conflict of interest).

Criteria for the judgment of PROBABLY LOW risk of bias:

There is insufficient information to permit a judgment of low risk of bias, but there is indirect evidence that suggests the study was free of support from a company, study author, or other entity having a financial interest in the outcome of the study, as described by the criteria for a judgment of low risk of bias.

Criteria for the judgment of PROBABLY HIGH risk of bias:

There is insufficient information to permit a judgment of high risk of bias, but there is indirect evidence that suggests the study was not free of support from a company, study author, or other entity having a financial interest in the outcome of the study, as described by the criteria for a judgment of high risk of bias.

Criteria for the judgment of HIGH risk of bias:

The study received support from a company, study author, or other entity having a financial interest in the outcome of the study. Examples of support include:

- Research funds;
- Chemicals provided at no cost;
- Writing services;
- Author/staff from study was employee or otherwise affiliated with company with financial interest;
- Company limited author access to the data;
- Company was involved in the design, conduct, analysis, or reporting of the study;
- Study authors claim a conflict of interest

Criteria for the judgment of NOT APPLICABLE (risk of bias domain is not applicable to study):

There is evidence that conflicts of interest are not capable of introducing risk of bias in the study.

9. Was the study apparently free of other problems that could put it at a risk of bias?

Criteria for a judgment of LOW risk of bias:

The study appears to be free of other sources of bias.

Criteria for the judgment of PROBABLY LOW risk of bias:

There is insufficient information to permit a judgment of low risk of bias, but there is indirect evidence that suggests the study was free of other threats to validity.

Criteria for the judgment of PROBABLY HIGH risk of bias:

There is insufficient information to permit a judgment of high risk of bias, but there is indirect evidence that suggests the study was not free of other threats to validity, as described by the criteria for a judgment of high risk of bias.

Criteria for the judgment of HIGH risk of bias:

There is at least one important risk of bias. For example, the study:

- Had a potential source of bias related to the specific study design used; or
- Stopped early due to some data-dependent process (including a formal-stopping rule); or
- The conduct of the study is affected by interim results (e.g. recruiting additional participants from a subgroup showing greater or lesser effect); or
- Has been claimed to have been fraudulent; or
- Had some other problem

Appendix VI. Instructions for grading the quality and strength of evidence

A. Grading Quality

Each of the categories to consider in downgrading or upgrading the evidence is described in detail below. Please record your results on the chart at the end of each category, including a brief explanation for your ratings.

Downgrade Categories

Category 1. Quality of Study Limitations (Risk of Bias)[46]

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

The evidence from studies can be rated down if most of the relevant evidence comes from studies that suffer from a high risk of bias. Risk of bias is rated by outcome across studies. Study limitations for each outcome for individual studies and across studies are summarized in the heat maps.

GRADE outlines the following principles for moving from risk of bias in individual studies to rating quality of evidence across studies.

1. In deciding on the overall quality of evidence, one does not average across studies (for instance if some studies have no serious limitations, some serious limitations, and some very serious limitations, one does not automatically rate quality down by one level because of an average rating of serious limitations). Rather, judicious consideration of the contribution of each study, with a general guide to focus on the high-quality studies is warranted.^a
2. This judicious consideration requires evaluating the extent to which each study contributes toward the estimate of magnitude of effect. The contribution that each study makes will usually reflect study sample size and number of outcome events. Larger studies with many events will

^a Note: Limitations to GRADE's risk of bias assessments as stated by GRADE: "First, empirical evidence supporting the criteria is limited. Attempts to show systematic difference between studies that meet and do not meet specific criteria have shown inconsistent results. Second, the relative weight one should put on the criteria remains uncertain. The GRADE approach is less comprehensive than many systems, emphasizing simplicity and parsimony over completeness. GRADE's approach does not provide a quantitative rating of risk of bias. Although such a rating has advantages, we share with the Cochrane Collaboration methodologists a reluctance to provide a risk of bias score that, by its nature, must make questionable assumptions about the relative extent of bias associated with individual items and fails to consider the context of the individual items."

contribute more, much larger studies with many more events will contribute much more.

3. One should be conservative in the judgment of rating down. That is, one should be confident that there is substantial risk of bias across most of the body of available evidence before one rates down for risk of bias.

4. The risk of bias should be considered in the context of other limitations. If, for instance, reviewers find themselves in a close-call situation with respect to two quality issues (risk of bias and, say, precision), GRADE suggests rating down for at least one of the two.

5. Notwithstanding the first four principles, reviewers will face close-call situations. You should acknowledge that you are in such a situation, make it explicit why you think this is the case, and make the reasons for your ultimate judgment apparent.

Rating for Risk of Bias (Study Limitations)		Rationale for your judgment
0 no change -1 decrease quality 1 level -2 decrease quality 2 levels		
Human		
Non-Human Mammalian		

Category 2. Indirectness of Evidence

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

Quality of evidence (your confidence in estimates of effect) may decrease when substantial differences exist between the population, the exposure, or the outcomes measured in research studies under consideration in the review.

Evidence is direct when it directly compares the exposures in which we are interested when applied to the populations in which we are interested and measures outcomes important to the study question (in GRADE the outcomes must be important to patients).

Based on GRADE [47], evidence can be indirect in one of three ways.^b

1. The population studied differs from the population of interest (the term applicability is often used for this form of indirectness). **Please note the Navigation Guide’s *a priori* assumption is that mammalian evidence of a health effect/lack of health effect is deemed to be direct evidence of human health with regards to directness of the population.** This is a marked departure from GRADE,^c based on empirical evidence in environmental health science that the reliability of experimental animal (mammalian) data for reproductive and developmental health has been well established through multiple studies of concordance between mammalian animals and humans after exposure to a variety of chemical agents [44, 48-51]. Presently, there is no example of a chemical agent that has adversely affected human reproduction or development but has not caused the same or similar adverse effects in mammalian animal models [44]. The National Academy of Sciences (NAS) has recognized the importance of animal data in identifying potential developmental risks. According to the NAS, studies of comparison between developmental effects in animals and humans find that “there is concordance of developmental effects between animals and humans and that humans are as sensitive or more sensitive than the most sensitive animal species [52].” GRADE states that in general, one should not rate down for population differences unless one has compelling reason to think that the biology in the population of interest is so different than the population tested that the magnitude of effect will differ substantially. According to GRADE, most often, this will not be the case. In applying this GRADE principle to the Navigation Guide, non-human evidence would be rated down as indirect when it is a biologically inappropriate non-human model system for the health outcome under study.
2. The intervention (exposure) tested may differ from the exposure of interest, i.e., a difference in the chemical, route and/or dose. Decisions regarding indirectness of populations and exposure depend on an understanding of whether biological or social factors are sufficiently different that one might expect substantial differences in the magnitude of effect. GRADE also states, “As with all other aspects of rating quality of evidence, there is a continuum of similarity of the intervention that will require judgment. It is rare, and usually unnecessary, for the intended populations and interventions to be identical to those in the studies, and we should only rate down if the differences are considered sufficient to make a difference in outcome likely.”

^b GRADE includes a fourth type of indirectness that occurs when there are no direct (i.e., head-to-head) comparisons between two or more interventions of interest. This criterion is not relevant to our study question related to toxicity of triclosan; it could be relevant to future case studies.

^c According to GRADE, in general, GRADE rates animal evidence down two levels for indirectness. They note that animal studies may, however, provide an important indication of drug toxicity. GRADE states, “Although toxicity data from animals does not reliably predict toxicity in humans, evidence of animal toxicity should engender caution in recommendations.” However, GRADE does not preclude rating non-human evidence as high quality. They state, “Another type of nonhuman study may generate high- quality evidence. Consider laboratory evidence of change in resistance patterns of bacteria to antimicrobial agents (e.g., the emergence of methicillin-resistant staphylococcus aureus-MRSA). These laboratory findings may constitute high-quality evidence for the superiority of antibiotics to which MRSA is sensitive vs. methicillin as the initial treatment of suspected staphylococcus sepsis in settings in which MRSA is highly prevalent.”

3. Outcomes may differ from those of primary interest, for instance, surrogate outcomes that are not themselves important, but measured in the presumption that changes in the surrogate reflect changes in an important outcome. The difference between desired and measured outcomes may relate to time frame. When there is a discrepancy between the time frame of measurement and that of interest, whether to rate down by one or two levels will depend on the magnitude of the discrepancy. Another source of indirectness related to measurement of outcomes is the use of substitute or surrogate endpoints in place of the exposed population's important outcome of interest. In general, the use of a surrogate outcome requires rating down the quality of evidence by one, or even two, levels. Consideration of the biology, mechanism, and natural history of the disease can be helpful in making a decision about indirectness. Surrogates that are closer in the putative causal pathway to the adverse outcomes warrant rating down by only one level for indirectness. GRADE states that rarely, surrogates are sufficiently well established that one should choose not to rate down quality of evidence for indirectness. In general, evidence based on surrogate outcomes should usually trigger rating down, whereas the other types of indirectness will require a more considered judgment.

Rating for Indirectness		Rationale for your judgment
0 no change -1 decrease quality 1 level -2 decrease quality 2 levels		
Human		
Non-Human Mammalian		

Category 3. Inconsistency of Evidence

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

According to Cochrane, “when studies yield widely differing estimates of effect (heterogeneity or variability in results) investigators should look for robust explanations for that heterogeneity. ...When heterogeneity exists and affects the interpretation of results, but authors fail to identify a plausible explanation, the quality of the evidence decreases.”

Based on GRADE [53], **a body of evidence is not rated up in quality if studies yield consistent results, but may be rated down in quality if inconsistent.** Their stated reason is that a consistent bias will lead to consistent, spurious findings.

GRADE suggests rating down the quality of evidence if large inconsistency (heterogeneity) in study results remains after exploration of a priori hypotheses that might explain heterogeneity. Judgment of the extent of heterogeneity is based on similarity of point estimates, extent of overlap of confidence intervals, and statistical criteria. GRADE’s recommendations refer to inconsistencies in effect size, specifically to relative measures (risk ratios and hazard ratios or odds ratios), not absolute measures.

Based on GRADE, reviewers should consider rating down for inconsistency when:

1. Point estimates vary widely across studies;
2. Confidence intervals (CIs) show minimal or no overlap;
3. The statistical test for heterogeneity-which tests the null hypothesis that all studies in a meta-analysis have the same underlying magnitude of effect- shows a low P-value;
4. The I^2 -which quantifies the proportion of the variation in point estimates due to among-study differences-is large. (I.e., the I^2 index quantifies the degree of heterogeneity in a meta-analysis).

GRADE states that inconsistency is important **only when it reduces confidence in results in relation to a particular decision.** Even when inconsistency is large, it may not reduce confidence in results regarding a particular decision. For example, studies that are inconsistent related to the magnitude of a beneficial or harmful effect (but are in the same direction) would not be rated down; in instances when results are inconsistent as to whether there is a benefit or harm of treatment, GRADE would rate down the quality of evidence as a result of variability in results, because the meaning of the inconsistency is so relevant to the decision to treat or not to treat.

Rating for Inconsistency	Rationale for your judgment
o no change	

-1 decrease quality 1 level		
-2 decrease quality 2 levels		
Human		
Non-Human Mammalian		

Category 4. Imprecision of Evidence

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

Cochrane states that when studies have few participants and few events, and thus have wide confidence intervals (CIs), authors can lower their rating of the quality of evidence. These ratings of precision are made as judgments by review authors. The ratings are made by looking across studies, or, if available, on the results of a meta-analysis.

GRADE defines evidence quality differently for systematic reviews and guidelines. For systematic reviews, quality refers to confidence in the estimates of effect. For guidelines, quality refers to the extent to which confidence in the effect estimate is adequate to support a particular decision [54]. For the purpose of step 3 of Navigation Guide, we will use the systematic review definition, because the decision phase does not occur until step 4 when recommendations for prevention are made. Thus, when reviewing the data for imprecision, evaluate your confidence in the estimate of the effect.

According to GRADE, to a large extent, CIs inform the impact of random error on evidence quality. Thus, when considering imprecision, the issue is whether the CI around the estimate of exposure effect is sufficiently narrow. If it is not, GRADE rates down the evidence quality by one level (for instance, from high to moderate). If the CI is very wide, GRADE might rate down by two levels.

Rating for Imprecision		Rationale for your judgment
0 no change -1 decrease quality 1 level -2 decrease quality 2 levels		
Human		
Non-Human Mammalian		

Category 5. Publication Bias

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

GRADE [46] and Cochrane [2] assess publication bias in a similar manner. Whereas “selective outcome reporting” is assessed for each study included in the review as part of the risk of bias assessment, “publication bias” is assessed on the body of evidence. GRADE states that “when an entire study remains unreported and the results relate to the size of the effect- publication bias- one can assess the likelihood of publication bias only by looking at a group of studies.”

Cochrane’s definition of publication bias is “the *publication* or *non-publication* of research findings depending on the nature and direction of the results.” Cochrane and GRADE are primarily concerned with *overestimates* of true effects of treatments or pharmaceuticals, especially related to “small studies effects”, i.e., the tendency for estimates of an intervention to be more beneficial in smaller studies. There is empirical evidence in the clinical sciences that publication and other reporting biases result in over estimating the effects of interventions [2].

In contrast, in environmental health, we are primarily concerned with *underestimating* the true effects of a chemical exposure, since in many cases population wide exposure has already occurred. We are also concerned that studies finding no association are less likely to be published.

Applying this inverted concern to GRADE’s assessment for publication bias, leads to these considerations when rating publication bias:

- Early *negative* studies, particularly if small in size, are suspect. (GRADE is concerned with early *positive* studies).
- Authors of systematic reviews should suspect publication bias when studies are uniformly small, particularly when sponsored by the industry. (Same as GRADE)
- Empirical examination of patterns of results (e.g., funnel plots) may suggest publication bias but should be interpreted with caution. (Same as GRADE)
- More compelling than any of these theoretical exercises is authors’ success in obtaining the results of some unpublished studies and demonstrating that the published and unpublished data show different results. (Same as GRADE)
- Comprehensive searches of the literature including unpublished studies, i.e., the grey literature, and a search for research in other languages are important to addressing publication bias. Note that Cochrane also states “comprehensive searching is not sufficient to prevent some substantial potential biases.”

Rating for Publication Bias		Rationale for your judgment
o no change -1 decrease quality 1 level -2 decrease quality 2 levels		
Human		
Non-Human Mammalian		

Upgrade Categories

GRADE states that the circumstances for upgrading likely occur infrequently and are primarily relevant to observational and other non-randomized studies. Although it is possible to rate up results from randomized controlled trials, GRADE has yet to find a compelling circumstance for

doing so [55].

GRADE specifies 3 categories for increasing the quality of evidence [55]

Category 6. Large Magnitude of Effect

Possible ratings: 0=no change; +1 or +2 upgrade 1 or 2 levels

Modeling studies suggests that confounding (from non-random allocation) alone is unlikely to explain associations with a relative risk (RR) greater than 2 (or less than 0.5), and very unlikely to explain associations with an RR greater than 5 (or less than 0.2). Thus, these are the definitions of “large magnitude of effect” used to upgrade 1 or 2 levels, respectively. Also, GRADE is more likely to rate up if the effect is rapid and out of keeping with prior trajectory; usually supported by indirect evidence. GRADE presents empirical evidence to support these conclusions, and states that “although further research is warranted, both modeling and empirical work suggest the size of bias from confounding is unpredictable in direction but bounded in size. Hence, the GRADE group has previously suggested guidelines for rating quality of evidence up by one category (typically from low to moderate) for associations greater than 2, and up by two categories for associations greater than 5.”

Applying the GRADE definitions of large magnitude of effect i.e., RR greater than 2 or 5 is problematic in environmental health because for dichotomous outcomes RR is a function of the exposure comparator; these definitions also are not applicable to results from continuous variables. At present, we do not have an empirically defined “large magnitude of effect.” Therefore, for the purpose of the triclosan case study, co-authors should assess whether the results indicate a large magnitude of effect using their expert judgment of “large effects” in environmental health and state their definition for discussion by the group.

Rating for Large Magnitude of Effect		Rationale for your judgment
o no change +1 increase quality 1 level +2 increase quality 2 levels		
Human		

Category 7. Dose-response

Possible ratings: 0=no change; +1 or +2 upgrade 1 or 2 levels

Possible considerations include consistent dose response gradients in one or multiple studies, and/or dose response across studies, depending on the overall relevance to the body of evidence.

Rating for Dose-Response		Rationale for your judgment
o no change +1 increase quality 1 level +2 increase quality 2 levels		
Human		

Category 8. Confounding Minimizes Effect

Possible ratings: 0=no change; +1 or +2 upgrade 1 or 2 levels

All plausible residual confounders or biases would reduce a demonstrated effect, or suggest a spurious effect when results show no effect. GRADE provides the following example of grading up evidence when observational studies have failed to demonstrate an association. Observational studies failed to confirm an association between vaccination and autism. This lack of association occurred despite the empirically confirmed bias that parents of autistic children diagnosed after the publicity associated with the article that originally suggested this relationship would be more likely to remember their vaccine experience than parents of children diagnosed before the publicity and presumably, than parents of non-autistic children. The negative findings despite this form of recall bias suggest rating up the quality of evidence.

Rating for Confounding Minimizes Effect		Rationale for your judgment
o no change +1 increase quality 1 level +2 increase quality 2 levels		

Human		
-------	--	--

The results of the reviewers' ratings by population will be compiled and discussed leading to a final decision on overall quality of human evidence. The rationale for the decision will be fully documented.

1. Final decision on overall quality of human evidence:

(Example: Moderate quality is upgraded 1 step to high for Xyz reason(s))

---- High

---- Moderate

---- Low

---- Very

2. Final decision on overall quality of non-human mammalian evidence:

(Example: High quality is downgraded 1 step to moderate for Xyz reason(s))

---- High

---- Moderate

---- Low

---- Very

3. Final decision on overall quality of non-human non-mammalian evidence:

(Example: High quality is downgraded 1 step to moderate for Xyz reason(s))

---- High

---- Moderate

---- Low

---- Very

B. Rate the Strength of Evidence

The evidence quality ratings will be translated into strength of evidence for each population based on a combination of four criteria: (1) Quality of body of evidence; (2) Direction of effect; (3) Confidence in effect; and (4) Other compelling attributes of the data that may influence certainty. The strength of evidence ratings are linked to Tables 1 and 2, below, where their meaning is defined.

C. Combine Strength of Evidence For Human and Non-human Evidence

The final step in the process is to combine the strength of the evidence according to the chart in Figure 1. Combining the strength of evidence for human and non-human data will produce an overall statement of toxicity, i.e., known to be toxic to fetal growth; probably toxic to fetal growth; possibly toxic to fetal growth; known to be not-toxic to fetal growth. These directions apply to combining two streams of evidence only. In situations where there is more than one stream of non-human evidence, for example, in vitro evidence and non-human mammalian evidence, the highest quality data should be used in this step.

Figure 1. Combining Streams of Human and Non-Human Evidence

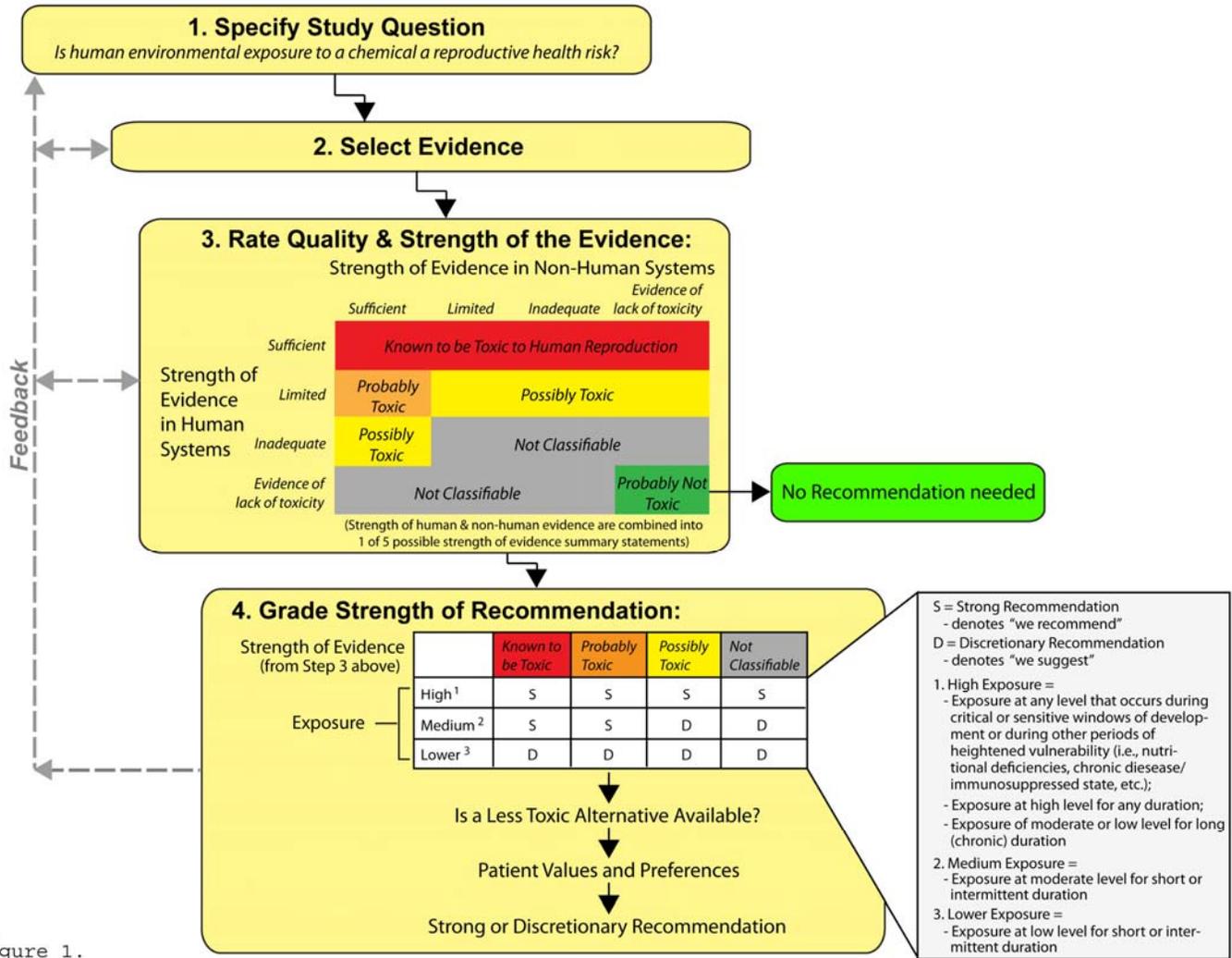


Figure 1.

Table 1. Strength of evidence definitions for human evidence^a

Strength Rating	Definition
Sufficient evidence of toxicity	A positive relationship is observed between exposure and outcome where chance, bias, and confounding can be ruled out with reasonable confidence. 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 ^b .
Limited Evidence of Toxicity	A positive relationship is observed between exposure and outcome where chance, bias, and confounding cannot be ruled out with reasonable confidence. 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 ^b . As more information becomes available, the observed effect could change, and this change may be large enough to alter the conclusion.
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	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 ^b . The conclusion is limited to the age at exposure and/or other conditions and levels of exposure studied.

^aThe Navigation Guide rates the quality and strength of evidence of human and non-human evidence streams separately as “sufficient”, “limited”, “inadequate” or “evidence of lack of toxicity” and then these two ratings are combined to produce one of five possible statements about the overall strength of the evidence of a chemical’s reproductive/developmental toxicity. The methodology is adapted from the criteria used by the International Agency for Research on Cancer (IARC) to categorize the carcinogenicity of substances [45] except as noted.

^bLanguage for the definitions of the rating categories were adapted from descriptions of levels of certainty provided by the U.S. Preventive Services Task Force Levels of Certainty Regarding Net Benefit. <http://www.uspreventiveservicestaskforce.org/uspstf07/methods/benefit.htm>

Table 2. Strength of evidence definitions for non-human evidence^a

Strength Rating	Definition
Sufficient Evidence of Toxicity	A positive relationship is observed between exposure and adverse outcome in multiple studies or a single appropriate study in a single species. ^b 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. ^c
Limited Evidence of Toxicity	The data suggest a positive relationship between exposure and adverse outcome, but there are important limitations in the quality of the body of evidence. 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. ^c As more information becomes available, the observed effect could change, and this change may be large enough to alter the conclusion.
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	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. ^d 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. ^c

^aThe Navigation Guide rates the quality and strength of evidence of human and non-human evidence streams separately as ‘sufficient’, ‘limited’, ‘inadequate’ or ‘evidence of lack of toxicity’ and then these two ratings are combined to produce one of five possible statements about the overall strength of the evidence of a chemical’s reproductive/developmental toxicity. The methodology is adapted from the criteria used by the International Agency for Research on Cancer (IARC) to categorize the carcinogenicity of substances (International Agency for Research on Cancer 2006) except as noted.

^bIARC’s criteria for sufficient evidence of carcinogenicity in animals requires multiple positive results (species, studies, sexes). The Navigation Guide integrates USEPA’s minimum criteria for animal data for a reproductive or developmental hazard, i.e., data demonstrating an adverse reproductive effect in a single appropriate, well-executed study in a single test species (U.S. Environmental Protection Agency 1996). The Navigation Guide also incorporates USEPA’s “sufficient evidence category” which includes data that “collectively provide enough information to judge whether or not a reproductive hazard exists within the context of effect as well as dose, duration, timing, and route of exposure. This category may include both human and experimental animal evidence” (U.S. Environmental Protection Agency 1996). The USEPA statement for developmental hazards is slightly different but includes the same relevant information regarding dose, duration, timing, etc. (U.S. Environmental Protection Agency 1991).

^cLanguage for the definitions of the rating categories were adapted from descriptions of levels of certainty provided by the U.S. Preventive Services Task Force Levels of Certainty Regarding Net Benefit (Sawaya et al. 2007).

^dBased on minimum data requirements according to USEPA Guidelines for Reproductive Toxicity (U.S. Environmental Protection Agency 1996).

ABOUT THE ARTICLE

Declarations of Interest: No competing interests reported

Sources of Support:

Internal: UCSF - none

External: UCSF through USEPA and CalEPA Contract

ACKNOWLEDGEMENTS

Kristina Thayer and Andrew Rooney in the Office of Health Assessment and Translation, National Toxicology Program for assistance with developing risk of bias evaluation criteria.

Kate Guyton in the National Center for Environmental Assessment Washington Division, National Center for Environmental Assessment, Office of Research and Development, US Environmental Protection Agency for assistance with identifying relevant toxicological databases.

Malcolm Macleod in the Experimental Neuroscience Centre for Clinical Brain Sciences, University of Edinburgh for assistance in developing animal-specific systematic review methodology.

Juleen Lam, Johns Hopkins University for helpful comments.

Ryan Babadi (RB) for assistance with literature search and reference screening.

References:

1. Guyatt, G.H., et al., *GRADE: an emerging consensus on rating quality of evidence and strength of recommendations*. BMJ, 2008. **336**(7650): p. 924-6.
2. Higgins, J.P.T. and S. Green, *Cochrane Handbook for Systematic Reviews of Interventions. Version 5.1.0 [Updated March 2011]*. 2011, The Cochrane Collaboration. Available from www.cochrane-handbook.org.
3. Woodruff, T.J., P. Sutton, and The Navigation Guide Work Group, *An Evidence-Based Medicine Methodology To Bridge The Gap Between Clinical And Environmental Health Sciences*. Health Affairs, 2011. **30**(5): p. 931-937.
4. Johnson, P., et al., *The Navigation Guide - Evidence-Based Medicine Meets Environmental Health: Systematic Review of Human Evidence for PFOA Effects on Fetal Growth*. Environmental Health Perspectives, In review.
5. Koustas, E., et al., *The Navigation Guide - Evidence-Based Medicine Meets Environmental Health: Systematic Review of Non-Human Evidence for PFOA Effects on Fetal Growth*. Environmental Health Perspectives, In press.
6. Cooney, C.M., *Triclosan Comes under Scrutiny*. Environmental Health Perspectives, 2010. **118**(6): p. A242-A242.
7. Thompson, A., et al., *The fate and removal of triclosan during wastewater treatment*. Water environment research : a research publication of the Water Environment Federation, 2005. **77**(1): p. 63-7.
8. Glaser, A., *The ubiquitous triclosan. A common antibacterial agent exposed*. Pesticides and You, 2004. **24**: p. 12-17.
9. Perencevich, E.N., M.T. Wong, and A.D. Harris, *National and regional assessment of the antibacterial soap market: A step toward determining the impact of prevalent antibacterial soaps*. American Journal of Infection Control, 2001. **29**(5): p. 281-283.
10. U.S. Environmental Protection Agency, *Reregistration eligibility decision for triclosan*. 2008: p. Case No. 2340.
11. Kolpin, D.W., et al., *Pharmaceuticals, hormones, and other organic wastewater contaminants in US streams, 1999-2000: A national reconnaissance*. Environmental Science & Technology, 2002. **36**(6): p. 1202-1211.
12. Adolfsson-Erici, M., et al., *Triclosan, a commonly used bactericide found in human milk and in the aquatic environment in Sweden*. Chemosphere, 2002. **46**(9): p. 1485-1489.
13. Calafat, A.M., et al., *Urinary concentrations of Triclosan in the US population: 2003-2004*. Environ Health Perspect, 2008. **116**(3): p. 303-307.
14. Wilding, B.C., K. Curtis, and K. Welker-Hood, *Hazardous Chemicals in Health Care: A Snapshot of Chemicals in Doctors and Nurses*. 2009, Physicians for Social Responsibility: Washington, D.C.
15. Wolff, M.S., et al., *Pilot study of urinary biomarkers of phytoestrogens, phthalates, and phenols in girls*. Environ Health Perspect, 2007. **115**(1): p. 116-21.
16. Robertshaw, H. and B. Leppard, *Contact dermatitis to triclosan in toothpaste*. Contact Dermatitis, 2007. **57**(6): p. 383-U4.
17. Schena, D., A. Papagrigoraki, and G. Girolomoni, *Sensitizing potential of triclosan and triclosan-based skin care products in patients with chronic eczema*. Dermatologic Therapy, 2008. **21**: p. S35-S38.
18. Aiello, A.E., E.L. Larson, and S.B. Levy, *Consumer antibacterial soaps: effective or just risky?* Clin Infect Dis, 2007. **45 Suppl 2**: p. S137-47.

19. Foran, C.M., E.R. Bennett, and W.H. Benson, *Developmental evaluation of a potential non-steroidal estrogen: triclosan*. Marine Environmental Research, 2000. **50**(1-5): p. 153-156.
20. Matsumura, N., et al., *Effects of nonylphenol and triclosan on production of plasma vitellogenin and testosterone in male South African clawed frogs (Xenopus laevis)*. Biological and Pharmaceutical Bulletin, 2005. **28**(9): p. 1748-1751.
21. Veldhoen, N., et al., *The bactericidal agent triclosan modulates thyroid hormone-associated gene expression and disrupts postembryonic anuran development (vol 80, pg 217, 2006)*. Aquatic Toxicology, 2007. **83**(1): p. 84-84.
22. Kumar, V., et al., *Alteration of testicular steroidogenesis and histopathology of reproductive system in male rats treated with triclosan*. Reproductive Toxicology, 2009. **27**(2): p. 177-185.
23. Wolff, M.S., et al., *Investigation of relationships between urinary biomarkers of phytoestrogens, phthalates, and phenols and pubertal stages in girls*. Environmental Health Perspectives, 2010. **118**(7): p. 1039-46.
24. U.S. Environmental Protection Agency. *Guidelines for Developmental Toxicity Risk Assessment*. 1991 [cited 2014 27 February]; Available from: <http://cfpub.epa.gov/ncea/cfm/recorddisplay.cfm?deid=23162#Download>.
25. U.S. Environmental Protection Agency. *Guidelines for reproductive toxicity risk assessment*. 1996 [cited 2014 27 February]; Available from: <http://www.epa.gov/raf/publications/pdfs/REPRO51.PDF>.
26. Dann, A.B. and A. Hontela, *Triclosan: environmental exposure, toxicity and mechanisms of action*. Journal of applied toxicology : JAT, 2011. **31**(4): p. 285-311.
27. James, M.O., et al., *Triclosan is a potent inhibitor of estradiol and estrone sulfonation in sheep placenta*. Environment International, 2010. **36**(8): p. 942-949.
28. Rodricks, J.V., et al., *Triclosan: a critical review of the experimental data and development of margins of safety for consumer products*. Critical Reviews in Toxicology, 2010. **40**(5): p. 422-84.
29. Fang, J.L., et al., *Occurrence, efficacy, metabolism, and toxicity of triclosan*. Journal of environmental science and health. Part C, Environmental carcinogenesis & ecotoxicology reviews, 2010. **28**(3): p. 147-71.
30. The Alliance for the Prudent Use of Antibiotics, *Triclosan*. 2011: Boston, MA.
31. Ciba Specialty Chemicals Corporation, *Triclosan Safety Studies Submitted to OTC Docket 81N-033P*. 2004, U.S. Food and Drug Administration: Washington, D.C.
32. U.S. Environmental Protection Agency. *Toxic Substances Control Act (TSCA) Section 8(e) Notices*. 2013 [cited 2013 December 3].
33. von Elm, E., et al., *The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies*. Journal of Clinical Epidemiology, 2008. **61**(4): p. 344-9.
34. Higgins, J. and J. Deeks, *Chapter 7: Selecting studies and collecting data*, in *Cochrane Handbook for Systematic Reviews of Interventions. Version 5.1.0*, J. Higgins and S. Green, Editors. 2011, The Cochrane Collaboration.
35. Guyatt, G.H., et al., *GRADE guidelines: 4. Rating the quality of evidence--study limitations (risk of bias)*. J Clin Epidemiol, 2011. **64**(4): p. 407-15.
36. Hooijmans, C.R., M. Leenaars, and M. Ritskes-Hoitinga, *A gold standard publication checklist to improve the quality of animal studies, to fully integrate the Three Rs, and to make systematic reviews more feasible*. 2010.
37. Kilkeny, C., et al., *Animal research: reporting in vivo experiments: the ARRIVE guidelines*. British Journal of Pharmacology, 2010. **160**(7): p. 1577-9.

38. Viswanathan, M., et al., *Assessing the Risk of Bias of Individual Studies in Systematic Reviews of Health Care Interventions.*, in *Agency for Healthcare Research and Quality Methods Guide for Comparative Effectiveness Reviews*. 2012.
39. Roseman, M., et al., *Reporting of conflicts of interest in meta-analyses of trials of pharmacological treatments*. JAMA, 2011. **305**(10): p. 1008-17.
40. Lundh, A., et al., *Industry sponsorship and research outcome*. Cochrane database of systematic reviews (Online), 2012. **12**.
41. Krauth, D., T.J. Woodruff, and L. Bero, *Instruments for assessing risk of bias and other methodological criteria of published animal studies: a systematic review*. Environmental Health Perspectives, 2013. **121**(9): p. 985-92.
42. Light, R.J. and D.B. Pillemer, *Summing up: the science of reviewing research*. 1984: Harvard University Press.
43. Duval, S. and R. Tweedie, *Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis*. Biometrics, 2000. **56**(2): p. 455-63.
44. Kimmel, C.A., et al., *Reliability of Experimental Studies for Predicting Hazards to Human Development*, in *NCTR Technical Report for Experiment No. 6015*. 1984: Jefferson, AR.
45. International Agency for Research on Cancer, *IARC Monographs on the Evaluation of Carcinogenic Risks to Humans: Preamble (amended January 2006)*. 2006, World Health Organization: Lyon, France.
46. Guyatt, G.H., et al., *GRADE guidelines: 5. Rating the quality of evidence-publication bias*. J Clin Epidemiol, 2011. **64**(12): p. 1277-1282.
47. Guyatt, G.H., et al., *GRADE guidelines: 8. Rating the quality of evidence-indirectness*. J Clin Epidemiol, 2011. **64**(12): p. 1303-1310.
48. Hemminki, K. and P. Vineis, *Extrapolation of the evidence on teratogenicity of chemicals between humans and experimental animals: chemicals other than drugs*. Teratog Carcinog Mutagen, 1985. **5**(4): p. 251-318.
49. Nisbet, I.C.T. and N.J. Karch, *Chemical Hazards to Human Reproduction*. 1983, Park Ridge, NJ: Noyes Data Corp.
50. Nemec, M.D., et al., *Significance, Reliability, and Interpretation of Developmental and Reproductive Toxicity Study Findings*, in *Developmental Reproductive Toxicology: A Practical Approach*. 2006, Informa Healthcare.
51. Newman, L.M., E.M. Johnson, and R.E. Staples, *Assessment of the Effectiveness of Animal Developmental Toxicity Testing for Human Safety*. Reproductive Toxicology, 1993. **7**(4): p. 359-390.
52. National Research Council (U.S.) Committee on Developmental Toxicology and National Research Council (U.S.) Commission on Life Sciences, *Scientific frontiers in developmental toxicology and risk assessment*. 2000, Washington, DC: National Academy Press. xviii, 327 p.
53. Guyatt, G.H., et al., *GRADE guidelines: 7. Rating the quality of evidence-inconsistency*. J Clin Epidemiol, 2011. **64**(12): p. 1294-1302.
54. Guyatt, G., et al., *GRADE guidelines: 6. Rating the quality of evidence--imprecision*. J Clin Epidemiol, 2011. **64**(12): p. 1283-93.
55. Guyatt, G.H., et al., *GRADE guidelines: 9. Rating up the quality of evidence*. J Clin Epidemiol, 2011. **64**(12): p. 1311-1316.