Environment International 92–93 (2016) 716–728

Contents lists available at ScienceDirect

Environment International

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

Application of the Navigation Guide systematic review methodology to the evidence for developmental and reproductive toxicity of triclosan

Paula I. Johnson [a,*], Erica Koustas [b], Hanna M. Vesterinen [a], Patrice Sutton [a], Dylan S. Atchley [a], Allegra N. Kim [c], Marlissa Campbell [c], James M. Donald [c], Saunak Sen [d], Lisa Bero [e,1], Lauren Zeise [c], Tracey J. Woodruff [a]

[a] University of California San Francisco, Program on Reproductive Health and the Environment, Oakland, CA, USA
[b] ORISE Post-doctoral Fellowship, U.S. Environmental Protection Agency, Office of Policy, National Center for Environmental Economics, Washington, D.C., USA
[c] Office of Environmental Health Hazard Assessment, California Environmental Protection Agency, CA, USA
[d] University of California San Francisco, Department of Epidemiology and Biostatistics, San Francisco, CA, USA
[e] University of California San Francisco, Department of Clinical Pharmacy, San Francisco, CA, USA

article info

Article history:
Received 29 July 2015
Received in revised form 8 March 2016
Accepted 9 March 2016
Available online 5 May 2016

Keywords:
Soap
Personal care
Antibacterial
Thyroid
Policy
Risk assessment

abstract

Background: There are reports of developmental and reproductive health effects associated with the widely used biocide triclosan.
Objective: Apply the Navigation Guide systematic review methodology to answer the question: Does exposure to triclosan have adverse effects on human development or reproduction?
Methods: We applied the first 3 steps of the Navigation Guide methodology: 1) Specify a study question, 2) Select the evidence, and 3) Rate quality and strength of the evidence. We developed a protocol, conducted a comprehensive search of the literature, and identified relevant studies using pre-specified criteria. We assessed the number and type of all relevant studies. We evaluated each included study for risk of bias and rated the quality and strength of the evidence for the selected outcomes. We conducted a meta-analysis on a subset of suitable data.
Results: We found 4282 potentially relevant records, and 81 records met our inclusion criteria. Of the more than 100 endpoints identified by our search, we focused our evaluation on hormone concentration outcomes, which had the largest human and non-human mammalian data set. Three human studies and 8 studies conducted in rats reported thyroxine levels as outcomes. The rat data were amenable to meta-analysis. Because only one of the human thyroxine studies quantified exposure, we did not conduct a meta-analysis of the human data. Through meta-analysis of the data for rats, we estimated for prenatal exposure a 0.09% (95% CI: −0.20, 0.02) reduction in thyroxine concentration per mg triclosan/kg-bw in fetal and young rats compared to control. For postnatal exposure we estimated a 0.31% (95% CI: −0.38, −0.23) reduction in thyroxine per mg triclosan/kg-bw, also compared to control. Overall, we found low to moderate risk of bias across the human studies and moderate to high risk of bias across the non-human studies, and assigned a "moderate/low" quality rating to the body of evidence for human thyroid hormone alterations and a "moderate" quality rating to the body of evidence for non-human thyroid hormone alterations.
Conclusion: Based on this application of the Navigation Guide systematic review methodology, we concluded that there was "sufficient" non-human evidence and "inadequate" human evidence of an association between triclosan exposure and thyroxine concentrations, and consequently, triclosan is "possibly toxic" to reproductive and developmental health. Thyroid hormone disruption is an upstream indicator of developmental toxicity. Additional endpoints may be identified as being of equal or greater concern as other data are developed or evaluated.
© 2016 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

Integration of the available scientific evidence to reach a strength-of-evidence conclusion about chemical toxicity is fundamental to developing hazard assessments for regulatory action, clinical guidelines, and safer alternatives to toxic chemicals. To this end, the Navigation Guide systematic review methodology was developed by a working group in 2009 to provide a transparent, reproducible framework to evaluate the quality and strength of evidence about the relationship between

* Corresponding author at: California Safe Cosmetics Program, Occupational Health Branch, California Department of Public Health, 850 Marina Bay Parkway, Richmond, CA 94804, USA.
E-mail address: Paula.Johnson@cdph.ca.gov (P.I. Johnson).
[1] Current affiliation: University of Sydney, Charles Perkins Centre, Medicines Use and Health Outcomes, New South Wales, Australia.

http://dx.doi.org/10.1016/j.envint.2016.03.009
0160-4120/©2016 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

environmental exposures and reproductive and developmental health (Woodruff and Sutton, 2011). Beginning in 2011, the National Toxicology Program (NTP) undertook a complementary effort to develop a framework for systematic reviews in environmental health (Rooney et al., 2014). In 2014 two reports by the National Academy of Sciences found that such methods of evidence integration reflect the approach that the U.S. Environmental Protection Agency (U.S. EPA) should adopt to determine whether environmental chemicals are harmful to human health (National Research Council, 2014a; National Research Council, 2014b). A report from the UK similarly recommended uptake of systematic methods of evidence integration by relevant European Union agencies, to increase transparency and decrease bias in regulatory rulemaking (Whaley, 2013). Since 2012, the NTP has been actively building the tools, expertise, and other infrastructure that will facilitate increased utilization of systematic review methodologies (Rooney et al., 2014; National Toxicology Program, 2015). The U.S. EPA has proposed steps to begin to incorporate principles of systematic review into its Integrated Risk Information System (IRIS) process (U.S. Environmental Protection Agency, 2014; The National Academies, 2012). A 2014 case study applying the Navigation Guide methodology to evaluate the human and non-human evidence of perfluorooctanoic acid (PFOA) on fetal growth demonstrated how the efforts under development by the NTP and consideration by the U.S. EPA are achievable (Koustas et al., 2014; Johnson et al., 2014; Lam et al., 2014; Woodruff and Sutton, 2014). The present case study was intended as part of ongoing proof-of-concept and an opportunity for the California Office of Environmental Health Hazard Assessment (OEHHA) to explore the Navigation Guide methodology on a broader range of outcomes. This systematic review evaluates the evidence for the effects of exposure to the widely-used biocide triclosan on endpoints of developmental and/or male or female reproductive toxicity.

Triclosan, or 2,4,4′-trichloro-2′-hydroxydiphenylether, is a synthetic, broad-spectrum anti-microbial agent developed over 50 years ago and introduced as a surgical scrub (Cooney, 2010). In 2013, there were 2000 antimicrobial consumer products, including soaps and other personal care products, dental products, clothing, paints, plastics and children's toys (Halden, 2014). A 2000 survey found that 76% of U.S. liquid soaps and 29% of bar soaps contained triclosan or an alternative antimicrobial triclocarban (Perencevich et al., 2001).

The FDA has the authority to regulate triclosan when used in personal care products and medical devices. As the FDA has not finalized its 1974 draft topical antimicrobial drug products Over-the-Counter Drug Monograph, triclosan is currently unregulated in personal care products (U.S. Food and Drug Administration, 2013). With intent to finalize the Monograph, the FDA proposed a new rule in 2013 that would require manufacturers to provide safety data and data that demonstrates the clinical benefit of using antibacterial soaps over plain soap and water (U.S. Food and Drug Administration, 2013). Pesticidal uses of triclosan come under the regulatory authority of U.S. EPA (U.S. EPA, 2015).

Exposure to triclosan is widespread in the U.S. population (Adolfsson-Erici et al., 2002; Calafat et al., 2008; Wilding et al., 2009; Wolff et al., 2007). There is also growing concern over triclosan's possible effects on public health, including direct health effects, e.g., skin irritation (Robertshaw and Leppard, 2007; Schena et al., 2008), endocrine disruption and associated reproductive effects as observed in animal experiments (Foran et al., 2000; Matsumura et al., 2005; Veldhoen et al., 2007; Stoker et al., 2010) and human studies (Wolff et al., 2010; Chen et al., 2013; Koeppe et al., 2013), and indirect effects, i.e., antibiotic resistance (Aiello et al., 2007).

This is the first systematic review of the human and animal evidence linking exposure to triclosan to adverse reproductive or developmental health endpoints. Past reviews of triclosan were expert-based narrative reviews, not systematic reviews, and/or primarily focused on assessing the risk of using personal care products containing triclosan, using exposure estimates based on certain concentrations of triclosan in the products (Rodricks et al., 2010; SCCS. Scientific Committee on Consumer Safety, 2011; Witorsch, 2014). In contrast, we did not estimate exposure or assess risk in the present review; we evaluated the evidence of the chemical's toxicity (i.e., hazard).

Based on the presence of triclosan in wide-ranging consumer products, the environment, and humans, and potential for human health effects, we applied the Navigation Guide systematic review methodology to evaluate the strength of the evidence relating triclosan exposure to developmental or reproductive health effects.

2. Methods

The Navigation Guide is based on best practices in evaluation of clinical evidence and adapts the evidence-based medicine methodology developed by Cochrane and the Grading of Recommendations Assessment Development and Evaluation (GRADE), tested and evaluated since the 1990s (Guyatt et al., 2011; Balshem et al., 2011). We assembled a team of reviewers with expertise in toxicology, epidemiology, environmental health, biology, statistics and systematic review, and developed a pre-specified protocol for conducting the systematic review (Johnson et al., 2013). Each of the protocol steps are described below and the protocol is available at http://prhe.ucsf.edu/prhe/pdfs/Triclosan%20Protocol.pdf

2.1. Specify the study question

Our objective was to answer the question: "Does exposure to triclosan have adverse effects on human development or reproduction?" We developed a "Participants," "Exposure," "Comparator" and "Outcomes" (PECO) statement, which is used as an aid to developing a strategy for answering the study question (Higgins and Green, 2011). Our PECO statement was:

2.1.1. Participants

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

2.1.2. Exposure

For developmental effects, we included one or more exposures to triclosan, by any route, which occurred during the following periods: 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).

For reproductive effects, we include one or more exposures to triclosan at any time preceding assessment of reproductive outcome.

2.1.3. Comparators

Comparable populations or subjects (human, non-human, tissues, organs, cell lines or components) exposed to vehicle-only treatment or lower levels of triclosan than the more highly exposed subjects.

2.1.4. Outcomes

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: 1) pre-conception exposure of parental or previous generations causing genetic mutation or

epigenetic changes, which in turn affect development of unexposed offspring, and 2) 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).

2.2. Select the evidence

2.2.1. Search methods

Our search was not limited by language or publication date. We searched several online databases (PubMed, ISI Web of Science, Biosis Previews, Embase and Toxline) on June 5, 2013 using the search terms in Table S1 (Supplemental material). We used the following databases to compile synonyms for triclosan: Medical Subject Headings (MeSH), PubChem, Sigma-Aldrich, 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&region = US; http://www.chemspider.com/Chemical-Structure.5363.html). We identified additional synonyms from several reviews and original research articles on triclosan (Rodricks et al., 2010; Dann and Hontela, 2011; James et al., 2010; Fang et al., 2010; Anon., 2011; Ciba Specialty Chemicals Corporation, 2004). We combined "triclosan" and its synonyms in a Boolean search using the "OR" statement. We searched 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 in MeSH headings (using the [mh] function in PubMed). We searched additional toxicological websites (June 17–25, 2013); the specific databases searched are provided in the Supplemental material (Table S2). We also hand-searched the reference lists of all included studies and used Web of Science to search for articles that cited the included studies.

2.2.2. Study selection criteria

We selected studies where triclosan was administered, measured or estimated and associations with developmental or reproductive outcomes were evaluated using a customized, structured form in DistillerSR (Evidence Partners; available at: http://www.systematic-review.net). Two of 5 possible reviewers (DA, RB, MC, AK, HV) independently conducted a title and abstract review of each reference from the literature search results to determine eligibility based on the criteria for inclusion. References not excluded based on the title and abstract were screened through full-text review by the title/abstract reviewers and a sixth reviewer (EK). An additional reviewer (PJ) screened 5% of the titles/abstracts and full-texts for quality assurance. In the case of differences between reviewers, the initial reviewers discussed the discrepancy and consulted another reviewer (PJ) if necessary to decide whether to exclude the reference.

We excluded studies if: 1. the report did not contain original data; 2. there was no triclosan exposure prior to the assessment of effect; 3. no developmental or reproductive outcomes were reported; or 4. there was no comparator (control group or exposure range comparison).

2.2.3. Data collection and management

We assessed the number of studies resulting from our search and the number of health outcomes. Two authors (DA, AK for human studies; EK, HV for non-human studies) independently extracted data and details of study design and outcome measures (see Supplemental material, Data extraction fields) from all included human and non-human mammalian articles into a Microsoft Access (2010) database.

We contacted an author of each included non-human mammalian study to request raw data from all relevant figures where data were only presented in graphical form and to obtain additional data which were pertinent to our study question but were missing or ambiguous. We contacted authors of human and non-human mammalian studies when the information provided in the study was unclear with respect to rating risk of bias domains.

2.3. Statistical analyses

We assessed study characteristics of included studies to determine suitability for use in a meta-analysis. We reported outcome measures and their standard errors (reported in the study or calculated from reported standard deviations and sample sizes) as a percentage normalized to the respective control groups, to have the same metric across studies. When meta-analysis was possible, we used a two-step modeling approach as described previously (Koustas et al., 2014). In the first step we analyzed each dataset separately using a linear mixed effects model and obtained a slope estimate of the dose–response effect and associated standard error. In the second step we combined the slope and standard error estimate from each dataset using a random effects model, producing an estimate of the overall mean change in thyroxine concentration per 1-unit increase in triclosan dose (mg/kg-bw-day), accounting for within- and between-study variability. We used Stata SE (Version 10: StataCorp LP, College Station, Texas, USA) to perform both steps in the analysis; we used the metareg function for step one and the metan function for step two. We evaluated statistical heterogeneity across study estimates in the meta-analysis using Cochran's Q statistic with $p \leq 0.05$ as our cut-off for statistical significance and I^2, (Higgins and Green, 2011) as previously described (Koustas et al., 2014; Johnson et al., 2014).

2.4. Rate the quality and strength of the evidence

We rated the quality and strength of the evidence according to the following steps: 1) We assessed the "risk of bias" (defined as study characteristics capable of introducing systematic error in the magnitude or direction of the results; Higgins and Green, 2011) for each included study; 2) we rated the quality of the evidence across studies; and 3) we rated the strength, or certainty, of the evidence across studies.

2.4.1. Assessing the risk of bias for each included study

We assessed risk of bias for the included human and non-human studies using revised instruments (Supplemental material, Instructions for making risk of bias determinations) that were previously developed for human and animal evidence (Koustas et al., 2014; Johnson et al., 2014), based on existing guidance from the Cochrane Collaboration's "Risk of Bias" tool and the Agency for Healthcare Research and Quality's (AHRQ) criteria that address selection bias and confounding, performance bias, attrition bias, detection bias, and reporting bias (Higgins and Green, 2011; Viswanathan et al., 2012). Because our body of human evidence included a study that was a subset of a randomized clinical trial (Cullinan et al., 2012), rather than evaluate that study for "baseline differences" as for the other observational studies, we evaluated that study for two different risk of bias domains which were part of our "Non-human experimental studies" risk of bias instrument (Supplemental material). We also included financial conflicts of interest as a potential source of bias based on data from studies on pharmacological treatments showing evidence of bias associated with funding source (Lundh et al., 2012; Krauth et al., 2013).

We assigned each risk of bias domain as "low risk of bias," "probably low risk of bias," "probably high risk of bias," "high risk of bias," or "not applicable" (risk of bias area not applicable to study) according to specific criteria as described in our risk of bias instruments (Supplemental material, Instructions for making risk of bias determinations). Review authors (DA, PJ, AK for human studies; EK, HV for non-human studies) independently recorded risk of bias determinations for each included study and discussed any discrepancies until consensus was reached.

We determined the important potential confounders or effect modifiers by which to determine risk of bias for the human studies by searching the included studies, the cited references and other known

relevant articles, such as studies using large datasets from the National Health and Nutrition Examination Survey (NHANES), for evidence of associations between potential confounders and triclosan exposure and the outcomes under study. Because age and body mass index (BMI) are associated with triclosan exposure and with thyroid hormone concentrations (Calafat et al., 2008; Chen et al., 2013; Lankester et al., 2013; Knudsen et al., 2005; Hollowell et al., n.d.), we assigned studies "low risk" of bias under the confounding domain if they accounted for potential confounding by age and BMI. Because triclosan is relatively non-persistent (half-life b 24 h), there is uncertainty in relying on a single urine measurement of triclosan to assess longer term exposure, and this reliance assumes that exposure is consistent over time. However, there is some evidence that a single urine triclosan measurement is a reasonably reliable estimate of exposure over time (Spearman correlation coefficient for measurements 3 months apart = 0.50) (Teitelbaum et al., 2008; Bertelsen et al., 2014). We considered this uncertainty and assumption in relation to each outcome in evaluating risk of bias under the exposure assessment domain for observational studies.

2.4.2. Rating the quality of evidence across studies

We separately rated the overall quality of the bodies of human and non-human evidence as "high," "moderate" or "low." The Navigation Guide follows the approach established by the GRADE method; i.e., we determined the final rating by first assigning a pre-specified quality rating to the bodies of evidence and then considered adjustments ("downgrades" or "upgrades") to the quality rating based on the characteristics of the included studies (Balshem et al., 2011). The quality ratings are not additive scores but serve as qualitative guidance in assessing the overall quality of evidence. GRADE guidelines are used to evaluate clinical interventions and assign an initial rating of "high" to bodies of evidence consisting of experimental human studies and an initial rating of "low" quality to observational studies (Balshem et al., 2011). We recognize, however, that not all observational studies are of low quality (Viswanathan et al., 2012; U.S. Environmental Protection Agency, 1996; International Agency for Research on Cancer, 2006) and that decisions in the context of environmental health may rely heavily on human observational data (Woodruff and Sutton, 2011). We therefore assigned an initial rating of "moderate" quality to the body of human evidence, which primarily consisted of observational studies, in consideration of the value and limitations of observational data in assessing associations between exposure and health outcomes in environmental health (Woodruff and Sutton, 2014). We assigned an initial rating of "high" quality to the experimental animal data, comparable to human randomized controlled trials and consistent with GRADE guidelines for experimental human studies, i.e. randomized controlled trials (Guyatt et al., 2011).

We assessed the overall bodies of human and non-human evidence for downgrading and upgrading the pre-specified quality ratings based on specific factors (Supplemental material, Table S2). These factors, based on GRADE guidelines (Balshem et al., 2011), were risk of bias, indirectness, inconsistency, imprecision, publication bias, large magnitude of effect, dose response and whether confounding minimizes the effect. Possible ratings were 0 (no change from initial quality rating), − 1 (1 level downgrade) or – 2 (2 level downgrade); + 1 (1 level upgrade) or + 2 (2 level upgrade). We each independently evaluated the quality of the evidence and then compared our ratings and rationale for each quality factor. We discussed our ratings as a group and recorded our rationale. Consistent with GRADE, we did not automatically add together the ratings for each downgrade and upgrade factor to create a score, e.g., a (− 1) downgrade for each of 2 factors does not necessarily translate into a (− 2) downgrade overall. Also consistent with GRADE, upgrades and downgrades were made only when there was compelling evidence to do so. We used judgment to decide the weight of each downgrade or upgrade in the final overall quality rating.

2.4.3. Rating the strength of the evidence across studies

We rated the overall strength of each body of evidence based on 4 considerations: (1) Quality of body of evidence (i.e., the rating from the previous step); (2) Direction of effect; (3) Confidence in effect (likelihood that a new study would change our conclusion); and (4) Other compelling attributes of the data that may influence certainty. We used these considerations to assign the overall strength rating, according to the definitions specified in the Navigation Guide for "sufficient evidence of toxicity," "limited evidence of toxicity," "inadequate evidence of toxicity," or "evidence of lack of toxicity" (Supplemental material, Tables S3 and S4), which are based on categories used by the International Agency for Research on Cancer (IARC). The Navigation Guide uses criteria and considerations used by IARC, the U.S. Preventive Services Task Force, and U.S. EPA for the type of evidence considered for each of its strength of evidence categories (U.S. Environmental Protection Agency, 1991; U.S. Environmental Protection Agency, 1996; International Agency for Research on Cancer, 2006; Sawaya et al., 2007). We each evaluated the strength of the evidence independently. We then convened to compare evaluations, resolve discrepancies by discussion, and record the collective rationale for decisions. We integrated the human and non-human evidence streams as specified in the Navigation Guide methodology, a process adapted from IARC's method which results in a single concise statement of health hazard (Woodruff and Sutton, 2011; International Agency for Research on Cancer, 2006). The result is 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 (Fig. S1).

3. Results

3.1. Included studies

Our search retrieved a total of 9485 records. After eliminating duplicates, 4282 unique records remained. By applying the specific predefined exclusion criteria, we excluded the majority of the irrelevant references (4034 abstracts excluded out of 4282 total) in under 18 h average for each reviewer. The remaining irrelevant references were excluded in under 6 h average during full-text screening. After application of the exclusion criteria, 81 articles remained: 24 invertebrate studies, 16 in vitro studies, 14 fish studies, 8 amphibian studies, 13 rodent studies, and 6 human studies (Fig. 1 and list in Supplemental material). In addition to the wide range of, and sparse data for, non-mammalian outcome measures, we did not have a developed method to assess the strength of the evidence for reproductive and developmental toxicity for these types of studies. Therefore, we limited our analysis to the mammalian (human and rodent) studies (Fig. 1; Tables S27 and S28). We also found numerous outcome measures (over 100 unique outcomes, including various endpoints at the cellular level) within the 13 rodent studies, with relatively sparse data for each outcome. However, most of the 6 human and 13 rodent studies focused on hormone modulation as an outcome measure, and thus we focused our analysis on that outcome. Thyroid hormone disruption is an upstream indicator of developmental toxicity (Miller et al., 2009; Woodruff et al., 2008; Crofton, 2008; Wise et al., 2012).

Three of 6 human studies reported associations between triclosan and thyroid hormones. The human studies spanned the years 2010 to 2013, had different study designs and ranged from 12 to 1831 study subjects from differing populations (Table 1). Eight of 13 rodent studies provided data on hormone levels following prenatal, prenatal plus postnatal, or postnatal-only exposure to triclosan (Table 2). Our search identified a rat study by Crofton et al. (2007), but because those data were included in the study by Paul et al. (2010a), we did not include the data reported in the Crofton et al. publication. We considered only the 3 human and 8 rat hormone studies in rating the quality and strength of the evidence.

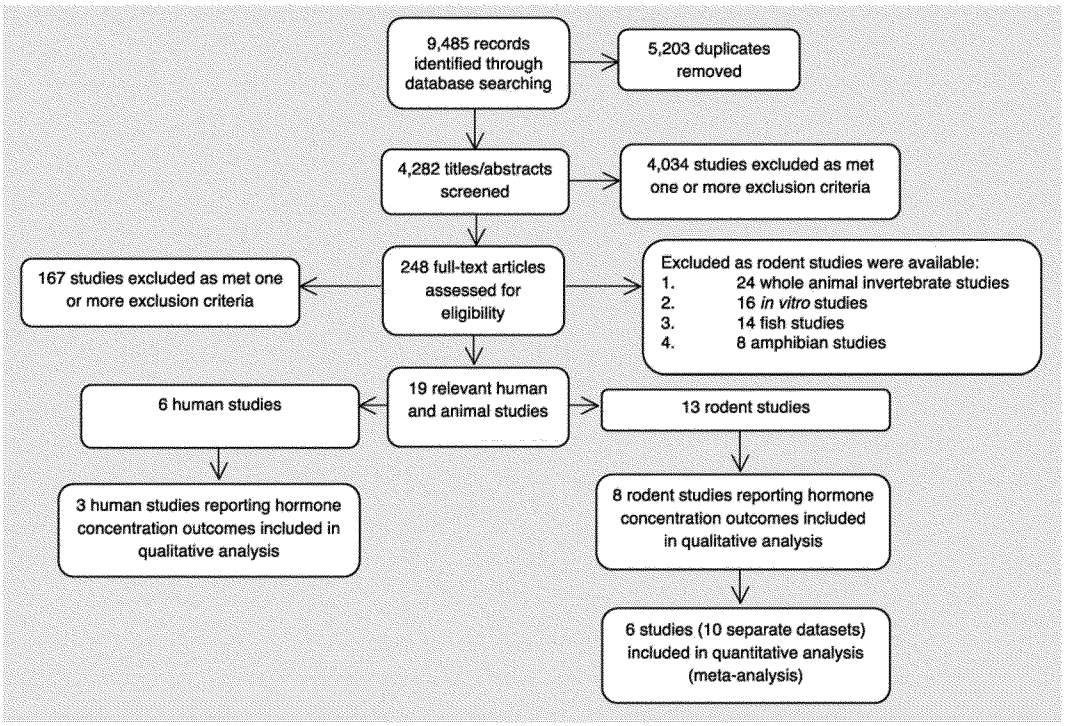

Fig. 1. Flowchart of the study-selection process.

3.2. Risk of bias assessment for individual studies

We assigned "low" or "probably low" risk of bias designations to the majority of the domains for the 3 included human hormone studies (Fig. 2). We assigned "probably high" risk of bias designations to the majority of the 8 included rat studies, particularly for the "allocation concealment" and "blinding" domains (Fig. 3). Additional detail on individual study characteristics and risk of bias designations is in the Supplemental material.

3.3. Data analysis

3.3.1. Human data

Because there were few studies and dissimilar types of data, we could not conduct a meta-analysis of the human data. Although 3 human studies measured thyroid hormones, only one quantified exposure (urinary triclosan from NHANES) (Koeppe et al., 2013), while in the other 2 of these studies, the exposure was use of toothpaste containing triclosan, and was not measured (Viswanathan et al., 2012; Paul et al., 2012).

3.3.2. Non-human mammalian data

Of 8 included rat hormone studies, 6 were amenable to meta-analysis for the outcome thyroxine concentration: 3 studies with 4 datasets where triclosan was administered during gestation, and from 4 studies with 6 datasets where triclosan was administered directly to the offspring during the postnatal developmental period or in both the pre- and postnatal periods.

The thyroxine studies had the following characteristics:

- Species: rat.
- Route of exposure: oral gavage.
- Outcome measurement: thyroxine concentration.

Table 1
Human studies reporting hormone concentration outcome (N = 3).

Study	Study design	Population	Location	Outcome measures	n	Exposure assessment
Koeppe et al. (2013)	Cross sectional	U.S. population (NHANES)	United States	Serum free T3 Serum total T3 Serum free T4 Serum total T4 Serum TSH Serum thyroglobulin	1831	Urinary triclosan
Cullinan et al. (2012)	Randomized controlled trial	Subset of cardiovascular and periodontal study cohort	United States	Serum TSH Serum free T4 Serum free T3 Antithyroglobulin antibody Antithyroid peroxidase antibody	132	Use of toothpaste containing 0.3% triclosan vs. placebo
Allmyr et al. (2009)	Case-crossover experiment	Adults	Sweden	Plasma 4b-hydroxycholesterol Plasma free T3 Plasma free T4 Plasma TSH	12	Use of toothpaste containing 0.3% triclosan

Table 2
Summary of studies of triclosan and rodent hormone concentrations (N = 8 studies and 10 data sets).

Study	Species	Strain	Route of administration	Time of administration	Time of assessment	Doses tested[a]	Total number[b]	Hormone concentration outcome[c,d]
Paul et al. (2012)	Rat	Long–Evans	Oral gavage	GD6–PND21	GD20	10–300	54 litters	Total T4[e]
					GD20		54	Total T4 (dams)
					PND22		95	Total T4 (dams)
Stoker et al. (2010)	Rat	Wistar	Oral gavage	PND19–21	PND21	1.18–75	48[f]	Total T4[e]; Free T4
				PND22–42	PND42	9.375–150	50[f]	Total T4[e]; Free T4; TSH
Paul et al. (2010b)	Rat	Long–Evans	Oral gavage	GD6–PND22	PND4	30–300	34 litters[g]	Total T4
					PND14		36 litters[h]	Total T4
					PND21		37 litters[h]	Total T4[e]
Rodriguez and Sanchez (2010)	Rat	Wistar	Drinking water	8 days prior to mating–PND21	GD5, GD10, GD15, GD20, PND5, PND10, PND15, PND20	1–50	32	Total T3 (dams); Total T4 (dams)
Paul et al. (2010b)[k]	Rat	Long–Evans	Oral gavage	PND27–29 (range of age at treatment) for 4 days	PND31–33 (range of age at assessment)	10–1000	120	Total T4[e,i]
						30–1000	40	Total T3[i]; TSH[i]
Zorrilla et al. (2009)	Rat	Wistar	Oral gavage	PND23–53	PND53	3–300	71[f]	Total T4[e,i]; Total T3; TSH; Total testosterone; Total androstenedione
Kumar et al. (2009)	Rat	Wistar	Oral gavage	Approx. PN week 10 for 60 days	Approx. PND130	20	16	Testosterone; Androstenedione; Pregnenolone; Follicle stimulating hormone; Luteinizing hormone
Axelstad et al. (2013)	Rat	Wistar	Oral gavage	GD7–PND16	GD15	75–300	32	Total T4 (dams)
					PND16	75–300	32	Total T4 (dams)
							32 litters[j]	Total T4 (males and females)[e]
				PND3–16	PND16	50–150	5 litters (38 animals)[j]	Total T4 (males and females)[e]

(GD) = gestational day.
(PND) = postnatal day.
(T4) = thyroxine.
(T3) = triiodothyronine.
(TSH) = thyroid-stimulating hormone.
[a] mg/kg-bw/day. A control group was included for each study.
[b] Number of animals, unless otherwise specified. See Supplemental study characteristics tables for number of animals per dose groups.
[c] Serum measurements presented, unless otherwise specified.
[d] Outcomes for gestational exposures are for offspring, unless otherwise noted.
[e] Outcome included in meta-analysis.
[f] Exact numbers analyzed not provided; value represents estimate based on numbers allocated.
[g] Samples collected from culled pups from each litter (to normalize litter size to 8) and pooled for analysis.
[h] One male and one female selected from each litter and sample pooled for analysis.
[i] Absolute and % of control values presented.
[j] Samples were pooled within litter (by sex). Nursing litters were culled to normalize litter size to 8 but not cross-fostered. Two litters were assigned to each of 3 groups. One of 2 control dams rejected litter, leaving 1 genetically homogeneous control litter. The control litter was reported to have higher T4 levels compared to historical laboratory controls.
[k] Our search identified a study by Crofton et al. (2007), but because those data were included in the study by Paul et al. (2010b), we did not include the data reported in the Crofton et al. publication.

- Time point of outcome measurement: various prenatal or postnatal times measured in days.

We reported thyroxine concentrations and their standard errors, as a percentage normalized to the concentration in the control group. We were unable to obtain raw data from studies that already reported normalized concentrations. The result was an estimate of the overall mean change in thyroxine concentration for a 1-unit increase in triclosan (mg/kg-bw-day), accounting for within- and between-study variability. We used only data from triclosan doses equal to or below 300 mg/kg-body weight (bw)-day. The dose was limited to focus on effects at lower tested doses and to minimize adverse impacts from responses at higher doses (such as litter loss) on the overall estimate and to account for the model assumptions of linearity. One dose group was therefore omitted: 1000 mg/kg-bw-day (Paul et al., 2010a).

Administration of triclosan to dams during gestation was not associated with a consistent dose response in the offspring; however, one study (Paul et al., 2012) evaluated thyroid hormone levels during gestation and showed a significant dose–response in fetuses (Fig. 4). The overall pooled meta-analysis estimate was a 0.09% reduction in thyroxine per mg/kg unit increase in triclosan (95% CI −0.20 to 0.02; I^2 = 22.8%; Fig. 4B). In contrast, there was a clear dose response for triclosan administered during the postnatal developmental time period (Fig. 5A) and the overall pooled meta-analysis estimate was a 0.31% reduction in thyroxine per mg/kg unit increase in triclosan (95% CI −0.38 to −0.23; I^2 = 61.5%; Fig. 5B). For other hormones (4 studies) we generally observed a trend towards a reduction in concentration, although there were limited data on each hormone and confidence intervals mostly overlapped (Supplemental material, Fig. S2).

3.4. Rating the quality and strength of the bodies of evidence for hormone modulation

3.4.1. Human evidence

We rated the overall quality of the human evidence "low to moderate." We rated the final overall strength of the human evidence "inadequate" (Table 3). Our rating of "inadequate" human evidence was based on insufficient evidence to assess the association between triclosan and human thyroid hormone concentrations. There were few studies (2 small studies and 1 large study) with inconsistent findings.

3.4.2 Non-human mammalian evidence.

Each factor considered in rating the overall quality of the non-human mammalian (rat) hormonal evidence was consistent among reviewers except for "risk of bias" where 9 reviewers rated (−1); 2 reviewers (0); and 1 reviewer (0/−1) (Table 4). Ultimately we reached consensus agreement to downgrade one level (to "moderate" quality) based on our concerns about risk of bias, as we had rated "probably

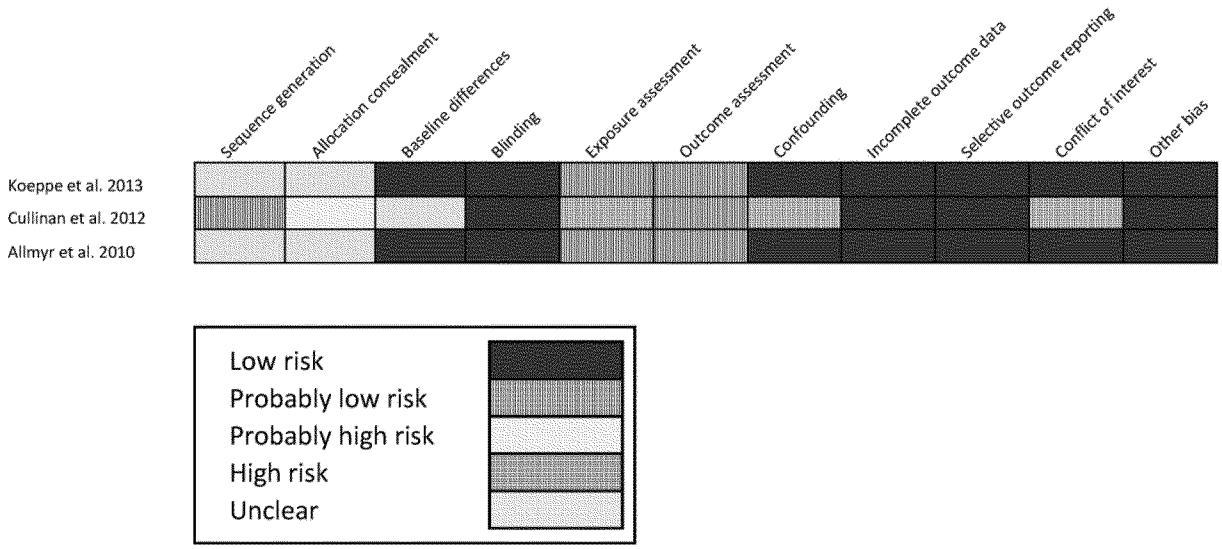

Fig. 2. Summary of risk of bias graphs for individual human studies. Review authors' judgments (low, probably low, probably high, and high risk) of bias for each risk of bias domain for each included human study (n = 3). The risk of bias results did not differ according to different specific outcome measures within the studies, and therefore results are presented by study. Cullinan et al. was a subset of a randomized controlled trial and was therefore evaluated under the "Sequence generation" and "Allocation" risk of bias domains and received a "N/A" (not applicable) rating under the "Baseline differences" domain. The other studies received a "N/A" rating under the "Sequence generation" and "Allocation" domains.

high" risk of bias across several domains, particularly for allocation concealment and blinding (Table 4). We also had consensus on the final overall strength of the rodent evidence (sufficient), based on consistency in the findings of the studies and the meta-analysis estimate of reduced thyroxine concentrations in relation to postnatal triclosan exposure (Table 4).

Based on our evaluation using the Navigation Guide criteria, we concluded that there was "sufficient" non-human evidence and "inadequate" human evidence of an association between triclosan exposure and thyroxine concentrations. Consequently, we concluded that triclosan is "possibly toxic" to reproductive and developmental health, based on the Navigation Guide evidence integration step (Fig. S1).

4. Discussion

We applied the Navigation Guide systematic review method to assess whether exposure to triclosan has adverse effects on human development or reproduction and found that triclosan is "possibly toxic" to reproductive and developmental health, based on its adverse impacts on the thyroid hormone thyroxine. Thyroid hormone disruption is an

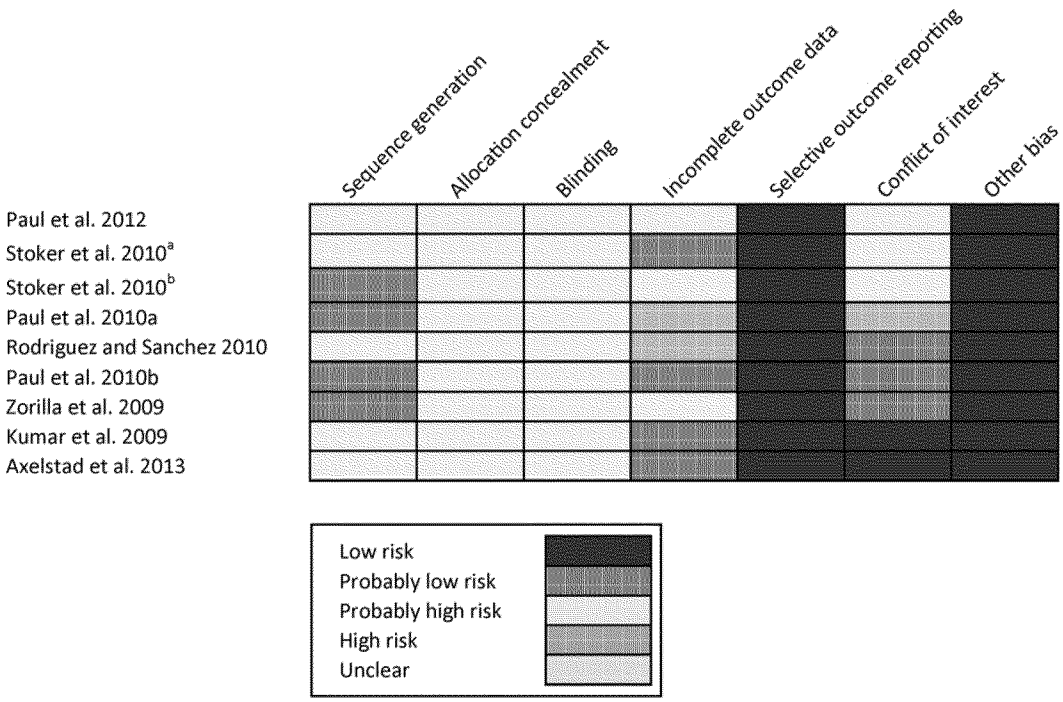

Fig. 3. Summary of risk of bias graphs for individual animal studies. Review authors' judgments (low, probably low, probably high, and high risk) of bias for each risk of bias domain for each included animal study (n = 8). Note Stoker et al. presents two experiments using two separate cohorts (pubertal assay and uterotrophic assay); each cohort was evaluated for risk of bias separately. [a]Stoker et al. pubertal assay cohort assessed free T4, total T4, and TSH. [b]Stoker et al. uterotrophic assay cohort assess free T4 and total T4.

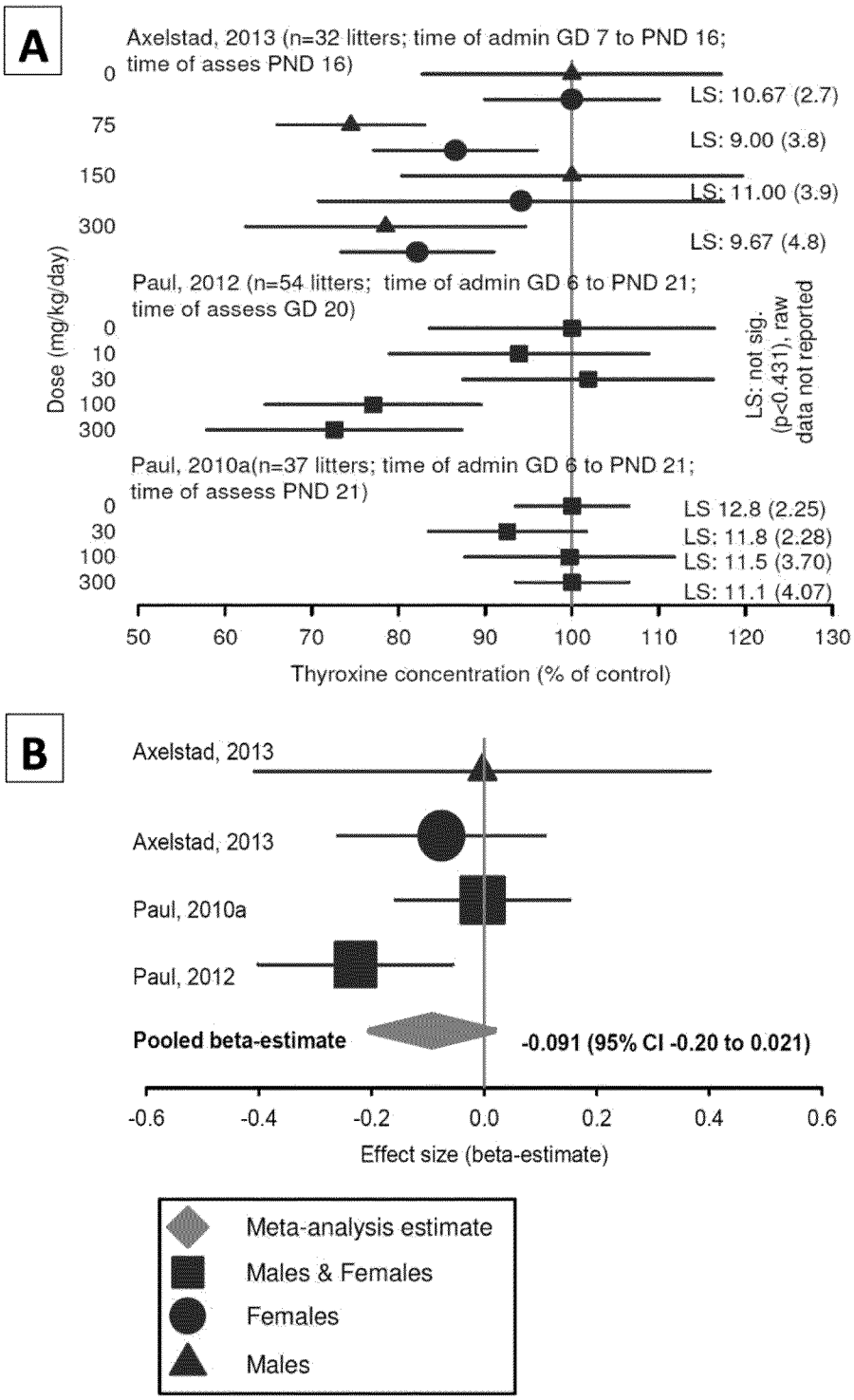

Fig. 4. Prenatal triclosan and thyroxine. A. Prenatal triclosan administration and thyroxine concentration as a percentage of the control group for doses up to 300 mg/kg/day. B. Prenatal beta-estimates for dose response and the random effects meta-analysis estimate. The vertical gray bar in A represents the line of no effect (the control group normalized to 100%); horizontal error bars represent 95% confidence intervals; in B, symbol sizes represent the log of the weight in the meta-analysis.

upstream indicator of developmental toxicity (Miller et al., 2009; Woodruff et al., 2008; Crofton, 2008; Wise et al., 2012).

One of the goals of this review and other case studies of applying the Navigation Guide methodology (Koustas et al., 2014; Johnson et al., 2014; Lam et al., 2014; Vesterinen et al., 2014) was to develop proof of concept of the use of improved methods of evidence integration in environmental health. Such an incremental methods testing approach has been successful in clinical medicine in developing an empirical basis for evidence-based medicine (Higgins and Green, 2011). The relatively few human studies in the triclosan case study revealed points of methodological consistency and inconsistency between the Navigation Guide and other methods of evidence integration related to how the terminology "possibly toxic" and "probably toxic" mapped to the human and non-human evidence.

Our overall quality rating system for non-human evidence was consistent with approaches adopted by the U.S. EPA for carcinogens

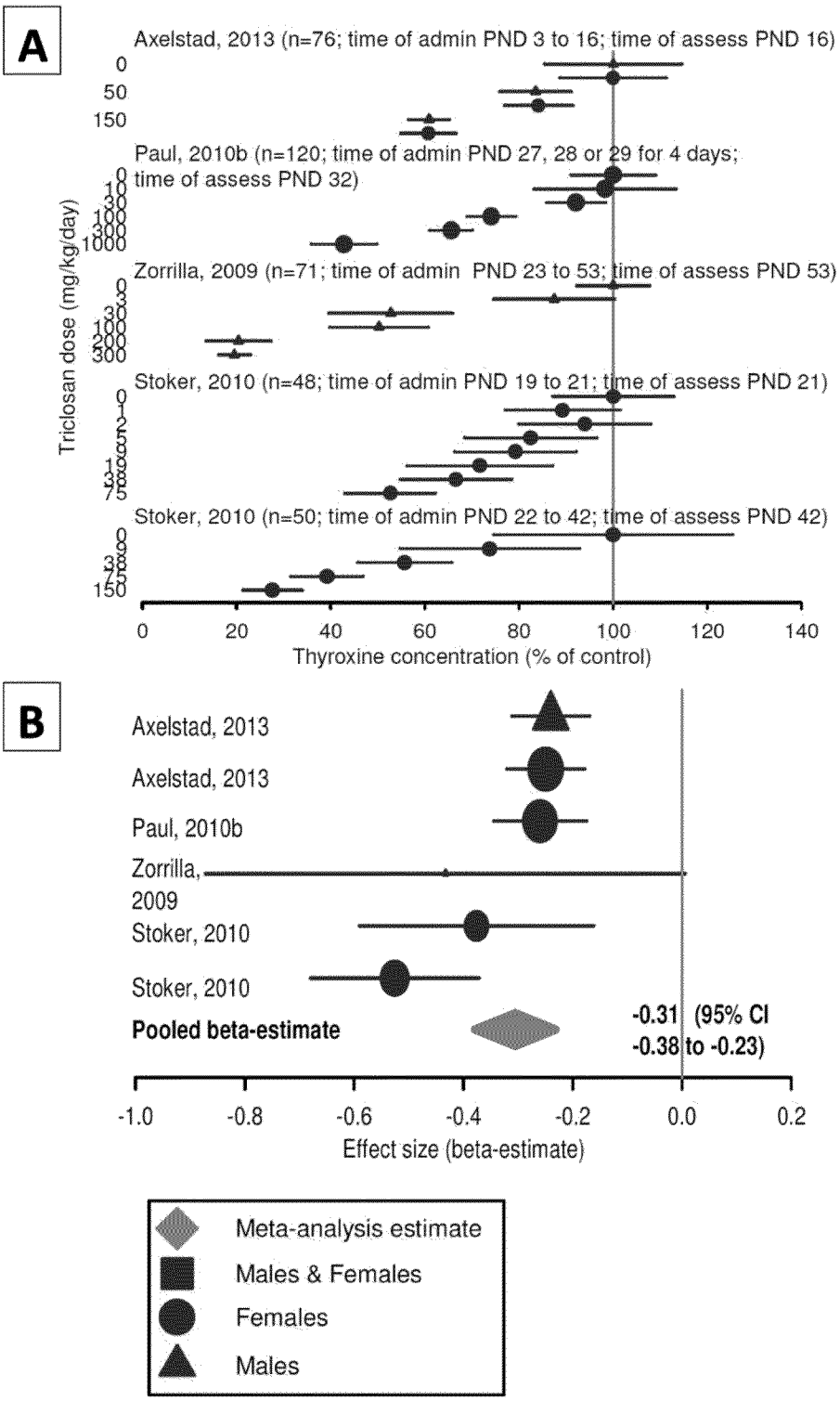

Fig. 5. Postnatal triclosan and thyroxine. A. Postnatal triclosan administration and thyroxine concentrations as a percentage of the control group for doses up to 300 mg/kg/day. B. Postnatal beta-estimates for dose response and the random effects meta-analysis estimate. The vertical gray bar in A represents the line of no effect (the control group normalized to 100%); horizontal error bars represent 95% confidence intervals; in B, symbol sizes represent the log of the weight in the meta-analysis.

and in the NTP-OHAT method in that it allowed for a finding of "sufficient" evidence based on positive findings in multiple studies or a single appropriate study in a single species (National Research Council, 2014b; Woodruff et al., 2008). However, the structure of our evidence integration table, modeled after the IARC evidence integration table for cancer (International Agency for Research on Cancer, 2006) does not align with U.S. EPA and NTP-OHAT when there was "insufficient" human evidence. As adapted from IARC's preamble (International Agency for Research on Cancer, 2006), in the absence of consideration of mechanistic data, the Navigation Guide evidence integration step requires both "limited" human and "sufficient" non-human evidence of toxicity in order for a chemical to be found to be "probably toxic" (Fig. S1). Current practice in U.S. EPA assessments of non-cancer health outcomes (Miller et al., 2009; Woodruff et al.,

Table 3
Summary of rating quality and strength of the human hormonal evidence.

Category	Downgrades	Rationale
Risk of bias	Eight (0); Four (−1)	Two of the three studies, one large and one small, had "low" or "probably low" risk of bias for all domains. However, some authors were more concerned about the potential risk of bias in the exposure assessment.
Indirectness	Six (0): Six (−1)	One study (Cullinan et al.) is of an older age group not representative of reproductive age where thyroid is a developmental or reproductive concern; Cullinan et al. exposure assessment by toothpaste use only is indirect. The concerns about this one study did not warrant a downgrade for some authors; but for some the concern, particularly for indirect exposure assessment, warranted a downgrade.
Inconsistency	Twelve (0)	The results of the 3 studies were consistent.
Imprecision	Twelve (0)	Although the Koeppe et al. study had some wide confidence intervals, most confidence intervals were sufficiently narrow.
Publication bias	Twelve (0)	There was variability in study size and there was a larger study (Koeppe et al.) showing no effect for some outcomes. A comprehensive literature search did not identify studies with conflicting results. There were not enough studies to utilize funnel plot analyses to assess publication bias.
	Upgrades	
Large magnitude of effect	Twelve (0)	All of the studies found null or minimal effects only.
Dose–response	Eleven (0); One (+1)	Most reviewers found minimal to no evidence of a dose–response gradient. One reviewer downgraded based on a statistically insignificant dose–response gradient.
Confounding minimizes effect	Twelve (0)	There was no evidence that residual confounding influenced results.
Overall quality of evidence (initial rating is "Moderate")	Seven (Moderate); Five (Low)	
Overall strength of evidence	Inadequate	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.

2008) and the NTP's framework do not require "sufficient" or "limited" human evidence to reach a comparable strength of evidence conclusion (National Research Council, 2014b). In the NTP-OHAT method, a chemical can be found to have a "presumed" hazard based on a combination of a "high" level of evidence in non-human studies and a "low," (equivalent to Navigation Guide "inadequate"), or "moderate" level of evidence in human studies (National Research Council, 2014b; National Toxicology Program, 2015) For future cases, we intend to revise the Navigation Guide method for evidence integration to better align with the labeling of the NTP-OHAT method (Rooney et al., 2014) and current practices at the U.S. EPA, such that "probably toxic," which more closely maps to "presumed," is reachable with strong non-human evidence.

This case study demonstrates that all conclusions in environmental health about a chemical's toxicity are limited by the available data. Of the few human studies on triclosan, even fewer presented results for the same outcome. For the non-human mammalian evidence, we found studies conducted at various stages of development and reporting over 100 unique outcome measures. For many endpoints, the data were too limited to assess and most data were not conducive to combining into meta-analysis. While conducting a meta-analysis is not an essential component of hazard or risk assessment, it can be a useful tool for synthesizing data. We narrowed the final analysis to the health outcome with the most data, which may not equate with the most sensitive health outcome or represent the best method of focusing an investigation. Our results were primarily based on postnatal effects in the non-human mammalian literature, as only one of the studies evaluated effects on thyroid hormones during gestation. This is a challenge as previous literature finds that thyroid hormone levels during gestation is an indicator of future neurodevelopment (Wise et al., 2012; Morreale, 2001; Mastorakos et al., 2007). Our findings also illustrate a strength of systematic reviews in that the method identifies research gaps which can inform how scarce research funding could be most efficiently and effectively targeted to answer a policy relevant question. A complete list of relevant studies is included in the Supplemental information and could be a starting point for identifying where research could be directed to strengthening the evidence base.

This was the first systematic review of the human and non-human mammalian evidence for triclosan and reproductive and developmental effects. One of the main strengths of systematic reviews is that the criteria and rationale for judgements and decisions are transparently documented. A different set of authors could presumably arrive at a different conclusion, but with this thorough documentation, the review is reproducible and the reader can understand what led to the difference.

The present review elucidates the potential hazard of triclosan and does not estimate exposure or conduct a quantitative risk assessment.

Table 4
Summary of rating quality and strength of the non-human mammalian hormonal evidence.

Category	Downgrades	Rationale
Risk of bias	Nine (−1); Two (0); One (0/−1)	(−1): There was "probably high" risk of bias across several domains; (0): Concern about overall risk of bias does not rise to the level of a downgrade; (0/−1): Most of the studies have "probably high" risk, rather than "high risk," and this was mostly due to unknown information about the studies.
Indirectness	Twelve (0)	Animal changes (in rodents) are reflective of what is seen in humans and the outcomes were directly relevant to humans.
Inconsistency	Twelve (0)	There was not substantial heterogeneity in studies across postnatal dosing for thyroxine; lack of consistency between post- and prenatal dosing has a biological explanation.
Imprecision	Twelve (0)	The confidence intervals were not wide for the thyroxine studies or the meta-analysis.
Publication bias	Twelve (0)	There were not enough studies to utilize funnel plot analyses to assess publication bias. However, we conducted a comprehensive search and found studies of variable sizes and funding sources. Studies include null findings as well as positive findings from studies with high risk for conflict of interest. On this basis we did not downgrade for publication bias.
Overall quality of evidence (initial rating is "High")	Moderate	We downgraded one level based on concerns about risk of bias.
Overall strength of evidence	Sufficient	We found sufficient evidence that exposure to triclosan alters hormone levels in rats, based on reduced thyroxine levels.

This is a distinction from previous reviews and risk assessments that appear to have reached conclusions differing from the current systematic review. Rodricks et al. concluded, based on estimates of a benchmark dose level and human exposure, that triclosan in consumer products is not expected to cause adverse effects (Rodricks et al., 2010). The Colgate-Palmolive Company-sponsored narrative review of endocrine disrupting activity of triclosan by Witorsch concluded that personal care products containing triclosan do not pose a risk of adverse effects from endocrine disruption (Witorsch, 2014). While both the present review and the Witorsch review found insufficient evidence in humans and evidence of a dose-dependent decrease in thyroxine in rats, our conclusions about the available evidence differed from Witorsch for several reasons. First, our criteria for reaching a decision about a chemical's toxicity were defined and stated before our review was undertaken. In our review we had consensus on the final overall strength of the rodent evidence (sufficient), based on consistency in the findings of the studies and the meta-analysis estimate of reduced thyroxine concentrations in relation to postnatal triclosan exposure (Tables 4 and S4). In contrast, the Witorsch narrative review had no predefined criteria for reaching its conclusion and ultimately discounted the rat findings on thyroxine because: (1) related findings were not present for other thyroid system endpoints, namely TSH, T3, thyroid histology or thyroid weight; (2) rats were not considered a proven model system for thyroid disruption; and (3) the mode of action for T4 disruption was unknown and/or inconsistent. We did not require that these three criteria be met in order to consider triclosan "possibly" toxic. We base this on previous literature identifying that thyroid hormone disruption, in particular thyroxine decrements, is an indicator of adverse effects (Miller et al., 2009; Woodruff et al., 2008; Crofton, 2008; Wise et al., 2012). In short, having consistent disruption of all thyroid system endpoints, in human studies (implicit if rats are to be discounted), and a documented mode of action sets a very high bar for demonstrating a chemical's toxicity. In addition, it is not consistent with the broad range of evidence evaluations by authoritative bodies such as U.S. EPA and IARC and is not necessary to make determinations about hazard (e.g., the mechanism of smoking is not known, but it is a carcinogen).

A second possible reason for the difference between our conclusion that triclosan is "possibly toxic" versus Witorsch's "TCS does not present a risk of endocrine disruptive health effects through exposure to personal care products" is that our review focuses on the potential hazard of triclosan and does not estimate exposure or conduct a risk assessment. Health Canada did not consider thyroid function in rats a critical effect for risk characterization of triclosan in humans, although they acknowledged the uncertainty in human relevance of triclosan-induced hypothyroxemia and the lack of developmental neurotoxicity data for triclosan (Health Canada, 2012). The European Union's Scientific Committee on Consumer Safety (SCCS) acknowledged differences between rats and humans with respect to thyroid hormone physiology and regulation, and they did not use the acceptable level of exposure, derived from rat studies, in assessing risk of thyroid hormone effects (SCCS. Scientific Committee on Consumer Safety, 2011). The SCCS conducted a risk assessment using exposure levels based on animal studies of other endpoints (e.g., hematotoxicity, reproductive effects) and concluded that triclosan is safe as used in some personal care products but not safe when considering aggregate exposures or high exposures resulting from the use of certain leave-on cosmetics such as body lotion (SCCS. Scientific Committee on Consumer Safety, 2011). None of these risk assessments included a systematic review of the reproductive and developmental hazard before undertaking the risk assessment (SCCS. Scientific Committee on Consumer Safety, 2011; Health Canada, 2012; Paul et al., 2013).

Thyroid hormone disruption is concerning because even small reductions in thyroxine in pregnant women can have adverse effects on neurodevelopment of children (Miller et al., 2009; Woodruff et al., 2008; Wise et al., 2012; Ghassabian et al., 2014; Henrichs et al., 2010). Because there is widespread exposure to triclosan, a finding that triclosan is "possibly toxic" has important public health implications.

Contrary to our previous systematic review of PFOA and fetal growth (Koustas et al., 2014; Johnson et al., 2014) our efforts to obtain additional unpublished information by contacting study authors were largely unsuccessful, and we did not receive a reason to explain the difference in response rates between the two reviews. This finding underscores the need for systemic change in how research findings are reported in environmental health such as by adoption of the Animal Research: Reporting of In Vivo Experiments (ARRIVE) and the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) guidelines, in addition to reporting further information as we describe in Vesterinen et al. (2013).

We did not downgrade the quality rating of the body of evidence for publication bias because we had no direct evidence that it existed. Because the body of literature on triclosan is relatively small, we were unable to evaluate publication bias using the funnel plot method typically used in systematic reviews in the healthcare field. As such, we cannot rule out that a publication bias exists.

As with our previous PFOA case study (Koustas et al., 2014) the majority of the included animal studies were "probably high risk of bias", particularly for the "allocation concealment" and "blinding" domains. This "worrisome truth" about the conduct and reporting of experimental animal studies in environmental health (Woodruff and Sutton, 2014) is also prevalent in the preclinical literature, and introduces bias into study findings (Bebarta et al., 2003; Landis et al., 2012; Macleod et al., 2004; McPartland et al., 2007; van der Worp and Macleod, 2011; van der Worp et al., 2007; Vesterinen et al., 2011; Holman et al., 2015).

There were other important limitations of some of the included studies. For example, the paper by Axelstad et al. (2013) reports on two separate experiments. One reasonably well-conducted experiment exposed pregnant and lactating rats to triclosan, and evaluated thyroxine levels in dams and their offspring. The second experiment involved direct dosing of nursing pups with triclosan in a corn oil vehicle. As the study authors point out, the results of the second experiment are compromised by genetic homogeneity among pups of the single surviving control litter, as well as by the high thyroxine levels in this control litter compared to their laboratory's historical controls. In addition, because the studies by Paul et al. combined males and females, they may have masked any sex-dependent differences in effect.

We designed our search to capture a wide range of outcomes by using chemical terms only and not limiting the search with outcome terms. This was an effective strategy because there were a relatively small number of studies on triclosan. Developing our PECO question and reference screening criteria was an undertaking that leveraged the extensive knowledge of the scientists at the CalEPA Office of Environmental Health Hazard Assessment. Our experience in developing these criteria points to the need for topic experts to be engaged in systematic reviews from the onset of the review.

Consistent with our previous case study (Koustas et al., 2014; Johnson et al., 2014), we found it was efficient to sort through a large number of studies captured through our search due to predefined exclusion criteria (derived from the PECO statement) and the use of Distiller software; on average it took approximately 15 s to screen each abstract and eliminate the majority of irrelevant studies. Screening potentially relevant full texts took on average 1.5 min per study.

While the efficiency and effectiveness of our screening methods expedited the review, the lack of tools to assess risk of bias for the diversity of evidence streams retrieved, i.e., invertebrate studies, in vitro studies, fish studies, and amphibian studies impeded inclusion of all of the relevant data. We lacked the time, resources, and expertise to develop the necessary assessment tools for the non-human non-mammalian evidence streams in the one year we had allocated to complete this case study. Hence, we were unable to include these studies in the review as we had initially set out to do. Risk of bias assessment tools for

model systems in environmental health is a critical research and development need in evidence integration. A critical requirement of evidence integration in environmental health is that each stream of evidence, i.e., human, non-human, mechanistic, etc., needs to be systematically reviewed, including for risk of bias for individual studies, before this evidence is integrated into the results. Future work will also look to establish precedents for efficient systematic assessment for chemicals with larger data sets, multiple inter-related endpoints that reflect disruption of fundamental developmental or reproductive processes, supported by a robust mechanistic literature.

In summary, we found that there was sufficient non-human evidence and inadequate human evidence of an association between triclosan exposure and thyroxine concentrations, and that triclosan is "possibly toxic" to reproductive and developmental health. Triclosan has a relatively sparse data set, with few human studies. Our conclusion was based on the most data rich endpoint, not necessarily the most sensitive endpoint, and it excluded consideration of non-mammalian data due to heterogeneity of these data and a corresponding lack of methods for assessing the quality of these studies. Our conclusion that triclosan is "possibly toxic" illustrates that current regulatory policies permit widespread exposure to environmental chemicals in the absence of evidence of safety.

Acknowledgements

We acknowledge the following persons for their contributions: Daniel Axelrad at U.S. Environmental Protection Agency (U.S. EPA) for assistance in selecting the study question and reviewing the protocol and findings; Martha Sandy and the late George Alexeeff of the California Environmental Protection Agency (CalEPA) for insightful comments and suggestions; Kathryn Guyton at the International Agency for Research on Cancer for assistance with developing evidence evaluation methodologies and identifying relevant toxicological databases; Kristina Thayer, Andrew Rooney and Abee Boyles in the Office of Health Assessment and Translation, National Toxicology Program for assistance with developing risk of bias and evidence evaluation criteria; Ryan Babadi (RB) and Adam D'Amico for assistance with literature searching; Kari Weber for assistance with data extraction; and Natalyn Daniels for assistance with manuscript preparation. Dr. Woodruff is also a faculty member in the Philip R. Lee Institute for Health Policy Studies at the University of California, San Francisco. The authors declare no competing financial interests. The views expressed in this paper are those of the authors and do not necessarily reflect the views or policies of the U.S. Environmental Protection Agency or the California Environmental Protection Agency (CalEPA).

This research was funded by the CalEPA's Office of Environmental Health Hazard Assessment (OEHHA) and by the U.S. EPA through a contract with Abt Associates (GAIA-0-6-UCSF17288), the National Institute of Environmental Health Sciences (ESO22841), U.S. EPA STAR grant (RD83543301), and the Fred Gellert Foundation. The research was supported in part by appointments to the Internship/Research Participation Program at the National Center for Environmental Economics, U.S. EPA, administered by the Oak Ridge Institute for Science and Education through an interagency agreement between the U.S. Department of Energy and U.S. EPA.

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.envint.2016.03.009 .

References

Adolfsson-Erici, M., Pettersson, M., Parkkonen, J., Sturve, J., 2002. Triclosan, a commonly used bactericide found in human milk and in the aquatic environment in Sweden. Chemosphere 46, 1485–1489.

Aiello, A.E., Larson, E.L., Levy, S.B., 2007. Consumer antibacterial soaps: effective or just risky? Clin. Infect. Dis. 45 (Suppl. 2), S137–S147.

Allmyr, M., Panagiotidis, G., Sparve, E., Diczfalusy, U., Sandborgh-Englund, G., 2009. Human exposure to triclosan via toothpaste does not change CYP3A4 activity or plasma concentrations of thyroid hormones. Basic Clin. Pharmacol. Toxicol. 105, 339–344.

The Alliance for the Prudent Use of Antibiotics, 2011. Triclosan. Author, Boston, MA.

Axelstad, M., Boberg, J., Vinggaard, A.M., Christiansen, S., Hass, U., 2013. Triclosan exposure reduces thyroxine levels in pregnant and lactating rat dams and in directly exposed offspring. Food Chem. Toxicol. 59, 534–540.

Balshem, H., Helfand, M., Schunemann, H.J., Oxman, A.D., Kunz, R., Brozek, J., et al., 2011. GRADE guidelines: 3. Rating the quality of evidence. J. Clin. Epidemiol. 64, 401–406.

Bebarta, V., Luyten, D., Heard, K., 2003. Emergency medicine animal research: does use of randomization and blinding affect the results? Acad. Emerg. Med. 10, 684–687.

Bertelsen, R.J., Engel, S.M., Jusko, T.A., Calafat, A.M., Hoppin, J.A., London, S.J., et al., 2014. Reliability of triclosan measures in repeated urine samples from Norwegian pregnant women. J. Exposure Sci. Environ. Epidemiol. 24, 517–521.

Calafat, A.M., Ye, X., Wong, L.Y., Reidy, J.A., Needham, L.L., 2008. Urinary concentrations of triclosan in the US population: 2003–2004. Environ. Health Perspect. 116, 303–307.

Chen, M., Tang, R., Fu, G., Xu, B., Zhu, P., Qiao, S., et al., 2013. Association of exposure to phenols and idiopathic male infertility. J. Hazard. Mater. 250–251, 115–121.

Ciba Specialty Chemicals Corporation, 2004. Triclosan Safety Studies Submitted to OTC Docket 81 N-033P. U.S. Food and Drug Administration, Washington, D.C.

Cooney, C.M., 2010. Triclosan comes under scrutiny. Environ. Health Perspect. 118, A242-A.

Crofton, K.M., 2008. Thyroid disrupting chemicals: mechanisms and mixtures. Int. J. Androl. 31, 209–223.

Crofton, K.M., Paul, K.B., Devito, M.J., Hedge, J.M., 2007. Short-term in vivo exposure to the water contaminant triclosan: evidence for disruption of thyroxine. Environ. Toxicol. Pharmacol. 24, 194–197.

Cullinan, M.P., Palmer, J.E., Carle, A.D., West, M.J., Seymour, G.J., 2012. Long term use of triclosan toothpaste and thyroid function. Sci. Total Environ. 416, 75–79.

Dann, A.B., Hontela, A., 2011. Triclosan: environmental exposure, toxicity and mechanisms of action. J. Appl. Toxicol. 31, 285–311.

Fang, J.L., Stingley, R.L., Beland, F.A., Harrouk, W., Lumpkins, D.L., Howard, P., 2010. Occurrence, efficacy, metabolism, and toxicity of triclosan. J. Environ. Sci. Health C Environ. Carcinog. Ecotoxicol. Rev. 28, 147–171.

Foran, C.M., Bennett, E.R., Benson, W.H., 2000. Developmental evaluation of a potential non-steroidal estrogen: triclosan. Mar. Environ. Res. 50, 153–156.

Ghassabian, A., El Marroun, H., Peeters, R.P., Jaddoe, V.W., Hofman, A., Verhulst, F.C., et al., 2014. Downstream effects of maternal hypothyroxinemia in early pregnancy: nonverbal IQ and brain morphology in school-age children. J. Clin. Endocrinol. Metab. 99, 2383–2390.

Guyatt, G.H., Oxman, A.D., Vist, G., Kunz, R., Brozek, J., Alonso-Coello, P., et al., 2011. GRADE guidelines: 4. Rating the quality of evidence—study limitations (risk of bias). J. Clin. Epidemiol. 64, 407–415.

Halden, R.U., 2014. On the need and speed of regulating triclosan and triclocarban in the United States. Environ. Sci. Technol. 48, 3603–3611.

Health Canada, 2012. Preliminary Assessment: Triclosan Chem Abs.

Henrichs, J., Bongers-Schokking, J.J., Schenk, J.J., Ghassabian, A., Schmidt, H.G., Visser, T.J., et al., 2010. Maternal thyroid function during early pregnancy and cognitive functioning in early childhood: the generation R study. J. Clin. Endocrinol. Metab. 95, 4227–4234.

Higgins, J.P.T., Green, S., 2011. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [Updated March 2011]: The Cochrane Collaboration. Available from www.cochrane-handbook.org.; (Available from: Available from www.cochrane-handbook.org).

Hollowell JG, Staehling NW, Flanders WD, Hannon WH, Gunter EW, Spencer CA, et al. Serum TSH, T4, and thyroid antibodies in the United States population (1988 to 1994): National Health and Nutrition Examination Survey (NHANES III). [other] 2013 [updated 2013/07/02/, 2014/03/25/21:35:21]; Available from: http://press.endocrine.org/doi/abs/10.1210/jcem.87.2.8182

Holman, L., Head, M.L., Lanfear, R., Jennions, M.D., 2015. Evidence of experimental bias in the life sciences: why we need blind data recording. PLoS Biol. 13, e1002190.

International Agency for Research on Cancer, 2006. IARC Monographs on the Evaluation of Carcinogenic Risks to Humans: Preamble (amended January 2006). World Health Organization, Lyon, France.

James, M.O., Li, W., Summerlot, D.P., Rowland-Faux, L., Wood, C.E., 2010. Triclosan is a potent inhibitor of estradiol and estrone sulfonation in sheep placenta. Environ. Int. 36, 942–949.

Johnson, P.I., Koustas, E., Vesterinen, H.M., Sutton, P., Atchley, D.S., Kim, A.N., et al., 2013. Applying the Navigation Guide Case Study #2 Reproductive and Developmental Effects of Exposure to Triclosan: A Systematic Review of the Evidence Protocol.

Johnson, P.I., Sutton, P., Atchley, D.S., Koustas, E., Lam, J., Sen, S., et al., 2014. The Navigation Guide—evidence-based medicine meets environmental health: systematic review of human evidence for PFOA effects on fetal growth. Environ. Health Perspect. 122, 1028–1039.

Knudsen, N., Laurberg, P., Rasmussen, L.B., Bülow, I., Perrild, H., Ovesen, L., et al., 2005. Small differences in thyroid function may be important for body mass index and the occurrence of obesity in the population. J. Clin. Endocrinol. Metab. 90, 4019–4024.

Koeppe, E.S., Ferguson, K.K., Colacino, J.A., Meeker, J.D., 2013. Relationship between urinary triclosan and paraben concentrations and serum thyroid measures in NHANES 2007–2008. Sci. Total Environ. 445–446, 299–305.

Koustas, E., Lam, J., Sutton, P., Johnson, P.I., Atchley, D.S., Sen, S., et al., 2014. The Navigation Guide—evidence-based medicine meets environmental health: systematic review of nonhuman evidence for PFOA effects on fetal growth. Environ. Health Perspect. 122, 1015–1027.

Krauth, D., Woodruff, T.J., Bero, L., 2013. Instruments for assessing risk of bias and other methodological criteria of published animal studies: a systematic review. Environ. Health Perspect. 121, 985–992.
Kumar, V., Chakraborty, A., Kural, M.R., Roy, P., 2009. Alteration of testicular steroidogenesis and histopathology of reproductive system in male rats treated with triclosan. Reprod. Toxicol. 27, 177–185.
Lam, J., Sutton, P., Koustas, E., Johnson, P., Atchley, D.S., Sen, S., et al., 2014. The Navigation Guide—evidence-based medicine meets environmental health: integration of animal and human evidence for PFOA effects on fetal growth. Environ. Health Perspect. 122, 1041–1051.
Landis, S.C., Amara, S.G., Asadullah, K., Austin, C.P., Blumenstein, R., Bradley, E.W., et al., 2012. A call for transparent reporting to optimize the predictive value of preclinical research. Nature 490, 187–191.
Lankester, J., Patel, C., Cullen, M.R., Ley, C., Parsonnet, J., 2013. Urinary triclosan is associated with elevated body mass index in NHANES. PLoS One 8.
Lundh, A., Sismondo, S., Lexchin, J., Busuioc, O.A., Bero, L., 2012. Industry sponsorship and research outcome. Cochrane database of systematic reviews (Online). p. 12.
Macleod, M.R., O'Collins, T., Howells, D.W., Donnan, G.A., 2004. Pooling of animal experimental data reveals influence of study design and publication bias. Stroke 35, 1203–1208.
Mastorakos, G., Karoutsou, E.I., Mizamtsidi, M., Creatsas, G., 2007. The menace of endocrine disruptors on thyroid hormone physiology and their impact on intrauterine development. Endocrine 31, 219–237.
Matsumura, N., Ishibashi, H., Hirano, M., Nagao, Y., Watanabe, N., Shiratsuchi, H., et al., 2005. Effects of nonylphenol and triclosan on production of plasma vitellogenin and testosterone in male South African clawed frogs (Xenopus laevis). Biol. Pharm. Bull. 28, 1748–1751.
McPartland, J.M., Glass, M., Pertwee, R.G., 2007. Meta-analysis of cannabinoid ligand binding affinity and receptor distribution: interspecies differences. Br. J. Pharmacol. 152, 583–593.
Miller, M.D., Crofton, K.M., Rice, D.C., Zoeller, R.T., 2009. Thyroid-disrupting chemicals: interpreting upstream biomarkers of adverse outcomes. Environ. Health Perspect. 117, 1033–1041.
Morreale de Escobar, G., 2001. The role of thyroid hormone in fetal neurodevelopment. J. Pediatr. Endocrinol. Metab. 14 (Suppl. 6), 1453–1462.
National Research Council, 2014a. Review of EPA's Integrated Risk Information System (IRIS) Process. National Academies Press, Washington, D.C.
National Research Council, 2014b. Review of the Environmental Protection Agency's State-of-the-Science Evaluation of Nonmonotonic Dose–Response Relationships as They Apply to Endocrine Disruptors. National Academies Press, Washington, D.C.
National Toxicology Program, 2015. In: U.S. Department of Health and Human Services (Ed.), Handbook for Conducting a Literature-Based Health Assessment Using OHAT Approach for Systematic Review and Evidence Integration. Office of Health Assessment and Translation, Division of National Toxicology Program, National Institute of Environmental Health Sciences.
Paul, K.B., Hedge, J.M., DeVito, M.J., Crofton, K., 2010a. Short-term exposure to triclosan decreases thyroxine in vivo via upregulation of hepatic catabolism in young Long–Evans rats. Toxicol. Sci. 113, 367–379.
Paul, K.B., Hedge, J.M., DeVito, M.J., Crofton, K.M., 2010b. Developmental triclosan exposure decreases maternal and neonatal thyroxine in rats. Environ. Toxicol. Chem. 29, 2840–2844.
Paul, K.B., Hedge, J.M., Bansal, R., Zoeller, R.T., Peter, R., DeVito, M.J., et al., 2012. Developmental triclosan exposure decreases maternal, fetal, and early neonatal thyroxine: a dynamic and kinetic evaluation of a putative mode-of-action. Toxicology 300, 31–45.
Paul, K.B., Thompson, J.T., Simmons, S.O., Vanden Heuvel, J.P., Crofton, K.M., 2013. Evidence for triclosan-induced activation of human and rodent xenobiotic nuclear receptors. Toxicol. in Vitro 27, 2049–2060.
Perencevich, E.N., Wong, M.T., Harris, A.D., 2001. National and regional assessment of the antibacterial soap market: a step toward determining the impact of prevalent antibacterial soaps. Am. J. Infect. Control 29, 281–283.
Robertshaw, H., Leppard, B., 2007. Contact dermatitis to triclosan in toothpaste. Contact Dermatitis 57 383-3U4.
Rodricks, J.V., Swenberg, J.A., Borzelleca, J.F., Maronpot, R.R., Shipp, A.M., 2010. Triclosan: a critical review of the experimental data and development of margins of safety for consumer products. Crit. Rev. Toxicol. 40, 422–484.
Rodriguez, P.E., Sanchez, M.S., 2010. Maternal exposure to triclosan impairs thyroid homeostasis and female pubertal development in Wistar rat offspring. J. Toxic. Environ. Health A 73, 1678–1688.
Rooney, A.A., Boyles, A.L., Wolfe, M.S., Bucher, J.R., Thayer, K.A., 2014. Systematic review and evidence integration for literature-based environmental health science assessments. Environ. Health Perspect.
Sawaya, G.F., Guirguis-Blake, J., LeFevre, M., Harris, R., Petitti, D., 2007. Update on methods: estimating certainty and magnitude of net benefit. U.S. Preventive Services Task Force. Available from: http://www.uspreventiveservicestaskforce.org/uspstf07/methods/benefit.htm#copyright
SCCS. Scientific Committee on Consumer Safety, 2011. Addendum to SCCP Opinion on Triclosan.
Schena, D., Papagrigoraki, A., Girolomoni, G., 2008. Sensitizing potential of triclosan and triclosan-based skin care products in patients with chronic eczema. Dermatol. Ther. 21 S35-SS8.
Stoker, T.E., Gibson, E.K., Zorrilla, L.M., 2010. Triclosan exposure modulates estrogen-dependent responses in the female wistar rat. Toxicol. Sci. 117, 45–53.
Teitelbaum, S.L., Britton, J.A., Calafat, A.M., Ye, X., Silva, M.J., Reidy, J.A., et al., 2008. Temporal variability in urinary concentrations of phthalate metabolites, phytoestrogens and phenols among minority children in the United States. Environ. Res. 106, 257–269.
The National Academies, 2012. Project information: review of the IRIS process. Washington, D.C. [cited 2012 June 21]; Available from: http://www8.nationalacademies.org/cp/projectview.aspx?key=49458
U.S. Environmental Protection Agency, 1991. Guidelines for Developmental Toxicity Risk Assessment. U.S. Environmental Protection Agency, Washington, D.C. ([cited 2014 27 February]; Available from: http://cfpub.epa.gov/ncea/cfm/recordisplay.cfm?deid=23162#Download).
U.S. Environmental Protection Agency, 1996. Guidelines for Reproductive Toxicity Risk Assessment. U.S. Environmental Protection Agency, Washington, D.C. ([cited 2014 27 February]; Available from: http://www.epa.gov/raf/publications/pdfs/REPRO51.PDF).
U.S. EPA, 2015. Office of Pesticide Programs. Pesticide reregistration status|Pesticides|US EPA. U.S. EPA, Office of Pesticide Programs [updated 2015/01/19; cited 2015 July 23]; Available from: http://www.epa.gov/pesticides/reregistration/status.htm
U.S. Food and Drug Administration, 2013. In: Services UDoHaH (Ed.), Federal Register: Safety and Effectiveness of Consumer Antiseptics; Topical Antimicrobial Drug Products for Over-the-Counter Human Use; Proposed Amendment of the Tentative Final Monograph; Reopening of Administrative Record; Proposed Rule (Washington D.C.).
US Environmental Protection Agency, 2014. Draft development materials for the integrated risk information system (IRIS) toxicological review of inorganic arsenic: bimonthly public meeting: June 2014. In: EPA U (Ed.), Washington. U. S. Environmental Protection Agency, D.C. (78 FR 76443) Available at: https://federalregister.gov/a/2013-29814.
van der Worp, H.B., Macleod, M.R., 2011. Preclinical studies of human disease: time to take methodological quality seriously. J. Mol. Cell. Cardiol. 51, 449–450.
van der Worp, H.B., Sena, E.S., Donnan, G.A., Howells, D.W., Macleod, M.R., 2007. Hypothermia in animal models of acute ischaemic stroke: a systematic review and meta-analysis. Brain 130, 3063–3074.
Veldhoen, N., Skirrow, R.C., Osachoff, H., Wigmore, H., Clapson, D.J., Gunderson, M.P., et al., 2007. The bactericidal agent triclosan modulates thyroid hormone-associated gene expression and disrupts postembryonic anuran development. Aquat. Toxicol. 83, 84.
Vesterinen, H.M., Egan, K., Deister, A., Schlattmann, P., Macleod, M.R., Dirnagl, U., 2011. Systematic survey of the design, statistical analysis, and reporting of studies published in the 2008 volume of the Journal of Cerebral Blood Flow and Metabolism. J. Cereb. Blood Flow Metab. 31, 1064–1072.
Vesterinen, H.M., Johnson, P.I., Koustas, E., Lam, J., Sutton, P., Woodruff, T.J., 2013. In support of EHP's proposal to adopt the ARRIVE guidelines. Environ. Health Perspect. 121, A325-A.
Vesterinen, H., Johnson, P., Atchley, D., Sutton, P., Lam, J., Zlatnik, M., et al., 2014. The relationship between fetal growth and maternal glomerular filtration rate: a systematic review. J. Matern-Fetal Neonatal Med. 1–6.
Viswanathan, M., Ansari, M., Berkman, N., Chang, S., Hartling, L., McPheeters, L., et al., 2012. Assessing the Risk of Bias of Individual Studies in Systematic Reviews of Health Care Interventions. Report No.: AHRQ Publication No. 12-EHC047-EF.
Whaley, P.W., 2013. Systematic Review and the Future of Evidence in Chemicals Policy: Réseau Environnement Santé.
Wilding, B.C., Curtis, K., Welker-Hood, K., 2009. Hazardous Chemicals in Health Care: A Snapshot of Chemicals in Doctors and Nurses. Physicians for Social Responsibility, Washington, D.C.
Wise, A., Parham, F., Axelrad, D.A., Guyton, K.Z., Portier, C., Zeise, L., et al., 2012. Upstream adverse effects in risk assessment: a model of polychlorinated biphenyls, thyroid hormone disruption and neurological outcomes in humans. Environ. Res. 117, 90–99.
Witorsch, R.J., 2014. Critical analysis of endocrine disruptive activity of triclosan and its relevance to human exposure through the use of personal care products. Crit. Rev. Toxicol. 44, 535–555.
Wolff, M.S., Teitelbaum, S.L., Windham, G., Pinney, S.M., Britton, J.A., Chelimo, C., et al., 2007. Pilot study of urinary biomarkers of phytoestrogens, phthalates, and phenols in girls. Environ. Health Perspect. 115, 116–121.
Wolff, M.S., Teitelbaum, S.L., Pinney, S.M., Windham, G., Liao, L., Biro, F., et al., 2010. Investigation of relationships between urinary biomarkers of phytoestrogens, phthalates, and phenols and pubertal stages in girls. Environ. Health Perspect. 118, 1039–1046.
Woodruff, T.J., Sutton, P., 2011. The Navigation Guide work group. An evidence-based medicine methodology to bridge the gap between clinical and environmental health sciences. Health Aff. (Millwood) 30, 931–937.
Woodruff, T.J., Sutton, P., 2014. The Navigation Guide systematic review methodology: a rigorous and transparent method for translating environmental health science into better health outcomes. Environ. Health Perspect. 122, A283.
Woodruff, T.J., Zeise, L., Axelrad, D.A., Guyton, K.Z., Janssen, S., Miller, M., et al., 2008. Meeting report: moving upstream-evaluating adverse upstream end points for improved risk assessment and decision-making. Environ. Health Perspect. 116, 1568–1575.
Zorrilla, L.M., Gibson, E.K., Jeffay, S.C., Crofton, K.M., Setzer, W.R., Cooper, R.L., et al., 2009. The effects of triclosan on puberty and thyroid hormones in male Wistar rats. Toxicol. Sci. 107, 56–64.

Supplemental Material

Application of the Navigation Guide Methodology: Systematic Review of the Evidence for Developmental and Reproductive Toxicity of Triclosan

Paula I. Johnson, Erica Koustas, Hanna Vesterinen, Patrice Sutton, Dylan S. Atchley, Allegra N. Kim, Marlissa Campbell, James M. Donald, Saunak Sen, Lisa Bero, Lauren Zeise and Tracey J. Woodruff

Table of Contents Page

Table S1. Search terms 3

Toxicological websites searched 6

Data extraction fields 8

Instructions for making risk of bias determinations

A. Non-human experimental studies 9

B. Human studies 19

Table S2. Factors for evaluating the overall quality of a body of evidence 29

Table S3. Strength of evidence definitions for human evidence 30

Table S4. Strength of evidence definitions for non-human evidence 32

Figure S1. Integration step for human and non-human evidence 34

Tables S5-S10. Human study characteristics and risk of bias 35

Tables S11-S26. Rodent study characteristics and risk of bias 40

Figure S2. Triclosan administration and non-thyroxine hormones 58

Table S27: Summary of endpoints measured in human studies 59

Table S28: Summary of endpoints measured in non-human mammalian studies 60

EPA-HQ-2018-0008760045261

List of included mammalian (human and rodent) studies 63

List of identified *in vitro*, fish, amphibian and invertebrate studies 64

References for Supplemental Material 70

EPA-HQ-2018-0008760045262

Table S1. Search terms. We conducted a literature search on June 5th 2013 in PubMed, Web of Science, Biosis Previews, Embase and Toxline using the following terms.

PubMed
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 C12H7Cl3O2 [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]

Web of Science & Biosis Previews
TS=(Triclosan* OR 3380-34-5 OR Irgasan OR "Colgate Total" OR

EPA-HQ-2018-0008760045263

(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 C12H7Cl3O2 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”

Embase
(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 C12H7Cl3O2 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”).ti,ab.

Toxline
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 C12H7Cl3O2 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”

Toxicological websites searched

- ATSDR Interaction Profiles http://www.atsdr.cdc.gov/interactionprofiles/
- ATSDR Toxicological Profiles http://www.atsdr.cdc.gov/toxpro2.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 Toxnet http://toxnet.nlm.nih.gov/newtoxnet/dart.htm
- EPA Acute Exposure Guideline Levels http://www.epa.gov/oppt/aegl/chemlist.htm
- EPA IRIS e-docket 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/psl1.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.ilsi.org/
- IPCS INCHEM http://www.inchem.org/
- ITER http://iter.ctcnet.net/publicurl/pub_search_list.cfm
- NIOSHTIC 2 http://www2.cdc.gov/nioshtic 2/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
- FDA Substance Registration System: http://fdasis.nlm.nih.gov/

Data extraction fields

- Study level data
    - Authors (human, animal)
    - Year of publication (human, animal)
    - Country of origin (human, animal)
    - Type of publication (human, animal)
    - Funding bodies (human, animal)
    - Age (human)
    - Ethnicity (human)
- Experiment level data
    - Species, strain, source and breeding protocol (animal)
    - Compound purity, preparation, supplier and acquisition (donated, purchased or not reported; animal)
    - Control or reference group used (human, animal)
    - Life stage at exposure (prenatal or postnatal; animal, human)(human)
    - Exposure metric (human)
    - Covariates (human)
    - Cohort name (human)
- Outcome level data
    - Outcome measure and units (human, animal)
    - Sample size per group and unit of analysis (human, animal)
    - Measure of central tendency (mean, median or mode), odds ratio of incidence and appropriate measure of variation (confidence intervals, standard error (SE) or standard deviation (SD) or other statistic; human, animal)
    - Time of administration and outcome assessment (human, animal)
    - Exposure or dose, units and route of exposure (human, animal)
    - Sex (human, animal)
    - The reported statistical test and p value used (human, animal)

Instructions for Making Risk of Bias Determinations

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.

EPA-HQ-2018-0008760045269

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

EPA-HQ-2018-0008760045270

- 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:

EPA-HQ-2018-0008760045271

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

EPA-HQ-2018-0008760045272

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:

EPA-HQ-2018-0008760045274

- 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.

EPA-HQ-2018-0008760045275

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.

EPA-HQ-2018-0008760045276

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);

EPA-HQ-2018-0008760045277

- 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 under-estimation 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):

EPA-HQ-2018-0008760045279

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.

EPA-HQ-2018-0008760045280

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

EPA-HQ-2018-0008760045282

- 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:

EPA-HQ-2018-0008760045283

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):

EPA-HQ-2018-0008760045285

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

EPA-HQ-2018-0008760045287

- 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

Table S2. Factors for evaluating the overall quality of a body of evidence

Downgrading Factors[a]	Summary of criteria for downgrading
Risk of bias	Study limitations – a substantial risk of bias across body of evidence
Indirectness	Evidence was not directly comparable to the question of interest (i.e., population, exposure, comparator, outcome)
Inconsistency	Widely different estimates of effect in similar populations (heterogeneity or variability in results)
Imprecision	Studies had few participants and few events (wide confidence intervals as judged by reviewers)
Publication Bias	Studies missing from body of evidence, resulting in an over or underestimate of true effects from exposure

Upgrading Factors[b]	Summary of criteria for upgrading
Large magnitude of effect	Upgraded if modeling suggested confounding alone unlikely to explain associations with large effect estimate as judged by reviewers
Dose response	Upgraded if consistent relationship between dose and response in one or multiple studies, and/or dose response across studies
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

[a] We evaluated all bodies of evidence for potential downgrading factors.

[b] We evaluated only the human body of evidence for potential upgrading factors.

Table S3. 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.

[a]The 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 [1] except as noted.

[b]Language 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.[2]

Table S4. 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]

[a]The 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) [1]except as noted.

[b]IARC'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) [3]. 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) [3]. 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) [4].

[c]Language 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) [2].

[d]Based on minimum data requirements according to USEPA Guidelines for Reproductive Toxicity (U.S. Environmental Protection Agency 1996) [3].

Figure S1. Integration step for human and non-human evidence

Strength of Evidence in Human Systems	Strength of the Evidence in Non-human Systems: *Sufficient*	*Limited*	*Inadequate*	*Evidence of Lack of Toxicity*
Sufficient	Known to be Toxic to Human Reproduction			
Limited	Probably Toxic	Possibly Toxic		
Inadequate	Possibly Toxic	Not Classifiable		
Evidence of Lack of Toxicity	Not Classifiable			Probably Not Toxic

Study characteristics and risk of bias designations for human studies

Table S5. Characteristics of Koeppe et al. 2013 [5]

Study Element	Description
Design	Cross-sectional study
Methods	Analyzed NHANES data to assess the relationship between triclosan exposure and thyroid function.
Participants	N = 1831 participants (ages ≥ 12 years) from the 2007-2008 NHANES with urinary biomarker data.
Exposure	Urinary triclosan concentrations (µg/g creatinine).
Outcomes	Serum free T3 Serum total T3 Serum free T4 Serum total T4 Serum thyroglobulin Serum TSH

Table S6. Risk of bias summary for Koeppe et al. 2013 [5]

Bias domain	Authors' judgment	Support for judgment
Baseline differences	Low risk	The strategy for recruiting participants was consistent across study groups and there was no evidence of baseline differences between groups.
Sequence generation	N/A	
Allocation concealment	N/A	
Blinding	Low risk	NHANES sampling strategy was blinded to exposure.
Exposure assessment	Probably low risk	NHANES data has general quality controls expected from all labs, however the values were not reported, aside from LOD; Exposure assessment relied on single urine sample and the half-life of triclosan is only several hours, however there is some supporting evidence for relying on a spot sample, i.e. similar triclosan concentrations over time, assuming consistent exposure.
Outcome assessment	Probably low risk	NHANES data has general quality controls expected from contract labs; however, the values were not reported.
Confounding	Low risk	Important potential confounders included (age, BMI and urinary creatinine); did not find evidence to require others.
Incomplete outcome data	Low risk	The study did not have incomplete outcome data.
Selective outcome reporting	Low risk	The study is free of suggestion of selective outcome reporting.
Conflict of interest	Low risk	The authors report no conflict of interest, and associated funds and persons appear to be from government and/or academia only and not from entities with financial interest in the outcome.
Other sources of bias	Low risk	No other potential biases are suspected.

Table S7. Characteristics of Cullinan et al. 2012 [6]

Study Element	Description

EPA-HQ-2018-0008760045296

Design	Randomized clinical trial
Methods	Investigated relationship between using triclosan toothpaste and thyroid function with data from a subset of the Cardiovascular and Periodontal study (CAPS), a randomized, double blind, placebo controlled, clinical trial over 5 years.
Participants	N = 132 CAPS participants recruited from Prince Charles Hospital (Brisbane, Australia) between 2000 and 2010 with available year 1 and year 5 serum samples.
Exposure	No direct measure of triclosan exposure; Use of toothpaste containing 0.3% triclosan vs. placebo
Outcomes	Serum TSH Serum free T4 Serum free T3 Serum anti-TGab Serum anti-TPOab

Table S8. Risk of bias summary for Cullinan et al. 2012 [6]

Bias domain	Authors' judgment	Support for judgment
Sequence generation	Probably low risk	Specified that the trial was randomized but do not provide any details on the method of randomization; Satisfactory analysis of age, gender and smoking similarities between cases and controls.
Allocation concealment	Probably high risk	No evidence or mention of allocation concealment.
Baseline differences	N/A	
Blinding	Low risk	Patients received blinded triclosan or placebo toothpaste; lab personnel were blinded to exposure status.
Exposure assessment	High risk	Exposure is assumed to be dependent on toothpaste treatment alone, but there are many other possible sources of triclosan; no exposure biomarker assessed.
Outcome assessment	Probably low risk	These are standard measurements by medical lab, but no information or citation provided on method reliability or validity (QA/QC), only name of lab given.
Confounding	High risk	Did not take into account important potential confounders (age or BMI).
Incomplete outcome data	Low risk	The study did not have incomplete outcome data.
Selective outcome reporting	Low risk	The study is free of suggestion of selective outcome reporting.
Conflict of interest	High risk	The study was funded by Colgate Palmolive, maker of triclosan-containing toothpaste.
Other sources of bias	Low risk	No other potential biases are suspected.

Table S9. Characteristics of Allmyr et al. 2010 [7]

Study Element	Description
Design	Case-crossover study
Methods	Participants were instructed to brush their teeth with toothpaste containing 0.3% triclosan twice a day for 14 days. Triclosan concentrations and measures of thyroid function were evaluated on day 1 and day 15.
Participants	N = 12 healthy adults at Karolinska Institute in Huddinge, Sweden.
Exposure	Plasma triclosan concentrations (ng/g) on day 1 and day 15.
Outcomes	Plasma 4b-hydroxychloesterol

EPA-HQ-2018-0008760045298

	Plasma free T3 Plasma free T4 Plasma TSH

Table S10. Risk of bias summary for Allmyr et al. 2010 [7]

Bias domain	Authors' judgment	Support for judgment
Baseline differences	Low risk	Subjects acted as own controls.
Sequence generation	N/A	
Allocation concealment	N/A	
Blinding	Low risk	Author confirmed outcome assessors were completely blinded to exposure status.
Exposure assessment	Probably high risk	Unknown if 14 days enough time to see effect; experiment may not have adequately controlled for non-toothpaste triclosan exposure; cited paper states standard recovery was only 46% in plasma, although repeatability high.
Outcome assessment	Probably low risk	Author provided citations for methods (for 4β-hydroxycholesterol); for hormones, these are standard measurements, but only coefficient of variation is given, no other QA/QC or citation.
Confounding	Low risk	Subjects were own controls, and thus confounding should not be a risk; BMI could possibly still be associated with triclosan during the 14 day period of exposure (if stored in fat), however, the initial experimental period was a "washout" (no exposure) and based on 1st triclosan measurement, there were very low initial triclosan levels detected.
Incomplete outcome data	Low risk	The study did not have incomplete outcome data.
Selective outcome reporting	Low risk	The study is free of suggestion of selective outcome reporting.
Conflict of interest	Low risk	Because the funding statement was unclear, the author was contacted to confirm that there was no support from entities with financial interest.
Other sources of bias	Low risk	No other potential biases are suspected.

Study characteristics and risk of bias designations for rodent studies

Table S11. Characteristics of Paul et al. 2012 (study ID 180). [8]

Study Element	Description
Participants	Long-Evans rats Timed-pregnant GD1 animals obtained from supplier Total number of dams allocated: 155
Exposure	Experimental groups: • Prenatal time point: dams treated with triclosan via daily gavage from GD6 to GD20 (experimental block 3) • Postnatal time point: dams treated with triclosan via daily gavage from GD6 to PND21 (experimental blocks 1 and 2) Exposure groups: • 3 dose groups = 10, 30, 100, 300 mg/kg/day Prenatal time point: 11, 11, 11, 10 animals for 10, 30, 100, 300 mg/kg/day dose groups Postnatal time point: 12, 22, 22, 18 animals for 10, 30, 100, 300 mg/kg/day dose groups • 1 control group = corn oil Prenatal time point: 11 animals Postnatal time point: 21 animals
Outcomes	1. Total T4 in offspring – blood sample collected from each fetus and pooled into single serum sample for each litter at GD20. Number of litters analyzed: • 11, 11, 11, 11, 10 for control, 10, 30, 100, and 300 mg/kg/day, respectively (experimental block 3) 2. Total T4 in dams – serum sample collected from each dam at GD20 and PND22. Number of dams analyzed: • 11, 11, 11, 11, 10 for control, 10, 30, 100, and 300 mg/kg/day, respectively at GD20 (experimental block 3) • 21, 12, 22, 22, 18 for control, 10, 30, 100, and 300 mg/kg/day, respectively at PND22 (experimental blocks 1 and 2)
Notes	Study incorporates samples (experimental block 1) from a previous publication (Paul et al. 2010a). Offspring outcomes (total T4) reported in both papers do not overlap, as we were only able to include the GD20 time point from this paper, which is from experimental block 3. We were unable to extract data for offspring T4 levels for PND4, PND14, and PND21 time points. Data for these time points are presented for Paul et al. 2010a. Total T4 in dams incorporates data from experimental block 1 from Paul et al. 2010a and experimental block 2 from Paul et al. 2012, and is reported here (not reported for Paul et al. 2010a).

Table S12. Risk of bias summary for Paul et al. 2012 (study ID 180) [8]

Bias domain	Authors' judgment	Support for judgment
Sequence generation	Probably high risk	"Dams were semi-randomly assigned to treatment groups by counter-balancing body weights to obtain equivalent group body weight means." Unclear what "semi-randomly" means.
Allocation concealment	Probably high risk	No discussion of allocation concealment for outcomes assessed.
Blinding	Probably high risk	No discussion of blinding for outcomes assessed.
Incomplete outcome data	Probably high risk	Authors provide total number allocated, numbers analyzed, and information on missing animals. However, some animals were originally used in another study (Paul 2010a) and it is unclear which experiment has missing animals.
Selective reporting	Low risk	No evidence of selective outcome reporting for outcomes assessed.
Conflict of interest	Probably high risk	"The authors declare that there are no conflicts of interest." However, study was funded in part by parties that have potential financial interests in the study outcome (PhRMA and BASF).
Other bias	Low risk	No other potential biases are suspected.

Table S13. Characteristics of Stoker et al. 2010 (study ID 756). [9]

Study Element	Description
Participants	Pubertal assay cohort: Wistar rats Timed-pregnant GD14 animals obtained from supplier Total number of female weanlings allocated: 50 Uterotrophic assay cohort: Wistar rats Dams and PND6 pups obtained from supplier Total number of female weanlings allocated: 60
Exposure	Experimental groups: • Pubertal assay cohort: female weanlings treated with triclosan via daily gavage from PND22-42 • Uterotrophic assay cohort: female weanlings treated with triclosan via daily gavage from PND19-21 Exposure groups: Pubertal assay cohort: • 4 dose groups = 9.375, 37.5, 75, 150 mg/kg/day; 10 animals/dose • 1 control group = corn oil; 10 animals Uterotrophic assay cohort: • 9 dose groups = 1.18, 2.34, 4.69, 9.375, 18.75, 37.5, 75, 150, 300 mg/kg/day; 6 animals/dose • 1 control group = corn oil; 6 animals
Outcomes	Pubertal assay: 1. Total T4 – serum sample collected from each animal at PND42. Number of animals analyzed: unclear (assumed same as number allocated) 2. Free T4 – serum sample collected from each animal at PND 42. Number of animals analyzed: unclear (assumed same as number allocated) 3. TSH – serum sample collected from each animal at PND 42. Number of animals analyzed: unclear (assumed same as number allocated) 4. Body weight – individual weights collected at PND30 and at PND42. Number of animals analyzed: • 10 for each exposure group Uterotrophic assay: 1. Total T4 – serum sample collected from each animal at PND21. Number of animals analyzed: unclear (assumed same as number allocated) 2. Free T4 – serum sample collected from each animal at PND 21. Number of animals analyzed: unclear (assumed same as number allocated; note that results for 150 and 300 mg/kg/day not presented).
Notes	Developmental and reproductive outcomes not included in review: for pubertal assay, uterus weight (blotted and wet), pituitary weight, liver weight, ovary weight, first estrus and vaginal opening; for uterotrophic assay, uterus weight (blotted and weight), columnar differentiation of uterine luminal epithelium, and increased cell

	height of uterine glands

Table S14. Risk of bias summary for Stoker et al. 2010 (study ID 756) [9]

Pubertal assay outcomes		
Bias domain	Authors' judgment	Support for judgment
Sequence generation	Probably high risk	"The female offspring were weaned on PND 21, ranked by body weight, and placed into treatment groups such that the mean body weight ± SD for all groups were similar. In addition, littermates were equally distributed between the treatment groups with 10 females per group"
Allocation concealment	Probably high risk	No discussion of allocation concealment for outcomes assessed.
Blinding	Probably high risk	No discussion of blinding for outcomes assessed. Note that a subjective grading was used for histopathology.
Incomplete outcome data	Low risk[a]	Authors provide numbers allocated and numbers analyzed, and these numbers match, suggesting there may be no missing data.
	Probably low risk[b]	Authors report numbers allocated but not numbers analyzed; however, methods indicate that all animals weighed were evaluated for outcome, and the number of animals weighed match the number of animals allocated, suggesting there may be no missing data.
Selective reporting	Low risk	No evidence of selective outcome reporting for outcomes assessed.
Conflict of interest	Probably high risk	Authors do not provide a statement denying conflicts of interest or information on funding source.
Other bias	Low risk	No other potential biases are suspected.
Uterotrophic assay outcomes		
Bias domain	Authors' judgment	Support for judgment
Sequence generation	Probably low risk	"The immature rats were weighed, weight ranked, and assigned randomly to each of the experimental and control groups."
Allocation concealment	Probably high risk	No discussion of allocation concealment for outcomes assessed.
Blinding	Probably high risk	No discussion of blinding for outcomes assessed. Note that a subjective grading was used for histopathology.
Incomplete outcome data	Probably high risk	Authors do not provide explanation for how selected animals for allocation, numbers analyzed, or information on missing animals.
Selective reporting	Low risk	No evidence of selective outcome reporting for outcomes assessed.

Conflict of interest	Probably high risk	Authors do not provide a statement denying conflicts of interest or information on funding source.
Other bias	Low risk	No other potential biases are suspected.

[a]For body weight outcomes
[b]For hormone concentration outcomes

Table S15. Characteristics of Paul et al. 2010a (study ID 803). [10]

Study Element	Description
Participants	Long-Evans rats Timed-pregnant GD1 animals obtained from supplier Total number of dams allocated: 40
Exposure	Dams treated with triclosan via daily gavage from GD6-PND22; subset of offspring sacrificed at PND4, PND14, and PND21. Exposure groups: • 3 dose groups = 30, 100, 300 mg/kg/day; 10 animals/dose • 1 control group = corn oil; 10 animals
Outcomes	1. Total T4 – blood sample collected from culled pups (to normalize litters to 8) and pooled into single serum sample for each litter at PND4. Number of litters analyzed: • 9, 9, 8, 8 for control, 30, 100, 300 mg/kg/day groups, respectively 2. Total T4 – blood sample collected from one male and one female pup and pooled into single serum sample for each litter at PND14. Number of litters analyzed: • 10, 10, 8, 8 for control, 30, 100, 300 mg/kg/day groups, respectively 3. Total T4 – blood sample collected from one male and one female pup and pooled into single serum sample for each litter at PND21. Number of litters analyzed: • 10, 10, 9, 8 for control, 30, 100, 300 mg/kg/day groups, respectively
Notes	Total T4 in dams not reported here; data is incorporated into data presented for Paul et al. 2012. Developmental and reproductive outcomes not included in review: litter size, viability index, gestation length, sex ratio, and eye opening.

EPA-HQ-2018-0008760045305

Table S16. Risk of bias summary for Paul et al. 2010a (study ID 803) [10]

Bias domain	Authors' judgment	Support for judgment
Sequence generation	Probably low risk	"Dams were randomly assigned to treatment groups by counter-balancing body weights."
Allocation concealment	Probably high risk	No discussion of allocation concealment for outcomes assessed.
Blinding	Probably high risk	No discussion of blinding for outcomes assessed.
Incomplete outcome data	High risk	Authors provide numbers allocated and some information on missing animals; however, the numbers reportedly analyzed do not match the numbers allocated, after consideration of missing animals. Also, only a subset of animals assessed, and unclear if subset was selected randomly.
Selective reporting	Low risk	No evidence of selective outcome reporting for outcomes assessed.
Conflict of interest	High risk	Authors do not provide a statement denying conflicts of interest, and study supported in part by parties that have potential financial interest in study outcome (funded in part by PhRMA grant and compound donated by Ciba).
Other bias	Low risk	No other potential biases are suspected.

Table S17. Characteristics of Rodriguez and Sanchez et al. 2010 (study ID 804). [11]

Study Element	Description
Participants	Wistar rats In-house breeding protocol Total number of females allocated: 56
Exposure	Females treated with triclosan via drinking water from 8 days prior to mating-PND21. On PND21, half of the pubertal cohort continued triclosan exposure to PND50 (dosed), and half were administered the control (non-dosed). Exposure groups: • 3 dose groups = 1, 10, 50 mg/kg/day; 14 animals (dams)/dose • 1 control group = water plus vehicle (NaOH to neutralize); 14 animals (dams)
Outcomes	1. Total T4 – serum sample collected from dams at GD5, 10, 15, 20 and PND5, 10, 15, 20. Number of dams analyzed: • 8 for each exposure group 2. Total T3 – serum sample collected from dams at GD5, 10, 15, 20 and PND5, 10, 15, 20. Number of dams analyzed: • 8 for each exposure group 3. Body weight – female offspring weighed at time of observed vaginal opening (dosed). Number of animals analyzed: • 9 for each exposure group 4. Body weight – female offspring weighed at time of observed first estrus (dosed). Number of animals analyzed: • 9 for each exposure group 5. Body weight – female offspring weighed at time of observed vaginal opening (non-dosed). Number of animals analyzed: • 9 for each exposure group 6. Body weight – female offspring weighed at time of observed first estrus (non-dosed). Number of animals analyzed: • 9 for each exposure group
Notes	Developmental and reproductive outcomes not included in review: number of implantation sites, litter size, live births index, 6-day survival index, weaning index, and gestation length (dams); sex ratio, vaginal opening, and first estrus (offspring).

Table S18. Risk of bias summary for Rodriguez and Sanchez 2010 (study ID 804) [11]

Bias domain	Authors' judgment	Support for judgment
Sequence generation	Probably high risk	No discussion of sequence generation for outcomes assessed.
Allocation concealment	Probably high risk	No discussion of allocation concealment for outcomes assessed.
Blinding	Probably high risk	No discussion of blinding for outcomes assessed.
Incomplete outcome data	High risk	For weight at first estrus and vaginal opening, authors provide numbers allocated and numbers analyzed. A subset of animals were used to measure outcomes, and "remaining females of each litter (4 rats) were randomly divided into non-dosed and dosed groups"; however unclear how determined which dosed animals would be used for in utero and lactational exposure, and which would be used for in utero, lactational, up until puberty exposure; also unclear how 4 rats per 12 litters (48) matches up with numbers presented in tables 2 and 3 (n=9 animals per group). For total T3 and T4, authors provide numbers allocated and numbers analyzed; however, the numbers analyzed are not consistent between methods and results. A subset of animals were used to measure outcomes, but the authors provide no description for how the subset was selected.
Selective reporting	Low risk	No evidence of selective outcome reporting for outcomes assessed.
Conflict of interest	Probably low risk	Authors do not provide a statement denying conflicts of interest, but the study does not appear to have been supported by a financially interested party.
Other bias	Low risk	No other potential biases are suspected.

EPA-HQ-2018-0008760045308

Table S19. Characteristics of Paul et al. 2010b (study ID 835). [12]

Study Element	Description
Participants	Long-Evans rats 21-23 day old female rats obtained from supplier Total number of females allocated: 120
Exposure	Females (27-29 days of age) treated with triclosan via daily gavage for four consecutive days (split over 3 experimental blocks). Exposure groups: • 5 dose groups = 10, 30, 100, 300, 1000 mg/kg/day For T4 and body weight outcomes (all experimental blocks): 8, 24, 24, 24, 16 animals for 10, 30, 100, 300, 1000 mg/kg/day groups, respectively For T3 and TSH outcomes (experimental block 3 only): 8 animals per exposure group for 30, 100, 300, 1000 mg/kg/day groups, respectively (no animals for 10 mg/kg/day group). • 1 control group for each experimental group = corn oil; 24 animals for all experimental blocks and 8 animals for experimental block 3
Outcomes	1. Total T4 – serum sample collected from each animal following 4 days of exposure (PND31-33, depending on age when began treatment). Number of animals analyzed: • 24, 8, 24, 24, 24, 16 for control, 10, 30, 100, 300, 1000 mg/kg/day groups, respectively 2. Total T3 – serum sample collected from each animal following 4 days of exposure (PND31-33, depending on age when began treatment). Number of animals analyzed: • 8 for each exposure group 3. Total TSH – serum sample collected from each animal following 4 days of exposure (PND31-33, depending on age when began treatment). Number of animals analyzed: • 8 for each exposure group 4. Body weight gain – females weighed following 4 days of exposure (PND31-33, depending on age when began treatment). Number of animals analyzed: • 24, 8, 24, 24, 24, 16 for control, 10, 30, 100, 300, 1000 mg/kg/day groups, respectively
Notes	Study incorporates samples (experimental blocks 1 and 2) from a previous publication (Crofton et al. 2007) [13], which was originally identified in the search, but later excluded as Paul et al. 2010b presents all data reported in the Crofton et al. 2007 publication.

Table S20. Risk of bias summary for Paul et al. 2010b (study ID 835) [12]

Bias domain	Authors' judgment	Support for judgment
Sequence generation	Probably low risk	"Rats were randomly assigned to treatment groups to balance body weights at start of dosing."
Allocation concealment	Probably high risk	No discussion of allocation concealment for outcomes assessed.
Blinding	Probably high risk	No discussion of blinding for outcomes assessed.
Incomplete outcome data	Probably low risk	Study uses samples from a cohort of animals described in Crofton et al. 2007 [13]. Authors for both studies provide numbers allocated and numbers analyzed, and these numbers match, suggesting there may be no missing data.
Selective reporting	Low risk	No evidence of selective outcome reporting for outcomes assessed.
Conflict of interest	Probably low risk	Authors do not provide a statement denying conflicts of interest, but the study does not appear to have been supported by a financially interested party.
Other bias	Low risk	No other potential biases are suspected.

EPA-HQ-2018-0008760045310

Table S21. Characteristics of Zorrilla et al. 2009 (study ID 989). [14]

Study Element	Description
Participants	Wistar rats Timed-pregnant GD13 animals obtained from supplier; male offspring treated with triclosan. Total number of male offspring allocated: 71
Exposure	Males treated with triclosan via daily gavage from PND23-53. Exposure groups: • 5 dose groups = 3, 30, 100, 200, 300 mg/kg/day (experimental block 1: 3, 30, 300 mg/kg/day; experimental block 2: 100, 200 mg/kg/day) 10, 10, 8, 8, 10 animals for 10, 30, 100, 200, 300 mg/kg/day dose groups, respectively • 1 control group for each experimental block = corn oil; 10 for experimental block 1 and 15 for experimental block 2.
Outcomes	1. Total T4 – serum sample collected from each male at PND53. Number of animals analyzed: unclear (assumed same as number allocated) 2. Total T3 – serum sample collected from each male at PND53. Number of animals analyzed: unclear (assumed same as number allocated) 3. Total TSH – serum sample collected from each male at PND53. Number of animals analyzed: unclear (assumed same as number allocated) 4. Total testosterone – serum sample collected from each male at PND53. Number of animals analyzed: unclear (assumed same as number allocated) 5. Total androstenedione – serum sample collected from each male at PND 53. Number of animals analyzed: unclear (assumed same as number allocated) 6. Body weight – males weighed at PND44. Number of animals analyzed: unclear (assumed same as number allocated) 7. Body weight – males weighed at PND53. Number of animals analyzed: unclear (assumed same as number allocated)
Notes	Developmental and reproductive outcomes not included in review: adrenal weight, epididymus weight, kidney weight, levator anti-bulbocavernosus (LABC) muscle weight, liver weight, anterior pituitary weight, seminal vesicle weight, testes weight, ventral prostate weight, and preputial separation.

Table S22. Risk of bias summary for Zorilla et al. 2009 (study ID 989) [14]

Bias domain	Authors' judgment	Support for judgment
Sequence generation	Probably low risk	"On PND3, the litters were randomly standardized to ten pups each to maximize uniformity in growth rates…Pups were also randomly assigned so that treatment groups had similar body weight means and variances."
Allocation concealment	Probably high risk	No discussion of allocation concealment for outcomes assessed.
Blinding	Probably high risk	No discussion of blinding for outcomes assessed. Note that for histology outcomes: "The slides were then randomly mixed and evaluated blind for scoring based on the range established." Unclear if applies to other outcomes.
Incomplete outcome data	Probably high risk	Authors provide numbers allocated, but do not provide numbers analyzed, or information on missing animals.
Selective reporting	Low risk	No evidence of selective outcome reporting for outcomes assessed.
Conflict of interest	Probably low risk	Authors do not provide a statement denying conflicts of interest, but the study does not appear to have been supported by a financially interested party.
Other bias	Low risk	No other potential biases are suspected.

Table S23. Characteristics of Kumar et al. 2009 (study ID 1020). [15]

Study Element	Description
Participants	Wistar rats Male animals obtained from supplier. Total number of male offspring allocated: 32
Exposure	Males treated with triclosan via daily gavage from 10 weeks of age for 60 days. Exposure groups: • 3 dose groups = 5, 10, 20 mg/kg/day; 8 animals/dose • 1 control group = PBS; 8 animals
Outcomes	1. Testosterone – serum sample collected from each male in control and 20 mg/kg/day dose groups after 60 day treatment. Number of animals analyzed: • 8 for each exposure group (control and 20 mg/kg/day) 2. Androstenedione – serum sample collected from each male in control and 20 mg/kg/day dose groups after 60 day treatment. Number of animals analyzed:

	• 8 for each exposure group (control and 20 mg/kg/day) 3. Pregnenolone – serum sample collected from each male in control and 20 mg/kg/day dose groups after 60 day treatment. Number of animals analyzed: • 8 for each exposure group (control and 20 mg/kg/day) 5. Follicle stimulating hormone – serum sample collected from each male in control and 20 mg/kg/day dose groups after 60 day treatment. Number of animals analyzed: • 8 for each exposure group (control and 20 mg/kg/day) 2. Luteinizing hormone – serum sample collected from each male in control and 20 mg/kg/day dose groups after 60 day treatment. Number of animals analyzed: • 8 for each exposure group (control and 20 mg/kg/day)
Notes	Reproductive outcomes not included in review: epididymis weight, seminal vesicle weight, testis weight, vas deferens weight, and ventral prostate weight.

Table S24. Risk of bias summary for Kumar et al. 2009 (study ID 1020) [15]

Bias domain	Authors' judgment	Support for judgment
Sequence generation	Probably high risk	No discussion of sequence generation for outcomes assessed.
Allocation concealment	Probably high risk	No discussion of allocation concealment for outcomes assessed.
Blinding	Probably high risk	No discussion of blinding for outcomes assessed.
Incomplete outcome data	Probably low risk	Authors provide numbers allocated and numbers analyzed, and these numbers match, suggesting there may be no missing data.
Selective reporting	Low risk	No evidence of selective outcome reporting for outcomes assessed.
Conflict of interest	Low risk	The study does not appear to have been supported by a financially interested party. For conflicts of interest: "None declared".
Other bias	Low risk	No other potential biases are suspected.

Table S25. Characteristics of Axelstad et al. 2013 (study ID 9817) [16]

Study Element	Description
Participants	Prenatal exposure assay: Wistar rats Timed-pregnant GD3 animals obtained from supplier Total number of dams allocated: 40 Postnatal exposure assay: Wistar rats Timed-pregnant GD16 animals obtained from supplier Total number of dams allocated: 6
Exposure	Experimental groups: • Prenatal plus lactational exposure assay: dams treated with triclosan via daily gavage from GD7-PND16 • Postnatal exposure assay: offspring treated with triclosan via daily gavage from PND3-16. Exposure groups: Prenatal exposure assay: • 3 dose groups = 75, 150, 300 mg/kg/day; 10 animals/dose group • 1 control group = corn oil; 10 animals Postnatal exposure assay: • 2 dose groups (1 litter per dose group) = 50, 300 mg/kg/day; 8 animals/litter dose group • 1 control group = corn oil; 8 animals from 1 litter
Outcomes	Prenatal exposure assay: 1. Total T4 – serum sample collected from each dam at GD15 and PND16. Number of dams analyzed: • 9, 7, 8, 8 for control, 75, 150, 300 mg/kg/day groups, respectively, for each time point 2. Total T4 – blood samples collected from each fetus, by gender, and pooled into a single serum female sample and male sample for each litter at PND16. Number of litters analyzed: • 9, 7, 8, 8 for control, 75, 150, 300 mg/kg/day groups, respectively 3. Body weight – weight collected from offspring at PND1, 6, 13, and 16. Number of litters analyzed: • 9, 7, 8, 8 for control, 75, 150, 300 mg/kg/day groups, respectively, for each time point 4. Body weight gain – body weight gain monitored for dams for following periods: GD7-21, GD7-PND1, and PND1-16. Number of dams analyzed: • 9, 7, 8, 8 for control, 75, 150, 300 mg/kg/day groups, respectively, for each time period

	Postnatal exposure assay: 1. Total T4 – serum sample collected from each pup at PND16. Number of litters and offspring analyzed: • 1 litter (8 pups), 2 litters (16 pups), 2 litters (14 pups) for control, 50, 150 mg/kg/day groups, respectively 2. Body weight – average body weight collected from each litter at PND 6, 13, and 16. Number of litters and offspring analyzed: • 2 litters (1 after day 7) (16 pups; 8 after day 7), 2 litters (16 pups), 2 litters (14 pups) for control, 50, 150 mg/kg/day groups, respectively
Notes	Nursing litters were culled to normalize litter size to 8 but not cross-fostered. Two litters were assigned to each of 3 groups. One of 2 control dams rejected litter, leaving 1 genetically homogeneous control litter. The control litter was reported to have higher T4 levels compared to historical laboratory controls. Two pups from 1 high-dose litter died on PND6. Developmental and reproductive outcomes not included in review: litter size, perinatal death, perinatal loss, postimplantation litter loss, prostate weight (male offspring), thyroid weight (male offspring), gestation length, sex ratio, anogenital distance (males and females), and presence of nipples (males and females).

Table S26. Risk of bias summary for Axelstad et al. 2013 (study ID 9817) [16]

Bias domain	Authors' judgment	Support for judgment
Sequence generation	Probably high risk	No discussion of sequence generation for outcomes assessed.
Allocation concealment	Probably high risk	No discussion of allocation concealment for outcomes assessed.
Blinding	Probably high risk	No discussion of blinding for outcomes assessed. Note that "…histological evaluations…were performed by a pathologist blinded to treatment groups." Unclear if applies to other outcomes.
Incomplete outcome data	Probably low risk	Authors provide numbers allocated and numbers analyzed, and provide information on missing animals.
Selective reporting	Low risk	No evidence of selective outcome reporting for outcomes assessed.
Conflict of interest	Low risk	The study does not appear to have been supported by a financially interested party. "The authors declare that there are no conflicts of interest."
Other bias	Low risk	No other potential biases are suspected.

Figure S2. Triclosan administration and non-thyroxine hormones. Point estimates represent the concentration of each hormone (labeled on left) as a percentage of the control group for each dose. The vertical gray bar represents the line of no effect (the control group normalized to 100%); horizontal error bars represent 95% confidence intervals.

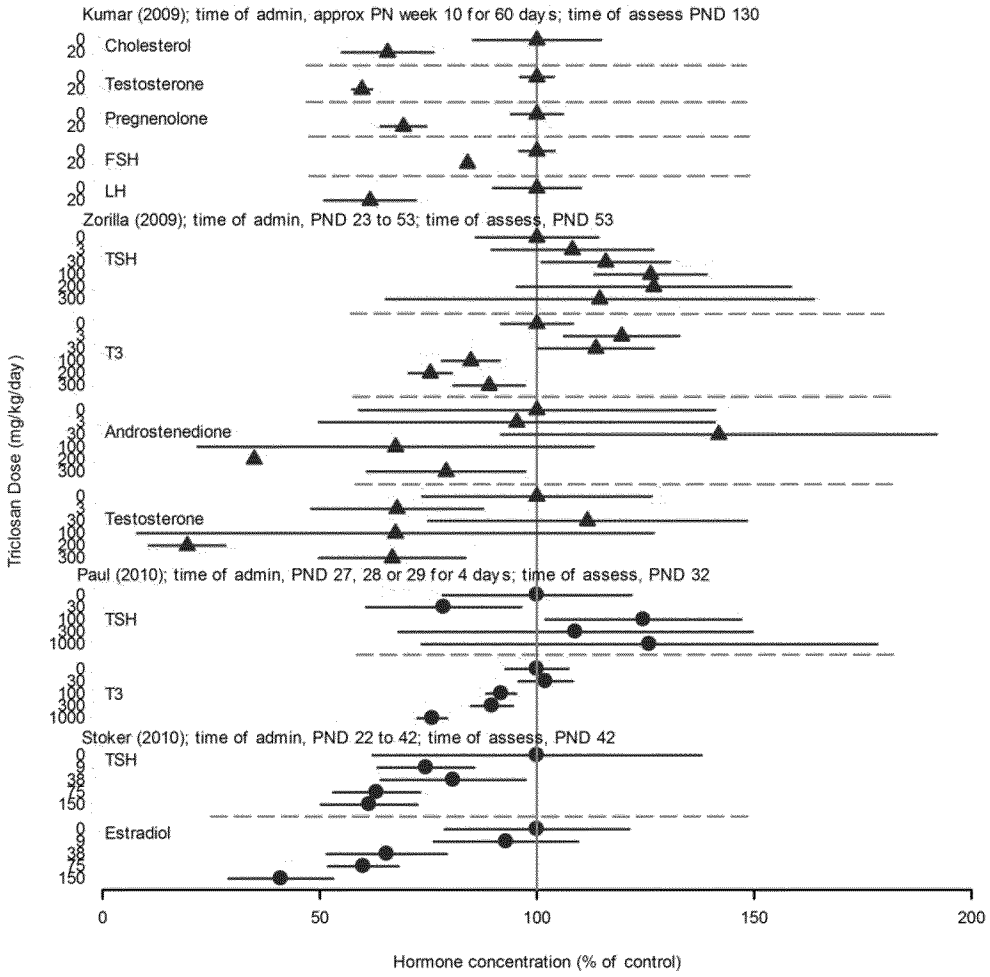

Table S27: Summary of endpoints measured in human studies[a]

Study	Thyroid/Hormone concentration[a]	Pubertal/ Reproductive
Koeppe et al. 2013	T3, T4, TSH, Thyroglobulin	N/A
Cullinan et al. 2012	T3, T4, TSH, Antithyroglobulin antibody, Antithyroid peroxidase antibody	N/A
Allmyr et al. 2010	T3, T4, TSH, 4b-Hydroxycholesterol	N/A
Buttke et al. 2012	N/A	Age at menarche
Wolff et al. 2010	N/A	Breast development, Pubic hair development
Chen et al. 2013	N/A	Idiopathic male infertility, Low semen volume, Low sperm concentration, Low total sperm per ejaculate

N/A = not applicable
(T4) = thyroxine
(T3) = triiodothyronine
(TSH) = thyroid-stimulating hormone
[a]In the present review, we used the Navigation Guide methodology to evaluate the quality and strength of the evidence for hormone concentration endpoints. The other endpoints are provided in this table for reference purposes. Not included are various endpoints measured at the cellular level.

Table S28: Summary of endpoints measured in non-human mammalian studies[a]

Study	Species	Hormone concentration[a]	Body size[b]	Reproductive	Organ weight[b]	Development	Histology/Morphology[b]
Louis et al. 2013	Rat	N/A	N/A	N/A	Uterus	N/A	N/A
Crawford and deCatanzaro 2012	Mouse	N/A	N/A	Implantation site number, litter size, gestation length	N/A	N/A	N/A
Paul et al. 2012	Rat	T4 (dams and offspring)	N/A	N/A	N/A	N/A	N/A
Stoker et al. 2010	Rat	T3, T4, TSH	Weight	N/A	Uterus, pituitary, liver, ovary	First estrus, vaginal opening	Columnar differentiation of luminal epithelium (uterus), increased cell height of glands (uterus)
Paul et al. 2010a	Rat	T4 (dams and offspring)	N/A	Litter size, viability index, gestation length	N/A	Sex ratio, eye opening	N/A
Rodriguez and Sanchez 2010	Rat	T3 (dams), T4 (dams)	Weight	Implantation site number, litter size, live birth index, 6-day survival index, weaning index, gestation length	N/A	Sex ratio, first estrus, vaginal opening	N/A
Paul et al. 2010b	Rat	T3, T4, TSH	Weight	N/A	Liver	N/A	N/A
Zorrilla et al. 2009	Rat	T3, T4, TSH, testosterone, androstenedione	Weight	N/A	Adrenal, epididymus, kidney, levator anti-bulbocavernosus (LABC) muscle, liver,	Preputial separation	N/A

					pituitary, seminal vesicle, testes, prostate		
Kumar et al. 2009	Rat	Testosterone, androstenedione, pregnenolone, follicle stimulating hormone, luteinizing hormone	N/A	N/A	Epididymus, seminal vesicle, testes, vas deferential, prostate	N/A	N/A
Russel and Montgomery 1980	Mouse	N/A	N/A	Litter size, number born, dams with litter, 12-day survival	N/A	N/A	N/A
Kawashima et al. 1987	Rat	N/A	Weight	Corpora lutea number, implantation site number, litter size, dams with live fetuses, dead implant number, implantation ratio	N/A	Number of fetuses with malformation, pyelectasis, skeletal malformation, skeletal variation, ossification state	N/A
Piekacz 1978	Rat and hamster	N/A	Length, weight (dams and offspring)	Litter size, resorptions	N/A	Sex ratio	N/A
Axelstad et al. 2013	Rat	T4 (dams and offspring)	Weight (dams and offspring)	Litter size, perinatal deaths, perinatal loss, post-implantation litter loss, gestation length	Prostate, thyroid (dams and offspring)	Sex ratio, anogenital distance, presence of nipples	N/A

N/A = not applicable
(T4) = thyroxine
(T3) = triiodothyronine
(TSH) = thyroid-stimulating hormone
[a]In the present review, we used the Navigation Guide methodology to evaluate the quality and strength of the evidence for hormone concentration endpoints. The other endpoints are provided in this table for reference purposes. Not included are various endpoints measured at the cellular level.
[b]Endpoint measured in offspring (gestational or early life exposures), unless otherwise stated.

List of included mammalian (human and rodent) studies

Human

1. Allmyr, M.; Panagiotidis, G.; Sparve, E.; Diczfalusy, U.; Sandborgh-Englund, G. Human exposure to triclosan via toothpaste does not change CYP3A4 activity or plasma concentrations of thyroid hormones. Basic & clinical pharmacology & toxicology. 105:339-344; 2009
2. Buttke, D.E.; Sircar, K.; Martin, C. Exposures to endocrine-disrupting chemicals and age of menarche in adolescent girls in NHANES (2003-2008). Environmental health perspectives. 120:1613-1618; 2012
3. Chen, M.; Tang, R.; Fu, G.; Xu, B.; Zhu, P.; Qiao, S.; Chen, X.; Xu, B.; Qin, Y.; Lu, C.; Hang, B.; Xia, Y.; Wang, X. Association of exposure to phenols and idiopathic male infertility. Journal of hazardous materials. 250-251:115-121; 2013
4. Cullinan, M.P.; Palmer, J.E.; Carle, A.D.; West, M.J.; Seymour, G.J. Long term use of triclosan toothpaste and thyroid function. The Science of the total environment. 416:75-79; 2012
5. Koeppe, E.S.; Ferguson, K.K.; Colacino, J.A.; Meeker, J.D. Relationship between urinary triclosan and paraben concentrations and serum thyroid measures in NHANES 2007-2008. The Science of the total environment. 445-446:299-305; 2013
6. Wolff, M.S.; Teitelbaum, S.L.; Pinney, S.M.; Windham, G.; Liao, L.; Biro, F.; Kushi, L.H.; Erdmann, C.; Hiatt, R.A.; Rybak, M.E.; Calafat, A.M.; Breast, C.; Environment Research, C. Investigation of relationships between urinary biomarkers of phytoestrogens, phthalates, and phenols and pubertal stages in girls. Environmental health perspectives. 118:1039-1046; 2010

Rodent

1. Axelstad, M.; Boberg, J.; Vinggaard, A.M.; Christiansen, S.; Hass, U. Triclosan exposure reduces thyroxine levels in pregnant and lactating rat dams and in directly exposed offspring. Food Chem Toxicol. 59:534-540; 2013
2. Crawford, B.R.; Decatanzaro, D. Disruption of blastocyst implantation by triclosan in mice: impacts of repeated and acute doses and combination with bisphenol-A. Reproductive toxicology. 34:607-613; 2012
3. Kawashima, K.; Nakaura, S.; Yamaguchi, M.; Tanaka, S.; Takanaka, A. [Effects of triclosan on fetal developments of rats]. Eisei Shikenjo Hokoku:28-32; 1987
4. Kumar, V.; Chakraborty, A.; Kural, M.R.; Roy, P. Alteration of testicular steroidogenesis and histopathology of reproductive system in male rats treated with triclosan. Reproductive toxicology. 27:177-185; 2009
5. Louis, G.W.; Hallinger, D.R.; Stoker, T.E. The effect of triclosan on the uterotrophic response to extended doses of ethinyl estradiol in the weanling rat. Reproductive toxicology. 36:71-77; 2013
6. Paul, K.B.; Hedge, J.M.; Bansal, R.; Zoeller, R.T.; Peter, R.; DeVito, M.J.; Crofton, K.M. Developmental triclosan exposure decreases maternal, fetal, and

early neonatal thyroxine: a dynamic and kinetic evaluation of a putative mode-of-action. Toxicology. 300:31-45; 2012
7. Paul, K.B.; Hedge, J.M.; Devito, M.J.; Crofton, K.M. Developmental triclosan exposure decreases maternal and neonatal thyroxine in rats. Environmental toxicology and chemistry / SETAC. 29:2840-2844; 2010a
8. Paul, K.B.; Hedge, J.M.; DeVito, M.J.; Crofton, K.M. Short-term exposure to triclosan decreases thyroxine in vivo via upregulation of hepatic catabolism in Young Long-Evans rats. Toxicological sciences : an official journal of the Society of Toxicology. 113:367-379; 2010b
9. Piekacz, H. [Effects of various preservative agents on the course of pregnancy and fetal development in experimental animals. Toxicological characteristics]. Rocz Panstw Zakl Hig. 29:469-481; 1978
10. Rodriguez, P.E.; Sanchez, M.S. Maternal exposure to triclosan impairs thyroid homeostasis and female pubertal development in Wistar rat offspring. Journal of toxicology and environmental health Part A. 73:1678-1688; 2010
11. Russell, L.B.; Montgomery, C.S. Use of the mouse spot test to investigate the mutagenic potential of triclosan (Irgasan DP300). Mutation research. 79:7-12; 1980
12. Stoker, T.E.; Gibson, E.K.; Zorrilla, L.M. Triclosan exposure modulates estrogen-dependent responses in the female wistar rat. Toxicological sciences : an official journal of the Society of Toxicology. 117:45-53; 2010
13. Zorrilla, L.M.; Gibson, E.K.; Jeffay, S.C.; Crofton, K.M.; Setzer, W.R.; Cooper, R.L.; Stoker, T.E. The effects of triclosan on puberty and thyroid hormones in male Wistar rats. Toxicological sciences : an official journal of the Society of Toxicology. 107:56-64; 2009

List of identified *in vitro*, fish, amphibian and invertebrate studies

In Vitro

1. Ahn, K.C.; Zhao, B.; Chen, J.; Cherednichenko, G.; Sanmarti, E.; Denison, M.S.; Lasley, B.; Pessah, I.N.; Kultz, D.; Chang, D.P.; Gee, S.J.; Hammock, B.D. In vitro biologic activities of the antimicrobials triclocarban, its analogs, and triclosan in bioassay screens: receptor-based bioassay screens. Environmental health perspectives. 116:1203-1210; 2008
2. Butt, C.M.; Wang, D.; Stapleton, H.M. Halogenated phenolic contaminants inhibit the in vitro activity of the thyroid-regulating deiodinases in human liver. Toxicological sciences : an official journal of the Society of Toxicology. 124:339-347; 2011
3. Chen, J.; Ahn, K.C.; Gee, N.A.; Gee, S.J.; Hammock, B.D.; Lasley, B.L. Antiandrogenic properties of parabens and other phenolic containing small molecules in personal care products. Toxicology and applied pharmacology. 221:278-284; 2007
4. Christen, V.; Crettaz, P.; Oberli-Schrammli, A.; Fent, K. Some flame retardants and the antimicrobials triclosan and triclocarban enhance the androgenic activity in vitro. Chemosphere. 81:1245-1252; 2010

5. Forgacs, A.L.; Ding, Q.; Jaremba, R.G.; Huhtaniemi, I.T.; Rahman, N.A.; Zacharewski, T.R. BLTK1 murine Leydig cells: a novel steroidogenic model for evaluating the effects of reproductive and developmental toxicants. Toxicological sciences : an official journal of the Society of Toxicology. 127:391-402; 2012
6. Gee, R.H.; Charles, A.; Taylor, N.; Darbre, P.D. Oestrogenic and androgenic activity of triclosan in breast cancer cells. Journal of applied toxicology : JAT. 28:78-91; 2008
7. Gomez, J.; Macina, O.T.; Mattison, D.R.; Zhang, Y.P.; Klopman, G.; Rosenkranz, H.S. Structural determinants of developmental toxicity in hamsters. Teratology. 60:190-205; 1999
8. Guo, L.W.; Wu, Q.; Green, B.; Nolen, G.; Shi, L.; Losurdo, J.; Deng, H.; Bauer, S.; Fang, J.L.; Ning, B. Cytotoxicity and inhibitory effects of low-concentration triclosan on adipogenic differentiation of human mesenchymal stem cells. Toxicology and applied pharmacology. 262:117-123; 2012
9. Henry, N.D.; Fair, P.A. Comparison of in vitro cytotoxicity, estrogenicity and anti-estrogenicity of triclosan, perfluorooctane sulfonate and perfluorooctanoic acid. Journal of applied toxicology : JAT. 33:265-272; 2013
10. Hinther, A.; Bromba, C.M.; Wulff, J.E.; Helbing, C.C. Effects of triclocarban, triclosan, and methyl triclosan on thyroid hormone action and stress in frog and mammalian culture systems. Environmental science & technology. 45:5395-5402; 2011
11. Honkisz, E.; Zieba-Przybylska, D.; Wojtowicz, A.K. The effect of triclosan on hormone secretion and viability of human choriocarcinoma JEG-3 cells. Reproductive toxicology. 34:385-392; 2012
12. Jacobs, M.N.; Nolan, G.T.; Hood, S.R. Lignans, bacteriocides and organochlorine compounds activate the human pregnane X receptor (PXR). Toxicology and applied pharmacology. 209:123-133; 2005
13. James, M.O.; Li, W.; Summerlot, D.P.; Rowland-Faux, L.; Wood, C.E. Triclosan is a potent inhibitor of estradiol and estrone sulfonation in sheep placenta. Environment international. 36:942-949; 2010
14. Jung, E.M.; An, B.S.; Choi, K.C.; Jeung, E.B. Potential estrogenic activity of triclosan in the uterus of immature rats and rat pituitary GH3 cells. Toxicology letters. 208:142-148; 2012
15. Kumar, V.; Balomajumder, C.; Roy, P. Disruption of LH-induced testosterone biosynthesis in testicular Leydig cells by triclosan: probable mechanism of action. Toxicology. 250:124-131; 2008
16. Svobodova, K.; Plackova, M.; Novotna, V.; Cajthaml, T. Estrogenic and androgenic activity of PCBs, their chlorinated metabolites and other endocrine disruptors estimated with two in vitro yeast assays. The Science of the total environment. 407:5921-5925; 2009

Fish

1. Foran, C.M.; Bennett, E.R.; Benson, W.H. Developmental evaluation of a potential non-steroidal estrogen: triclosan. Marine environmental research. 50:153-156; 2000

2. Fritsch, E.B.; Connon, R.E.; Werner, I.; Davies, R.E.; Beggel, S.; Feng, W.; Pessah, I.N. Triclosan impairs swimming behavior and alters expression of excitation-contraction coupling proteins in fathead minnow (Pimephales promelas). Environmental science & technology. 47:2008-2017; 2013
3. Houtman, C.J.; Booij, P.; van der Valk, K.M.; van Bodegom, P.M.; van den Ende, F.; Gerritsen, A.A.; Lamoree, M.H.; Legler, J.; Brouwer, A. Biomonitoring of estrogenic exposure and identification of responsible compounds in bream from Dutch surface waters. Environmental toxicology and chemistry / SETAC. 26:898-907; 2007
4. Ishibashi, H.; Matsumura, N.; Hirano, M.; Matsuoka, M.; Shiratsuchi, H.; Ishibashi, Y.; Takao, Y.; Arizono, K. Effects of triclosan on the early life stages and reproduction of medaka Oryzias latipes and induction of hepatic vitellogenin. Aquatic toxicology. 67:167-179; 2004
5. Kim, J.W.; Ishibashi, H.; Yamauchi, R.; Ichikawa, N.; Takao, Y.; Hirano, M.; Koga, M.; Arizono, K. Acute toxicity of pharmaceutical and personal care products on freshwater crustacean (Thamnocephalus platyurus) and fish (Oryzias latipes). The Journal of toxicological sciences. 34:227-232; 2009
6. Liang, X.; Nie, X.; Ying, G.; An, T.; Li, K. Assessment of toxic effects of triclosan on the swordtail fish (Xiphophorus helleri) by a multi-biomarker approach. Chemosphere. 90:1281-1288; 2013
7. Muth-Kohne, E.; Wichmann, A.; Delov, V.; Fenske, M. The classification of motor neuron defects in the zebrafish embryo toxicity test (ZFET) as an animal alternative approach to assess developmental neurotoxicity. Neurotoxicology and teratology. 34:413-424; 2012
8. Nassef, M.; Kim, S.G.; Seki, M.; Kang, I.J.; Hano, T.; Shimasaki, Y.; Oshima, Y. In ovo nanoinjection of triclosan, diclofenac and carbamazepine affects embryonic development of medaka fish (Oryzias latipes). Chemosphere. 79:966-973; 2010
9. Oliveira, R.; Domingues, I.; Koppe Grisolia, C.; Soares, A.M. Effects of triclosan on zebrafish early-life stages and adults. Environmental science and pollution research international. 16:679-688; 2009
10. Parrott, J.L.; Bennie, D.T. Life-cycle exposure of fathead minnows to a mixture of six common pharmaceuticals and triclosan. Journal of toxicology and environmental health Part A. 72:633-641; 2009
11. Pinto, P.I.S.; Guerreiro, E.M.; Power, D.M. Triclosan interferes with the thyroid axis in the zebrafish (Danio rerio). Toxicology Research. 2:60; 2013
12. Raut, S.A.; Angus, R.A. Triclosan has endocrine-disrupting effects in male western mosquitofish, Gambusia affinis. Environmental toxicology and chemistry / SETAC. 29:1287-1291; 2010
13. Rostkowski, P.; Horwood, J.; Shears, J.A.; Lange, A.; Oladapo, F.O.; Besselink, H.T.; Tyler, C.R.; Hill, E.M. Bioassay-directed identification of novel antiandrogenic compounds in bile of fish exposed to wastewater effluents. Environmental science & technology. 45:10660-10667; 2011
14. Schultz, M.M.; Bartell, S.E.; Schoenfuss, H.L. Effects of triclosan and triclocarban, two ubiquitous environmental contaminants, on anatomy,

physiology, and behavior of the fathead minnow (Pimephales promelas). Archives of environmental contamination and toxicology. 63:114-124; 2012

Amphibian

1. Fort, D.J.; Mathis, M.B.; Hanson, W.; Fort, C.E.; Navarro, L.T.; Peter, R.; Buche, C.; Unger, S.; Pawlowski, S.; Plautz, J.R. Triclosan and thyroid-mediated metamorphosis in anurans: differentiating growth effects from thyroid-driven metamorphosis in Xenopus laevis. Toxicological sciences : an official journal of the Society of Toxicology. 121:292-302; 2011
2. Fort, D.J.; Rogers, R.L.; Gorsuch, J.W.; Navarro, L.T.; Peter, R.; Plautz, J.R. Triclosan and anuran metamorphosis: no effect on thyroid-mediated metamorphosis in Xenopus laevis. Toxicological sciences : an official journal of the Society of Toxicology. 113:392-400; 2010
3. Harada, A.; Komori, K.; Nakada, N.; Kitamura, K.; Suzuki, Y. Biological effects of PPCPs on aquatic lives and evaluation of river waters affected by different wastewater treatment levels. Water science and technology : a journal of the International Association on Water Pollution Research. 58:1541-1546; 2008
4. Marlatt, V.L.; Veldhoen, N.; Lo, B.P.; Bakker, D.; Rehaume, V.; Vallee, K.; Haberl, M.; Shang, D.; van Aggelen, G.C.; Skirrow, R.C.; Elphick, J.R.; Helbing, C.C. Triclosan exposure alters postembryonic development in a Pacific tree frog (Pseudacris regilla) Amphibian Metamorphosis Assay (TREEMA). Aquatic toxicology. 126:85-94; 2013
5. Matsumura, N.; Ishibashi, H.; Hirano, M.; Nagao, Y.; Watanabe, N.; Shiratsuchi, H.; Kai, T.; Nishimura, T.; Kashiwagi, A.; Arizono, K. Effects of nonylphenol and triclosan on production of plasma vitellogenin and testosterone in male South African clawed frogs (Xenopus laevis). Biological & pharmaceutical bulletin. 28:1748-1751; 2005
6. Palenske, N.M.; Nallani, G.C.; Dzialowski, E.M. Physiological effects and bioconcentration of triclosan on amphibian larvae. Comparative biochemistry and physiology Toxicology & pharmacology : CBP. 152:232-240; 2010
7. Smith, G.R.; Burgett, A.A. Effects of three organic wastewater contaminants on American toad, Bufo americanus, tadpoles. Ecotoxicology. 14:477-482; 2005
8. Veldhoen, N.; Skirrow, R.C.; Osachoff, H.; Wigmore, H.; Clapson, D.J.; Gunderson, M.P.; Van Aggelen, G.; Helbing, C.C. The bactericidal agent triclosan modulates thyroid hormone-associated gene expression and disrupts postembryonic anuran development. Aquatic toxicology. 80:217-227; 2006

Invertebrate

1. Amorim, M.J.; Oliveira, E.; Soares, A.M.; Scott-Fordsmand, J.J. Predicted no effect concentration (PNEC) for triclosan to terrestrial species (invertebrates and plants). Environment international. 36:338-343; 2010
2. Anselmo, H.M.; Koerting, L.; Devito, S.; van den Berg, J.H.; Dubbeldam, M.; Kwadijk, C.; Murk, A.J. Early life developmental effects of marine persistent

organic pollutants on the sea urchin Psammechinus miliaris. Ecotoxicol Environ Saf. 74:2182-2192; 2011
3. Anselmo, H.M.; van den Berg, J.H.; Rietjens, I.M.; Murk, A.J. Inhibition of cellular efflux pumps involved in multi xenobiotic resistance (MXR) in echinoid larvae as a possible mode of action for increased ecotoxicological risk of mixtures. Ecotoxicology. 21:2276-2287; 2012
4. Binelli, A.; Parolini, M.; Pedriali, A.; Provini, A. Antioxidant Activity in the Zebra Mussel (Dreissena polymorpha) in Response to Triclosan Exposure. Water, Air, and Soil Pollution. 217:421-430; 2011
5. Borgmann, U.; Bennie, D.T.; Ball, A.L.; Palabrica, V. Effect of a mixture of seven pharmaceuticals on Hyalella azteca over multiple generations. Chemosphere. 66:1278-1283; 2007
6. Brown, J.; Bernot, M.J.; Bernot, R.J. The influence of TCS on the growth and behavior of the freshwater snail, Physa acuta. J Environ Sci Health A Tox Hazard Subst Environ Eng. 47:1626-1630; 2012
7. Cortez, F.S.; Seabra Pereira, C.D.; Santos, A.R.; Cesar, A.; Choueri, R.B.; Martini Gde, A.; Bohrer-Morel, M.B. Biological effects of environmentally relevant concentrations of the pharmaceutical Triclosan in the marine mussel Perna perna (Linnaeus, 1758). Environ Pollut. 168:145-150; 2012
8. Delorenzo, M.E.; Keller, J.M.; Arthur, C.D.; Finnegan, M.C.; Harper, H.E.; Winder, V.L.; Zdankiewicz, D.L. Toxicity of the antimicrobial compound triclosan and formation of the metabolite methyl-triclosan in estuarine systems. Environ Toxicol. 23:224-232; 2008
9. Dussault, E.B.; Balakrishnan, V.K.; Sverko, E.; Solomon, K.R.; Sibley, P.K. Toxicity of human pharmaceuticals and personal care products to benthic invertebrates. Environmental toxicology and chemistry / SETAC. 27:425-432; 2008
10. Fraker, S.L.; Smith, G.R. Direct and interactive effects of ecologically relevant concentrations of organic wastewater contaminants on Rana pipiens tadpoles. Environ Toxicol. 19:250-256; 2004
11. Gaume, B.; Bourgougnon, N.; Auzoux-Bordenave, S.; Roig, B.; Le Bot, B.; Bedoux, G. In vitro effects of triclosan and methyl-triclosan on the marine gastropod Haliotis tuberculata. Comparative biochemistry and physiology Toxicology & pharmacology : CBP. 156:87-94; 2012
12. Karimullina, E.; Li, Y.; Ginjupalli, G.K.; Baldwin, W.S. Daphnia HR96 is a promiscuous xenobiotic and endobiotic nuclear receptor. Aquatic toxicology. 116-117:69-78; 2012
13. Lin, D.; Xie, X.; Zhou, Q.; Liu, Y. Biochemical and genotoxic effect of triclosan on earthworms (Eisenia fetida) using contact and soil tests. Environ Toxicol. 27:385-392; 2012
14. Matozzo, V.; Costa Devoti, A.; Marin, M.G. Immunotoxic effects of triclosan in the clam Ruditapes philippinarum. Ecotoxicology. 21:66-74; 2012a
15. Matozzo, V.; Formenti, A.; Donadello, G.; Marin, M.G. A multi-biomarker approach to assess effects of Triclosan in the clam Ruditapes philippinarum. Marine environmental research. 74:40-46; 2012b

16. Morin, S.; Proia, L.; Ricart, M.; Bonnineau, C.; Geiszinger, A.; Ricciardi, F.; Guasch, H.; Romani, A.M.; Sabater, S. Effects of a bactericide on the structure and survival of benthic diatom communities. Vie et Milieu. 60:109-116; 2010
17. Orvos, D.R.; Versteeg, D.J.; Inauen, J.; Capdevielle, M.; Rothenstein, A.; Cunningham, V. Aquatic toxicity of triclosan. Environmental toxicology and chemistry / SETAC. 21:1338-1349; 2002
18. Pannu, M.W.; O'Connor, G.A.; Toor, G.S. Toxicity and bioaccumulation of biosolids-borne triclosan in terrestrial organisms. Environmental toxicology and chemistry / SETAC. 31:646-653; 2012
19. Park, H.G.; Yeo, M.K. The toxicity of triclosan, bisphenol A, bisphenol A diglycidyl ether to the regeneration of cnidarian, Hydra magnipapillata. Molecular & Cellular Toxicology. 8:209-216; 2012
20. Perron, M.M.; Ho, K.T.; Cantwell, M.G.; Burgess, R.M.; Pelletier, M.C. Effects of triclosan on marine benthic and epibenthic organisms. Environmental toxicology and chemistry / SETAC. 31:1861-1866; 2012
21. Riva, C.; Cristoni, S.; Binelli, A. Effects of triclosan in the freshwater mussel Dreissena polymorpha: a proteomic investigation. Aquatic toxicology. 118-119:62-71; 2012
22. Schnug, L.; Jakob, L.; Hartnik, T. The toxicity of a ternary biocide mixture to two consecutive earthworm (Eisenia fetida) Generations. Environmental toxicology and chemistry / SETAC. 32:937-947; 2013
23. Tamura, I.; Kagota, K.; Yasuda, Y.; Yoneda, S.; Morita, J.; Nakada, N.; Kameda, Y.; Kimura, K.; Tatarazako, N.; Yamamoto, H. Ecotoxicity and screening level ecotoxicological risk assessment of five antimicrobial agents: triclosan, triclocarban, resorcinol, phenoxyethanol and p-thymol. Journal of applied toxicology : JAT. 33:1222-1229; 2013
24. Tatarazako, N.; Ishibashi, H.; Teshima, K.; Kishi, K.; Arizono, K. Effects of triclosan on various aquatic organisms. Environ Sci. 11:133-140; 2004

EPA-HQ-2018-0008760045329

References for Supplemental Material

1. International Agency for Research on Cancer. IARC Monographs on the Evaluation of Carcinogenic Risks to Humans: Preamble (amended January 2006). Lyon, France: World Health Organization; 2006.
2. Sawaya GF, Guirguis-Blake J, LeFevre M, Harris R, Petitti D. Update on Methods: Estimating Certainty and Magnitude of Net Benefit. U.S. Preventive Services Task Force,; 2007; Available from: http://www.uspreventiveservicestaskforce.org/uspstf07/methods/benefit.htm#copyright
3. U.S. Environmental Protection Agency. Guidelines for reproductive toxicity risk assessment. Washington, DC: U.S. Environmental Protection Agency; 1996 [cited 2014 27 February]; Available from: http://www.epa.gov/raf/publications/pdfs/REPRO51.PDF.
4. U.S. Environmental Protection Agency. Guidelines for Developmental Toxicity Risk Assessment. Washington, DC: U.S. Environmental Protection Agency; 1991 [cited 2014 27 February]; Available from: http://cfpub.epa.gov/ncea/cfm/recordisplay.cfm?deid=23162#Download.
5. Koeppe ES, Ferguson KK, Colacino JA, Meeker JD. Relationship between urinary triclosan and paraben concentrations and serum thyroid measures in NHANES 2007-2008. Sci Total Environ. 2013; 445-446:299-305.
6. Cullinan MP, Palmer JE, Carle AD, West MJ, Seymour GJ. Long term use of triclosan toothpaste and thyroid function. Sci Total Environ. 2012; 416:75-9.
7. Allmyr M, Panagiotidis G, Sparve E, Diczfalusy U, Sandborgh-Englund G. Human exposure to triclosan via toothpaste does not change CYP3A4 activity or plasma concentrations of thyroid hormones. Basic & clinical pharmacology & toxicology. 2009; 105:339-44.
8. Paul KB, Hedge JM, Bansal R, Zoeller RT, Peter R, DeVito MJ, et al. Developmental triclosan exposure decreases maternal, fetal, and early neonatal thyroxine: a dynamic and kinetic evaluation of a putative mode-of-action. Toxicology. 2012; 300:31-45.
9. Stoker TE, Gibson EK, Zorrilla LM. Triclosan exposure modulates estrogen-dependent responses in the female wistar rat. Toxicol Sci. 2010; 117:45-53.
10. Paul KB, Hedge JM, DeVito MJ, Crofton KM. Developmental triclosan exposure decreases maternal and neonatal thyroxine in rats. Environ Toxicol Chem. 2010; 29:2840-4.
11. Rodriguez PE, Sanchez MS. Maternal exposure to triclosan impairs thyroid homeostasis and female pubertal development in Wistar rat offspring. J Toxicol Environ Health A. 2010; 73:1678-88.
12. Paul KB, Hedge JM, DeVito MJ, Crofton K. Short-term Exposure to Triclosan Decreases Thyroxine In Vivo via Upregulation of Hepatic Catabolism in Young Long-Evans Rats. Toxicol Sci. 2010; 113:367-79.
13. Crofton KM, Paul KB, Devito MJ, Hedge JM. Short-term in vivo exposure to the water contaminant triclosan: Evidence for disruption of thyroxine. Environ Toxicol Pharmacol. 2007; 24:194-7.
14. Zorrilla LM, Gibson EK, Jeffay SC, Crofton KM, Setzer WR, Cooper RL, et al. The effects of triclosan on puberty and thyroid hormones in male Wistar rats. Toxicol Sci. 2009; 107:56-64.
15. Kumar V, Chakraborty A, Kural MR, Roy P. Alteration of testicular steroidogenesis and histopathology of reproductive system in male rats treated with triclosan. Reprod Toxicol. 2009; 27:177-85.
16. Axelstad M, Boberg J, Vinggaard AM, Christiansen S, Hass U. Triclosan exposure reduces thyroxine levels in pregnant and lactating rat dams and in directly exposed offspring. Food Chem Toxicol. 2013; 59:534-40.

EPA-HQ-2018-0008760045330