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′-hydroxydiphenyl ether, 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 < 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.

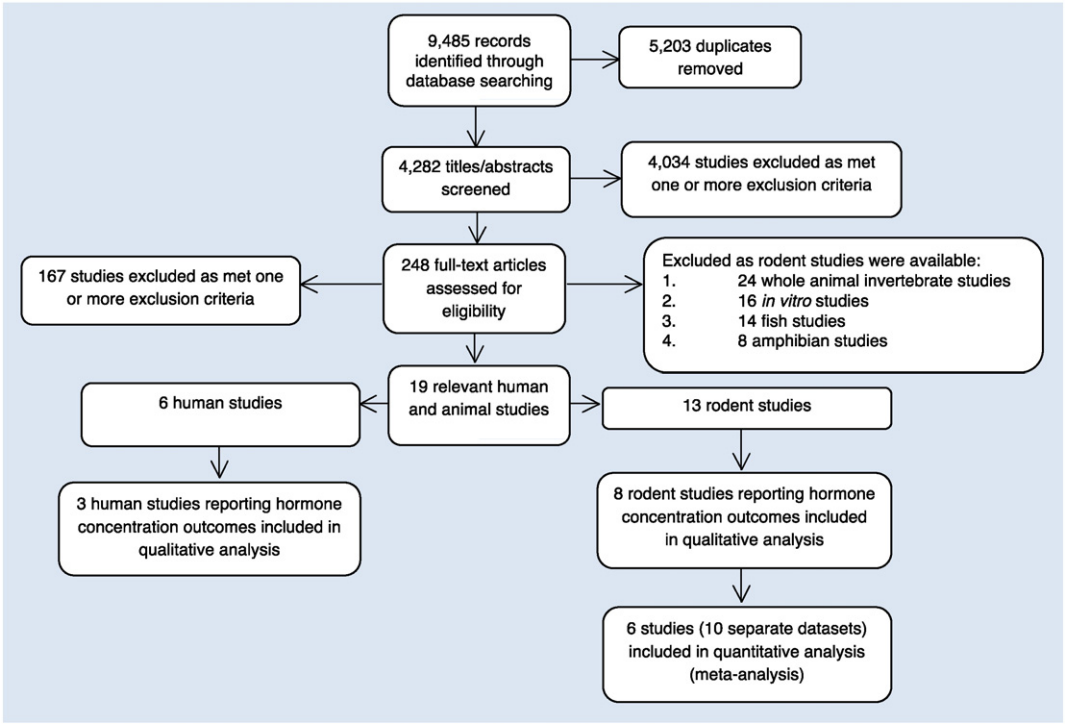

Fig. 1. Flowchart of the study-selection process.

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

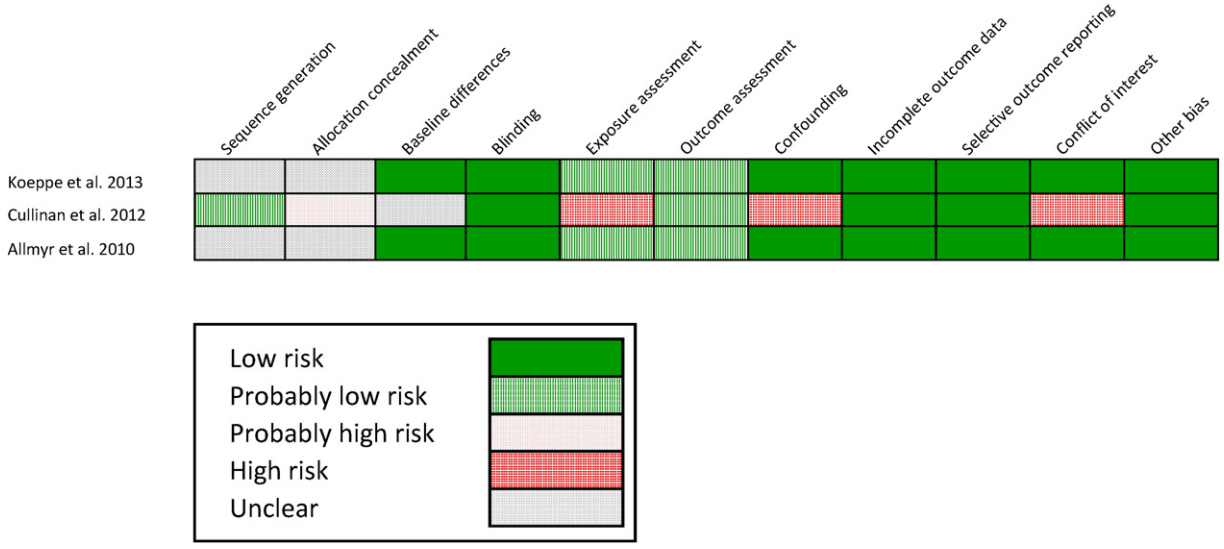

Fig. 2. Summary of risk of bias graphs for individual human studies. Review authors' judgments (low, probably low, probably high, and high risk) of bias for each risk of bias domain for each included human study ($n = 3$). The risk of bias results did not differ according to different specific outcome 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

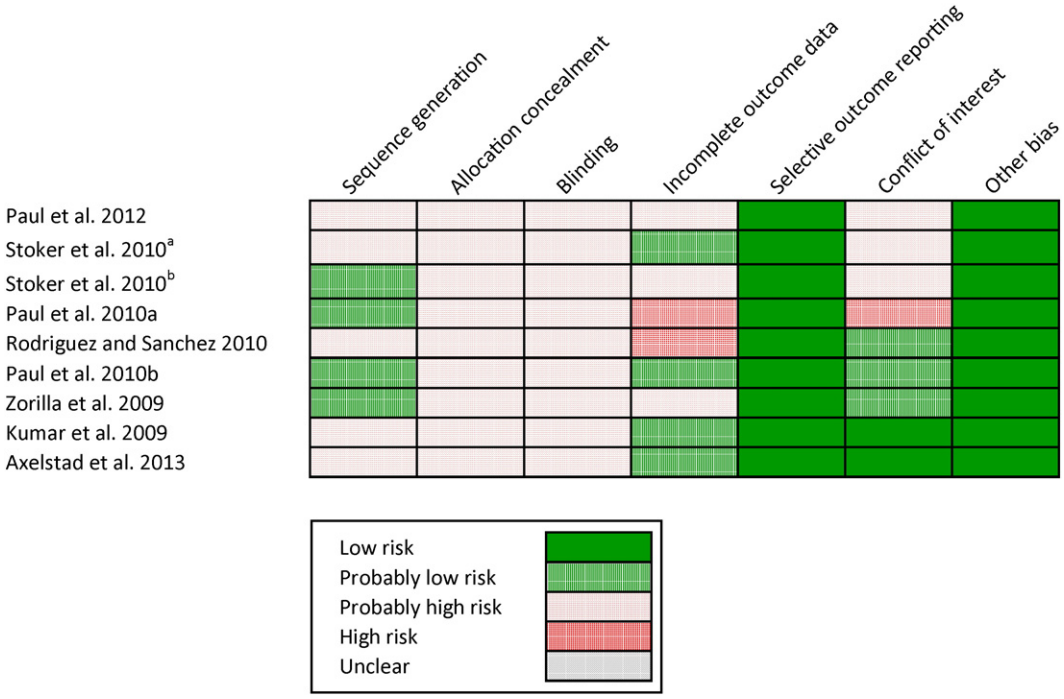

Fig. 3. Summary of risk of bias graphs for individual animal studies. Review authors' judgments (low, probably low, probably high, and high risk) of bias for each risk of bias domain for each included animal study ($n = 8$). Note Stoker et al. presents two experiments using two separate cohorts (pubertal 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.

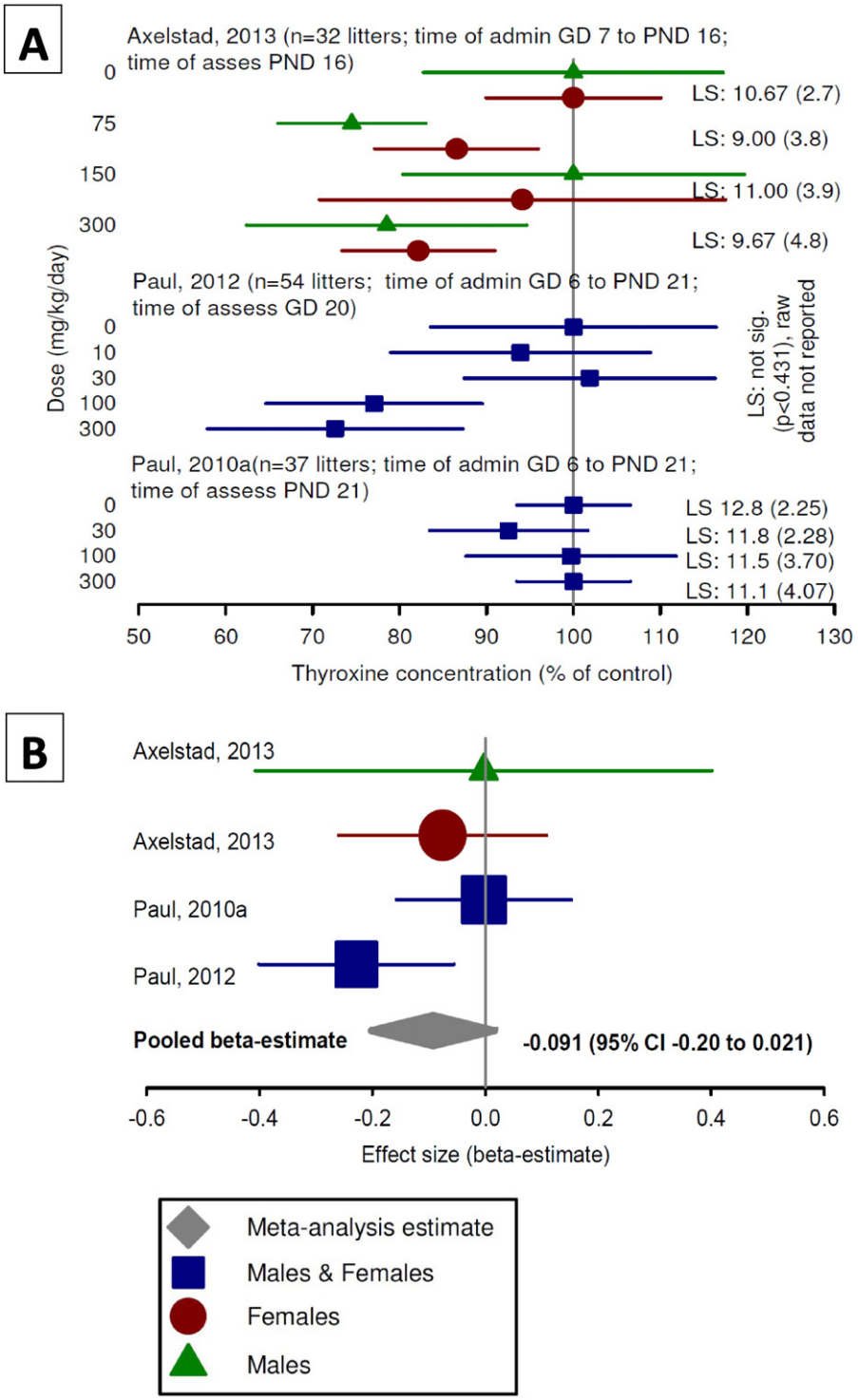

Fig. 4. Prenatal triclosan and thyroxine. A. Prenatal triclosan administration and thyroxine concentration as a percentage of the control group for doses up to 300 mg/kg/day. B. Prenatal beta-estimates for dose response and the random effects meta-analysis estimate. The vertical gray bar in A represents 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

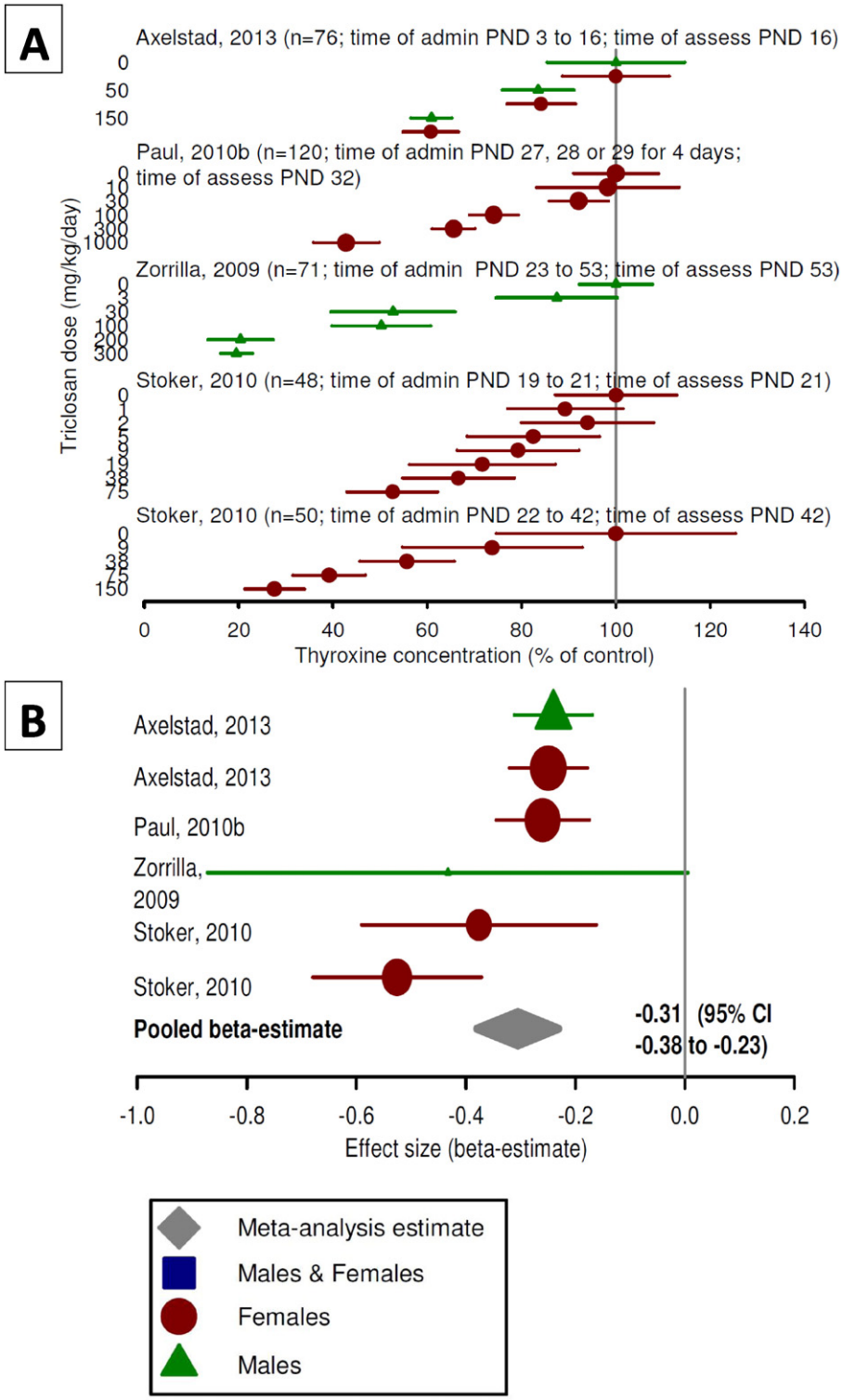

Fig. 5. Postnatal triclosan and thyroxine. A. Postnatal triclosan administration and thyroxine concentration as a percentage of the control group for doses up to 300 mg/kg/day. B. Postnatal beta-estimates for dose response and the random effects meta-analysis estimate. The vertical gray bar in A represents 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-UCSF 17288), 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 InterventionsVersion 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. Departmentof Healthand 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.