Critical Reviews in Toxicology

ISSN: 1040-8444 (Print) 1547-6898 (Online) Journal homepage: https://www.tandfonline.com/loi/itxc20

Systematic review of the literature on triclosan and health outcomes in humans

Michael Goodman, Daniel Q. Naiman & Judy S. LaKind

To cite this article: Michael Goodman, Daniel Q. Naiman & Judy S. LaKind (2018) Systematic review of the literature on triclosan and health outcomes in humans, Critical Reviews in Toxicology, 48:1, 1-51, DOI: 10.1080/10408444.2017.1350138

To link to this article: https://doi.org/10.1080/10408444.2017.1350138

CRITICAL REVIEWS IN TOXICOLOGY, 2018
VOL. 48, NO. 1, 1–51
https://doi.org/10.1080/10408444.2017.1350138

Taylor & Francis
Taylor & Francis Group

REVIEW ARTICLE

OPEN ACCESS Check for updates

Systematic review of the literature on triclosan and health outcomes in humans

Michael Goodman[a], Daniel Q. Naiman[b] and Judy S. LaKind[c,d]

[a]Department of Epidemiology, Rollins School of Public Health, Emory University, Atlanta, GA, USA; [b]Department of Applied Mathematics & Statistics, The Johns Hopkins University, Baltimore, MD, USA; [c]LaKind Associates, LLC, Catonsville, MD, USA; [d]Department of Epidemiology and Public Health, University of Maryland School of Medicine, Baltimore, MD, USA

ABSTRACT

The ability of epidemiologic evidence to inform regulatory decisions is largely dependent on the coherence and quality of the published literature. This systematic review examines the quality and consistency of studies assessing health outcomes associated with exposure to triclosan (TCS), an antimicrobial chemical with a short physiologic half-life. We used elements of the Biomonitoring, Environmental Epidemiology, and Short-Lived Chemicals instrument to evaluate aspects of study quality. Each methodological domain – overall design, exposure assessment, and data analysis – was categorized according to three tiers where Tier 1 indicated the highest quality. We also examined consistency of methods, results and reporting as considerations for weight of evidence (WOE) assessment. Studies were considered sufficiently comparable if they addressed the same or similar research questions. Forty-two studies met the criteria for inclusion. Only one randomized cross-over clinical trial of TCS was assigned to Tier 1 for all three domains. Most other studies were assigned to Tier 3 for at least one domain. Although the available literature examined more than 100 different health endpoints and reported hundreds of different measures of association, few studies were considered comparable. For reported measures of association, most were not significantly different from the null; the few statistically significant results represented isolated findings without a discernable across- or within-study pattern. We conclude that the current body of epidemiologic literature does not allow a meaningful WOE assessment due to methodological limitations of individual studies and lack of inter-study consistency. On the other hand, methodologically stronger studies may be used to inform future research.

ARTICLE HISTORY
Received 29 March 2017
Revised 25 June 2017
Accepted 29 June 2017

KEYWORDS
Triclosan; epidemiology; systematic review; ICC; short-lived chemical; anti-androgen; BEES-C; PI(E)CO

Table of contents

CONTACT Judy S. LaKind lakindassoc@gmail.com LaKind Associates, LLC, 106 Oakdale Avenue, Catonsville, MD 21228, USA

Introduction

Systematic assessment of evidence from epidemiologic studies investigating potential health effects of environmental chemicals is playing an increasing role in regulatory decision-making (USEPA 2010; NAS 2014; OHAT 2015). However, the utility of epidemiologic data as the basis for regulation is largely dependent on the strength, coherence, and quality of the published studies (OHAT 2015; USEPA 2016). While the epidemiologic studies exploring health outcomes and associations with environmental chemical exposures are numerous, recent publications, and governmental reports indicate that informed decision-making may be difficult due to the variable quality of the available evidence and the lack of consistency in terms of specific hypotheses tested, study methods, and reporting of results (Goodman et al. 2010, 2014a, 2014b; Youngstrom et al. 2011; LaKind et al. 2014a, 2014b, 2015a, 2015b; Rooney et al. 2014; USEPA 2016). In this communication, we offer a systematic examination of the epidemiologic literature on one specific chemical of interest – triclosan (TCS).

TCS (2,4,4′-trichloro-2′-hydroxydiphenyl ether) is a broad-spectrum antimicrobial agent that has been in use for over 40 years (Calafat et al. 2008; Yueh and Tukey 2016). TCS is used as an antiseptic, disinfectant, or preservative in medical and personal care products such as hand soaps and shampoos, mouthwash, toothpaste, and cosmetics, and in household items such as cutting boards (Witorsch and Thomas 2010; Yueh and Tukey 2016). TCS was found in the majority of urine samples obtained via population-based surveys both in the US and in Canada (Calafat et al. 2008; Health Canada 2015) and, as shown in Figure 1, a number of international regulatory activities in recent years have focused on this chemical.

Toxicological studies suggest that TCS can impact endocrine function, thyroid hormone homeostasis, and antibiotic resistance (Yueh and Tukey 2016). As reviewed in detail elsewhere (Witorsch and Thomas 2010), mechanism of action of TCS may involve interactions with aryl hydrocarbon and sex hormone receptors, and effects on the hypothalamo-pituitary-thyroid axis. However, *in vitro* and *in vivo* findings have not been consistently coherent leaving room for uncertainty in predicting health outcomes.

Widespread human exposure and concerns about health effects make TCS an emerging chemical of interest and an increasing focus of research. At the same time, several important aspects of TCS biomonitoring present a challenge for observational epidemiologic studies. For example, a factor that may introduce error in TCS exposure assessment is the likelihood of sample contamination because the chemical is widespread in the environment and can be commonly found in laboratory and clinical settings (Barr et al. 1999; Needham et al. 2007; Calafat and Needham 2008, 2009). In addition, stability of urinary TCS metabolites, which are typically measured in biomonitoring studies, may depend on the proper handling, freezing and storage of samples (Ye et al. 2007; Provencher et al. 2014).

Perhaps the most distinguishing feature of TCS is its short physiological half-life, which means that it is rapidly absorbed, metabolized, and eliminated (primarily via urine) with a median excretion half-life of 11 h after oral intake (Yueh and Tukey 2016). In addition, TCS exposures vary over time following changes in health status, product use, activity or location (Pleil and Sobus 2013). For these reasons, full characterization of exposure to TCS may require continuous biomonitoring, which involves taking multiple samples from the same individual at different times (NRC 2006).

This review examines the quality, quantity, and consistency of studies assessing the associations between TCS exposure and various health outcomes in humans. In assessing study quality, we focus on the critical issues of study design, exposure assessment (with specific emphasis on exposure misclassification due to insufficient number of measures) and data analysis. We then examine consistency of the TCS literature in terms of specific hypotheses tested, methods, results and reporting, and assess its utility for weight of evidence (WOE) assessment. In reviewing the literature on TCS, we also aim to share observations pertaining to methodological issues that are applicable to other ubiquitous chemicals with short physiological half-lives.

Quality assessment considerations

In conducting this review, we relied on elements of the Biomonitoring, Environmental Epidemiology, and Short-Lived

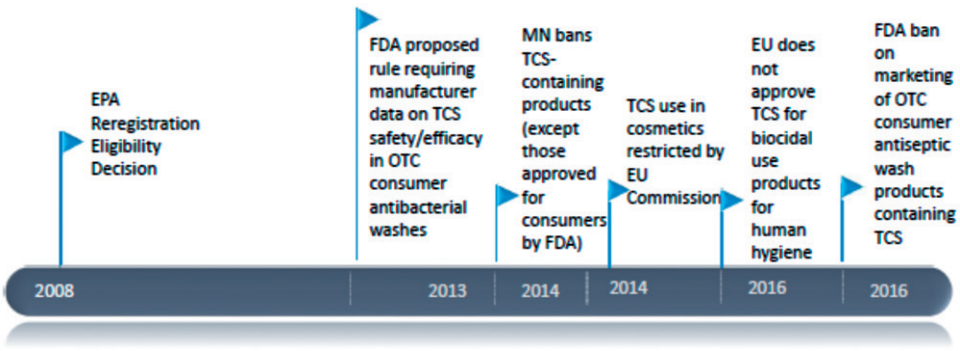

Figure 1. Regulatory actions related to TCS (EU: European Commission; FDA: Food and Drug Administration; MN: Minnesota; OTC: over-the-counter) (Beyond Pesticides 2016; EU 2016; USEPA 2010; USFDA 2016).

Chemicals (BEES-C) instrument, which was developed specifically to guide the transparent assessment of the quality of epidemiologic studies focused on exposure to short-lived environmental chemicals such as TCS (LaKind et al. 2014a). The BEES-C instrument provides a framework for evaluating three main domains of study quality: (i) study design, (ii) exposure assessment (including biomarker selection and measurement), and (iii) data analysis. Each of the above domains is categorized according to three tiers where Tier 1 indicates the highest quality and Tier 3 is reserved for study aspects that are considered suboptimal.

Study design

In assessing design quality, we focused on two distinct, but related issues: threats to validity and ability to establish the temporal relation between exposure of interest (in this case TCS) and the outcome of concern. There is general consensus that experimental studies (particularly randomized clinical trials) have fewer threats to validity and provide more definitive evidence compared to observational studies addressing the same research question (Guyatt et al. 2000a, 2000b; Petticrew and Roberts 2003). The main purpose of randomization is to avoid confounding by ensuring that participant characteristics are evenly distributed across treatment groups (Roberts and Torgerson 1998). In addition, a well-designed and properly implemented experimental study guarantees that the probability of being enrolled does not depend on the probability of being assigned to a particular treatment group, thereby minimizing another threat to validity – selection bias (Kahan et al. 2015).

An additional important attribute of experimental studies is that they leave no room for ambiguity with respect to temporal relation between exposure and outcome. By contrast, in observational research, establishing temporality is only possible in "incidence" studies, which identify health-related events such as new cases of disease at the time of onset or at least at the time of detection (Pearce 2012). Observational incidence studies may use a cohort and various versions of a case-control (particularly nested case-control) design. Regardless of design, however, the main feature of incidence studies is the ability to establish the proper sequence of exposure and outcome. If exposures are likely to vary over time (as in many biomonitoring studies of short-lived chemicals), a useful approach in some circumstances may be a longitudinal study that assesses the relation between repeated measures of exposure (e.g. TCS) and repeated measures of health biomarkers (e.g. levels of thyroid hormones).

Unlike incidence studies, a "prevalence" study cannot establish temporality. A typical prevalence study relies on cross-sectional design, which ascertains the exposure and health outcome information simultaneously (Rothman and Greenland 1998). Case-control studies also fall into this category if they use prevalent rather than incident cases (LaKind et al. 2014a).

Another consideration when assessing strength and weakness of study design is the need to capture critical windows of exposure relevant to the outcome of interest. For TCS, this critical window is usually not known; however, when a health outcome is expected to occur within a relatively short time after exposure (e.g. a change in health biomarker levels following intentional TCS administration in a clinical trial), conclusions about a presence or absence of a causal link may be justified.

Exposure assessment

In biomonitoring studies of ubiquitous, short-lived chemicals such as TCS, perhaps the most important source of exposure misclassification is an insufficient number of samples (LaKind et al. 2014a). It has been shown that reliance on a single random sample to characterize exposure to a short-lived chemical may result in substantial misclassification (i.e. over- or under-estimation of the true exposure level) and may not provide sufficient information to state with confidence that the measured exposure preceded the outcome (Preau et al. 2010; LaKind et al. 2012, 2014a; Wielgomas 2013; Goodman et al. 2014a).

The BEES-C instrument (LaKind et al. 2014a, 2015b) offers specific guidance for assessing this and other issues (e.g. analyte stability and sample contamination) as possible sources of exposure misclassification and for assigning studies to quality tiers. In this review, we focus on the number of samples measured for TCS in each study, as this is likely to have the greatest impact on exposure misclassification. (In assessing the utility of one measurement for understanding longer-term exposures, researchers often rely on intra-class correlation coefficients, or ICCs – either previously reported values or ones developed from their own study data. This is explored further in the Discussion section. In addition, it is often the case that publications do not include sufficient information to assess whether avoidance of sample contamination and analyte stability were properly considered; this issue is also revisited in the Discussion section of this paper.

Data analysis and reporting of results

While essential aspects of data analysis are not specific to TCS or other short-lived chemicals, an assessment of study quality would be incomplete without an evaluation of the data analysis methods. The BEES-C instrument emphasizes the importance of distinguishing between predictive and explanatory analyses (LaKind et al. 2014a). The goal of predictive analysis is to maximize model fit and a decision on whether to retain a particular covariate is based on statistical considerations (Bellazzi and Zupan 2008). By contrast, in an explanatory (hypothesis testing) analysis, inclusion and exclusion of control variables (confounders, mediators or effect modifiers) are driven, at least in part, by *a priori* reasoning (Concato et al. 1993; Hernan et al. 2002; Beran and Violato 2010).

Considerations of analytic strategy and proper handling of confounding factors depend on the hypothesis under study. For example, confounding variables that are commonly considered in studies of neurodevelopmental effects in children include participant age, race, gender, socioeconomic status (SES), and maternal variables such as age, education, and IQ

(Amler et al. 2006). Similarly, factors that need to be considered in studies of respiratory effects may include, in addition to demographic and SES variables, personal and family history of allergies or asthma, occupation, and exposure to tobacco smoke (Beasley et al. 2015).

It is important to keep in mind that the results of observational studies are inevitably subject to uncertainty due to unaccounted threats to validity and various data handling decisions and assumptions. The magnitude of uncertainty can be formally assessed through quantitative sensitivity analyses, which are considered an important part of an analysis plan (Greenland 1996).

Methods

Overall approach

For this review, we followed the Assessment Tool of Multiple Systematic Reviews (AMSTAR) (Shea et al. 2007). AMSTAR is similar to the often-cited Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) instrument (Liberati et al. 2009); however, unlike PRISMA, AMSTAR is not focused on health care interventions and is dedicated primarily to quality of review methods rather than completeness of reporting.

Eight of the 11 AMSTAR items are relevant to this review: (1) *a priori* statement of research questions and inclusion/exclusion criteria; (2) duplicate literature searches by two or more coauthors; (3) the use of at least two electronic search engines followed by a supplemental search of reviews, textbooks, and secondary references with keywords and MeSH (Medical Subject Headings) statements reported in the methods section; (4) a list of studies excluded from the review; (5) a summary of study characteristics that met the inclusion criteria; (6) a formal assessment of individual study quality; (7) consideration of study quality in drawing conclusions; and (8) a statement of sources of support (Shea et al. 2007).

Two AMSTAR items omitted from this list relate specifically to meta-analyses which are not part of this review. The third omitted item calls for consideration of reports and other non-peer reviewed papers. As the goal of this review was to the assess TCS literature in terms of its utility for WOE assessment, it is important to keep in mind that many organizations charged with this task (e.g. the National Toxicology Program [NTP]) use only "publicly accessible, peer-reviewed information in its evaluations" (OHAT 2015). For this reason, we relied only on the published literature that would be considered by such organizations.

Identification and selection of studies

Electronic data sources PubMed and Web of Science Core Collection™ were used to conduct the initial literature search. For the PubMed search, the following keywords and Medical Subjects Heading (MeSH) terms were used: ("triclosan" [MeSH Terms] OR "triclosan" [All Fields]) AND ("epidemiology" [Subheading] OR "epidemiology" [All Fields] OR "epidemiology" [MeSH Terms]). The results of this search were limited to human studies. A separate search utilized the following terms: ("triclosan" [MeSH Terms] OR "triclosan" [All Fields]) AND ("toxicity" [Subheading] OR "toxicity" [All Fields]) AND ("humans" [MeSH Terms] OR "humans" [All Fields] OR "human" [All Fields]). For the Web of Science search, we used broad search criteria with only two keywords: "triclosan" and "epidemiology."

References of retrieved articles were reviewed to identify publications not captured by the electronic search. The search and selection of relevant studies was conducted independently by two study authors (JSL, MG) with all disagreements resolved by consensus. The criteria for inclusion into the review were as follows:

1. research of human populations,
2. a comparison of one or more levels of exposure to TCS in relation to a specified reference category (independent variable),
3. data included health-related outcomes (dependent variables),
4. the association between TCS exposure and outcome was either reported or could be assessed based on the information provided in the publication, and
5. publication appeared in English prior to 15 August 2016 (end of literature search).

Publications on antimicrobial efficacy of TCS in different products (e.g. toothpaste and sutures), descriptive studies that did not assess the associations between TCS and health outcomes and papers that did not include data on humans were excluded. A list of studies retrieved and evaluated, but excluded from the review, and the reasons for exclusions are provided in Supplement 1.

Data extraction

We followed the literature review protocol described in previous publications (Goodman et al. 2014a; LaKind et al. 2014b). Each study that met the inclusion criteria was examined independently by two coauthors (MG and JSL). Information extracted from each study for the purposes of this review included the following:

1. Description of the study population: size, composition, source, and location
2. Study design: e.g. clinical trial, cohort, cross-sectional, case-control, or other
3. Type of TCS exposure: e.g. intervention- or biomonitoring-based, and for the latter, source of specimen and number and timing of samples
4. The type of exposure variable (e.g. ordinal, binary, or continuous) used in the analysis
5. Health outcomes of interest
6. Statistical methods for assessing the association of interest and the list of covariates
7. Results: the reported measure of association (e.g. an odds ratio [OR] for binary endpoints, or a linear regression coefficient [β] for continuous endpoints), and a measure used for statistical inference (e.g. a 95% confidence interval [CI] or a p value).

If the result of interest was not reported in the paper, whenever possible, the measures of association and 95% CIs were calculated by one of the authors (MG). When the result was reported in a qualitative fashion, the corresponding text was reproduced verbatim. The data from each study were tabulated, and the resulting summary tables were again cross-checked with disagreements resolved by consensus.

Assessment of strength of evidence (data quality) from individual studies

We used components of the BEES-C instrument (LaKind et al. 2014a) for evaluating three main domains of study quality: (i) study design, (ii) exposure assessment (with specific emphasis on the number of samples and/or quantitative assessment of measurement error), and (iii) data analysis. Each of these domains is categorized according to three tiers where Tier 1 indicates the highest quality and Tier 3 is reserved for study aspects that are considered suboptimal.

Study design

When evaluating quality of the overall study design, clinical trials of TCS were included in Tier 1 because they present the lowest threat to validity and clearly establish the temporal relation between intervention and outcome. Cohort studies and case-control studies that used incident cases were placed in Tier 2; these studies are generally able to establish the correct sequence of exposure and outcome, but are more vulnerable to bias than clinical trials. Case-control studies with prevalent cases and cross-sectional studies were assigned to Tier 3 because in these studies both exposure and outcome are ascertained at the same time, and in some instances exposure assessment follows rather than precedes the outcome of interest. Publications that included both a cohort and cross-sectional design (depending on the analysis) were assigned to Tier 2.

Exposure assessment

In Tier 1 studies, the exposure assessment was based on a sufficient number of samples per individual to estimate exposure over the appropriate duration, or used adequate long-term sampling (e.g. multiple 24-h urine collections). If only one sample was used to characterize TCS exposure, the study was assigned to Tier 1 if it provided quantitative evidence that probability of misclassification was small. A Tier 2 study included more than one sample but without explicit evaluation of error or had more than one TCS measure, but used each measure as a separate observation. The randomized experimental studies were assigned to Tier 1 if they provided evidence of increased TCS levels following intervention (relative to the reference group) thereby leaving little room for exposure misclassification. Otherwise, experimental studies were assigned to Tier 2. In a Tier 3 study, exposure was based on a single sample without considering error or there were no biomonitoring data.

Data analysis and reporting of results

We reviewed the data analysis sections of each paper and made decisions regarding assignment to one of the three tiers for the data analysis domain based on the general considerations outlined in the BEES-C guidelines (LaKind et al. 2014a). Tier 1 was reserved for studies that used randomization or demonstrated comprehensive evaluation of control variables and included sensitivity analyses. Tier 2 ratings described studies that applied appropriate statistical methods but did not include sensitivity analyses. Studies that did not control for extraneous factors or considered only a few potential confounders (e.g. age and BMI) without justification for covariate selection were assigned to Tier 3. If a publication included several analyses, the tier assignment was based on the analysis of the highest quality.

Assessment of consistency across studies

All studies were organized according to nine broad categories of outcomes of interest: (1) allergies and asthma, (2) measures of overweight and obesity, (3) female reproductive endpoints, (4) male reproductive endpoints, (5) birth outcomes, (6) puberty endpoints in girls, (7) thyroid function, (8) sex hormone levels, and (9) miscellaneous. The last category included a variety of outcomes for which only one or two studies were conducted and assessment of consistency across studies was not possible. In instances when the methods section of a publication indicated evaluation of multiple outcomes but the results section reported only selected findings, the non-reported results were also documented. This approach allowed us to search for reasonably homogeneous groups of comparable articles and evaluate consistency of methods, results and reporting within each group. Studies were considered sufficiently comparable if they addressed the same or similar research questions. The comparability was assessed using the PI(E)CO format, where P specifies the target population, I(E) determines the intervention (or exposure), C denotes the comparison group, and O identifies the outcome of interest (Centre for Reviews and Dissemination 2016). The within-study consistency was assessed by comparing results of different analyses testing the same hypothesis, but using different exposure metrics or analytic approaches.

Results

Overview of studies

We identified 89 studies from our literature search; of those, 43 studies met the criteria for inclusion in the current review (Supplement 1). Of the 43 studies, 24 were conducted in North America (Wolff et al. 2008, 2010, 2015; Clayton et al. 2011; You et al. 2011; Buttke et al. 2012; Savage et al. 2012, 2014; Koeppe et al. 2013; Lankester et al. 2013; Buser et al. 2014; Spanier et al. 2014; Adgent and Rogan 2015; Li et al. 2015; Ihde et al. 2015; Shiue 2015a, 2015b; Velez et al. 2015; Ashley-Martin et al. 2016; Buckley et al. 2016; Poole et al. 2016; Scinicariello and Buser 2016; Smarr et al. 2017; Geer et al. 2017), two studies in the Caribbean (Watkins et al. 2015; Aker et al. 2016), 10 studies in Europe (Allmyr et al. 2009; Chevrier et al. 2012; Philippat et al. 2012, 2014; Bertelsen et al. 2013; Buhl et al. 2014; Den Hond et al. 2013, 2015; Geens et al. 2015; Lassen et al. 2016), four in Asia (Chen et al.

2013; Wang et al. 2015; Xue et al. 2015; Zhu et al. 2016), and three in Australia (Cullinan et al. 2012, 2015a, 2015b).

The literature on TCS exposures and health outcomes is quite recent and rapidly expanding. Most studies included in the review were published within the last three years, and none prior to 2008.

Not all studies represented independent data sources. For example, 14 US studies used data from the National Health and Nutrition Examination Survey (NHANES) (Clayton et al. 2011; You et al. 2011; Buttke et al. 2012; Savage et al. 2012, 2014; Koeppe et al. 2013; Lankester et al. 2013; Buser et al. 2014; Spanier et al. 2014; Adgent and Rogan 2015; Li et al.

Tier 1: Design: clinical trials/ Exposure: assessment based on a sufficient number of samples per individual to estimate exposure over the appropriate duration, or through the use of adequate long-term sampling; for chemicals for which one sample may be sufficient to fully characterize exposure, inclusion of evidence that errors can be considered sufficiently small; randomized experimental studies with inclusion of increased TCS levels following intervention/ Data analysis: randomization or demonstrated comprehensive evaluation of control variables and included sensitivity analyses.

Tier 2: Design: cohort studies and case-control studies with incident cases; studies with both cohort and cross-sectional designs/ Exposure: > one sample but without explicit evaluation of error or had > one TCS measure but used each measure as a separate observation; other experimental studies/ Data analysis: applied appropriate statistical methods but did not include sensitivity analyses.

Tier 3: Design: case-control studies with prevalent cases and cross-sectional studies/ Exposure: based on a single sample without considering error or there were no biomonitoring data/ Data analysis: no control for extraneous factors or considered only a few potential confounders without justification for covariate selection.

Publication	Design	Exposure assessment	Data analysis
Adgent and Rogan 2015	Tier 3	Tier 3	Tier 1
Aker et al. 2016	Tier 2	Tier 1	Tier 1
Allmyr et al. 2009	Tier 1	Tier 1	Tier 3
Ashley-Martin et al. 2016	Tier 2	Tier 3	Tier 2
Bertelsen et al. 2013	Tier 3	Tier 3	Tier 2
Buckley et al. 2016	Tier 2	Tier 3	Tier 1
Buhl et al. 2014	Tier 3	Tier 3	Tier 3
Buser et al. 2014	Tier 3	Tier 3	Tier 2
Buttke et al. 2012	Tier 3	Tier 3	Tier 3
Chen et al. 2013	Tier 3	Tier 3	Tier 3
Chevrier et al. 2012	Tier 2	Tier 3	Tier 2
Clayton et al. 2011	Tier 3	Tier 3	Tier 2
Cullinan et al. 2012	Tier 1	Tier 3	Tier 1
Cullinan et al. 2015a	Tier 1	Tier 3	Tier 1
Cullinan et al. 2015b	Tier 1	Tier 3	Tier 1
Den Hond et al. 2013	Tier 3	Tier 3	Tier 2
Den Hond et al. 2015	Tier 3	Tier 3	Tier 2
Geens et al. 2015	Tier 3	Tier 2	Tier 3
Geer et al. 2017	Tier 2	Tier 2	Tier 2
Ihde et al. 2015	Tier 3	Tier 3	Tier 3
Koeppe et al. 2013	Tier 3	Tier 3	Tier 2
Lankester et al. 2013	Tier 3	Tier 3	Tier 2
Lassen et al. 2016	Tier 2	Tier 3	Tier 1
Li et al. 2015	Tier 3	Tier 3	Tier 2
Philippat et al. 2012	Tier 2	Tier 3	Tier 1
Philippat et al. 2014	Tier 2	Tier 3	Tier 1
Poole et al. 2016	Tier 1	Tier 1	Tier 1
Savage et al. 2012	Tier 3	Tier 3	Tier 2
Savage et al. 2014	Tier 3	Tier 3	Tier 2
Scinicariello and Buser 2016	Tier 3	Tier 3	Tier 2
Shiue 2015a	Tier 3	Tier 3	Tier 2
Shiue 2015b	Tier 3	Tier 3	Tier 2
Smarr et al. 2017	Tier 2	Tier 1	Tier 1
Spanier et al. 2014	Tier 3	Tier 3	Tier 2
Velez et al. 2015	Tier 3	Tier 3	Tier 2
Wang et al. 2015	Tier 3	Tier 3	Tier 3
Watkins et al. 2015	Tier 2	Tier 1	Tier 1
Wolff et al. 2008	Tier 2	Tier 3	Tier 2
Wolff et al. 2010	Tier 2	Tier 3	Tier 2
Wolff et al. 2015	Tier 2	Tier 3	Tier 2
Xue et al. 2015	Tier 3	Tier 3	Tier 3
You et al. 2011	Tier 3	Tier 3	Tier 2
Zhu et al. 2016	Tier 3	Tier 3	Tier 2

Figure 2. Results of evaluation of quality of individual TCS epidemiology studies using elements of the BEES-C instrument.

2015; Shiue 2015a, 2015b; Scinicariello and Buser 2016), and most used at least partially overlapping survey years. Similarly, three Australian papers (Cullinan et al. 2012, 2015a, 2015b) were based on the same study, as were two publications from Puerto Rico (Watkins et al. 2015; Aker et al. 2016), and three papers by Wolff et al. (2008, 2010, 2015).

Assessment of strength of evidence: quality of studies

Study design

As shown in Figure 2, more than half of studies ($n = 25$) were assigned to Tier 3 based on their overall design. Twenty two studies were cross-sectional and the remaining three (Chen et al. 2013; Den Hond et al. 2015; Wang et al. 2015) used case-control design, but included prevalent rather than incident cases.

Tier 2 included 13 publications. Twelve of those publications (Wolff et al. 2008, 2010, 2015; Philippat et al. 2012, 2014; Watkins et al. 2015; Aker et al. 2016; Ashley-Martin et al. 2016; Buckley et al. 2016; Lassen et al. 2016; Smarr et al. 2017; Geer et al. 2017) were based on seven different cohorts assembled in France, Puerto Rico, Denmark, Canada, and various parts of the US. One paper (Chevrier et al. 2012) presented a case-control study that used incident cases identified among members of a French cohort.

Five experimental studies were included in Tier 1. One of the methodologically strongest studies in this category was the double-blind, randomized, crossover trial assessing the effect of TCS on human microbiome (Poole et al. 2016). In that study, participants were randomly assigned to receive TCS-containing or TCS non-containing household and personal care products for 4 months and then switched exposure arms for an additional 4 months.

Three publications were based on the double-blind, randomized, placebo-controlled, prospective clinical trial Cardiovascular and Periodontal Study (CAPS) conducted in Australia among patients who had history of cardiovascular disease (Cullinan et al. 2012, 2015a, 2015b). CAPS participants were randomly assigned into one of two groups: one group received placebo toothpaste, and the other received toothpaste containing 0.3% TCS. All subjects were followed for five years.

The fifth publication in this group was the single arm Swedish study (Allmyr et al. 2009) that examined changes in cytochrome P450 3A4 (CYP3A4) activity and thyroid hormone concentrations before and after a two-week intervention that involved use of TCS-containing toothpaste.

Exposure assessment

Almost all of the studies ($N = 37$) reviewed here based their exposure assessment on one urine sample or included no measures of TCS (Figure 2). Only one of the studies with a single measurement (Smarr et al. 2017) provided an exploration of error associated with the use of one sample instead of generically describing error (e.g. with a discussion of ICCs; this is further described in the Discussion section). Thus, 36 of the 43 papers were assigned to Tier 3.

Geens et al. (2015) obtained multiple urine samples at different times in the study but each was used as a separate observation; thus, this publication was assigned to Tier 2. Similarly, Geer et al. (2017) analyzed third trimester urine samples and cord blood samples separately and was assigned to Tier 2. The remaining five publications were assigned to Tier 1. Two of those (Watkins et al. 2015; Aker et al. 2016) were longitudinal cohort studies that included three samples obtained at different times and examined associations with health outcomes, taking all three TCS measures into consideration. Allmyr et al. (2009) conducted an experimental study comparing health biomarker levels before and after intentional TCS administration and observed significantly higher plasma TCS levels following the 14-day study intervention (0.009–0.81 ng/g and 26–296 ng/g: ranges before and after exposure, respectively). Poole et al. (2016) used a cross-over experimental study design and included measurements of TCS in five urine samples for each study period with clear evidence of increased exposure during the intervention phase. Smarr et al. (2017) used one sample in a study of reduced fecundity but explored the extent of error by conducting a separate analysis restricted to couples with a shorter time to pregnancy.

While we did not formally assign quality tiers for steps taken to avoid sample contamination and to ensure analyte stability, we note that only two studies specifically reported the incorporation of field and lab blanks in their methods (Velez et al. 2015; Ashley-Martin et al. 2016). None of the studies offered documentation of analyte stability. Fourteen studies used NHANES data, and did not provide documentation for either of these issues. Documentation for prevention of contamination of samples during laboratory operations is available (Calafat et al. 2008; CDC 2013), but no similar documentation describing these precautions during sample collection was available nor was information given on sample handling and storage conditions.

Data analysis

When assessing the quality of data analysis, Tiers 1 and 2 included 12 and 23 publications, respectively (Figure 2). The remaining eight publications failed to control for extraneous factors or considered only a few selected basic covariates such as age and BMI; these were assigned to Tier 3 (Figure 2).

Among papers included in Tier 1, particularly notable were studies that performed longitudinal analyses with repeated measures of TCS exposure, outcomes of interest, or both. For example, Aker et al. (2016) used data from a cohort study of pregnant women in Puerto Rico to examine associations between urinary TCS concentrations measured during three visits at 16–20, 20–24, and 24–28 weeks gestation and serum levels of thyroid and sex hormones at visits 1 and 3.

Although all cross sectional studies included in the present review were assigned to Tier 3 in the overall design and the exposure assessment domains, one cross-sectional study was placed in Tier 1 with regards to its data analysis methods. Adgent and Rogan (2015) analyzed the data from the 2005–2008 NHANES to examine the association between urinary TCS and enterolactone (an indirect marker of gut microflora and antibiotic use). The linear regression analyses in that

study controlled for a number of potential confounders including patient demographic characteristics, SES, dietary habits, and frequency of bowel movement. The authors also conducted several sensitivity analyses, which treated TCS as continuous, binary, and ordinal variables.

Assessment of consistency

The 43 studies included in the present review examined more than 100 different health endpoints and reported hundreds of different measures of association across nine broad categories: (1) allergies and asthma, (2) measures of overweight and obesity, (3) female reproductive endpoints, (4) male reproductive endpoints, (5) birth outcomes, (6) puberty endpoints in girls, (7) thyroid function, (8) sex hormone levels, and (9) miscellaneous. In the following sections of the review, we examine each of these categories and assess inter- and intra-study consistency of methods and reported results. Tables 1–9 summarize the studies in each of the nine groups using the PI(E)CO format.

Allergies and asthma

Table 1 summarizes six studies that evaluated the association between TCS and various indicators of allergy and asthma (Clayton et al. 2011; Savage et al. 2012, 2014; Bertelsen et al. 2013; Buhl et al. 2014; Spanier et al. 2014). All six studies were cross sectional. In five studies, TCS exposure was based on measured urinary levels from a single sample. The sixth study (Buhl et al. 2014) defined TCS exposure as a positive response to a skin patch test. All six studies were included in Tier 3 with regards to their exposure assessment methods. The studies examined 24 different endpoints such as various categories of wheeze and asthma, evidence of sensitization to numerous allergens, rhinitis, dermatitis and eczema. Most of the analyses did not show significant associations and the results demonstrated no consistent patterns across studies. Three studies reported that persons with higher TCS levels may have higher prevalence of inhalation allergies; however, two of those studies (Savage et al. 2012; Spanier et al. 2014) used the same 2005–2006 NHANES data. These same studies did not find consistent evidence of an association with asthma prevalence.

Measures of overweight and obesity

Nine studies assessed the association between TCS exposure and measures of overweight or obesity (e.g. body mass index [BMI], waist circumference [WC]) (Table 2). All but three studies in this group were cross-sectional, and all except one were assigned to Tier 3 with respect to the quality of exposure assessment.

In the previously cited double-blind, randomized, crossover trial assessing the effect of TCS on human microbiome (Poole et al. 2016), the authors also examined whether the participants demonstrated any weight gain during the TCS vs. non-TCS intervals (each interval was 4 months duration). Six subjects gained weight (defined as increase of at least 0.6%) in the TCS phase but lost weight or stayed at the same weight in the non-TCS phase; however this result was not statistically significant ($p = 0.22$).

A study of 173 mother–child pairs (Buckley et al. 2016) included measurements of TCS in a single spot urine sample collected from prospective mothers at 25–40 weeks gestation and followed their children postnatally to calculate percent body fat at three visits between 4 and 9 years of age. Mixed effects linear models with repeated measures were used to examine the association between prenatal TCS and percent body fat across the follow up period. Additional models examined the same endpoint at each visit and extensive sensitivity analyses were used to examine the effects of different strategies of controlling for urine dilution. None of the results demonstrated a significant departure from the null value; however these findings are difficult to examine for inter-study consistency because no other studies used percent body fat as the endpoint of interest.

Another cohort study (Wolff et al. 2015) followed 1170 girls (ages 6–8 years) with annual examinations over a period of up to seven years. TCS was measured at baseline in a single urine sample collected at enrollment. Body mass index percentile was examined at the last pre-pubertal visit. None of the results demonstrated a statistically significant association.

The results of the cross-sectional studies did not reveal any discernable pattern. Most analyses yielded findings consistent with the null hypothesis. In two NHANES studies that observed significant associations (Lankester et al. 2013; Li et al. 2015), the results were in opposite directions.

Female reproductive endpoints

Two cross-sectional studies and one cohort study examined the association between TCS exposure and female reproductive health (Table 3). The cohort study by Smarr et al. (2017) was assigned to Tier 1 for exposure characterization as described previously. The cross-sectional studies by Wang et al. (2015) and Velez et al. (2015) were assigned to Tier 3 with respect to exposure assessment methods, as they both relied on one TCS measurement with no assessment of error in exposure characterization.

Wang et al. (2015) examined urine TCS concentrations in women who had medically unexplained spontaneous abortion in mid-gestation (cases) and in women who had no history of spontaneous abortion and had at least one living child (controls). Using subjects whose urinary TCS was $<$ LOD as the reference category, the authors reported significantly elevated ORs for "medium" and "high" TCS exposure. While these findings led the authors to conclude that "TCS-exposure might cause spontaneous abortion," this causal conclusion based on the human data from Wang et al. is not warranted because exposure in this study was measured after the event of interest and because methods of study design, exposure assessment and data analysis were consistent with Tier 3.

A cohort of women enrolled in the Longitudinal Investigation of Fertility and the Environment (LIFE) study provided a single spot urine sample within 2 months of stopping contraception and the participants were followed to

Table 1. Summary of epidemiologic studies evaluating association between TCS exposure and allergic conditions (bolded text indicates a statistically significant result).

Publication; study design	Study population	Exposure assessment	Outcome definition	Analysis; measure of association	Results
Bertelsen et al. (2013); cross-sectional	Ten-year old members ($n = 623$) of a birth cohort in Oslo (preferential sampling of subjects with current asthma), Norway	Single urine sample at 10 years of age; <LOD and quartiles > LOD or continuous log_{10}-transformed LOD: 2.3 μg/l % > LOD: 47 TCS, μg/l ref: <2.3 Q1: 2.3–4.3 Q2: 4.4–11.3 Q3: 11.4–121 Q4: >121	Positive SPT and/or elevated IgE for 15 different allergens Parent-reported current allergic rhinitis Current asthma ascertained based on parental report plus exercise-induced FEV1 decreased by ≥10%, one of various symptoms or current use of asthma medication	Logistic regression adjusted for urine SG, parental history of allergy, income, maternal education; OR (95% CI)	SPT or IgE: <LOD (ref) Q1: 0.96 (0.54, 1.7) Q2: 1.5 (0.86, 2.7) Q3: 1.4 (0.83, 2.5) Q4: **2.0 (1.1, 3.4)** *p*-trend=**0.03** Per log_{10}: 1.2 (1.0, 1.4) SPT only: <LOD (ref) Q1: 1.1 (0.62, 2.0) Q2: 1.4 (0.77, 2.6) Q3: 0.99 (0.54, 1.8) Q4: 1.7 (0.96, 3.0) *p*-trend =0.09 Per log_{10}: 1.1 (0.97, 1.3) IgE only: <LOD (ref) Q1: 0.94 (0.52, 1.7) Q2: 1.6 (0.91, 2.8) Q3: 1.4 (0.82, 2.5) Q4: **2.2 (1.3, 3.8)** *p*-trend=**0.008** Per log_{10}: 1.3 (1.1, 1.5) Rhinitis: <LOD (ref) Q1: 0.72 (0.37, 1.4) Q2: 1.4 (0.73, 2.5) Q3: 1.1 (0.57, 2.0) Q4: **1.9 (1.1, 3.4)** *p*-trend=**0.02** Per log_{10}: 1.2 (0.97, 1.4) Asthma: <LOD (ref) Q1: 1.1 (0.60, 2.1) Q2: 0.99 (0.51, 1.9) Q3: 1.1 (0.59, 2.0) Q4: 0.89 (0.45, 1.8) *p*-trend =0.7 Per log_{10}: 1.0 (0.88, 1.3) Inhalant allergens: <LOD (ref) Q1: 1.1 (0.64, 2.0) Q2: 1.7 (0.94, 3.0) Q3: 1.4 (0.83, 2.5) Q4: **2.1 (1.2, 3.6)** *p*-trend=**0.02** Per log_{10}: 1.2 (1.1, 1.4) Seasonal allergens: <LOD (ref) Q1: 0.94 (0.50, 1.7) Q2: 1.6 (0.85, 2.9) Q3: 1.2 (0.65, 2.2) Q4: **2.1 (1.2, 3.7)**

(*continued*)

Table 1. Continued

Publication; study design	Study population	Exposure assessment	Outcome definition	Analysis; measure of association	Results
					p-trend=**0.02** Per log_{10}: 1.2 (1.0, 1.4) Perennial allergens: <LOD (ref) Q1: 1.4 (0.77, 2.5) Q2: 1.6 (0.84, 2.9) Q3: 1.6 (0.89, 2.9) Q4: 1.5 (0.82, 2.8) *p*-trend =0.4 Per log_{10}: 1.2 (1.0, 1.4) Food allergens: <LOD (ref) Q1: 0.59 (0.25, 1.4) Q2: 1.0 (0.49, 2.1) Q3: 0.63 (0.28, 1.4) Q4: 1.3 (0.64, 2.5) *p*-trend =0.3 Per log_{10}: 0.99 (0.81, 1.2) No effect modification by sex
Buhl et al. (2014); cross sectional	Patients ($n = 112\,594$) reported to the Information Network of Departments of Dermatology, 1993–2012, Germany, Switzerland, and Austria	Positive patch skin reaction to TCS; yes vs. no No measurements	Occupational dermatitis, atopic dermatitis, site specific dermatitis, allergic rhinitis, asthma	Crude associations; PR (95% CI)[a] [a]Calculated based on data provided in the paper, logistic regression analyses reported in the article cannot be interpreted	Occupational dermatitis: **0.43 (0.30, 0.62)** Atopic dermatitis: 0.63 (0.47, 0.84) Hand dermatitis: 0.45 (0.34, 0.60) Leg dermatitis: **2.41 (1.91, 3.05)** Face dermatitis: 0.94 (0.71, 1.24) Allergic rhinitis: 0.81 (0.61, 1.08) Allergic asthma: 0.88 (0.56, 1.38)
Clayton et al. (2011); cross-sectional	Adults and children ($n = 1444$) included in the 2003–2006 National Health and Nutrition Examination Surveys, USA	Single urine sample; ln-transformed continuous LOD: 2.3 ng/ml % > LOD: ~75 (2003–2004 data) TCS, ng/ml Mean: 116.3	Self-reported diagnoses of allergies or hay fever	Logistic regression adjusted for age, sex, Cr, race, BMI, family income and education OR (SE), *p* value	Children (≤18 years) **1.257 (0.097), $p < 0.01$** Adults (≥18 years) 0.996 (0.074), $p > 0.05$
Savage et al. (2012); cross sectional	Children ($n = 860$) included in the 2005–2006 National Health and Nutrition Examination Survey, USA	Single urine sample; tertiles Log-transformed TCS as continuous variable LOD: NR % > LOD: NR TCS, ng/ml T1: 1.15–5.7 T2: 5.8–36.8 T3: 36.9–2670	Elevated serum IgE for 15 different aeroallergens and four food allergens. Doctor-diagnosed asthma based on self-report Atopic asthma and atopic wheeze based on self-report +1 positive (≥35 kU/L) IgE level Nonatopic asthma based on doctor-diagnosed asthma and negative IgE test results Nonatopic wheeze base on wheezing in past 12 months and negative IgE test results	Logistic for elevated IgE and other binary outcomes adjusted for Cr, sex, age, race/ethnicity, and poverty index ratio; OR (95% CI) Linear regression for IgE level used as continuous variable adjusted for the same covariates β (95% CI)	Aeroallergy: T1 (ref) T2: **1.85 (1.15–2.98)** T3: **1.73 (1.11–2.69)** *p*-trend=**0.006** Food allergy: T1 (ref) T2: 1.53 (0.78–3.02) T3: **2.39 (1.10–5.18)** *p*-trend=**0.03** Aeroallergy (males): T1 (ref) T2: 2.15 (0.99–4.69) T3: **2.37 (1.01–5.53)**

(*continued*)

Table 1. Continued

Publication; study design	Study population	Exposure assessment	Outcome definition	Analysis; measure of association	Results
			IgE as continuous variable		*p*-trend=**0.046** Food allergy (males): T1 (ref) T2: 2.01 (0.88–4.62) T3: **3.91 (1.24–12.31)** *p*-trend=**0.02** Aeroallergy (females): T1 (ref) T2: 1.61 (0.83–3.11) T3: 1.16 (0.62–2.19) *p*-trend =0.57 Food allergy (females): T1 (ref) T2: 1.14 (0.45–2.89) T3: 1.24 (0.48–3.20) *p*-trend =0.64 Atopic asthma: T1 (ref) T2: 1.88 (0.74–4.76) T3: 2.41 (0.87–6.71) *p*-trend =0.07 Atopic wheeze: T1 (ref) T2: 1.95 (0.78–4.90) T3: 2.56 (0.76–8.62) *p*-trend =0.12 Non-atopic asthma: T1 (ref) T2: 0.63 (0.24–1.67) T3: 0.29 (0.7–1.17) *p*-trend =0.09 Non-atopic wheeze: T1 (ref) T2: **0.27 (0.10–0.75)** T3: 0.30 (0.07–1.24) *p*-trend =0.08 Ig E continuous[a] T1: (ref) T2: 0.054 (−0.064 to 0.172) T3: 0.070 (−0.062 to 0.203) *p*-trend =0.26 [a]From supplemental tables
Savage et al. (2014); cross sectional	Children and adults ($n = 7303$ for asthma prevalence; $n = 639$ for asthma control) included in the 2005–2006, 2007–2008 and 2009–2010 NHANES, USA	Single urinary measure; tertiles log-transformed continuous variable LOD =2.3 ng/ml % > LOD: 76–83 TCS, ng/ml T1: 1.13–4.9 T2: 5–27.9 T3: 28.2–3170	Self-reported asthma Among those with asthma, self-reported asthma attacks in last 12 months	Logistic regression adjusted for Cr level, age, race, sex, and poverty-index ratio; OR (95% CI)	Asthma prevalence: T1 (ref) T2: not reported T3: 1.14 (0.89–1.46) *p*-trend: NR Asthma attacks in last year: T1 (ref) T2: **1.89 (1.15, 3.13)** T3: **1.79 (1.01, 3.15)** *p*-trend: NR
Spanier et al. (2014); cross sectional	Children ($n = 837$) included in the 2005–2006 National Health	Single urinary measure; quartiles LOD: NR	Elevated IgE for 15 different aeroallergens and four food	Logistic regression adjusted for Cr level, age, race, sex, and	Aeroallergy: Q1 (ref)

(*continued*)

Table 1. Continued

Publication; study design	Study population	Exposure assessment	Outcome definition	Analysis; measure of association	Results
	and Nutrition Examination Survey, USA	% >LOD: NR TCS, ng/ml Q1: <3.8 Q2: 3.8–12.14 Q3: 12.15–53.99 Q4: ≥54.0 GM (95% CI): 15.5 (13.0, 18.6)	allergens Atopic asthma and atopic wheeze based on self-report + one IgE level ≥35 kU/L Self-reported eczema	poverty-index ratio; OR (95% CI)	Q2: **1.84 (1.16, 2.92)** Q3: **1.70 (1.10, 2.62)** Q4: **1.91 (1.02, 3.57)** *p*-trend: NR Eczema: Q1 (ref) Q2: 0.86 (0.32, 2.30) Q3: 0.64 (0.21, 1.91) Q4: 0.68 (0.29, 1.57) *p*-trend: NR Food allergy (eczema): Q1 (ref) Q2: 1.60 (0.33, 7.85) Q3: **21.21 (2.70, 166.71)** Q4: 8.17 (0.93, 71.59) *p*-trend: NR Food allergy (no eczema): Q1 (ref) Q2: 1.12 (0.47, 2.68) Q3: 1.04 (0.44, 2.43) Q4: 2.03 (0.85, 4.85) *p*-trend: NR Atopic asthma: Q1 (ref) Q2: 1.36 (0.74, 2.48) Q3: 1.74 (0.53–5.67) Q4: **2.47 (1.03–5.92)** *p*-trend: NR Atopic wheeze: Q1 (ref) Q2: 0.96 (0.23, 4.04) Q3: 1.88 (0.88, 4.01) Q4: 2.16 (0.54, 8.58) *p*-trend: NR Non-atopic asthma: Q1 (ref) Q2: **0.23 (0.08, 0.69)** Q3: 0.50 (0.19, 1.32) Q4: 0.28 (0.06, 1.41) *p*-trend: NR Non-atopic wheeze: Q1 (ref) Q2: **0.31 (0.11, 0.85)** Q3: **0.27 (0.08, 0.98)** Q4: 0.33 (0.08, 1.37) *p*-trend: NR

CI: confidence interval; Cr: creatinine; GM: geometric mean; IgE: Immunoglobulin E; ISAAC: International Study of Asthma and Allergies in Childhood; LOD: limit of detection; NR: not reported; OR: odds ratio; SG: specific gravity; SPT: skin prick test; TCS: triclosan.

Table 2. Summary of epidemiologic studies evaluating association between TCS exposure and measures of overweight/obesity measures (bolded text indicates a statistically significant result).

Publication; study design	Study population	Exposure assessment	Outcome definition	Analysis; measure of association	Results
Buckley et al. (2016); cohort	Mother–child pairs ($n = 173$) from The Mount Sinai Children's Environmental Health Study, USA	Single spot urine sample collected at 25–40 weeks gestation ln-transformed and Cr standardized as follows 1. ln-Cr is modeled on age, race/ethnicity, education, pre-pregnancy BMI, height 2. TCS level is divided by ratio of observed/predicted Cr 3. ln-transformed result of step 2 is used as the unit of exposure; tertiles or continuous LOD: 2.27 μg/L %>LOD: 79.8 TCS, μg/L GM =14.5 Min: <LOD 25th %ile: 2.90 75th %ile: 46.5 Max: 1790	Percent body fat (fat mass/weight ×100) at three visits between 4 and 9 years of age Obesity defined as BMI >85th sex and age-specific percentile[a] [a]Results not available	Mixed effects linear model with repeated measures adjusted for ln-Cr and urine collection calendar date; maternal race/ethnicity, age, education, work status, smoking during pregnancy, height, pre-pregnancy BMI, and adequacy of gestational weight gain; prenatal summed DEHP metabolites; breastfeeding; months of age and physical activity at follow-up; and, for overall models, child's sex; β (95% CrI) per standard deviation or by quartile	Continuous across visits Overall: 0.42 (−0.52, 1.36) Girls: 0.24 (−1.01, 1.48) Boys: 0.66 (−0.77, 2.08) Visit 1[a] Overall: 0.62 (−0.60, 1.86) Girls: 1.01 (−0.55, 2.56) Boys: 0.04 (−1.86, 1.97) Visit 2[a] Overall: 0.76 (−0.27, 1.79) Girls: 0.51 (−0.88, 1.88) Boys: 1.12 (−0.52, 2.79) Visit 3[a] Overall: 0.51 (−0.64, 1.67) Girls: 0.39 (−1.11, 1.89) Boys: 0.68 (−1.07, 2.42) By Cr-adjustment method[a] None: 0.53 (−0.50, 1.57) Adjustment: 0.50 (−0.53, 1.57)[b] Correction: 0.41 (−0.65, 1.46)[b] Standardization: 0.37 (−0.57, 1.30)[b] Adjustment + standardization: 0.42 (−0.52, 1.36)[b] Accounting for loss to follow-up[a] Overall: 0.56 (−0.41, 1.54) Girls: NR Boys: NR Tertiles across visits[a] Overall: T1 (ref) T2: –0.30 (−2.78, 2.20) T3: 0.81 (−1.69, 3.27) Girls: T1 (ref) T2: −0.55 (−3.78, 2.68) T3: 0.87 (−2.17, 3.90) Boys: T1 (ref) T2: −0.07 (−3.45, 3.26) T3: 0.80 (−2.63, 4.21) [a]From supplemental tables [b]Adjustment = inclusion for ln-Cr as a covariate; Correction = TCS expressed as μg/g Cr; Standardization: as described in the "exposure assessment" without including Cr in the model Adjustment + standardization = primary analysis

(*continued*)

Table 2. Continued

Publication; study design	Study population	Exposure assessment	Outcome definition	Analysis; measure of association	Results
Buser et al. (2014); cross sectional	Children and adolescents ($n = 1298$) included in the 2007–2010 National Health and Nutrition Examination Survey (NHANES), USA	Single spot urinary measure; quartiles Ln-transformed (data not shown) LOD: NR %>LOD: 79 TCS, ng/L, GM (SE) All: 13.19 (0.99) Normal/underweight: 13.99 (1.36) Overweight: 12.92 (1.39) Obese 11.00 (0.77)	Age- and sex-specific BMI *z*-score, WC, and obesity/overweight Overweight defined as *z*-score >85th percentile; obesity defined as BMI *z*-score ≥95th percentile	Linear models for BMI *z*-score and WC adjusted for age, sex, race/ethnicity, calorie intake, television and video game use (children), recreational activity (adolescents), serum cotinine, income level, and urinary Cr; β (95% CI) Logistic regression for overweight and obesity adjusted for the same covariates; OR (95% CI)	BMI z-score, all Q1 (ref) Q2: −0.02 (−0.20, 0.16) Q3: 0.12 (−0.12, 0.36) Q4: −0.11 (−0.32, 0.09) *p*-trend =0.21 WC, all Q1 (ref) Q2: −0.13 (−2.21, 1.95) Q3: 1.24 (−1.60, 4.08) Q4: **−2.44 (−4.53, −0.34)** *p*-trend=**0.01** BMI z-score, children (6–11 years) Q1 (ref) Q2: −0.15 (−0.50, 0.20) Q3: 0.05 (−0.30, 0.40) Q4: 0.02 (−0.26, 0.31) *p*-trend =0.47 WC, children (6–11 years) Q1 (ref) Q2: −1.05 (−3.85 1.75) Q3: 0.07 (−3.32, 3.47) Q4: −0.58 (−3.07, 1.90) *p*-trend =0.76 BMI z-score, adolescents (12–19 years) Q1 (ref) Q2: 0.08 (−0.16, 0.32) Q3: 0.21 (−0.07, 0.49) Q4: −0.14 (−0.40, 0.13) *p*-trend =0.11 WC, adolescents (12–19 years) Q1 (ref) Q2: 1.49 (−1.94, 4.91) Q3: 2.85 (−0.77, 5.46) Q4: −2.11 (−4.81, 0.59) *p*-trend=**0.03** Obesity, all Q1 (ref) Q2: 0.91 (0.59, 1.41) Q3: 1.34 (0.77, 2.35) Q4: 0.74 (0.51, 1.09) *p*-trend: NR Overweight, all Q1 (ref) Q2: 1.25 (0.68, 2.31) Q3: 1.28 (0.73, 2.24) Q4: 1.16 (0.63, 2.16) *p*-trend: NR Obesity, children (6–11 years) Q1 (ref) Q2: 0.57 (0.30, 1.09)

(*continued*)

Table 2. Continued

Publication; study design	Study population	Exposure assessment	Outcome definition	Analysis; measure of association	Results
					Q3: 1.21 (0.56, 2.62) Q4: 0.84 (0.41, 1.73) *p*-trend: NR Overweight, children (6–11 years) Q1 (ref) Q2: 1.34 (0.49, 3.70) Q3: 1.17 (0.60, 2.29) Q4: 1.24 (0.52, 2.92) *p*-trend: NR Obesity, adolescents (12–19 years) Q1 (ref) Q2: 1.36 (0.69, 2.67) Q3: 1.62 (0.78, 3.37) Q4: 0.81 (0.44, 1.50) *p*-trend: NR Overweight, adolescents (12–19 years) Q1 (ref) Q2: 1.18 (0.44, 3.18) Q3: 1.34 (0.56, 3.21) Q4: 0.99 (0.47, 2.10) *p*-trend: NR
Den Hond et al. (2013); cross sectional	Adolescents ($n = 193$) included in the Flemish Environment and Health Study (FLEHS II, 2007–2011), Belgium	Single first morning urine sample; quartiles LOQ: 0.1 μg/L %>LOD: 100 TCS, μg/L Median (P25–P75): 1.30 (0.54–4.91) Min: 0.10 Max: 706 TCS, μg/g Cr Median (P25–P75): 0.92 (0.37–3.69) Min: 0.07 Max: 882	One time BMI measure	Linear regression (not specified)	"Urinary TCS levels were not significantly associated with age, BMI, educational level, smoking or urbanization"
Geens et al. (2015); cross sectional	Overweight/obese patients ($n = 151$) undergoing weight loss treatment and lean individuals ($n = 43$) selected among staff and volunteers, Endorup trial, Belgium	For obese subjects, three 24-h urine samples collected at baseline (0 months) and at 6 and 12 months of the study; single void (second void of day) measure at 3 months For lean subjects a one-time single void measure; continuous variables LOD: NR %>LOD: >99 TCS, ng/ml, median (10th, 90th) *All* 0 *M*: 1.5 (0.3, 73) 3 *M*: 1.3 (0.3, 110) 6 *M*: 1.9 (0.3, 205) 12 *M*: 1.9 (0.3, 250)	BMI and WC at 0, 3, and 6 months and BMI only at 12 months for obese subjects. One-time measures for lean subjects	Linear regression for WC adjusted for age, weight loss (for 3 and 6 months in obese individuals) and gender (in the overall analysis) β (95% CI) Pearson correlation for log-TCS vs. BMI; correlation coefficients (*p* values)	WC 0 months all obese: 0.01 (−0.16, 0.17) 0 months obese men: −0.01 (−0.32, 0.31) 0 months obese women: −0.01 (−0.21, 0.20) 3 months all obese: NR 3 months obese men: NR 3 months obese women: NR 6 months all obese: NR 6 months obese men: NR 6 months obese women: NR All lean: 0.14 (–0.17, 0.46) Lean men: –0.24 (–0.82, 0.33) Lean women: 0.33 (–0.03, 0.68)

(*continued*)

Table 2. Continued

Publication; study design	Study population	Exposure assessment	Outcome definition	Analysis; measure of association	Results
		Lean 0.9 (0.3, 52) TCS, μg/g Cr, median (10th, 90th) *All* 0 *M*: 1.4 (0.3, 52) 3 *M*: NR 6 *M*: 2.3 (0.3, 212) 12 *M*: 3.8 (0.4, 280) *Lean* NR			BMI[a] 0 months obese: 0.011 (0.895) 3 months obese: 0.120 (0.271) 6 months obese: 0.011 (0.940) 12 months obese: −0.244 (0.146) All lean: −0.087 (0.580) [a]From supplemental table
Lankester et al. (2013); cross sectional	Adults ($n = 4037$) included in the 2003–2008 National Health and Nutrition Examination Survey (NHANES), USA	Single spot urine sample; binary (<LOD, vs. ≥LOD) and modified quartiles (<LOD and low, medium, and high by tertiles) LOD: 2.3 ng/ml %>LOD: 77.3 TCS, ng/mL Median: 11.6	One time BMI measure	Linear regression adjusted for BPA, sex, race, SES, sex-race interaction and sex-SES interaction; β (95% CI)	Binary <LOD (ref) >LOD: **0.83 (0.40, 1.27)** Modified quartiles, Cr-adjusted <LOD (ref) Low: **1.37 (0.74, 2.01)** Medium: **0.95 (0.32, 1.57)** High: 0.19 (–0.32, 0.71) Models without Cr- or SG-adjustment yielded "similar" results (no data given)
Li et al. (2015); cross sectional	Children and adults ($n = 7964$) included in the 2003–2010 National Health and Nutrition Examination Survey (NHANES), USA	Single spot urinary measure; Cr-adjusted, continuous *z*-score Age- and gender-specific quartiles LOD: 2.3 ng/ml %>LOD: 78 TCS, μg/L *GM (95% CI)* Male: 23.6 (21.3, 26.0) Female: 21.7 (18.9, 24.5) TCS, μg/g Cr *GM (95% CI)* Male: 18.7 (16.8, 20.6) Female: 20.3 (17.8, 22.9)	One time BMI and WC	Linear regression model adjusted for sex, age, race, poverty to income ratio, cotinine levels, and urinary BPA concentrations "where appropriate" β (95% CI)	BMI All 6–19-year olds: **−0.47 (−0.69, −0.26)** Boys: **−0.34 (−0.05, −0.64)** Girls: **−0.62 (−0.31, −0.94)** All adults: **−0.58 (−0.79, −0.36)** Men: **−0.42 (−0.06, −0.77)** Women: **-0.71 (–0.34, -1.07)** WC All 6–19-year olds: **−1.09 (−1.64, −0.54)** Boys: **−0.92 (−0.09, −1.74)** Girls: **−1.32 (−0.54, −2.09)** All adults: **−1.54 (−2.08, −1.00)** Men: **−1.35 (−0.48, −2.22)** Women: **−1.68 (−0.86, −2.50)** BMI by TCS quartile (6–19 year olds) *p*-trend=**0.0001** BMI by TCS quartile (adults) *p*-trend=**0.0001** WC by TCS quartile (6–19 year olds) *p*-trend=**0.001** WC by TCS quartile (adults) *p*-trend <**0.0001**

(*continued*)

Table 2. Continued

Publication; study design	Study population	Exposure assessment	Outcome definition	Analysis; measure of association	Results
Poole et al. (2016); cross-over experiment	Healthy adult volunteers ($n = 14$) randomized to 4 months of triclosan products (toothpaste, soap, dishwashing liquid) followed by non-triclosan products, or the same protocol in reverse (double blind), USA	TCS interval vs. non-TCS phase; binary LOD: NR %>LOD: 100 (TCS phase), 72 (nTCS phase) TCS, pg/μl Median (Q1–Q3) at baseline: 24.4 (7.2–59.4) Median at end of TCS phase: 1548 Median at end of non-TCS phase: 14.6	Weight gain (yes vs. no) during TCS vs. non-TCS intervals	Crude OR (95% CI)	5.0 (0.6, 236)
Wolff et al. (2015); cohort	Girls ($n = 1170$) included in the Breast Cancer and Environment Research Program (BCERP) Puberty Study, USA	Single urine sample collected at baseline (6–8 years of age); Cr-adjusted, continuous LOD = NR %>LOD: >80 TCS, μg/g Cr GM (95% CI): 18.1 (16.5, 19.8) TCS, μg/g Cr 10th, 90th %iles: 3, 174	BMI percentile at last B1 and last PH1 visit	ANOVA comparing TCS across BMI categories; GMs (95% CI) for each category	TCS by BMI percentile at B1 <50th: 19.8 (16.9, 23.2) 50–85th: 17.4 (14.7, 20.4) ≥85th: 17.4 (14.8, 20.3) Not significant TCS by BMI percentile at PH1 <50th: 19.5 (16.6, 22.9) 50–85th: 17.3 (14.7, 20.3) ≥85th: 17.7 (15.1, 20.7) Not significant
Xue et al. (2015); cross-sectional[a] [a]Authors characterize this as a "prospective case-control study", but all cases represent prevalent obesity	Overweight or obese ($n = 49$) and non-obese ($n = 27$) children recruited from an endocrinology clinic, India	Single spot urine sample; Cr-adjusted and unadjusted, continuous LOQ: 0.10 ng/ml %>LOD: 100 TCS, ng/ml Range: 0.220–2570 GM ± SD: 9.55 ± 314 AM: 88.3 TCS, μg/g Cr Range: 0.150–1710 GM ± SD: 6.33 ± 208 AM: 58.5	Overweight defined as BMI >85th percentile; obesity defined as BMI >95th percentile Non-obese "control" children had BMI at 50th percentile	*t*-Test comparing mean TCS in overweight/obese and non-obese children; means ± SD for each category (*p* value) Logistic regression adjusted for age, gender, family income, parental education, antenatal complications, maternal pre-eclampsia, physical activity, fast-food frequency, and beverage frequency; OR (95% CI)	t-Test not adjusted for Cr Cases: 47.5 ± 78.1 Controls: 162 ± 513 (0.89) t-Test Cr-adjusted Cases: 31.5 ± 51.8 Controls: 108 ± 340 (0.93) Logistic regression NR

AM: arithmetic mean; B1: stage 1 of breast development; BMI: body mass index; BPA: bisphenol A; CI: confidence interval; Cr: creatinine; CrI: credible interval (between 2.5th and 97.5th percentiles of the posterior distribution); GM: geometric mean; LOD: limit of detection; LOQ: limit of quantitation; NR: not reported; OR: odds ratio; PH1: stage 1 of pubic hair development; SD: standard deviation; TCS: triclosan; WC: waist circumference.

Table 3. Summary of epidemiologic studies evaluating association between TCS exposure and female reproductive function (bolded text indicates a statistically significant result).

Publication; study design	Study population	Exposure assessment	Outcome definition	Analysis; measures of association	Results
Smarr et al. (2017); cohort	Female partners in 501 reproductive-age couples enrolled in the Longitudinal Investigation of Fertility and the Environment (LIFE) study, USA	Single spot urine sample at baseline (within 2 months of stopping contraception); binary (above vs. below LOD) and quartiles and ln-transformed continuous LOQ: 2 μg/g Cr %>LOQ: 93 TCS, μg/g Cr Median (IQR): 16.8 (5.32, 67.5)	Time to pregnancy	Cox proportional odds model adjusted for age, difference between partners' age, each partner's Cr, BMI, serum cotinine, race/ethnicity, and income, and male TCS FOR (95% CI)	Binary 1.01 (0.95–1.06) Quartiles (not adjusted for male TCS)[a] Q1: 1.0 (ref) Q2: 0.94 (0.66, 1.35) Q3: 0.83 (0.57, 1.22) Q4: 0.86 (0.59, 1.25) *p*-trend =0.35 Quartiles (adjusted for male TCS)[a] Q1: 1.0 (ref) Q2: 0.95 (0.66, 1.38) Q3: 0.82 (0.54, 1.23) Q4: 0.83 (0.53, 1.30) *p*-trend =0.33 "No associations were observed with couple fecundity when chemicals were modeled continuously." [a]From supplemental tables
Velez et al. (2015); cross-sectional	Pregnant women ($n = 1699$; 1583 in final model) recruited into the Maternal-Infant Research on Environmental Chemicals Study, Canada	Single spot urine sample during the 1st trimester of pregnancy; log transformed continuous (rescaled by SD), quartiles. Binary (cutoff at 75th percentile) LOD: 0.12 ng/ml %>LOD: 99.9 TCS, ng/ml %<LOD: 0.1 Median: 8.3 Min: <LOD Max: 6784 GM (95% CI): 11.93 (10.67, 13.4) 75th%: 71.7	Time to pregnancy (by recall)	Cox proportional odds model adjusted for urine SG, maternal age, maternal smoking, education, income and BMI; FOR (95% CI)	Continuous 0.94 (0.88, 1.01) Quartiles Q1: 1.0 (ref) Q2: 1.10 (0.91, 1.31) Q3: 1.10 (0.92, 1.32) Q4: 0.89 (0.74, 1.07) *p*-trend: NR Binary **0.84 (0.72, 0.97)**
Wang et al. (2015); cross-sectional (prevalence case-control)	Cases ($n = 113$) women with mid-gestation history of SA, controls ($n = 339$) women with no history of SA, all participants recruited from Nanjing Medical University hospitals, NJMU Birth Cohort, China	Single urine sample; collected within 12 hours after SA "low-median- high" (definitions not given) LOD: 0.9 ng/ml %>LOD: 57.52 (cases), 32.74 (controls) TCS, ng/ml *Cases* GM (95% CI): 2.50 (1.49, 3.51) 50th: 1.40 75th: 5.43 90th: 37.39 95th: 95.94 *Controls* GM (95% CI): 0.99 (0.86, 1.15) 50th: <LOD 75th: 1.65 90th: 8.28 95th: 23.56	History of SA (yes vs. no)	Logistic regression adjusted for BMI, drinking status and Cr OR (95% CI)	"Low chemical TCS": 1.0 (ref) "Median chemical TCS": **2.00 (1.08, 3.70)** "High chemical TCS": **2.36 (1.29, 4.34)** *p*-trend=**0.003**

CI: confidence interval; Cr: creatinine; FOR: fecundability odds ratio; IQR: interquartile range; LOD: limit of detection; LOQ: limit of quantitation; NR: not reported; OR: odds ratio; SA: spontaneous abortion; SG: specific gravity; TCS: triclosan.

Table 4. Summary of epidemiologic studies evaluating association between TCS exposure and male reproductive function (bolded text indicates a statistically significant result).

Publication; study design	Study population	Exposure assessment	Outcome definition	Analysis; measures of association	Results
Chen et al. (2013); cross-sectional (prevalence case-control)	Infertile men ($n = 877$) compared to fertile controls ($n = 713$), recruited from hospitals in China	Single urine sample; low (<LOD), medium (<median), high with and without Cr (as continuous variable) adjustment LOD: 0.9 ng/ml %>LOD: ~48 TCS, ng/ml Control / Case %>LOD 47.8 / 48.0 GM 1.593 / 1.707 50th < LOD / < LOD 75th 3.960 / 5.014 90th 31.727 / 32.219 95th 57.466 / 65.916	Any infertility (case); Low semen volume (<1.5 ml) Low sperm concentration ($<15 \times 10^6$/ml) Low sperm number per ejaculate ($<39 \times 10^6$)	Logistic regression for each outcome compared to controls, adjusted for age, BMI and Cr[a]; OR (95% CI) [a]If TCS level is not Cr-adjusted	Any infertility, 877 cases[a] Low: 1.0 (ref) Medium: 0.93 (0.72, 1.19) High: 1.23 (0.96, 1.59) p-trend =0.166 Any infertility (Cr-adjusted TCS)[a] Low: 1.0 (ref) Medium: 1.03 (0.81, 1.32) High: **1.31 (1.02, 1.68)** p-trend= **0.033** Low semen volume, 43 cases Low: 1.0 (ref) Medium: 0.59 (0.24, 1.43) High: 1.27 (0.59, 2.70) p-trend =0.776 Low semen volume (Cr-adjusted TCS)[a] Low: 1.0 (ref) Medium: 0.47 (0.20, 1.07) High: 0.93 (0.45, 1.93) p-trend =0.713 Low sperm concentration, 170 cases Low: 1.0 (ref) Medium: 0.84 (0.55, 1.29) High: 1.22 (0.80, 1.86) p-trend =0.500 Low sperm concentration (Cr-adjusted TCS)[a] Low: 1.0 (ref) Medium: 0.98 (0.64, 1.48) High: 1.14 (0.75, 1.73) p-trend =0.564 Low total sperm number per ejaculate, 182 cases Low: 1.0 (ref) Medium: 0.71 (0.46, 1.09) High: 1.06 (0.70, 1.60) p-trend =0.904 Low total sperm number per ejaculate (Cr-adjusted TCS)[a] Low: 1.0 (ref) Medium: 0.96 (0.64, 1.45) High: 1.05 (0.69, 1.58) p-trend =0.841 [a]From supplemental tables
Den Hond et al. (2015); cross-sectional (prevalence case-control)	Cases, men with low TMC ($n = 40$); controls, men with normal TMC ($n = 80$), all recruited from fertility clinics in Belgium	Single urinary measure; expressed as Interquartile interval; ln-transformed continuous LOD: NR %>LOD: NR TCS, µg/L, GM (P25, P75)	Low TMC defined using $<20 \times 10^6$ cutoff; sperm concentration (10^6/ml), sperm motility (%), sperm morphology (% normal), all measured twice	Logistic regression for interquartile TCS change and case/control status adjusted for age, BMI, smoking, ETS exposure and educational level and Cr; OR (95% CI)	Low TMC[b] 0.99 (0.62–1.59) Sperm concentration[a] −0.08 (SE: 0.08); $p = 0.33$ Sperm motility[a] −0.45 (1.20); $p = 0.71$

(*continued*)

Table 4. Continued

Publication; study design	Study population	Exposure assessment	Outcome definition	Analysis; measures of association	Results
		Controls: 2.8 (0.5, 5.5) Cases: 2.6 (0.5, 15.6)		Mixed linear regression for ln-transformed TCS and sperm quality measures adjusted for age, BMI, smoking, Cr, and for concentration and motility only, period of abstinence, volume of ejaculation, pH of sperm; β (SE), p value	Sperm morphology[a] +0.01 (0.13); $p = 0.99$ [a]From supplemental table [b]"Subfertility" defined two ways: no reproduction in past 12 months, and increased subfertility in men, defined on the basis of semen parameters
Smarr et al. (2017); cohort	Male partners in 501 reproductive-age couples enrolled in the Longitudinal Investigation of Fertility and the Environment (LIFE) study, USA	Single spot urinary measure at baseline (within 2 months of stopping contraception); binary (above vs. below LOD) and quartiles and continuous; ln-transformed LOQ: 2 μg/g Cr %>LOQ: 87 TCS, μg/g Cr Median (IQR): 16.2 (4.41, 64.4) TCS, ng/ml Median (IQR): 17.8 (4.42, 77.1)	Time to pregnancy	Cox proportional odds model adjusted for female age and TCS, difference between partners' age, and both partner's Cr, BMI, serum cotinine, race/ethnicity, and income; FOR (95% CI)	Binary 1.01 (0.94–1.07) Quartiles (not adjusted for female TCS)[a] Q1: 1.0 (ref) Q2: 0.97 (0.66, 1.41) Q3: 0.95 (0.65, 1.38) Q4: 1.02 (0.69, 1.51) p-trend =0.95 Quartiles (adjusted for female TCS)[a] Q1: 1.0 (ref) Q2: 1.12 (0.74, 1.69) Q3: 1.04 (0.68, 1.58) Q4: 1.21 (0.76, 1.92) p-trend =0.49 "No associations were observed with couple fecundity when chemicals were modeled continuously." [a]From supplemental tables
Zhu et al. (2016); cross-sectional	Volunteers ($n = 471$) recruited from male reproductive health clinics in China	Single spot urinary measure Cr-adjusted and ln-transformed, tertiles and continuous LOD: 0.1 μg/L %>LOD: 96.4 TCS, ng/ml GM: 1.12 10th: 0.20 25th: 0.50 50th: 1.12 75th: 3.38 90th: 14.35 95th: 33.37 Max: 98.01 TCS, ng/mg Cr GM: 0.99 10th: 0.21 25th: 0.41 50th: 0.97 75th: 2.95 90th: 12.23 95th: 21.13 Max: 131.22	Semen volume (mL) Concentration (10^6/mL) Total sperm count (10^6) Sperm motility: percent forward moving (%) Number of forward moving (10^6) Morphology: Normal sperm morphology (%) Number of normal sperm (10^6) VAP (μm/s)	Linear regression adjusted for age, BMI, abstinence period, education, income, current smoking, drinking and exposure to other chemicals or heavy metals; regression coefficient (95% CI)	Sperm volume T1: −0.16 (−0.63, 0.32) T2: −0.28 (−1.07, 0.52) T3: −0.24 (−0.50, 0.03) Continuous: −0.20 (−0.43, 0.02) Concentration T1: **−0.21 (−0.41, −0.01)** T2: 0.15 (−0.19, 0.49) T3: 0.04 (−0.09, 0.16) Continuous: −0.01 (−0.11, 0.08) Total sperm count T1: **−0.25 (−0.48, −0.02)** T2: 0.03 (−0.32, 0.39) T3: −0.06 (−0.19, 0.08) Continuous: −0.09 (−0.20, 0.01) Motility: Percent forward moving T1: −0.99 (−4.74, 2.76) T2: 0.66 (−6.42, 7.74) T3: −1.55 (−3.94, 0.84)

(*continued*)

Table 4. Continued

Publication; study design	Study population	Exposure assessment	Outcome definition	Analysis; measures of association	Results
		TCS, ng/mg Cr T1: <0.66 T2: 0.66–2.33 T3: ≥2.33			Continuous: −1.19 (−3.11, 0.72) Motility: Number of forward moving T1: **−0.35 (−0.68, −0.03)** T2: −0.18 (−0.72, 0.35) T3: −0.12 (−0.28, 0.04) Continuous: **−0.17 (−0.32, −0.02)** Morphology: Percentage of normal T1: **−1.64 (−3.05, −0.23)** T2: **2.76 (0.25, 5.28)** T3: −0.13 (−0.96, 0.70) Continuous: −0.39 (−1.09, 0.32) Morphology: Number of normal T1: **−0.48 (−0.80, −0.16)** T2: 0.25 (−0.24, 0.74) T3: −0.06 (−0.23, 0.12) Continuous: −0.14 (−0.28, 0.01) VAP T1: −0.24 (−1.03, 0.56) T2: 0.36 (−1.02, 1.75) T3: 0.02 (−0.46, 0.51) Continuous: 0.03 (−0.36, 0.42)

CI: confidence interval; Cr: creatinine; ETS: environmental tobacco smoke; FOR: fecundability odds ratio; GM: geometric mean; LOD: limit of detection; LOQ: limit of quantitation; OR: odds ratio; TCS: triclosan; TMC: total motility count, calculated as sample volume × sperm concentration × total percentage of grade A and B motility spermatozoa; VAP: average path velocity.

Table 5. Summary of epidemiologic studies evaluating association between TCS exposure and birth outcomes (bolded text indicates a statistically significant result).

Publication; study design	Study population	Exposure assessment	Outcome definition	Analysis; measure of association	Results
Chevrier et al. (2012); case-control	Newborns with hypospadias ($n = 21$) or undescended testis ($n = 50$) each matched to three controls of the same residential area, and both gestational age and date of prenatal (maternal) urine collection; selected from EDEN and PELAGIE cohorts, France	Single morning urine sample collected prenatally; tertiles LOD: 2.3 μg/L %>LOD: 83.2 TCS, μg/L after correction for sampling conditions[a] Median: 17 Max: 1320 TCS, μg/L T1: <4 T2: 4–51 T3: ≥51 [a]From supplemental table	Hypospadias or undescended testis identified within the first days of life	Conditional logistic regression adjusted for maternal age, parity, educational level, gestational age, Cr; OR (95% CI)	Undescended testis, 38 cases, 113 controls T1: (ref) T2: 0.59 (0.2, 1.7) T3: 0.74 (0.3, 2.1) Undescended testis (term births only), 35 cases, 104 controls[a] T1: (ref) T2: 0.56 (0.2, 1.7) T3: 0.62 (0.2, 1.9) Hypospadias, 10 cases, 30 controls T1: (ref) T2: 3.33 (0.40, 28.95)[b] T3: 2.22 (0.29, 18.37)[b] [a]From supplemental table [b]Calculated from data provided in the article (95% CI computed based on Fisher's exact test)
Geer et al. (2017); cohort and cross-sectional	Pregnant women ($n = 185$) and their newborns recruited at a prenatal clinic, USA	Single spot maternal urine sample, collected prenatally ($n = 185$) and single cord plasma level ($n = 38$); log-transformed, Cr-corrected LOD: NR %>LOD: 100 Urine TCS, μg/L[a] Min: 0.16 Mean: 166.69 Max: 11909.3 Cord plasma TCS, μg/L[a] Min: 0.16 Mean: 8.38 Max: 127.6 [a]From Supplement	BW, GA, BL, HC – all continuous LBW (<2500 g) and PTB (<37 week) – both binary	Linear regression models for continuous outcomes adjusted for maternal nativity, age, alcohol and tobacco use, neonate gender, TCC, and BuPB; β (95% CI) Logistic regression for binary outcomes adjusted for neonate gender; OR (95% CI)	Maternal TCS – continuous outcomes BW: NR GA: 0.22 (−0.16, 0.60) BL: 0.39 (−0.24, 1.02) HC: 0.16 (−0.16, 0.48) Cord blood TCS – continuous outcomes BW: −64.93 (−354.67, 224.81) GA: 0.37 (−0.78, 1.53) BL: 0.65 (−1.21, 2.51) HC: −0.09 (−0.87, 0.70) Maternal TCS – binary outcomes LBW: 0.72 (0.39, 1.32) PTB: 0.95 (0.60, 1.50) Cord blood TCS – binary outcomes LBW: 1.32 (0.33, 5.32) PTB: 1.40 (0.41, 4.77)
Lassen et al. (2016); cohort	Mother–child pairs ($n = 514$) selected from the Odense Child Cohort study, Denmark	Single morning maternal urine sample collected around 28 weeks gestation; osmolality-corrected; quartiles and continuous, $\log_2$-transformed LOD: NR %>LOD: 83 TCS, ng/ml Median: 0.88 95th %ile: 428 Max: 2614	BW, BL, HC, AC; AGDs, AGDl, penile width (boys only) at 3 months of age	Linear regression models for BW, BL, HC, and AC adjusted for GA, maternal smoking, parity and pre-pregnancy BMI. Linear models for AGD and penile width adjusted for weight for age SD-score and post-conceptional age; β (95% CI)	BW in boys: Q1: (ref) Q2: −62 (−199, 74) Q3: 29 (−108, 166) Q4: −81 (−218, 56) p-trend =0.49 Continuous: −5.2 (−18.5, 8.2) BW in girls: Q1: (ref) Q2: 50 (−90, 190) Q3: 25 (−116, 167) Q4: 36 (−104, 176)

(*continued*)

Table 5. Continued

Publication; study design	Study population	Exposure assessment	Outcome definition	Analysis; measure of association	Results
					p-trend =0.71 Continuous: 4.7 (−7.3, 17.0) BL in boys: Q1: (ref) Q2: 0.1 (−0.5, 0.7) Q3: **0.7 (0.1, 1.3)** Q4: 0.1 (−0.5, 0.8) *p*-trend =0.29 Continuous: 0.01 (−0.05, 0.07) BL in girls: Q1: (ref) Q2: −0.5 (−1.1, 0.1) Q3: −0.4 (−1.0, 0.2) Q4: −0.2 (−0.8, 0.4) *p*-trend =0.66 Continuous: −0.004 (−0.06, 0.05) HC in boys: Q1: (ref) Q2: −0.2 (−0.7, 0.4) Q3: −0.3 (−0.9, 0.2) Q4: **−0.7 (−1.2, −0.1)** *p*-trend=**0.01** Continuous: **−0.06 (−0.11, −0.002)** HC in girls: Q1: (ref) Q2: −0.4 (−0.9, 0.2) Q3: −0.3 (−0.8, 0.2) Q4: −0.2 (−0.7, 0.3) *p*-trend =0.56 Continuous: −0.01 (−0.05, 0.04) AC in boys: Q1: (ref) Q2: −0.3 (−0.9, 0.3) Q3: −0.4 (−1.0, 0.2) Q4: −0.6 (−1.2, 0.0) *p*-trend =0.07 Continuous: −0.05 (−0.11, 0.01) AC in girls: Q1: (ref) Q2: 0.3 (−0.4, 0.9) Q3: 0.1 (−0.6, 0.8) Q4: −0.03 (−0.7, 0.6) *p*-trend =0.81 Continuous: 0.00 (−0.06, 0.06) AGDs in boys: Q1: (ref) Q2: −0.9 (−2.8, 0.9) Q3: **−2.3 (−4.1, −0.4)**

(*continued*)

Table 5. Continued

Publication; study design	Study population	Exposure assessment	Outcome definition	Analysis; measure of association	Results
					Q4: −1.3 (−3.1, 0.6) *p*-trend =0.08 Continuous: −0.16 (−0.34, 0.02) AGDs in girls: Q1: (ref) Q2: −0.5 (−2.0, 1.0) Q3: −0.4 (−1.8, 1.1) Q4: −0.3 (−1.8, 1.1) *p*-trend =0.69 Continuous: −0.02 (−0.15, 0.10) AGDI in boys: Q1: (ref) Q2: 0.3 (−1.7, 2.3) Q3: −1.2 (−3.3, 0.8) Q4: −1.3 (−3.4, 0.8) *p*-trend =0.11 Continuous: −0.20 (−0.39, 0.00) AGDI in girls: Q1: (ref) Q2: −0.6 (−2.4, 1.2) Q3: 0.1 (−1.6, 1.9) Q4: 0.1 (−1.7, 1.8) *p*-trend =0.79 Continuous: 0.01 (−0.15, 0.17) Penile width (boys only) Q1: (ref) Q2: 0.0 (−0.4, 0.5) Q3: −0.1 (−0.5, 0.3) Q4: −0.2 (−0.6, 0.3) *p*-trend =0.32 Continuous: −0.02 (−0.06, 0.02)
Philippat et al. (2012); cohort	Newborns with hypospadias or undescended testis each matched to three controls (total $n = 191$) of the same residential area, and both gestational age and date of prenatal (maternal) urine collection; selected from EDEN cohort, France	Single urine sample collected prenatally ln-transformed and standardized for conditions of sampling including hour of sampling, time between sample collections and freezing, season and day of sampling, and GA at collection; tertiles LOD: 2.3 μg/L %>LOD: 84.1 TCS, μg/L 5th %ile: 1.6 50th %ile: 24.1 95th %ile: 634.0	BW, BL, HC from maternity records	Linear regression models "corrected for overrepresentation of cases … by a weighting approach" adjusted for gestational duration, maternal pre-pregnancy weight and height, smoking, education and parity, recruitment center, and Cr. Models for HC also adjusted for delivery mode; β (95% CI)	BW all subjects T1: (ref) T2: –58 (–194, 78) T3: –40 (–171, 90) *p*-trend =0.68 Continuous: –6 (–31, 19) BW controls only[a] T1: (ref) T2: –74 (–236, 86) T3: –22 (–180, 136) *p*-trend =0.91 Continuous: –4 (–36, 28) BW all subjects non-standardized TCS[a] T1: (ref) T2: –38 (–173, 96) T3: –27 (–159, 106) *p*-trend =0.86 Continuous: –9 (–34, 17)

(*continued*)

Table 5. Continued

Publication; study design	Study population	Exposure assessment	Outcome definition	Analysis; measure of association	Results
					BL all subjects T1: (ref) T2: −0.2 (−0.9, 0.4) T3: 0.1 (−0.6, 0.7) *p*-trend =0.68 Continuous: 0.0 (−0.1, 0.1) BL controls only T1: (ref) T2: NR T3: NR *p*-trend = NR Continuous: NR BL all subjects non-standardized TCS T1: (ref) T2: NR T3: NR *p*-trend = NR Continuous: NR HC all subjects T1: (ref) T2: 0.2 (−0.3, 0.7) T3: −0.2 (−0.7, 0.3) *p*-trend =0.31 Continuous: −0.1 (−0.2, 0.0) HC controls only[a] T1: (ref) T2: 0.2 (−0.4, 0.8) T3: −0.2 (−0.8, 0.4) *p*-trend =0.35 Continuous: −0.1 (−0.2, 0.1) HC all subjects non-standardized TCS[a] T1: (ref) T2: 0.2 (−0.3, 0.7) T3: −0.3 (−0.8, 0.2) *p*-trend =0.09 Continuous: −0.1 (−0.2, 0.0) [a]From supplemental table
Philippat et al. (2014); cross sectional and cohort	Mother–child pairs limited to boys ($n = 520$) selected from the EDEN cohort, France	Single spot morning (at home or at hospital) urine sample during gestation; ln-transformed and standardized for conditions of sampling including hour and day of sampling, gestational age at collection, duration of storage at room temperature, and Cr level; tertiles (results NR) LOD: 2.3 μg/L %>LOD: 81 TCS, μg/L 5th %ile: <LOD 50th %ile: 30 95th %ile: 755	Fetal BD, HC, FL, length, AC, weight; Height and weight for newborns (from hospital records) and at 1 and 3 years; Child weight and height at 4, 8, 12, 24, 36 months, HC, AC at 3 years[a]	Random effect linear (to account for mother–son pair). All models adjusted for maternal and paternal height, pre-pregnancy weight, maternal smoking and education, recruitment center, GA at measurement, parity. Model for HC at birth also adjusted for measurement age. Models for postnatal growth also adjusted for breastfeeding duration; increase in 1 IQR of ln-transformed phenol-standardized concentrations; β (95% CI)	BD 1st u/s: 0.16 (−0.17, 0.48) 2nd u/s: −0.15 (−0.53, 0.23) 3rd u/s: −0.42 (−0.89, 0.04) Rate 1st to 2nd u/s: −0.03 (−0.07, 0.00) Rate 2nd to 3rd u/s: −0.03 (−0.07, 0.02) BD non-standardized TCS[a] 1st u/s: 0.18 (−0.1, 0.49) 2nd u/s: −0.10 (−0.5, 0.27) 3rd u/s: −0.38 (−0.8, 0.08) Rate 1st to 2nd u/s: NR Rate 2nd to 3rd u/s: NR HC 2nd u/s: 0.06 (−1.23, 1.35)

(*continued*)

Table 5. Continued

Publication; study design	Study population	Exposure assessment	Outcome definition	Analysis; measure of association	Results
			[a]The results in the article are shown for 6 months and not for 4 or 8 months		3rd u/s: −1.21 (−2.69, 0.28) Birth: −1.20 (−2.63, 0.33) 36 months: −1.02 (−2.80, 0.76) Rate 2nd to 3rd u/s: −0.07 (−0.23, 0.09) Rate 3rd u/s to birth: −0.07 (−0.44, 0.30) Rate birth to 36 months: NR HC non-standardized TCS[a] 2nd u/s: 0.10 (−1.2, 1.36) 3rd u/s: −1.17 (−2.6, 0.28) Birth: −1.10 (−2.6, 0.34) 36 months: −0.95 (−2.7, 0.79) Rate 2nd to 3rd u/s: NR Rate 3rd u/s to birth: NR Rate birth to 36 months: NR FL 2nd u/s: −0.01 (−0.30, 0.28) 3rd u/s: −0.32 (−0.65, 0.01) Rate 2nd to 3rd u/s: −0.01 (−0.05, 0.02) FL non-standardized TCS[a] 2nd u/s: 0.01 (−0.3, 0.29) 3rd u/s: −0.31 (−0.6, 0.01) Rate 2nd to 3rd u/s: NR Length Birth: −0.16 (−2.60, 2.27) 6 months: 0.00 (−2.71, 2.72) 12 months: −0.22 (−3.40, 2.96) 24 months: −0.17 (−4.01, 3.66) 36 months: 0.19 (−4.40, 4.79) Rate 0–6 months: −0.05 (−0.15, 0.06) Rate 6–12 months: −0.01 (−0.07, 0.04) Rate 12–24 months: −0.01 (−0.03, 0.02) Rate 24–36 months: 0.00 (−0.02, 0.03) Length non-standardized TCS[a] Birth: −0.18 (−2.6, 2.20) 6 months: 0.02 (−2.6, 2.68) 12 months: −0.19 (−3.3, 2.92) 24 months: −0.20 (−3.9, 3.54) 36 months: 0.09 (−4.4, 4.58) Rate 0–6 months: NR Rate 6–12 months: NR Rate 12–24 months: NR

(*continued*)

Table 5. Continued

Publication; study design	Study population	Exposure assessment	Outcome definition	Analysis; measure of association	Results
					Rate 24–36 months: NR AC 2nd u/s: 0.17 (−1.42, 1.77) 3rd u/s: **−2.39 (−4.31, −0.47)** 36 months: 2.66 (−1.21, 6.53) Rate 2nd to 3rd u/s: **−0.21 (−0.40, −0.01)** Rate 3rd u/s to birth (or 36 months): NR AC non-standardized TCS[a] 2nd u/s: 0.26 (−1.3, 1.82) 3rd u/s: **−2.28 (−4.2, −0.41)** 36 months: 2.41 (−1.4, 6.20) Rate 2nd to 3rd u/s: NR Rate 3rd u/s to birth (or 36 months): NR Weight 2nd u/s: −1.91 (−24.0, 20.2) 3rd u/s: **−35.2 (−61.8, −8.58)** Birth: 4.60 (−49.0, 58.3) 6 months: 17.5 (−95.7, 131) 12 months: 21.9 (−120, 164) 24 months: 65.6 (−114, 245) 36 months: 119 (−103, 340) Rate 2nd to 3rd u/s: **−3.26 (−6.12, −0.41)** Rate 3rd u/s to birth: **7.03 (0.33, 13.7)** Rate 0–6 months: −0.85 (−4.28, 2.58) Rate 6–12 months: 0.05 (−1.93, 2.04) Rate 12–24 months: 0.53 (−0.67, 1.72) Rate 24–36 months: 0.71 (−0.54, 1.97) Weight non-standardized TCS[a] 2nd u/s: −0.59 (−22.1, 21.0) 3rd u/s: **−34.4 (−60.4, −8.45)** Birth: 5.57 (−46.9, 58.0) 6 months: 16.2 (−94.1, 127) 12 months: 21.2 (−117, 159) 24 months: 62.6 (−112, 237) 36 months: 112 (−104, 328) Rate 2nd to 3rd u/s: NR Rate 3rd u/s to birth: NR Rate 0–6 months: NR Rate 6–12 months: NR Rate 12–24 months: NR Rate 24–36 months: NR Post-natal weight adjusted for BW[a]

(*continued*)

Table 5. Continued

Publication; study design	Study population	Exposure assessment	Outcome definition	Analysis; measure of association	Results
					6 months: 1.17 (–105, 108) 12 months: 5.58 (–130, 142) 24 months: 49.3 (–125, 224) 36 months: 102 (–115, 319) Rate 6–12 months: NR Rate 12–24 months: NR Rate 24–36 months: NR Post-natal weight adjusted for height[a] 6 months: 16.4 (–75.0, 108) 12 months: 25.7 (–86.4, 138) 24 months: 68.4 (–71.6, 208) 36 months: 113 (–58.4, 285) Rate 6–12 months: NR Rate 12–24 months: NR Rate 24–36 months: NR [a]From supplemental table
Wolff et al. (2008); cohort	Mother–child pairs ($n = 367$) included in the Children's Environmental Health Study, USA	Single urine sample collected prenatally; ln-transformed and restricted to observations with Cr $\geq$20 mg/dL; Continuous; tertiles with Cr-corrected values LOD: 2.27 µg/L %>LOD: 77.4 TCS, µg/L Min: LOD 25th: 2.9 50th: 11 75th: 42 Max: 1790	BW, BL, HC, GA	Generalized linear models adjusted for race, infant sex, GA (except models with GA as outcome), ln-Cr, maternal smoking and education, marital status and pre-pregnancy BMI; β (95% CI)	BW Overall model: –11 (–34, 11) Interaction with sex: –23 (–68, 22) BL Overall model: –0.04 (–0.17, 0.08) Interaction with sex: –0.26 (–0.51, 0.001) HC Overall model: –0.04 (–0.13, 0.04) Interaction with sex: NR GA Overall model: 0.00 (–0.10, 0.09) Interaction with sex: NR

AC: abdominal circumference; AGDI: long anogenital distance from center of anus to either the top of clitoris (girls) or penile base (boys); AGDs: short anogenital distance from center of anus to either posterior fourchette of vagina (in girls) or posterior base of scrotum (boys); BD: biparietal diameter; BL: birth length; BuPB: butylparaben; BW: birth weight; CI: confidence interval; Cr: creatinine; FL: femoral length; GA: gestational age; HC: head circumference; IQR: interquartile range; LBW: low birth weight; LOD: limit of detection; NR: not reported; OR: odds ratio; PTB: preterm birth; TCC: triclocarban; TCS: triclosan; u/s: ultrasound.

Table 6. Summary of epidemiologic studies evaluating association between TCS exposure and puberty outcomes in girls (bolded text indicates a statistically significant result).

Publication; study design	Study population	Exposure assessment	Outcome definition	Analysis; measure of association	Results
Buttke et al. (2012); cross sectional	Girls 12–16 years of age ($n = 440$) included in the 2003–2008 National Health and Nutrition Examination Survey (NHANES), USA	Single spot urine sample; ln-transformed Cr-corrected, continuous LOD: NR %>LOD: at least 75 TCS, µg/g GM (95% CI): 12.4 (8.98, 17.2)	Age in years of menarche reported by a parent or guardian	Cox proportional hazards model adjusted for BMI (or BMI percentile) and race/ethnicity; HR (95% CI)	Adjusted for BMI 0.97 (0.90, 1.05) Adjusted for BMI percentile 1.0 (0.91, 1.09)
Wolff et al. (2010); cohort	Girls 6–8 years of age at enrollment ($n = 1151$) included in The Breast Cancer and Environment Research Centers (BCERC) study, USA	Single urinary measure at baseline (6–8 years of age); continuous log-TCS with Cr in model; Cr-adjusted, quintiles LOD: 2.3 µg/l %>LOD: 82.2 TCS, µg/L[a] Median: 11.4 Min: LOD Max: 4550 TCS, µg/g Cr[a] Median: 14.7 Min: LOD Max: 2535 [a]From Supplement	Evidence of B2+ and PH2+ at 1 year follow up	Poisson regression, adjusted for age, race/ethnicity, BMI, guardian education, season of urine collection, and study site; PRs (95% CI)	B2± by TCS quintile (n=981) Q1: (ref) Q2: 0.99 (0.93, 1.05) Q3: 0.97 (0.92, 1.03) Q4: 1.01 (0.95, 1.07) Q5: 1.03 (0.97, 1.10) *p*-trend =0.242 PH2± by TCS quintile (n=963) Q1: (ref) Q2: **0.93 (0.87, 0.99)** Q3: **0.93 (0.87, 0.99)** Q4: **0.89 (0.83, 0.94)** Q5: 0.94 (0.88, 1.00) *p*-trend=**0.02** B2± for continuous TCS NR PH2± for continuous TCS NR
Wolff et al. (2015); cohort	Girls 6–8 years of age at enrollment ($n = 1157$)[a] included in the Breast Cancer and Environment Research Program (BCERP) Puberty Study, USA[b] [a]From supplemental tables [b]Extension of Wolff et al. (2010)	Single urinary measure at baseline (6–8 years of age); continuous ln-TCS and Cr-adjusted, quintiles LOD: NR %>LOD: >80 TCS, µg/g Cr GM: 18 10th %ile: 3 90th %ile: 174	Time to B2 and time to PH2 during 7 years of follow up	Cox proportional hazards model adjusted for ln-Cr (for continuous ln-TCS only), race/ethnicity and caregiver education; HR (95% CI)	Age at B2 for continuous ln-TCS (n=1129) **1.05 (1.01, 1.09)** Age at B2 for Cr-adjusted TCS quintiles (n=1128) Q1: (ref) Q2: 0.96 (0.79–1.17) Q3: 1.05 (0.86–1.28) Q4: 1.17 (0.96–1.42) Q5: 1.17 (0.96–1.43) *p*-trend=**0.027** Age at PH2 for continuous ln-TCS (n=1125) 1.01 (0.97, 1.05) Age at PH2 for Cr-adjusted TCS quintiles NR

B2: stage 2 of breast development; BMI: body mass index; CI: confidence interval; Cr: creatinine; GM: geometric mean; HR: hazard ratio; LOD: limit of detection; NR: not reported; PH2: stage 2 of pubic hair development; PR: prevalence ratio; TCS: triclosan.

Table 7. Summary of epidemiologic studies evaluating association between TCS exposure and sex hormone concentrations (bolded text indicates a statistically significant result).

Publication; study design	Study population	Exposure assessment	Outcome definition	Analysis; measures of association	Results
Aker et al. (2016); cohort and cross sectional	Pregnant women ($n = 106$) from an ongoing prospective cohort (Puerto Rico Test site for Exploring Contamination Threats – PROTECT), hospital- and prenatal clinic-based; Puerto Rico	Three spot urine samples at 16–20; 20–24; and 24–28 weeks gestation; specific gravity-adjusted; ln-transformed, continuous LOD: NR %>LOD: NR TCS concentration: NR	Serum levels of E, P, SHBG, and E/P ratio at visit 1 (16–20 weeks gestation) and visit 3 (24–28 weeks gestation)	Linear mixed model for longitudinal analyses (visits 1 and 3), separate linear regression model (cross-sectional) for each visits 1 and 3, all models adjusted for maternal age, education and BMI; Percent change in hormone relative to population median hormone level (95% CI) associated with interquartile range increase in TCS	E Mixed model: −3.48 (4.93, −0.03)[b] Visit 1: −3.49 (−15.38, 8.40)[a] Visit 3: −7.30 (−25.00, 10.40)[a] P: Mixed model: −5.66 (−12.81, 2.08) Visit 1: −7.26 (−16.72, 3.27)[a] Visit 3: −2.19 (−13.20, 10.20)[a] SHBG: Mixed model: −1.44 (−8.16, 5.27) Visit 1: 3.48 (−7.66, 14.62)[a] Visit 3: 3.58 (−8.98, 16.14)[a] E/P Mixed model: −1.69 (−12.09, 8.71) Visit 1: −2.06 (−17.64, 13.52)[a] Visit 3: −9.20 (−23.94, 5.53)[a] [a]From supplemental tables [b]Estimate is outside 95% CI
Den Hond et al. (2015); cross sectional (prevalence case-control)	Cases: men with low TMC ($n = 40$) Controls: men with normal TMC ($n = 80$) All recruited from fertility clinics; Belgium	Single urine sample; expressed as interquartile interval; ln-transformed continuous LOD: NR %>LOD: NR TCS, μg/L *Controls* GM (25th %ile, 75th %ile): 2.8 (0.5, 5.5) *Cases* GM (25th %ile, 75th %ile): 2.6 (0.5, 15.6)	Serum free T, total T, free E, total E, LH, FSH, SHBG, and inhibin B	Linear regression adjusted for age, BMI, smoking, Cr; models adjusted for fasting serum (SHBG), timing of blood draw (SHBG and T). β (p value) per ln-change in TCS % change/50% TCS increase	SHBG Per ln change in TCS: −0.01 (0.55)[a] % change per 50% TCS increase: NR Total T Per ln change in TCS: 0.00 (0.74)[a] % change per 50% TCS increase: NR Free T Per ln change in TCS: 0.02 (0.35)[a] % change per 50% TCS increase: NR Total E Per ln change in TCS: 0.00 (0.59)[a] % change per 50% TCS increase: NR Free E Per ln change in TCS: 0.02 (0.38)[a] % change per 50% TCS increase: NR LH Per ln change in TCS: **0.03 (0.048)**[a] % change per 50% TCS increase: **1.2 (0.048)** FSH Per ln change in TCS: 0.02 (0.42)[a]

(*continued*)

Table 7. Continued

Publication; study design	Study population	Exposure assessment	Outcome definition	Analysis; measures of association	Results
					% change per 50% TCS increase: NR Inhibin B Per ln change in TCS: **−0.03 (0.046)**[a] % change per 50% TCS increase: **−1.4 (0.046)** [a]From supplemental tables "… we found that none of the correlations were statistically significant at the 0.1 level after accounting for multiple testing."
Ihde et al. (2015); cross-sectional	Pre-pubertal children ($n = 50$) who attended a pediatric clinic, USA	Single urine sample LOD: 0.399 ng/ml %>LOD: 6% TCS, ng/ml Min: 0.2 Median: 18.5 IQR: 112.75 Max: 2207.0	Urine estrogen metabolites: 2-OHE1, 2-OHE2, sum of 2-OHE1 and 2-OHE2, 2-OMeE1, 4-OHE1, 4-OMeE1, and 16a-OHE1	Spearman correlation coefficients	
Poole et al. (2016); cross-over experiment	Healthy adult volunteers ($n = 14$) randomized to 4 months of TCS products (toothpaste, soap, dishwashing liquid) followed by non-TCS products, or the same protocol in reverse (double blind), USA	TCS interval vs. non-TCS interval; binary LOD: NR %>LOD: 100 (TCS phase) %>LOD:72 (nTCS phase) TCS, pg/μl Median (Q1–Q3) at baseline: 24.4 (7.2–59.4) Median at end of TCS phase: 1548 Median at end of nTCS phase: 14.6	Serum free T, total T	Wilcoxon's test to examine differences across each period; p value	Free T $p = 0.16$ Total T[a] $p = 0.13$ [a]From supplemental table
Scinicariello and Buser (2016); cross sectional	Children and adolescents ($n = 588$) included in the 2011–2012 National Health and Nutrition Examination Survey (NHANES), USA	Single spot urine sample; quartiles LOD: 2.3 ng/ml %>LOD: NR TCS, ng/ml GM (SE) male child: 8.79 (1.80) male adolescent: 8.55 (1.67) female child: 6.30 (0.72) female adolescent: 12.27 (3.05)	Serum total T	Linear regression adjusted for age, race/ethnicity, cotinine, Cr, cholesterol, income, obesity, season and time of sample collection, physical activity (for adolescents only, BP-3, BPA, and parabens; β (95% CI)	Male children Q1: (ref) Q2: −37.50 (−72.75, 41.91) Q3: −17.30 (−62.09, 80.40) Q4: −16.18 (−51.32, 177.32) p-trend=**0.01** Male adolescents Q1: (ref) Q2: −24.42 (−55.51, 28.40) Q3: −36.87 (−60.94, 2.02) Q4: −9.52 (−39.35, 33.64) p-trend =0.08 Female children Q1: (ref) Q2: 22.14 (−34.30, 2.02) Q3: −35.60 (−58.10, −1.00) Q4: −17.30 (−56.83, 60.00) p-trend =0.06 Female adolescents Q1: (ref) Q2: −3.92 (−27.39, 25.86) Q3: −12.19 (−37.50, 24.61) Q4: −10.42 (−34.30, 22.14) p-trend =0.88

2-OHE1: 2-hydroxyestrone; 2-OHE2: 2-hydroxyestradiol; 2-OMeE1: 2-methoxyestrone; 4-OHE1: 4-hydroxyestrone; 4-OMeE1: 4-methoxyestrone; 16a-OHE1: 16 alpha-hydroxyestrone; BP-3: benzophenone-3; BPA: bisphenol A; CI: confidence interval; Cr: creatinine; E: estradiol; FSH: follicle stimulating hormone; GM: geometric mean; IQR: interquartile range; LOD: limit of detection; NR: not reported; P: progesterone; SHBG: sex hormone-binding globulin; LH: luteinizing hormone; T: testosterone; TCS: triclosan.

Table 8. Summary of epidemiologic studies evaluating association between TCS exposure and measures of thyroid function (bolded text indicates a statistically significant result).

Publication; study design	Study population (N, location)	Exposure assessment	Outcome definition	Analysis; measures of association	Results
Aker et al. (2016); cohort and cross sectional	Pregnant women ($n = 106$) from an ongoing prospective hospital- and prenatal clinic-based study (Puerto Rico Testsite for Exploring Contamination Threats), Puerto Rico	Three urine samples collected at 16–20; 20–24; and 24–28 weeks gestation; SG-standardized; ln-transformed, continuous LOD: 2.3 ng/ml[a] %>LOD: 93.3[a] [a]From Watkins et al. (2015); SG-corrected	Serum levels of TSH, FT3 and FT4 at visit 1 (16–20 weeks gestation) and visit 3 (24–28 weeks gestation)	Linear mixed model for longitudinal analyses (visits 1 and 3), separate linear model (cross-sectional) for each visit, all models adjusted for maternal age, education and BMI; TSH ln-transformed; percent change in hormone level relative to the median value per interquartile increase in TCS (95% CI)	TSH Mixed model: 4.90 (−5.28, 16.17) Visit 1: −6.23 (−20.78, 10.99)[a] Visit 3: 6.87 (−10.52, 27.63)[a] FT3: Mixed model: −1.07 (−4.78, 2.64) Visit 1: 0.62 (−4.92, 6.15)[a] Visit 3: −3.15 (−8.49, 2.18)[a] FT4: Mixed model: −3.44 (−8.60, 1.71) Visit 1: −1.53 (−8.04, 4.99)[a] Visit 3: −4.12 (−12.49, 4.25)[a] [a]From supplemental tables
Allmyr et al. (2009); before-and-after experiment	Healthy employees or students ($n = 12$) using toothpaste with 0.3% triclosan at the Karolinska Institute, Sweden	Difference in plasma levels before and after 14-day trial of toothpaste use; continuous variable LOQ: NR %>LOD: 100 TCS, ng/g *Pre-exposure* Range: 0.009–0.81 *Post-exposure* Range: 26–296	Difference in plasma levels of TSH, FT3 and FT4 before and after exposure interval	Linear regression; β (95% CI)[a] [a]Calculated by MG based on data provided in supplemental materials	TSH −0.001 (−0.006, 0.004) FT3 −0.003 (−0.007, 0.001) FT4 −0.002 (−0.011, 0.008)
Cullinan et al. (2012); placebo controlled experiment	CVD patients ($n = 132$) (Cardiovascular and Periodontal Study) enrolled in a randomized, double blind, placebo controlled, clinical trial of 4-year use of 0.3% triclosan in toothpaste vs. placebo, Australia	Triclosan intervention vs. placebo; binary No TCS measurements	Serum levels of TSH, FT3, FT4, and thyroid autoantibodies (anti-TGab and anti-TPOab) at year 1 and year 5 of the trial	*t*-Test comparing hormone levels between intervention and placebo groups in year 1, ANCOVA for the same comparisons in year 5 with year 1 levels used as a covariate, Fischer's Exact test to compare intervention and placebo groups with respect to percentage of patients outside reference ranges and percentages that moved outside and into reference range in between years 1 and 5; *p* values	TSH Year 1 *t*-test: $p = 0.09$ Year 5 ANCOVA: $p = 0.54$ Year 1: % outside range: $p = 0.43$ Year 5: % outside range: $p = 1.00$ % moved outside range by year 5: $p = 1.00$ % moved into range by year 5: $p = 1.00$ FT3 Year 1 *t*-test: $p = 0.47$ Year 5 ANCOVA: $p = 0.89$ Year 1: % outside range: NC Year 5: % outside range: NC % moved outside range by year 5: NC % moved into range by year 5: NC FT4 Year 1 *t*-test: $p = 0.12$ **Year 5 ANCOVA: $p = 0.01$** Year 1: % outside range: $p = 0.61$ Year 5: % outside range: $p = 0.61$ % moved outside range by year 5: $p = 0.48$

(*continued*)

Table 8. Continued

Publication; study design	Study population (N, location)	Exposure assessment	Outcome definition	Analysis; measures of association	Results
					% moved into range by year 5: $p = 1.0$ anti-TGab Year 1 *t*-test: $p = 0.90$ Year 5 ANCOVA: $p = 0.56$ Year 1: % outside range: $p = 1.0$ Year 5: % outside range: $p = 0.5$ % moved outside range by year 5: NC % moved into range by year 5: NC anti-TPOab Year 1 *t*-test: $p = 0.38$ Year 5 ANCOVA: $p = 0.88$ Year 1: % outside range: $p = 0.49$ Year 5: % outside range: $p = 1.00$ % moved outside range by year 5: $p = 0.5$ % moved into range by year 5: NC
Geens et al. (2015); cross sectional	Obese adult patients ($n = 151$) undergoing weight loss treatment and lean individuals ($n = 43$) selected among staff and volunteers, Endorup trial study, Belgium	For obese subjects, three 24-h urine samples collected at baseline (0 months) and at 6 and 12 months of the study, and a single void measure at 3 months. For lean subjects, a one-time single void measure; log-transformed; continuous variables. LOD: NR %>LOD: >99 TCS, ng/ml *Obese, median* 0 months: 1.5 3 months: 1.3 6 months: 1.9 12 months: 1.9 *Lean, median* 0.9	Serum TSH and FT4 at 0, 3, 6, and 12 months for obese subjects. Apparently one-time measure for lean subjects	Linear regression for 0, 3, and 6 months, adjusted for age, weight loss (for 3 and 6 months in obese individuals) and gender (in the overall analysis) β (95% CI)	TSH 0 months all obese: 0.01 (−0.17, 0.17) 0 months obese men: −0.15 (−0.46, 0.16) 0 months obese women: 0.08 (−0.12, 0.29) 3 months all obese: −0.01 (−0.24, 0.22) 3 months obese men: −0.25 (−0.69, 0.19) 3 months obese women: 0.06 (−0.22, 0.34) 6 months all obese: 0.10 (−0.18, 0.37) 6 months obese men: −0.22 (−1.01, 0.56) 6 months obese women: 0.19 (−0.13, 0.52) All lean: 0.08 (−0.25, 0.42) Lean men: 0.28 (−0.44, 1.01) Lean women: 0.03 (−0.39, 0.45) FT4 0 months all obese: **−0.18 (−0.34, −0.01)** 0 months obese men: −0.11 (−0.42, 0.20) 0 months obese women: **−0.22 (−0.42, −0.02)**

(*continued*)

Table 8. Continued

Publication; study design	Study population (N, location)	Exposure assessment	Outcome definition	Analysis; measures of association	Results
					3 months all obese: −0.01 (−0.03, 0.01) 3 months obese men: 0.06 (−0.40, 0.53) 3 months obese women: **−0.30 (−0.55, −0.05)** 6 months all obese: −0.09 (−0.41, 0.22) 6 months obese men: −0.05 (−0.78, 0.69) 6 months obese women: −0.16 (−0.52, 0.21) All lean: −0.07 (−0.40, 0.26) Lean men: 0.02 (−0.72, 0.77) Lean women: −0.12 (−0.54, 0.29)
Koeppe et al. (2013); cross sectional	Adults and adolescents ($n = 1831$) included in the 2007–2008 National Health and Nutrition Examination Survey (NHANES), USA	Single spot urine sample; ln-transformed; continuous and quartiles for non-linear relationships LOD: 2.3 μg/L %>LOD: 81.1 (adolescent male), 86 (adolescent female), 78.5 (adult male), 80.8 (adult female) TCS, μg/g Cr Median Adolescent males: 10.4 Adolescent females: 11.5 Adult males: 10.6 Adult females: 15.6	One-time serum FT3 and FT4 (ln-transformed), and TT3 and TT4; TSH, thyroglobulin; continuous	Linear regression sample-weighted and adjusted for age, sex, BMI and ln-Cr; also age-and sex-stratified; β (95% CI)	Ln-FT3 Adolescents: 0.002 (−0.01, 0.01) Adults: −0.001 (−0.003, 0.001) Adolescent males: 0.004 (−0.01, 0.02) Adolescent females: −0.001 (−0.01, 0.009) Adult males: −0.002 (−0.01, 0.002) Adult females: 0.000 (−0.01, 0.004) TT3 Adolescents: **1.96 (0.05, 3.88)** Adults: −0.006 (−0.88, 0.86) Adolescent males: 1.82 (−0.51, 4.15) Adolescent females: 1.90 (−1.07, 4.87) Adult males: 0.11 (−1.08, 1.30) Adult females: −0.10 (−1.57, 1.37) Ln–FT4 Adolescents: −0.002 (−0.02, 0.01) Adults: 0.000 (−0.01, 0.01) Adolescent males: −0.007 (−0.02, 0.004) Adolescent females: 0.004 (−0.02, 0.03) Adult males: 0.007 (−0.003, 0.02) Adult females: −0.007 (−0.02, 0.002) TT4 Adolescents: 0.008 (−0.11, 0.12)

(*continued*)

Table 8. Continued

Publication; study design	Study population (N, location)	Exposure assessment	Outcome definition	Analysis; measures of association	Results
					Adults: –0.005 (–0.07, 0.06) Adolescent males: –0.03 (–0.15, 0.09) Adolescent females: 0.03 (–0.17, 0.23) Adult males: –0.005 (–0.08, 0.07) Adult females: –0.008 (–0.11, 0.09) TSH No significant findings, data not reported Thyroglobulin No significant findings, data not reported
Poole et al. (2016); randomized cross-over experiment	Healthy adult volunteers ($n = 14$) randomized to 4 months of TCS products (toothpaste, soap, dishwashing liquid) followed by non-TCS products, or the same protocol in reverse (double blind), USA	TCS interval vs. non-TCS interval; binary LOD: NR %>LOD: 100 (TCS phase), 72 (nTCS phase) TCS, pg/μl Median (Q1–Q3) at baseline: 24.4 (7.2–59.4) Median at end of TCS phase: 1548 Median at end of nTCS phase: 14.6	Serum TSH and FT4[a] measured at baseline and the end of each interval [a]From Supplemental Table 1	Wilcoxon's test to examine differences across each period	TSH $p = 0.18$ free FT4 NC[a] [a]From supplemental table. According to text, "No significant differences by phase were found" for TSH and T4

Cr: creatinine; TSH: thyroid stimulating hormone; FT3: free T3 (triiodothyronine); TT3: total T3; FT4: free T4 (thyroxine); TT4: total T4; anti-TGab: antithyroglobulin antibody; anti-TPOab: antithyroid peroxidase antibody; ANCOVA: analysis of covariance; CI: confidence interval; LOD: limit of detection; LOQ: limit of quantitation; NC: not calculated; NR: not reported; SG: specific gravity; TCS: triclosan.

Table 9. Summary of epidemiologic studies evaluating association between TCS exposure miscellaneous health outcomes (bolded text indicates a statistically significant result).

Publication; study design	Study population	Exposure assessment	Outcome definition	Analysis; measures of association	Results
Adgent and Rogan (2015); cross sectional	Adult participants ($n = 2441$) included in the 2005–2008 National Health and Nutrition Examination Surveys, USA	Single spot urine sample; ln-transformed continuous; binary (detected vs. non-detected); quintiles; Cr-adjusted (results not shown) and unadjusted LOD: 2.3 ng/ml %>LOD: 80 TCS, ng/ml Min: <2.3 Median: 12.1 Max: 3620 Q1: non-detect Q2: 2.3 to <7.4 Q3: 7.4 to <21.6 Q4: 21.6 to <104.5 Q5: 104.5+	Urinary enterolactone (an indirect measure of antimicrobial exposures)	Linear regression adjusted for sex, age, race/ethnicity, BMI, poverty income ratio, dietary fiber, BM/day, education, cotinine, and ln-Cr; β (95% CI)	Continuous ln-TCS 0.01 (−0.04, 0.06) Detected vs. non detected TCS All: 0.07 (−0.15, 0.30) Women: 0.31 (−0.07, 0.70) Men: -0.18 (−0.47, 0.11) *p*-interaction =0.02 TCS quintiles Q1: (ref) Q2: −0.02 (−0.33, 0.28) Q3: 0.20 (−0.06, 0.46) Q4: 0.07 (−0.27, 0.41) Q5: 0.06 (−0.21, 0.34) *p*-trend: NR
Allmyr et al. (2009); before-and-after experiment	Healthy employees or students ($n = 12$) using toothpaste with 0.3% TCS at the Karolinska Institute, Sweden	Difference in plasma levels before and after 14-day trial of toothpaste use; continuous variable LOQ: NR %>LOD: 100 TCS, ng/g *Pre-exposure* Range: 0.009–0.81 *Post-exposure* Range: 26–296	Difference in plasma 4β-hydroxycholesterol (a marker for CYP3A4 induction) before and after exposure interval	Linear regression; β (95% CI)[a] [a]Calculated by MG based on data provided in supplemental materials	0.028 (−0.021, 0.076)
Ashley-Martin et al. (2016); cohort	Mother–newborn pairs ($n = 1219$) from Maternal-Infant Research on Environmental Chemicals Study, Canada	Single urine sample at 6–13 weeks gestation; quartiles or log-transformed continuous LOD =0.12 µg/L %>LOD: 99.4 TCS, µg/L $\text{Median}_{\text{sg-adj:}}$ 9.52 IQR: 10.24 Q1: <2.30 Q2: 2.30 to <8.90 Q3: 8.90 to <77.09 Q4: ≥77.09	Cord plasma immune markers: IgE, IL-33, TSLP, IL-33/TSLP ratio	Logistic regression adjusted for specific gravity, maternal age, and household income; OR (95% CI)	IgE Q1: (ref) Q2: 1.50 (0.95, 2.38) Q3: 1.29 (0.82, 2.06) Q4: 1.03 (0.63, 1.68) *p*-trend: NR IL-33 Q1: (ref) Q2: 1.41 (0.92, 2.17) Q3: 1.21 (0.78, 1.87) Q4: 1.32 (0.85, 2.05) *p*-trend: NR TSLP Q1: (ref) Q2: 1.24 (0.82, 1.89) Q3: 1.13 (0.74, 1.71) Q4: 1.08 (0.70, 1.66) *p*-trend: NR IL-33/TSLP Q1: (ref) Q2: 1.35 (0.85, 2.15) Q3: 1.21 (0.76, 1.92) Q4: 1.17 (0.72, 1.89) *p*-trend: NR No significant associations with TCS modeled as a continuous variable. No effect modification by sex

(*continued*)

Table 9. Continued

Publication; study design	Study population	Exposure assessment	Outcome definition	Analysis; measures of association	Results
Clayton et al. (2011); cross sectional	Adults and children ($n = 620$) included in the 2003–2004 National Health and Nutrition Examination Surveys, USA	Single spot urine sample; ln-transformed continuous only included participants with levels > LOD LOD: 2.3 ng/ml %>LOD: NR TCS, ng/ml (NHANES 2003–2006) Mean (95% CI): 116.3 (105.43, 127.11)	Serum CMV-specific IgG	Linear regression adjusted for age, sex, Cr, race, BMI, family income, and education; β (SE)	Children (≤18 years) 0.012 (0.046), $p > 0.05$ Adults (≥18 years) −0.022 (0.032), $p > 0.05$
Cullinan et al. (2015a); placebo controlled experiment	CVD patients ($n = 383$) enrolled in Cardiovascular and Periodontal Study, a randomized, double blind, placebo-controlled, prospective clinical trial of 4-year use of 0.3% triclosan in toothpaste vs. placebo; Australia	Triclosan intervention vs. placebo; binary No TCS measurements	Changes in blood biomarkers (systemic, kidney function, liver function) between baseline and end of study (measured at baseline and annually). Biomarkers included TC, LDL, HDL, TG, CRP, ESR, Hb, WCC, potassium, eGFR, glucose, bilirubin, ALP, AST, ALT, GGT	Mixed linear model regression with patient-level random effect adjusted for sex, age, hypertension, diabetes, periodontal status, smoking, and use of statins or anti-inflammatory medications; p value only[a] [a]Analysis based on $n = 165$ in TCS group and $n = 163$ in placebo group	TC p=**0.03** greater decrease in TCS group LDL p=**0.04** decrease in TCS group only HDL $p = 0.05$ TG $p = 0.52$ CRP $p = 0.85$ ESR $p = 0.06$ Hb $p = 0.82$ WCC $p = 0.93$ Potassium $p = 0.46$ eGFR $p = 0.77$ Glucose $p = 0.15$ Bilirubin $p = 0.96$ ALP $p = 0.15$ AST $p = 0.51$ ALT $p = 0.81$ GGT $p = 0.46$
Cullinan et al. (2015b); placebo controlled experiment	CVD patients ($n = 438$) enrolled in Cardiovascular and Periodontal Study, a randomized, double blind, placebo-controlled, prospective clinical trial of 5-year use of 0.3% triclosan in toothpaste vs. placebo; Australia	Triclosan intervention vs. placebo; binary No measurements	SAE during follow up	Logistic regression[a] adjusted for gender, age and BMI; OR (95% CI) [a]Analysis based on $n = 223$ in TCS group and $n = 215$ in placebo group	Any SAE 1.165 (0.80, 1.71) Cardiovascular SAE 0.863 (0.56, 1.34)
Poole et al. (2016); placebo controlled cross-over experiment	Healthy adult volunteers ($n = 14$) randomized to 4 months of TCS products (toothpaste,	TCS interval vs. non-TCS interval; binary LOD: NR	Changes in plasma biomarkers and oral and gut microbiome between baseline and end of	Microbiome: Linear model with mixed effects Biomarkers: Wilcoxon's test to	Oral/gut microbiome for different taxa all p values >0.42

(*continued*)

Table 9. Continued

Publication; study design	Study population	Exposure assessment	Outcome definition	Analysis; measures of association	Results
	soap, dishwashing liquid) followed by non-TCS products, or the same protocol in reverse (double blind), USA	%>LOD: 100 (TCS phase), 72 (nTCS phase) TCS, pg/μl (5 urine samples per study period) Median (Q1–Q3) at baseline: 24.4 (7.2–59.4) Median at end of TCS phase: 1548 Median at end of nTCS phase: 14.6	study. Biomarkers included adiponectin, adipsin, C-peptide, CRP, Cr, ESR, ghrelin, GIP, GLP1, glucagon, glucose, IL-10, IL-6, leptin, MCP-1, resistin, serpin E1, TG, TNF-α, visfatin	examine differences across study period; *p* value	Adiponectin[a] $p = 0.24$ Adiposin[a] $p = 0.38$ C-peptide[a] $p = 0.4$ Cr[a] $p = 0.66$ CRP $p = 0.65$ ESR[a] $p = 0.26$ GIP $p = 0.53$ Ghrelin[a] $p = 0.14$ GLP1[a] $p = 0.75$ Glucagon[a] $p = 0.66$ Glucose[a] $p = 0.22$ IL-10[a] $p = 0.73$ IL-6[a] $p = 0.59$ Insulin $p = 0.81$ Leptin $p = 0.15$ MCP-1[a] $p = 0.79$ Resistin[a] $p = 0.49$ Serpin E1[a] $p = 0.48$ TG $p = 0.39$ TNF-α[a] p=NC Visfatin[a] p=NC [a]From supplemental tables
Shiue (2015a); cross-sectional	Children, adolescents and adults ($n = 5892$, age 12–80 years) included in the 2011–2012 National Health and Nutrition Examination Survey (NHANES), USA	Single spot urine sample; log-transformed; continuous LOD: NR %>LOD: NR	Self-reported health characterized as: (1) "very good, excellent" vs. "good, fair, poor" (2) "very good, excellent, "good" vs. "fair, poor" (3) as a continuous variable	Logistic regression for binary outcomes; OR (95% CI) Linear regression for continuous outcomes; β (95% CI) All adjusted for age, sex, BMI, income to poverty ratio, cotinine, alcohol status, physical activity, and Cr	Very good, excellent vs. good, fair, poor **0.93 (0.89, 0.98)** Very good, excellent, good vs. fair, poor **0.87 (0.80, 0.96)** Continuous **−0.04 (−0.06, −0.02)**

(*continued*)

Table 9. Continued

Publication; study design	Study population	Exposure assessment	Outcome definition	Analysis; measures of association	Results
Shiue et al. (2015b); cross-sectional	Adults ($n = 4566$, age 30–80 years) included in the 2011–2012 National Health and Nutrition Examination Survey (NHANES), USA	Single spot urine sample; log-transformed; continuous LOD: NR %>LOD: NR	Self-reported oral health characterized as: (1) gum disease (yes/no) (2) bone loss (yes/no) (3) loose teeth (yes/no) (4) tooth "does not look right" (yes/no) (5) overall health of teeth (good, fair, poor) (6) tooth ache (rare, occasional, often)	Logistic regression adjusted for Cr, age, sex, BMI, ratio of family income to poverty ratio, serum cotinine and sampling weight; OR (95% CI)	Gum disease (n=4481) 0.97 (0.85, 1.10) Bone loss (n=4531) NR Loose teeth (n=4557) **0.81 (0.70, 0.93)** Tooth "does not look right" (n=4557) **0.84 (0.75, 0.95)** Overall health of teeth (n=4554) Good vs. fair: NR Good vs. poor: NR Mouth ache (n=4562) Rare vs. occasional: 0.90 (0.81, 1.00) Rare vs. often: **0.82 (0.70, 0.98)**
Watkins et al. (2015); cohort and cross-sectional	Pregnant women ($n = 106$) from an ongoing Puerto Rico Testsite for Exploring Contamination Threats study, hospital- and prenatal clinic-based; Puerto Rico	Three spot urine samples at 16–20 (visit 1); 20–24 (visit 2); and 24–28 weeks (visit 3) gestation; ln-transformed, continuous LOD: 2.3 ng/ml %>LOD: 93.3 $TCS_{SG\text{-}adj}$, ng/ml GM: 27.3 5th %ile: 1.53 25th %ile: 4.90 50th %ile: 26.5 75th %ile: 148 95th %ile: 913	Biomarkers of oxidative stress (in urine) and inflammation (in plasma) at visits 1 and 3: OHdG, isoprostane, IL-1β, IL-6, IL-10, TNF-α, CRP	Linear mixed model with patient-level random intercept adjusted for SG, study visit, maternal pre-pregnancy BMI and education; % difference of IQR (95% CI)	OHdG (n=52) −2.20 (−15.6, 13.3) Isoprostane (n=52) 0.52 (−16.5, 21.1) IL-1β (n=101) 14.4 (−4.51, 37.1) IL-6 (n=101) **31.5 (8.05, 59.9)** IL-10 (n=101) 13.6 (−5.61, 36.8) TNF-α (n=101) −0.93 (−10.2, 9.29) CRP (n=101) 8.93 (−8.50, 29.7)
You et al. (2011); cross-sectional	Adults and children without known renal disease ($n = 2573$) included in the 2003–2006 National Health and Nutrition Examination Surveys, USA	Single spot urine sample, continuous LOD: NR %>LOD: NR	Renal function, $eGFR_{MDRD}$ and $eGFR_{CKD\text{-}EPI}$ categorized as ≥90 (normal), 60–90 (mild impairment) and <60 (CKD) mL/min/1.73 m^2	Regression models adjusted for education, income, household size, BMI, WC, smoking, alcohol, calorie intake, daily activities, CVD and Cr; GM (95% CI) of TCS by eGFR category	$eGFR_{MDRD}$, all ≥90: 19.40 (13.66, 27.56) 60–90: 14.68 (11.17, 19.29) <60: 13.76 (8.62, 21.98) *p*-trend=**0.003** $eGFR_{MDRD}$, males ≥90: 23.05 (14.44, 36.79) 60–90: 17.85 (12.21–26.09) <60: 14.42 (7.69, 27.03) *p*-trend =0.07 $eGFR_{MDRD}$, females ≥90: 18.84 (9.01, 39.98) 60–90: 13.53 (6.91, 26.49) <60: 14.87 (6.91, 32.02) *p*-trend =0.07 Interaction with sex: $p = 0.89$ $eGFR_{MDRD}$, ≥65 years ≥90: 9.92 (4.30, 22.88) 60–90: 5.74 (3.19, 10.34) <60: 7.41 (3.51, 15.66) *p*-trend =0.91 $eGFR_{MDRD}$, ≤65 years

(*continued*)

Table 9. Continued

Publication; study design	Study population	Exposure assessment	Outcome definition	Analysis; measures of association	Results
					≥90: 21.74 (13.99, 33.78) 60–90: 16.80 (11.69, 24.16) <60: 8.45 (3.97, 17.99) *p*-trend =0.008 Interaction with age: *p*=**0.02** $eGFR_{CKD\text{-}EPI}$, all ≥90: 18.75 (13.36, 26.31) 60–90: 14.21 (10.04, 20.12) <60: 13.56 (8.21, 22.41) *p*-trend=**0.01** $eGFR_{CKD\text{-}EPI}$, males ≥90: 23.84 (11.12, 51.11) 60–90: 21.25 (8.38, 53.87) <60: 14.51 (5.48, 38.40) *p*-trend =0.14 $eGFR_{CKD\text{-}EPI}$, females ≥90: 18.82 (9.18, 38.59) 60–90: 11.58 (5.94, 22.55) <60: 16.52 (6.52, 41.87) *p*-trend=**0.03** Interaction with sex: *p* = 0.66 $eGFR_{CKD\text{-}EPI}$, ≥65 years ≥90: 12.17 (5.78, 25.60) 60–90: 5.42 (2.83, 10.40) <60: 6.23 (3.05, 12.72) *p*-trend =0.19 $eGFR_{CKD\text{-}EPI}$, ≤65 years ≥90: 20.88 (13.66, 31.91) 60–90: 15.98 (9.99, 25.57) <60: 7.98 (3.92, 16.25) *p*-trend=**0.02** Interaction with age: *p*=**0.04**

ALP: alkaline phosphatase; ALT: alanine transaminase; AST: aspartate transaminase; BM: bowel movement; CI: confidence interval; CRP: C-reactive protein; CVD: cardiovascular disease; CYP3A4: cytochrome P450 3A4; eGFR: estimated glomerular filtration rate; $eGFR_{CKD\text{-}EPI}$: eGFR calculated using Chronic Kidney Disease Epidemiology Collaboration formula; $eGFR_{MDRD}$: eGFR calculated using Modification of Diet in Renal Disease formula; ESR: erythrocyte sedimentation rate; GGT: γ-glutamyl transferase; GIP: gastric inhibitory polypeptide; GLP: glucagon-like peptide; GM: geometric mean; Hb: hemoglobin; HDL: high density lipoprotein; IgG: immunoglobulin G; IL-interleukin; IQR: interquartile range; LDL: low-density lipoprotein; LOD: limit of detection; MCP: monocyte chemotactic protein; NR: not reported; OR: odds ratio; SAE: serious adverse event; TC: total cholesterol; TCS: triclosan; TG: triglycerides; TNF: tumor necrosis factor; WCC: white cell count.

pregnancy (Smarr et al. 2017). The outcome of interest was time to pregnancy. The study included a variety of analyses that treated TCS as a binary, continuous or ordinal (quartiles) variable after controlling for multiple potential confounders including male partner's TCS. None of the analyses demonstrated a significant association between baseline TCS and time to pregnancy and there was no evidence of a dose-response. As noted earlier, a methodological strength of the Smarr et al. (2017) study is the ability to examine the extent of exposure assessment error by comparing the overall result to the result of a separate analysis restricted to couples with a shorter time to pregnancy.

Velez et al. (2015) conducted a cross-sectional analysis that examined the association between urinary TCS measured during first trimester of pregnancy and time to pregnancy (ascertained by recall). The results were mixed; only one analysis, which used urinary TCS levels as a binary variable (using 75th percentile as the cutoff), demonstrated a statistically significant association with longer time to pregnancy; however, this result was not confirmed when TCS was treated as a continuous or ordinal variable.

Male reproductive endpoints

Endpoints examined in this group of studies included male infertility or subfertility (two studies), time to pregnancy (one study), and a variety of measures of semen quality such as volume, concentration, total sperm count, and sperm motility and morphology (three studies). All but one study in this group were cross-sectional (Table 4). Only one cohort study (Smarr et al. 2017) evaluated the association between TCS levels at baseline and time to partner's pregnancy; the results showed no evidence of association in any of the analyses.

The two cross-sectional studies assessing the association between male subfertility or infertility and TCS did not demonstrate consistent results. One study conducted in China (Chen et al. 2013) observed a significant positive association in one of the two analyses (creatinine-adjusted only), whereas the second study conducted in Belgium (Den Hond et al. 2015) reported a null result.

Studies that examined the various measures of semen quality were also inconsistent. The studies by Chen et al. (2013) and Den Hond et al. reported null results. Another cross-sectional study from China (Zhu et al. 2016) reported some statistically significant results among 24 TCS tertile-specific regression coefficients, but all of these results were observed in the lowest rather the highest tertiles of TCS exposure and the significant associations varied in terms of magnitude and direction.

Birth outcomes

Five cohort studies and one case-control study examined the association between prenatal exposure to TCS and birth outcomes (Table 5). All of these studies were assigned to Tier 3 with respect to their exposure assessment methods.

Only one of six studies, a case-control study from France, focused on birth defects – hypospadias and undescended testes. The five cohort studies reported data for 11 different endpoints: gestational age, preterm birth, low birth weight for gestational age, weight, body length, head circumference, abdominal circumference, two alternative measures of anogenital distance, penile width, and femoral length. These endpoints were assessed at different ages pre- and postnatally, among different population subgroups, and for different exposure metrics, which resulted in over 130 individual measures of association.

Despite the large number of reported results, consistency across studies is difficult to assess because very few of these results are directly comparable. For example, Wolff et al. (2008) and Geer et al. (2017) reported the associations between prenatal TCS and birth weight in boys and girls adjusted for sex, Philippat et al. (2012) and Philippat et al. (2014) excluded girls from their analyses, and Lassen et al. (2016) reported findings separately for boys and girls.

Very few of the results were statistically significantly different from the null value and the statistically significant results lacked inter- or intra-study consistency. For example, the significant association between prenatal TCS and head circumference in the Lassen et al. (2016) study was observed in boys, but not girls; the two other studies limited to boys (Philippat et al. 2012, 2014) reported null results, as did the two studies that examined boys and girls together (Wolff et al. 2008; Geer et al. 2017).

Puberty endpoints in girls

Three studies evaluated the hypothesis that TCS exposure may influence onset of puberty in girls (Table 6). Two of those studies were based on the same Breast Cancer and Environment Research Program (BCERP) cohort (Wolff et al. 2010, 2015) and the third study (Buttke et al. 2012) used data from the 2003–2008 NHANES.

The participants in the BCERP cohort study were enrolled at the age of 6–8 years and were followed for up to seven years. TCS was measured at baseline in a single urine sample collected at enrollment. The main endpoints of interest were the presence and timing of breast and pubic hair development (Tanner stages B2 and PH2, respectively). The first analysis of this cohort (Wolff et al. 2010) evaluated the association between urinary TCS at baseline and evidence of B2 or PH2 at one year of follow up. TCS was expressed as an ordinal (quintiles) or a continuous variable, although the results for continuous TCS were not reported. In the analyses by quintile, no association was observed for B2, but there was evidence of an inverse relation between TCS and PH2. In the second publication (Wolff et al. 2015), which described the same cohort but used time-to-event analysis, the results for B2 demonstrated a positive association with TCS (this was not present in the Wolff et al. 2010 study). By contrast, the previously observed inverse association for PH2 was no longer present in Wolff et al. (2015).

The NHANES-based study (Buttke et al. 2012) ascertained age at menarche among 12- to 16-year-old girls based on information provided by a parent or guardian. TCS was measured in a single urine sample obtained at the time of the interview. No association between TCS and age at menarche was observed; this study was assigned to Tier 3 for all three

methodological domains (design, exposure assessment and data analysis).

Sex hormone concentrations

Table 7 summarizes the five studies that investigated the association between TCS and sex hormone levels. The cohort study by Aker et al. (2016) conducted longitudinal analyses using serum levels of estrogen, progesterone, sex-hormone-binding globulin (SHBG), and estrogen/progesterone ratio as the endpoints of interest. There was no evidence of an association in any of the analyses. Similarly, the cross-over trial by Poole et al. (2016) found no evidence that TCS exposure affects serum levels of free or total testosterone.

The remaining three studies in this group used cross-sectional design. Den Hond et al. (2015) reported a positive significant association of TCS with luteinizing hormone and an inverse significant association with inhibin B in men, but no association with any other measures (SHBG, FSH, free and total estrogen and free and total testosterone). An NHANES-based study (Scinicariello and Buser 2016) reported an inverse significant trend for association of TCS with total serum testosterone, but only among male children and not among male adolescents or female participants of any age group. Ihde et al. (2015) observed no statistically significant associations between TCS and urine estrogen metabolite levels in pre-pubertal children.

Thyroid function

Six studies examined the association between TCS and thyroid hormone levels (Table 8). Three of these studies were experimental and three were observational. One observational study was based on a cohort of pregnant women in Puerto Rico and the remaining two observational studies were cross sectional.

A single arm Swedish trial (Allmyr et al. 2009) examined changes in thyroid hormone concentrations before and after a two-week use of a TCS-containing toothpaste. Plasma TCS levels increased from 0.009–0.81 ng/g before the intervention to 26–296 ng/g at the completion of the trial. Despite this increase in TCS exposure, the levels of thyroid stimulating hormone (TSH), free thyroxine (FT4), and free triiodothyronine (FT3) remained largely unchanged.

Another experimental study was based on the double-blind, randomized, placebo-controlled, CAPS trial (Cullinan et al. 2012). Subjects assigned to receive either TCS or placebo toothpaste were compared with respect to changes in TSH, FT3, FT4, and thyroid antibodies at year 1 and year 5 of the trial. Additional analyses also compared the two groups with respect to the percentages of participants whose thyroid hormone values moved outside the reference range. There was no evidence that TCS toothpaste had a deleterious effect on thyroid function. Only one of 31 different analyses demonstrated statistically significant difference between two groups (FT4 at 5 years); however, the difference was opposite of the hypothesized direction with TCS group showing higher levels relative to the placebo group.

The third experimental study in this group was a double-blind, randomized, crossover trial of TCS incorporated in household and personal care products (Poole et al. 2016). One of the analyses in that study compared serum TSH and FT4 levels at baseline and at the end of each (TCS and placebo) 4-month interval. No significant effect on either of these markers was found, despite an almost 100-fold difference in TCS exposure between the two study intervals.

Aker et al. (2016) used data from a cohort study of pregnant women in Puerto Rico to examine associations between urinary TCS concentrations measured during three visits at 16–20, 20–24, and 24–28 weeks gestation and serum levels of thyroid hormones at visits 1 and 3. A distinguishing feature of this study is the availability of three TCS measures and two estimates of thyroid hormones. This allowed the authors to conduct longitudinal analyses assessing the association between TCS and each of the thyroid function markers (TSH, FT3, and FT4) across the follow up. Neither the longitudinal analyses nor the cross sectional analyses carried out separately for each visit showed a significant association between TCS and thyroid hormone levels.

The results of the two cross-sectional studies demonstrated greater variability. Geens et al. (2015) measured TCS in urine samples collected at baseline and at 3, 6, and 12 months of follow-up from obese adult patients undergoing weight loss treatment. Serum TSH and FT4 were also measured at each visit. In addition, one-time TCS and thyroid hormone measures were obtained from non-obese volunteers. No associations were observed for TSH. The results for FT4 were inconsistent. There was a significant inverse association with TCS, but this association was only found among all participants or obese women, and not among obese men or non-obese subjects of either sex. For obese women, the association was only present at baseline and at 3 months, but not at the 6 month follow up.

In another cross-sectional study based on the 2007–2008 NHANES data (Koeppe et al. 2013), the authors conducted multiple analyses evaluating associations between urinary TCS levels and serum FT3, FT4, TSH, thyroglobulin, and total T3 and T4. Analyses were conducted separately by age and by sex. Only one of 24 reported results (the association between TCS and total T3 among all adolescents) showed a statistically significant departure from the null and the direction of that association was positive, opposite of the hypothesized direction.

Miscellaneous

The studies of miscellaneous health endpoints are presented in Table 9. This group included 11 different publications (many already discussed in the previous sections). Four of these publications were clinical trials (Allmyr et al. 2009; Cullinan et al. 2015a, 2015b, Poole et al. 2016), two were cohort studies (Watkins et al. 2015; Ashley-Martin et al. 2016) and the remaining five studies were cross sectional (all NHANES-based). Only two of the studies examined a similar endpoint. A cross-sectional analysis of the 2003–2006 NHANES data (You et al. 2011) observed that persons with lower estimated glomerular filtration rate (a measure of kidney function) tended to have lower levels of urinary TCS (a measure of excretion). By contrast, the placebo-controlled

trial of TCS toothpaste (Cullinan et al. 2015a, 2015b) found no evidence that eGFR differed across the two study groups ($p = 0.77$).

All other endpoints examined in this group of studies varied widely, and for this reason this group cannot be evaluated with respect to consistency or patterns of results. These studies are included in the current review for completeness.

Discussion

In the past two decades, the focus in environmental epidemiology has shifted from chemicals with long physiologic half-lives (e.g. dioxins and polychlorinated biphenyls) to those that are metabolized and/or eliminated within hours or days. Short-lived chemicals present unique challenges for exposure assessment and for study design and interpretation (Goodman et al. 2014a; LaKind et al. 2014b). TCS is an important short-lived chemical because of widespread human exposure and public concern surrounding its use (http://www.nytimes.com/2013/12/17/health/fda-to-require-proof-that-antibacterial-soaps-are-safe.html). In this review, we systematically examined the body of published epidemiologic literature on TCS and a range of health endpoints. We focused on two aspects of this literature: the quality of the publications as defined by elements in the BEES-C instrument (LaKind et al. 2014a, 2015b) and – for a given health outcome – the extent to which we observed inter-study and intra-study consistency in terms of specific research questions, methods, results, and reporting.

Quality of the available evidence

One lesson of the current review is the relative paucity of high quality studies capable of providing convincing evidence for or against a causal link between TCS and any of the outcomes studied. Only one study, a randomized cross-over clinical trial (Poole et al. 2016) was assigned to Tier 1 for all three domains. Most studies (38 of 43) were assigned to Tier 3 for at least one of the domains.

In conducting a quality assessment of individual studies, we found that problems with the extant literature fall into two main categories: limited ability to establish the relevant temporal relationship between exposure to TCS and outcomes of interest and inadequate exposure assessment increasing the likelihood of misclassification.

Study design and temporal relationship

As the main and perhaps the only inarguable property of a causal link is the temporal relationship between exposure and outcome (Potischman and Weed 1999; Rothman and Greenland 2005), cross-sectional and other prevalence-based studies are usually inadequate for confirming or refuting causal associations (Adami et al. 2011). Most of the studies in this review fall into this category.

Clinical trials may offer unequivocal evidence that the exposure (e.g. via intentional administration of TCS products) truly preceded the outcome (e.g. as measured by a change in hormone levels). On the other hand, clinical trials have a limited ability to address relevant research questions related to long-term outcomes. Moreover, testing hypothesized detrimental effects of TCS in a clinical trial may be unethical, unless there is a demonstrable equipoise between hypothesized harm and potential benefit (Nycum and Reid 2007).

When a clinical trial is unethical or not feasible, cohort studies may present the best design option. As TCS exposures vary over time, even cohort studies in which exposure assessment preceded outcome ascertainment may not be able to adequately establish the temporal relationship, thereby limiting causal inferences. Consider, for example, cohort studies that examined the association between prenatal TCS levels and neonatal endpoints (e.g. birth weight, head circumference, or anogenital distance) assessed either at birth or postnatally. Fetal head circumference increases from about 10 to 22 cm between gestational weeks 14 and 24, at a rate of 2.2 cm per week. In the next 10 weeks of gestation, the average head circumference growth rate is around 0.7 cm/week; and for the remainder of a full-term pregnancy (34–40 weeks), the growth slows down further to about 0.5 cm/week (Papageorghiou et al. 2014). If a maternal urine sample is obtained in late, rather than early, pregnancy, the resulting measure may miss the most relevant exposure window, and the study may be unable to establish the true temporal relation between the exposure and the outcome.

Exposure assessment

The quality of exposure assessment is a major determinant of the overall quality of any environmental epidemiology study (LaKind et al. 2014a). Areas of uncertainty in biomonitoring studies of TCS, or similar short-lived and ubiquitous chemicals, include the integrity of the sample itself as well as the optimal sample number and appropriate timing of sample collection required to adequately assign longer-term exposure status.

As TCS is widespread in the environment, in laboratory settings and even in sample collection equipment (Barr et al. 1999; Needham et al. 2007; Calafat and Needham 2008, 2009), studies must include approaches that ensure that samples are not affected by contamination during collection, shipping, storage, and analysis (USEPA 2008, 2009; Ye et al. 2013). This is particularly important for environmental phenols such as bisphenol A, TCS, and some parabens, which may be present in numerous consumer and household products.

Some short-lived chemicals can be broken down by blood enzymes, absorbed by storage containers or degrade when freezing of samples is delayed or if the samples are thawed and re-frozen (Ye et al. 2007; Provencher et al. 2014). TCS in samples stored at 4 °C and −70 °C has been reported to be stable for up to 7 and 180 days, respectively (Ye et al. 2007); to our knowledge, TCS stability outside of these storage parameters has not been explored.

Once converted to the parent compound, it is not possible to distinguish TCS derived from the biological sample and that from environmental contamination (Ye et al. 2007).

These observations indicate that documentation of analyte stability constitutes an important element of exposure assessment in studies of TCS.

Given the importance of sample integrity to the reliability and quality of the exposure assessment, it is surprising that few of the publications reviewed here documented steps taken to avoid sample contamination or demonstrate analyte stability. For example, Wang et al. (2015) included participants who likely provided urine samples in a health care setting where TCS might have been widely used (MacIsaac et al. 2014), yet gave no documentation on efforts taken to prevent sample contamination during collection, shipping, storage, or analysis.

Information showing that TCS levels remained stable from the time of the biological sample collection until the time of laboratory analysis was also not provided in this literature. For example, Lassen et al. (2016) recruited women into their study starting in January 2010 and some samples were not analyzed until December 2012. No information was given on the length of time the samples were maintained at room temperature or whether any assessment of analyte stability was conducted in order to ascertain that two years is an acceptable storage time.

We recognize that certain aspects of study design may not have been included in a publication but were still documented elsewhere (e.g. in laboratory reports). In general, we did not attempt to seek these out but rather relied on what was provided in each publication or associated publication supplement.

Perhaps the most important methodological problem in studies of short-lived chemicals is the use of a single measure as a surrogate for longer-term exposure. Error and the resulting misclassification associated with this approach is well-documented for bisphenol A and several short-lived pesticides (Teeguarden et al. 2011; Attfield et al. 2014; Morgan et al. 2016).

One approach used by the scientific community to assess temporal stability in exposure is to consider the intraclass correlation coefficient, or ICC, for the chemical of interest. The ICC is defined as the proportion of the overall variation in measurements that is explained by differences between the subject distributions (in other words, how much variation is related to within person variability and how much is related to variability between people). ICC values range from zero to one, with a value for the ICC of zero meaning that all variation is due to variability within the subject, and a value of one indicating that all variation is due to variability between subjects. The ICC values are typically divided into three categories, with an ICC <0.40 indicating poor reproducibility, an ICC of 0.40 to <0.75 indicating fair to good reproducibility, and an ICC of 0.75 or greater indicating excellent reproducibility (Fleiss 1981; Rosner 2000). Researchers will generally assume that an ICC value closer to 1 supports the use of one measurement of a short-lived chemical based on

Table 10. TCS ICCs in the published literature.

Reference	*N*/participants	Cohort	*N*/urine samples/type and time	ICC, unadjusted (95% CIs)	ICC, SG-adjusted (95% CIs)	ICC, creatinine-adjusted
Meeker et al. (2013)	105	Pregnant women, Puerto Rico	3 spot during pregnancy	0.42 (0.30, 0.55)	0.47 (0.35, 0.59)	NA
Teitelbaum et al. (2008)	35	6–10 year old children in NYC	Up to 7 spot over 6 months	0.35	NA	0.39
Weiss et al. (2015)	80	Pregnant women in Ottawa	Five during pregnancy. One 2–3 months post-partum. First sample: pooled 24-hour sample from serial spot urine voids. Remainder: spot	Similar to adjusted	within a week-day = 0.77 within a week-end day = 0.79 across the study period = 0.50	NA
Geens et al. (2015)	151 case/ 43 control	Overweight/obese and 43 lean adult men and women, Belgium	Prior to a diet program or bariatric surgery, and then 3, 6, and 12 months later. either 24-h or single void samples	NA	NA	0.476[a]
Philippat et al. (2013)	71	Pregnant women, NYC	Three serial spot samples; 1st sample collected following amniocentesis; second two taken two or more weeks apart	0.56	0.58	0.61
Bertelsen et al. (2014)	45	Pregnant women, Norway	Spot samples at gestational weeks 17, 23, and 29	0.45	NA	0.49
Koch et al. (2014)	8	Husband–wife pairs, Belgium	All urine voids over a six day period	Spot = 0.934 24-h = 0.982	NA	Spot = 0.957 24-h = 0.987
Lassen et al. (2013)	33	Men, Denmark	Two spot, three first morning and three 24-hour samples over a 3-month period	24-h samples: 0.55 (0.35, 0.73) (collected∼40 days apart)	NA	NA

NA: not available.
[a]It is unclear whether this value is based on adjusted or unadjusted levels; further a value of 0.495 is reported in the Supplement to Geens et al.

the assumption that the variability within a subject is small (i.e. one measurement represents long-term exposure). As published ICCs for TCS range from approximately 0.3 to 0.9 (Table 10), these TCS ICC values suggest reproducibility could be described as either poor, fair or excellent, depending on which published value is used.

This wide range of ICCs reflects the complications associated with relying on ICC values when exploring intra-individual variability for urinary TCS. For example, as noted by Koch et al. (2014), "In many cases the relatively high ICC values in the current study derived from large between-individual differences in exposure levels rather than from low within-individual variation … " Koch et al. further observe that a single spot sample may "substantially over- or under-represent average exposure to that individual." Similarly, referring to ICCs of around 0.5–0.6, Philippat et al. (2013) stated: "Although a correlation of this magnitude is greater than reported for many other rapidly metabolized compounds, it may still result in bias due to exposure misclassification if the putative window of susceptibility is distant from the sample collection period … To limit the effects of exposure misclassification, studies of the effects of phenol prenatal exposure on health should try to collect multiple urine samples during pregnancy."

We further explore issues surrounding the use of the ICC as the foundation for using one measurement to approximate longer term exposures here, and in detail in Supplement 2 of this paper. We ask two questions: How should a particular ICC value be interpreted in terms of misclassification probabilities? What issues ought to be considered when interpreting the ICC that is reported in a given study?

Misclassification probabilities. The following is an illustration from the TCS literature. Lassen et al. (2013) obtained serial measurements of urinary TCS levels in 33 individuals (while raw individual data for the 33 men are not given, data are shown in Figure 1 in the publication). Serial measurements for some of the men showed urinary TCS levels ranging over approximately three orders of magnitude during the 3-month period. Regarding one's ability to use one urinary measure to assign a study participant to a correct exposure tertile, Lassen et al. found that for men with three serial morning urine measurements, 13 (42%) of the 31 men had all three measurements in the same tertile within three months, while 6% had no measurements in the same tertile. For 24-h urine samples collected over a three month period, only 25% of the men had all three measurements in the same tertile.

To quantify reproducibility, consider a study in which we obtain urinary chemical measurements for 15 subjects. We can then determine the probability that the subjects are classified in their correct tertiles based on their measurements, depending on the true ICC and the number of measurements per subject. For example, if the true ICC is 0.50, and there are two measurements per subject, the chance that a subject is correctly classified is 64.9% (Figure 3). When the number of

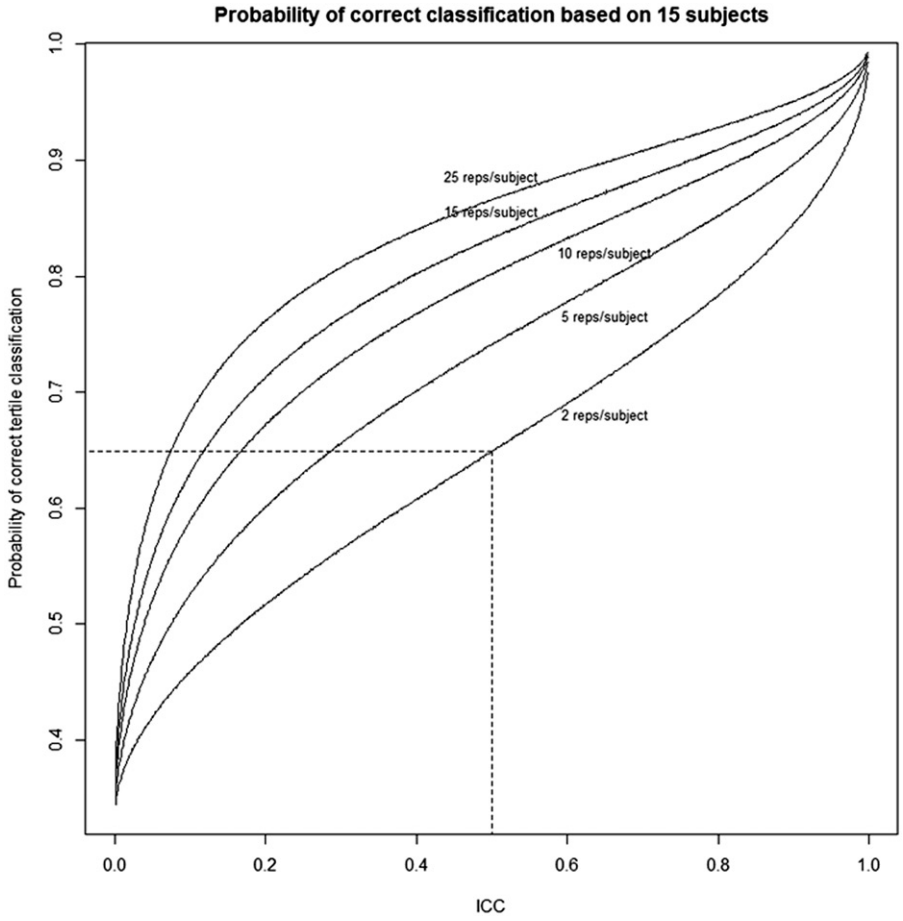

Figure 3. Probability that the sample mean for a subject falls in same tertile as that subject's true tertile, based on a study of 15 subjects. The probability is dependent on the number of measurements per subject. Each probability is based on a Monte Carlo simulation of the one-way ANOVA model with 100 000 trials.

measurements per subject is 25, this proportion becomes 86.6%. In addition, as the true ICC increases, the chance of correctly classifying the tertile of a subject increases. As the ICC approaches one, we should attain correct prediction 100% of the time. At the other extreme, when the ICC is zero, predicting the correct tertile should be equivalent to random guessing, hence a $33\frac{1}{3}$% prediction rate. In summary, the lower the value of the true ICC and the fewer the number of urine samples measured, the greater the likelihood that study will incorrectly characterize participants' exposures. Even for an ICC value that is considered to indicate "good" reproducibility, a large portion of the study population may been misclassified regarding their exposures.

Issues to be considered when interpreting reported ICCs. It is important to recognize that the ICC itself is based on modeling assumptions. Most researchers estimate ICCs under the one-way ANOVA random effects assumptions, but in fact these assumptions can be violated in a variety of ways. Specifically, measurement distributions can have different variances, they may not have the same shape after accounting for differences in spread, and they may vary over time. It is also important to recall that the ICC is an estimated quantity, and consequently it has a sampling distribution that depends on the numbers of subjects and number of replicate measurements per subject. An additional violation of ANOVA assumptions is that there is reason to expect that concentrations for a subject measured over time will exhibit serial correlation, with pairs of measurements collected at closer times tending to be more similar than pairs measured over longer time gaps. When the ICC estimation is based on measurements that are tightly clustered in time, this can lead to under-estimation of within subject variance, and hence an over-estimate of the ICC. Finally, ignoring covariates that account for between-subject variation is another potential source of over-estimation of the ICC (see Supplement 2).

In summary, the reliance on literature-based ICCs can give false confidence in our ability to properly characterize exposures to short-lived chemicals such as TCS using one measurement. Ultimately, research on temporal variability for an array of population types will need to be conducted. In addition, a discussion on "acceptable" levels of misclassification is needed.

Consistency of methods results and reporting across studies

Replication of findings is crucial for drawing inferences, making regulatory decisions, and developing policy. In the field of clinical medicine, there is consensus that replication of studies, and specifically clinical trials, is required for approval of drugs and medical devices (Berlin and Colditz 1999). By contrast, in environmental research most data on humans are observational and thus the conditions within a study are far less controlled.

Compared to their colleagues who perform clinical trials, researchers conducting observational studies have far more leeway in terms of selecting measures of exposure and outcome, choosing statistical methods, and deciding which of the analyses should be reported (Goodman et al. 2010). In those circumstances, a systematic weight-of-the-evidence assessment of observational research is only possible if the literature is sufficiently coherent to allow a PI(E)CO review (Centre for Reviews and Dissemination 2016).

One of the problems with reviewing the literature on TCS is a lack of clear *a priori* understanding of its expected effects on human health. Consider the general concern that TCS may act as an endocrine disruptor capable of altering sex hormone function (Dann and Hontela 2011). As the concept of "endocrine disruption" may have multiple possible manifestations, an important next step in the assessment of evidence from human studies is to formulate one or more specific hypotheses.

An *a priori* hypothesis stemming from clinical or basic science evidence may be described as "the cornerstone of any epidemiological inquiry" (LaKind et al. 2014a). Studies that rely on data with multiple variables offer an opportunity for multiple simultaneous hypothesis testing, which may increase efficiency, but presents a challenge with respect to proper interpretation of results. For example, data from *in vitro* assays and results of *in vivo* animal studies appear to be in disagreement on whether or not TCS is expected to exert androgenic or anti-androgenic activity (Foran et al. 2000; Chen et al. 2007; Kumar et al. 2009; Witorsch 2014). Without strong *a priori* evidence, it may be reasonable to examine two competing hypotheses: (1) higher levels of TCS exposure are associated with higher levels of testosterone in humans and (2) TCS exposure is inversely related to testosterone concentrations. If TCS indeed affects testosterone levels in either direction, a coherent body of human literature, assessed within a PI(E)CO framework, should support one of the two competing hypotheses.

A closer examination of the literature on TCS and testosterone levels shows that a PI(E)CO review is not possible. Three studies measured testosterone levels in relation to TCS exposure, but their methods of data collection, analysis, and reporting were too discordant to allow a side-by-side comparison of results. With respect to the population of interest (P), one study included healthy adult volunteers of both sexes (Poole et al. 2016), the other included men with low sperm motility counts and controls whose counts were normal (Den Hond et al. 2015), and the third was limited to children and adolescents (Scinicariello and Buser 2016). The intervention (I) or exposure (E) categories and the comparison groups (C) also differed across studies. The first study was a cross-over clinical trial that compared intervals of intentional TCS administration to the corresponding no-exposure interval for the same subjects (Poole et al. 2016). The second was a cross-sectional study that expressed urine TCS as a natural logarithm-transformed continuous variable (Den Hond et al. 2015). The third study also used cross-sectional design, but divided urinary TCS levels into quartiles (Scinicariello and Buser 2016). In terms of outcome definitions, the article by Poole et al. (2016) focused on free testosterone (although total testosterone results are available in the Supplementary materials), Scinicariello and Buser (2016) measured total testosterone only, and Den Hond et al. (2015) reported data for both. The analyses also differed across studies. Poole et al. (2016) performed a Wilcoxon test and reported the results as p values

only, Den Hond et al. (2015) used linear models and reported regression coefficients and Scinicariello and Buser (2016) expressed their results as ORs for each quartile using the first quartile as the reference category. Moreover, Scinicariello and Buser (2016) reported the association between TCS and total testosterone separately for male children, male adolescents, female children and female adolescents further complicating the inter-study comparisons.

As another example, a recent review of animal studies concluded that TCS administration results in decreased concentrations of T4 in rats (Johnson et al. 2016). This observation presents a rare opportunity to examine the human literature with a clear *a priori* formulated hypothesis that higher levels of TCS exposure should produce lower T4. Three clinical trials addressed this hypothesis. Of those, two studies (Allmyr et al. 2009; Poole et al. 2016) found no discernable effect of TCS, and one study (Cullinan et al. 2012) reported that the T4 changes after 5 years of TCS intervention, although clinically not relevant, were in the opposite of the hypothesized direction when compared to the placebo group. Cullinan et al. (2012) reported that the difference between TCS and placebo arms reflected "reduced levels in the placebo group but no change in the triclosan group." Based on their results, the authors concluded that "triclosan toothpaste had no detectable effect on thyroid function." Of the three observational studies addressing the same hypothesis, two (Koeppe et al. 2013; Aker et al. 2016) reported null results, and one (Geens et al. 2015) observed a significant inverse association that was limited to all obese participants and obese women at 0 months and obese women at 3 months.

Similar problems with attempting a PI(E)CO review can be observed with the two available studies on exposure to TCS and time to pregnancy (Velez et al. 2015; Smarr et al. 2017). The results from these studies were inconsistent, as were the study design and exposure assessment approaches. Regarding exposure, Smarr et al. (2017) measured TCS in a single spot urine sample collected within 2 months of stopping contraception while Velez et al. (2015) used a spot sample collected during the first trimester. Whether these differences impacted the reported measures of association is unknown, but other research has suggested that many factors affect TCS levels in pregnant women (e.g. time of day of urine sample collection, time since previous void, day of the week and season of sample collection, number of previous pregnancies) (Arbuckle et al. 2015; Weiss et al. 2015) and it is possible that not accounting for these could impact study findings.

When assessing consistency of reported results, it is also important to consider publication bias, defined as the "tendency on the parts of investigators or editors to fail to publish study results on the basis of the direction or strength of the study findings" (Dickersin and Min 1993). It has been demonstrated that studies with positive significant results are more likely to be published, and the tendency towards publication bias is more pronounced in observational studies than in clinical trials (Easterbrook et al. 1991). It is therefore plausible that the literature on TCS is also affected by publication bias. For example, in performing this review we identified three very similar studies that evaluated the 2011–2012 NHANES data to assess three separate outcomes (oral health, memory problems, and self-reported health status) in relation to urinary levels of heavy metals, phthalates, phenols, thiocyanate, parabens, pesticides, polyaromatic hydrocarbons, and polyfluorinated compounds (Shiue 2015a, 2015b, 2015c). Although all three analyses were based on the same NHANES data and evaluated the same chemicals, only two papers (Shiue 2015a, 2015b) included results for TCS, although there were over 500 NHANES participants with both TCS exposure and memory problem data.

Another closely related concept is selective within-study reporting defined as "selection on the basis of the results of a subset of the original variables recorded for inclusion in a publication" (Dwan et al. 2008). Consider, for example, a study from the French EDEN mother–child cohort (Philippat et al. 2014). The study included children of both sexes, but the results were presented for boys only. The number of girls available for analysis was about the same as that for boys and there is little evidence that the proposed mechanisms of effect on fetal growth (e.g. via glucocorticoid or thyroid action) are only relevant to boys. The authors justify exclusion of girls by citing two papers (Wolff et al. 2008; Harley et al. 2013) as supporting sex-specific effects. However, a closer look at these two papers fails to identify a discernable pattern. In the analyses of birth weight, Wolff et al. (2008) report a significant inverse association for 2,5-dichlorophenol in boys but not in girls, and a positive association for benzophenone-3 among boys, but an inverse association among girls. Harley et al. (2013) report inverse association between bisphenol A and body size at 9 years of age among girls, and no evidence of associations in boys. Taken together, these data do not appear to provide a good reason for limiting analysis or reporting to any particular sex.

As reviewed elsewhere (Kyzas et al. 2007), it is a common practice to highlight positive and statistically significant results in the abstract or conclusion sections of articles. Most studies included in the present review reported some positive findings. For example, Lassen et al. (2016) presented 39 different quartile-specific measures of association and 13 measures of associations that treated TCS as a continuous variable. Most results were consistent with the null hypothesis; some were inverse and some were positive. Among four statistically significant ORs, three were inverse, and one was positive. The tests for interaction were not statistically significant, justifying an overall sex-adjusted analysis. The authors concluded that their findings were "compatible with an anti-androgenic effect of prenatal TCS exposure on fetal growth in boys," but another likely explanation of the observed results is random variation stemming from multiple comparisons.

The above examples illustrate that a narrative review of stated conclusions may leave a false impression that the studies are in agreement. The inconsistencies only become evident when results from different analyses are organized and presented systematically.

Conclusions

Although the available epidemiological literature on TCS included examinations of more than 100 different health

endpoints and reported on hundreds of different measures of association, very few studies addressed similar specific research questions and used sufficiently consistent methods to allow a meaningful comparison of results. Of the reported measures of association, most were not significantly different from the null, and the few statistically significant results represented isolated findings without a discernable across- or within-study pattern.

This review indicates that the existing body of epidemiological literature on TCS does not allow a meaningful weight of the evidence assessment due to methodological limitations of individual studies and a lack of inter-study consistency. For this reason, our proposals for future research are focused on ways of improving the primary studies since the quality of the review is inevitably tied to the quality and coherence of the underlying literature.

In designing future research, it is important to take into consideration methodologically stronger studies. For example, studies assessing relatively short-term changes in health biomarkers in response to TCS exposure would benefit from following the example of the existing experimental trials described in this review. If experimental trials are not feasible, future observational studies may consider emulating some of the aspects of existing prospective cohorts, particularly those with the highest quality exposure assessment methods. It is important to point out, however, that the methods of assessing exposure to TCS are a matter of considerable uncertainty due to the overall lack of knowledge on how to properly characterize exposure to short-lived chemicals. While there appears to be a consensus that a one-time spot sample leaves too much room for misclassification, the optimal number and the correct timing of samples required for the state-of-the-art exposure assessment are not known. Further research on best approaches for characterizing exposures to short-lived chemicals with exposures impacted by the time-dependent and product-use variations by humans is needed to address this issue and to allow interpretation of these data to match our analytical capabilities.

One approach would be to conduct studies of the associations between TCS and health outcomes that are preceded by a series of methodological exposure characterization studies that could then inform well-designed epidemiological research. These methodological studies could employ timed sample collection protocols, describe various exposure profiles by measuring peak and background concentrations, and take into consideration variable product use in humans.

In the meantime, the present systematic review demonstrates that the body of published epidemiological literature on TCS is not sufficiently concordant in terms of methodology or results nor is it of sufficient quality to inform WOE assessments. The issues identified here are in no way unique to TCS and are applicable to other chemicals with similar properties. With the growing reliance on human data for public health decision-making, we are at an important juncture in terms of developing guidance for future epidemiological studies capable of producing more conclusive evidence for or against the hypothesis that TCS – or chemicals with similar physico-chemical properties – adversely affect human health.

Acknowledgements

The authors thank Dr. Brian Slezak of Colgate-Palmolive for his minor editorial comments on the manuscript. We also thank the six anonymous reviewers selected by the journal Editor for their thoughtful comments on the manuscript.

Declaration of interest

The employment affiliations of the authors are shown on the cover page. The authors have sole responsibility for the writing and content of this paper. LaKind Associates is a private consulting firm specializing in strategic risk management, assessment of human exposures and health risks, biomonitoring, state-of-the-science reviews, and environmental regulatory review; LaKind Associates consults to governmental and private sectors. The academic institutions with which the authors are affiliated are traditional academic institutions. All authors consult to both industry and government and conduct research with industry, government, and academia. The research was supported via funding from Colgate-Palmolive Company to LaKind Associates, with subcontracts to the two other coauthors. Colgate-Palmolive was not involved in the design, collection, management, analysis, or interpretation of the information in the manuscript; or in the preparation of the manuscript. This work is the exclusive professional work product of the three authors. While Colgate-Palmolive had the opportunity to offer comments on the draft manuscript, the findings and conclusions in this manuscript are those of the authors and do not necessarily represent the views of Colgate-Palmolive or any institutions with which the authors are affiliated. None of the authors have appeared in any legal or regulatory proceedings within the past 5 years related to the contents of the paper nor have been engaged to make such appearances in the future.

Supplemental material

Supplemental material for this article is available online here.

Funding

The research was supported via funding from Colgate-Palmolive Company to LaKind Associates, with subcontracts to the two other coauthors.

References

Adami H-O, Berry CL, Breckenridge CB, Smith LL, Swenberg JA, Trichopoulos D, Weiss NS, Pastoor TP. 2011. Toxicology and epidemiology: improving the science with a framework for combining toxicological and epidemiological evidence to establish causal inference. Toxicol Sci. 122:223–234.

Adgent MA, Rogan WJ. 2015. Triclosan and prescription antibiotic exposures and enterolactone production in adults. Environ Res. 142:66–71.

Aker AM, Watkins DJ, Johns LE, Ferguson KK, Soldin OP, Anzalota Del Toro LV, Alshawabkeh AN, Cordero JF, Meeker JD. 2016. Phenols and parabens in relation to reproductive and thyroid hormones in pregnant women. Environ Res. 151:30–37.

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.

Amler RW, Barone S, Jr, Belger A, Berlin CM Jr, Cox C, Frank H, Goodman M, Harry J, Hooper SR, Ladda R, et al. 2006. Hershey Medical Center Technical Workshop Report: optimizing the design and interpretation of epidemiologic studies for assessing neurodevelopmental effects from in utero chemical exposure. NeuroToxicology. 27:861–874.

Arbuckle TE, Weiss L, Fisher M, Hauser R, Dumas P, Bérubé R, Neisa A, LeBlanc A, Lang C, Ayotte P, et al. 2015. Maternal and infant exposure

to environmental phenols as measured in multiple biological matrices. Sci Total Environ. 508:575–584.
Ashley-Martin J, Dodds L, Arbuckle TE, Marshall J. 2016. Prenatal triclosan exposure and cord blood immune system biomarkers. Int J Hyg Environ Health. 219:454–457.
Attfield KR, Hughes MD, Spengler JD, Lu C. 2014. Within- and between-child variation in repeated urinary pesticide metabolite measurements over a 1-year period. Environ Health Perspect. 122:201–206.
Barr DB, Barr JR, Driskell WJ, Hill RH Jr, Ashley DL, Needham LL, Head SL, Sampson EJ. 1999. Strategies for biological monitoring of exposure for contemporary-use pesticides. Toxicol Ind Health. 15:168–179.
Beasley R, Semprini A, Mitchell EA. 2015. Risk factors for asthma: is prevention possible? Lancet. 386:1075–1085.
Bellazzi R, Zupan B. 2008. Predictive data mining in clinical medicine: current issues and guidelines. Int J Med Inform. 77:81–97.
Beran TN, Violato C. 2010. Structural equation modeling in medical research: a primer. BMC Res Notes. 3:267.
Berlin JA, Colditz GA. 1999. The role of meta-analysis in the regulatory process for foods, drugs, and devices. JAMA. 281:830–834.
Bertelsen RJ, Engel SM, Jusko TA, Calafat AM, Hoppin JA, London SJ, Eggesbø M, Aase H, Zeiner P, Reichborn-Kjennerud T, et al. 2014. Reliability of triclosan measures in repeated urine samples from Norwegian pregnant women. J Expo Sci Environ Epidemiol. 24:517–521.
Bertelsen RJ, Longnecker MP, Lovik M, Calafat AM, Carlsen KH, London SJ, Lodrup Carlsen KC. 2013. Triclosan exposure and allergic sensitization in Norwegian children. Allergy. 68:84–91.
Beyond Pesticides. 2016. Triclosan. Washington (DC). [accessed 2017 July 8]. http://www.beyondpesticides.org/programs/antibacterials/triclosan/fda-2016-decision-and-history.
Buckley JP, Herring AH, Wolff MS, Calafat AM, Engel SM. 2016. Prenatal exposure to environmental phenols and childhood fat mass in the Mount Sinai Children's Environmental Health Study. Environ Int. 91:350–356.
Buhl T, Fuchs T, Geier J. 2014. Contact hypersensitivity to triclosan. Ann Allergy Asthma Immunol. 113:119–120.
Buser MC, Murray HE, Scinicariello F. 2014. Association of urinary phenols with increased body weight measures and obesity in children and adolescents. J Pediatr. 165:744–749.
Buttke DE, Sircar K, Martin C. 2012. Exposures to endocrine-disrupting chemicals and age of menarche in adolescent girls in NHANES (2003–2008). Environ Health Perspect. 120:1613–1618.
Calafat AM, Needham LL. 2008. Factors affecting the evaluation of biomonitoring data for human exposure assessment. Int J Androl. 31:139–143.
Calafat AM, Needham LL. 2009. What additional factors beyond state-of-the-art analytical methods are needed for optimal generation and interpretation of biomonitoring data? Environ Health Perspect. 117:1481–1485.
Calafat AM, Ye X, Wong L-Y, Reidy JA, Needham LL. 2008. Urinary concentrations of triclosan in the U.S. population: 2003–2004. Environ Health Perspect. 116:303–307.
CDC (Centers for Disease Control and Prevention). 2013. Laboratory procedure manual. Analyte: benzophenone-3, bisphenol A, 2,4-dichlorophenol, 2,5-dichlorophenol, methyl-, ethyl-, propyl-, and butyl parabens, triclosan. Atlanta (GA): Organic Analytical Toxicology Branch, Division of Laboratory Sciences, National Center for Environmental Health.
Centre for Reviews and Dissemination. 2016. Guidance notes for registering a systematic review protocol with PROSPERO. York, UK: University of York.
Chen J, Ahn KC, Gee NA, Gee SJ, Hammock BD, Lasley BL. 2007. Antiandrogenic properties of parabens and other phenolic containing small molecules in personal care products. Toxicol Appl Pharmacol. 221:278–284.
Chen M, Tang R, Fu G, Xu B, Zhu P, Qiao S, Chen X, Xu B, Qin Y, Lu C, et al. 2013. Association of exposure to phenols and idiopathic male infertility. J Hazard Mater. 250–251:115–121.
Chevrier C, Petit C, Philippat C, Mortamais M, Slama R, Rouget F, Calafat AM, Ye X, Silva MJ, Charles MA, et al. 2012. Maternal urinary phthalates and phenols and male genital anomalies. Epidemiology. 23:353–356.
Clayton EM, Todd M, Dowd JB, Aiello AE. 2011. The impact of bisphenol A and triclosan on immune parameters in the U.S. population, NHANES 2003-2006. Environ Health Perspect. 119:390–396.
Concato J, Feinstein AR, Holford TR. 1993. The risk of determining risk with multivariable models. Ann Intern Med. 118:201–210.
Cullinan MP, Palmer JE, Carle AD, West MJ, Seymour GJ. 2012. Long term use of triclosan toothpaste and thyroid function. Sci Total Environ. 416:75–79.
Cullinan MP, Palmer JE, Carle AD, West MJ, Westerman B, Seymour GJ. 2015a. The influence of a triclosan toothpaste on adverse events in patients with cardiovascular disease over 5-years. Sci Total Environ. 508:546–552.
Cullinan MP, Palmer JE, Faddy MJ, Westerman B, Carle AD, West MJ, Seymour GJ. 2015b. The influence of triclosan on biomarkers of cardiovascular risk in patients in the Cardiovascular and Periodontal Study (CAPS): a randomized controlled trial. J Periodontol. 86:847–855.
Dann AB, Hontela A. 2011. Triclosan: environmental exposure, toxicity and mechanisms of action. J Appl Toxicol. 31:285–311.
Den Hond E, Paulussen M, Geens T, Bruckers L, Baeyens W, David F, Dumont E, Loots I, Morrens B, de Bellevaux BN, et al. 2013. Biomarkers of human exposure to personal care products: results from the Flemish Environment and Health Study (FLEHS 2007-2011). Sci Total Environ. 463–464:102–110.
Den Hond E, Tournaye H, De Sutter P, Ombelet W, Baeyens W, Covaci A, Cox B, Nawrot TS, Van Larebeke N, D'Hooghe T. 2015. Human exposure to endocrine disrupting chemicals and fertility: a case-control study in male subfertility patients. Environ Int. 84:154–160.
Dickersin K, Min YI. 1993. Publication bias: the problem that won't go away. Ann N Y Acad Sci. 703:135–146.
Dwan K, Altman DG, Arnaiz JA, Bloom J, Chan AW, Cronin E, Decullier E, Easterbrook PJ, Von Elm E, Gamble C, et al. 2008. Systematic review of the empirical evidence of study publication bias and outcome reporting bias. PLoS One. 3:e3081.
Easterbrook PJ, Berlin JA, Gopalan R, Matthews DR. 1991. Publication bias in clinical research. Lancet. 337:867–872.
EU. 2016. Commission Implementing Decision (EU) 2016/110 of 27 January 2016 not approving triclosan as an existing active substance for use in biocidal products for product-type 1; [accessed 2017 July 8]. http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv%3AOJ.L_.2016.021.01.0086.01.ENG.
Fleiss J. 1981. Statistical methods for rates and proportions. New York: Wiley.
Foran CM, Bennett ER, Benson WH. 2000. Developmental evaluation of a potential non-steroidal estrogen: triclosan. Mar Environ Res. 50:153–156.
Geens T, Dirtu AC, Dirinck E, Malarvannan G, Van Gaal L, Jorens PG, Covaci A. 2015. Daily intake of bisphenol A and triclosan and their association with anthropometric data, thyroid hormones and weight loss in overweight and obese individuals. Environ Int. 76:98–105.
Geer LA, Pycke BF, Waxenbaum J, Sherer DM, Abulafia O, Halden RU. 2017. Association of birth outcomes with fetal exposure to parabens, triclosan and triclocarban in an immigrant population in Brooklyn, New York. J Hazard Mater. 323(Pt A):177–183.
Goodman M, LaKind JS, Mattison DR. 2014a. Do phthalates act as obesogens in humans? A systematic review of the epidemiological literature. Crit Rev Toxicol. 44:151–175.
Goodman M, Mandel JS, DeSesso JM, Scialli AR. 2014b. Atrazine and pregnancy outcomes: a systematic review of epidemiologic evidence. Birth Defects Res B Dev Reprod Toxicol. 101:215–236.
Goodman M, Squibb K, Youngstrom E, Anthony LG, Kenworthy L, Lipkin PH, Mattison DR, LaKind JS. 2010. Using systematic reviews and meta-analyses to support regulatory decision making for neurotoxicants: lessons learned from a case study of PCBs. Environ Health Perspect. 118:727–734.
Greenland S. 1996. Basic methods for sensitivity analysis of biases. Int J Epidemiol. 25:1107–1116.
Guyatt GH, Haynes RB, Jaeschke RZ, Cook DJ, Green L, Naylor CD, Wilson MC, Richardson WS. 2000a. Users' Guides to the Medical Literature:

XXV. Evidence-based medicine: principles for applying the Users' Guides to patient care. Evidence-Based Medicine Working Group. JAMA. 284:1290–1296.

Guyatt GH, Naylor D, Richardson WS, Green L, Haynes RB, Wilson MC, Cook DJ, Jaeschke RZ. 2000b. What is the best evidence for making clinical decisions? JAMA. 284:3127–3128.

Harley KG, Schall RA, Chevrier J, Tyler K, Aguirre H, Bradman A, Holland NT, Lustig RH, Calafat AM, Eskenazi B. 2013. Prenatal and postnatal bisphenol A exposure and body mass index in childhood in the CHAMACOS cohort. Environ Health Perspect. 121:514–520.

Health Canada. 2015. Third report on human biomonitoring of environmental chemicals in Canada. Ottawa: Minister of Health; [accessed 2016 Dec 19]. https://www.canada.ca/en/health-canada/services/environmental-workplace-health/reports-publications/environmental-contaminants/third-report-human-biomonitoring-environmental-chemicals-canada.html.

Hernan MA, Hernandez-Diaz S, Werler MM, Mitchell AA. 2002. Causal knowledge as a prerequisite for confounding evaluation: an application to birth defects epidemiology. Am J Epidemiol. 155:176–184.

Ihde ES, Loh JM, Rosen L. 2015. Association of environmental chemicals & estrogen metabolites in children. BMC Endocr Disord. 15:83.

Johnson PI, Koustas E, Vesterinen HM, Sutton P, Atchley DS, Kim AN, Campbell M, Donald JM, Sen S, Bero L, et al. 2016. Application of the navigation guide systematic review methodology to the evidence for developmental and reproductive toxicity of triclosan. Environ Int. 92–93:716–728.

Kahan BC, Rehal S, Cro S. 2015. Risk of selection bias in randomised trials. Trials. 16:405.

Koch HM, Aylward LL, Hays SM, Smolders R, Moos RK, Cocker J, Jones K, Warren N, Levy L, Bevan R. 2014. Inter- and intra-individual variation in urinary biomarker concentrations over a 6-day sampling period. Part 2: personal care product ingredients. Toxicol Lett. 231:261–269.

Koeppe ES, Ferguson KK, Colacino JA, Meeker JD. 2013. Relationship between urinary triclosan and paraben concentrations and serum thyroid measures in NHANES 2007–2008. Sci Total Environ. 445–446:299–305.

Kumar V, Chakraborty A, Kural MR, Roy P. 2009. Alteration of testicular steroidogenesis and histopathology of reproductive system in male rats treated with triclosan. Reprod Toxicol. 27:177–185.

Kyzas PA, Denaxa-Kyza D, Ioannidis JP. 2007. Almost all articles on cancer prognostic markers report statistically significant results. Eur J Cancer. 43:2559–2579.

LaKind JS, Goodman M, Barr DB, Weisel CP, Schoeters G. 2015b. Lessons learned from the application of BEES-C: systematic assessment of study quality of epidemiologic research on BPA, neurodevelopment, and respiratory health. Environ Int. 80:41–71.

LaKind JS, Goodman M, Makris SL, Mattison DR. 2015a. Improving concordance in environmental epidemiology: a three-part proposal. J Toxicol Environ Health B Crit Rev. 18:105–120.

LaKind JS, Goodman M, Mattison DR. 2014b. Bisphenol A and indicators of obesity, glucose metabolism/type 2 diabetes and cardiovascular disease: a systematic review of epidemiologic research. Crit Rev Toxicol. 44:121–150.

LaKind JS, Goodman M, Naiman DQ. 2012. Use of NHANES data to link chemical exposures to chronic diseases: a cautionary tale. PLoS One. 7:e51086.

LaKind JS, Sobus JR, Goodman M, Barr DB, Fürst P, Albertini RJ, Arbuckle TE, Schoeters G, Tan Y-M, Teeguarden J, et al. 2014a. A proposal for assessing study quality: biomonitoring, Environmental Epidemiology, and Short-Lived Chemicals (BEES-C) Instrument. Environ Int. 73C:195–207.

Lankester J, Patel C, Cullen MR, Ley C, Parsonnet J. 2013. Urinary triclosan is associated with elevated body mass index in NHANES. PLoS One. 8:e80057.

Lassen TH, Frederiksen H, Jensen TK, Petersen JH, Main KM, Skakkebæk NE, Jørgensen N, Kranich SK, Andersson AM. 2013. Temporal variability in urinary excretion of bisphenol A and seven other phenols in spot, morning, and 24-h urine samples. Environ Res. 126:164–170.

Lassen TH, Frederiksen H, Kyhl HB, Swan SH, Main KM, Andersson AM, Lind DV, Husby S, Wohlfahrt-Veje C, Skakkebaek NE, et al. 2016. Prenatal triclosan exposure and anthropometric measures including anogenital distance in Danish infants. Environ Health Perspect. 124:1261–1268.

Li S, Zhao J, Wang G, Zhu Y, Rabito F, Krousel-Wood M, Chen W, Whelton PK. 2015. Urinary triclosan concentrations are inversely associated with body mass index and waist circumference in the US general population: experience in NHANES 2003–2010. Int J Hyg Environ Health. 218:401–406.

Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JP, Clarke M, Devereaux PJ, Kleijnen J, Moher D. 2009. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. J Clin Epidemiol. 62:e1–e34.

MacIsaac JK, Gerona RR, Blanc PD, Apatira L, Friesen MW, Coppolino M, Janssen S. 2014. Health care worker exposures to the antibacterial agent triclosan. J Occup Environ Med. 56:834–839.

Meeker JD, Cantonwine DE, Rivera-González LO, Ferguson KK, Mukherjee B, Calafat AM, Ye X, Anzalota Del Toro LV, Crespo-Hernández N, Jiménez-Vélez B, et al. 2013. Distribution, variability, and predictors of urinary concentrations of phenols and parabens among pregnant women in Puerto Rico. Environ Sci Technol. 47:3439–3447.

Morgan MK, Sobus JR, Barr DB, Croghan CW, Chen FL, Walker R, Alston L, Andersen E, Clifton MS. 2016. Temporal variability of pyrethroid metabolite levels in bedtime, morning, and 24-h urine samples for 50 adults in North Carolina. Environ Res. 144:81–91.

NAS (National Academy of Sciences). 2014. Review of EPA's Integrated Risk Information System (IRIS). Process Committee to Review the IRIS Process; Board on Environmental Studies and Toxicology; Division on Earth and Life Studies; National Research Council. 78-0-309-30414-6.

National Research Council (NRC). 2006. Human biomonitoring for environmental chemicals. Washington (DC): The National Academies Press.

Needham LL, Calafat AM, Barr DB. 2007. Uses and issues of biomonitoring. Int J Hyg Environ Health. 210:229–238.

Nycum G, Reid L. 2007. The harm-benefit tradeoff in "bad deal" trials. Kennedy Inst Ethics J. 17:321–350.

OHAT (Office of Health Assessment and Translation). 2015. Handbook for conducting a literature-based health assessment using OHAT approach for systematic review and evidence integration. Research Triangle Park (NC): National Toxicology Program, National Institute of Environmental Health Sciences.

Papageorghiou AT, Ohuma EO, Altman DG, Todros T, Ismail LC, Lambert A, Jaffer YA, Bertino E, Gravett MG, Purwar M, et al. 2014. International standards for fetal growth based on serial ultrasound measurements: the Fetal Growth Longitudinal Study of the INTERGROWTH-21st Project. The Lancet. 384:869–879.

Pearce N. 2012. Classification of epidemiological study designs. Int J Epidemiol. 41:393–397.

Petticrew M, Roberts H. 2003. Evidence, hierarchies, and typologies: horses for courses. J Epidemiol Community Health. 57:527–529.

Philippat C, Botton J, Calafat AM, Ye X, Charles MA, Slama R, ES Group. 2014. Prenatal exposure to phenols and growth in boys. Epidemiology. 25:625–635.

Philippat C, Mortamais M, Chevrier C, Petit C, Calafat AM, Ye X, Silva MJ, Brambilla C, Pin I, Charles MA, et al. 2012. Exposure to phthalates and phenols during pregnancy and offspring size at birth. Environ Health Perspect. 120:464–470.

Philippat C, Wolff MS, Calafat AM, Ye X, Bausell R, Meadows M, Stone J, Slama R, Engel SM. 2013. Prenatal exposure to environmental phenols: concentrations in amniotic fluid and variability in urinary concentrations during pregnancy. Environ Health Perspect. 121:1225–1231.

Pleil JD, Sobus JR. 2013. Estimating lifetime risk from spot biomarker data and intraclass correlation coefficients (ICC). J Toxicol Environ Health A. 76:747–766.

Poole AC, Pischel L, Ley C, Suh G, Goodrich JK, Haggerty TD, Ley RE, Parsonnet J. 2016. Crossover control study of the effect of personal care products containing triclosan on the microbiome. mSphere. 1:e00056-15.

Potischman N, Weed DL. 1999. Causal criteria in nutritional epidemiology. Am J Clin Nutr. 69:1309S–1314S.

Preau JL, Jr, Wong LY, Silva MJ, Needham LL, Calafat AM. 2010. Variability over 1 week in the urinary concentrations of metabolites of diethyl

phthalate and di(2-ethylhexyl) phthalate among eight adults: an observational study. Environ Health Perspect. 118:1748–1754.

Provencher G, Bérubé R, Dumas P, Bienvenu JF, Gaudreau E, Bélanger P, Ayotte P. 2014. Determination of bisphenol A, triclosan and their metabolites in human urine using isotope-dilution liquid chromatography–tandem mass spectrometry. J Chromatogr A. 1348:97–104.

Roberts C, Torgerson D. 1998. Randomisation methods in controlled trials. BMJ. 317:1301.

Rooney AA, Boyles AL, Wolfe MS, Bucher JR, Thayer KA. 2014. Systematic review and evidence integration for literature-based environmental health science assessments. Environ Health Perspect. 122:711–718.

Rosner B. 2000. Fundamentals of biostatistics. 5th ed. Pacific Grove (CA): Duxbury.

Rothman KJ, Greenland S. 1998. Modern epidemiology. Philadelphia (PA): Lippincott Williams and Wilkins.

Rothman KJ, Greenland S. 2005. Causation and causal inference in epidemiology. Am J Public Health. 95 Suppl 1:S144–S150.

Savage JH, Johns CB, Hauser R, Litonjua AA. 2014. Urinary triclosan levels and recent asthma exacerbations. Ann Allergy Asthma Immunol. 112:179–181e172.

Savage JH, Matsui EC, Wood RA, Keet CA. 2012. Urinary levels of triclosan and parabens are associated with aeroallergen and food sensitization. J Allergy Clin Immunol. 130:453–460e457.

Scinicariello F, Buser MC. 2016. Serum testosterone concentrations and urinary bisphenol A, benzophenone-3, triclosan, and paraben levels in male and female children and adolescents: NHANES 2011–2012. Environ Health Perspect. 124:1898–1904.

Shea BJ, Grimshaw JM, Wells GA, Boers M, Andersson N, Hamel C, Porter AC, Tugwell P, Moher D, Bouter LM. 2007. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. BMC Med Res Methodol. 7:10.

Shiue I. 2015c. Arsenic, heavy metals, phthalates, pesticides, hydrocarbons and polyfluorinated compounds but not parabens or phenols are associated with adult remembering condition: US NHANES, 2011–2012. Environ Sci Pollut Res. 22:6381–6386.

Shiue I. 2015a. Urinary arsenic, heavy metals, phthalates, pesticides, polyaromatic hydrocarbons but not parabens, polyfluorinated compounds are associated with self-rated health: USA NHANES, 2011–2012. Environ Sci Pollut Res Int. 22:9570–9574.

Shiue I. 2015b. Urinary heavy metals, phthalates, phenols, thiocyanate, parabens, pesticides, polyaromatic hydrocarbons but not arsenic or polyfluorinated compounds are associated with adult oral health: USA NHANES, 2011–2012. Environ Sci Pollut Res Int. 22:15636–15645.

Smarr MM, Sundaram R, Honda M, Kannan K, Buck Louis GM. 2017. Urinary concentrations of parabens and other antimicrobial chemicals and their association with couples' fecundity. Environ Health Perspect. 125:730–736.

Spanier AJ, Fausnight T, Camacho TF, Braun JM. 2014. The associations of triclosan and paraben exposure with allergen sensitization and wheeze in children. Allergy Asthma Proc. 35:475–481.

Teeguarden JG, Calafat AM, Ye X, Doerge DR, Churchwell MI, Gunawan R, Graham MK. 2011. Twenty-four hour human urine and serum profiles of bisphenol a during high-dietary exposure. Toxicol Sci. 123:48–57.

Teitelbaum SL, Britton JA, Calafat AM, Ye X, Silva MJ, Reidy JA, Galvez MP, Brenner BL, Wolff MS. 2008. Temporal variability in urinary concentrations of phthalate metabolites, phytoestrogens and phenols among minority children in the United States. Environ Res. 106:257–269.

USEPA (US Environmental Protection Agency). 2008. National functional guidelines for superfund organic methods data review. OSWER 9240.1-48 USEPA-540-R-08-01; [accessed 2016 Dec 17]. http://nepis.epa.gov/Exe/ZyPURL.cgi?Dockey=P100C7ST.TXT.

USEPA (US Environmental Protection Agency). 2009. Region III fact sheet quality control tools: blanks; [accessed 2016 Dec 17]. https://www.epa.gov/sites/production/files/2015-06/documents/blanks.pdf.

USEPA (US Environmental Protection Agency). 2010. Framework document: DRAFT framework for incorporating human epidemiologic & incident data in health risk assessment; [accessed 2016 Dec 19]. https://www.regulations.gov/document?D=EPA-HQ-OPP-2009-0851-0004.

USEPA (US Environmental Protection Agency). 2016. Glyphosate issue paper: evaluation of carcinogenic potential EPA's Office of Pesticide Programs.

USFDA (US Food and Drug Administration). 2016. FDA issues final rule on safety and effectiveness of antibacterial soaps; [accessed 2016 July 8]. https://www.fda.gov/newsevents/newsroom/pressannouncements/ucm517478.htm.

Velez MP, Arbuckle TE, Fraser WD. 2015. Female exposure to phenols and phthalates and time to pregnancy: the Maternal-Infant Research on Environmental Chemicals (MIREC) Study. Fertil Steril. 103:1011–1020e1012.

Wang X, Chen X, Feng X, Chang F, Chen M, Xia Y, Chen L. 2015. Triclosan causes spontaneous abortion accompanied by decline of estrogen sulfotransferase activity in humans and mice. Sci Rep. 5:18252.

Watkins DJ, Ferguson KK, Anzalota Del Toro LV, Alshawabkeh AN, Cordero JF, Meeker JD. 2015. Associations between urinary phenol and paraben concentrations and markers of oxidative stress and inflammation among pregnant women in Puerto Rico. Int J Hyg Environ Health. 218:212–219.

Weiss L, Arbuckle TE, Fisher M, Ramsay T, Mallick R, Hauser R, LeBlanc A, Walker M, Dumas P, Lang C. 2015. Temporal variability and sources of triclosan exposure in pregnancy. Int J Hyg Environ Health. 218:507–513.

Wielgomas B. 2013. Variability of urinary excretion of pyrethroid metabolites in seven persons over seven consecutive days—implications for observational studies. Toxicol Lett. 221:15–22.

Witorsch RJ, Thomas JA. 2010. Personal care products and endocrine disruption: a critical review of the literature. Crit Rev Toxicol. 40:1–30.

Witorsch RJ. 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 MS, Engel SM, Berkowitz GS, Ye X, Silva MJ, Zhu C, Wetmur J, Calafat AM. 2008. Prenatal phenol and phthalate exposures and birth outcomes. Environ Health Perspect. 116:1092–1097.

Wolff MS, Teitelbaum SL, McGovern K, Pinney SM, Windham GC, Galvez M, Pajak A, Rybak M, Calafat AM, Kushi LH, et al. 2015. Environmental phenols and pubertal development in girls. Environ Int. 84:174–180.

Wolff MS, Teitelbaum SL, Pinney SM, Windham G, Liao L, Biro F, Kushi LH, Erdmann C, Hiatt RA, Rybak ME, 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.

Xue J, Wu Q, Sakthivel S, Pavithran PV, Vasukutty JR, Kannan K. 2015. Urinary levels of endocrine-disrupting chemicals, including bisphenols, bisphenol A diglycidyl ethers, benzophenones, parabens, and triclosan in obese and non-obese Indian children. Environ Res. 137:120–128.

Ye X, Bishop AM, Reidy JA, Needham LL, Calafat AM. 2007. Temporal stability of the conjugated species of bisphenol A, parabens, and other environmental phenols in human urine. J Expo Sci Environ Epidemiol. 17:567–572.

Ye X, Zhou X, Hennings R, Kramer J, Calafat AM. 2013. Potential external contamination with bisphenol A and other ubiquitous organic environmental chemicals during biomonitoring analysis: an elusive laboratory challenge. Environ Health Perspect. 121:283–286.

You L, Zhu X, Shrubsole MJ, Fan H, Chen J, Dong J, Hao CM, Dai Q. 2011. Renal function, bisphenol A, and alkylphenols: results from the National Health and Nutrition Examination Survey (NHANES 2003–2006). Environ Health Perspect. 119:527–533.

Youngstrom E, Kenworthy L, Lipkin PH, Goodman M, Squibb K, Mattison DR, Anthony LG, Makris SL, Bale A, Raffaele KC, et al. 2011. A proposal to facilitate weight-of-evidence assessments: harmonization of neurodevelopmental environmental epidemiology studies (HONEES). Neurotoxicol Teratol. 33:354–359.

Yueh MF, Tukey RH. 2016. Triclosan: a widespread environmental toxicant with many biological effects. Annu Rev Pharmacol Toxicol. 56:251–272.

Zhu W, Zhang H, Tong C, Xie C, Fan G, Zhao S, Yu X, Tian Y, Zhang J. 2016. Environmental exposure to triclosan and semen quality. Int J Environ Res Public Health. 13:224.