

# Systematic Review and Meta-Analysis of Early-Life Exposure to Bisphenol A and Obesity-Related Outcomes in Rodents

Pim Nicolaas Hubertus Wassenaar,<sup>1</sup> Leonardo Trasande,<sup>2,3,4,5,6</sup> and Juliette Legler<sup>1,7</sup>

<sup>1</sup>Department of Chemistry and Biology, Institute for Environmental Studies, Vrije Universiteit Amsterdam, Amsterdam, Netherlands

<sup>2</sup>Department of Pediatrics, New York University School of Medicine, New York, New York, USA

<sup>3</sup>Department of Environmental Medicine, New York University School of Medicine, New York, New York, USA

<sup>4</sup>Department of Population Health, New York University School of Medicine, New York, New York, USA

<sup>5</sup>New York University Wagner School of Public Service, New York, New York, USA

<sup>6</sup>New York University Global Institute of Public Health, New York, New York, USA

<sup>7</sup>Department of Life Sciences, Division of Biosciences, Institute of Environment, Health and Societies, College of Health and Life Sciences, Brunel University London, Uxbridge, UK

**BACKGROUND:** Early-life exposure to bisphenol A (BPA) has been implicated to play a role in the development of obesity.

**OBJECTIVE:** A systematic review with meta-analyses of experimental rodent studies was conducted to answer the following question: does early-life exposure to BPA affect the obesity-related outcomes body weight, fat (pad) weight, and circulating and tissue levels of triglycerides, free fatty acids (FFA), and leptin?

**METHODS:** The methodology was prespecified in a rigorous protocol using the Systematic Review Centre for Laboratory Animal Experimentation (SYRCLE) approach. Using PubMed and EMBASE, we identified 61 articles that met the inclusion criteria. The risk of bias and the methodological quality of these articles were assessed using the SYRCLE Risk of Bias tool, and a confidence-rating methodology was used to score the quality of evidence. Meta-analyses were performed using random effect models and standardized mean differences (SMDs), or, where possible, mean differences (MDs) were calculated.

**RESULTS:** Overall summary estimates indicated significant positive associations between BPA and fat weight [SMD = 0.67 (95% CI: 0.53, 0.81)], triglycerides [SMD = 0.97 (95% CI: 0.53, 1.40)], and FFA [SMD = 0.86 (95% CI: 0.50, 1.22)], and a nonsignificant positive association with leptin levels [MD = 0.37 (95% CI: -0.14, 0.87)] and a significant negative association with body weight were estimated [MD = -0.22 (95% CI: -0.37, -0.06)]. Subgroup analyses revealed stronger positive associations for most outcome measures in males and at doses below the current U.S. reference dose of 50 µg/kg/d compared with doses above the reference dose. It should be noted that there was substantial heterogeneity across studies for all outcomes assessed and that there was insufficient information to assess risk of bias for most studies.

**CONCLUSIONS:** Findings from our systematic review suggest that early-life exposure to BPA may increase adiposity and circulating lipid levels in rodents. <https://doi.org/10.1289/EHP1233>

## Introduction

The prevalence of obesity is increasing worldwide, with currently >600,000,000 obese adults (WHO 2015). Obesity is defined by the World Health Organization (WHO) as “abnormal or excessive fat accumulation that may impair health” and is classified using the body mass index (BMI; WHO 2015). In addition to an increase in body weight or fat accumulation, obesity is also related to increased levels of circulating triglycerides, free fatty acids (FFA), and leptin (Boden 2008; Considine et al. 1996; Galic et al. 2010; Subramanian and Chait 2012). Although energy imbalance is considered the major cause of obesity, an accumulating body of evidence suggests that other risk factors such as exposure to endocrine-disrupting chemicals (EDCs) also contribute to the development of obesity. In particular, early-life exposure to obesogens may result in a higher susceptibility to developing obesity (Grün and Blumberg 2006; Heindel et al.

2015). Obesogens are chemicals that alter hormonal pathways that regulate lipid metabolism and thereby stimulate adipocyte differentiation and a predisposition to obesity; and/or increase the susceptibility to obesity and related metabolic disorders (Grün and Blumberg 2006). Several chemicals are suggested to have putative obesogenic effects, including bisphenol A (BPA), which may exert obesogenic effects through various pathways, including by its activity as an estrogen and glucocorticoid receptor agonist, as well as by interference with thyroid hormone pathways and by activation of peroxisome proliferator-activated receptor- $\gamma$  (PPAR $\gamma$ ; Ahmed and Atlas 2016; Rubin 2011). BPA is a high-production-volume chemical that is mainly used to make plastics and epoxy resins that are used in food packaging, coatings, and linings. BPA can leach into food, leading to widespread human exposure (NTP 2010). BPA has been detected in 93% of urine samples in the United States (Calafat et al. 2005) as well as in amniotic fluid, neonatal blood, placenta, cord blood, and human breast milk (Vandenberg et al. 2007). The U.S. Environmental Protection Agency (EPA) has set the reference dose (RfD), that is to say, the tolerable daily BPA exposure for the human population without an appreciable risk of deleterious effects during the lifetime, at 50 µg/kg/d, based on a lowest observed adverse effect level (LOAEL) in rodent studies of 50 mg/kg/d (U.S. EPA 1988). Recently, the European Food Safety Authority (EFSA) reevaluated the toxicological data for BPA in rodents and identified adverse effects on liver and kidney as well as on the mammary gland at levels <50 mg/kg/d and subsequently lowered the current tolerable daily intake level (TDI, equivalent to RfD in the United States) to 4 µg/kg/d (EFSA 2015).

To our knowledge, only a limited number of animal studies examining the metabolic or obesogenic effects of prenatal

---

Address correspondence to J. Legler, Institute for Environment, Health and Societies, Brunel University London, Uxbridge, UK. Telephone: 44-1895-266789. Email: [juliette.legler@brunel.ac.uk](mailto:juliette.legler@brunel.ac.uk)

Supplemental Material is available online (<https://doi.org/10.1289/EHP1233>).

The authors declare they have no actual or potential competing financial interests.

Received 21 October 2016; Revised 27 June 2017; Accepted 5 July 2017; Published 5 October 2017.

**Note to readers with disabilities:** *EHP* strives to ensure that all journal content is accessible to all readers. However, some figures and Supplemental Material published in *EHP* articles may not conform to 508 standards due to the complexity of the information being presented. If you need assistance accessing journal content, please contact [ehponline@niehs.nih.gov](mailto:ehponline@niehs.nih.gov). Our staff will work with you to assess and meet your accessibility needs within 3 working days.

**Table 1.** PECO statement (population, exposure, comparator and outcomes).

| Variable   | Description  |
|------------|--|
| Population | Experimental rodent studies  |
| Exposure   | Early-life exposure to bisphenol A (during gestation and/or lactation up to postnatal day 21)  |
| Comparator | Animals exposed to vehicle-only treatment  |
| Outcomes   | Body weight, fat (pad) weights, triglyceride levels, free fatty acids levels and leptin levels |

exposure to BPA were considered in the reevaluation of the rodent data for establishing regulatory limits (EFSA 2015) despite mounting evidence for a role for BPA in altering body-weight homeostasis via effects on the neuroendocrine system, the pancreas, the liver and/or adipocyte tissue (reviewed in Le Corre et al. 2015; Rubin 2011; Vom Saal et al. 2012). The aim of this study was to systematically review experimental rodent studies reporting early-life exposure to BPA and metabolic outcomes in order to provide a more rigorous evaluation of the existing rodent data. We reviewed and performed meta-analyses of rodent studies reporting pre- or perinatal exposure to BPA and the following obesity-related outcome measures: body weight (because there is nothing equivalent to BMI for use in rodent models), fat (pad) weight, circulating or tissue levels of triglycerides and FFA, and circulating levels of leptin (Table 1). We determined the quality of the studies and rated the confidence of the evidence using established methodologies.

## Methods

The methodology of this systematic review was prespecified in a protocol and followed the guidelines of the Systematic Review Centre for Laboratory Animal Experimentation (SYRCLE; de Vries et al. 2015). This protocol has been published on the SYRCLE website (Wassenaar and Legler 2015) as “The effects of early-life exposure to endocrine disrupting chemicals on obesity development in rodents: a systematic review” (2015) (also see Supplemental Material).

### Search Strategy

Articles published before September 21, 2015 were identified in MEDLINE (via PubMed) and EMBASE. Following the SYRCLE methodology (Leenaars et al. 2012), a comprehensive search strategy was developed and included the search components “endocrine disrupting chemicals,” “obesity,” and “rodents” (see Excel Table S1). The initial search component for “endocrine disrupting chemicals” was broader than BPA only because this design may allow investigation of other EDCs in the future. To detect all rodent studies, animal search filters (de Vries et al. 2014; Hooijmans et al. 2010b) were modified to only include rodents. In addition, the reference lists of the included articles and those of relevant reviews were screened manually for potentially relevant new articles.

### Selection of Papers

Study selection consisted of two screening phases. The first selection was based on title and abstract screening, and the second selection was based on a full-text screening. One reviewer (P.W.) conducted the two screening phases, and in case of doubt, a second reviewer (J.L.) was consulted.

Studies were selected for full-text screening when they met the inclusion criteria. In case of doubt, articles were also analyzed based on their full text. Studies were included in this systematic review when they met all of the following criteria: *a*) original full paper that presented unique data; *b*) exposure to BPA; *c*) obesity-related article or at least one of the outcome measures was examined (body weight, fat pad weights, triglyceride levels, FFA

levels, or leptin levels); *d*) experimental rodent study; *e*) perinatal exposure via maternal or direct pup exposure [during gestation and/or lactation up to postnatal day (PND) 21]. Studies were excluded if they met one of the following criteria: *a*) not an original paper; *b*) exposure to a chemical other than BPA; *c*) no disease or outcome of interest (no obesity-related outcome); *d*) not a rodent study; *e*) not perinatal exposure (paternal exposure, exposure after PND21, and measurements in unborn fetuses were excluded); *f*) outcomes not measured in F1 generation; *g*) unhealthy or genetically altered rodents (data measured after ovariectomy were considered unhealthy and were not extracted); *h*) outcomes were measured after diet was altered to high-fat diet during follow-up. In addition, selection was restricted to English-language articles.

### Study Characteristics and Data Extraction

The following characteristics were extracted from the included studies: bibliographic data (authors, year of publication, journal of publication, conflict of interest section, and funding source), animal model characteristics (species, strain, and sex), exposure characteristics (chemical, life stage and duration, dose, frequency, and route of exposure), study design characteristics (number of animals in experimental and control groups, duration of follow-up, and timing of data collection), and outcome measures [types of outcomes measured: body weight, fat (pad) weights, triglyceride levels, FFA levels and/or leptin levels, and compartment of outcome measured]. From each study, we considered each analysis of a specific outcome measure with a specific dose and/or sex as a separate individual comparison. In addition, analyses in different fat pads (even when derived from the same animal) and analyses with different time windows of exposure were considered as separate comparisons. Consequently, multiple comparisons could have been included from one study. One reviewer extracted the data (P.W.), and in case of doubt, a second reviewer was consulted (J.L.).

All outcome data were collected for each individual comparison as the mean, standard deviation (SD), and number of animals per group. When raw data or group averages with standard error (SE) and number of animals per group were reported, the mean and SD values were recalculated. When the group size was reported as a range, the smallest number of animals was used for the meta-analysis in the interest of conservative estimates. In cases where data were presented graphically and not as text or in tables, data were extracted in pixels using a digital screen ruler. In cases of missing or unclear data, including SDs or SEs that were not provided or that could not be extracted from figures owing to many crossing lines, the authors were contacted via email for the original data. If no author contact details were available, or if no response was obtained from the authors within three weeks after repeated contact, the data were omitted from analysis. When outcomes of comparisons were measured at different time points, the time point with greatest efficacy was used (i.e., the time point with the strongest association with the outcome). The time point of greatest efficacy was selected over the other time point(s) when the absolute difference between the means of the exposure and control groups divided by the sum of the SDs was the highest. By using the absolute difference, the

direction of the effect was not considered in the selection of the time point of greatest efficacy.

### **Risk of Bias and Methodological Quality Assessment**

To assess the risk of bias in the included studies, SYRCLE's Risk of Bias (RoB) tool was used; this tool is specifically designed for animal studies (Hooijmans et al. 2014). The RoB tool consists of ten questions which can be used to detect selection, performance, detection, attrition, and reporting bias in the included studies (see Figure 2 for a complete list of the ten items included in the RoB tool). The items in the RoB tool were scored with "yes," indicating low risk of bias; "no," indicating high risk of bias; or "unclear," indicating that the item was not reported, and therefore, the risk of bias was unknown. For the scoring of these items, we applied the same signaling questions as those proposed by Hooijmans et al. (2014). Briefly, Item 1 of the RoB tool, sequence generation, was scored with "yes" when authors clearly described a random component in the sequence generation, such as the use of a computer random number generator. A "no" score was provided when a nonrandom approach was applied, such as allocation by judgement or preference. Item 2 of the RoB tool, baseline similarities, was analyzed based on age and body weight. Baseline similarities were scored with "yes" when the age of directly exposed animals was reported, when directly exposed animals were randomly distributed across exposure and control groups according to body weight, or when both of these occurred. Item 3 focused on allocation concealment and was scored based on the applied method to conceal the allocation sequence. Methods such as sequentially numbered, opaque, sealed envelopes were considered adequate; however, allocation based on animal number, for example, was not considered adequate because the investigator might have been able to foresee the assignments. Item 4, random housing, was scored with "yes" when animals were randomly housed during the experiment. Blinding of caregivers, Item 5, focused on the blinding of caregivers to knowledge about which intervention each animal received. Appropriate blinding included identical housing for the exposure and control groups, whereas differences in housing were considered inappropriate blinding. Item 6, random outcome assessment, focused on the method applied to select animals for outcome assessment, and Item 7 focused on the blinding of the outcome assessor (for instance, whether similar assessment methods were applied to both the exposure and the control groups). Item 8, incomplete outcome data, focused on whether all animals were included in the analyses and whether reasons for dropouts were clearly explained. Item 9, selective outcome reporting, focused on whether results of all outcomes mentioned in the methods section were reported in the results section and vice versa. Items 6–9 were only analyzed for the following outcome measures: body weight, fat (pad) weights, triglyceride levels, FFA levels, and leptin levels. In addition, in Item 10, other potential sources of bias were scored, including risks of additional additives that were added during dosing and design-specific risks.

In addition to the RoB assessment, four items were added to check the methodological quality. Two of these items specifically focused on potential litter effects. In theory, potential litter effects could have been covered under Item 10 of the RoB tool; however, we decided to report these items separately because these are key issues for perinatal exposure studies. One of the items focused on whether the litter or the individual offspring were used as statistical unit, and the other item focused on whether effects on litter size were observed after exposure to BPA compared with control conditions. These items were translated into the following two questions: "Was intralitter correlation controlled for by using the litter as statistical unit (instead of

offspring)?" and "Was the study free of potential intralitter correlation caused by effects on litter size?" These items were scored with "yes," "no," or "unclear." In addition, we included two overall study quality indicators to acquire additional lower-tier information on the reporting quality of the studies. Because animal studies are known for their poor reporting quality in comparison with randomized clinical trials, it is likely that many items of the RoB tool are not reported or are poorly reported (Hooijmans et al. 2014). The two overall study quality indicators scored whether any randomization was reported for any level of the experiment and whether any blinding was reported for any level of the experiment. These items were scored "yes" when reported and "no" when not reported. One reviewer (P.W.) conducted the RoB and methodological quality assessment, and in case of doubt, a second reviewer (J.L.) was consulted. The results of the RoB tool and the additional methodological quality items were used in the sensitivity analyses and in the confidence rating as described below.

### **Data Synthesis and Statistical Analysis**

Meta-analyses were performed using Review Manager (RevMan) v5.3 (The Cochrane Community) if at least five studies reported on a specific outcome measure, which was the case for all outcomes studied. No criteria were set for the number of comparisons that had to be included. Mean differences (MDs; the mean of the experimental group minus the mean of the control group) were calculated for outcome measures that reported data on the same scale, or when all data could be converted to the same scale, including body weight and leptin. If data were not reported on the same scale and could not be converted, standardized mean differences (SMDs; the mean of the experimental group minus the mean of the control group divided by the pooled SD of the two groups) were calculated, including for fat weight, triglycerides, and FFA. In the meta-analyses, random effect models were used to account for the anticipated heterogeneity for all outcome measures. Positive SMDs and MDs indicated an increase in the outcome measures after BPA exposure, whereas negative SMD and MDs indicated a decrease in the outcome measure. Heterogeneity was assessed using  $I^2$  and was represented on a scale ranging from 0% to 100%, where  $\leq 50\%$  was considered as no serious heterogeneity between studies, 50–75% was considered as moderate heterogeneity, and  $>75\%$  was considered as substantial heterogeneity (NTP 2015). The significance level of the meta-analyses was set at  $p < 0.05$ . Subgroup analyses were performed to assess the influence of variables and to explore possible causes of heterogeneity. Subgroups were predefined in the protocol, and subgroup analyses were only performed when at least three studies could be included per subgroup. No criteria were set for the number of comparisons that had to be included. In some cases, subgroups that contained fewer than three studies, such as specific strains or routes of exposure, were combined in a subgroup called "Others" to generate a subgroup with three or more studies. The following subgroup variables were assessed for all outcome measures: animal species (rats or mice), strains, sex (male or female), time window of exposure (perinatal, prenatal, or postnatal), dosage of treatment (below or above the RfD of 50  $\mu\text{g}/\text{kg}/\text{d}$ ), route of exposure (gavage, oral, diet, drinking water, or subcutaneous injections), timing of outcome measurement (before or after PND 21, the normal weaning period), and frequency of exposure (daily or constant exposure, for example, via constant availability of BPA in drinking water or diet). Some studies provided birth weight data for mixed sexes, whereas at later time points, sex-specific body weight data were provided. We included only the sex-specific data because they are considered to be more informative (i.e., they provide information on



sex-specific effects as well as information on later time points). Furthermore, for the outcome measures triglyceride, FFA, and leptin, effects on different compartments were analyzed if at least three studies could be included (e.g., levels in serum/plasma and hepatic tissue). For the outcome measure fat weight, subgroup analyses were conducted to analyze the effects on different fat pads. All subgroup differences were analyzed with a test for subgroup differences in RevMan. To correct for multiple testing, the significance levels of the subgroup analyses were adjusted to  $p < 0.01$ . For fat weight, where multiple fat pads from individual animals were often analyzed, no additional correction was applied in the analysis to account for nonindependency. The results are presented as SMDs or MDs with 95% confidence intervals (CIs), heterogeneity value  $I^2$  and number of studies with number of comparisons in parentheses [i.e.,  $n = \text{number of studies (number of comparisons)}$ ].

Potential publication bias was assessed by visually inspecting funnel plots for asymmetry for outcome measures containing at least 10 studies. In addition, lag time for “negative” studies and conflict of interest sections with funding sources were investigated. A publication lag time for “negative” studies may be present because “positive” studies tend to be published earlier (NTP 2015). Therefore, we visually inspected the effect sizes on year of publication to identify whether a lag time for “negative” studies was likely to be present. Furthermore, publication bias could be present when studies are uniformly sponsored by industry or by nongovernmental organizations and/or when authors have a conflict of interest (NTP 2015). We characterized publication bias as “undetected” or as “strongly suspected” (i.e., when clear asymmetry was observed in funnel plots, when clear lag time was observed for negative studies, when the majority of studies had conflict of interest issues, or for combinations of any or all of these) in line with the assessment of the Office of Health Assessment and Translation (OHAT; NTP 2015).

To evaluate the robustness of our results, we performed a series of three sensitivity analyses. First, to assess the impact of the latest time point on interpreting study results and possible transient effects, we selected the latest measured time point instead of the time point with the greatest efficacy. Second, a sensitivity analysis was performed by excluding studies of potential high bias and poor reporting quality, that is to say, by excluding studies that did not receive a single “yes” score on Items 1–12 of the risk of bias and methodological quality indicators. A third sensitivity analysis was conducted by excluding studies that were not free of potential litter effects. The excluded studies were selected based on a “no” answer to either of the two included methodological quality items (Items 13–14).

### Confidence Rating

The quality of evidence of the outcomes of the systematic review was rated using the confidence rating methodology described by OHAT (NTP 2015). The OHAT confidence rating methodology is primarily based on the Grading of Recommendation, Assessment, Development, and Evaluation (GRADE) approach, in which four confidence ratings are used: high, moderate, low, and very low. The quality of evidence was rated for each outcome measure separately, in which initial confidence was rated based on the presence of four key study design features: *a*) controlled exposure; *b*) exposure before outcome development; *c*) outcome assessment on the individual level; and *d*) inclusion of a comparison group. Experimental animal studies usually have all of these features and therefore, the included studies received an initial rating of “high confidence.” Subsequently, five factors were assessed that could reduce the confidence rating (risk of bias, unexplained inconsistency, indirectness,

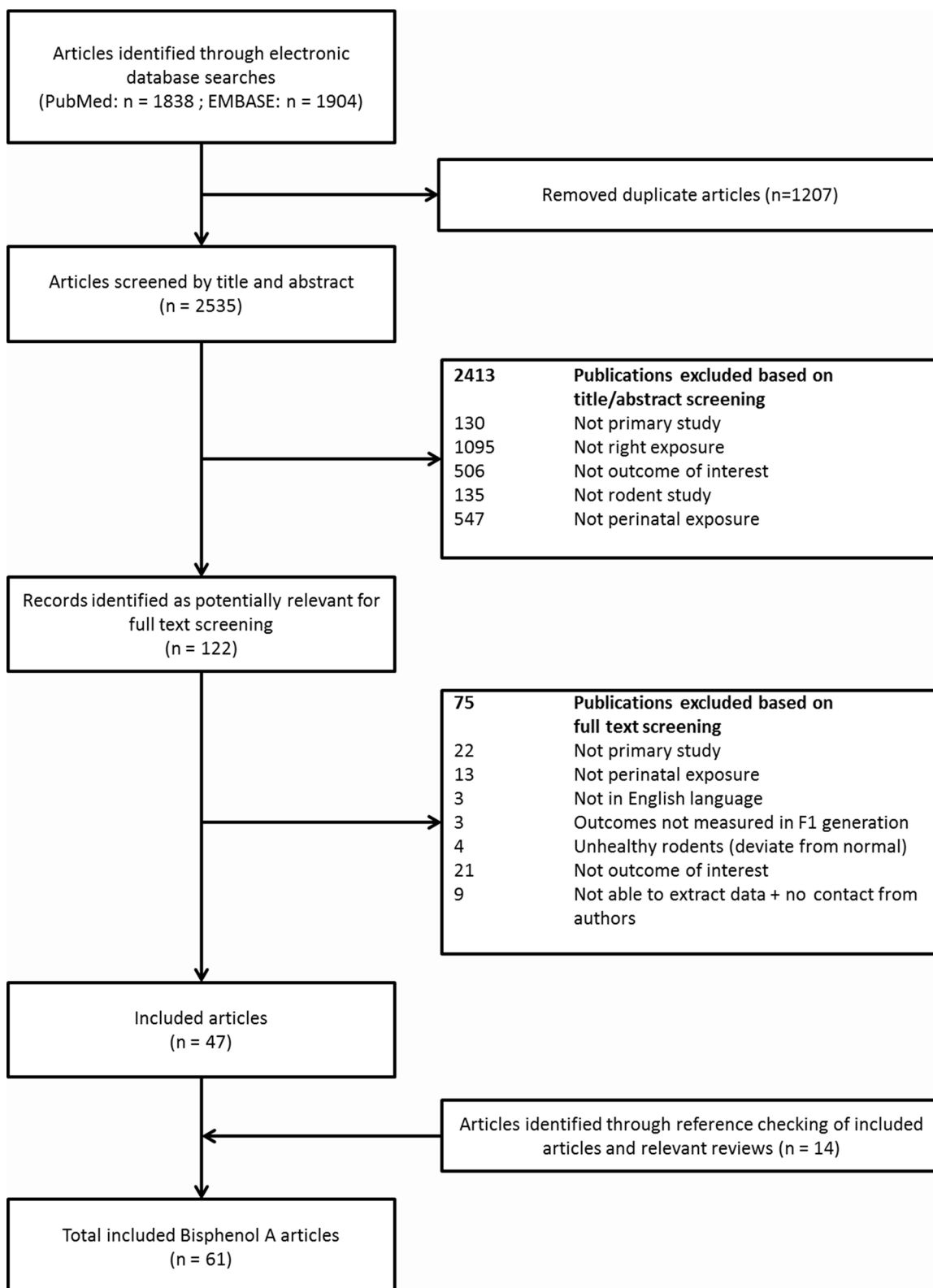
imprecision, and publication bias), followed by four factors that could increase the confidence rating (large magnitude of effect, dose response, plausible confounding, and consistency across study designs).

Briefly, for downgrading confidence based on risk of bias, the results of the SYRCLE RoB tool were considered as well as the results of the sensitivity analyses. We downgraded this item based on the RoB tool when predominantly “no” and “unclear” scores were provided for the included studies and/or when the direction of the overall effect changed after sensitivity analyses. Unexplained inconsistency was addressed by considering similarity of point estimates, overlap of confidence intervals between studies, and statistical heterogeneity (i.e.,  $I^2 > 50\%$  is moderate heterogeneity and  $>75\%$  is substantial heterogeneity). Indirectness was rated based on multiple aspects: the relevance of the animal models to the outcome of concern for humans (i.e., effects on body weight, fat weight, triglycerides, FFA, and leptin), the directness of the end point to the primary health outcome (i.e., obesity), and the relevance of the route of exposure. Duration of treatment and time window between exposure and outcome measurement were not rated under indirectness because defining a “too-short” time window is difficult. In line with OHAT recommendations, studies conducted with rats or mice were considered relevant for humans and were not downgraded on this specific aspect of indirectness (i.e., the first aspect). Imprecision was assessed based on the 95% confidence intervals (i.e., overlap with null or not). Publication bias was assessed similarly to the manner described above, as “undetected” or “strongly suspected” using funnel plots or when there were other indications of potential publication bias (i.e., lag time for “negative” studies or conflicts of interest). The factor “large effect magnitude” for upgrading confidence is difficult to assess because relatively “small” effects can have major public health impacts on a population basis. Furthermore, when using SMDs, the estimates are not directly related to actual physiological differences; therefore, it is difficult to determine whether the magnitude of effect estimates has any relevance to public health. Therefore, no threshold effect was set, and consequently, this factor was not used to upgrade confidence within this study. Dose–response effects were assessed by visually inspecting the effect sizes of individual study estimates. Indications of dose–response effects needed to be present consistently both within and across studies to be upgraded. The factor “plausible confounding” primarily applies to observational studies (NTP 2015). Because we considered animal studies in the present study, we decided not to use this factor to upgrade confidence. Within this study, confidence was upgraded for “consistency across study designs” when results were consistent across multiple animal models (i.e., rats and mice) and across multiple strains based on  $p$ -values for subgroup differences. Because only rodents were considered in this systematic review, we rated “consistency across species” only as a half confidence level and not as a full confidence level. Additionally, in line with the OHAT confidence rating methodology, we did not downgrade twice for what was essentially the same limitation, for example, whether wide confidence intervals resulted from unexplained inconsistency or from imprecision. In addition, in the case that two domains were borderline for downgrading, the body of evidence was downgraded once for a single factor to account for both partial concerns.

## Results

### Study Selection and Characteristics

Figure 1 shows the flow chart of the study selection process. Using the comprehensive search strategies, 2,535 unique articles



**Figure 1:** Flow chart of study selection process.

were identified from PubMed and EMBASE. After title and abstract screening, 122 articles were selected for full-text screening. Of these 122 publications, 47 publications were included in the review. In 18 of the 122 articles, data reporting was unclear, and the authors were contacted (see Excel Table S2). For 4 of these articles, data were provided; for 14 articles, no reply was

received. Of these 14 articles, 9 articles were fully excluded because body weight data were unclear and no information was presented on other outcome measures relevant to this review. For the remaining 5 studies, we were able to extract data of sufficient quality for outcomes other than body weight. In addition, after screening the reference lists of the included articles and of

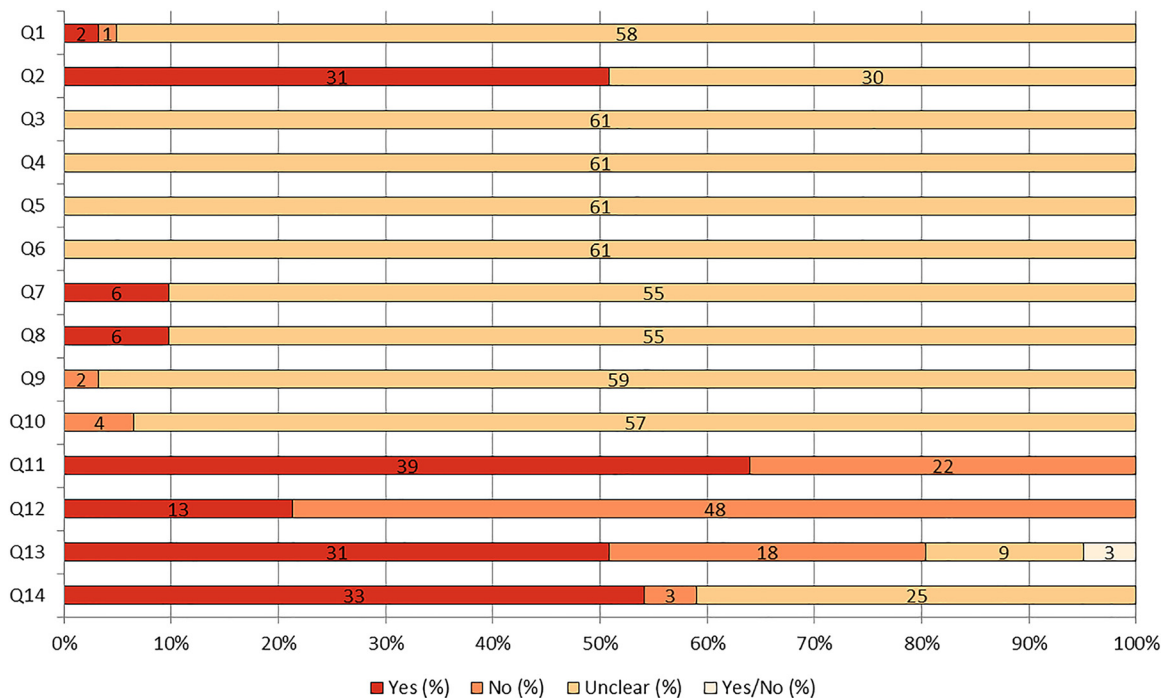
relevant reviews, an additional 14 articles were included, resulting in a total of 61 articles (see Excel Table S3 for the characteristics of all included studies).

Of the 61 included studies, 55 reported effects of BPA on body weight, 13 on fat weight, eight on triglyceride levels, seven on FFA levels, and eight on leptin levels. These outcome measures included 190, 117, 17, 18, and 34 independent comparisons, respectively, totaling 376 independent comparisons. FFA were mostly examined in blood (71% of the studies reported serum/plasma concentrations) but were also examined in hepatic (29%) or fat tissue (14%). The same is true for triglycerides (75% of the studies reported serum/plasma concentrations, and 38% reported hepatic tissue concentrations). Leptin levels were only reported in blood (100% in serum/plasma). The outcome fat weight was examined for several fat pads including the intraabdominal fat pads: retroperitoneal (23% of the studies), (peri)gonadal (62%), (peri)renal (38%), and mesenteric fat (31%). In addition, mammary gland fat (8%), subcutaneous fat (15%), brown adipose tissue (BAT; 31%), and total fat weight (31%) were reported. In all BPA exposure studies, outcomes were measured between PND0 and PND540, and the exposure dose ranged from 0.2 µg/kg/d to 655 mg/kg/d.

### Risk of Bias and Methodological Quality Assessments

The main observation from the risk of bias and methodological quality assessments is the many “unclear” scores, indicating that

most items were not sufficiently reported, resulting in an unknown risk of bias (Figure 2). The individual scores of the RoB tool and the methodological quality indicators of each included study are provided (see Excel Table S4). With respect to selection bias (Figure 2; Q1–Q3), the sequence generation process was reported in only three studies (5%; Q1). Although many studies mentioned that the animals were randomly assigned to exposure groups, the randomization method was unclear. As a result, the risk of bias on Item 1 could not be judged for many articles. Baseline similarities were reported more often (49%; Q2), whereas information about allocation concealment was not reported at all (Q3). None of the articles reported on random housing and blinding of caregivers (Figure 2; Q4 and Q5, respectively). As a result, performance bias could not be judged. Regarding detection bias (Figure 2; Q6 and Q7), none of the studies described a random outcome assessment for relevant outcome measures (Q6). In addition, the outcome assessor was reported to have been blinded in six studies (10%; Q7). Incomplete outcome data were adequately addressed in six studies (10%; Q8), resulting in a low risk of attrition bias for these studies. With respect to reporting bias (Q9), a high risk was identified for two studies (3%). All other studies were scored with an unclear risk of bias on this item. Additionally, other potential sources of bias were identified in four articles (7%; Q10). For two studies this included additional additives during dosing. Rats received 15% sucrose in their drinking water in addition to BPA exposure (Xu et al. 2011) or ten subcutaneous injections of corn oil (Ichihara et al. 2003). In two



**Figure 2:** Results of the risk of bias and methodological quality indicators for all included studies. The items in the Systematic Review Centre for Laboratory Animal Experimentation (SYRCLE) Risk of Bias assessment (Q1–Q10) were scored with “yes” indicating low risk of bias, “no” indicating high risk of bias, or “unclear” indicating that the item was not reported, resulting in an unknown risk of bias (Hooijmans et al. 2014). Q1–Q3 consider selection bias, Q4–Q5 consider performance bias, Q6–Q7 consider detection bias, Q8 considers attrition bias, Q9 considers reporting bias, and Q10 considers other biases. The overall study quality indicators (Q11–Q12) were scored with “yes” when reported or “no” when not reported. The methodological quality indicators focusing on potential intralitter correlation (Q13–Q14) were scored with “yes,” “no,” or “unclear.” Q, question. Q1: Was the allocation sequence adequately generated and applied?; Q2: Were the groups similar at baseline or were they adjusted for confounders in the analysis?; Q3: Was the allocation to the different groups adequately concealed?; Q4: Were the animals randomly housed during the experiment?; Q5: Were the caregivers and/or investigators blinded from knowledge which intervention each animal received during the experiment?; Q6: Were animals selected at random for outcome assessment?; Q7: Was the outcome assessor blinded?; Q8: Were incomplete outcome data adequately addressed?; Q9: Are reports of the study free of selective outcome reporting?; Q10: Was the study apparently free of other problems that could result in high risk of bias?; Q11: Was it stated that the experiment was randomized at any level?; Q12: Was it stated that the experiment was blinded at any level?; Q13: Was intralitter correlation controlled for by using the litter as statistical unit (instead of offspring)?; Q14: Was the study free of potential intralitter correlation caused by effects on litter size?

other studies, animals were delivered by cesarean section (Howdeshell and vom Saal 2000; Nagao et al. 2002).

In addition to risk of bias, four study quality indicators were used to assess the methodological quality of the studies. In 64% of the studies, randomization at any level of the experiment was reported (Figure 2; Q11), whereas blinding was reported in only 21% of the studies (Q12). Assessment of litter effects revealed that the litter was used as a statistical unit in 51% of the studies, the offspring was used as a statistical unit in 30% of the studies, and in 15% of the studies, it was unclear whether litter or offspring was used as statistical unit (Q13). In addition, three studies received both a “yes” and a “no” score (5%). Two of these studies used the litter as a statistical unit for measurements before PND21 and the offspring as a statistical unit for measurements after PND21. The third study used the litter as a statistical unit for one outcome measure, but the offspring was used as a statistical unit for other relevant outcome measures (see Excel Table S4). Effects were observed on litter size after exposure to BPA in only three studies (5%; Q14).

### Effects of BPA on Obesity-Related Outcomes

**Body weight.** Out of 55 studies, a total of 190 comparisons investigating the effects of BPA on body weight could be included in

the meta-analysis. These studies reported body weight in grams; therefore, MDs were calculated for the effects of BPA on body weight. Early-life exposure to BPA was associated with significantly lower body weight based on the overall summary estimate [MD = -0.22 (95% CI: -0.37, -0.06); Table 2]. There was substantial heterogeneity among the studies ( $I^2 = 86\%$ ; Table 2). A forest plot shows the individual effect estimates for the 190 comparisons of BPA exposure with body weight (see Excel Figure S1). Subgroup analysis showed that heterogeneity was very high for all estimates (i.e.,  $I^2 > 75\%$ ) with the exception of the estimate for some strains. Associations did not vary between mice and rats ( $p$ -value for subgroup differences = 0.69), although they varied across strains (Table 2). Nonsignificant positive associations were estimated for three mouse strains, a significant positive association was estimated for Wistar rats (with substantial heterogeneity,  $I^2 = 92\%$ ), and null or negative associations were estimated for all other strains, including a significant negative association for F344 rats ( $I^2 = 35\%$ ;  $p$ -value for subgroup differences < 0.0001). When based on females only, a significant negative association was estimated [MD = -0.41 (95% CI: -0.65, -0.17)] in contrast with a null summary estimate for males [MD = -0.05 (95% CI: -0.27, 0.17);  $p$ -value for subgroup differences = 0.03]. Associations varied across time window of exposure, with null estimates for perinatal and prenatal exposure,

**Table 2.** Effects of early-life exposure to bisphenol A on body weight from random effects meta-analyses.

| Analysis                      | Subgroups             | MD (g) (95% CI)      | $I^2$ | No. studies (no. comparisons) | Test for subgroup differences |
|-------------------------------|-----------------------|----------------------|-------|-------------------------------|-------------------------------|
| Overall                       |                       | -0.22 (-0.37, -0.06) | 86%   | 55 (190)                      |                               |
| Species                       | Rats                  | -0.06 (-0.39, 0.27)  | 86%   | 35 (111)                      | $p = 0.69$                    |
|                               | Mice                  | -0.14 (-0.33, 0.05)  | 85%   | 20 (79)                       |                               |
| Strain                        | C57BL/6 (mice)        | 0.11 (-0.56, 0.78)   | 83%   | 5 (27)                        | $p < 0.0001$                  |
|                               | CD-1 (mice)           | 0.19 (-0.20, 0.58)   | 67%   | 4 (15)                        |                               |
|                               | CF-1 (mice)           | 0.51 (-0.14, 1.17)   | 59%   | 4 (11)                        |                               |
|                               | ICR (mice)            | -0.02 (-0.21, 0.16)  | 85%   | 3 (12)                        |                               |
|                               | F344 (rats)           | -0.83 (-1.16, -0.50) | 35%   | 3 (11)                        |                               |
|                               | Long evans (rats)     | -0.32 (-0.65, 0.02)  | 50%   | 6 (12)                        |                               |
|                               | Sprague-Dawley (rats) | -0.21 (-0.74, 0.33)  | 87%   | 16 (62)                       |                               |
|                               | Wistar (rats)         | 8.16 (2.23, 14.09)   | 92%   | 8 (14)                        |                               |
|                               | Others <sup>d</sup>   | -0.53 (-1.37, 0.31)  | 84%   | 7 (26)                        |                               |
| Sex                           | Males                 | -0.05 (-0.27, 0.17)  | 86%   | 46 (111)                      | $p = 0.03$                    |
|                               | Females               | -0.41 (-0.65, -0.17) | 86%   | 34 (79)                       |                               |
| Time window of exposure       | Perinatal             | 0.09 (-0.20, 0.38)   | 84%   | 36 (127)                      | $p = 0.04$                    |
|                               | Prenatal              | 0.02 (-0.13, 0.18)   | 82%   | 11 (40)                       |                               |
|                               | Postnatal             | -3.21 (-5.70, -0.71) | 89%   | 9 (23)                        |                               |
| Frequency of exposure         | Daily                 | -0.34 (-0.50, -0.18) | 85%   | 40 (128)                      | $p < 0.0001$                  |
|                               | Constant              | 0.99 (0.37, 1.61)    | 88%   | 15 (62)                       |                               |
| Route of exposure             | Gavage                | -0.57 (-0.89, -0.25) | 76%   | 15 (48)                       | $p = 0.0004$                  |
|                               | Oral                  | -0.22 (-0.56, 0.11)  | 88%   | 11 (40)                       |                               |
|                               | Diet                  | 0.08 (-0.40, 0.56)   | 79%   | 8 (49)                        |                               |
|                               | Drinking water        | 4.67 (2.17, 7.17)    | 91%   | 9 (22)                        |                               |
|                               | Subcutaneous          | -0.37 (-0.71, -0.03) | 89%   | 12 (31)                       |                               |
| Dose of exposure <sup>b</sup> | ≤50 µg/kg             | 0.16 (-0.03, 0.36)   | 87%   | 35 (72)                       | $p = 0.003$                   |
|                               | >50 µg/kg             | -0.40 (-0.72, -0.09) | 86%   | 35 (112)                      |                               |
| Time of outcome measure       | PND0–PND21            | -0.29 (-0.40, -0.17) | 81%   | 18 (59)                       | $p = 0.01$                    |
|                               | >PND21                | 0.49 (-0.12, 1.11)   | 87%   | 47 (131)                      |                               |

Note: Effect sizes are expressed as the MD in grams with 95% CIs calculated using random effects models. From each study, we considered each analysis with a specific dose, sex, and/or time window of exposure as a separate individual comparison.  $I^2$  is a measure of heterogeneity. Positive MDs represent an increase in the outcome measure after exposure. Negative MDs represent a decrease in the outcome measure after exposure. Tests for subgroup differences were conducted using Review Manager [RevMan v5.3 (The Cochrane Community)]. CI, confidence interval; MD, mean difference; PND, postnatal day.

<sup>a</sup>Includes OF-1, NMRI, and Avy Agouti mice and Donryu, Alderley Park, and Holtzman rats.

<sup>b</sup>The results from Anderson et al. (2013) were excluded from this analysis because exposure was expressed as diet concentration, and it was not possible to estimate dose concentrations.



**Table 3.** Effects of early-life exposure to bisphenol A on fat weight from random effects meta-analyses.

| Analysis                      | Subgroups             | SMD (95% CI)       | <i>I</i> <sup>2</sup> | No. studies (no. comparisons) | Test for subgroup differences |
|-------------------------------|-----------------------|--------------------|-----------------------|-------------------------------|-------------------------------|
| Overall                       |                       | 0.67 (0.53, 0.81)  | 46%                   | 13 (117)                      |                               |
| Species                       | Rats                  | 1.12 (0.71, 1.52)  | 66%                   | 4 (19)                        | <i>p</i> = 0.01               |
|                               | Mice                  | 0.57 (0.44, 0.71)  | 36%                   | 9 (98)                        |                               |
| Strain                        | C57BL/6 (mice)        | 0.71 (0.56, 0.85)  | 0%                    | 3 (52)                        | <i>p</i> = 0.0001             |
|                               | CD-1 (mice)           | 0.18 (−0.05, 0.41) | 15%                   | 3 (28)                        |                               |
|                               | Sprague-Dawley (rats) | 1.06 (0.63, 1.48)  | 68%                   | 3 (17)                        |                               |
|                               | Others <sup>b</sup>   | 0.99 (0.46, 1.51)  | 76%                   | 4 (20)                        |                               |
| Sex                           | Males                 | 0.69 (0.56, 0.81)  | 19%                   | 13 (86)                       | <i>p</i> = 0.44               |
|                               | Females               | 0.52 (0.11, 0.93)  | 72%                   | 9 (31)                        |                               |
| Time window of exposure       | Perinatal             | 0.65 (0.51, 0.79)  | 37%                   | 10 (95)                       | <i>p</i> = 0.44               |
|                               | Prenatal              | 0.81 (0.43, 1.19)  | 67%                   | 3 (22)                        |                               |
| Frequency of exposure         | Daily                 | 0.91 (0.61, 1.21)  | 68%                   | 5 (36)                        | <i>p</i> = 0.05               |
|                               | Constant              | 0.58 (0.44, 0.72)  | 23%                   | 8 (81)                        |                               |
| Route of exposure             | Diet                  | 0.57 (0.43, 0.71)  | 13%                   | 5 (70)                        | <i>p</i> = 0.13               |
|                               | Drinking water        | 0.68 (0.21, 1.15)  | 58%                   | 3 (11)                        |                               |
|                               | Others <sup>c</sup>   | 0.91 (0.61, 1.21)  | 68%                   | 5 (36)                        |                               |
| Dose of exposure <sup>d</sup> | ≤50 µg/kg             | 0.76 (0.55, 0.97)  | 38%                   | 9 (53)                        | <i>p</i> = 0.66               |
|                               | >50 µg/kg             | 0.70 (0.52, 0.88)  | 43%                   | 7 (58)                        |                               |
| Fat pads                      | Retroperitoneal fat   | 0.68 (−0.09, 1.45) | 45%                   | 3 (7)                         | <i>p</i> = 0.67               |
|                               | (Peri)gonadal fat     | 0.72 (0.36, 1.07)  | 65%                   | 8 (27)                        |                               |
|                               | (Peri)renal fat       | 0.63 (0.36, 0.89)  | 34%                   | 5 (20)                        |                               |
|                               | Mesenteric fat        | 0.82 (0.55, 1.10)  | 0%                    | 4 (14)                        |                               |
|                               | Brown adipose tissue  | 0.94 (0.47, 1.41)  | 52%                   | 4 (15)                        |                               |
|                               | Total fat             | 0.35 (−0.14, 0.85) | 70%                   | 4 (14)                        |                               |
|                               | Others <sup>d</sup>   | 0.67 (0.42, 0.92)  | 0%                    | 3 (20)                        |                               |

Note: Effect sizes are expressed as the SMD with 95% CIs calculated using random effects models. From each study, we considered each analysis with a specific dose, sex, and/or time window of exposure as a separate individual comparison, as well as analyses of different fat pads (even when derived from the same animal). *I*<sup>2</sup> is a measure of heterogeneity. Positive SMDs represent an increase in the outcome measure after exposure. Negative SMDs represent a decrease in the outcome measure after exposure. Tests for subgroup differences were conducted using Review Manager [RevMan v5.3 (The Cochrane Community)]. CI, confidence interval; SMD, standardized mean difference.

<sup>a</sup>Results of Anderson et al. (2013) were excluded from this analysis because exposure was expressed as diet concentration, and it was not possible to estimate dose concentrations.

<sup>b</sup>Includes OF-1, MF-1, and Avy Agouti mice and Wistar rats.

<sup>c</sup>Includes subcutaneous, oral, and gavage exposure.

<sup>d</sup>Includes mammary and subcutaneous fat pads.

and a significant negative association was estimated for postnatal exposure (*p*-value for subgroup differences = 0.04). Associations with body weight were in opposite directions when based on studies with daily exposure [e.g., via gavage or subcutaneous injection; MD = −0.34 (95% CI: −0.50, −0.18)] versus constant exposure via diet or drinking water [MD = 0.99 (95% CI: 0.37, 1.61); *p*-value for subgroup differences = 0.0001]. In addition, significant positive associations were estimated for exposure via drinking water, and null or negative associations were estimated for all other exposure routes, including significant negative associations for subcutaneous exposure and exposure via gavage (*p*-value for subgroup differences = 0.0004). Exposure doses <50 µg/kg/d were positively associated with body weight [MD = 0.16 (95% CI: −0.03, 0.36)], whereas exposure doses >50 µg/kg/d were significantly associated with a lower body weight [MD = −0.40 (95% CI: −0.72, −0.09); *p*-value for subgroup differences = 0.003]. Furthermore, measurements before PND21 were associated with significantly lower body weight [MD = −0.29 (95% CI: −0.40, −0.17)], whereas measurements after PND21 were estimated to have a nonsignificant positive association with body weight [MD = 0.49 (95% CI: −0.12, 1.11); *p*-value for subgroup differences = 0.01].

**Fat weight.** For fat weight, 13 studies were included, consisting of 117 comparisons. Only SMDs could be calculated for fat weight because data were reported on deviating scales, including

weight (grams) and percentages relative to control. Overall, exposure to BPA was significantly associated with an increased fat weight in rodents [SMD = 0.67 (95% CI: 0.53, 0.81); Table 3] with a relatively low heterogeneity that could be related to the fact that several fat pads derived from the same animal were included (*I*<sup>2</sup> = 46%). A forest plot is provided showing the individual effect estimates for BPA exposure and fat weight (see Excel Figure S2). Subgroup analyses revealed that associations varied between species with stronger associations in rats, although both estimates indicated a significant positive association (*p*-value for subgroup differences = 0.01; Table 3). Data for rats [SMD = 1.12 (95% CI: 0.71, 1.52)] were based on moderately heterogeneous data (*I*<sup>2</sup> = 66%), and data for mice [SMD = 0.57 (95% CI: 0.44, 0.71)] were based on data with lower heterogeneity (*I*<sup>2</sup> = 36%). This difference in association of species is also reflected in the different associations of strains. A nonsignificant positive association was estimated for CD-1 mice and significant positive associations were estimated for C57BL/6 mice, Sprague-Dawley rats, and others (*p*-value for subgroup differences = 0.0001). No difference in association was estimated for sex (i.e., males and females; *p*-value for subgroup differences = 0.44) or for time window of exposure (i.e., perinatal or prenatal; *p*-value for subgroup differences = 0.44). Estimates for both daily exposure [SMD = 0.91 (95% CI: 0.61, 1.21)] and constant exposure [SMD = 0.58 (95% CI: 0.44, 0.72)] were significantly



**Table 4.** Effects of early-life exposure to bisphenol A on triglyceride levels from random effects meta-analyses.

| Analysis         | Subgroups           | SMD (95% CI)       | <i>I</i> <sup>2</sup> | No. studies (no. comparisons) | Test for subgroup differences |
|------------------|---------------------|--------------------|-----------------------|-------------------------------|-------------------------------|
| Overall          |                     | 0.97 (0.53, 1.40)  | 58%                   | 8 (18)                        |                               |
| Species          | Rats                | 1.34 (0.44, 2.24)  | 74%                   | 5 (8)                         | <i>p</i> = 0.23               |
|                  | Mice                | 0.73 (0.32, 1.15)  | 27%                   | 3 (10)                        |                               |
| Strain           | Wistar (rats)       | 2.00 (0.98, 3.02)  | 60%                   | 3 (5)                         | <i>p</i> = 0.01               |
|                  | Others <sup>a</sup> | 0.63 (0.24, 1.02)  | 37%                   | 5 (13)                        |                               |
| Sex              | Males               | 1.16 (0.69, 1.63)  | 55%                   | 7 (15)                        | <i>p</i> = 0.004              |
|                  | Females             | 0.05 (−0.54, 0.65) | 0%                    | 3 (3)                         |                               |
| Dose of exposure | ≤50 µg/kg           | 1.45 (0.78, 2.13)  | 63%                   | 6 (10)                        | <i>p</i> = 0.01               |
|                  | >50 µg/kg           | 0.45 (0.04, 0.87)  | 18%                   | 4 (8)                         |                               |
| Compartment      | Serum/plasma        | 0.76 (0.39, 1.13)  | 34%                   | 6 (14)                        | <i>p</i> = 0.22               |
|                  | Hepatic             | 1.94 (0.10, 3.79)  | 83%                   | 3 (4)                         |                               |

Note: Effect sizes are expressed as the SMD with 95% CIs calculated using random effects models. From each study, we considered each analysis with a specific dose, sex, and/or time window of exposure as a separate individual comparison. *I*<sup>2</sup> is a measure of heterogeneity. Positive SMDs represent an increase in the outcome measure after exposure. Negative SMDs represent a decrease in the outcome measure after exposure. Tests for subgroup differences were conducted using Review Manager [RevMan v5.3 (The Cochrane Community)]. CI, confidence interval; SMD, standardized mean difference.

<sup>a</sup>Includes OF-1 and C57BL/6JxFVB mice and Sprague-Dawley rats.

associated with fat weight (*p*-value for subgroup differences = 0.05) and were based on moderately and lowly heterogeneous data, respectively (*I*<sup>2</sup> = 68% and 23%). Estimated associations for route of exposure, including exposure via diet and drinking water, did not vary (*p*-value for subgroup differences = 0.13). In addition, no difference in association was estimated for different doses (i.e., doses below or above the RfD of 50 µg/kg/d; *p*-value for subgroup differences = 0.66) or for different fat pads (*p*-value for subgroup differences = 0.67). No subgroup analyses of timing of outcome measurement were conducted because fewer than three studies could be included in the subgroup with measurements before PND21.

**Triglycerides.** In total, 8 studies consisting of 18 comparisons could be included in the meta-analysis. Only SMDs could be

calculated for triglycerides because data were reported on deviating scales, including tissue concentrations (milligrams/gram), volume concentrations (milligrams/milliliter), and percentages relative to control. Overall, BPA exposure was positively associated with triglyceride levels [SMD = 0.97 (95% CI: 0.53, 1.40); Table 4] with moderate heterogeneity among the studies (*I*<sup>2</sup> = 58%). A forest plot shows the individual effect estimates of BPA exposure with triglyceride levels (see Excel Figure S3). Associations did not vary between mice and rats (*p*-value for subgroup differences = 0.23), although associations varied across strains. A significant positive association was estimated for Wistar rats [SMD = 2.00 (95% CI: 0.98, 3.02)] with moderate heterogeneity (*I*<sup>2</sup> = 60%), and a weaker association was estimated for the other strains, which have been combined in one subgroup

**Table 5.** Effects of early-life exposure to bisphenol A on free fatty acid levels from random effects meta-analyses.

| Analysis                | Subgroups           | SMD (95% CI)       | <i>I</i> <sup>2</sup> | No. studies (no. comparisons) | Test for subgroup differences |
|-------------------------|---------------------|--------------------|-----------------------|-------------------------------|-------------------------------|
| Overall                 |                     | 0.86 (0.50, 1.22)  | 36%                   | 7 (17)                        |                               |
| Species                 | Rats                | 0.76 (0.08, 1.43)  | 53%                   | 5 (7)                         | <i>p</i> = 0.65               |
|                         | Mice                | 0.94 (0.52, 1.36)  | 22%                   | 3 (10)                        |                               |
| Strain                  | Wistar (rats)       | 0.99 (−0.19, 2.17) | 68%                   | 3 (4)                         | <i>p</i> = 0.81               |
|                         | Others <sup>a</sup> | 0.84 (0.48, 1.20)  | 23%                   | 5 (13)                        |                               |
| Sex                     | Males               | 1.24 (0.88, 1.59)  | 0%                    | 5 (11)                        | <i>p</i> = 0.0003             |
|                         | Females             | 0.14 (−0.33, 0.61) | 0%                    | 3 (6)                         |                               |
| Frequency of exposure   | Daily               | 0.97 (0.15, 1.78)  | 61%                   | 5 (7)                         | <i>p</i> = 0.78               |
|                         | Constant            | 0.84 (0.48, 1.19)  | 5%                    | 3 (10)                        |                               |
| Dose of exposure        | ≤50 µg/kg           | 1.57 (1.05, 2.09)  | 0%                    | 4 (6)                         | <i>p</i> = 0.001              |
|                         | >50 µg/kg           | 0.53 (0.19, 0.87)  | 0%                    | 4 (11)                        |                               |
| Time of outcome measure | PND0–PND21          | 0.38 (−0.30, 1.05) | 27%                   | 3 (4)                         | <i>p</i> = 0.11               |
|                         | >PND21              | 1.02 (0.61, 1.42)  | 32%                   | 5 (13)                        |                               |
| Compartment             | Serum/plasma        | 0.97 (0.65, 1.30)  | 0%                    | 5 (12)                        | <i>p</i> = 0.51               |
|                         | Others <sup>b</sup> | 0.63 (−0.35, 1.61) | 66%                   | 3 (5)                         |                               |

Note: Effect sizes are expressed as the SMD with 95% CIs calculated using random effects models. From each study, we considered each analysis with a specific dose, sex, and/or time window of exposure as a separate individual comparison. *I*<sup>2</sup> is a measure of heterogeneity. Positive SMDs represent an increase in the outcome measure after exposure. Negative SMDs represent a decrease in the outcome measure after exposure. Tests for subgroup differences were conducted using Review Manager [RevMan v5.3 (The Cochrane Community)]. CI, confidence interval; PND, postnatal day; SMD, standardized mean difference.

<sup>a</sup>Includes OF-1, Avi Agouti, and C57BL/6JxFVB mice and Sprague-Dawley rats.

<sup>b</sup>Includes hepatic and fat tissue.

**Table 6.** Effects of early-life exposure to bisphenol A on leptin levels from random effects meta-analyses.

| Analysis                       | Subgroups | MD (ng/ml) SMD (95% CI) | $I^2$ | No. studies (no. comparisons) | Test for subgroup differences |
|--------------------------------|-----------|-------------------------|-------|-------------------------------|-------------------------------|
| Overall                        |           | 0.37 (−0.14, 0.87)      | 50%   | 8 (34)                        |                               |
| Species                        | Rats      | −0.30 (−1.27, 0.67)     | 64%   | 3 (10)                        | $p = 0.08$                    |
|                                | Mice      | 0.72 (0.09, 1.35)       | 43%   | 5 (24)                        |                               |
| Sex                            | Males     | 0.69 (0.14, 1.25)       | 46%   | 8 (24)                        | $p = 0.03$                    |
|                                | Females   | −1.03 (−2.49, 0.43)     | 60%   | 5 (10)                        |                               |
| Time window of exposure        | Perinatal | 0.46 (−0.07, 1.00)      | 50%   | 5 (23)                        | $p = 0.8$                     |
|                                | Prenatal  | 0.26 (−1.24, 1.76)      | 51%   | 3 (11)                        |                               |
| Frequency of exposure          | Daily     | −0.01 (−0.85, 0.83)     | 59%   | 5 (17)                        | $p = 0.16$                    |
|                                | Constant  | 0.75 (0.10, 1.41)       | 38%   | 3 (17)                        |                               |
| Route of exposure <sup>a</sup> | Gavage    | −0.30 (−1.27, 0.68)     | 63%   | 3 (10)                        | $p = 0.08$                    |
|                                | Diet      | 0.75 (0.10, 1.41)       | 38%   | 3 (17)                        |                               |
| Dose of exposure <sup>b</sup>  | ≤50 μg/kg | 0.18 (−0.53, 0.89)      | 51%   | 7 (20)                        | $p = 0.26$                    |
|                                | >50 μg/kg | 0.82 (−0.04, 1.67)      | 67%   | 3 (8)                         |                               |

Note: Effect sizes are expressed as the MD in ng/ml with 95% CIs calculated using random effects models. From each study, we considered each analysis with a specific dose, sex, and/or time window of exposure as a separate individual comparison.  $I^2$  is a measure of heterogeneity. Positive MDs represent an increase in the outcome measure after exposure. Negative MDs represent a decrease in the outcome measure after exposure. Tests for subgroup differences were conducted using Review Manager RevMan v5.3 (The Cochrane Community). CI, confidence interval; MD, mean difference.

<sup>a</sup>Measurements of Angle et al. (2013); oral exposure, five comparisons) and Alonso-Magdalena et al. (2010; subcutaneous exposure, two comparisons) were excluded from this subgroup analysis because there were not three studies that could be included in an “Others” subgroup.

<sup>b</sup>Results of Anderson et al. (2013) were excluded from this analysis because exposure was expressed as diet concentration, and it was not possible to estimate dose concentrations.

[SMD = 0.63 (95% CI: 0.24, 1.02);  $I^2 = 37%$ ;  $p$ -value for subgroup differences = 0.01]. Furthermore, BPA exposure was associated with elevated triglyceride levels in males [SMD = 1.16 (95% CI: 0.69, 1.63);  $I^2 = 55%$ ], but not in females [SMD 0.05 (95% CI: −0.54, 0.65);  $I^2 = 0%$ ;  $p$ -value for subgroup differences = 0.004]. Associations between BPA and triglyceride levels varied according to exposure concentrations, such that concentrations <50 μg/kg/d were associated with higher triglyceride levels [SMD = 1.45 (95% CI: 0.78, 2.13);  $I^2 = 63%$ ] than concentrations >50 μg/kg/d [SMD = 0.45 (95% CI: 0.04, 0.87);  $I^2 = 18%$ ;  $p$ -value for subgroup differences = 0.01]. Furthermore, associations did not vary for circulating triglyceride levels (in serum/plasma) and triglyceride levels in hepatic tissues ( $p$ -value for subgroup differences = 0.22). No subgroup analyses of time window of exposure, frequency of exposure, route of exposure, or timing of outcome measurement were conducted because fewer than three studies could be included in the different subgroups.

**Free fatty acids.** Out of 7 studies, a total of 17 comparisons investigating the effects of BPA on FFA were included in the meta-analysis. Because FFA were reported on deviating scales, including tissue concentrations (milligrams/gram), volume concentrations (milligrams/milliliter), and percentages relative to control, only SMDs could be calculated. The summary estimate for BPA exposure on FFA levels over all studies indicated a significant positive association [SMD = 0.86 (95% CI: 0.50, 1.22);  $I^2 = 36%$ ; Table 5]. A forest plot shows the individual effect estimates of BPA exposure with FFA levels (see Excel Figure S4). Associations were similar when based on studies of mice versus rats ( $p$ -value for subgroup differences = 0.65), and of Wistar rats versus other strains ( $p$ -value for subgroup differences = 0.81) and on studies of daily versus constant exposure ( $p$ -value for subgroup differences = 0.78), but they were stronger for studies of males [SMD = 1.24 (95% CI: 0.88, 1.59);  $I^2 = 0%$ ] than females [SMD = 0.14 (95% CI: −0.33, 0.61);  $I^2 = 0%$ ;  $p$ -value for subgroup differences = 0.0003]. In addition, the association varied by dose, such that BPA doses <50 μg/kg/d were associated with higher FFA levels [SMD = 1.57 (95% CI: 1.05,

2.09);  $I^2 = 0%$ ] than doses >50 μg/kg/d [SMD = 0.53 (95% CI: 0.19, 0.87);  $I^2 = 0%$ ;  $p$ -value for subgroup differences = 0.001]. Furthermore, associations were comparable for measurements before PND21 and measurements after PND21 ( $p$ -value for subgroup differences = 0.11) and for serum/plasma and other compartments (including hepatic and fat tissue;  $p$ -value for subgroup differences = 0.51). No subgroup analyses on time window of exposure or route of exposure were conducted because fewer than three studies could be included in the different subgroups.

**Leptin.** For leptin, 34 comparisons from 8 studies could be included in the meta-analysis. Because these studies consistently reported leptin in nanograms/milliliter, MDs could be calculated. The overall estimate indicated a nonsignificant positive association of BPA exposure with leptin levels [MD = 0.37 (95% CI: −0.14, 0.87);  $I^2 = 50%$ ; Table 6]. A forest plot shows the individual effect estimates of BPA exposure with leptin levels (see Excel Figure S5). When based on mice only, there was a significant positive association [MD = 0.72 (95% CI: 0.09, 1.35);  $I^2 = 43%$ ], in contrast with a nonsignificant negative association for rats only [MD = −0.30 (95% CI: −1.27, 0.67);  $I^2 = 64%$ ;  $p$ -value for subgroup differences = 0.08]. A significant positive association was also estimated for males [MD = 0.69 (95% CI: 0.14, 1.25);  $I^2 = 46%$ ], whereas a nonsignificant negative association was estimated for females [MD = −1.03 (95% CI: −2.49, 0.43);  $I^2 = 60%$ ]. Associations were similar when based on studies of perinatal versus prenatal exposure ( $p$ -value for subgroup differences = 0.8) and for exposure doses below and above 50 μg/kg/d ( $p$ -value for subgroup differences = 0.26). Estimates for constant exposure indicated a significant positive association [MD = 0.75 (95% CI: 0.10, 1.41);  $I^2 = 38%$ ], in contrast with a null summary for daily exposure [MD = −0.01 (95% CI: −0.85, 0.83);  $I^2 = 59%$ ;  $p$ -value for subgroup differences = 0.16]. A significant positive association was also estimated for exposure via diet [MD = 0.75 (95% CI: 0.10, 1.41);  $I^2 = 38%$ ], whereas a nonsignificant negative association was estimated for exposure via gavage [MD = −0.30 (95% CI: −1.27, 0.68);  $I^2 = 63%$ ;  $p$ -value for subgroup differences = 0.08]. No subgroup analyses

**Table 7.** Results of the sensitivity analyses on the overall effects of early-life exposure to bisphenol A on body weight, fat weight, triglycerides, free fatty acids, and leptin.

| Outcome measure  | Effect size | Overall effect       | 1                  |                    | 2                    |                    | 3                    |                    |
|------------------|-------------|----------------------|--------------------|--------------------|----------------------|--------------------|----------------------|--------------------|
|                  |             |                      | Effect             | No. studies/ $I^2$ | Effect               | No. studies/ $I^2$ | Effect               | No. studies/ $I^2$ |
| Body weight      | MD          | -0.22 (-0.37, -0.06) | 0.17 (-0.22, 0.55) | 55/78%             | -0.32 (-0.57, -0.07) | 42/84%             | -0.26 (-0.45, -0.07) | 33/83%             |
| Fat weight       | SMD         | 0.67 (0.53, 0.81)    | 0.66 (0.52, 0.80)  | 13/46%             | 0.63 (0.48, 0.78)    | 9/38%              | 0.62 (0.32, 0.91)    | 8/60%              |
| Triglycerides    | SMD         | 0.97 (0.53, 1.40)    | —                  | —                  | 0.89 (0.41, 1.37)    | 4/56%              | 1.58 (0.65, 2.52)    | 4/69%              |
| Free fatty acids | SMD         | 0.86 (0.50, 1.22)    | —                  | —                  | 0.76 (0.40, 1.11)    | 4/24%              | 0.59 (-0.10, 1.27)   | 4/56%              |
| Leptin           | MD          | 0.37 (-0.14, 0.87)   | —                  | —                  | 0.39 (-0.18, 0.96)   | 7/52%              | -0.17 (-1.17, 0.83)  | 6/55%              |

Note: Effect sizes are expressed as the SMD or as the MD with 95% CIs. Positive SMDs and MDs represent an increase in the outcome measure after exposure. Negative SMDs and MDs represent a decrease in the outcome measure after exposure. Results of the three sensitivity analyses are provided: 1 = selecting measurement of latest time point rather than highest efficacy; 2 = excluding studies which did not record a "yes" score on the risk of bias and the two overall study quality indicators; 3 = excluding studies which did not correct for potential litter effects. —, Analysis could not be conducted; CI, confidence interval; MD, mean difference; SMD, standardized mean difference.

on strain, timing of outcome measurements, or compartments were conducted because fewer than three studies could be included in the different subgroups.

### Publication Bias

The presence of publication bias was assessed using funnel plots for body weight and fat weight because these outcome measures contained  $\geq 10$  studies. Visual analysis of funnel plots did not suggest substantial publication bias. For body weight, negative studies with a moderate sample size might have been slightly underestimated (see Excel Figure S6A). Nevertheless, this funnel plot does not indicate "strongly suspected" publication bias. In the funnel plot of fat weight, either small studies showing a decreased fat weight appeared to be slightly underestimated or small studies showing an increased fat weight appeared to be slightly over-represented, but this funnel plot also does not indicate "strongly suspected" publication bias (see Excel Figure S6B). Furthermore, in the vast majority of the studies, no conflict-of-interest issues were observed (see Excel Table S4), and no lag phase of "negative" studies was observed for all outcome measures.

### Sensitivity Analyses

Sensitivity analyses had a relatively minor impact on the associations of the overall analyses for most outcome measures (Table 7). In the first sensitivity analysis, in which the time points with greatest efficacy were replaced with the latest measured time points, 89 comparisons were replaced for body weight, and 7 comparisons were replaced for fat weight. For the other outcome measures, all time points with the greatest efficacy were the latest measured time points. This sensitivity analysis changed the overall association of body weight from a significant negative association to a nonsignificant positive association. For the second and third sensitivity analyses, 13 and 23 studies could have been excluded, respectively (see Excel Table S4). Excluding the studies that were not free of potential litter effects resulted in minor changes on the overall association with FFA levels; the overall association changed from a significant positive association to a nonsignificant positive association (Table 7). The same sensitivity analysis changed the direction of the overall estimate for leptin from a nonsignificant positive association to a nonsignificant negative association. Further, no remarkable differences were observed.

### Confidence Rating

The quality of evidence for all outcome measures was downgraded because of serious concerns about risk of bias owing to the many unclear scores and because of serious concerns about unexplained inconsistency (Table 8). Body weight was downgraded for unexplained inconsistency because of varying point estimates, minimal or no overlap of confidence intervals between studies, and substantial heterogeneity ( $I^2 > 75\%$ ), whereas fat weight and leptin were only downgraded because of varying point estimates and minimal overlap of confidence intervals between studies. In contrast, triglyceride was downgraded for unexplained inconsistency because of varying point estimates and moderate heterogeneity ( $I^2 > 50\%$ ), and FFA was downgraded only because of varying point estimates. In addition, the body of evidence for FFA was upgraded by half a confidence level for consistency across species because no differences were estimated between rats and mice and across strains. For all outcome measures, there were indications of dose-response-related effects either across or within some studies. However, the consistency of these indications was not considered to be sufficient to

**Table 8.** Quality of the evidence of the overall effects of bisphenol A on the investigated obesity related outcome measures using the Office of Health Assessment and Translation confidence rating methodology (NTP 2015).

| Outcome measure  | Body of evidence (animal studies)    | Factors for downgrading |                           |              |                | Factors for upgrading |                        |               |                                   | Final confidence rating |
|------------------|--------------------------------------|-------------------------|---------------------------|--------------|----------------|-----------------------|------------------------|---------------|-----------------------------------|-------------------------|
|                  |                                      | Risk of bias            | Unexplained inconsistency | Indirectness | Imprecision    | Publication bias      | Magnitude <sup>d</sup> | Dose response | Residual confounding <sup>e</sup> |                         |
| Body weight      | Initial high confidence (55 studies) | ↓ <sup>b</sup>          | ↓ <sup>c</sup>            | —            | —              | — <sup>d</sup>        | — <sup>e</sup>         | —             | —                                 | Low                     |
| Fat weight       | Initial high confidence (13 studies) | ↓ <sup>f</sup>          | ↓ <sup>g</sup>            | —            | —              | — <sup>d</sup>        | — <sup>e</sup>         | —             | —                                 | Low                     |
| Triglyceride     | Initial high confidence (8 studies)  | ↓ <sup>f</sup>          | ↓ <sup>h</sup>            | —            | —              | — <sup>d</sup>        | — <sup>e</sup>         | —             | —                                 | Low                     |
| Free fatty acids | Initial high confidence (7 studies)  | ↓ <sup>f</sup>          | ↓ <sup>i</sup>            | —            | —              | — <sup>d</sup>        | — <sup>e</sup>         | —             | ↑ <sup>j</sup>                    | Low to Moderate         |
| Leptin           | Initial high confidence (8 studies)  | ↓ <sup>b</sup>          | ↓ <sup>g</sup>            | —            | — <sup>k</sup> | — <sup>d</sup>        | — <sup>e</sup>         | —             | —                                 | Low                     |

Note: —, no concern, or not present; ↓, serious concern; ↑, sufficient to upgrade evidence.

<sup>a</sup>The factors “large effect magnitude” and “residual confounding” were not assessed in this study and consequently were not used to upgrade the evidence.

<sup>b</sup>Serious concern because of many “unclear” scores and a change in direction of the association after sensitivity analyses.

<sup>c</sup>Serious concern because of varying point estimates, minimal or no overlap of confidence intervals between studies, and substantial heterogeneity ( $I^2 > 75\%$ ).

<sup>d</sup>No strongly suspected publication bias observed.

<sup>e</sup>Indications for dose–response effects either within or across studies, but the consistency of these indications was not considered sufficient to upgrade the confidence.

<sup>f</sup>Serious concern because of many “unclear” scores.

<sup>g</sup>Serious concern because of varying point estimates and minimal overlap of confidence intervals between studies.

<sup>h</sup>Serious concern because of varying point estimates and moderate heterogeneity ( $I^2 > 50\%$ ).

<sup>i</sup>Serious concern because of varying point estimates.

<sup>j</sup>No subgroup differences were estimated across species and strains.

<sup>k</sup>Body of evidence was already downgraded for unexplained inconsistency and additional downgrading for imprecision was not considered appropriate (NTP 2015).

upgrade the confidence of evidence for any of the outcome measures. Based on this stringent confidence rating, the quality of evidence was rated as low for the outcome measures body weight, fat weight, triglyceride, and leptin and was rated as low-to-moderate for FFA (Table 8).

## Discussion

To our knowledge, this is the first systematic review with meta-analysis of exposure to BPA and obesity-related outcomes in rodents. Our analysis provides evidence in support of obesogenic effects of early-life exposure to BPA because the obesity-related outcomes fat weight, triglyceride levels, and FFA levels showed significant positive associations with BPA exposure, and the obesity-related outcome leptin showed a nonsignificant positive association. A significant negative association for body weight was estimated. As described in more detail below, there was substantial heterogeneity across studies for all outcomes assessed, and information was insufficient to assess the risk of bias for most studies.

The meta-analysis of fat weight indicated a significant positive overall association with early-life BPA exposure, although no differences between specific fat pads were found. In humans, increases in intraabdominal fat, also known as “dysfunctional adipose tissue,” are particularly significant given the greater risk for metabolic syndrome and cardiometabolic disease due to the accumulation of excess visceral adipose tissue (Després and Lemieux 2006; McCarthy 2014). Multiple biological mechanisms have been proposed to underlie the adipogenic effects of BPA, including an estrogen receptor-dependent mechanism, as well as activation other factors such as insulin growth factor-1, thyroid receptor/retinoic X receptor, and PPAR $\gamma$  (Alonso-Magdalena et al. 2015).

Sex-specific subgroup analyses revealed sex-specific associations, with significant positive associations between BPA exposure and triglyceride and FFA levels observed in males; for females, negative associations were found for body weight and leptin levels. No sex-specific association was estimated for fat weight. Several factors might contribute to differences in the effects of BPA between males and females, including hormonal differences, genetic differences in xenobiotic metabolism, and sex-specific placental responses to environmental factors such as EDCs (Babelova et al. 2015; Gabory et al. 2013; Richter et al. 2007).

Early-life exposure to doses lower than the RfD (50  $\mu\text{g}/\text{kg}/\text{d}$ ) showed a stronger positive association with body weight, triglycerides, and FFA than higher doses (U.S. EPA 1988). In contrast, no association with dose was estimated for fat weight or leptin levels. Low dose effects and nonmonotonic dose–response curves have been reported previously for BPA (Angle et al. 2013; Vandenberg 2014; Vom Saal et al. 2012). In addition, it is possible that very high-exposure doses of BPA might affect overall fitness, resulting in toxicity and in related decreases in some outcome measures.

Subgroup analyses of species and strains did not reveal consistent associations across outcome measures. Rats were associated with higher fat weight after BPA exposure compared with mice, although no differences in species-specific associations were estimated for the other outcome measures. In addition, no consistent differences were estimated for frequency of exposure, route of exposure, or timing of outcome measurements. The differences in association observed for these subgroups are likely to be confounded by other study characteristics and not (directly) related to causal differences in biologic susceptibility. Furthermore, no differences in association were observed for the biological compartments of triglycerides and FFA (i.e.,



blood vs. hepatic levels) or for the time window of exposure across all outcome measures.

### **Strengths and Limitations**

The strength of this systematic review is that many published studies were available, and as a result, many subgroups could be analyzed using a prespecified methodology with a rigorous protocol. We followed the systematic review methodology developed by SYRCLE ([www.syracle.nl](http://www.syracle.nl)), which has been specifically designed to evaluate animal studies, although other methodologies are also available, including those of the Collaborative Approach to Meta-Analysis and Review of Animal Data from Experimental Studies (CAMARADES; [www.camarades.info](http://www.camarades.info)) and OHAT (<http://www.ntp.niehs.nih.gov/go/38673>). Our risk of bias and methodological quality assessments revealed that many studies insufficiently reported their methodology, resulting in an unknown risk of bias. Therefore, this systematic review is limited by the poor reporting quality of the included animal studies. Although comparable to animal studies in other research fields (Hooijmans et al. 2015; Wever et al. 2012), the lack of sufficient reporting is notable and should be improved in future studies. Several checklists to improve reporting quality are available, such as the ARRIVE guidelines (Kilkenny et al. 2010) and the Gold Standard Publication Checklist (Hooijmans et al. 2010a); these should be used when submitting manuscripts.

A substantial amount of heterogeneity is present in the data. Heterogeneity was not notably reduced after subgroup analyses, indicating that variation in the design and quality of the included studies are the main sources of heterogeneity in this systematic review. To account for the heterogeneity, a random effects model was used. Furthermore, we also note that differences in all subgroup-specific estimates should be interpreted with caution given the potential for confounding by study characteristics that might be related to BPA exposures, outcomes, and the subgroup of interest, such as the sex or species of the animal models used in each study. For fat weight, triglycerides, and FFA, differences in SMD estimates among subgroups may have also been confounded by study characteristics related to the standard deviation of the original study estimates, and differences in SMDs between subgroups do not necessarily reflect differences in the magnitudes of associations (i.e., actual differences in the measured outcome between treatment and control groups). In addition, it should be noted that for the outcome measure fat weight, where multiple fat pads from individual animals were often included, no additional correction was applied in the analysis to account for nonindependency.

The funnel plots did not reveal severe asymmetry; therefore, publication bias was assessed as “undetected.” It should be noted, however, that funnel plots have substantial limitations, particularly when multiple end points related to a specific outcome measure are reported in the same study (Guyatt et al. 2011). Therefore, the results of the funnel plots, and the related publication bias, should be interpreted with caution.

Although a very large amount of literature was identified and screened ( $n = 2,535$ ), it is likely that we did not identify all body-weight data available in the literature. Because body weight is a general parameter that is included in most rodent studies, it is often not reported in the title and abstract sections. Consequently, these articles would have been missed during the screening processes of this systematic review.

This systematic review may have been limited by the restriction to English language articles. However, the impact of this restriction appeared to be limited because only three non-English articles were excluded, although it also depends on the comprehensiveness of the PubMed and EMBASE databases with regard

to the indexing of non-English articles. In addition, this study only focused on rodents because these species are specifically relevant for risk-assessment purposes given that their relationship to human biology has been extensively analyzed, and risk-assessment methodology is highly adapted to rodent studies. Future research on additional species would be valuable to compare the effects of perinatal BPA exposure in other species. Furthermore, we did not examine the effects of diet in our review, and we excluded studies that altered the diet during follow-up, for example, by giving a high-fat diet challenge. We excluded these studies because of the increased complexity associated with these additional variables. Nevertheless, a future systematic review investigating the obesogenic effects of early-life BPA exposure and diet would be extremely useful because the developmental time period is critically sensitive to both nutritional and environmental influences that can affect the etiology of obesity (Heindel and Schug 2013). Moreover, although there was rigorous consultation between reviewers when deciding which studies and data to include in this systematic review, the use of two independent reviewers is preferred for screening purposes, study quality assessment, and data extraction because the use of one reviewer might result in more errors (Buscemi et al. 2006).

Initially, we intended to conduct a prespecified subgroup sensitivity analysis excluding studies with a high risk of selection bias at “baseline similarities” for outcome measures that could not be analyzed using different time points. However, none of the included articles had a high risk of bias on this item, and excluding all studies with an unclear risk of bias would have resulted in too few studies to conduct the meta-analyses. Therefore, we deviated from the prespecified protocol and included two additional sensitivity analyses that were conducted for all outcome measures.

### **Risk Assessment Implications**

In this systematic review, BPA was associated with several obesity-related outcomes in rodents at doses  $<50 \mu\text{g}/\text{kg}/\text{d}$ , which is the current RfD for BPA in the United States (U.S. EPA 1988). In Europe, the TDI of  $4 \mu\text{g}/\text{kg}/\text{d}$  set by EFSA in 2015 is mainly based on the adverse effects of BPA exposure on kidneys in mice, in which a 10% change in kidney weight is expected to occur at a concentration of  $8,960 \mu\text{g}/\text{kg}/\text{d}$  (EFSA 2015). Although our findings are subject to a number of limitations and should be interpreted with caution, we believe that they support the need to reexamine BPA safety levels. A similar conclusion, albeit based on different end points, was recently drawn by the Dutch National Institute for Public Health and the Environment (Rijksinstituut voor Volksgezondheid en Milieu; RIVM 2016). A reevaluation of safety levels might be further supported by the fact that BPA has recently been identified as an EDC within Europe based on adverse interactions of BPA with reproductive function, mammary gland development, cognitive function, and metabolism (ECHA 2017). Further, two longitudinal birth cohorts have recently reported associations between prenatal BPA exposure (measured in maternal urine) and obesity-related outcomes, including positive associations with waist circumference and BMI Z-scores at 4 y of age in  $\geq 344$  children from Sabadell, Spain (Valvi et al. 2013), and positive associations with fat mass index, percent body fat, and waist circumference at 7 y of age in 375 children from New York, New York (Hoepner et al. 2016). An expert panel estimated that the potential effects of prenatal BPA exposure on childhood obesity may result in substantial obesity-related health and economic costs (Legler et al. 2015), and in another analysis, it was estimated that the costs of obesity (in children and in adults) that might be attributed

to childhood BPA exposure may outweigh the costs of using safer alternatives to BPA in food-associated uses (Trasande 2014).

It is important to note that although this review focused on BPA, other chemicals with putative obesogenic effects have also been identified. For instance, the BPA analog bisphenol S (Helies-Toussaint et al. 2014) and several other chemical classes have been identified as potential obesogenic chemicals; these include phthalates, organotins, perfluorinated alkyl acids, brominated flame retardants, (non)-dioxin-like polychlorinated biphenyls, and several pesticides (Heindel et al. 2015).

## Conclusions

This systematic review provides evidence of obesogenic effects of early-life exposure to BPA, indicating significant positive associations with fat weight, triglycerides, and FFA as well as a nonsignificant positive association with leptin levels. In contrast, a significant negative association with body weight was estimated. Subgroup analyses revealed positive associations for most outcome measures in males compared with females, as well as stronger positive associations at doses below the current U.S. RfD of 50 µg/kg/d compared with doses above the RfD (U.S. EPA 1988). It should be noted that there was substantial heterogeneity across studies for all outcomes assessed and that information was insufficient to assess the risk of bias for most studies. We recommend further research on the sex-specific effects of BPA, including interaction with diet, as well as elucidation of the underlying mechanisms. In conclusion, our findings provide evidence in support of obesogenic effects of early-life exposure to BPA.

## Acknowledgments

This project received funding from the European Community's Seventh Framework Programme (FP7/2007–2013; grant agreement OBELIX no. 227391) and from Netherlands Organization for Scientific Research (NWO-VIDI 864.09.005). L.T. also acknowledges support from the National Institute of Environmental Health Sciences/National Institutes of Health (R01ES022972).

## References

Ahmed S, Atlas E. 2016. Bisphenol S- and bisphenol A-induced adipogenesis of murine preadipocytes occurs through direct peroxisome proliferator activated receptor gamma activation. *Int J Obes Relat Metab Disord* 40(10):1566–1573, PMID: 27273607, <https://doi.org/10.1038/ijo.2016.95>.

Alonso-Magdalena P, Quesada I, Nadal Á. 2015. Prenatal exposure to BPA and offspring outcomes. *Dose Response* 13(2):1559325815590395, PMID: 26676280, <https://doi.org/10.1177/1559325815590395>.

Alonso-Magdalena P, Vieira E, Soriano S, Menes L, Burks D, Quesada I, et al. 2010. Bisphenol A exposure during pregnancy disrupts glucose homeostasis in mothers and adult male offspring. *Environ Health Perspect* 118(9):1243–1250, PMID: 20488778, <https://doi.org/10.1289/ehp.1001993>.

Anderson OS, Peterson KE, Sanchez BN, Zhang Z, Mancuso P, Dolinoy DC. 2013. Perinatal bisphenol A exposure promotes hyperactivity, lean body composition, and hormonal responses across the murine life course. *FASEB J* 27(4):1784–1792, PMID: 23345456, <https://doi.org/10.1096/fj.12-223545>.

Angle BM, Do RP, Ponzi D, Stahlhut RW, Drury BE, Nagel SC, et al. 2013. Metabolic disruption in male mice due to fetal exposure to low but not high doses of bisphenol A (BPA): evidence for effects on body weight, food intake, adipocytes, leptin, adiponectin, insulin and glucose regulation. *Reprod Toxicol* 42:256–268, <https://doi.org/10.1016/j.reprotox.2013.07.017>.

Babelova A, Burckhardt BC, Salinas-Riester G, Pommerenke C, Burckhardt G, Henjakovic M. 2015. Next generation sequencing of sex-specific genes in the livers of obese ZSF1 rats. *Genomics* 106(4):204–213, PMID: 26200819, <https://doi.org/10.1016/j.ygeno.2015.07.006>.

Boden G. 2008. Obesity and free fatty acids. *Endocrinol Metab Clin North Am* 37(3):635–646, PMID: 18775356, <https://doi.org/10.1016/j.ecl.2008.06.007>.

Buscemi N, Hartling L, Vandermeer B, Tjosvold L, Klassen TP. 2006. Single data extraction generated more errors than double data extraction in systematic reviews. *J Clin Epidemiol* 59(7):697–703, PMID: 16765272, <https://doi.org/10.1016/j.jclinepi.2005.11.010>.

Calafat AM, Kuklennyk Z, Reidy JA, Caudill SP, Ekong J, Needham LL. 2005. Urinary concentrations of bisphenol A and 4-nonylphenol in a human reference population. *Environ. Health Perspect* 113(4):391–395, PMID: 15811827, <https://doi.org/10.1289/ehp.7534>.

Considine RV, Sinha MK, Heiman ML, Kriauciunas A, Stephens TW, Nyce MR, et al. 1996. Serum immunoreactive-leptin concentrations in normal-weight and obese humans. *N Engl J Med* 334(5):292–295, PMID: 8532024, <https://doi.org/10.1056/NEJM199602013340503>.

de Vries RBM, Hooijmans CR, Langendam MW, Leenaars M, Ritskes-Hoitinga M, Wever KE. 2015. A protocol format for the preparation, registration and publication of systematic reviews of animal intervention studies. *Evidence-based Preclin Med* 2(1):1–9, <https://doi.org/10.1002/ebm2.7>.

de Vries RBM, Hooijmans CR, Tillema A, Leenaars M, Ritskes-Hoitinga M. 2014. Updated version of the Embase search filter for animal studies. *Lab Anim* 48(1):88, PMID: 23836850, <https://doi.org/10.1177/0023677213494374>.

Després J-P, Lemieux I. 2006. Abdominal obesity and metabolic syndrome. *Nature* 444(7121):881–887, PMID: 17167477, <https://doi.org/10.1038/nature05488>.

ECHA (European Chemical Agency). 2017. Agreement of the member state committee on the identification of bisphenol A as a substance of very high concern. <https://echa.europa.eu/documents/10162/ac9efb97-c06b-d1a7-2823-5dc69208a238> [accessed 21 June 2017].

EFSA (European Food Safety Authority). 2015. Factsheet: Scientific opinion on bisphenol A. [http://www.efsa.europa.eu/sites/default/files/corporate\\_publications/files/factsheetbpa150121.pdf](http://www.efsa.europa.eu/sites/default/files/corporate_publications/files/factsheetbpa150121.pdf) [accessed 14 December 2015].

Gabory A, Roseboom TJ, Moore T, Moore LG, Junien C. 2013. Placental contribution to the origins of sexual dimorphism in health and diseases: sex chromosomes and epigenetics. *Biol Sex Differ* 4(1):5, PMID: 23514128, <https://doi.org/10.1186/2042-6410-4-5>.

Galic S, Oakhill JS, Steinberg GR. 2010. Adipose tissue as an endocrine organ. *Mol Cell Endocrinol* 316(2):129–139, PMID: 19723556, <https://doi.org/10.1016/j.mce.2009.08.018>.

Grün F, Blumberg B. 2006. Environmental obesogens: organotins and endocrine disruption via nuclear receptor signaling. *Endocrinology* 147(6):S50–S55, PMID: 16690801, <https://doi.org/10.1210/en.2005-1129>.

Guyatt GH, Oxman AD, Montori V, Vist G, Kunz R, Brozek J, et al. 2011. GRADE guidelines: 5. Rating the quality of evidence—publication bias. *J Clin Epidemiol* 64(12):1277–1282, PMID: 21802904, <https://doi.org/10.1016/j.jclinepi.2011.01.011>.

Heindel JJ, Newbold RR, Schug TT. 2015. Endocrine disruptors and obesity. *Nat Rev Endocrinol* 11(11):653–661, PMID: 26391979, <https://doi.org/10.1038/nrendo.2015.163>.

Heindel JJ, Schug TT. 2013. The perfect storm for obesity. *Obesity (Silver Spring)* 21(6):1079–1080, PMID: 23784921, <https://doi.org/10.1002/oby.20222>.

Helies-Toussaint C, Peyre L, Costanzo C, Chagnon M-C, Rahmani R. 2014. Is bisphenol S a safe substitute for bisphenol A in terms of metabolic function? An *in vitro* study. *Toxicol Appl Pharmacol* 280(2):224–235, PMID: 25111128, <https://doi.org/10.1016/j.taap.2014.07.025>.

Hoepner LA, Whyatt RM, Widen EM, Hassoun A, Oberfield SE, Mueller NT, et al. 2016. Bisphenol A and adiposity in an inner-city birth cohort. *Environ Health Perspect* 124(10):1644–1650, PMID: 27187982, <https://doi.org/10.1289/EHP205>.

Hooijmans CR, Geessink FJ, Ritskes-Hoitinga M, Scheffer G. 2015. A systematic review and meta-analysis of the ability of analgesic drugs to reduce metastasis in experimental cancer models. *Pain* 156(10):1835–1844, PMID: 26181303, <https://doi.org/10.1097/j.pain.0000000000000296>.

Hooijmans CR, Leenaars M, Ritskes-Hoitinga M. 2010a. A gold standard publication checklist to improve the quality of animal studies, to fully integrate the three Rs, and to make systematic reviews more feasible. *Altern Lab Anim* 38(2):167–182, PMID: 20507187.

Hooijmans CR, Rovers MM, de Vries RB, Leenaars M, Ritskes-Hoitinga M, Langendam MW. 2014. SYRCL's risk of bias tool for animal studies. *BMC Med Res Methodol* 14:43, PMID: 24667063, <https://doi.org/10.1186/1471-2288-14-43>.

Hooijmans CR, Tillema A, Leenaars M, Ritskes-Hoitinga M. 2010b. Enhancing search efficiency by means of a search filter for finding all studies on animal experimentation in PubMed. *Lab Anim* 44(3):170–175, PMID: 20551243, <https://doi.org/10.1258/la.2010.009117>.

Howdeshell KL, vom Saal FS. 2000. Developmental exposure to bisphenol A: Interaction with endogenous estradiol during pregnancy in mice. *Am Zool* 40(3):429–437, <https://doi.org/10.1093/icb/40.3.429>.

Ichihara T, Yoshino H, Imai N, Tsutsumi T, Kawabe M, Tamano S, et al. 2003. Lack of carcinogenic risk in the prostate with transplacental and lactational exposure to bisphenol A in rats. *J Toxicol Sci* 28(3):165–171, PMID: 12974608, <https://doi.org/10.2131/jts.28.165>.

- Kilkenny C, Browne WJ, Cuthill IC, Emerson M, Altman DG. 2010. Improving bioscience research reporting: The ARRIVE guidelines for reporting animal research. *PLoS Biol* 8(6):e1000412, PMID: 20613859, <https://doi.org/10.1371/journal.pbio.1000412>.
- Le Corre L, Besnard P, Chagnon M-C. 2015. BPA, an energy balance disruptor. *Crit Rev Food Sci Nutr* 55(6):769–777, PMID: 24915348, <https://doi.org/10.1080/10408398.2012.678421>.
- Leenaars M, Hooijmans CR, van Veggel N, ter Riet G, Leeflang M, Hooft L, et al. 2012. A step-by-step guide to systematically identify all relevant animal studies. *Lab Anim* 46(1):24–31, PMID: 22037056, <https://doi.org/10.1258/la.2011.011087>.
- Legler J, Fletcher T, Govarts E, Porta M, Blumberg B, Heindel JJ, et al. 2015. Obesity, diabetes, and associated costs of exposure to endocrine-disrupting chemicals in the European Union. *J Clin Endocrinol Metab* 100(4):1278–1288, PMID: 25742518, <https://doi.org/10.1210/jc.2014-4326>.
- McCarthy HD. 2014. Measuring growth and obesity across childhood and adolescence. *Proc Nutr Soc* 73(2):210–217, PMID: 24423184, <https://doi.org/10.1017/S0029665113003868>.
- Nagao T, Saito Y, Usumi K, Yoshimura S, Ono H. 2002. Low-dose bisphenol A does not affect reproductive organs in estrogen-sensitive C57BL/6N mice exposed at the sexually mature, juvenile, or embryonic stage. *Reprod Toxicol* 16(2):123–130, PMID: 11955943, [https://doi.org/10.1016/S0890-6238\(02\)00003-5](https://doi.org/10.1016/S0890-6238(02)00003-5).
- NTP (National Toxicology Program). 2010. Bisphenol A (BPA). Bisphenol A factsheet. [https://www.niehs.nih.gov/health/assets/docs\\_a\\_e/bisphenol\\_a\\_bpa\\_508.pdf](https://www.niehs.nih.gov/health/assets/docs_a_e/bisphenol_a_bpa_508.pdf) [accessed 26 November 2015].
- NTP. 2015. OHAT systematic review. <http://ntp.niehs.nih.gov/go/38673> [accessed 9 January 2017].
- Richter CA, Birnbaum LS, Farabolini F, Newbold RR, Rubin BS, Talsness CE, et al. 2007. In vivo effects of bisphenol A in laboratory rodent studies. *Reprod Toxicol* 24(2):199–224, PMID: 17683900, <https://doi.org/10.1016/j.reprotox.2007.06.004>.
- RIVM [Rijksinstituut voor Volksgezondheid en Milieu (Dutch National Institute for Public Health and the Environment)]. 2016. "Bisphenol A: Part 2. Recommendations for risk management (RIVM Report 2015-0192)." [http://www.rivm.nl/en/Documents\\_and\\_publications/Scientific/Reports/2016/maart/Bisphenol\\_A\\_Part\\_2\\_Recommendations\\_for\\_risk\\_management](http://www.rivm.nl/en/Documents_and_publications/Scientific/Reports/2016/maart/Bisphenol_A_Part_2_Recommendations_for_risk_management).
- Rubin BS. 2011. Bisphenol A: an endocrine disruptor with widespread exposure and multiple effects. *J Steroid Biochem Mol Biol* 127(1–2):27–34, PMID: 21605673, <https://doi.org/10.1016/j.jsbmb.2011.05.002>.
- Subramanian S, Chait A. 2012. Hypertriglyceridemia secondary to obesity and diabetes. *Biochim Biophys Acta* 1821(5):819–825, PMID: 22005032, <https://doi.org/10.1016/j.bbali.2011.10.003>.
- Trasande L. 2014. Further limiting bisphenol A in food uses could provide health and economic benefits. *Health Aff* 33(2):316–323, PMID: 24452104, <https://doi.org/10.1377/hlthaff.2013.0686>.
- U.S. EPA (U.S. Environmental Protection Agency). 1988. Integrated Risk Information System: Bisphenol A. [http://cfpub.epa.gov/ncea/iris/index.cfm?fuseaction=iris.showQuickview&substance\\_nmbr=0356](http://cfpub.epa.gov/ncea/iris/index.cfm?fuseaction=iris.showQuickview&substance_nmbr=0356) [accessed 15 December 2015].
- Valvi D, Casas M, Mendez MA, Ballesteros-Gómez A, Luque N, Rubio S, et al. 2013. Prenatal bisphenol A urine concentrations and early rapid growth and overweight risk in the offspring. *Epidemiology* 24(6):791–799, PMID: 24036610, <https://doi.org/10.1097/EDE.0b013e3182a67822>.
- Vandenberg LN. 2014. Non-monotonic dose responses in studies of endocrine disrupting chemicals: bisphenol A as a case study. *Dose Response* 12(2):259–276, PMID: 24910584, <https://doi.org/10.2203/dose-response.13-020.Vandenberg>.
- Vandenberg LN, Hauser R, Marcus M, Olea N, Welshons WV. 2007. Human exposure to bisphenol A (BPA). *Reprod Toxicol* 24(2):139–177, PMID: 17825522, <https://doi.org/10.1016/j.reprotox.2007.07.010>.
- Vom Saal FS, Nagel SC, Coe BL, Angle BM, Taylor JA. 2012. The estrogenic endocrine disrupting chemical bisphenol A (BPA) and obesity. *Mol Cell Endocrinol* 354(1–2):74–84, PMID: 22249005, <https://doi.org/10.1016/j.mce.2012.01.001>.
- Wassenaar PNH, Legler J. 2015. Systematic review protocol: the effects of early life exposure to endocrine disrupting chemicals on obesity development in rodents. [https://issuu.com/radboudumc/docs/the\\_effects\\_of\\_early-life\\_exposure\\_e-28355229/48259284](https://issuu.com/radboudumc/docs/the_effects_of_early-life_exposure_e-28355229/48259284) [accessed 10 December 2015].
- Wever KE, Menting TP, Rovers M, van der Vliet JA, Rongen GA, Masereeuw R, et al. 2012. Ischemic preconditioning in the animal kidney, a systematic review and meta-analysis. *PLoS One* 7(2):e32296, PMID: 22389693, <https://doi.org/10.1371/journal.pone.0032296>.
- WHO (World Health Organization). 2015. Obesity and overweight. Fact sheet. <http://www.who.int/mediacentre/factsheets/fs311/en/> [accessed 26 November 2015].
- Xu X, Tan L, Himi T, Sadamatsu M, Tsutsumi S, Akaike M, et al. 2011. Changed preference for sweet taste in adulthood induced by perinatal exposure to bisphenol A-A probable link to overweight and obesity. *Neurotoxicol Teratol* 33(4):458–463, PMID: 21704699, <https://doi.org/10.1016/j.ntt.2011.06.002>.