Substance Name: Butyl 4-hydroxybenzoate

EC Number: 202-318-7

CAS Number: 94-26-8

MEMBER STATE COMMITTEE SUPPORT DOCUMENT FOR IDENTIFICATION OF

BUTYL 4-HYDROXYBENZOATE

AS A SUBSTANCE OF VERY HIGH CONCERN BECAUSE OF ITS ENDOCRINE DISRUPTING PROPERTIES (ARTICLE 57(F) - HUMAN HEALTH) PROPERTIES

Adopted on 28 May 2020

CONTENTS

TABLES

ABBREVIATIONS

AFC	Antral follicle count
AGD	Anogenital distance
AhR	Aryl hydrocarbon receptor
AO	Adverse outcome
AR	Androgen receptor
BP	Butylparaben
BPA	Bisphenol A
CI	Confidence interval
CoRAP	Community rolling action plan
CYP	Cytochrome P450
DAG	Directed acyclic graph
DSP	Daily sperm production
E2	Estradiol
EATS	Estrogenic, androgenic, thyroidal and steroidogenic
EC50	The half maximal effective concentration
ED	Endocrine disrupting
EDC	Endocrine disrupting chemical
EFSA	European Food Safety Authority
ER	Estrogen receptor
ERC	Environmental release category
ESR	Estrogen receptor
FLG	Filaggrin gene
FOR	Fecundability odds ratio
FSH	Follicle-stimulating hormone
FT	Free testosterone
GA	Gestational age
GD	Gestation day
GLP	Good laboratory practice
GPER	G protein coupled estrogen receptor
GPR	G protein coupled receptor
GREB	Growth Regulating Estrogen Receptor Binding
GW	Gestational week
HE	Histological evaluation
HSD	Hydroxysteroid dehydrogenase
IC50	The half maximal inhibitory concentration
IHC	Immunohistochemistry
IVF	*In vitro* fertilisation
JRC	Joint Research Centre
KE	Key event
KER	Key event relationship
LBW	Low birth weight
LGA	Large for gestational age
LH	Luteinising hormone
LOAEL	Lowest observed adverse effect level
LOD	Limit of detection
LOEC	Lowest observed effect concentration
LOEL	Lowest observed effect level
LOQ	Limit of quantification
MIE	Molecular initiating event
MoA	Mode of action
MP	Methylparaben

NOAEL	No observed adverse effect level
NOEC	No observed effect concentration
NOEL	No observed effect level
OV	Ovarian volume
OVX	Ovariectomised rat
PC	Product category
PCOS	Polycystic ovarian syndrome
PD	Pup day
PE	Preeclampsia
PHBA	Parahydroxybenzoic acid / 4-hydroxybenzoic acid
PHHA	p-hydroxyhippuric acid
PND	Postnatal day
PP	Propylparaben
PROC	Process category
PTB	Preterm birth
REC	Relative effect concentration
s.c.	Subcutaneous injection
SCCS	Scientific Committee on Consumer Safety
SD	Sprague Dawley
SG	Specific gravity
SGA	Small for gestational age
SHBG	Sex hormone binding globulin
TBG	Thyroxine-binding globulin
TSH	Thyroid-stimulating hormone
TTP	Time to pregnancy
VO	Vaginal opening

IDENTIFICATION OF A SUBSTANCE OF VERY HIGH CONCERN ON THE BASIS OF THE CRITERIA SET OUT IN REACH ARTICLE 57

Substance Names: Butyl 4-hydroxybenzoate (Butylparaben)

EC Number: 202-318-7

CAS number: 94-26-8

- The substance is identified as a substance of equivalent level of concern to those of other substances listed in points (a) to (e) of Article 57 of Regulation (EC) No 1907/2006 (REACH) according to Article 57(f) of REACH Regulation.

Summary of how the substance meets the criteria set out in Article 57 of the REACH Regulation

Butyl 4-hydroxybenzoate (commonly referred to as butylparaben) is identified as a substance of very high concern in accordance with Article 57(f) of Regulation (EC) 1907/2006 (REACH) because of its endocrine disrupting properties for which there is scientific evidence of probable serious effects to human health which gives rise to an equivalent level of concern to those of other substances listed in points (a) to (e) of Article 57 REACH.

Endocrine disrupting (ED) properties of butylparaben relevant for human health:

The adverse health effect is reduced sperm count and quality (number of normal sperm) as observed in rodent studies using perinatal exposure. These effects are considered severe as similar effects in humans could cause sub- and infertility. Effects are irreversible and are shown to occur later in life after exposure in the perinatal period only. There is supportive evidence from studies on adverse effects on sperm count and quality in rodents following pubertal and/or adult exposure, although there are some inconsistencies between these studies.

There is sufficient evidence to conclude that butylparaben acts via endocrine mode(s) of action according to a mode of action analysis including an evaluation of biological plausibility. The evidence is strongest for estrogenic activity shown *in vitro* and *in vivo*, but some studies also point to altered steroidogenesis and androgen receptor (AR) antagonism.

Based on the above conclusion, evidence that the substance is of an equivalent level of concern includes:

Sub- and infertility are not only detrimental to the propagation of the species, but also have a major impact on quality of life. Fertility treatment and counselling carry high societal costs.

No safe concentration/level can be derived from the available data on adverse reproductive effects via endocrine modes of action. Two of the available studies show reduced sperm count or quality in perinatally exposed rats at the lowest tested dose, and no NOAEL can be determined for this endpoint.

Registration dossiers submitted for the substance? Yes

Justification

1. Identity of the substance and physical and chemical properties

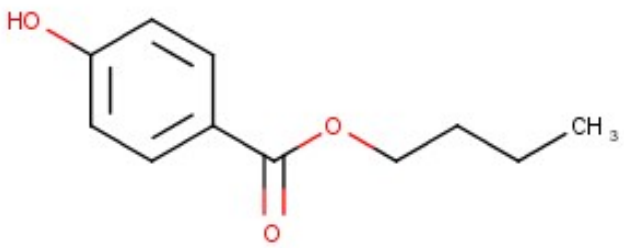
1.1 Name and other identifiers of the substance

Table 1: Substance identity

EC number:	202-318-7
EC name:	Butyl 4-hydroxybenzoate
CAS number (in the EC inventory):	94-26-8
CAS name:	N/A
IUPAC name:	Butyl p-hydroxybenzoate
Index number in Annex VI of the CLP Regulation	N/A
Molecular formula:	C11H14O3
Molecular weight range:	194.23
Synonyms:	Butylparaben

Substance type: mono-constituent

Structural formula:



1.2 Physicochemical properties

Table 2: Overview of physicochemical properties (based on the registration information[1])

Property	Description of key information	Value	Reference/source of information
Physical state at 20°C and 101.3 kPa	Visual inspection	Solid: White particulate/powder	ECHA dissemination site[1]: 29.11.2019
Melting/freezing point	Data from handbook	68.5 °C	D.R. CRC Handbook of Chemistry and Physics 86TH Edition 2005-2006
Boiling point	differential scanning calorimetry method	>= 330 - <= 337 °C at 102.4 kPa	OECD Guideline 103
Vapour pressure		0.002 Pa at 20°C, 0.005 Pa at 25°C, 0.113 Pa at 50°C	EPI Suite version 4.11, Mpbpwin v. 1.43 REACH Guideline on QSARs R.6
Density		1.2365 g/cm3 at 20.0 °C	OECD Guideline 109
Water solubility	Data from handbook	207 mg/L at 20 °C	Yalkowsky, F. H. and He, Y.; Handbook of Aqueous Solubility Data, CRC Press, Boca Raton, FL; ISBN 0-8493-1532-8
Partition coefficient n-octanol/water (log value)	Data from handbook	log Pow = 3.57	Hansch. C., A. Leo and D. Hoekman. 1995. Exploring QSAR. Hydrophobic, Electronic, and Steric Constants. ACS Professional Reference Book. Washington, DC: American Chemical Society.

[1] https://echa.europa.eu/registration-dossier/-/registered-dossier/25335 (information as disseminated on 29 November 2019)

2 Harmonised classification and labelling

There is no harmonised classification for butylparaben and the substance is therefore not covered in Annex VI to the CLP Regulation.

3 Environmental fate properties

Not relevant for the identification of the substance as SVHC in accordance with Article 57 point (f) REACH due to its endocrine disrupting properties for human health.

4 Human health hazard assessment

Butylparaben is considered a substance of very high concern due to its endocrine disrupting properties. The evaluation of endocrine disrupting properties is carried out taking into account *inter alia* the Guidance for the identification of endocrine disruptors in the context of Regulations (EU) No 528/2012 and (EC) No 1107/2009 (ECHA/EFSA 2018), and the WHO/IPCS definition of an endocrine disruptor (WHO/IPCS, 2002) as interpreted by the JRC Endocrine Expert Advisory Group (2013). The ECHA/EFSA document is intended to provide guidance to applicants and assessors of the competent regulatory authorities on the implementation of the scientific criteria for the determination of endocrine disrupting properties of or biocidal products and plant protection products. Here, similar approach is taken for a REACH substance in order to provide a systematic overview of the available data. Consequently, the justification part of the report is accompanied by the following five annexes:

- Annex I - Lines of evidence for adverse effects and endocrine activity
- Annex II – Mode of action analysis
- Annex III – Detailed study information
- Annex IV – Data overview (ED Guidance table)
- Annex V – Human epidemiology studies

An overview of available studies is given in Section 4.5 but study details are presented in Annexes III to V.

4.1 Toxicokinetics

Butylparaben is metabolised to parahydroxybenzoic acid (PHBA) and a large proportion of PHBA is excreted as p-hydroxyhippuric acid (PHHA, the glycine conjugate of PHBA). Recent studies (Moos *et al*. 2016) indicate the presence of other metabolites in human urine. This report does not include a review of metabolism, but this is noted as some metabolites are also tested *in vitro* (see Section 4.5).

4.2 Acute toxicity

Information from ECHA's dissemination site[2]:

A read-across approach was conducted on the source substance isobutyl 4-hydroxybenzoate for acute toxicity:

[2] https://echa.europa.eu/registration-dossier/-/registered-dossier/25335 (information as disseminated on 29 November 2019)

Acute toxicity after a single oral application was tested in female rats, which received 2000 mg/kg bw (two groups of three females). All animals survived until the end of the study. The most relevant clinical findings in the animals treated with the test item at a dose of 2000 mg/kg bw were reduced spontaneous activity, prone position, moving the bedding, ataxia, hunched posture, piloerection and half eyelid closure. All symptoms recovered within 1 day post-dose.

Throughout the 14-day observation period, the weight gain of the animals was within the normal range of variation for this strain.

At necropsy, no macroscopic findings were observed in any animal of any step in the acute toxicity test.

4.3 Repeated dose toxicity

No information at ECHA's dissemination site. This document focuses on endocrine disrupting properties and therefore repeated dose toxicity studies related to reproductive and endocrine disrupting targets were included. See discussion in Section 4.5 and study descriptions in Annex IV.

4.4 Toxicity for reproduction

No information at ECHA's dissemination site. This document focuses on endocrine disrupting properties, and therefore relevant reproductive and developmental toxicity studies were included in the assessment. Details on those studies are presented in Section 4.5 and Annexes III to V.

Effects on fertility related parameters (sperm count, sperm quality, testicular function) and development (birth parameters, growth, early reproductive development and late life effects on male reproduction) of rodents exposed to butylparaben have been described in peer-reviewed publications.

Effects on human fertility and development have been described in studies examining exposures in adulthood and during pregnancy and childhood in relation to health effects early or late in life.

4.5 Endocrine Activity

This section presents an assessment of butylparaben as an SVHC of "Equivalent level of concern" due to endocrine disrupting properties.

For butylparaben, the Lines of Evidence for endocrine disrupting properties are presented in Annex I and a Mode of Action analysis is presented in Annex II in accordance with the ECHA/EFSA 2018 guidance. Annex III provides information on the literature search as well as a list of relevant publications. Annex IV (excel sheet) provides information from all relevant studies. The results of these analyses are provided in Section 4.5.1 and 4.5.2.

4.5.1 Lines of evidence

The analysis of lines of evidence (Annex I) for endocrine disrupting properties of butylparaben was carried out for effects of perinatal exposure and pubertal/adult exposure separately. Lines of evidence focus on male reproductive toxicity, with only effects in exposed males or male offspring of exposed females presented.

4.5.1.1 In silico

The Danish QSAR Database[3] predicts butylparaben as an Estrogen Receptor (ER) α binder and activator (Exp.: POS and Battery: POS_IN) negative in Androgen receptor antagonism (Exp.: NEG, Battery: NEG), and inconclusive in Arylhydrocarbon (AhR) Activation, pregnant X receptor and CYP34A induction. Butylparaben was predicted as a moderate binder to the Estrogen Receptor (OH-group), and both weak and strong binder to metabolites from *in vivo* Rat metabolism.

The ToxCast Database predicts butylparaben as being active in 111 out of 738 assays. Looking at results specifically in EDSP21, which contains classical endocrine activity sensitive assays such as AR, ER, thyroid and steroidogenesis assays, butylparaben is active in 15 out of 28 ER related assays.

These predictions are supporting evidence for estrogenic activity (estrogen receptor binding and activation) of butylparaben.

4.5.1.2 *In vitro* and *in vivo* mechanistic studies

Studies on endocrine activity of butylparaben (estrogen receptor activation, altered steroidogenesis, androgen receptor antagonism) have been described in peer-reviewed publications. Details on those studies are presented in Annex IV following the guidance from ECHA/EFSA (2018).

In vitro and *in vivo* mechanistic studies presented in the lines of evidence analysis revealed that there is sufficient evidence for estrogenicity of butylparaben, and there is supporting evidence for altered steroidogenesis and anti-androgenic effects based on less consistent data.

Several studies show an estrogen receptor agonistic response similar to estrogen (Gonzalez *et al*, 2018; Pop *et al*, 2018; Watanabe *et al*, 2013), and several studies show effects on growth of estrogen sensitive cells (Khanna & Darbre, 2013; Charles & Darbre, 2013; Gonzalez *et al*, 2018; Pop *et al*, 2018; Williams *et al*, 2019, van Meeuwen *et al*. 2008) or tissues (uterotropic assay *in vivo*: Routledge *et al*, 1998; Hossaini *et al*, 2000; Lemini *et al*, 2003; Lemini *et al*, 2004; Goswami & Kalita, 2016; Vo & Jeung, 2009). Two uterotrophic studies reported no effect (Guerra *et al*, 2017a; Shaw & DeCatanzaro, 2009). One study showed no effect after 21 days of exposure, but changes in uterine histology at the prepubertal stage of a postnatal female rat model (Vo *et al*, 2010). The parent compound is more active than the metabolite PHBA (as reviewed by (Boberg *et al*, 2010).

With regard to altered steroidogenesis, one study shows upregulation of aromatase/Cyp19a1 gene expression and increased aromatase enzyme activity as well as increased estradiol levels in different cell lines (Williams *et al*, 2019). Another study reports decreased aromatase activity in microsomes from human placenta after exposure (van Meeuwen *et al*. 2008). Other studies show no increase in estradiol levels in other cell lines, and the response may be cell type specific (Taxvig *et al*, 2008; Wróbel & Gregoraszczuk, 2013; Guerra *et al*, 2016).

Studies on anti-androgenic activity show inconsistent findings. Effects are observed in some (Chen *et al*, 2007; Pop *et al*, 2016) but not all studies (Kjærstad *et al*, 2010; Watanabe *et al*, 2013), possibly due to different study design. Studies on androgen activity consistently show no androgen activity (agonism) of butylparaben (Gonzalez *et al*, 2018; Chen *et al*, 2007; Pop *et al*, 2016).

There are no data from Hershberger studies on butylparaben, but a study on the structurally related propyl 4-hydroxybenzoate (propylparaben) showed clear evidence of antiandrogenic effect in a Hershberger assay (Özdemir *et al*. 2018). The toxicological profiles of butylparaben

[3] Danish (Q)SAR Database (2019), http://qsar.food.dtu.dk, assessed 29 November 2019

and propylparaben are considered similar as data from Ahn *et al*. (2012), Darbre & Khanna (2013), Khanna *et al*. (2014) and Lee *et al*. (2017) consistently suggests that read-across between those substances can be justified. Furthermore, SCCS (2013) concluded that toxicokinetic data from *in vitro* and *in vivo* rat studies can be justified for read-across between butylparaben and propylparaben.

Another study (Watanabe *et al*, 2013) shows that incubation of butylparaben with rat liver microsomes reduced the activation of ERalpha and ERbeta compared to exposure without liver microsomes. The activity of the metabolite 4-hydroxybenzoic acid (PHBA) on ERalpha and ERbeta was assessed and found not to be active on ERalpha and weakly active on ERbeta.

Studies on development and fertility related parameters of rodents exposed to butylparaben during early development or in puberty and adulthood have been described in peer-reviewed publications. Details on those studies are presented in Annex IV following the guidance from ECHA/EFSA (2018).

4.5.1.3 Endocrine activity *in vivo* and adversity

Perinatal exposure

For perinatal exposure, the lines of evidence analysis revealed supporting data on endocrine activity of butylparaben, as there are some indications of the effect of butylparaben on the androgen-sensitive endpoint anogenital distance (AGD) in male offspring. Moderate reduction of AGD was seen in two studies evaluating AGD at postnatal day (PND) 1, using doses of 400 mg/kg bw/day or above (Zhang *et al*, 2014; Boberg *et al*, 2016), but not in other studies using lower doses (Kang *et al*, 2002; Guerra *et al*, 2017b) or using high doses but evaluating AGD at gestation day (GD) 21 (Taxvig *et al*, 2008). Inconsistency between studies may be due to different exposure periods, dose levels and measuring sensitivity.

Supporting evidence of endocrine activity *in vivo* was seen for altered serum hormone levels. Two studies showed increases in estradiol at PND 21 and PND 90 at high doses (Zhang *et al*, 2014; Zhang *et al*, 2016). Another study showed increased testosterone levels at high doses at PND 45 and PND 75. In addition, a small decrease was seen in intermediate dose on PND 75 (Maske *et al*. 2020) Two studies showed reduced testosterone levels at selected ages at high doses, while a third showed an increase at an intermediate dose level (Zhang *et al*, 2014; Zhang *et al*, 2016; Guerra *et al*, 2017b).

For luteinising hormone (LH) and follicle-stimulating hormone (FSH) levels, data were inconsistent between studies with different exposure periods, routes and dose levels (Zhang *et al*, 2014; Guerra *et al*, 2017b; Maske *et al*., 2020). Collectively, these hormone data point to endocrine activity differing depending on study design.

In addition, there is good evidence that butylparaben adversely affects sperm parameters. Several studies show reduced sperm counts or quality in adult male rats after exposure to butylparaben in the perinatal period only (Kang *et al*, 2002; Zhang *et al*, 2014; Boberg *et al*, 2016; Guerra *et al*, 2017b; Maske *et al*., 2020).

Four studies showed effects on sperm count, whereas one study did not (Guerra *et al*, 2017b). That study had fewer animals than (some of) the other studies, and did not use as high doses as two of the studies showing reduced sperm count. However, that study showed effects on other sperm parameters including reduced number of sperm with progressive motility and increased numbers of abnormal sperm (Guerra *et al*, 2017b). Sperm motility was investigated in two other studies showing a reduced percentage of motile sperm (Kang *et al*, 2002; Maske *et al*., 2020).

Whereas studies by Kang *et al*. (2002), Boberg *et al*. (2016) and Maske *et al*. (2020), initiated exposure at GD 6 or 7 the study by Guerra *et al*. (2017b) started exposure at GD 12. This may explain some differences between studies in patterns of later in life effects in male offspring. Two of the studies including doses of 400 mg/kg bw/day or above showed reduced sperm counts at these doses (Zhang *et al*, 2014; Boberg *et al*, 2016), whereas the Maske *et al*. (2020) study saw a statistically significant effect on sperm count (67% of control numbers) at 100 mg/kg bw/day but not at 1000 mg/kg bw/day. However, even though not statistically significant, the average sperm count in the high dose group was approximately 80% of control, and the variance in the control group was notable. At lower doses some studies showed effects and other studies did not. The results on sperm motility were considered supportive evidence. Differences in study design (route and timing of exposure) may in part explain these differences.

For testicular histopathology, no effects were seen on foetal testis in two studies (Fisher *et al*, 1999; Taxvig *et al*, 2008), whereas signs of histological effects on seminiferous tubules of prepubertal testes were seen in one study (Kang *et al*, 2002) and change in adult testes were seen in four out of five studies (Zhang *et al*, 2014; Zhang *et al*, 2016; Guerra *et al*, 2017b; Maske *et al*. 2020). As different histological endpoints were examined in each study, these findings are considered supporting evidence and not sufficient as evidence of adverse effects.

Thus, there is sufficient evidence of decreased sperm count and decreased number of normal sperm (adverse effects) and supporting evidence of altered hormone levels (endocrine activity *in vivo*) and altered sperm motility, altered AGD in some studies and reports of histological changes in testes in some studies (adverse effects) (see Annex I).

In addition, a few studies examined effects in female rodents exposed in utero or early postnatally, but findings were not considered sufficient evidence of adverse effects. These studies are presented in Annex III and summarised here:

In one study (Ahn *et al*, 2012), increased uterus weight and altered uterine gene expression were observed in female SD rats exposed from PND 1 to 7 by subcutaneous injections to butylparaben. In the ovaries, numbers of primordial follicles increased, primary follicle numbers decreased and changes in gene expression were seen for cell specific markers and factors involved in steroidogenesis. The authors concluded that the examined parabens inhibited the early phase of folliculogenesis. Another study (Maske *et al*, 2018) showed delayed puberty and reduced estrous cycle length in Holzman rats exposed from GD 6 to PND 21 by subcutaneous injections to butylparaben. Hormone levels and ovarian follicle numbers at the highest doses (100 and 1000 mg/kg bw/day) were altered, but the pattern of changes differed with age and dose. Adult ovarian expression of genes for estrogen receptors and a steroidogenesis related factor were altered at the middle dose (100 mg/kg bw/day). Uterus histology was affected at the two highest doses (100 and 1000 mg/kg bw/day), and weights of adrenal, hypothalamus, pituitary, ovary and uterus were altered in different directions depending on age and doses. In mated adult offspring, increased pre-and post-implantation loss was seen. In Wistar rats orally exposed from GD 7 to PND 22, reduced AGD was seen in females at the two highest doses, 100 and 500 mg/kg bw/day (Boberg *et al*, 2016; male effects described in Annex I and IV). Ovary weights were reduced at PND 17, and mammary gland outgrowth was increased at PND 22 (100 and 500 mg/kg bw/day). These studies support both the endocrine activity of butylparaben and the plausibility that this can lead to adverse effects *in vivo*, but are not considered sufficient evidence of adverse effects, and these studies are therefore not included in the lines of evidence analysis (Annex I).

No epidemiological studies examined the relationship between butylparaben exposure in utero and the effects on male reproductive parameters (hormone levels, sperm parameters) later in life (see Annex V).

One study provided supporting evidence for endocrine effects as it showed a negative association between maternal urinary butyl paraben levels and maternal serum levels of estradiol and the estradiol/progesterone ratio (Aker *et al*, 2019a).

A larger study (Aker *et al*, 2016) showed a borderline trend of lower maternal testosterone levels with higher maternal butylparaben levels and a significant negative association between maternal butylparaben and sex hormone binding globulin (SHBG). These studies are supportive evidence for an endocrine disrupting activity of butylparaben during pregnancy.

One study (Fernandez *et al*, 2016) observed no association between placental butylparaben levels and congenital malformations of the male genitalia (cryptorchidism and hypospadias). Residues of butylparaben were however, more frequently detected in cases versus controls. This is supporting evidence for adverse effects of butylparaben exposure during pregnancy.

Overall, there is sufficient evidence for endocrine activity of butylparaben *in vitro* supported by *in vivo* evidence in studies using perinatal exposure. *In vitro*, the evidence is strongest for estrogenic activity, whereas supporting evidence of altered steroid synthesis and anti-androgenic activity points to the possibility of multiple modes of action.

There is sufficient evidence of adverse effect *in vivo* (decreased sperm count and decreased number of normal sperm), supported by evidence of endocrine activity *in vivo* (altered hormone levels) and adverse effects (altered anogenital distance, decreased sperm motility, altered testicular histopathology).

According to ECHA/EFSA (2018), a positive conclusion on a) Adversity based on 'Estrogenic, androgenic, thyroidal and steroidogenic (EATS)-mediated' parameters and b) Mechanistic tests at OECD Conceptual Framework level 2 or 3 (*in vitro* and *in vivo* mechanistic) lead to the next step of performing a Mode of Action analysis. No further data on mode of action or effects of butylparaben in relation to perinatal exposure are considered necessary.

Pubertal/adult exposure

For pubertal/adult exposure, the lines of evidence analysis revealed supporting evidence of endocrine activity *in vivo* (altered hormone levels) and supporting evidence of adverse effects (decreased sperm count and sperm quality, altered testicular histopathology) in male rodents. The conclusion that there is not sufficient evidence of adverse effects is based on inconsistent data from studies with some variation in study design.

For altered hormone levels, one study (Riad *et al*, 2018) showed increased serum estradiol after 8 weeks of dosing, and four studies showed reduced testosterone levels at selected ages at high doses (Oishi, 2001; Oishi, 2002; Hoberman *et al*, 2008; Riad *et al*, 2018). For LH levels, decreases were seen in two studies but at different ages and doses (Hoberman *et al*, 2008; Riad *et al*, 2018). For FSH levels, data differed between studies using different dose levels (Hoberman *et al*, 2008; Riad *et al*, 2018). Collectively, these data on hormone levels point to endocrine activity differing depending on study design.

Adverse effects (reduced sperm numbers) were seen in three smaller studies (Oishi, 2001; Garcia *et al*, 2017; Riad *et al*, 2018) but not in one larger study (Hoberman *et al*, 2008). Previous evaluations of the Scientific Committee on Consumer Safety (SCCS) (2013) concluded that the negative study (Hoberman *et al*, 2008) is not accepted as "unarguable refutation of the Oishi findings" (Oishi, 2001). Studies from Garcia *et al* (2017) and Riad *et al* (2018) were published in 2017 and 2018, respectively, and thus not included in SCCS evaluations. Data on sperm motility and morphology can be considered supportive information, together with the reduced sperm number, despite inconsistency between studies. For testicular histopathology different parameters were assessed in all studies, except for one Hoberman *et al*, 2008), and the reported findings can also be considered supporting evidence of adverse effects.

In addition, a few studies examined effects in female rodents exposed in puberty or adulthood, but the findings were not considered sufficient evidence of adverse effects. These studies are presented in Annex III and are summarised here:

In a study exposing females from PND 21 to 40 by oral gavage changes in uterine histology as well as reduced number of corpora lutea was seen. No effects were seen on uterine weight (Vo *et al*. 2010). In 8-week old female Sprague Dawley (SD) rats (Lee *et al*, 2017) exposed for 5 weeks by gavage to one dose of butylparaben, estrous cycle length was affected, and changes in ovarian expression of selected genes related to steroidogenesis and ovary development were altered at the end of dosing. Serum FSH was increased and the number of secondary follicles and Graafian follicles were reduced. In another study (Pollock *et al*, 2017), estradiol levels were increased 6, 8 and 10 hours after a single dose of butylparaben. These studies support both the endocrine activity of butylparaben and the plausibility that this can lead to adverse effects *in vivo*, but are not considered sufficient evidence of adverse effects, and these studies are therefore not included in the lines of evidence analysis (Annex I).

A few epidemiological studies examined the relationship between butylparaben exposure in adults and male reproductive parameters (hormone levels, sperm parameters) (see Annex V).

One study (Smarr *et al*, 2018) observed an inverse association between male urinary butylparaben levels and sperm concentration and sperm motility. This paper by Smarr *et al* (2018) had a "moderate" quality/reliability score due to not adjusting models for abstinence time and not reporting clearly the butylparaben detection rate.

Three studies (Joensen *et al*, 2018; Meeker *et al*, 2011; Nishihama *et al*, 2017) observed no significant association between male urine butylparaben and classical sperm parameters (concentration, total count, motility) although the paper by Meeker *et al* (2011) observed a significant association between butylparaben and increased sperm DNA damage. However, especially the papers written by Joensen *et al* (2018) and Nishihama *et al* (2017) had "low" quality/reliability scores due to small study size and/or low butylparaben detection rate and are likely underpowered. The paper by Meeker *et al* (2011) also had a low butylparaben detection rate. The paper by Jurewicz *et al*. (2017) observed a positive association between male urinary butylparaben and sperm XY disomy, however, this paper was scored "low" as butylparaben could only be detected in 14/156 men. The paper by Buck Louis *et al* (2018) studied male seminal plasma butylparaben levels in relation to time-to-pregnancy and found no associations. This study had a butylparaben detection rate of 55%.

This is supporting evidence for adverse effects of butylparaben exposure in adulthood. It is noted that especially studies in males were challenged by a low butylparaben detection rate as men seem to be less exposed to butylparaben than adult women are (Annex V).

Overall, there is sufficient evidence of endocrine activity *in vitro* supported by evidence of endocrine activity *in vivo* with pubertal/adult exposure. *In vitro*, the evidence is strongest for estrogenic activity, whereas supporting evidence of altered steroid synthesis and anti-androgenic activity points to the possibility of multiple modes of action. There is supporting evidence of endocrine activity *in vivo* (altered hormone levels).

There is supporting evidence of adverse effects after pubertal/adult exposure (decreased sperm count, number of normal sperm, motility, altered testicular histopathology). The conclusion that there is not sufficient evidence of adverse effects is based on inconsistent data from studies with some variation in study design.

According to ECHA/EFSA (2018), a negative conclusion on a) Adversity based on 'EATS-mediated' parameters and a positive conclusion on b) Mechanistic tests at OECD Conceptual Framework level 2 or 3 (*in vitro* and *in vivo* mechanistic) lead to the next step of performing Mode of Action analysis. As all 'EATS-mediated' parameters in relation to adverse effects have not been investigated, additional information e.g. from level 3, 4 or 5 studies may need to be

generated in relation to pubertal/adult exposure before a conclusion on the endocrine disrupting properties can be made.

4.5.2 Mode of action analysis

The mode of action (MoA) analysis (see Annex II for further details) was carried out for effects of perinatal exposure and pubertal/adult exposure separately. The steps in the mode of action analysis include an overview of key events, an analysis of biological plausibility of key event relationships, and considerations on dose and temporal concordance, human relevance and uncertainties.

Perinatal exposure

For perinatal exposure, the analysis leads to the conclusion that butylparaben acts via multiple modes of action, and it is biologically plausible that estrogen receptor activation (directly or due to increased estradiol levels) and androgen receptor antagonism during development leads to adverse effects on male reproductive function following perinatal exposure.

The mode of action of butylparaben is based on "EATS-mediated adversity", and therefore the substance is considered to be an endocrine disrupter. No alternative non-endocrine mode of action is demonstrated.

The table below presents the results of the Mode of action analysis for perinatal exposure:

Mode of action analysis	There is sufficient evidence of endocrine activity (estrogen receptor activation and possibly altered steroidogenesis and androgen receptor antagonism) and adverse effects (decreased sperm count and quality).
Biological plausibility	It is biologically plausible that adverse effects are due to the endocrine activity of butylparaben.
Dose and temporal concordance	In each study, indicators of key events related to endocrine activity are affected at the same doses causing adverse effects. Between studies, there are differences in effective doses. Key events are observed in the hypothesised order, i.e. *in vivo* indicators of endocrine activity are seen in developing animals, and adverse effects are seen in adulthood.
Essentiality	Essentiality has not been investigated.
Human relevance	Human relevance is assumed, as there are no data indicating that these endocrine modes of action are not relevant to humans.
Uncertainties	The uncertainty analysis highlights that the evidence base for butylparaben is relatively limited, yet there is consistency between different studies on both an endocrine mode of action and adverse effects.
Conclusion: Butylparaben likely acts via multiple modes of action and it is biologically plausible that the endocrine activity of butylparaben leads to the observed adverse effects on the male reproductive system. For estrogen receptor activation (directly or due to increased estradiol levels) the evidence for each key event relationship is considered "High", except the step "Increased estrogen receptor signalling to Altered reproductive development of offspring", for which the evidence is considered "Moderate to high". For Androgen receptor antagonism, evidence for all key event relationships is considered "High". There is sufficient dose and temporal concordance between key events, and effects are assumed relevant to humans.	

Pubertal/adult exposure

For pubertal/adult exposure, the analysis leads to the conclusion that butylparaben acts via multiple modes of action, and it is biologically plausible that the endocrine activity of butylparaben leads to adverse effects on the male reproductive system following pubertal/adult exposure. However, there is insufficient data to reach a conclusion on potential adverse effects on the male reproductive system following pubertal/adult exposure.

The postulated mode of action of butylparaben is based on "EATS mediated adversity", but there is not sufficient data on adverse effects to conclude on endocrine disruption for pubertal/adult exposure. No alternative non-endocrine mode of action is demonstrated.

The table below presents the result of the Mode of action analysis for pubertal/adult exposure:

Mode of action analysis	There is sufficient evidence of endocrine activity (estrogen receptor activation and possibly altered steroidogenesis and androgen receptor antagonism). There is only supportive evidence for adverse effects (decreased sperm count and quality, testicular histopathology).
Biological plausibility	It is biologically plausible that potential adverse effects are due to the endocrine activity of butylparaben.
Dose and temporal concordance	In each *in vivo* study, indicators of key events related to endocrine activity (altered hormone levels) are affected at the same doses causing adverse effects on sperm parameters. Between studies, there are however some differences in effective doses, possibly depending on study design and sensitivity. Key events are observed in the hypothesised order.
Essentiality	Essentiality has not been investigated.
Human relevance	Human relevance is assumed, as there are no data indicating that these endocrine modes of action are not relevant to humans.
Uncertainties	This uncertainty analysis highlights that the evidence base for butylparaben is relatively limited. Lack of consistency between different *in vivo* studies on adverse effects are a major uncertainty.
Conclusion: Butylparaben likely acts via multiple modes of action and it is biologically plausible that the endocrine activity of butylparaben leads to adverse effects on the male reproductive system following pubertal/adult exposure. For estrogen receptor activation (directly or due to increased estradiol levels) the evidence for each key event relationship is considered "High", except the step "Increased estrogen receptor signalling to Altered testicular function", for which the evidence is considered "moderate". For Androgen receptor antagonism, the evidence for each key event relationship is considered "High", except the step "Reduced androgen receptor signalling to Altered testicular function", for which the evidence is considered "moderate". However, there is insufficient data to reach a conclusion on potential adverse effects on the male reproductive system following pubertal/adult exposure. There is sufficient dose and temporal concordance between key events, and effects are assumed relevant to humans.	

The <u>overall conclusion</u> is that butylparaben can be considered an endocrine disrupter for human health. This is based on sufficient information on an endocrine mode of action and the adverse effects of butylparaben following perinatal exposure.

5 Environmental hazard assessment

Environmental data for butylparaben was not reviewed and thus not included in this document.

6 Conclusions on the SVHC Properties

6.1 Equivalent level of concern assessment

6.1.1 Summary of the data provided

The data provided in Section 4.5 and Annexes I (lines of evidence analysis), II (mode of action analysis) and III (data summaries) are summarised here.

The endocrine disrupting properties of butylparaben are established taking into account *inter alia* the guidance of ECHA/EFSA (2018).

6.1.1.1 Lines of evidence

The analysis of lines of evidence (Annex I) for endocrine disrupting properties of butylparaben was carried out for effects of perinatal exposure and pubertal/adult exposure separately.

Endocrine activity *in silico* and *in vitro*

In silico, the predictions found supportive evidence on endocrine activity of butylparaben, as there are alerts and positive results on estrogen receptor binding and activation.

In vitro, the lines of evidence analysis revealed that there is sufficient evidence for estrogenicity of butylparaben, and there is supporting evidence for altered steroidogenesis and anti-androgenic effects based on less consistent data.

Endocrine activity *in vivo* and adversity

Perinatal exposure

There is good evidence that butylparaben adversely affects sperm parameters. Several studies show reduced sperm counts and/or quality in adult male rats after exposure to butylparaben in the perinatal period only (Section 4.5.1 and Annex I). Supporting evidence of adverse effects are the altered AGD in some studies, altered sperm motility and reports of histological changes in testis in some studies (see Annex I). A few studies reporting effects on female reproductive tissues of perinatally exposed rodents are considered as supporting evidence of endocrine activity and plausibility of *in vivo* effects.

No epidemiological studies examined the relationship between butylparaben exposure *in utero* and effects on male reproductive parameters (hormone levels, sperm parameters) later in life.

According to ECHA/EFSA (2018), a positive conclusion on a) Adversity based on 'EATS-mediated' parameters and b) Mechanistic tests at OECD Conceptual Framework level 2 or 3 (*in vitro* and *in vivo* mechanistic) leads to the next step of performing a Mode of Action analysis. No further data on mode of action or effects of butylparaben in relation to perinatal exposure are considered necessary.

Pubertal/adult exposure

For pubertal/adult exposure, the lines of evidence analysis revealed supporting evidence of endocrine activity *in vivo* (altered hormone levels) and supporting evidence of adverse effects (decreased sperm count, decreased number of normal sperm and motility, altered testicular histopathology). The conclusion that there is not sufficient evidence of adverse effects is based on inconsistent data from studies with some variation in study design (Section 4.5.1 and Annex I). Studies reporting (mainly uterotrophic) effect on female reproductive tissues after adult exposure are considered supporting evidence of endocrine activity.

A few epidemiological studies examined the relationship between butylparaben exposure in adults and male reproductive parameters (hormone levels, sperm parameters) (see Section 4.5 and Annex V). This is considered as supporting evidence for adverse effects of butylparaben exposure in adulthood.

According to ECHA/EFSA (2018), a negative conclusion on a) Adversity based on 'EATS-mediated' parameters and a positive conclusion on b) Mechanistic tests at OECD Conceptual Framework level 2/3 (*in vitro* and *in vivo* mechanistic) leads to the next step of performing a Mode of Action analysis. As all 'EATS-mediated' parameters in relation to adverse effects have not been investigated, additional information e.g. from level 3, 4 or 5 studies may need to be generated in relation to pubertal/adult exposure before a conclusion on the endocrine disrupting properties can be made.

6.1.1.2 Mode of action analysis

The mode of action analysis (Annex II) was carried out for effects of perinatal exposure and pubertal/adult exposure separately. The steps in the mode of action analysis included an overview of key events, an analysis of the biological plausibility of key event relationships, and considerations on dose and temporal concordance, human relevance and uncertainties.

Perinatal exposure

For perinatal exposure, the analysis leads to the conclusion that butylparaben acts via multiple modes of action, and it is biologically plausible that the endocrine activity of butylparaben leads to the observed adverse effects on the male reproductive system following perinatal exposure.

The mode of action of butylparaben is based on "EATS-mediated adversity", and the substance can be considered an endocrine disrupter. No alternative non-endocrine mode of action is demonstrated. For estrogen receptor activation (directly or due to increased estradiol levels) the evidence for each key event relationship is considered "High", except the step "Increased estrogen receptor signalling to Altered reproductive development of offspring", for which the evidence is considered "Moderate to high". For androgen receptor antagonism, evidence for all key event relationships is considered "High". There is sufficient dose and temporal concordance between key events, and effects are assumed to be relevant for humans.

Pubertal/adult exposure

For pubertal/adult exposure, the analysis leads to the conclusion that butylparaben acts via multiple modes of action, and it is biologically plausible that the endocrine activity of butylparaben leads to adverse effects on the male reproductive system following pubertal/adult exposure. However, the data is insufficient to reach a conclusion on potential adverse effects on the male reproductive system following pubertal/adult exposure.

The postulated mode of action of butylparaben is based on "EATS mediated adversity", but there is insufficient data on adverse effects to conclude on the endocrine disruption. No alternative non-endocrine mode of action is demonstrated. For estrogen receptor activation (directly or due to increased estradiol levels) the evidence for each key event relationship is considered "High", except the step "Increased estrogen receptor signalling to altered testicular function", for which the evidence is considered "moderate". For Androgen receptor antagonism, the evidence for each key event relationship is considered "High", except the step "Reduced androgen receptor signalling to altered testicular function", for which the evidence is considered "moderate". However, the data is insufficient to reach a conclusion on potential adverse effects on the male reproductive system following pubertal/adult exposure.

6.1.2 Equivalent level of concern assessment

Butylparaben exposure gives rise to an equivalent level of concern to substances listed in Article 57 points (a) to (e) due to its endocrine disrupting properties for human health.

The adverse health effects are reduced sperm count and quality as observed in rodent studies (Section 4.5). These effects are considered severe as similar effects in humans could cause sub- and infertility. Effects are irreversible and are shown to occur later in life after exposure in the perinatal period only.

Sub- and infertility is not only detrimental to the propagation of the species, but also has a major impact on quality of life. Fertility treatment and counselling carries high societal costs.

No safe concentration/level can be derived from the available data on adverse reproductive effects via endocrine modes of action. Two of the available studies show reduced sperm count or quality in perinatally exposed rats at the lowest tested dose, and no NOAEL (no observed adverse effect level) can be determined for this endpoint. A study used by the EU Scientific Committee on Consumer Safety (SCCS, 2013) for safety evaluation did not include evaluation of these endpoints (Fisher *et al*, 1999).

6.1.3 Conclusion on whether the substance gives rise to an equivalent level of concern

Butyl 4-hydroxybenzoate (commonly referred to as butylparaben) is identified as a substance of very high concern in accordance with Article 57(f) of Regulation (EC) 1907/2006 (REACH) because of its endocrine disrupting properties for which there is scientific evidence of probable serious effects to human health which gives rise to an equivalent level of concern to those substances listed in points (a) to (e) of Article 57 REACH.

Endocrine disrupting (ED) properties of butylparaben relevant for human health:

The adverse health effect is reduced sperm count and quality (number of normal sperm) as observed in rodent studies using perinatal exposure. These effects are considered severe as similar effects in humans could cause sub- and infertility. Effects are irreversible and are shown to occur later in life after exposure in the perinatal period only. There is supportive evidence from studies on adverse effects on sperm count and quality in rodents following pubertal and/or adult exposure, although there are some inconsistencies between these studies.

There is sufficient evidence to conclude that butylparaben acts via endocrine mode(s) of action according to a mode of action analysis including an evaluation of biological plausibility. The evidence is strongest for estrogenic activity shown *in vitro* and *in vivo*, but some studies also point to altered steroidogenesis and androgen receptor antagonism.

Based on the above conclusion, evidence that the substance is of an equivalent level of concern includes:

Sub- and infertility are not only detrimental to the propagation of the species, but also have a major impact on quality of life. Fertility treatment and counselling carry high societal costs.
No safe concentration/level can be derived from the available data on adverse reproductive effects via endocrine modes of action. Two of the available studies show reduced sperm count or quality in perinatally exposed rats at the lowest tested dose, and no NOAEL can be determined for this endpoint.

References

Ahn HJ, An BS, Jung EM, Yang H, Choi KC, Jeung EB. (2012) ´Parabens inhibit the early phase of folliculogenesis and steroidogenesis in the ovaries of neonatal rats´, *Molecular Reproduction and Development*, 79(9), pp. 626-636. doi: 10.1002/mrd.22070.

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

Aker AM, Ferguson KK, Rosario ZY, Mukherjee B, Alshawabkeh AN, Calafat AM, Cordero JF, Meeker JD. (2019a) A repeated measures study of phenol, paraben and Triclocarban urinary biomarkers and circulating maternal hormones during gestation in the Puerto Rico PROTECT cohort. Environ Health. 2019 Apr 2;18(1):28.

Aker AM, Ferguson KK, Rosario ZY, Mukherjee B, Alshawabkeh AN, Calafat AM, Cordero JF, Meeker JD. (2019b) The associations between prenatal exposure to triclocarban, phenols and parabens with gestational age and birth weight in northern Puerto Rico. Environmental Research 2019. 169:41-51

Akingbemi BT. (2005) Estrogen regulation of testicular function. Reprod Biol Endocrinol 2005; 3:51.

Alam, M. S. and Kurohmaru, M. (2014) 'Disruption of Sertoli cell vimentin filaments in prepubertal rats: An acute effect of butylparaben *in vivo* and *in vitro*', *Acta Histochemica*. Elsevier GmbH., 116(5), pp. 682–687. doi: 10.1016/j.acthis.2013.12.006.

Berger K, Gunier RB, Chevrier J, Calafat AM, Ye X, Eskenazi B, Harley KG. (2018) Associations of maternal exposure to triclosan, parabens, and other phenols with prenatal maternal and neonatal thyroid hormone levels. Environ Res. 2018 Aug;165:379-386.

Boberg J, Axelstad M, Svingen T, Mandrup K, Christiansen S, Vinggaard AM, Hass U. (2016) 'Multiple Endocrine Disrupting Effects in Rats Perinatally Exposed to Butylparaben', *Toxicological Sciences*, 152(1), pp. 244–256. doi: 10.1093/toxsci/kfw079.

Buck Louis GM, Smarr MM, Sun L, Chen Z, Honda M, Wang W, Karthikraj R, Weck J, Kannan K. (2018) Endocrine disrupting chemicals in seminal plasma and couple fecundity. Environ Res. 2018 May;163:64-70.

Charles AK, Darbre PD. (2013) 'Combinations of parabens at concentrations measured in human breast tissue can increase proliferation of MCF-7 human breast cancer cells', *Journal of Applied Toxicology*, 33(5), pp. 390–398. doi: 10.1002/jat.2850.

Chen J, Ahn KC, Gee NA, Gee SJ, Hammock BD, Lasley BL. (2007) 'Antiandrogenic properties of parabens and other phenolic containing small molecules in personal care products', *Toxicology and Applied Pharmacology*, 221(3), pp. 278–284. doi: 10.1016/j.taap.2007.03.015.

Danish (Q)SAR Database (2019), http://qsar.food.dtu.dk, assessed 29. November, 2019

Delbès G, Levacher C, Habert R. (2006) Estrogen effects on fetal and neonatal testicular development. Reproduction. 2006 Oct;132(4):527-38.

Delbes G, Levacher C, Pairault C, Racine C, Duguenne C, Krust A, Habert R. (2004) Estrogen receptor beta-mediated inhibition of male germ cell line development in mice by endogenous estrogens during perinatal life. Endocrinology 2004; 145:3395-3403.

Delbes G, Levacher C, Duguenne C, Racine C, Pakarinen P, Habert R. (2005) Endogenous

estrogens inhibit mouse fetal Leydig cell development via estrogen receptor alpha. Endocrinology 2005; 146:2454-2461.

Engeli RT, Rohrer SR, Vuorinen A, Herdlinger S, Kaserer T, Leugger S, Schuster D, Odermatt A. (2017) 'Interference of Paraben Compounds with Estrogen Metabolism by Inhibition of 17β-Hydroxysteroid Dehydrogenases', *International Journal of Molecular Sciences*, 18(9), p. 2007. doi: 10.3390/ijms18092007.

ECHA/EFSA 2018. ECHA (European Chemicals Agency) and EFSA (European Food Safety Authority) with the technical support of the Joint Research Centre (JRC). Guidance for the identification of endocrine disruptors in the context of Regulations (EU) No 528/2012 and (EC) No 1107/2009. EFSA Journal 2018;16(6):5311, 135 pp. https://doi.org/10.2903/j.efsa.2018.5311. ECHA-18-G-01-EN.

EU (2017) Commission Delegated Regulation (EU) 2017/2100 of 4 September 2017 setting out scientific criteria for the determination of endocrine-disrupting properties pursuant to Regulation (EU) No 528/2012 of the European Parliament and Council. OJ L 301, 17.11.2017, p. 1–5. Available online: http://data.europa.eu/eli/reg_del/2017/2100/oj

EU (2018) Commission Regulation (EU) 2018/605 of 19 April 2018 setting out scientific criteria for the determination of endocrine disrupting and amending Annex II to Regulation (EC) 1107/2009. OJ L 101, 20.4.2018, p. 33–36. Available online: http://data.europa.eu/eli/reg/2018/605/oj

Fernandez MF, Arrebola JP, Jimenez-Diaz I, Saenz JM, Molina-Molina JM, Ballesteros O, Kortenkamp A, and Olea N. (2016) Bisphenol A and other phenols in human placenta from children with cryptorchidism or hypospadias. Reprod.Toxicol. 2016;59:89-95

Fisher JS, Turner KJ, Brown D, Sharpe RM. (1999) 'Effect of neonatal exposure to estrogenic compounds on development of the excurrent ducts of the rat testis through puberty to adulthood.', *Environmental Health Perspectives*, 107(5), pp. 397–405. doi: 10.1289/ehp.99107397.

Garcia T, Schreiber E, Kumar V, Prasad R, Sirvent JJ, Domingo JL, Gómez M. (2017) 'Effects on the reproductive system of young male rats of subcutaneous exposure to n-butylparaben', *Food and Chemical Toxicology*, 106, pp. 47–57. doi: 10.1016/j.fct.2017.05.031.

Geer LA, Pycke BFG, Waxenbaum J, Sherer DM, Abulafia O, Halden RU. (2017) Association of birth outcomes with fetal exposure to parabens, triclosan and triclocarban in an immigrant population in Brooklyn, New York. J Hazard Mater. 2017 Feb 5;323(Pt A):177-183

Gonzalez TL, Moos RK, Gersch CL, Johnson MD, Richardson RJ, Koch HM, Rae JM. (2018) 'Metabolites of n-Butylparaben and iso-Butylparaben Exhibit Estrogenic Properties in MCF-7 and T47D Human Breast Cancer Cell Lines', *Toxicological Sciences*, 164(1), pp. 50–59. doi: 10.1093/toxsci/kfy063.

Goswami & Kalita (2016) ´Alteration in uterine histoarchitecture and uterine protein following butylparaben exposure in adult mice´, *Int. Res. J. Pharm*, 7(2), pp. 25-30. doi: 10.7897/2230-8407.07215.

Guerra MT, Furlong HC, Kempinas WG, Foster WG. (2016) 'Effects of *in vitro* exposure to butylparaben and di-(2 ethylhexyl) phthalate, alone or in combination, on ovarian function', *Journal of Applied Toxicology*, 36(9), pp. 1235–1245. doi: 10.1002/jat.3335.

Guerra MT, Sanabria M, Cagliarani SV, Leite GAA, Borges CDS, Kempinas WDG. (2017a) ´Long-term effects of in utero and lactational exposure to butyl paraben in female rats´, *Environmental Toxicology*, 32, pp. 776-788. doi: 10.1002/tox.22277.

Guerra MT, Sanabria M, Leite GAA, Borges CS, Cucielo MS, Anselmo-Franci JA, Fister WG, Kempinas WG. (2017b) 'Maternal exposure to butyl paraben impairs testicular structure and sperm quality on male rats', *Environmental Toxicology*, 32(4), pp. 1273–1289. doi: 10.1002/tox.22323.

Guo J, Wu C, Lu D, Jiang S, Liang W, Chang X, Xu H, Wang G, Zhou Z. (2017) Urinary paraben concentrations and their associations with anthropometric measures of children aged 3 years. Environ Pollut. 2017 Mar;222:307-314

Hoberman AM, Schreur DK, Leazer T, Daston GP, Carthew P, Re T, Loretz L, Mann P. (2008) 'Lack of effect of butylparaben and methylparaben on the reproductive system in male rats', *Birth Defects Research Part B: Developmental and Reproductive Toxicology*, 83(2), pp. 123–133. doi: 10.1002/bdrb.20153.

Hossaini A, Larsen JJ, Larsen JC. (2000)´Lack of oestrogenic effects of food preservatives (Parabens) in uterotrophic assays´, *Food and Chemical Toxicology*, 38(4), pp. 319-323. doi: 10.1016/s0278-6915(99)00160-x.

Jefferson WN, Couse JF, Banks EP, Korach KS, Newbold RR. (2000) Expression of estrogen receptor beta is developmentally regulated in reproductive tissues of male and female mice. Biol Reprod 2000; 62:310-317

JRC (Joint Research Centre) (2013). Key Scientific issues relevant to the identification and characterisation of endocrine disrupting substances – Report of the Endocrine Disruptors Expert Advisory Group (ED EAG). Eds. Munn S. and Gourmenou M. Pp 32. Available at: https://ec.europa.eu/jrc/en/publication/eur-scientific-and-technical-research-reports/key-scientific-issues-relevant-identification-and-characterisation-endocrine-disrupting

JRC (Joint Research Centre) (2016). Screening methodology to identify potential endocrine disruptors according to different options in the context of an impact assessment. In: JRC Science for policy reports. Publications Office of the European Union, Luxembourg. 54 pp. https://doi.org/10.2788/73203

Joensen UN, Jørgensen N, Thyssen JP, Szecsi PB, Stender S, Petersen JH, Andersson AM, Frederiksen H. (2018) Urinary excretion of phenols, parabens and benzophenones in young men: Associations to reproductive hormones and semen quality are modified by mutations in the Filaggrin gene. Environ Int. 2018 Dec;121(Pt 1):365-374.

Joseph A, Shur BD, Hess RA. (2011) Estrogen, efferent ductules, and the epididymis. Biol Reprod. 2011 Feb;84(2):207-17. doi: 10.1095/biolreprod.110.087353.

Jurewicz J, Radwan M, Wielgomas B, Klimowska A, Kałużny P, Radwan P, Jakubowski L, Hanke W. (2017) Environmental exposure to parabens and sperm chromosome disomy. Int J Environ Health Res. 2017 Oct;27(5):332-343

Kang KS, Che JH, Ryu DY, Kim TW, Li GX, Lee YS (2002) 'Decreased Sperm Number and Motile Activity on the F1 Offspring Maternally Exposed to Butyl p-Hydroxybenzoic Acid (Butyl Paraben)', *Journal of Veterinary Medical Science*, 64(3), pp. 227–235. doi: 10.1292/jvms.64.227.

Khanna S, Darbre PD. (2013) 'Parabens enable suspension growth of MCF-10A immortalized, non-transformed human breast epithelial cells', *Journal of Applied Toxicology*, 33(5), pp. 378–382. doi: 10.1002/jat.2753.

Khanna S, Dash PR, Darbre PD. (2014) 'Exposure to parabens at the concentration of maximal proliferative response increases migratory and invasive activity of human breast cancer cells *in vitro*', *Journal of Applied Toxicology*, 34(9), pp. 1051–1059. doi: 10.1002/jat.3003.

Kjaerstad, MB, Taxvig C, Andersen HR, Nelleman C. (2010) 'Mixture effects of endocrine disrupting compounds *in vitro*', *International Journal of Andrology*, 33(2), pp. 425–433. doi: 10.1111/j.1365-2605.2009.01034.x.

Klopčič I, Kolšek K, Dolenc MS. (2015) 'Glucocorticoid-like activity of propylparaben, butylparaben, diethylhexyl phthalate and tetramethrin mixtures studied in the MDA-kb2 cell line', *Toxicology Letters*, 232(2), pp. 376–383. doi: 10.1016/j.toxlet.2014.11.019.

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

Lassurguère J, Livera G, Habert R, Jégou B. (2003) Time- and dose-related effects of estradiol and diethylstilbestrol on the morphology and function of the fetal rat testis in culture. Toxicol Sci 2003; 73:160-169

Lee JH, Lee M, Ahn C, Kang HY, Tran DN, Jeung EB. (2017) ´Parabens Accelerate Ovarian Dysfunction in a 4-Vinylcyclohexene Diepoxide-Induced Ovarian Failure Model´, *International Journal of Environmental Research and Public Health*, 14(2), pp. E161. doi: 10.3390/ijerph14020161.

Lehraiki A, Chamaillard C, Krust A, Habert R, Levacher C. (2011) Genistein impairs early testosterone production in fetal mouse testis via estrogen receptor alpha. Toxicol *In vitro* 2011; 25:1542-1547.

Lemini C, Jaimez R, Ávila ME, Franco Y, Larrea F, Lemus AE. (2003) ´*In vivo* and *in vitro* estrogen bioactivities of alkyl parabens´, *Toxicological and Industrial Health*, 19(2-6), pp. 69-79. doi: 10.1191/0748233703th177oa.

Lemini C, Hernández A, Jaimez R, Franco Y, Avila ME, Castell A. (2004) ´Morphometric analysis of mice uteri treated with the preservatives methyl, ethyl, propyl, and butylparaben´, *Toxicological and Industrial Health,* 20(6-10), pp. 123-132. doi: 10.1191/0748233704th202oa.

Martin OV, Shialis T, Lester JN, Scrimshaw MD, Boobis AR, Voulvoulis N. (2008) Testicular dysgenesis syndrome and the estrogen hypothesis: a quantitative meta-analysis. Environ Health Perspect. 2008 Feb;116(2):149-57. doi: 10.1289/ehp.10545.

Martinez RM, Hauser R, Liang L, Mansur A, Adir M, Dioni L, Racowsky C, Bollati V, Baccarelli AA, Machtinger R. (2019) Urinary concentrations of phenols and phthalate metabolites reflect extracellular vesicle microRNA expression in follicular fluid. Environ Int. 2019 Feb;123:20-28.

Maske P, Dighe V, Vanage G. (2018) ´n-butylparaben exposure during perinatal period impairs fertility of the F1 generation female rats´, *Chemosphere*, 213, pp. 114-123. doi: 10.1016/j.chemosphere.2018.08.130.

Maske P, Dighe V, Mote C, Vanage G (2020) n-Butylparaben exposure through gestation and lactation impairs spermatogenesis and steroidogenesis causing reduced fertility in the F1 generation male rats, Environ. Pollut. 2020; 256: 112957 (pp. 1-11). doi: 10.1016/j.envpol.2019.112957.

McLachlan JA. (2001) Environmental signaling: what embryos and evolution teach us about endocrine disrupting chemicals. Endocr Rev. 2001 Jun;22(3):319-41.

Meeker JD, Yang T, Ye X, Calafat AM, and Hauser R. (2011) Urinary concentrations of parabens and serum hormone levels, semen quality parameters, and sperm DNA damage. Environ.Health Perspect. 2011;119(2):252-257

Meistrich ML, Hughes TJ, Bruce WR. (1975) Alteration of epididymal sperm transport and maturation by oestrogen and testosterone. Nature 1975;258:145-147.

Messerlian C, Mustieles V, Minguez-Alarcon L, Ford JB, Calafat AM, Souter I, Williams PL, Hauser R (2018) Environment and Reproductive Health (EARTH) Study Team. Preconception and prenatal urinary concentrations of phenols and birth size of singleton infants born to mothers and fathers from the Environment and Reproductive Health (EARTH) study. Environ Int. 2018 May;114:60-68.

Minguez-Alarcon L, Chiu YH, Messerlian C, Williams PL, Sabatini ME, Toth TL, Ford JB, Calafat AM, and Hauser R. (2016) Urinary paraben concentrations and *in vitro* fertilization outcomes among women from a fertility clinic. Fertil.Steril. 2016;105(3):714-721.

Moos RK, Angerer J, Dierkes G, Brüning T, and Koch HM. (2016) Metabolism and elimination of methyl, iso- and n-butyl paraben in human urine after single oral dosage. Arch Toxicol 2016;90:2699–2709. doi: 10.1007/s00204-015-1636-0

Nielsen M, Björnsdóttir S, Høyer PE, Byskov AG. (2000) Ontogeny of oestrogen receptor alpha in gonads and sex ducts of fetal and newborn mice. J Reprod Fertil 2000; 118:195-204.

Nishihama Y, Toshima H, Yoshinaga J, Mizumoto Y, Yoneyama M, Nakajima D, Shiraishi H, Tokuoka S. (2017) Paraben exposure and semen quality of Japanese male partners of subfertile couples. Environ Health Prev Med. 2017 Mar 15;22(1):5.

Nishihama Y, Yoshinaga J, Iida A, Konishi S, Imai H, Yoneyama M, Nakajima D, Shiraishi H. (2016) Association between paraben exposure and menstrual cycle in female university students in Japan. Reprod Toxicol. 2016 Aug;63:107-13

O'Donnell L, Robertson KM, Jones ME, Simpson ER. (2001) Estrogen and spermatogenesis. Endocr Rev 2001; 22:289-318.

OECD (2018), Revised Guidance Document 150 on Standardised Test Guidelines for Evaluating Chemicals for Endocrine Disruption, OECD Series on Testing and Assessment, No. 150, OECD Publishing, Paris, https://doi.org/10.1787/9789264304741-en.

O'Hara L and Smith LB. (2015) "Androgen receptor roles in spermatogenesis and infertility". Best Prac Res Clin Endocrinol Metab 2015; 29:595-605

Oishi S. (2001) 'Effects of butylparaben on the male reproductive system in rats', *Toxicology and Industrial Health*, 17(1), pp. 31–39. doi: 10.1191/0748233701th093oa.

Oishi S. (2002) 'Effects of butyl paraben on the male reproductive system in mice', *Archives of Toxicology*, 76(7), pp. 423–429. doi: 10.1007/s00204-002-0360-8.

Pengpeng J, Wang X, Chang F, Bai Y, Li Y, Zhou R, Chen L. (2013) Low dose bisphenol A impairs spermatogenesis by suppressing reproductive hormone production and promoting germ cell apoptosis in adult rats. J Biomed Res 2013; 27:135-144.

Philippat C, Botton J, Calafat AM, Ye X, Charles MA, and Slama R. (2014) Prenatal exposure to phenols and growth in boys. Epidemiology. 2014;25(5):625-635.

Pollock T, Weaver RE, Ghasemi R, deCatanzaro D. (2017) 'Butyl paraben and propyl paraben modulate bisphenol A and estradiol concentrations in female and male mice', *Toxicology and Applied Pharmacology*. Elsevier Inc., 325, pp. 18–24. doi: 10.1016/j.taap.2017.04.001.

Pop A, Drugan T, Gutleb AC, Lupu D, Cherfan J, Loghin F, Kiss B. (2016) 'Individual and combined *in vitro* (anti)androgenic effects of certain food additives and cosmetic preservatives', *Toxicology in vitro*. Elsevier Ltd, 32, pp. 269–277. doi: 10.1016/j.tiv.2016.01.012.

Pop A, Drugan T, Gutleb AC, Lupu D, Cherfan J, Loghin F, Kiss B. (2018) 'Estrogenic and anti-estrogenic activity of butylparaben, butylated hydroxyanisole, butylated hydroxytoluene

and propyl gallate and their binary mixtures on two estrogen responsive cell lines (T47D-Kbluc, MCF-7)', *Journal of Applied Toxicology*, 38(7), pp. 944–957. doi: 10.1002/jat.3601.

Pollack AZ, Mumford SL, Krall JR, Carmichael AE, Sjaarda LA, Perkins NJ,Kannan K, Schisterman EF. (2018) Exposure to bisphenol A, chlorophenols, benzophenones, and parabens in relation to reproductive hormones in healthy women: A chemical mixture approach. Environ Int. 2018 Nov;120:137-144.

Prusakiewicz JJ, Harville HM, Zhang Y, Ackermann C, Voorman RL. (2007) 'Parabens inhibit human skin estrogen sulfotransferase activity: Possible link to paraben estrogenic effects', *Toxicology*, 232(3), pp. 248–256. doi: 10.1016/j.tox.2007.01.010.

Riad MA, Abd-Rabo MM, Aziz SAAE, Behairy AME, Badawy MM. (2018) 'Reproductive toxic impact of subchronic treatment with combined butylparaben and triclosan in weanling male rats', *Journal of Biochemical and Molecular Toxicology*, 32(3), p. e22037. doi: 10.1002/jbt.22037.

Routledge EJ, Parker J, Odum J, Ashby J, Sumpter JP. (1998) ´Some alkyl hydroxy benzoate preservatives (parabens) are estrogenic´, *Toxicology and applied pharmacology*, 153(1), pp. 12-19. doi: 10.1006/taap.1998.8544.

Sathyanarayana S, Swan SH, Farin FM, Wilkerson HW, Bamshad M, Grady R, Zhou C, Schwartz SM. (2012) A pilot study of the association between genetic polymorphisms involved in estrogen signaling and infant male genital phenotypes. Asian J. Androl. 14, 766–772.

SCCS (Scientific Committee on Consumer Safety), Opinion on parabens, 03 May 2013.

Schwartz CL, Christiansen S, Vinggaard AM, Axelstad M, Hass U, Svingen T. (2019) Anogenital distance as a toxicological or clinical marker for fetal androgen action and risk for reproductive disorders. Arch Toxicol. 2019 Feb;93(2):253-272. doi: 10.1007/s00204-018-2350-5.

Scott HM, Hutchison GR, Mahood, IK, Hallmark N, Welsh M, De Gendt K, Verhoeven G, O'Shaughnessy P, Sharpe RM. (2007) Role of androgens in fetal testis development and dysgenesis. Endocrinology 2007; 148:2027-2036

Shaw J, de Catanzaro D. (2009) ´Estrogenicity of parabens revisited: impact of parabens on early pregnancy and an uterotrophic assay in mice´, *Reproductive Toxicology*, 28(1), pp. 26-31. doi: 10.1016/j.reprotox.2009.03.003.

Smarr MM, Honda M, Kannan K, Chen Z, Kim S, Louis GMB. (2018) Male urinary biomarkers of antimicrobial exposure and bi-directional associations with semen quality parameters. Reprod Toxicol. 2018 Apr;77:103-108.

Smith KW, Souter I, Dimitriadis I, Ehrlich S, Williams PL, Calafat AM, and Hauser R. (2013) Urinary paraben concentrations and ovarian aging among women from a fertility center. Environ.Health Perspect. 2013;121(11-12):1299-1305.

Steinberger E, Duckett GE. (1965) Effect of estrogen or testosterone on initiation and maintenance of spermatogenesis in the rat. Endocrinology 1965;76:1184-1189.

Tan KA, De gendt K, Atanassova N, Walker M, Sharpe RM, Saunders PT, Denolet E, Verhoeven G. (2005) The role of androgens in sertoli cell proliferation and functional maturation: studies in mice with total or Sertoli cell-selective ablation of the androgen receptor. Endocrinology 2005; 146:2674-2683.

Taxvig C, Vinggaard AM, Hass U, Axelstad M, Boberg J, Hansen PR, Frederiksen H, Nellemann C. (2008) 'Do Parabens Have the Ability to Interfere with Steroidogenesis?', *Toxicological Sciences*, 106(1), pp. 206–213. doi: 10.1093/toxsci/kfn148.

van Meeuwen *et al*. (2008) 'Aromatase inhibiting and combined estrogenic effects of parabens and estrogenic effects of other additives in cosmetics', Toxicology and Applied Pharmacology, 230, pp. 372-382. doi: 10.1016/j.taap.2008.03.002.

Vo TT, Jeung EB. (2009), ´An evaluation of estrogenic activity of parabens using uterine calbindin-d9k gene in an immature rat model´, *Toxicological Sciences*, 112(1), pp. 68-77. doi: 10.1093/toxsci/kfp176.

Vo *et al*. (2010) 'Potential estrogenic effect(s) of parabens at the prepubertal stage of a postnatal female rat model´, Reproductive Toxicology, 29, pp. 306-316. doi: 10.1016/j.reprotox.2010.01.013.

Vo TTB, Jung EM, Choi KC Yu FH, Jeung EB. (2011) 'Estrogen receptor α is involved in the induction of Calbindin-D9k and progesterone receptor by parabens in GH3 cells: A biomarker gene for screening xenoestrogens', *Steroids*. Elsevier Inc., 76(7), pp. 675–681. doi: 10.1016/j.steroids.2011.03.006.

Watanabe Y, Kojima H, Takeuchi S, Uramaru N, Ohta S, Kitamura S. (2013) 'Comparative study on transcriptional activity of 17 parabens mediated by estrogen receptor α and β and androgen receptor', *Food and Chemical Toxicology*. Elsevier Ltd, 57, pp. 227–234. doi: 10.1016/j.fct.2013.03.036.

WHO/IPCS, (2002). World Health Organization, International Programme on Chemical Safety. Global assessment of the state-of-the-science of endocrine disruptors. WHO/PCS/EDC/02.2

Williams GP, Darbre PD. (2019) 'Low-dose environmental endocrine disruptors, increase aromatase activity, estradiol biosynthesis and cell proliferation in human breast cells', *Molecular and Cellular Endocrinology*. Elsevier, 486, pp. 55–64. doi: 10.1016/j.mce.2019.02.016.

Wróbel A, Gregoraszczuk EŁ. (2013) 'Effects of single and repeated *in vitro* exposure of three forms of parabens, methyl-, butyl- and propylparabens on the proliferation and estradiol secretion in MCF-7 and MCF-10A cells', *Pharmacological Reports*, 65(2), pp. 484–493. doi: 10.1016/S1734-1140(13)71024-7.

Wróbel AM, Gregoraszczuk EŁ. (2014a) 'Actions of methyl-, propyl- and butylparaben on estrogen receptor-α and -β and the progesterone receptor in MCF-7 cancer cells and non-cancerous MCF-10A cells', *Toxicology Letters*, 230(3), pp. 375–381. doi: 10.1016/j.toxlet.2014.08.012.

Wróbel AM, Gregoraszczuk EŁ. (2014b) 'Differential effect of methyl-, butyl- and propylparaben and 17β-estradiol on selected cell cycle and apoptosis gene and protein expression in MCF-7 breast cancer cells and MCF-10A non-malignant cells', *Journal of Applied Toxicology*, 34(9), pp. 1041–1050. doi: 10.1002/jat.2978.

Wróbel AM, Gregoraszczuk, EŁ. (2015) 'Action of methyl-, propyl- and butylparaben on GPR30 gene and protein expression, cAMP levels and activation of ERK1/2 and PI3K/Akt signaling pathways in MCF-7 breast cancer cells and MCF-10A non-transformed breast epithelial cells', *Toxicology Letters*, 238(2), pp. 110–116. doi: 10.1016/j.toxlet.2015.08.001.

Yang H, Nguyen TT, An BS, Choi KC, Jeung EB. (2011) 'Synergistic effects of parabens on the induction of calbindin-D 9k gene expression act via a progesterone receptor-mediated pathway in GH3 cells', *Human & Experimental Toxicology*, 31(2), pp. 134–144. doi: 10.1177/0960327111422402.

Zhang GL, Zhang XF, Feng YM, Li L, Huynh E, Sun XF, Sun ZY, Shen W. (2013a) Exposure to bisphenol A results in decline in mouse spermatogenesis. Reprod Feril Dev 2013; 25:847-859.

Zhang L, Dong L, Ding S, Qiao P, Wang C, Zhang M, Zhang L, Du Q, Li Y, Tang N, Chang B. (2014) 'Effects of n-butylparaben on steroidogenesis and spermatogenesis through changed E2 levels in male rat offspring', *Environmental Toxicology and Pharmacology*. Elsevier B.V., 37(2), pp. 705–717. doi: 10.1016/j.etap.2014.01.016.

Zhang L, Ding S, Qiao P, Dong L, Yu M, Wang C, Zhang M, Zhang L, Li Y, Tang N, Chang B. (2016) 'n- butylparaben induces male reproductive disorders via regulation of estradiol and estrogen receptors', *Journal of Applied Toxicology*, 36(9), pp. 1223–1234. doi: 10.1002/jat.3291.

Zhang Z, Sun L, Hu Y, Jiao J, Hu J. (2013b) 'Inverse antagonist activities of parabens on human oestrogen-related receptor γ (ERRγ): *In vitro* and in silico studies', *Toxicology and Applied Pharmacology*. Elsevier Inc., 270(1), pp. 16–22. doi: 10.1016/j.taap.2013.03.030.

Özdemir E, Barlas N, Çetinkaya MA. (2018) Assessing the antiandrogenic properties of propyl paraben using the Hershberger bioassay. Toxicol Res. 2018 Jan 25;7(2):235-243. doi: 10.1039/c7tx00319f.

Annex I - Additional information: Lines of Evidence for Adverse Effects and Endocrine Activity

Methods

The lines of evidence for adverse effects and endocrine activity of butylparaben are presented in the tables below. These tables clarify the available evidence as well as the sources of information. The tables are extracted from information in an Excel-sheet built on the Excel template for reporting the available information relevant for endocrine disrupter (ED) assessment under the Biocides and Plant Protection Products Regulations (Appendix E of ECHA/EFSA 2018).

The information is organised into lines of evidence of a) endocrine activity generated *in vitro* (Table A1-A3) or *in vivo* (Table A4-A5) and b) endocrine-related adverse effects based on *in vivo* endpoints in males (Table A4-A5). To obtain a better overview of *in vivo* data, Table A4 presents studies using perinatal exposures to butylparaben and Table A5 presents studies using pubertal and/or adult exposure to butylparaben. Studies on females are not included here, as the lines of evidence analysis focuses on male reproductive toxicity.

Endocrine-related adverse effects are grouped according to estrogenic, androgenic, thyroidal and steroidogenic (EATS) modalities, based on OECD (2018), and adapting the Joint Research Centre's (JRC) screening methodology to identify potential endocrine disruptors (JRC, 2016). Below, the reporting tables include information on endocrine modalities E, A and S. For each type of *in vitro* assay it is indicated which endpoints relate to which endocrine modality. For *in vivo* assays, it cannot be determined whether endpoints related to modalities E, A or S. Information on T (thyroidal) endpoints *in vivo* or *in vitro* is not included here.

With regards to mechanistic understanding, *in vivo* evidence (including hormone levels) is regarded as stronger evidence of an endocrine disrupting mode of action compared to *in vitro* assays (JRC 2016). This is due to the reasoning that *in vivo* mechanistic effects incorporate absorption, distribution, metabolism and excretion of the chemical, and as *in vivo* mechanistic effects are more closely linked to the manifestation of the adversity (JRC 2016). Mammalian *in vivo* adverse effects can be divided into EATS specific endpoints and non-specific adversity, i.e. endpoints considered sensitive to, but not specific of adversity caused though EATS pathways. A third type of adversity is general adversity including systemic toxicity and effects unrelated to EATS pathways. Among EATS-mediated effects are the endpoints decrease in male AGD and reduction in sperm parameters (sperm number, sperm motility, sperm morphology) (OECD 2018).

Epidemiological data are included as supportive evidence. As described by ECHA/EFSA (2018), any available epidemiological studies should be considered, but cannot be used to override or dismiss evidence of adversity found in laboratory studies, nor can they replace laboratory studies (ECHA/EFSA 2018).

Tables

Table A1: Lines of evidence for estrogen receptor activation by butylparaben

Reference	Grouping	Line of evidence	Species/cell lines	Observed effects (pos/neg)	Assessment of each line of evidence
Lines of evidence for **endocrine activity *in vitro***					
Gonzalez *et al*. 2018	*In vitro* mechanistic	Estrogen receptor transactivation	MCF-7 cells co-transfected with ERE-luciferase reporter construct	Transcriptional activity was induced by estradiol, butylparaben (BP) and the BP metabolite 3-hydroxy *n*-butyl 4-hydroxybenzoate	Sufficient evidence for estrogen receptor activation. Butylparaben (BP) and the BP metabolite 3-hydroxy *n*-butyl 4-hydroxybenzoate showed estrogen receptor agonistic response similar to estrogen. The anti-estrogenic effects seen in the study by Pop *et al*. (2018) differ from other studies but authors write that there is no toxicity of exposure. The hydroxylated isomers of BP (n-butyl-3-hydroxybenzoate and n-butyl-2-hydroxybenzoate) also activated the estrogen receptor, however, to a lower degree than BP.
Pop *et al*. 2018	*In vitro* mechanistic	Estrogen receptor transactivation	T47D-Kbluc	Transcriptional activity was induced (agonist mode, no estradiol present) in the dose-range 0.3-10 µM, whereas an anti-estrogenic response was observed in the dose-range 60-100 µM. Cell viability assay showed no toxicity.	
Pop *et al*. 2018	*In vitro* mechanistic	Estrogen receptor transactivation	T47D-Kbluc	Transcriptional activity was reduced (antagonist mode, estradiol present in 30 pM) at approximately 60 µM. Cell viability assay showed no toxicity.	
Watanabe *et al*. 2013	*In vitro* mechanistic	Estrogen receptor transactivation	Transiently transfected CHO cells (Erα and Erβ)	ERα and ERβ agonistic activity corresponding to that of E2 after BP exposure. The activity of the hydroxylated isomers of BP (n-butyl-3-hydroxybenzoate and n-butyl-2-hydroxybenzoate) showed lower estrogenic activities than BP	
Khanna & Darbre 2013	*In vitro* mechanistic	Estrogenic response in target cells	MCF-10A	MCF-10A growth is estrogen dependent. BP stimulated proliferation and increase in colonies in a similar manner as estradiol at all doses tested	Supporting data of estrogen mode of action. Proliferation of MCF-10A, MCF-7 as well as ZR-75-1 is estrogen dependent. Exposure to BP induces proliferation in the studies by Khanna & Darbre (2013), Charles & Darbre (2013), Gonzalez *et al*. (2018), Pop *et al*. (2018), Van Meeuwen *et al*. 2008 and Williams *et al*. (2019)
Charles & Darbre 2013	*In vitro* mechanistic	Estrogenic response in target cells	MCF-7	Proliferation in the presence of BP after 7 and 14 days of exposure.	
Pop *et al*. 2018	*In vitro* mechanistic	Estrogenic response in target cells	MCF-7	Proliferation increased in the dose-range 0.1-10 µM, whereas a reduced relative proliferation was seen in the interval 10-60 µM. Cell viability assay showed no toxicity.	
Pop *et al*. 2018	*In vitro* mechanistic	Estrogenic response in target cells	MCF-7	An anti-estrogenic effect (antagonist mode, estradiol present in 10 pM) on relative proliferation was seen in the range 0.3 – 200 µM. Cell viability assay showed no toxicity.	

Gonzalez *et al*. 2018	*In vitro* mechanistic	Estrogen receptor binding	MCF-7, T47D, MDA-MB-231	3-hydroxy *n*-butyl 4-hydroxybenzoate, a BP metabolite, induced proliferation of the estrogen dependent MCF-7 cell line in a concentration related manner. Co-treatment with the anti-estrogen ICI 182, 780 inhibited proliferation after both BP and 3-hydroxy *n*-butyl 4-hydroxybenzoate exposure, showing that the induced proliferation was estrogen receptor mediated. This was further supported by a study in MDA-MB-231, a non-ERα expressing cell line, where neither BP nor its metabolite 3-hydroxy *n*-butyl 4-hydroxybenzoate induced proliferation. In T47D cells proliferation was seen only at the highest dose. In addition, mRNA levels of GREB1, a critical down stream target of ERα signaling, was induced by BP and its metabolite 3-hydroxy *n*-butyl 4-hydroxybenzoate in manner similar to estradiol. Treatment with the anti-estrogen ICI 182, 780 blocked the induction	similar to the response of estradiol. Furthermore, addition of anti-estrogen (Gonzalez *et al*, 2018, Williams *et al*, 2019) inhibited cell proliferation as well as expression of downstream ERα target GREB1. Together these data indicated activation of estrogen pathway via estrogen receptor. The anti-estrogenic effects seen at high doses in the study by Pop *et al*. (2018) differ from the other studies, but authors write that there is no toxicity of exposure.
Williams *et al*. 2019	*In vitro* mechanistic	Estrogenic response in target cells	MCF-7, ZR-75-1, HMF3A	MCF-7 and ZR-75-1 are estrogen dependent for proliferation and exposure to BP increased proliferation. Co-treatment with the aromatase inhibitor letrozole reduced the overall cell proliferation and addition of the anti-estrogen ICI 182, 780 totally suppressed cell proliferation. HMF3A (ERα negative, ERβ positive) was not affected by exposure.	
Van Meeuwen *et al*. 2008	*In vitro* mechanistic	Estrogenic response in target cells	MCF-7	Increased proliferation in the presence of BP after 6 days of exposure. Decreased proliferation with co-exposure to anti-androgen, indicating ER dependent effect.	
Goswami & Kalita 2016	*In vivo* mechanistic	Uterus weight & histopath (uterotrophic assay)	Mouse, 7 d, Subcutaneous injection (s.c.)	Increased number of uterine glands, increased uterine weight and histological alterations, as well as increased endometrial and mycometrium thickness and total tissue protein. Estradiol (positive control) stimulated uterus at all endpoints included. No Lowest observed effect level (LOEL), as all doses of butylparaben caused adverse effect.	Data supports an estrogen mode of action. The weight of the uterus in immature and ovariectomised (OVX) rodents increases as a response to estradiol via an ER-responsive pathway. In the uterotrophic assay, immature or OVX rodents were used to investigate if butylparaben can increase uterine weight. In the studies by Routledge *et al*. (1998), Hossaini *et al*.
Guerra *et al*. 2017a	*In vivo* mechanistic	Uterus weight (uterotrophic assay)	Rat, 3 d, s.c.	Butylparaben did not affect uterus weight. Estradiol (positive control) increased uterus weight.	
Routledge *et al*. 1998	*In vivo* mechanistic	Uterus weight (uterotrophic assay)	Rat, 3 d, s.c./oral	Increased uterus weight after s.c. exposure. Oral exposure did not affect uterine weight. Estradiol (positive control) increased uterus weight.	
Hossaini *et al*. 2000	*In vivo* mechanistic	Uterus weight (uterotrophic	Rat/mouse, 3 d, s.c.	No effect in mice (only one dose tested). In rats, increased uterus weight was recorded after exposure. Estradiol	

		assay)		(positive control) increased uterus weight.	(2000), Lemini *et al*. (2003), Lemini *et al*. (2004), Goswami & Kalita (2016), Vo & Jeung (2009) exposure to butylparaben increased the uterine weight, whereas two studies (Shaw & deCatanzaro, 2009; Guerra *et al*, 2017a) reported no effect. One study showed no effect after 21 days of exposure, but changes in uterine histology (Vo *et al*. 2010). Additionally, co-exposure with fulvestrant (anti-estrogen; Vo & Jeung, 2009) removed the stimulating effect of butylparaben. Together these data indicate activation of the estrogen pathway via estrogen receptors, in a similar manner to estradiol.
Lemini *et al*. 2003	*In vivo* mechanistic	Uterus weight (uterotrophic assay)	Rat/mouse, 3 d, s.c.	Increased uterus weight was recorded in both immature rats and mice as well as in adult OVX mice. Estradiol (positive control) increased uterus weight.	
Lemini *et al*. 2004	*In vivo* mechanistic	Uterus weight (uterotrophic assay)	Mouse, 3 d, s.c.	Increased uterus weight after exposure. Estradiol (positive control) increased uterus weight.	
Shaw & deCatanzaro 2009	*In vivo* mechanistic	Uterus weight (uterotrophic assay)	Mouse, 3 d, s.c.	No effect on uterus weight. Estradiol (positive control) increased uterus weight.	
Vo & Jeung 2009	*In vivo* mechanistic	Uterus weight (uterotrophic assay)	Rat, 3 d, s.c.	Increased uterus weight after exposure. After co-exposure with fulvestrant (anti-estrogen) the stimulating effect of butylparaben disappeared. Estradiol (positive control) increased uterus weight.	
Vo *et al*. 2010	*In vivo* mechanistic	Uterus weight	Rat, 21 days	No effect on relative uterus weight after 21 days exposure, but increased thickness of endometrium, which may be a more relevant endpoint than uterus weight as the rat passes puberty. Decrease in numbers of corpora lutea and no effect on time of vaginal opening, estrous cyclicity or relative ovary weight.	

Table A2: Lines of evidence for altered steroidogenesis by butylparaben.

Reference	Grouping	Line of evidence	Species/cell lines	Observed effects (pos/neg)	Assessment of each line of evidence
Lines of evidence for **endocrine activity *in vitro***					
Taxvig *et al*. 2008	*In vitro* mechanistic	Hormone level	H295R	Progesterone levels were increased. No effect on estradiol and testosterone levels. Hormones were measured after 48 h exposure	Supporting evidence of altered steroidogenesis. Data supporting estrogen promoting activities of BP was seen in the study by Williams *et al*. (2019) where estradiol levels were increased, Cytochrome P450 (CYP)19A1 gene expression was upregulated, and aromatase enzyme activity increased after BP exposure. However, there are also studies supporting no estrogen promoting activities (Taxvig *et al*, 2008; Wróbel & Gregoraszczuk, 2013; Guerra *et al*, 2016). A discrepancy was noted in the effects on the precursor progesterone as the study by Taxvig *et al*. (2008) sees an increase, whereas the study by Guerra *et al*. (2016) does not see an effect. It should be kept in mind that different types of cell lines are used, and a clear pattern is difficult to deduce – the response may be cell type specific.
Wróbel & Gregoraszczuk 2013	*In vitro* mechanistic	Hormone level	MCF-7 and MCF-10A	BP exposure stimulated estradiol secretion from MCF-7 cells at the lowest concentration. BP exposure reduced estradiol secretion at all doses tested in MCF-10A cells	
Guerra *et al*. 2016	*In vitro* mechanistic	Hormone level	Primary culture of pre-antral follicles (mouse)	Estradiol levels measured after 8 and 12 days of BP exposure. No effect seen compared to control group	
Guerra *et al*. 2016	*In vitro* mechanistic	Hormone level	Primary human granulosa cells	Progesterone levels (precursor product for estradiol synthesis) not affected after 24, 48, 72 or 96 hours	
Williams *et al*. 2019	*In vitro* mechanistic	Hormone level and enzyme activity	MCF-7, ZR-75-1 breast cancer cells and HMF3A breast fibroblasts	BP exposure increased estradiol levels in all examined cell lines. CYP19A1 gene expression was upregulated and aromatase enzyme activity increased after BP exposure in all examined cell lines. Co-treatment with the aromatase inhibitor letrozole inhibited all aromatase enzyme activity.	
van Meeuwen *et al*. 2008	*In vitro* mechanistic	Enzyme activity	Microsomes from human placenta	Aromatase activity in human placental microsomes was reduced after 45 minutes exposure.	

Table A3: Lines of evidence for androgen receptor antagonism by butylparaben.

Reference	Grouping	Line of evidence	Species/cell lines	Observed effects (pos/neg)	Assessment of each line of evidence (e.g. supporting data, sufficient evidence)
Lines of evidence for **endocrine activity *in vitro*.**					
Kjærstad *et al*. 2010	*In vitro* mechanistic	Androgen receptor transactivation	AR-transfected Chinese Hamster Ovary cells	No AR-antagonistic effect	Studies on anti-androgenic activity showed inconsistent findings. Effects were observed in some but not all studies, possibly due to different study designs. Studies on androgen activity showed sufficient evidence for BP not having androgen activity.
Gonzalez *et al*. 2018	*In vitro* mechanistic	Androgenic response in target cells	LNCaP	3-hydroxy *n*-butyl 4-hydroxybenzoate, a BP metabolite, did not induce proliferation	
Watanabe *et al*. 2013	*In vitro* mechanistic	Androgen receptor transactivation	Transiently transfected CHO cells (AR)	No AR-antagonistic effect	
Chen *et al*. 2007	*In vitro* mechanistic	Androgen receptor transactivation	Stably transfected HEK 293 cells (AR) (2933Y cells)	Anti-androgenic activity at highest concentration (10 µM). No androgenic activity.	
Pop *et al*. 2016	*In vitro* mechanistic	Androgen receptor transactivation	Transfected MDA-kb2 human breast cancer cells (ATCC CRL-2713)	Anti-androgenic activity at three highest doses (approximately in the interval 50-100 µM, read from graph). IC50 = 58.5 µM. No androgenic activity	

Table A4: Lines of evidence for endocrine activity of butylparaben leading to adverse effect in male offspring following perinatal exposure.

Reference	Grouping	Line of evidence	Species	Exposure period	Route of exposure	Effect dose (mg/kg bw/day)	Observed effects (pos/neg)	Assessment of each line of evidence
Lines of evidence for **endocrine activity with perinatal** exposure. All listed endpoints relate to the modality EAS.								
Zhang *et al*. 2014	*In vivo* mechanistic	Estradiol level	Rat	GD 7 - PND 21	Oral	400	Increase. Measured in male offspring on PND 21, 35, 49, 90, 180. PND 21: E2 increased in two highest dose groups (400 and 1000 mg/kg bw/day). Increased in highest dose group (1000 mg/kg bw/ day) on PND 35, 49 and 90. No effect on PND 180. Absolute (pg/ml)	Supporting evidence of endocrine activity. E2 levels increased at high doses in prepuberty and in young adults in two studies. Not changed later in adulthood in one study, and not changed in prepuberty in another study. In a fourth study, significant decrease was seen in prepuberty, whereas increase was seen post puberty. In adult animals, decrease was seen, but not in a dose-dependent manner. Some consistency between two studies with different exposure periods and dose levels.
Zhang *et al*. 2016	*In vivo* mechanistic	Estradiol level	Rat	GD 7 - PND 21	Oral	400	Increase. Measured in male offspring PND 21 and 90. PND21: increased in 400 and 1000 mg/kg bw/day groups. PND90: increased in 1000 mg/kg bw/day group. , Absolute (pg/ml))	
Boberg *et al*. 2016	*In vivo* mechanistic	Estradiol level	Rat	GD 7 - PD 22	Oral	-	No effect. Serum estradiol levels measured in male offspring at pup day (PD)16 when effects on gene expression of Cyp19a1 (aromatase) was seen.	
Maske *et al*. 2020	*In vivo* mechanistic	Estradiol level	Rat	GD 6 – PND 21	Direct (s.c.)	10	Change. Estradiol measured in male offspring on PND 30, 45 and 75. PND 30: Decrease at 10, 100 and 1000 mg/kg bw/day. PND 45: Increase at 1000 mg/kg bw/day PND 75: Decrease at 10 mg/kg bw/day, but not at higher doses.	

Taxvig *et al*. 2008	*In vivo* mechanistic	Testosterone level	Rat	GD7-GD21	Direct (s.c.)	-	No effects on testosterone at GD 21 in male foetuses at doses of 200 and 400 mg/kg bw/day (measured in both plasma and testes), Absolute (nM, ng/testis) and testosterone production ex vivo (ng/testis/5 hours)	Supporting evidence of endocrine activity. Testosterone levels reduced at selected ages and dose groups (high doses) in two studies, but increased in two other studies. No effect on foetal testosterone levels. Inconsistency of data may be due to different exposure periods, routes and dose levels.
Zhang *et al*. 2014	*In vivo* mechanistic	Testosterone level	Rat	GD 7 - PND 21	Oral	400	Decrease. Measured in male offspring on PND 21, 35, 49, 90, 180. Testosterone levels decreased at all ages in high dose group (1000 mg/kg bw/day). At PND 35 and 49, 400 mg/kg bw/day exposure also decreased testosterone levels. Absolute (ng/ml)	
Zhang *et al*. 2016	*In vivo* mechanistic	Testosterone level	Rat	GD 7 - PND 21	Oral	1000	Decrease. Measured in male offspring PND 21 and 90. Decreased in 1000 mg/kg bw/day at both ages. The reduction was approximately 36% on PND21. The reduction seemed dose dependent on PND90. Absolute (ng/ml)	
Guerra *et al*. 2017b	*In vivo* mechanistic	Testosterone level	Rat	GD 12 - PND 21	Direct (s.c.)	200	Increase. Measured in male offspring on PND 110. Serum levels of testosterone increased in high dose group (200 mg/kg bw/day). Absolute (pg/ml)	
Maske *et al*. 2020	*In vivo* mechanistic	Testosterone level	Rat	GD 6- PND 21	Direct (s.c.)	100	Change. Measured in male offspring on PND 30, 45 and 75. PND 30: no effect PND 45: increase at 100 and 1000 mg/kg bw/day PND 75: decrease at 100 (small effect) and increase at 1000 mg/kg bw/day	

Zhang *et al*. 2014	*In vivo* mechanistic	Luteinising Hormone (LH) level	Rat	GD 7 - PND 21	Oral	400	Measured in male offspring on PND 21, 35, 49, 90, 180. PND 21 and 49: decreased in two highest dose groups (400 and 1000 mg/kg bw/day). PND 35: decreased in highest dose group (1000 mg/kg bw/day). PND 90 and 180: increased in highest dose group (1000 mg/kg bw/day). Absolute (MIU/ml)	Supporting evidence of endocrine activity. LH levels decreased at high doses around the time of puberty, but increased in adulthood. LH levels decreased in adulthood in another study. Inconsistency in the data may be due to different exposure periods, routes and dose levels.
Guerra *et al*. 2017b	*In vivo* mechanistic	Luteinising Hormone (LH) level	Rat	GD 12 - PND 21	Direct (s.c.)	200	Measured in male offspring on PND 110. Serum levels of LH decreased in high dose group (200 mg/kg bw/day). Absolute (pg/ml)	
Maske *et al*. 2020	*In vivo* mechanistic	Luteinising Hormone (LH) level	Rat	GD 6- PND 21	Direct (s.c.)	1000	Measured in male offspring on PND 45 and 75. Increased in 1000 mg/kg bw/day group at both ages.	
Zhang *et al*. 2014	*In vivo* mechanistic	Follicle Stimulating Hormone (FSH) level	Rat	GD 7 - PND 21	Oral	400	Measured in male offspring on PND 21, 35, 49, 90, 180. PND 35: Decreased in the two highest dose groups (400 and 1000 mg/kg bw/day). PND 90: Increased levels in the highest dose group (1000 mg/kg bw/day). Absolute (MIU/ml)	Supporting evidence of endocrine activity. FSH levels decreased at high doses around the time of puberty, but increased in adulthood. FSH levels decreased in adulthood in another study. Inconsistency in the data may be due to different exposure periods, routes and dose levels.
Guerra *et al*. 2017b	*In vivo* mechanistic	Follicle Stimulating Hormone (FSH) level	Rat	GD 12 - PND 21	Direct (s.c.)	200	Measured in male offspring on PND 110. Serum levels of FSH decreased in high dose group (200 mg/kg bw/day). Absolute (ng/ml)	
Lines of evidence for **adverse effect with perinatal exposure**. All listed endpoints relate to the modality EAS.								
Kang *et al*. 2002	EATS mediated	Anogenital distance	Rat	GD6- PND20	Direct (s.c.)	-	No effect at PND 1 at doses of 100 and 200 mg/kg bw/day	Supporting evidence of endocrine activity and adverse effects. Inconsistency between the studies may be due to different exposure periods, dose levels and
Taxvig *et al*. 2008	EATS mediated	Anogenital distance	Rat	GD7- GD21	Direct (s.c.)	-	No effect at GD 21 at doses of 200 and 400 mg/kg bw/day	
Zhang *et*	EATS	Anogenital	Rat	GD 7 –	Oral	400	Reduced at PND 1 and 21 at	

al. 2014	mediated	distance		PND 21			400 and 1000 mg/kg bw/day and not at 64 or 160 mg/kg bw/day	measuring sensitivity. Effects were seen in two studies evaluating AGD at PND 1 and using doses above 200 mg/kg bw/day.
Boberg *et al*. 2016	EATS mediated	Anogenital distance	Rat	GD 7 – PD 22	Oral	100	Reduced of AGD and AGDi at PND 1 at 100 and 500 mg/kg bw/day	
Guerra *et al*. 2017b	EATS mediated	Anogenital distance	Rat	GD 12 – PND 21	Direct (s.c.)	-	No effect at PND 1 at doses of 10, 100, 200 mg/kg bw/day.	
Fisher *et al*. 1999	EATS-mediated	Testis histopathology	Rat	PND 2-18	Direct (s.c.)	-	No effect at PND 18. Only one low dose (2 mg/kg bw/day) and evaluation of aquaporins and rete testis morphology only.	Supportive evidence of adverse effect. No effects on foetal testis in two studies. Signs of histological effects on seminiferous tubules of prepubertal testes in one study. Change in adult testes in four out of four studies, but different endpoints examined.
Kang *et al*. 2002	EATS-mediated	Testis histopathology	Rat	GD6-PND20	Direct (s.c.)	100	Decrease. The numbers of spermatogonia, spermatocytes, round spermatids and elongated spermatids were investigated in seminifereous tubules. Round and elongated spermatids were reduced in both exposure groups (100 and 200 mg/kg bw/day).	
Taxvig *et al*. 2008	EATS-mediated	Testis histopathology	Rat	GD7-GD21	Direct (s.c.)	-	No effect on foetal testis histopathology at GD 21.	
Zhang *et al*. 2014	EATS-mediated	Testis histopathology	Rat	GD 7 – PND 21	Oral	400	Change. PND 21: reduced and loosely arranged germ cells, reduced layers of seminiferous tubules in the two highest dose groups (400 and 1000 mg/kg bw/day). No obvious effects on Leydig cells. PND 90: Expanded lumens of the seminiferous tubules, reduced layer of seminiferous tubules, and reduced number of spermatocyte cells in the two highest dose groups (400 and 1000 mg/kg	

							bw/day)	
Zhang *et al*. 2016	EATS-mediated	Testis histopathology	Rat	GD 7 – PND 21	Oral	400	Change. As above, same study as Zhang *et al*. (2014).	
Boberg *et al*. 2016	EATS-mediated	Testis histopathology	Rat	GD 7 – PD 22	Oral	-	No effect. Examined at PD 90, control and high dose, no morphometrical analyses.	
Guerra *et al*. 2017b	EATS-mediated	Testis histopathology	Rat	GD 12 – PND 21	Direct (s.c.)	-	No effect at GD 20 on number of foetal Leydig cells, diameter of seminiferous cords, number of gonocytes/cord	
Guerra *et al*. 2017b	EATS-mediated	Testis histopathology	Rat	GD 12 – PND 21	Direct (s.c.)	100	Increase in number of Leydig cells in interstitium of adult testes at PND 110 in two highest dose groups (100 and 200 mg/kg bw/day). No effect on testicular morphometry PND 110 (number of Sertoli cells, nuclear Leydig cell volume, nuclear Leydic cell area)	
Guerra *et al*. 2017b	EATS-mediated	Testis histopathology (molecular markers)	Rat	GD 12 – PND 21	Direct (s.c.)	200	Decrease in IHC staining intensity on estrogen receptor(ESR)1 and AR in adult: reduced ESR1 in elongated spermatids in stage I – VI and rounded spermatids on stage VII – VIII and reduced AR in Sertoli cell nuclei in stages VII – VIII in high dose group (200 mg/kg bw/day).	
Guerra *et al*. 2017b	EATS-mediated	Testis histopathology	Rat	GD 12 – PND 21	Direct (s.c.)	10	Change. Spermatogenesis kinetics PND 110. Percentage of seminiferous tubules in stage I – VI decreased in high dose group (200 mg/kg bw/day) and stages VII-VIII increased in low and high dose (10 and 200 mg/kg bw/day).	

Maske *et al*. 2020	*In vivo* mechanistic	Testis histopathology	Rat	GD 6-PND 21	Direct (s.c.)	10	Change. Seminiferous tubules and germinal layers were evaluated on PND 30, 45 and 75. Effects were seen on tubules (degenerative) and the germ layer (arrangement) on PND 30 (10, 100 and 1000 mg/kg bw) and on PND 45 and 75 degenerative changes on tubules and reduced spermatogenesis was seen at 10, 100 and 1000 mg/kg bw	
Fisher *et al*. 1999	EATS-mediated	Testis weight	Rat	PND 2-18	Direct (s.c.)	-	No effect. Absolute weight, one low dose of 2 mg/kg bw/day.	No clear pattern of effects on testis weight in prepuberty or adulthood.
Kang *et al*. 2002	EATS-mediated	Testis weight	Rat	GD6-PND20	Direct (s.c.)	100	Change. Investigated at PND 21, 49, 70 and 90. 100 mg/kg bw/day led to increase on PND 21 and decrease on PND 49. 200 mg/kg bw/day led to increase on PND 90.	
Zhang *et al*. 2014	EATS-mediated	Testis weight	Rat	GD 7 – PND 21	Oral	400	Decrease. Absolute weights measured PND 21, 35, 49, 90, 180. Reduced in the two highest dose groups (400 and 1000 mg/kg bw/day) on PND 21, 35, and 49. BW also affected at these ages. On PND 90 affected in the three highest dose groups (160, 400 and 1000 mg/kg bw/day), no effect on BW at this age. No effect on testis weight on PND 180.	
Zhang *et al*. 2016	EATS-mediated	Testis weight	Rat	GD 7 – PND 21	Oral	-	No effect. Relative weights PND 21 and 90. Same study as Zhang *et al*. (2014), but the study by Zhang *et al*. (2014) is possibly a subgroup of animals from	

							Zhang *et al*. (2016), therefore different effects (BW at PD 90 affected in the study by Zhang *et al*. (2016) but not Zhang *et al*. (2014)).	
Boberg *et al*. 2016	EATS-mediated	Testis weight	Rat	GD 7 – PD 22	Oral	-	No effect. Testis weighed in offspring on PD16, 22, and 80 – 90. Absolute weights analysed using body weight as covariate in ANOVA.	
Guerra *et al*. 2017b	EATS-mediated	Testis weight	Rat	GD 12 – PND 21	Direct (s.c.)	-	No effect. Absolute weight PND 110	
Maske *et al*. 2020	*In vivo* mechanistic	Testis weight	Rat	GD 6-PND 21	Direct (s.c.)	-	No effect. Testis weighed in male offspring on PND 30, 45 and 75.	
Kang *et al*. 2002	EATS-mediated	Sperm numbers	Rat	GD6-PND20	Direct (s.c.)	100	Decrease. Sperm count in caudal epididymis. Decreased to 50% of control in both exposure groups (100 and 200 mg/kg bw/day).	Sufficient evidence for adverse effect. One study (Guerra *et al*, 2017b) did not show effect at doses up to 200 mg/kg bw/day and another study (Zhang *et al*, 2014) showed effect only at the highest doses of 400 and 1000 mg/kg bw/day. In Maske *et al*. (2020) effect was statistically significant in the middle dose (100 mg/kg bw/day), but the high dose (1000 mg/kg bw/day) is likely affected as the average value is approximately 80% of control and the variance in the control group is rather big. Effects are considered serious and irreversible.
Zhang *et al*. 2014	EATS-mediated	Sperm numbers	Rat	GD 7 – PND 21	Oral	400	Decrease. Numbers of sperm in cauda epididymis and daily sperm production on PND 90 reduced in two high dose groups (400 and 1000 mg/kg bw/day, no effect at 64 or 160 mg/kg bw/day).	
Boberg *et al*. 2016	EATS-mediated	Sperm numbers	Rat	GD 7 – PD 22	Oral	10	Decrease. Number of sperm in cauda epididymis measured on PD 90. Reduced in all dose groups (10, 100, 500 mg/kg bw/day).	
Guerra *et al*. 2017b	EATS-mediated	Sperm numbers	Rat	GD 12 – PND 21	Direct (s.c.)	-	No effect. Sperm counts on PND 110 (testis and epididymis) at doses of 10, 100 or 200 mg/kg bw/day	
Maske *et al*. 2020	*In vivo* mechanistic	Sperm numbers	Rat	GD 6-PND 21	Direct (s.c.)	100	Decrease. Sperm count was investigated on PND 75 and a decrease was seen at 100	

							mg/kg bw/day (67% of control), but not at 1000 mg/kg bw/day (however, the average value in this group is approximately 80% of control and the variance in the control group is rather big)	
Kang *et al*. 2002	EATS-mediated	Sperm motility	Rat	GD6-PND20	Direct (s.c.)	100	Decrease. Sperm motile activity (%) was decreased in both exposure groups of 100 and 200 mg/kg bw/day.	Supporting evidence for adverse effect. Reduced sperm motility in three studies but lack of dose-response and different methods applied.
Guerra *et al*. 2017b	EATS-mediated	Sperm motility	Rat	GD 12 – PND 21	Direct (s.c.)	10	Decrease. Motile sperm with progressive trajectory (%) at PND 110: decreased in low dose only (10 mg/kg bw/day) while slight decreases at 100 and 200 mg/kg bw/day were not statistically significant. Motile sperm with non-progressive trajectory (%) at PND 110: increased in low dose group (10 mg/kg bw/day) while slight increases at 100 and 200 mg/kg bw/day were not statistically significant. No change in % non-motile sperm at PND 110.	
Maske *et al*. 2020	*In vivo* mechanistic	Sperm motility	Rat	GD 6-PND 21	Direct (s.c.)	100	Decrease. Sperm motility was investigated on PND 75 and a decrease was seen at 100 mg/kg bw/day, but not in the the high dose group (1000 mg/kg bw/day)	

Guerra *et al*. 2017b	EATS-mediated	Sperm morphology	Rat	GD 12 – PND 21	Direct (s.c.)	10	Decrease. Normal morphology (%) at PND 110: decrease in all dose groups (10, 100, 200 mg/kg bw/day). Abnormal head (characteristic curvature missing) (%) at PND 110: increase in all dose groups (10, 100, 200 mg/kg bw/day).	Sufficient evidence for adverse effect. Reduced percentage of normal sperm in one study; same effect size at all doses. Effects are considered to be serious and irreversible.
Guerra *et al*. 2017b	Sensitive to, but not diagnostic of, EATS	Fertility (mammals)	Rat	GD 12 – PND 21	Direct (s.c.)	-	No effect on fertility with natural mating (age not clear) or in utero artificial insemination (PND 110).	Some effects on fertility in rats. Not considered evidence of low severity of other adverse effects (sperm count, motility and morphology)
Maske *et al*. 2020	*In vivo* mechanistic	Fertility	Rat	GD 6-PND 21	Direct (s.c.)		Change. Naïve females were sired by exposed male offspring. Decrease in mean no. of implantation sites (100 mg/kg bw/day). Increase in % pre-implantation loss (PIL) in 100 and 1000 mg/kg bw/day as well as increase in % post-implanation loss (POL) at 1000 mg/kg bw/day. No effect on copulation, time taken for copulation, no. of copulated females showing resorptions, or mean no. of CL	
Integrated assessment of lines of evidence for reduced sperm count and quality in male offspring: There is sufficient evidence of adverse effects (decreased sperm count and decreased number of normal sperm). There is supporting evidence of endocrine activity *in vivo* (altered hormone levels, altered anogenital distance, decreased sperm motility and altered testicular histopathology).								

Table A5: Lines of evidence for endocrine activity of butylparaben leading to adverse effect with pubertal/adult exposure.

Reference	Grouping	Line of evidence	Species/cell lines	Exposure (weeks, starting age)	Route of exposure	Effect dose (mg/kg bw/day)	Observed effects (pos/neg)	Assessment of each line of evidence
Lines of evidence for **endocrine activity with pubertal and/or adult exposure**. All listed endpoints relate to the modality EAS.								
Pollock *et al*. 2017	*In vivo* mechanistic	Estradiol level	Mouse	Single injection	Direct (s.c.)	79.5	Increased urinary estradiol level at single dose of 79.5 BP injection after 8 h in males (and at several time points in females). Injected with BP and assessed over 12 hours. Absolute (ng E2/ml urine) and Relative (ng E2/mg creatinine)	Supporting evidence of endocrine activity. Different measures as one study evaluated serum estradiol and the other evaluated urinary estradiol.
Riad *et al*. 2018	*In vivo* mechanistic	Estradiol level	Rat	8 (starting PND 19-22)	Oral	50	Increase in serum estradiol. Absolute (pg/ml) at the end of dosing. One dose only.	
Oishi 2001	*In vivo* mechanistic	Testosterone level	Rat	8 (starting PND 19-21)	Oral	103	Decrease in testosterone in the 103 and 1026 mg/kg bw/day exposure groups. Testosterone concentration in the 1026 mg/kg bw/day group was 34% of the control value. Measured as ng T/ml serum	Supporting evidence of endocrine activity. All four studies showed decreased testosterone levels, but at different doses and ages. The largest study (Hoberman *et al*, 2008) only showed effects at one age.
Oishi 2002	*In vivo* mechanistic	Testosterone level	Mouse	10 (starting PND 27-29)	Oral	1504	Decrease in highest dose group. Measured as ng T/ml serum	
Hoberman *et al*. 2008	*In vivo* mechanistic	Testosterone level	Rat	8 (starting PND 22)	Oral	109.3	Decrease in testosterone level at week 3, increase at week 9, but no change at week 5 and 7. Week 3: Reduced in two highest doses (109.3 and 1087.6), to almost 50% of control in highest dose. Week 9: Increased in highest dose (1087.6 mg/kg bw/day). No effects week 5 and 7. Absolute (ng/ml)	
Riad *et al*. 2018	*In vivo* mechanistic	Testosterone level	Rat	8 (starting PND 19-22)	Oral	50	Decrease in testosterone to approximately 50% of control (serum). Absolute (ng/ml) at the end of dosing. One dose only.	
Hoberman	*In vivo*	Follicle	Rat	8 (starting	Oral	1087.6	Increase in FSH level measured in	Supporting

et al. 2008	mechanistic	Stimulating Hormone (FSH) level		PND 22)			week 3, 5, 7 and 9. Week 9: Increased in highest dose (1087.6 mg/kg bw/day). No effects week 3, 5, and 7.	evidence of endocrine activity. Inconsistent findings between studies, but very different doses applied.
Riad *et al*. 2018	*In vivo* mechanistic	Follicle Stimulating Hormone (FSH) level	Rat	8 (starting PND 19-22)	Oral	50	Decreased serum FSH, Activity measure (IU/l) at the end of dosing. One dose only.	
Hoberman *et al*. 2008	*In vivo* mechanistic	Luteinising Hormone (LH) level	Rat	8 (starting PND 22)	Oral	10.9	Decrease in LH level at week 5: Reduced in low and high dose groups (10.9 and 1087.6 mg/kg bw/day). No effects week 3, 7, and 9.	Supporting evidence of endocrine activity, but effect only seen at one age in study Hoberman *et al*. (2008).
Riad *et al*. 2018	*In vivo* mechanistic	Luteinising Hormone (LH) level	Rat	8 (starting PND 19-22)	Oral	50	Decrease serum LH at the end of dosing. One dose only.	
Lines of evidence for **adverse effect of butylparaben with pubertal and/or adult exposure**. All listed endpoints relate to the modality EAS.								
Oishi 2002	EATS-mediated	Testis histopathology	Mouse	10 (starting PND 27-29)	Oral	14.4	Change in appearance of spermatogenic stages (I-VI, VII-VIII, IX-XII). Stage I-VI increased in all groups from 14.4, stage VII-VIII decreased in middle and highest exposure group (146, 1504 mg/kg bw/day), Relative (percentage of total numbers).	Supporting evidence of adverse effects. Different parameters assessed in all studies and lack of histological changes in the largest study (Hoberman *et al*, 2008).
Oishi 2002	EATS-mediated	Testis histopathology	Mouse	10 (starting PND 27-29)	Oral	14.4	Change in numbers of spermatids. The numbers of spermatogonia, spermatocytes, round spermatids and elongated spermatids were investigated. Elongated spermatids decreased in in all dose groups. Round spermatids increased in highest dose group	
Hoberman *et al*. 2008	EATS-mediated	Testis histopathology	Rat	8 (starting PND 22)	Oral	-	No effects at doses up to 1087.6 mg/kg bw/day. Expert judgement	
Alam & Kurohmaru 2014	EATS-mediated	Testis histopathology	Rat	Single administration (PND 21)	Oral	1000	Change in immunohistochemical markers. Vimentin filaments: affected 6 and 24 h after exposure. Actin and alpha tubulin: No effect.	

Alam & Kurohmaru 2014	EATS-mediated	Testis histopathology	Rat	Single administration (PND 21)	Oral	1000	Detachment and displacement of spermatogenic cells from Sertoli cells.	
Riad *et al*. 2018	EATS-mediated	Testis histopathology	Rat	8 (starting PND 19-22)	Oral	50	Change. Reduced Leydig cell population, subcapsular blood vessels as well as interstitial vasculature was dilated and congested. Seminiferous tubules showed detachment of the spermatogenic lineage from their basement membrane as well as spermatogenic arrest., Expert judgement)	
Oishi 2001	EATS-mediated	Testis weight	Rat	8 (starting PND 19-21)	Oral	-	No effects on absolute and relative weight	Consistent data showing no effect on testis weight
Oishi 2002	EATS-mediated	Testis weight	Mouse	10 (starting PND 27-29)	Oral	-	No effects on absolute and relative weight	
Hoberman *et al*. 2008	EATS-mediated	Testis weight	Rat	8 (starting PND 22)	Oral	-	No effects on absolute weight	
Riad *et al*. 2018	EATS-mediated	Testis weight	Rat	8 (starting PND 19-22)	Oral	-	No effects on relative weight	
Oishi 2001	EATS-mediated	Sperm numbers	Rat	8 (starting PND 19-21)	Oral	10.4	Decrease in cauda epidydidmis sperm counts (counts/cauda) in all exposure groups and the concentration (counts/g cauda) was reduced in the highest exposure group. In the testis, both daily sperm production (DSP) and efficiency (daily sperm production/g testis) was reduced in all exposure groups in a dose dependent manner. In the highest dose group DSP and DSP/g testis were 61.3% and 63.3% of the control group, respectively.	Supporting evidence of adverse effect. Reduced sperm numbers in three smaller studies but not in one larger study. Evaluation of Hoberman study (2008) by SCCS led to conclusion that the study is not accepted as "unarguable refutation of the Oishi
Hoberman *et al*. 2008	EATS-mediated	Sperm numbers	Rat	8 (starting PND 22)	Oral	-	No effect on testicular spermatid concentration, count/g testis. No effect on cauda epididymal sperm concentration, count/g cauda	
Garcia *et al*. 2017	EATS-mediated	Sperm numbers	Rat	8 (3 days per week starting PND 42)	Direct (s.c.)	150/-	Decrease in spermatozoa counts in epididymis (count/g epididymis): annotation in table indicates reduced sperm count at all doses	

							(37-51% of controls), but it is in comparison to the naive control. All doses show numerically lower spermatozoa counts than vehicle (oil) control (50-69% of oil control), but with no indication of statistical significance. Testicular spermatid counts (count/g testis) were significantly reduced at the highest dose of 600 mg/kg bw/day (62% of control) when compared to naïve controls. In comparison to vehicle (oil) controls there was only slight reduction (88% of oil control) in testicular spermatid counts.	findings" (Oishi, 2001)
Riad *et al*. 2018	EATS-mediated	Sperm numbers	Rat	8 (starting PND 19-22)	Oral	50	Decrease in sperm count (sperm/g cauda) to approximately 55% of the control group.	
Hoberman *et al*. 2008	EATS-mediated	Sperm motility	Rat	8 (starting PND 22)	Oral	-	No effect on vas sperm motility, Relative (% motile)	Supporting evidence of adverse effect. Reduced sperm motility in one smaller study but not in one larger study. Evaluation of Hoberman study (2008) by SCCS led to conclusion that the study is not accepted as "unarguable refutation of the Oishi findings" (Oishi,
Riad *et al*. 2018	EATS-mediated	Sperm motility	Rat	8 (starting PND 19-22)	Oral	50	Decrease in sperm motility to 51.8% of controls. Relative (percent motility).	

								2001).
Hoberman *et al*. 2008	EATS-mediated	Sperm morphology	Rat	8 (starting PND 22)	Oral	-	No effect on abnormal sperm endpoint, Relative (% abnormal)	Supporting evidence of adverse effect. Increased number of abnormal sperm in one smaller study but not in one larger study. Evaluation of Hoberman study (2008) by SCCS led to conclusion that the study is not accepted as "unarguable refutation of the Oishi findings" (Oishi, 2001).
Garcia *et al*. 2017	EATS-mediated	Sperm morphology	Rat	8 (3 days per week starting PND 42)	Direct (s.c.)	300	Change. For sperm morphology the percentage of normal was decreased and the percentage of abnormal increased in the two highest dose groups (300 and 600 mg/kg bw/day) compared to the vehicle (oil) control. All doses were affected compared to the naïve control. Relative (% abnormal)	
Integrated assessment of lines of evidence for reduced sperm count and quality in adults: There is supporting evidence of endocrine activity *in vivo* (altered hormone levels). There is supporting evidence of adverse effect (decreased sperm count, number of normal sperm and motility, and altered testicular histopathology).								

Annex II – Additional information: Mode of action analysis

Introduction

To determine whether butylparaben can be considered an endocrine disrupter it is necessary to describe endocrine-related adverse effects and endocrine activity. In addition, a link between the two is established based on biological plausibility as it relates to current scientific knowledge, using a Weight of Evidence approach.

To ensure transparency, a mode of action analysis was carried out taking into account EFSA/ECHA 2018, section 3.5. This guidance document highlights that both biological plausibility and empirical support are weighted, however biological plausibility is the most influential consideration.

For clarity, the mode of action analysis was carried out for perinatal and adult exposure in two separate sections.

Tables

Perinatal exposure

The information from the analysis of Lines of evidence (Annex I) is ordered and mapped, and descriptions of key events support the postulated mode of action. For each possible mode of action, a summary table is presented (Tables B1-B3).

Table B1: Summary table on key events for mode of action of butylparaben with perinatal exposure – <u>estrogen receptor activation</u>.

Summary of hypothesis: The molecular initiating event (MIE) is activation of the estrogen receptor(s). In developing males, increased estrogen receptor signaling results in altered testicular development in offspring and subsequently altered testicular function in adulthood. In turn, reduced sperm count and quality are observed in offspring.		
	Brief description of key event (KE)	**Supporting evidence**
MIE	Molecular: Activation of estrogen receptor	High. Lines of evidence show sufficient evidence for endocrine activity related to estrogen receptor activation. Several studies show estrogen receptor agonistic response similar to estrogen (Gonzalez *et al*, 2018; Pop *et al*, 2018; Watanabe *et al*, 2013).
KE1	Increased estrogen receptor signaling	High. Several studies show effects on growth of estrogen sensitive cells (Khanna & Darbre, 2013; Charles & Darbre, 2013; Gonzalez *et al*, 2018; Pop *et al*, 2018; Williams *et al*, 2019, van Meeuwen *et al*. 2008) or tissues (uterotropic assay *in vivo*; Routledge *et al*, 1998; Hossaini *et al*, 2000; Lemini *et al*, 2003; Lemini *et al*, 2004; Goswami & Kalita, 2016; Vo & Jeung, 2009).
KE2	Organ: Altered reproductive development of male offspring	Moderate. Reduced AGD in males at PND 1 and 21 (Zhang *et al*, 2014; Boberg *et al*, 2016), but other studies showed no effect on AGD at PND 1 (Kang *et al*, 2002; Guerra *et al*, 2017b) or in fetal males GD 21 (Taxvig *et al*, 2008). Inconsistency between studies on AGD may be due to different exposure periods, dose levels and measuring sensitivity. The two studies including doses of 400 mg/kg bw/day or above both showed reduced sperm counts at these doses (Zhang *et al*, 2014; Boberg *et al*, 2016). A dose of 100 mg/kg bw/day reduced AGD in one study (Boberg *et al*, 2016), but in other studies doses in the same range (10 to 200 mg/kg bw/day) did not affect AGD (Kang *et al*, 2002; Zhang *et*

		al, 2014; Guerra *et al*, 2017b). No changes in fetal testis histology (Boberg *et al*, 2016; Guerra *et al*, 2017b). Signs of histological effects on seminiferous tubules of prepubertal testes in one study (Zhang *et al*, 2014).
KE3	Organ: Altered testicular and epididymal function of adult offspring	Moderate. Altered serum levels of T, estradiol (E2) (and LH, FSH; increase or decrease depending on study design) (Zhang *et al*, 2014; Zhang *et al*, 2016; Guerra *et al*, 2017b). Altered adult testicular histopathology (increased number of Leydig cells and possible change in spermatogenesis kinetics (Guerra *et al*, 2017b), and reduced number of round and elongated spermatids (Kang *et al*, 2002). Altered testicular expression of hormone receptors (altered expression of ERalpha and ERbeta mRNA (Kang *et al*, 2002); possibly reduced protein expression of ERalpha and AR in some cell types and spermatogenic stages (Guerra *et al*, 2017b). No reports of change in epididymal histology.
Adverse Outcome (AO) 1	Organ: Reduced sperm count and quality of offspring's	High. All studies using perinatal exposure caused altered sperm count and/or quality, though different parameters were affected in different studies. Reduced epididymal sperm count (50-75% of control; Kang *et al*, 2002; Boberg *et al*, 2016) but no change in epididymal sperm count in another study (Guerra *et al*, 2017b). Reduced sperm motility (60% of control; Kang *et al*, 2002) and reduced percentage of progressive motile sperm (low dose only, Guerra *et al*, 2017b). Increased percentage of sperm with head abnormalities and reduced percentage of normal sperm (Guerra *et al*, 2017b).
AO2	Organism: Impaired fertility of male offspring	Low evidence for effect in rodents, but high plausibility that impaired sperm count and quality in humans lead to impaired fertility (see Biological plausibility table below). No effect on fertility assessed by natural mating or artificial insemination (Guerra *et al*, 2017b).

Table B2: Summary table on key events for mode of action of butylparaben with perinatal exposure – <u>altered steroidogenesis</u>

Summary of hypothesis: The molecular initiating event is altered steroidogenesis/increased estradiol level. In developing males, increased estrogen receptor signaling results in altered testicular development in offspring and subsequently altered testicular function in adulthood. In turn, reduced sperm count and quality are observed in offspring.		
	Brief description of key event (KE)	**Supporting evidence**
MIE	Molecular: Altered steroidogenesis/increased estradiol level	Moderate. Lines of evidence show supporting evidence for endocrine activity related to altered steroidogenesis. One study shows upregulation of aromatase/Cyp19a1 gene expression and increased aromatase enzyme activity as well as increased estradiol levels in different cell lines (Williams *et al*, 2019). Other studies show no increase of estradiol levels in other cell lines, and the response may be cell type specific (Taxvig *et al*, 2008; Wróbel & Gregoraszczuk; 2013; Guerra *et al*, 2016, van Meeuwen *et al*. 2008).
KE1	Increased estrogen receptor signaling	High. (See Table B1)
KE2	Organ: Altered reproductive development of male offspring	Moderate. (See Table B1)
KE3	Organ: Altered testicular and epidiymal function of adult offspring	Moderate. (See Table B1)
AO1	Organ: Reduced sperm count and quality of offspring	High. (See Table B1)
AO2	Organism: Impaired fertility of	Low evidence for effect in rodents, but high plausibility

	male offspring	that impaired sperm count and quality in humans leads to impaired fertility. (See Table B1)

Table B3: Summary table on key events for mode of action of butylparaben with perinatal exposure – androgen receptor antagonism

Summary of hypothesis: The molecular initiating event is antagonism of the androgen receptor. In developing males, reduced androgen receptor signaling results in altered testicular development in offspring and subsequently altered testicular function in adulthood. In turn, reduced sperm count and quality are observed in offspring.		
	Brief description of key event (KE)	**Supporting evidence**
MIE	Molecular: Androgen receptor antagonism	Moderate. Lines of evidence show supporting evidence for endocrine activity related to androgen receptor antagonism. Studies on anti-androgenic activity show inconsistent findings. Effect in some (Chen *et al*, 2007; Pop *et al*, 2016) but not all studies (Kjærstad *et al*, 2010; Watanabe *et al*, 2013), possibly due to different study design. Studies on androgen activity consistently show no androgen activity (agonism) of butylparaben (Gonzalez *et al*, 2018; Chen *et al*, 2007; Pop *et al*, 2016).
KE1	Molecular: Reduced androgen receptor signaling	Moderate. There are some indications of effect of butylparaben on the androgen-sensitive endpoint anogenital distance in male offspring's. Moderate reduction of AGD was seen in two studies evaluating AGD at PND 1 and using doses of 400 mg/kg bw/day or above (Zhang *et al*, 2014; Boberg *et al*, 2016), but not in other studies using lower doses (Kang *et al*, 2002; Guerra *et al*, 2017b) or using high doses but evaluating AGD at GD 21 (Taxvig *et al*, 2008). Inconsistency between studies may be due to different exposure periods, dose levels and measuring sensitivity. There are no data from Hershberger studies on butylparaben, but a study on the structurally related propylparaben showed clear evidence of anti-androgenic effect in a Hershberger assay (Özdemir *et al*, 2018).
KE2	Organ: Altered reproductive development of offspring	Moderate. (See Table B1)
KE3	Organ: Altered testicular and epidiymal function of adult offspring	Moderate. (See Table B1)
AO1	Organ: Reduced sperm count and quality of offspring	High. (See Table B1)
AO2	Organism: Impaired fertility of male offspring	Low evidence for effect in rodents, but high plausibility that impaired sperm count and quality in human leads to impaired fertility (see Biological plausibility table below). (See Table B1)

In addition to the observations made in studies using butylparaben, the conclusions on biological plausibility of these modes of action are strengthened by evidence from other endocrine disrupters affecting sperm count and quality of offspring. This includes understanding of physiology, endocrinology and toxicology, and information from studies on other chemicals or knock-out models. This information is listed below:

Table B4: Analysis of biological plausibility of estrogen receptor activation leading to impaired fertility of male offspring

Summary of hypothesis: The molecular initiating event is activation of the estrogen receptor(s). In developing males, increased estrogen receptor signaling results in altered testicular development and function. In turn, reduced sperm count and quality are observed in offspring.		
	Brief description of	**Supporting evidence**

	key event relationship (KER)	
MIE to KE1	Estrogen receptor activation to Increased transcription of estrogen receptor regulated genes	High. Estrogen receptor activation leads to increased estrogen receptors signaling.
KE1 to KE2	Increased estrogen receptor signaling to Altered reproductive development of offspring	Moderate to high. ERalpha is expressed in foetal Leydig cells (Nielsen *et al*, 2000) and has regulatory effects on steroidogenesis; endogenous estrogens inhibit testicular development and function in fetal/neonatal life (Delbes *et al*, 2006; Delbes *et al*, 2005). Exogenous 'estrogens' lead to decreased testosterone levels in rodents (Delbes *et al*, 2004; Delbes *et al*, 2005; Lassurguere *et al*, 2003; Lehraiki *et al*, 2011). In turn, reduced testosterone levels in male foetus may cause failure to masculinise male foetus (Stewart *et al*, 2018; Schwartz *et al*, 2019). Human evidence is scarce, but one study has shown association between polymorphism of ESR1 (coding for ERalpha) and short AGD in boys (Sathyanarayana *et al*, 2012). ERbeta is expressed in late gestation gonocytes (Jefferson *et al*, 2000) and regulates apopototic/mitotic rate during late gestation (Delbes *et al*, 2004), and ERbeta activation by estrogenic compounds could cause altered gonocyte proliferation (Delbes *et al*, 2006). There are several examples of exogenous estrogens altering male reproductive development in rodents. Developmentally estrogenised male mice display retained or cryptorchid testes, decrease in sperm number, increase in abnormal sperm, retained Müllerian ducts, epididymal cysts, hypospadias, and prostatic disease, and such phenotype has been seen, in whole or in part, in mice, rats, hamsters, and humans exposed to estrogens *in utero* (McLachlan *et al*, 2001). Biological pathways leading from estrogen receptor activation to effects on testes are currently not well-described and may include alterations of androgen-dependent processes such as suppression of testosterone production and downregulation of the expression of the androgen receptor protein in reproductive target tissues including the testes (Martin *et al*, 2007). The epididymis is highly responsive to androgens, but estrogen has a predominant role in efferent ductules and initial segment epididymis (Joseph *et al*, 2011). Interference with estrogen receptor signaling can thus affect epididymal development both directly and secondary to altered androgen receptor signaling.
KE2 to KE3	Altered reproductive development of offspring to Altered testicular and epididymal function in adult offspring	High. Correct development of the reproductive system in early life is essential to achieve optimal reproductive function in adulthood. It is highly biologically plausible that impaired reproductive development is a cause of altered testicular and epididymis function in adulthood.
KE3 to AO1	Altered testicular and epididymal function in adult offspring to Reduced sperm count and quality in offspring	High. Correct function of testis and epididymis is necessary for an optimal sperm count and quality (motility, morphology)
AO1 to AO2	Reduced sperm count and quality in offspring to Impaired fertility of	High. There is clear evidence that impaired sperm count and quality in humans leads to impaired fertility. In rodents, reproductive function is less sensitive to reductions in sperm count and

	male offspring	quality.

Table B5: Analysis of biological plausibility of altered steroidogenesis/ increased estradiol level leading to impaired fertility of male offspring.

Summary of hypothesis: The molecular initiating event is altered steroidogenesis/increased estradiol level causing increased estradiol levels and in turn activation of the estrogen receptor(s). In developing males, increased estrogen receptor signaling results in altered testicular development and function. In turn, reduced sperm count and quality are observed in offspring.		
	Brief description of key event relationship (KER)	**Supporting evidence**
MIE to KE1	Altered steroidogenesis/increased estradiol level to Increased estrogen receptor signaling	High. When estradiol levels are increased it is biologically plausible that increased estrogen receptors signaling occurs.
KE1 to KE2	Increased estrogen receptor signaling to Altered reproductive development of offspring	Moderate to high. (see Table B4)
KE2 to KE3	Altered reproductive development of offspring to Altered testicular and epidiymal function in adult offspring	High. (see Table B4)
KE3 to AO1	Altered testicular and epidiymal function in adult offspring to Reduced sperm count and quality in offspring	High. (see Table B4)
AO1 to AO2	Reduced sperm count and quality in offspring to Impaired fertility of male offspring	High. (see Table B4)

Table B6: Analysis of biological plausibility of androgen receptor antagonism leading to impaired fertility of male offspring.

Summary of hypothesis: The molecular initiating event is antagonism of the androgen receptor. In developing males, impaired androgen signaling altered testicular development and function. In turn, reduced sperm count and quality are observed in offspring.		
	Brief description of key event relationship (KER)	**Supporting evidence**
MIE to KE1	Androgen receptor antagonism to Reduced androgen receptor signaling	High. Antagonism of androgen receptor leads to reduced androgen receptor signaling.
KE1 to KE2	Reduced androgen receptor signaling To Altered reproductive development of offspring	High. Reduced transcription of androgen regulated genes in peritubular cells can cause reduced proliferation of Sertoli cells (SC). AR knockout mice show progressive loss of Sertoli cells, but SC-specific KO mice do not (suggesting SC independent AR action) (Tan *et al*, 2005; Scott *et al*, 2007). The epididymis is highly responsive to androgens, but estrogen has a predominant role in efferent ductules and initial segment epididymis (Joseph *et al*, 2011). Interference with androgen receptor signaling can thus affect epididymal development both directly and secondary to altered estrogen receptor signaling.
KE2 to KE3	Altered reproductive development of offspring To Altered testicular and epidiymal function in adult offspring	High. SC support gonocytes and spermatogenesis in postnatal testis, so reduced SC number can cause reduced spermatogenesis/fertility. Impaired epididymal development can cause impaired epididymal function.

KE3 to AO1	Altered testicular and epidiymal function in adult offspring to Reduced sperm count and quality in offspring	High. Correct function of testis and epididymis is necessary for an optimal sperm count and quality (motility, morphology).
AO1 to AO2	Reduced sperm count and quality in offspring to Impaired fertility of male offspring	High. There is clear evidence that impaired sperm count and quality in humans leads to impaired fertility. In rodents, reproductive function is less sensitive to reductions in sperm count and quality.

Table B7: Other considerations for key event relationships of perinatal exposure.

Dose and temporal concordance	**Comment**
Dose	In each *in vivo* study, indicators of key events related to endocrine activity (e.g. altered hormone levels and altered AGD) were affected at the same doses causing adverse effects on sperm parameters. Between studies, there are however some differences in effective doses, possibly depending on study design and sensitivity.
Temporal concordance	Regarding temporal concordance, it is noted that key events are observed in the hypothesised order. *In vivo* indicators of key events related to endocrine activity are seen in developing animals (e.g. altered hormone levels in prepuberty and altered AGD at birth), and adverse effects on sperm parameters are seen in adulthood, i.e. long after the end of exposure. This is in line with the conclusion that developmental changes in the male reproductive system are the cause of adult adverse effects. There is strong empirical support for temporal concordance and moderate support for dose-response concordance.
Essentiality	**Comment**
	For determining essentiality it should be demonstrated whether or not downstream KEs and/or the adverse effect is prevented/decreased if an upstream event is experimentally blocked. It was not examined whether counteracting the endocrine related key events would prevent adverse effects of butylparaben with perinatal exposure.
Human relevance	**Comment**
	There are no data indicating that these endocrine modes of action are not relevant to humans. Thus, human relevance is assumed by default. No epidemiological studies examined the relationship between butylparaben exposure in utero and effect on male reproductive parameters (hormone levels, sperm parameters) later in life. A few studies showed supportive evidence for an endocrine disrupting activity of butylparaben during pregnancy (see Annex V).
Identified uncertainties	**Comment**
No one- or two-generation studies have been performed	Butylparaben has not been tested in one- or two-generation studies. The evidence on adverse effects comes from studies using perinatal exposure only. This is a disadvantage in that there is no evaluation of adverse outcome following exposure during the full reproductive cycle. On the other hand, the available knowledge from studies using perinatal exposures only is an advantage in determining mode of action, as it is clear that effects originate from developmental changes in male reproductive development.
Different effect levels/no effect levels	Adverse effects (reduced sperm count and quality) are seen at different dose levels in different studies. Differences in study design may explain

observed in different *in vivo* studies.	some differences between studies in patterns of late-life effects in male offspring. The two studies including doses of 400 mg/kg bw/day or above both showed reduced sperm counts at these doses (Zhang *et al*, 2014; and Boberg *et al*, 2016), but at lower doses some studies showed effect and other studies did not.
Lack of a clear description of biological pathways leading from estrogen receptor activation or androgen receptor antagonism in utero to adverse effects on testis function in adulthood.	It is biologically plausible that the alteration in sperm count and quality observed for butylparaben is due to endocrine disruption during development. The uncertainty in describing biological pathways also applies to several other estrogenic or anti-androgenic substances for which perinatal exposures lead to reductions in sperm count and/or quality. It is stated in EFSA/ECHA 2018 guidance on identification of endocrine disrupters that to conclude on the biological plausibility of the link, it may not be necessary to have demonstrated the whole sequence of events leading to the adverse effect for the substance under evaluation. Existing knowledge from endocrinology or toxicology may be sufficient to assess the link and conclude on the biological plausibility between adverse effects and the endocrine activity. It is noted that in some cases, "*the MoA analysis could be very simple; when an adverse effect is 'EATS-mediated', the biologically plausible link is already pre-established in the absence of information proving the contrary (i.e. a fully developed non-ED MoA). This is because, in the case of 'EATS-mediated' parameters, where the pattern of effects is deemed adverse, the biological plausibility that the adverse effects are caused via an EATS-mediated MoA is high, based on existing knowledge and theory (i.e. coherence analysis), and as such, it may not be necessary to generate further empirical data on the substance under evaluation to substantiate the link between the observed adverse effect(s) and an endocrine-mediated MoA*." (ECHA/EFSA 2018, section 3.5.2)

Pubertal/adult exposure

The information from the analysis of Lines of evidence (Annex I) is ordered and mapped, and descriptions of key events support the postulated mode of action. For each possible mode of action, a summary table is presented (Tables B8-B10).

Table B8: Summary table on key events for mode of action of butylparaben with adult exposure.

Summary of hypothesis: The molecular initiating event is activation of the estrogen receptor(s). In young or adult males, increased estrogen signaling results in altered testicular function. In turn, reduced sperm count and quality are observed in adulthood.		
	Brief description of key event (KE)	**Supporting evidence**
MIE	Molecular: Estrogen receptor activation (alpha and beta)	High. (See Table B1)
KE1	Molecular: Increased estrogen receptor signaling	High. (See Table B1)
KE2	Organ: Altered testicular function	Moderate. Lines of evidence analysis shows supporting evidence of effect on hormone levels and histopathology. One study showed increased serum estradiol in rats (Riad *et al*, 2018). Four studies showed decreased testosterone levels, but at different doses and ages (Oishi, 2001; Oishi, 2002; Hoberman *et al*, 2008; Riad *et al*, 2018). The largest study (Hoberman *et al*, 2008) showed effect at only one dose. Changes in LH and FSH levels were observed in two studies, but in different directions or at different dose levels (Hoberman *et al*, 2008; Riad *et al*, 2018). Different histological parameters were assessed in different studies showing effect on testis histology in exposed animals (Oishi, 2002; Alam & Kurohmaru, 2014; Riad *et al*, 2018), and these parameters were not evaluated in the largest study (Hoberman *et al*, 2008) showing no histological changes in testes (Hoberman *et al*, 2008). There were no effects on testis weight (Oishi, 2001; Oishi, 2002; Hoberman *et al*, 2008; Riad *et al*, 2018).
AO1	Organ: Reduced sperm count and quality of adults	Moderate. Lines of evidence analysis shows supporting evidence of adverse effects after pubertal/adult exposure (decreased sperm count and decreased number of normal sperm). The conclusion that there is not sufficient evidence of adverse effects is based on inconsistent data from studies with some variation in study design.
AO2	Organism: Impaired fertility of males	No data.

Table B9: Summary table on key events for mode of action of butylparaben with adult exposure.

Summary of hypothesis: The molecular initiating event is altered steroidogenesis/increased estradiol levels and in turn increased activation of the estrogen receptor(s). In young or adult males, increased estrogen signalling results in altered testicular function. In turn, reduced sperm count and quality are observed in adulthood.		
	Brief description of key event (KE)	**Supporting evidence**
MIE	Molecular: altered steroidogenesis/increased estradiol levels	Moderate. (See Table B2)
KE1	Molecular: Increased estrogen	High.

	receptor signaling	(See Table B1)
KE2	Organ: Altered testicular function	(See Table B8)
AO1	Organ: Reduced sperm count and quality of adults	(See Table B8)
AO2	Organism: Impaired fertility of males	(See Table B8)

Table B10: Summary table on key events for mode of action of butylparaben with adult exposure.

Summary of hypothesis: The molecular initiating event is antagonism of the androgen receptor. In young or adult males, impaired androgen signaling results in altered testicular function. In turn, reduced sperm count and quality are observed in adulthood.		
	Brief description of key event (KE)	**Supporting evidence**
MIE	Molecular: Androgen receptor antagonism	Moderate. (See Table B3)
KE1	Molecular: Reduced transcription of androgen regulated genes	Moderate. (See Table B3).
KE2	Organ: Altered testicular function	(See Table B8)
AO1	Organ: Reduced sperm count and quality of adults	(See Table B8)
AO2	Organism: Impaired fertility of males	(See Table B8)

In addition to the observations made in studies using butylparaben, the conclusions on biological plausibility of these modes of action are strengthened by evidence from other endocrine disrupters affecting sperm count and quality with adult exposure. This includes understanding of physiology, endocrinology and toxicology, and information from studies on other chemicals or knock-out models. This information is listed below:

Table B11: Analysis of biological plausibility of estrogen receptor activation leading to impaired fertility of adult males.

Summary of hypothesis: The molecular initiating event is activation of the estrogen receptor(s). In young or adult males, increased estrogen signaling results in altered testicular function. In turn, reduced sperm count and quality is observed.		
	Brief description of key event relationship (KER)	**Supporting evidence**
MIE to KE1	Estrogen receptor activation to Increased estrogen receptor signaling	High (see Table B4)
KE1 to KE2	Increased estrogen receptor signaling to Altered testicular function	Moderate. Estrogen receptors are expressed in mature testes and epididymides of both rodents and humans and estrogen signaling is involved in the control of spermatogenesis (O'Donnell *et al*, 2001; Akingbemi 2005). Administration of estrogen to mature males can perturb spermatogenesis (Steinberger & Ducket 1965; Meistrich *et al*, 1975) and affect epididymal function. In addition, signal disruption can also occur via interference with the hypothalamo-pituitary-gonadal axis, for example by estrogen-induced changes in gonadotropin production.

		Studies on adult exposures to estrogens support these key event relationships. Exposure of adult male mice to estradiol benzoate did not affect testicular sperm count, but significantly reduced sperm count in the epididymis (Meistrich *et al*, 1975). Exposure to Bisphenol A (BPA) in adult mice cause reduced sperm count and quality (Zhang *et al*, 2013a; Pengpeng *et al*, 2013).
KE2 to AO1	Altered testicular function to Reduced sperm count and quality of offspring	High. Correct function of testis and epididymis is necessary for an optimal sperm count and quality (motility, morphology).
AO1 to AO2	Reduced sperm count and quality of offspring to Impaired fertility of males	High. There is clear evidence that impaired sperm count and quality in humans leads to impaired fertility. In rodents, reproductive function is less sensitive to reductions in sperm count and quality.

Table B12: Analysis of biological plausibility of altered steroidogenesis leading to impaired fertility of adult males.

Summary of hypothesis: The molecular initiating event is altered steroidogenesis/increased estradiol levels and in turn activation of the estrogen receptor(s). In young or adult males, increased estrogen signaling results in altered testicular function. In turn, reduced sperm count and quality is observed.		
	Brief description of key event relationship (KER)	**Supporting evidence**
MIE to KE1	Altered steroidogenesis/increased estradiol level to Increased estrogen receptor signaling	High. When estradiol levels are increased it is biologically plausible that increased estrogen receptors signaling occurs.
KE1 to KE2	Increased estrogen receptor signaling to Altered reproductive development of offspring	Moderate. (see Table B4)
KE2 to AO1	Altered testicular function to Reduced sperm count and quality of offspring	High. Correct function of testis and epididymis is necessary for an optimal sperm count and quality (motility, morphology).
AO1 to AO2	Reduced sperm count and quality of offspring to Impaired fertility of males	High. There is clear evidence that impaired sperm count and quality in humans leads to impaired fertility. In rodents, reproductive function is less sensitive to reductions in sperm count and quality.

Table B13: Analysis of biological plausibility of estrogen receptor activation leading to impaired fertility of adult males.

Summary of hypothesis: The molecular initiating event is androgen receptor antagonism. In young or adult males, reduced androgen signaling results in altered testicular function. In turn, reduced sperm count and quality are observed.		
	Brief description of key event relationship (KER)	**Supporting evidence**
MIE to KE1	Androgen receptor antagonism to Reduced androgen receptor signaling	High. Antagonism of androgen receptor leads to reduced androgen receptor signaling.
KE1 to KE2	Reduced androgen receptor signaling To	Moderate. Testosterone signaling through the Androgen receptor is required for spermatogenesis in

	Altered testicular function	mature males, evidenced by various mouse knockout models and human patients with AR mutations (O'Hara & Smith, 2015). AR is expressed by most somatic cells of the adult testis (but not germ cells) and blockage of AR signaling in Sertoli and peritubular cells perturbs spermatogenic progression. Some chemicals with anti-androgenic mode of action cause reduced sperm count in rats exposed in adulthood.
KE2 to AO1	Altered testicular function to Reduced sperm count and quality of offspring	High. Correct function of testis and epididymis is necessary for an optimal sperm count and quality (motility, morphology).
AO1 to AO2	Reduced sperm count and quality of offspring to Impaired fertility of males	High. There is clear evidence that impaired sperm count and quality in humans leads to impaired fertility. In rodents, reproductive function is less sensitive to reductions in sperm count and quality.

Table B14: Other considerations for key event relationships of adult exposure.

Dose and temporal concordance	**Comment**
Dose	In each *in vivo* study, indicators of key events related to endocrine activity (altered hormone levels) are affected at the same doses causing adverse effects on sperm parameters. Between studies, there are however some differences in effective doses, possibly depending on study design and sensitivity.
Temporal concordance	Regarding temporal concordance, it is noted that key events are observed in the hypothesised order. In the studies showing effects, *in vivo* indicators of key events related to endocrine activity (altered hormone levels) are seen during exposure (though at selected ages only), and adverse effects on sperm parameters are seen at sacrifice in adulthood.
Essentiality	**Comment**
	For determining essentiality it should be demonstrated whether or not downstream KEs and/or the adverse effect is prevented/decreased if an upstream event is experimentally blocked. It has not been examined whether the counteracting the endocrine related key events prevents adverse effects of butylparaben with pubertal/adult exposure.
Human relevance	**Comment**
	There are no data indicating that these endocrine modes of action are not relevant to humans. Thus, human relevance is assumed by default. Only few epidemiological studies examined the relationship between butylparaben exposure in adults and male reproductive parameters, and one observed an inverse association between male urinary butylparaben levels and sperm concentration and sperm motility, while three observed no associations.
Identified uncertainties	**Comment**
No one- or two-generation studies or 90-day subchronic toxicity studies have been performed	Butylparaben was not tested in one- or two-generation studies or 90-day subchronic toxicity studies. The (limited) evidence on adverse effects comes from studies using pubertal/adult exposure only.
Different effect levels/no effect levels observed in different *in vivo* studies.	Adverse effects (reduced sperm count and quality) are not seen in all studies. Differences in study design may explain some differences between studies.

Lack of a clear description of biological pathways leading from estrogen receptor activation or androgen receptor antagonism to adverse effects on testis function in adulthood.	It is biologically plausible that the alteration in sperm count and quality observed for butylparaben is due to endocrine disruption in adulthood. The uncertainty in describing biological pathways applies also to several other estrogenic or anti-androgenic substances for which pubertal/adult exposures lead to reductions in sperm count and/or quality. It is stated in EFSA/ECHA 2018 guidance on identification of endocrine disrupters that to conclude on the biological plausibility of the link, it may not be necessary to have demonstrated for the substance under evaluation the whole sequence of events leading to the adverse effect. Existing knowledge from endocrinology or toxicology may be sufficient to assess the link and come to a conclusion on the biological plausibility between adverse effects and the endocrine activity. It is noted that in some cases, "*the MoA analysis could be very simple; when an adverse effect is 'EATS-mediated', the biologically plausible link is already pre-established in the absence of information proving the contrary (i.e. a fully developed non-ED MoA). This is because, in the case of 'EATS-mediated'* *parameters, where the pattern of effects is deemed adverse, the biological plausibility that the adverse effects are caused via an EATS-mediated MoA is high, based on existing knowledge and theory (i.e. coherence analysis), and as such, it may not be necessary to generate further empirical data on the substance under evaluation to substantiate the link between the observed adverse effect(s) and an endocrine-mediated MoA*." (ECHA/EFSA 2018, section 3.5.2)

Annex III - Detailed study information

In vivo studies

Detailed study information on *in vivo* studies are summarised below:

Exposure Period	Reference	Year	System	Method	Relevant Endpoints & Effects (Effects Are Annotated With Arrows Or 'Affected')	Effects	Noel/Loel	Comments/Notes	Klimisch Score
Perinatal exposure	Fisher *et al*.	1999	Wistar rats	Neonatal repeated, s.c injection (PND 2-18). Dose: 2 mg/ kg bw /day, n= 6.	Testis weight Testis histopathology	No effects reported	No observed effect level) NOEL = 2 mg/kg bw /day	BP was a gift and the purity of the compound was not reported. Also, only one dose tested.	*Reliability 2.* - Acceptable study, which meets basic scientific principles - One shortcoming is that purity of BP is not reported. - Single dose tested

	Kang *et al*.	2002	Sprague-Dawley rats	Development of male reproductive system, s.c. (GD6-PND20). Dose 110, 200 mg/ kg bw/day, n = 5-7 for organ weight/histology, 5 form sperm parameters and 3 for gene expression	Pups: *Live births* ↓ *Surviving to weaning* ↓ AGD Weight: *Testis* ↓↑ *Prostate* ↓ *Seminal vesicle* ↓ Sperm: *Numbers* ↓ *Motility* ↓ *Morphology* ↓ ERα and ERβ expression in testis ↓↑		NOAEL = 100 mg/kg bw/day	The study is well described. Purity of BP not reported. Two doses of BP included.	*Reliability 2.* - Acceptable, well-documented study - One shortcoming is that purity of BP is not reported, but likely of an acceptable purity
	Taxvig *et al*.	2008	Wistar rats	Development of male reproductive system, s.c. (GD7-21). Dose: 200, 400 mg/kg bw(day, n = 13-18	AGD Hormones: *Testosterone* *Progesterone* *Cortisol* Histopahtology: *Testis* *Adrenal*	No effects reported	x	Hormones measured in dams GD 21: 17α-hydroxyprogesterone and progesterone, no effects of exposure. Female AGD also measured with no effects reported.	*Reliability 2.* - Acceptable, well-documented study, comparable to guideline standards - *In vivo* and *in vitro* study included - *In vitro* study on effects on steroidogenesis conducted based on the OECD test guideline system H295R
	Zhang *et al*.	2014	Wistar rats	Development of male reproductive system, oral (gavage) (GD7-PND21). Dose: 64, 160, 400, 1000 mg/ kg bw/day, n =7-8.	Pups: *Sex ratio* ↓ *Body weight* ↓ AGD ↓ Puberty (delayed) Weight: *Testis* ↓ *Epididymis* ↓ *Seminal vesicle* ↓	Effects on AGD, sexual maturation, hormone levels, sperm parameters and testicular health	NOAEL 64 mg/kg bw/day	Dams: FSH and LH ↑ Offspring affected at several ages (for many endpoints PND 21, 35, 49, 90, 180. Male offspring: sex ratio affected (fewer males) Bw significantly decreased from PND 0-49 in dose groups 400	*Reliability 2.* - Acceptable, well-documented study - Basic data given, comparable to guideline standards

					Hormones: *Testosterone* ↓ *Estradiol* ↑ *Progesterone* ↑ *LH* ↓↑ *FSH* ↓↑ Sperm numbers and daily sperm production ↓ Histopathology testis (affected PND 21 and 90)			and 1000 mg/kg/day, but not affected PND 90-180. Weight of testis, epididymis and seminal vesicles decreased, however several overlaps with reduced BW and relative weights not reported. AGD shortened on PND1 and 21 (also coincides with reduced BW), with significant differences at 400 and 1000 mg/kg/day. AGD PND1 (mm): 3.98 ± 0.55 (0 mg/kg/day), 3.5 ± 0.53 (400 mg/kg/day) and 3.35 ± 0.56 (1000 mg/kg/day), AGD PND21 (mm): 20.44 ± 1.46 (0 mg/kg/day), 19.79 ± 0.42 (400 mg/kg/day), 18.12 ± 0.95 (1000 mg/kg/day) Testis histopathology affected on PND 21 and 90 with, fex, reduced and loosely arranged germ cells, reduced layers of seminiferous tubules, reduced numbers of spermatocytes. No obvious effects on Leydig cells. Testosterone hormone levels in male offspring were significantly reduced at PND 21, 35, 49, 90, and 180 in offspring with maternal dose of 1000 mg/kg/day. In offspring with maternal dose of 400 mg/kg/day, testosterone levels were significantly reduced at PND 35 and 49, but normalised at PND 90 and 180. Male estradiol levels were significantly increased at PND 21, 35, 49, and 90 (1000 mg/kg/day).	

	Zhang *et al*.	2016	Wistar rats	Mechanisms of ED and reproductive disorders, oral (gavage) (GD7-PND21). Dose: 64, 160, 400, 1000 mg/kg bw/day, n = 7-8.	Body weight ↓ Weight: *Testis* *Epididymis* ↓ *Seminal vesicle* Hormones: *Testosterone* ↓ *Estradiol* ↑ Gene expression: *Star, P450scc, Sult1e1* (affected) Gene and protein expression: *Erα*, *Erβ*, *Ar* (affected) Methylation of Erα promoter ↓ Histopathology testis (affected)	Effects on hormone levels, testicular health, expression of key steroidogenic enzymes and receptors.	NOAEL = 160 mg/kg bw/day (effects are seen at protein level at this dose)	Data is possibly based on the same animal study as Zhang *et al*. 2014.	*Reliability 2.* - Acceptable, well-documented study, which meets basic scientific principles - Concern: some data seem to already have been reported in Zhang *et al*. 2014 and are likely reported here again without reference to the previous study.

	Boberg *et al*.	2016	Wistar rats	Development of male reproductive system, oral (gavage) (GD7-21 and PD1-22). Dose: 10, 100, 500 mg/kg bw/day, n = 18.	AGD and AGDi ↓ Nipple retention Puberty Weight: *Testis* *Prostate* ↓ *Seminal vesicle* ↓ *Epididymis* *LABC* *Bulbourethral gland* Sperm numbers ↓ Histopathology: *Epididymis* *Testis* *Prostate* (affected) Gene expression *Cyp19a1* ↓ Hormones - Estradiol	Effects on AGD, sperm parameters and prostate health.	Lowest oberserved adverse effect level (LOAEL) = 10 mg/kg bw/day	No significant effects on bw in offspring observed. No data on hormone levels provided. AGD and AGDi shortened in both males and females. Male AGD significantly reduced in dose groups of 100 and 500 mg/bw/day. Male AGD (mm), PND1: 3.96 ± 0.1 (0 mg/kg bw/day), 3.77 ± 0.2 (100 mg/kg bw/day), 3.69 ± 0.3 (500 mg/kg bw/day). Number of sperm in cauda significantly reduced in all dose groups. Genes (cell markers, receptors (Ar, Fshr, Lhr), steroidogenesis) were investigated in testis PD 16 and in adulthood. Down regulation of Cyp19a1 in all exposure groups was seen on PD16. Not other effects seen on gene expression. Hormone levels (estradiol measured PD16 males and PD 22 females): no effect. Mammary gland was investigated in females. PD 22: higher number of TEBs in two highest dose groups (100, 500 mg/kg bw/day). Increased outgrowth towards the lymphnode in 100 mg/kg bw/day. Adult: no clear effects seen.	*Reliability 2.* - Acceptable, well-documented study, comparable to guideline standards

	Guerra *et al*	2017b	Wistar rats	Male reproductive development, s.c. (GD 12 - PND21). Dose: 10, 100, 200 mg/kg/day, n = 8/group.	AGD Nipple retention Puberty Weight: Pituitary *Testis* *Epididymis* *Prostate* *Seminal vesicle* *Vas deferens* Histopathology: *Fetal testis* *PND 110 testis ↓ Leydig cells* Hormones: *Testosterone* ↑ *FSH* ↓ *LH* ↓ Sperm: *Spermatogenesis kinetics* ↑↓ *Sperm counts* *Motile sperm* ↓ *Non-motile sperm* *Normal morphology* ↓ *Abnormal morphology* ↑ Testis morphometry (no of cells) ESR1 and AR protein in testis ↓ Sexual behaviour Fertility	Effects on hormon levels, sperm parameters and protein levels of receptors in testis.	NOAEL = 10 mg/kg/day		*Reliability 2.* - Acceptable, well-documented study, comparable to guideline standards - One shortcoming is that purity of BP is not reported, but likely of an acceptable purity
	Maske *et al*	2020	Holtzman rats	Male reproductive development, s.c. (GD6 – PND21). Dose: 10, 100, 1000 mg/ kg bw/day, n = 6-10 / group.	Testicular descent (delayed) Puberty (delayed) Weight: *Testis* *Epididymis* ↓ *Seminal vesicles* ↑ *Prostate* ↑ *Hypothalamus* ↓ *Pituitary* ↓↑ Hormones: *Testosterone* ↓↑ *Estradiol* ↓↑ *LH* ↑	Effects on testicular descent, puberty, weight of reproductive organs, hormone levels, sperm parameters, fertility and gene expression in testis.	LOAEL = 10 mg/kg bw/day	Testicular descent was delayed in the two highest dose groups and balanopreputial separation (puberty) delayed in the 10 mg/kg bw group. Weight of reproductive organs was affected at several ages and doses, not necessarily dose-response. Hormone levels were	*Reliability 2.* - Acceptable, well-documented study - One shortcoming is that purity of BP is not reported, but likely of an acceptable purity

					Testis histopathology (affected) Sperm: *Motility* ↓ *Sperm count* ↓ *Daily sperm prod.* ↓ *Sperm transit time* ↓ Fertility (affected) Gene expression: *Ar* ↑ *Erα* ↑ *Erβ* ↑ *Insl3* ↑ *Star* ↓			also affected at several ages. Sperm related parameters were affected and pattern of reduced motility, count, production etc was seen. Reduced fertility was seen in naïve females mated with the exposed males. Gene expression in tesis was affected.	
Postnatal exposure Start prepubertal Exposure for 8-10 weeks	Oishi	2001	Wistar rats	Repeated dose, oral (diet) (8 weeks from PND 19-21). Dose: 10.4 ± 3.07, 103 ± 31.2, 1026 ± 310 mg/kg bw/day , n = 8.	Weight: *Testis* *Epididymis* ↓ *Prostate* *Seminal vesicle* ↓ *Preputial glands* Sperm numbers (testis and cauda) ↓ Testosterone ↓		LOAEL = 10.4 mg/kg bw/day (0.01%)	Absolute and relative weights of epididymides were decreased in a dose-dependent manner and the decrease was significant at the two highest dosing groups. The sperm count of the group receiving the highest dose was 58.2% of control values. Testosterone levels were reduced dose-dependently and significantly reduced at 10.4 mg/kg bw/day.	*Reliability 2.* - Acceptable, well-documented study performed based on basic scientific principles - European Union Scientific Committee on Consumer products (SCCP) has acknowledged some doubts on the results of this study

	Oishi	2002	CD-1 ICR mice	Repeated dose, oral (diet) (10 weeks from PND 27-29). Dose: 14.4 ± 3.60, 146 ± 35.9, 1504 ± 357 mg/kg bw/day, n = 8	Weight: T*estis* *Epididymis* ↓ *Prostate* *Seminal vesicle* *Preputial glands* Sperm morphology: *Type and stage* (affected) Testosterone ↓		NOAEL = 14.4 mg/kg bw/day	There were no treatment-related effects of butyl paraben on the liver, ventral prostates, seminal vesicles, and preputial glands (both in terms of absolute weight and relative to body weight) in any of the study groups. Testosterone levels were reduced dose-dependently and significantly reduced at 1504 mg/kg bw/day.	*Reliability 2.* - Acceptable, well-documented study performed based on basic scientific principles - European Union Scientific Committee on Consumer products (SCCP) has acknowledged some doubts on the results of this study
	Hoberman *et al*.	2008	Wistar rats	Repeated dose, oral (diet) (Start PND22. Continued for 8 weeks). Dose: 10.9 ± 0.4, 109.3 ± 8.2, 1087.6 ± 67.8 mg/kg bw/day, n=8.	Weight: *Testis* *Epididymis* *Prostate* *Seminal vesicle* Sperm: *Numbers* *Motility* *Morphology* Histopathology: *Epididymis* *Testis* *Prostate* *Seminal vesicle* Hormones: *Testosterone* ↓ *FSH* ↑ *LH* ↓	Effects on hormone levels	NOAEL = 1086.6 mg/kg bw/day (10000 ppm)	Note: metabolism in skin included. None of the parameters evaluated for butylparaben showed dosage-dependent adverse effects. A significant reduction of testerone levels was not reported.	*Reliability 2.* - Comparable to guideline study, and performed under GLP (good laboratory practice) conditions, - European Union Scientific Committee on Consumer products (SCCP) has evaluated this study report and concluded that the study "cannot be considered as scientifically valid"

	Riad *et al*.	2018	Wistar rats	Repeated dose, oral (p.o.) (start PND 19-21, 8 weeks). Dose: 50 mg/kg BP, n = 6.	Weight: *Testis* *Prostate* *Seminal vesicle* Sperm: *Sperm numbers* ↓ *Sperm motility* ↓ Hormones: *Testosterone* ↓ *Estradiol* ↑ *LH* ↓ *FSH* ↓ *Testosterone/LH* ↓ *Testosterone/Estradiol* ↓ Testis DNA damage ↑ Histopathology testis (affected)	Effects on hormone levels, sperm parameters and testis DNA damage	LOAEL = 50 mg/kg	Butylparaben caused significant elevation of estradiol levels, and significantly decreased testosterone levels.	*Reliability 2.* - Acceptable, well documented publication, which meets basic scientific principles - One shortcoming is that purity of BP is not reported, but is likely of an acceptable purity. - Only single dose investigated
Prepubertal	Alam & Kurohmaru	2014	Sprague-Dawley rats	Reproductive toxicity, oral (single administration) (3 week old male rats). Dose: 1000 mg/kg bw, n= 8.	Histopathology: *Testis - detachment and displacement of spermatogenic cells from Sertolic cells* IHC: *Testis - vimentin filaments were affected 6 and 24 h after exposure. No effect on the microtubule network.*	Effects on testicular histology	LOAEL = 1000 mg/kg bw/day	Evaluation of vimentin filaments, actin and alpha-tubulin (IHC) showed that the Sertoli cell vimentin filaments were affected by exposure, without changes in the microtubule network. Also, histological evaluation (HE) showed detachment and displacement of spermatogenic cells from away from Sertoli cells.	*Reliability 2.* - Acceptable, well-documented study - *In vivo* and *in vitro* study included - *In vivo* study: no dose relationship (single dose included)

	Guerra *et al*.	2017a	Wistar rats	Female reproductive development and uterotrophic assay, s.c (GD12-GD20 and GD12 to end of lactation (PND20)) Dose: 10, 100, or 200 mg/kg (E2 positive control), n =7 (uterotrophic) n = 7-9 (repro dev)Estradiol positive control (10 µg/kg bw)	No effect on uterine weight Positive control (estradiol) ↑ uterine weight. No effect on no of delivered pups, body weight, AGD, nipple retention, vaginal opening (VO), first estrous (or BW at VO and first estrous), estrous cycling. FSH increased at 10 mg/kg/day. No effects were seen on organ weights and BW at PND 75. No effects were seen on no of germ cells (PND 20) or follicles (adulthood). Some effects on sexual behaviour, but not statistically significant (200 mg/kg/day). 50% gestational rate (200 mg/kg/day) but data from pregnant animals were comparable to controls.	No effects on uterine weight. Effects on FSH levels and sexual behaviour.	LOAEL = 10 mg/kg /day	Uterotrophic study; perinatal exposure study	*Reliability 2.* - Acceptable, well-documented study, comparable to guideline standards
	Routledge *et al*.	1998	Alpk:AP rats	Uterotrophic assay. Immature (21-22 days old) and adult (6-8 weeks) OVX rats. Doses and exposure route not entirely clear, but both oral and s.c. exposure. Estradiol used as positive control.	No effect after oral administration Uterine weight ↑ (s.c. exposure) Positive control (estradiol) ↑ uterine weight	Increased uterine weight		The method and results section is a bit difficult to deduce, both in relation to doses used and exposure route.	*Reliability 2.* - Acceptable, design comparable to guideline standards

	Hossaini *et al*.	2000	B6D2 F1 mice & Wistar rats	Uterotrophic assay, s.c. (3 days administration, PND 18-20 in both species). Dose: 100 mg/kg bw/ day (mice) 400, 600 mg/kg bw/day (rats). Estradiol used as positive control (0.1 mg/kg bw/day) for both species	Mice: No effect on uterine weight Rats: Uterine weight ↑	No effects in mice, uterine weight increased in rats	NOAEL = 100 mg/kg bw/day	No effect after s.c exposure in mice (but only one dose tested). In rats, 400 mg/kg bw/day increased uterus wet weight but not weight mg/bw. However, 600 mg/kg bw/ day increased both wet weight and relative weight. Estradiol increased uterus weight in both species.	*Reliability 2.* - Acceptable, well-documented study, comparable to guideline standards
	Lemini *et al*.	2003	CD1 mice & Wistar rats	Uterotrophic assay, s.c. (3 days administration, Immature 21 days old in both species. Adults in OVX study). Dose: 0.7, 7, 21, 70, 210 mg/kg bw in immature mice and rats (0.7 mg/kg bw/day dose not included in OVX mice). Estradiol positive control (10 µg/kg) in both species	Immature: Mice: Uterine weight ↑ Rats: Uterine weight ↑ OVX adult: Uterine weight ↑	Uterine weight increased in all models	NOAEL = 0.7 mg/kg bw/ day, LOAEL 7 mg/kg bw/day	In immature mice uterine weight was increased from 7 mg/kg bw/day and upwards. In OVX mice uterine weight was increased from 21 mg/kg bw/day. In immature rats effects were seen from 70 mg/kg bw/day. The authors write that they noticed that s.c. administration of the high doses diluted in oil formed a subcutaneous deposit, which is likely to affect bioavailability.	*Reliability 2.* - Acceptable, well-documented study, comparable to guideline standards

	Vo & Jeung	2009	Sprague-Dawley rats	Uterotrophic assay, s.c. (3 days administration, immature rats PND 14-16). Dose: 62.5, 250, 1000 mg/kg bw/day, n = 8. Ethinyl estradiol used as positive control (1 mg/kg bw/day)	Uterine weight ↑ Positive control (estradiol) ↑ uterine weight	Increased uterine weight	NOEL = 250 mg/kg bw/day	Uterus weight was increased in the highest dose group as well as in the positive control group. When the anti-estrogen fulvestrant was administered together with the high dose of paraben, the stimulating effect on uterus disappeared indicating that BP uterine effect is mediated via ER dependent pathway.	*Reliability 2.* - Acceptable, well-documented study, comparable to guideline standards
	Ahn *et al*.	2012	Sprague-Dawley rats	Uterus and ovary, s.c. injections (PND 1-7). Dose: 62.5, 250, 1000 mg/kg bw/day, n = 5 Estradiol included as positive control	On PND 8 uterine weight was increased (1000 mg/kg bw/day). No effects seen on ovary weight. Uterine CaBP-9k expression was increased (250 and 1000 mg/kg bw/day). Primordial follicle numbers increased (1000 mg/kg bw/day), early primary follicle numbers decreased (62.5, 250, 1000 mg/kg bw/day), and similar changes were seen with estradiol and propylparaben. Upregulated levels of *Amh, Foxl2, Kitl* (250 and 1000 mg/kg bw/day), downregulated levels of *Star* (250 and 1000 mg/kg bw/day*),* upregulated (250 mg/kg bw/day) and downregulated *Cyp11a1* (1000 mg/kg bw/day).	Adverse effects on the ovary and steroidogenesis.	LOAEL = 62.5 mg/kg bw/day		
	Vo *et al*.	2010	Sprague-Dawley rats	Female reproductive endpoints, oral gavage (PND 21-	No effect on VO, relative uterus weight, relative ovary weight, estrous cycling. A	Adverse effects were seen on the ovary.	LOAEL = 62.5 mg/kg bw/dat		

				40). Dose: 62.5, 250, 1000 mg/ kg bw/day, n = 10/group Estradiol included as positive control	sginificant decrease of corpora lutea was seen in the 62.5 and 1000 mg/kg bw/ day groups. In the uterus, thickness of morphometric measurement was significantly increased in all dose groups.				
Postpubertal exposure	Garcia *et al*.	2017	Sprague-Dawley rats	Sperm parameters, s.c. 57 days, 3 alternating days per week (start 6 weeks old). Dose: 0 (Both naÏve control and vehicle exposed control), 150, 300, 600 mg/kg bw/day, n = 8-10.	***NOTE: NaÏve control used for stat analysis*** Prostate weight ↑ Sperm numbers (affected, ↓↑) Sperm morphology: *Normal* ↓ *Abnormal* ↑	Effects on prostate/testis health and sperm parameters.	LOAEL = 150 mg/kg bw/day	**NOTE**: It seems as if the naïve control (no vehicle) was used for stat analysis. However, some of the endpoints have been analysed in relation to the vehicle control. See paper for details. No effects on butylparaben on bw gain and relevative organ weight changes were reported.	*Reliability 3*. - The study is well documented, but the statistical design is not appropriate for toxicological assessment. Both a naÏve control and a vehicle exposed control were included and statistical analysis seems to be mainly conducted on the naÏve control (and unclear when it is vehicle/naÏve control that is used and if both controls are actually used for analysis of all endpoints) - The results are therefore difficult to interpret

	Pollock *et al*.	2017	CF1 mice	Pharmacokinetic effects E2, s.c. (one injection). Dose: 1, 3, 9, mg (35, 103.3, 310 mg/kg, females) (26.9, 79.5, 242.1 mg/kg, males), n = 10/group.	Urinary estradiol concentrations were measured (both sexes) after BP exposure. In males E2 levels were increased after 3 mg exposure at 8 h. In females estradiol levels were increased after 3 mg exposure at 6, 8, and 10 h.	Effects on estradiol levels	NOAEL = 1 mg (26.9 mg/kg in males)	Kinetic study	*Reliability 5.* - This study is a kinetic study and not a toxicology study and reliability can therefore not be evaluated in terms of toxicology studies - However, contributes with additional information on the potential effect of BP on excretion of estradiol
	Goswami & Kalita	2016	Swiss albino mice	Effects on uterus, s.c., 7 days (adult). Dose: 0, 10, 50, 100 mg/kg bw, n ≥ 5. Estradiol used as positive control (0.001mg/kg bw)	Uterine glands ↑ Uterine weight ↑ Endometrial and mycometrium thickness ↑ Total tissue protein ↑ Histological alterations	Effects on uterus	LOEL = 10 mg/kg bw	Increased number of uterine glands (10, 50, 100 mg/kg bw/day). Increased uterine weight and histological alterations (50 and 100 mg/kg bw/day). Increased endometrial and mycometrium thickness and total tissue protein (100 mg/kg bw/day). Estradiol stimulated all endpoints.	*Reliability 2.* - Acceptable, well-documented study, comparable to guideline standards
	Lemini *et al*.	2004	CD1 mice	Uterotrophic assay, s.c. 3 days (adult). Dose: 0, 70, 210 mg/kg bw, n ≥ 6. Estradiol used as positive control (10 µg/kg bw)	Relative uterine weight ↑	Effect on uterus (increased weight)	LOEL = 70 mg/kg bw/day	The absolute uterine weight was not affected, however relive uterine weight was increased in both dose groups as well as in the positive control.	*Reliability 2.* - Acceptable, well-documented study, comparable to guideline standards

	Shaw & deCatanzaro	2009	CF1 & CD1 mice	Uterotrophic assay, s.c. 3 days (adult). Dose: 0.735, 7.35, and 35 mg/day, n= 7-8. Estradiol used as positive control (500 ng/day). The two mouse strains were tested separately.	No effect on uterine weight Positive control (estradiol) ↑ uterine weight	No effect on uterine weight	NOEL = 35 mg/day	All animals seem to have received the same dose (not taking bodyweight into account).	*Reliability 2.* - Acceptable, well-documented study, comparable to guideline standards
	Lee *et al*.	2017	Sprague-Dawley rats	Female reproductive endpoints, oral (5 weeks, dissolved in corn oil, I assume by gavage) (8 weeks old. Dose: 100 mg/kg/day, n = 6.	Affected estrous cycle length after exposure. Upregulated Amh mRNA levels, but no effect on Foxl2 and Kitlg. Downregulation of Hydroxysteroid dehydrogenase(Hsd)3b1 and Cyp19a1. Downregulation of Lhr. Increased FSH levels in serum and decrease in secondary follicles and Graafian follicles.	Adverse effects on the ovary	LOAEL = 100 mg/kg/day	Poor quality. It is difficult to deduce the actual results. Adult exposure	

	Maske *et al*.	2018	Holtzman rats	Fertility study in F1 females, s.c. (GD6-PND21). Doses: 10, 100, 1000 mg/kg bw/day, n = 15.	F1 females: Increase bw at all time points from birth to PND75 (10 mg/kg bw/day). Delayed VO (100, 1000 mg/kg bw/day). Reduced estrous cycle length (10, 1000 mg/kg bw/day), E2 level reduced at all age measured (100 mg/kg bw/day). Testosterone and progesterone levels were affected at several ages, and significance only found in some cases. Fertility was affected (increased pre- and post-implantation loss at 100, 1000 mg/kg bw/day). Increased number of days before copulation was noted in all exposed groups. Different ovarian follicle types were affected at different ages and effects seen at both 100 and 1000 mg/kg bw/day. Ovarian gene expression of ERα and Star was upregulated (100 mg/kg bw/day. Weight of adrenal gland, hypothalamus, pituitary, ovary, and uterus were all affected at different ages.	Adverse effects on ovary, fertility and hormones.	NOAEL = 10 mg/kg bw/day (estrous cycling)		

In vitro studies

Detailed study information on *in vitro* studies are summarised below:

Category	Reference	Year	System & Method	Relevant endpoints & Effects	Comments/notes	Modality
Androgen/anti-androgen related effects	Alam & Kurohmaru	2014	Primary rat Sertoli cell culture Dose: 1, 100, 1000 µM Duration of exposure: 6 and 24 h	Histological evaluation: Vacuoles in the cytoplasm ↑ Immunohistochemistry (IHC): Disruption of vimentin filaments Vimentin protein expression ↓		A
	Kjærstad *et al*.	2010	AR reporter gene assay (CHO cells) agonism mode (co-exposure with AR agonist R1881) Dose: 0.03-30 µM Duration of exposure: Not reported	AR antagonism (no effect)	Seems as AR agnoism mode was not tested for BP.	A
	Chen *et al*.	2007	AR reporter gene assay (stably transfected HEK 293 cells, (2933Y cells)). Dose: 0.0001, 0.001, 0.01, 0.1, 1, 10 µM Duration of exposure: 16 h	Anti-androgenic activity at highest concentration (10 µM). No androgenic activity.		A

	Pop *et al*.	2018	AR reporter gene assay (transfected MDA-kb2 human breast cancer cells (ATCC CRL-2713)). Dose: 0.5-100 µM (estimated from graph) Duration of exposure: 24 h	Anti-androgenic activity at three highest doses (approximately in the interval 50-100 µM, read from graph). IC50 = 58.5 µM. No androgenic activity	The doses are estimated from the graph and therefore not exact.	A
Proliferation, gene- and protein expression	Khanna & Darbre	2013	MCF-10A human breast epithelial cells Dose: 10 µM Duration of exposure: 17 days	Cell proliferation ↑ Number of colonies ↑ - *Effects similar to estradiol (positive control)*	One concentration (10 µM) tested for cell proliferation and a range tested for number of colonies. Range approximated from graph.	E

	Wróbel & Gregoraszczuk	2014a	MCF-7 human breast cancer cells MCF-10A human breast epithelial cells Dose: 0.02 µM Duration of exposure: 24 hours	MCF-7: Genes related to G1/S-phase and cell cycle progression inhibitors (affected) G2/M phase Corresponding proteins (no effect) Apoptosis regulation, caspase 8 ↑ MCF-10A: Genes related to G1/Sphase, G2/M phase and cell cycle progression inhibitors (affected) Proteins Cyclin D, Cyclin E, Cyclin A upregulated, p21 ↓ Cyclin B1 (no effect) Apoptosis regulation, Bcl-XL ↑ *- Generally, the effects were similar to estradiol (positive control), but not for all endpoints*		E
	Zhang *et al*.	2013b	hERRγ transfected into *E.coli* Dose: 0.0001, 0.001, 0.01, 0.1, 1, 10, 100 µM Duration of exposure: Not reported	Inverse antagonism on ERRγ ↑	LOEL 0.1 µm, relative effect concentration (50%) set to 0.309 µM	E

	Charles & Darbre	2013	MCF-7 human breast cancer cells Dose: Not reported Duration of exposure: 7 and 14 days	Effects on proliferation compared to E2: After 7 days Lowest observed effect concentration (LOEC): 0.7 µM No observed effect concentration (NOEC): 0.5 µM EC50 2 µM After 14 days LOEC 0.5 µM NOEC 0.2 µM EC50 1 µM - *Effects similar to estradiol (positive control)*	The range of concentrations tested is not reported.	E

	Khanna *et al*.	2014	MCF-7 human breast cancer cells, T-47-D human breast cancer cells, ZR-75-1 human breast cancer cells Dose: 10 µM Duration of exposure: 7 days and 20 weeks	MCF-7: Motility: 7 days 20 weeks ↑ (increase greater than with E2) Motility after co-exposure with anti-estrogen ↓ Migration ↑ Matrix degradation ↑ Protein expression of E-cadherin, β-catenin: 7 days 20 weeks ↓ Protein expression of ERα: 7 days ↓ 20 weeks: lower levels than under E2 deprivation conditions and only slightly higher than when the cells were maintained with E2. T-47-D : Motility: 7 days 20 weeks ↑ ZR-75-1: Motility: 7 days ↑ 20 weeks ↑ *- Effects similar to estradiol (positive control)*		E

	Wróbel & Gregoraszczuk	2014b	MCF-7 human breast cancer cells MCF-10A human breast epithelial cells Dose: 0.02 µM Duration of exposure: 6, 24, 48 and 72 hours	MCF-7: ERα gene exp ↑ (24h) ERα protein exp ↑ (48h) ERβ gene exp ↑ (24h) ERβ protein exp ↑ (48h) Co-exp with ER blocker prevented effect for both receptors PGR gene exp ↑ (24h) PGR protein exp MCF-10A: ERα gene exp ERα protein exp ERβ gene exp ERβ protein exp ↑ (72h) Co-exp with ER blocker did not affect the effect seen on protein exp at 72h PGR gene exp ↑ (24h) PGR protein exp *- Generally, the effects were similar to estradiol (positive control), but not for all endpoints*	Progesterone receptor (PGR)	E

	Wróbel & Gregoraszczuk	2015	MCF-7 human breast cancer cells MCF-10A human breast epithelial cells Dose: 0.02 µM Duration of exposure: 6, 24, 48 and 72 hours (gene and protein), 15 min (cAMP), 0-120 min (phosphorylation ERK1/2 and AKT)	MCF-7: GPR30 gene exp ↑ (24h) GPR30 protein exp cAMP levels Phosphorylation of ERK1/2 ↑ Phosphorylation of AKT ↓ (only shortly, stabilized again) MCF-10A: GPR30 gene exp ↑ (24h) GPR30 protein exp ↑ (72h, different from E2) cAMP levels Phosphorylation of ERK1/2 ↑ Phosphorylation of AKT *- Effects similar to estradiol (positive control) except for GPR30 protein upregulation at 72h in the MCF-10A cells.*	BP acted similar to E2 on these endpoints, except for GPR30 protein upregulation at 72h in the MCF-10A cells. G protein-coupled estrogen receptor 1 (GPER) is also known as G protein-coupled receptor 30 (GPR30). Responsible for some of the rapid effects of estradiol.	E

	Gonzalez *et al*.	2018	MCF-7 & T47D human breast cancer cell lines, LNCaP, MDA-MB-231 (Erα negative) Dose: Different doses within the range 0.00001 - 30 µM (BP and metabolite 3-OH) Duration of exposure: various lengths depending on endpoint	Proliferation: MCF-7 ↑ (3-OH) T47D ↑ (3-OH) LNCaP no effect (3-OH) MDA-MB-231 (Erα negative) no effect (BP and 3-OH) Co-exposure with anti-estrogen (proliferation): MCF-7 ↓ (BP and 3-OH) Gene expression GREB ↑ Co-exp with anti-estrogen (gene expression GREB1) ↓ Reporter gene assay (ERE transcriptional activity) ↑ *- Estrogenic activity (estradiol used as positive control for some endpoints and similar effects seen)*	GREB1 is a critical downstream target of Erα signaling. 3-OH seem to be less potent than BP.	E

	Pop *et al*.	2018	T47D-Kbluc and MCF-7 breast cancer cells Dose: 0.3-100 µM Duration of exposure: 24 h (reporter gene assay), 72 h (proliferation)	T47D (reporter gene assay, estrogen sensitive): Low dose ↑ (estrogenic response) High dose ↓ (anti-estrogenic response) T47D (reporter gene assay, antagonist mode by presence of E2): High dose ↓ (anti-estrogenic response) MCF-7 (proliferation): Low dose ↑ (estrogenic response) High dose ↓ (anti-estrogenic response) MCF-7 (proliferation, antagonist mode by presence of E2): High dose ↓ (anti-estrogenic response) - *Estrogenic activity at lower concentrations and anti-estrogenic at higher concentrations*	T47D-Kbluc cells naturally expresses Erα and Erβ and stably transfected with a triplet estrogen-response element - promoter - luciferase reporter gene construct)	E

	Yang *et al*.	2011	Pituitary lactosomatotrophic GH3 cells Dose: 10 µM Duration of exposure: 24 h	Gene expression (same pattern as E2): CaBP-9k ↑ Erα Progesterone receptor B ↑ Protein expression (same pattern as E2, except progesterone receptor where E2 upregulated): CaBP-9k ↑ Erα Progesterone receptor B *- Effects similar to estradiol (positive control)*	The CaBP-9k gene is controlled by sex hormones and is upregulated by estrogen and down regulated by progesterone during the estrous cycle and early pregnancy in the rat uterus. This indicates that the pathway may not be through Erα, but maybe progesterone receptor.	E

	Vo *et al*.	2011	Pituitary lactosomatotrophic GH3 cells Dose: 0.1, 1, 10, 100 µM and 100 µM on its own Duration of exposure: 24 h	Gene and protein expression: CaBP-9k ↑ Erα (no clear effects) Progesterone receptor ↑ Co-treatment with anti-estrogen: CaBP9k and PR same as vehicle control (BP cannot induce a response) ERE activity is induced by exposure and co-traeatment with antiestrogen reduced the response on ERE activity - *Effects similar to estradiol (positive control)*	The CaBP-9k gene is controlled by sex hormones and is upregulated by estrogen and down regulated by progesterone during the estrous cycle and early pregnancy in the rat uterus. The results indicate that the effect of BP on CaBP-9k and PR expression may be due to binding to and signalling through estrogen receptor.	E
	Watanabe *et al*.	2013	Erα, Erβ, AR transactivation assays Dose: 0.01-10 µM Duration of exposure: 24 h	Erα and Erβ agonistic response ↑ - *Effects similar to estradiol (positive control)* Erα relative effect concentration (REC) 20%: 0.29 µM Erβ elative effect concentration (REC) 20%: 0.15 µM AR antagonism (no response) Hydroxylated isomers showed lower, or no effect than BP		E

	Van Meeuwen *et al*.	2008	MCF-7 human breast cancer cells Microsomes from human placenta Dose: 0.0001-100 µM (proliferation), 0.000001-0.1 µM (inhibited proliferation), 0.1-1000 µM (aromatase inhibition) Duration of exposure: 6 days (MCF-7), 45 minutes (microsomes)	Proliferation: MCF-7 ↑ Co-exposure with anti-estrogen (proliferation): MCF-7 ↓ Aromatase inhibition: Enzyme activity ↓	The proliferative stimulation produced by BP was via ER, as shown by co-exposure with an anti-estrogen. Aromatase activity in human placental microsomes was reduced after exposure.	E, S
Hormone production & aromatase	Taxvig *et al*.	2008	H295R steroidogenesis assay Dose: 0.1, 0.3, 1, 3, 10, 30 µM Duration of exposure: 48 h	Hormone levels: Progesterone ↑ (30µM) Estradiol Testosterone	Maybe a tendency towards decrease in E and T levels at the highest concentration (30 µM)	E, T, S
	Wróbel & Gregoraszczuk	2013	MCF-7 human breast cancer cells MCF-10A human breast epithelial cells Dose: 0.0002, 0.002, 0.02, 0.2, 2 µM Duration of exposure: 24 and 72 hours	MCF-7: Gene and protein expression Cyp19a1 Estradiol excretion ↓ (lowest dose only) MCF-10A: Gene and protein expression Cyp19a1 ↓ Estradiol excretion ↓		E, S

	Guerra *et al*.	2016	Primary culture of pre-antral mouse follicles and primary human granulosa cell cultures Dose: 0.01, 0.1, 1, 10 µM Duration of exposure: up to 12 h for follicles, up to 96 h for granulosa cells	Morphology/growth/developmental pattern of follicles (no effect) Estradiol production from follicles (not affected) Protesterone production from granulosa cells (no effect)		E, S
	Williams *et al*.	2019	MCF-7 and ZR-75-1 breast cancer cells and HMF3A breast fibroblast (Erα negative)	MCF-7, ZR-75-1 , HMF3A: Cyp19a1 gene expression ↑ Aromatase activity ↑ Estradiol ↑ MCF-7, ZR-75-1: Proliferation ↑ Co-exposure with aromatase inhibitor proliferation ↓	These data indicate that BP act on aromatase to induce estradiol synthesis and thereby stimulate Erα positive cell proliferation.	E, S
	Klopčič *et al*.	2015	MDA-kb2 stably transformed with murine mammalian tumor virus luciferase Dose: 0.01 and 1 µM Duration of exposure: 24 h	Glucocorticoid-like activity ↑	Co-treatment with flutamide was used to ensure that the registered effects was not due to androgen agonist activity.	S

Hormone metabolism	Prusakiewicz *et al*.	2007	Cytosolic protein from skin, pooled human liver cytosolic fraction, normal human epidermal keratinocytes (NHEKs) Dose: 1 - 1000 µM Duration of exposure: 30 min	Inhibition of estradiol sulfation (SULT) (arrow ↓ indicates inhibition) Liver cytosol ↓ Skin cytosol ↓ NHEK cells ↓ Inhibition of estrone sulfation (SULT) NHEK ↓ Inhibition of DHEA sulfation (SULT) NHEK	Experiments showed that the results were not confounded by hydrolysis of BP	S
	Engeli *et al*.	2017	Lysate of human embryonic kidney cells (HEK-293) Dose: 20 µM Duration of exposure: not reported	Activity of 17β-HSD2 ↓ (estradiol to estrone) Activity of 17β-HSD1 ↓ (estrone to estradiol)	Inhibition of 17β-HSD2 prevents local inactivation of estradiol.	S

Annex IV – Data overview (ED Guidance table)

Annex IV is provided as a separate Excel file.

Annex E _ EDGD table_May2020.xlsm

Annex V – Human epidemiology studies

Study name:	The Puerto Rico Testsite for Exploring Contamination Threats (PROTECT)
Paper title:	Phenols and Parabens in relation to Reproductive and Thyroid Hormones in Pregnant Women
Authors:	Aker *et al*
Publication year:	2016

		Score
Study design and conduct		Low
study type	prospective	
study year	2010-2012	
additional comments related to study design	Although data is from a prospective cohort of pregnant women it is in this manuscript treated as cross-sectional data (it investigates associations between exposure and outcome biomarkers measured at the same time) although it also takes advantage of the availability of samples taken at two different timepoints during pregnancy in linear mixed model.	
Study population		Moderate
sampling method	convenience sample, representative	
study size	106	
age range	18-40 years	
sex	pregnant women	
other population characteristics	pregnant women sampled at two time points: within 16-20 Gestational week (GW) and 24-28 GW	
quality of provided information on population characteristics	well defined	
additional comments related to study population	Study population seems to be representative of pregnant women in the recruitment area. More information on recruitment and inclusion available in previous publications. Relevant exclusion criteria used. Moderate score mainly due to the moderate study size. Note that the study population in this paper is from the same study as in Aker *et al* 2019.	
Exposure assessment		High
method(s) used for exposure assessment	biomonitoring, human samples	
Validated methods used?	Yes	
Is the timing between exposure and outcomes assessment appropriate?	Yes	
If HBM:		
matrix and sample type	urine, random spot	
Validated biomarker measured?	Yes	
Adjusted for urinary dilution?	yes, specific gravity	
measured concentrations (median, range)	not presented, but the inter-quartile range is given as 3.30 ng/mL	
% samples with BP <LOD/LOQ	not reported	

LOD/LOQ for BP	0.1 ng/mL	
additional comment related to exposure assessment	Exposure biomarkers analysed at CDC with high quality validated method (online solid phase extraction- HPLC-isotope dilution MS/MS).	
		Score
Outcome assessment		High
outcome(s) assessed	Maternal serum hormone levels at time of sampling: estradiol, progesterone, SHBG, estradiol/progesterone ratio, Thyroid-stimulating hormone (TSH), fT3, fT4	
quality of outcome assessment	standardised and validated outcome assessment	
additional comment related to outcome assessment	fT3 and fT4 measured by LC-MS/MS following ultrafiltration step which is gold standard for measuring free thyroid hormones.	
Confounder control		Moderate
Is information available for confounders relevant to the scientific questions asked? (comment if needed)	Yes Only mentioned which confounders were included in the final model (variables found to change the main effect estimate >10% were retained in the final models) - not complete list of which confounders were considered/tested. E.g was parity tested but not changing main effect estimate?	
Are confounders clearly indicated?	Yes	
Are confounders adequately controlled for?	Yes	
additional comment related to confounder control		
Statistical analysis		Moderate
methods used for investigating associations	Linear Mixed models with the study ID as random factor allowing including data from both sampling time points. Also cross-sectional analyses by multiple linear regression models stratified by sampling time (visit) also performed as sensitivity tests. Correlation between continuous variables by Pearson correlation Coefficients.	
Suitability of used methods?	Yes	
Maximised use of data?	Yes	
only descriptive statistics or/and bivariate analysis	No	
Appropriate control for confounders?	Yes	
Unadjusted and adjusted estimates presented?	only adjusted estimates presented	
Sensitivity tests and interaction analysis conducted?	Yes	
multiple testing issues	Yes	
additional comment related to the statistical analysis	Statistical power is probably an issue because of the relatively small study size. Also risk of Type I error (chance findings) due to multiple testing. Due to these issues the statistical analysis only scores Moderate.	

		Score
Reporting		Moderate
Key elements of M&M and results are reported in sufficient detail?	Yes	
A plausible mechanism for the association under investigation is provided?	No	
Are the conclusion made justified by the data shown?	Yes	
additional comment related to the reporting of the study	Actual measured levels of BP and number of samples with BP above LOD (limit of detection) were not presented in the paper (also not in supplementary material). This makes it difficult to judge if the results of the study can be translated to be representative also for other populations as it is not clear if this is a particular highly exposed population or not. A plausible mechanism cannot clearly be given. The results are discussed appropriately in relation to previous human as well as *in vitro* and animal studies. However, there is not clear consistency between studies and therefore no clear plausible mechanism although effects on steroid production is suggested.	
Key findings		
What are the key findings?	Maternal urinary BP levels were significantly negatively associated with maternal serum levels of estradiol (-8.5% per IQR, p=0.05) and the estradiol/progesterone ratio (-9.3% per IQR, p=0.04 (no effect on progesterone)) and significantly positively associated with maternal serum levels of fT4 (5.6% per IQR, p=0.01)	
Any secondary findings?	BP was only moderately correlated to methylparaben and propylparaben and not correlated to any of the other measured phenols	
effect size in relation to biological relevance	Preeclampsia (PE) is associated with 50-70% decreased estrogen levels and aberrant placenta production of estrogens could play a key role in PE symptoms. While the effect estimates observed in this study for the association to estradiol (8-9% decrease in estradiol per IQR increase in BP) is not of the same size it cannot be excluded that it may affect placenta function (or reflect placenta function). The biological relevance of an isolated association to increased fT4 is uncertain. If this change was biologically relevant a concurrent lowering of TSH would have been expected.	
	overall score:	Moderate

Study name:	The Puerto Rico Testsite for Exploring Contamination Threats (PROTECT)
Paper title:	The associations between prenatal exposure to triclocarban, phenols and parabens with gestational age (GA) and birth weight in northern Puerto Rico
Authors:	Aker *et al*.
Publication year:	2019b

		Score
Study design and conduct		High
study type	prospective	
study year	2011-2017	
additional comments related to study design		
Study population		High
sampling method	convenience sample, representative	
study size	922 participants	
age range	18-40	
sex	pregnant women	
other population characteristics	pregnant women with urine sampled up to three time points: within 16-20 ; 20-24 and 24-28 GW	
quality of provided information on population characteristics	well defined	
additional comments related to study population	They write that 922 women are included but most models reported are based on max 750 cases. Note it is the same study population as in Aker *et al* 2016 and Aker *et al* 2019a	
Exposure assessment		High
method(s) used for exposure assessment	biomonitoring, human samples	
Validated methods used?	Yes	
Is the timing between exposure and outcomes assessment appropriate?	Yes	
If HBM:		
matrix and sample type	urine, random spot	
Validated biomarker measured?	Yes	
Adjusted for urinary dilution?	yes, specific gravity	
measured concentrations (median, range)	median 0.2 ng/mL, IQR= 1.1 ng/mL	
% samples with BP <LOD/LOQ	26%	
LOD/LOQ for BP	0.1 ng/mL	
additional comment related to exposure assessment	They sample at up to three different time points during pregnancy giving a better estimate of exposure. Measured at CDC by online solid phase extraction- HPLC- isotope dilution MS/MS.	

		Score
Outcome assessment		High
outcome(s) assessed	gestational age (GA), preterm birth, birth size (SGA (small for GA), large for GA) LGA)	
quality of outcome assessment	standardised and validated outcome assessment	
additional comment related to outcome assessment		
Confounder control		High
Is information available for confounders relevant to the scientific questions asked?	Yes	
(comment if needed)	yes	
Are confounders clearly indicated?	Yes	
Are confounders adequatley controlled for?	Yes	
additional comment related to confounder control	Although regarding specific gravity of the individual urine samples: as they have up to three urine samples per woman they calculate the average biomarker concentration for use in models - and also include the average specific gravity as a covariate. In this situation it would probably be more correct to adjust the individual measurements for the specific gravity (SG) of the same sample and then use the average of the adjusted concentrations in the models. In this paper the authors mention that they did consider parity as a confounder but it did not affect the effect estimates.	
Statistical analysis		High
methods used for investigating associations	Multiple linear regression to examine associations to continuous outcome variables (gestational age, birth weight) and logistic regressions models for binary outcome variables (preterm or not, SGA or not, LGA or not)	
Suitability of used methods?	Yes	
Maximised use of data?	Yes	
only descriptive statistics or/and bivariate analysis	No	
Appropriate control for confounders?	Yes	
Unadjusted and adjusted estimates presented?	only adjusted estimates presented	
Sensitivity tests and interaction analysis conducted?	Yes	
multiple testing issues	Yes	
additional comment related to the statistical analysis	Due to the potential risk of Type I error due to multiple testing it could be argued that the scoring should be only moderate. But it is a large study with thorough and sound statistical analysis. It could possibly be a more maximal use of data to not average biomarker exposure but use mixed models with subject ID as a random factor as they did in the two other papers. Also slight concern if it is appropriate to use average SG as covariate in the models.	

		Score
Reporting		High
Key elements of M&M and results are reported in sufficient detail?	Yes	
A plausible mechanism for the association under investigation is provided?	Yes	
Are the conclusion made justified by the data shown?	Yes	
additional comment related to the reporting of the study	They suggest that the associations they observed between paraben exposure and maternal hormones could be related also to the decreased risk of SGA seen here (and increased GA seen for methyl- and propyl paraben). They also mention that BP *in vitro* has been shown to activate adipogenesis as a potential explanation of the decreased risk of being born SGA.	
Key findings		
What are the key findings?	BP (and methyl and propyl paraben) were all significantly associated with a lower risk of being born SGA. BP was not associated to GA or risk of being born preterm although both methylparaben and propylparaben was significantly associated to 1.5-2 day longer GA/IQR and lower risk of being preterm. For MP and PP the association was strongest to exposure biomarkers from urine collected GW 16-20. Fetal sex did not modify any of these associations.	
Any secondary findings?		
effect size in relation to biological relevance	Being born SGA is a risk factor for many conditions later in life so a decreased risk of being born SGA may be biologically relevant.	
	overall score:	High

Study name:	The Puerto Rico Testsite for Exploring Contamination Threats (PROTECT)
Paper title:	A repeated measures study of phenol, paraben and triclocarban urinary biomarkers and circulating maternal hormones during gestation in the Puerto Rico PROTEC cohort
Authors:	Aker *et al*
Publication year:	2019a

		Score
Study design and conduct		Low
study type	prospective	
study year	2012-2017	
additional comments related to study design	Although data is from a prospective cohort of pregnant women it is in this manuscript treated as cross-sectional data (it investigates associations between exposure and outcome biomarkers measured at the same time) although it also takes advantage of the availability of samples taken at two different time points during pregnancy in linear mixed model.	
Study population		High
sampling method	convenience sample, representative	
study size	602	
age range	18-40 years (mean 26.5y)	
sex	Pregnant women	
other population characteristics	pregnant women sampled at two timepoints: within 16-20 GW and 24-28 GW	
quality of provided information on population characteristics	well defined	
additional comments related to study population	Note that the study population in this paper is from the same study as in Aker *et al* 2016, and Aker *et al* 2019b	
Exposure assessment		Moderate
method(s) used for exposure assessment	biomonitoring, human samples	
Validated methods used?	Yes	
Is the timing between exposure and outcomes assessment appropriate?	Yes	
If HBM:		
matrix and sample type	urine, random spot	
Validated biomarker measured?	Yes	
Adjusted for urinary dilution?	yes, specific gravity	
measured concentrations (median, range)	1st sampling: median: 0.25 ng/mL (95%ile: 39.1); 2nd sampling: median: 0.2 ng/mL (95%ile: 32.6)	
% samples with BP <LOD/LOQ	24% and 34%, respectively, at the two sampling times	

LOD/LOQ for BP	0.1 ng/mL	
additional comment related to exposure assessment	Measured at CDC by online solid phase extraction- HPLC-isotope dilution MS/MS. Based on Table 2 the IQR were respectively 2.1 ng/ml and 0.81 ng/ml at the two timepoints. This is lower than the IQR of BP in Aker *et al*. 2016.	
		Score
Outcome assessment		High
outcome(s) assessed	Maternal serum hormone levels at time of sampling: estriol, progesterone, testosterone, SHBG, progesterone/estriol ratio, corticotropin-releasing hormone, TSH, T3, fT3, T4, fT4	
quality of outcome assessment	standardised and validated outcome assessment	
additional comment related to outcome assessment	Since one of the significant associations observed in Aker 2016 was between BP and estradiol it is surprising that maternal estradiol is not included in the outcomes studied in this paper. fT4 was in this study measured by a direct immunoassay, which may be less accurate than methods separating free and bound T4 before analysis especially during pregnancy where thyroid binding globulin is increased.	
Confounder control		Moderate
is information available for confounders relevant to the scientific questions asked? (comment if needed)	Yes	
	yes - but only the covariates included in the final model are listed (specific gravity, study visit, BMI at first visit, maternal age, passive smoking, socio-economy). So unclear if e.g. parity was considered but tested and shown not to affect the effect estimates < 10 %.	
are confounders clearly indicated?	Yes	
are confounders adequatly controlled for ?	Yes	
additional comment related to confounder control	Lack of description of the full list of confounders considered. Since outcomes are maternal hormone levels and parity is known to influence pregnancy steroid levels it seems a relevant confounder to test. Maybe they did and found that it did not change the effect estimate but they do not describe if they did.	
Statistical analysis		Moderate
methods used for investigating associations	Linear Mixed models with the study ID as random factor allowing to include data from both sampling time points. Also cross-sectional analyses by multiple linear regression models stratified by sampling time (visit) also performed as sensitivity tests. Correlation between continuous variables by Pearson correlation Coefficients.	
suitability of used methods?	Yes	
maximised use of data?	Yes	
only descriptive statistics or/and bivariate analysis	No	
appropriate control for confounders?	Yes	

unadjusted and adjusted estimates presented?	only adjusted estimates presented	
sensitivity tests and interaction analysis conducted?	Yes	
multiple testing issues	Yes	
additional comment related to the statistical analysis	For GW 24-28 it is stated in table 2 that almost 34% of the samples had BP < LOD yet the 25%ile is set to 0.1 (not <LOD). It is unclear how values below LOD were handled in the statistical analysis and how data could be categorized into quartiles. Risk of type I error (chance findings) due to multiple testing. For these reasons the statistical analysis only score moderate, even though this is a relatively large study and methodology is sound.	
		Score
Reporting		Moderate
key elements of M&M and results are reported in sufficient detail?	Yes	
a plausible mechanism for the association under investigation is provided?	No	
are the conclusion made justified by the data shown?	No	
additional comment related to the reporting of the study	No conclusion is made on the observed association of BP and decreased maternal SHBG and testosterone. It is strange that the authors did not include estradiol in the analyses as, in a preliminary study including only 106 women from PROTECT, they observed a significant association between BP and decreased estradiol. Lack of reporting full list of confounders considered.	
Key findings		
what are the key findings?	BP was associated with a decrease in maternal SHBG (-5.3% change per BP IQR, p=0.01). Also associated with a tendency of lower testosterone (-6.8% change per BP IQR, p=0.06) and estriol (-5.2%, p=0.13).	
any secondary findings?	Note that the association of BP with increased fT4 observed in Aker *et al* 2016 could not be seen in this larger study of the same study population. Also the association with SHBG seen in Aker *et al* 2019 was not evident at all in the smaller study group described by Aker *et al* 2016. Thus, there is not consistency between the findings of the smaller study (Aker *et al* 2016) and this larger study (Aker *et al* 2019) of the same study population. The only observations that hint in the same direction is the tendency of BP to be associated with lower maternal sex steroids. BP only moderately associated to Methylparaben and propyl paraben in accordance with Aker *et al* 2016.	
effect size in relation to biological relevance	uncertain	
	overall score:	Moderate

Study name:	Center for the health assessment of mothers and children of Salina (CHAMACOS)
Paper title:	Associations of maternal exposure to triclosan, parabens, and other phenols with prenatal maternal and neonatal thyroid hormone levels
Authors:	Berger *et al*
Publication year:	2018

		Score
Study design and conduct		Moderate
study type	longitudinal	
study year	1999-2000	
additional comments related to study design	Although it is a longitudinal study part of the study is analysed cross-sectionally (e.g exposure biomarkers vs. Maternal hormone levels). Urine collected at two timepoints during pregnancy: GW 14 (5-28) and 27 (21-39).	
Study population		High
sampling method	convenience sample, representative	
study size	338 (out of 601 totally enrolled)	
age range	<18 years	
sex	pregnant women	
other population characteristics	Additional inclusion criteria: qualified for California's low-income health insurance program (MediCal)	
quality of provided information on population characteristics	well defined	
additional comments related to study population	Study population generally from low income households (61% < federal poverty level).	
Exposure assessment		High
method(s) used for exposure assessment	biomonitoring, human samples	
validated methods used?	Yes	
Is the timing between exposure and outcomes assessment appropriate?	Yes	
If HBM:		
matrix and sample type	urine, random spot	
validated biomarker measured?	Yes	
adjusted for urinary dilution?	yes, specific gravity	
measured concentrations (median, range)	median (SG adjusted) 0.3 ng/mL, 95th%ile: 27.2 ng/mL	
% samples with BP <LOD/LOQ	44%	
LOD/LOQ for BP	0.2 ng/mL	
additional comment related to exposure assessment	Measured at CDC by solid phase extraction HPLC-MS/MS.	

		Score
Outcome assessment		High
outcome(s) assessed	Maternal thyroid hormone levels, neonatal TSH levels	
quality of outcome assessment	standardised and validated outcome assessment	

additional comment related to outcome assessment	fT4 measured by RIA following direct equilibrium dialysis (more precise than direct measurement by immunoassay)	
Confounder control		High
is information available for confounders relevant to the scientific questions asked? (comment if needed)	Yes yes, confounders considered: maternal age, education and country of birth and household income.	
are confounders clearly indicated?	Yes	
are confounders adequatly controlled for ?	Yes	
additional comment related to confounder control		
Statistical analysis		Low
methods used for investigating associations	Multiple linear regression. In addition to models adjusting for confounders also models adjusting for confounders + other chemical exposures were constructed with chemicals selected based on Bayesian Model Averaging.	
suitability of used methods?	Yes	
maximised use of data?	No	
only descriptive statistics or/and bivariate analysis	No	
appropriate control for confounders?	Yes	
unadjusted and adjusted estimates presented?	both unadjusted and adjusted estimates presented	
sensitivity tests and interaction analysis conducted?	Yes	
multiple testing issues	Yes	
additional comment related to the statistical analysis	They have two pregnancy measurements of parabens but use the average of the two measurements in the statistical analysis. No clear description on how they handle the many samples with BP values below LOD in the statistical analysis. Risk of Type I error due to multiple testing but also risk of Type II error due to low detection rate leading to lower statistical power	
		Score
Reporting		Low
key elements of M&M and results are reported in sufficient detail?	No	
a plausible mechanism for the association under investigation is provided?	No	
are the conclusion made justified by the data shown?	Yes	
additional comment related to the reporting of the study	Missing information on how they in the statistical analysis handle values below LOD	
Key findings		
what are the key findings?	No significant association of BP with maternal thyroid hormones nor with neonate thyroid hormones	
any secondary findings?		
effect size in relation to biological relevance	Not relevant	
	overall score:	Low

Study name:	Longitudinal investigation of fertility and the environment (LIFE) study
Paper title:	Endocrine Disrupting Chemicals in Seminal Plasma and Couple Fecundity
Authors:	Germaine M. Buck Louis, Melissa M. Smarr, Liping Sun, Zhen Chen, Masato Honda, Wei Wang, Rajendiran Karthikraj, Jennifer Weck, and Kurunthachalam Kannan
Publication year:	2018

		Score
Study design and conduct		High
study type	prospective	
study year	2005-2009	
additional comments related to study design	Male partners in couples participating in a study of pregnancy planners	
Study population		Moderate
sampling method	Other	
study size	501 males of male-female couples; 339 with residual seminal fluid from second semen sample for endocrine disrupting chemical (EDC) quantification	
age range	≥18 years, reproductive age	
sex	males	
other population characteristics	Inclusion criteria: male partners aged ≥18 years; in a committed relationship; no physician-diagnosed infertility; ability to communicate in English or Spanish; and couple off contraception for ≤2 months.	
quality of provided information on population characteristics	well defined	
additional comments related to study population	Male partners in couples participating in a study of pregnancy planners. Recruitment through fishing and hunting license registries and marketing databases for these interests.	
Exposure assessment		Moderate
method(s) used for exposure assessment	biomonitoring, human samples	
validated methods used?	yes	
Is the timing between exposure and outcomes assessment appropriate?	yes	
If HBM:		
matrix and sample type	seminal plasma	
validated biomarker measured?	yes	
adjusted for urinary dilution?	not relevant	
measured concentrations (median, range)	median 0.02 ng/mL (IQR <LOD - 0.07)	
% samples with BP <LOD/LOQ	45%	
LOD/LOQ for BP	0.012 ng/mL	

additional comment related to exposure assessment	Seminal plasma generally not the matrix used to measure exposure especially as levels generally are lower in seminal plasma than in e.g. urine, but may be relevant in a study of TTP. Participants also provided urine samples and other chemicals have in this study previously been measured in urine. It is unclear why urine was not used for exposure assessment in this paper.	
		Score
Outcome assessment		High
outcome(s) assessed	Time to pregnancy (TTP)	
quality of outcome assessment	standardised and validated outcome assessment	
additional comment related to outcome assessment		
Confounder control		Moderate
is information available for confounders relevant to the scientific questions asked? (comment if needed)	Yes yes, confounders included: male age, BMI, and serum cotinine (biomarker for smoking).	
are confounders clearly indicated?	Yes	
are confounders adequatly controlled for ?	Yes	
additional comment related to confounder control	Medical history of the men (data available from questionnaires) did not seem to be considered as confounder or exclusion criteria.	
Statistical analysis		Moderate
methods used for investigating associations	Cox discrete-time survival analytic techniques were utilised to estimate fecundability odds ratios (FORs) and corresponding 95% confidence intervals (CIs), with separate models run for each chemical	
suitability of used methods?	Yes	
maximised use of data?	Yes	
only descriptive statistics or/and bivariate analysis	No	
appropriate control for confounders?	Yes	
unadjusted and adjusted estimates presented?	both unadjusted and adjusted estimates presented	
sensitivity tests and interaction analysis conducted?	No	
multiple testing issues	Yes	
additional comment related to the statistical analysis	Multiple testing issues not addressed. Many samples below LOD weakens the statistical power. In this study they also have information on medical and reproductive history of participants from questionnaires. It might have been relevant to exclude men with a known cause of infertility/subfertility (e.g. history of cryptorchidism) from this study looking at associations between male exposure and TTP.	
		Score
Reporting		Moderate
key elements of M&M and results are reported in sufficient detail?	Yes	

a plausible mechanism for the association under investigation is provided?	No	
are the conclusion made justified by the data shown?	Yes	
additional comment related to the reporting of the study		
Key findings		
what are the key findings?	No association of any EDC in seminal fluid and TTP	
any secondary findings?	No	
effect size in relation to biological relevance		
	overall score:	Moderate

Study name:	
Paper title:	Bisphenol A and other phenols in human placenta from children with cryptorchidism or hypospadias
Authors:	Mariana F. Fernández, Juan P. Arrebola, Inmaculada Jiménez-Díaz, José María Sáenz, José Manuel Molina-Molina, Oscar Ballesteros, Andreas Kortenkamp, Nicolás Olea
Publication year:	2015

		Score
Study design and conduct		Moderate
study type	Case-control	
study year	2000 to 2002	
additional comments related to study design	The study is a nested case-control study in a prospective cohort.	
Study population		Moderate
sampling method	Random, representative sample	
study size	51 controls + 28 cases	
age range	Newborns examined at birth (cases were also examined at 1 month)	
sex	Male	
other population characteristics	-	
quality of provided information on population characteristics	Well defined	
additional comments related to study population	Although the case-control ratio in the study was 1:3, they were not able to find more than 114 controls that matched the criteria (parity, GA and DOB), and only 51 out of those had adequate biological sample material to be included. There is no power calculation, but based on the confidence intervals of the ORs, the study seems under-powered.	
Exposure assessment		Moderate
method(s) used for exposure assessment	Biomonitoring, human samples	
validated methods used?	Yes	
Is the timing between exposure and outcomes assessment appropriate?		
If HBM:		
matrix and sample type	Placenta	
validated biomarker measured?		
adjusted for urinary dilution?	Not applicable	
measured concentrations (median, range)	0.44 ng/g placenta (Limit of quantification, LOQ-1.60 ng/g placenta)	
% samples with BP <LOD/LOQ	27%	
LOD/LOQ for BP	0.06 ng/g	
additional comment related to exposure assessment	Uncertain if BP levels in placenta tissue can be considered as a validated biomarker for paraben exposure but it is a relevant tissue for assessing fetal exposure. Exposure assessment was done in term placenta and the sensitive window of development related to cryptorchidism and hypospadias are in 1-2 trimester.	

		Score
Outcome assessment		High
outcome(s) assessed	Congenital cryptorchidism and/or hypospadia	
quality of outcome assessment	well defined	
additional comment related to outcome assessment	outcomes assessed by examination	
Confounder control		Moderate
is information available for confounders relevant to the scientific questions asked? (comment if needed)	Yes Cases and controls were matched on gestational age, date of birth and parity. Confounders included in the final adjusted model: maternal age and newborn birthweight.	
are confounders clearly indicated?	Yes	
are confounders adequatly controlled for?		
additional comment related to confounder control	The authors state that confounding variables were selected if significantly associated with outcomes in bivariate analyses or changed the beta coefficient by >20% in the multivariable analysis. However, they do not report the full list of confounders that they considered and tested. The authors also discuss that the relatively small sample size prevented adjustment for some potential con-founders and residual confounding therefore cannot be ruled out.	
Statistical analysis		Moderate
methods used for investigating associations	Logistic regression: estimated crude and adjusted odds ratios (ORs). Models of butyl paraben levels used as either a continuous variable or in tertiles.	
suitability of used methods?	Yes	
maximised use of data?	Yes	
only descriptive statistics or/and bivariate analysis	No	
appropriate control for confounders?	Yes	
unadjusted and adjusted estimates presented?	Both unadjusted and adjusted estimates presented	
sensitivity tests and interaction analysis conducted?	No	
multiple testing issues	Yes	
additional comment related to the statistical analysis	Statistical methodology is sound and well-adjusted, but due to study size the power is limited.	

		Score
Reporting		Moderate
key elements of M&M and results are reported in sufficient detail?	Yes	
a plausible mechanism for the association under investigation is provided?		
are the conclusion made justified by the data shown?		
additional comment related to the reporting of the study	Negative observation for BP and authors did not discuss mechansims.	
Key findings		
what are the key findings?	No associations between levels of BP in placenta at term and congenital malformations (cryptorchidism/hypospadia). Residues of BP were, however, more frequently detected in cases than in	

	controls (85.7% vs. 66.7%, respectively, P = 0.054).	
any secondary findings?	No	
effect size in relation to biological relevance	inconclusive	
	overall score:	Moderate

Study name:	
Paper title:	Association of birth outcomes with fetal exposure to parabens, triclosan and triclocarban in an immigrant population in Brooklyn, New York
Authors:	Laura A. Geer, Benny F. G. Pycke, Joshua Waxenbaum, David M. Sherer, Ovadia Abulafia, Rolf U. Halden
Publication year:	2017

		Score
Study design and conduct		High
study type	Longitudinal	
study year	2007 to 2009	
additional comments related to study design		
Study population		Moderate
sampling method	Convenience sample, non-representative	
study size	177	
age range	18 to 45 years	
sex	Female	
other population characteristics	Recruited from an urban immigrant population	
quality of provided information on population characteristics	Well defined	
additional comments related to study population		
Exposure assessment		High
method(s) used for exposure assessment	Biomonitoring, human samples	
validated methods used?	Yes	
Is the timing between exposure and outcomes assessment appropriate?	Yes	
If HBM:		
matrix and sample type	Urine, random spot	
validated biomarker measured?	Yes	
adjusted for urinary dilution?	Yes, creatinine	
measured concentrations (median, range)	Cord blood plasma: 0.07 ng/mL (range: 0.01-0.35). Urine: 4.72 ng/mL (range: 0.02-146.61)	
% samples with BP <LOD/LOQ	not reported	
LOD/LOQ for BP	not reported	
additional comment related to exposure assessment	Matrix and sample type: Urine, random spot. Cord blood plasma. It is not reported when during pregnancy/which trimester the urine samples were collected.	

		Score
Outcome assessment		Moderate
outcome(s) assessed	Birth outcomes: birth weight, Gestational age (GA) at birth, body length, head circumference	
quality of outcome assessment	Register or medical record non-confirmed	
additional comment related to outcome assessment		

Confounder control		High
is information available for confounders relevant to the scientific questions asked? (comment if needed)	Yes	
	Confounders considered: Maternal age, nativity, neonate gender, alcohol and tobacco. They adjusted for confounders that were independently associated with the outcomes variable, or which changed the magnitude of the effects size by at least 5% when included in multiple linear regression models.	
are confounders clearly indicated?	Yes	
are confounders adequatly controlled for ?	Yes	
additional comment related to confounder control	Co-occurring pollutants that were not highly correlated (correlation coefficient < 0.6) were adjusted for in final multi-pollutant models.	
Statistical analysis		Moderate
methods used for investigating associations	Adjusted multiple linear regression. The relationship between pollutant predictors and the dichotomous outcomes preterm birth (PTB), birth at <37 weeks, and low birth weight (LBW), birth weight <2,500 g, were analysed using logistic regression.	
suitability of used methods?	Yes	
maximised use of data?	Yes	
only descriptive statistics or/and bivariate analysis	No	
appropriate control for confounders?	Yes	
unadjusted and adjusted estimates presented?	Only adjusted estimates presented	
sensitivity tests and interaction analysis conducted?	Yes	
multiple testing issues	Yes	
additional comment related to the statistical analysis	Sensitivity test performed but no interaction analysis. An odds ratio for BP and being born premature of +60 is very high and considering also the very wide confidence interval, we speculate that this result may be driven by an outlier and also may have been modelled on few cases of babies born premature. However, no information is given on the number of cases for this dichotomous outcome. The authors should have performed sensitivity test with exclusion of outliers.	
		Score
Reporting		Moderate
key elements of M&M and results are reported in sufficient detail?	No	
a plausible mechanism for the association under investigation is provided?	Yes	
are the conclusion made justified by the data shown?	Yes	

additional comment related to the reporting of the study	Lacking information on detection rate of BP in the samples and the limit of detection. Lacking information on incidence rate of being born preterm in the study group. Authors discuss that parabens have been associated with oxidative stress and that this could be a possible mechanism behind the association of BP with fetal growth restriction.	
Key findings		
what are the key findings?	Cord blood BP was associated with decreased GA at birth, in weeks (β = -3.04 95% CI -5.09, -0.99). Cord blood BP was associated with increased odds of preterm birth (PTB) with OR = 60.77, 95% CI 2.60, 1417.93. Cord blood BP was marginally associated with decreased birth weight (β = -480.40, 95% CI -976.68, 15.89, p = 0.057). Urine BP was associated with decreased GA at birth, in weeks (β = −0.36, 95% CI −0.72, -0.01).	
any secondary findings?		
effect size in relation to biological relevance	effect sizes observed are relatively large and potentially of biological relevance.	
	overall score:	Moderate

Study name:		
Paper title:	Urinary paraben concentrations and their associations with anthropometric measures of children aged 3 years	
Authors:	Jianqiu Guo, Chunhua Wu, Dasheng Lu, Shuai Jiang, Weijiu Liang, Xiuli Chang, Hao Xu, Guoquan Wang, Zhijun Zhou	
Publication year:	2017	

		Score
Study design and conduct		Low
study type	Cross-sectional	
study year	2012 to 2013	
additional comments related to study design	Language and gramatical errors in the text makes it difficult to interpret in detail how the study is performed	
Study population		Moderate
sampling method		
study size	436	
age range	The specific age of the children (mean/median, range) is not presented in the paper but they are around 3 years old.	
sex	Both males and females	
other population characteristics		
quality of provided information on population characteristics	not sufficiently defined	
additional comments related to study population	It is not clear from the description whether this is a random sample or not: "During July 2012-April 2013, 498 children were recruited in our study when they visited Sheyang Maternal and Child Health Care Centre. All children's mothers had previously participated in our longitudinal cohort study during pregnancy at hospital." Thus, the children are the offspring of mothers who previously participated in a longitudinal cohort study during pregnancy but it also states that the children were recruited during a hospital visit, so it remains unclear whether the children were healthy or not. They most likely were but not stated clearly.	
Exposure assessment		High
method(s) used for exposure assessment	Biomonitoring, human samples	
validated methods used?	Yes	
Is the timing between exposure and outcomes assessment appropriate?	Yes	
If HBM:		
matrix and sample type	Urine, random spot	
validated biomarker measured?	Yes	
adjusted for urinary dilution?	Yes, specific gravity	
measured concentrations (median, range)	0.05 ng/mL (< LOD to 5.59)	
% samples with BP <LOD/LOQ	2,50%	
LOD/LOQ for BP	0.01 ng/mL	

additional comment related to exposure assessment	Detection limit approximately 10 times lower than most other studies measuring BP in urine, which may explain the higher detection rate for BP as levels measured seemed to be at similar level as other studies. Based on measured BP levels they also calculate estimated daily intakes.	
		Score
Outcome assessment		High
outcome(s) assessed	Z scores for weight, height, weight for height and BMI.	
quality of outcome assessment	Standardised and validated outcome assessment	
additional comment related to outcome assessment	All children were measured by pediatric physicians blinded to exposure measures	
Confounder control		Moderate
is information available for confounders relevant to the scientific questions asked? (comment if needed)	Yes Confounders included: maternal body mass index, paternal body mass index, child's gender, maternal education, family annual income, inhabitation, feeding pattern, smoking status, time spent playing outdoors, sampling season, (child's sex x log - (each paraben)) and birth outcome measures (weight, length or ponderal index).	
are confounders clearly indicated?	Yes	
are confounders adequatly controlled for ?	Yes	
additional comment related to confounder control	It is stated by the authors that covariates were selected for the final statistical model in terms of sociodemographic and biological considerations, and in relation to concentrations of urinary parabens or measured body size of children ($p < 0.1$). Not clear how this should be interpreted which may be due to language barrier. From the footnotes of the tables it seems that they included all the above listed confounders, which seems a lot for a study this size and potentially could weaken the statistical power.	
Statistical analysis		Moderate
methods used for investigating associations	Adjusted generalised linear models.	
suitability of used methods?	Yes	
maximised use of data?	Yes	
only descriptive statistics or/and bivariate analysis	No	
appropriate control for confounders?	Yes	
unadjusted and adjusted estimates presented?	Only adjusted estimates presented	
sensitivity tests and interaction analysis conducted?	Yes	
multiple testing issues	Yes	
additional comment related to the statistical analysis	Models both included all children and stratified by sex. Concern about overadjustment by including more than 10 confounders and interaction terms in the models.	
		Score
Reporting		Moderate

key elements of M&M and results are reported in sufficient detail?	Yes	
a plausible mechanism for the association under investigation is provided?		
are the conclusion made justified by the data shown?	Yes	
additional comment related to the reporting of the study		
Key findings		
what are the key findings?	There were no significant associations between BP and anthropometric measurements.	
any secondary findings?	Higher urinary ∑ paraben concentrations were associated with increased height z-score: (β = 0.25, 95% CI: 0.01, 0.49, p = 0.04), (β = 0.28, 95% CI: 0.04, 0.53, p = 0.02), and (β = 0.44, 95% CI: 0.20, 0.68; p < 0.01) in respectively second, third and fourth quartiles compared to the first quartile. In the adjusted linear model, increased ∑ paraben concentration was associated with increased height z-score only in boys.	
effect size in relation to biological relevance	Uncertain	
	overall score:	Moderate

Study name:	Male Reproductive Health Study
Paper title:	Urinary excretion of phenols, parabens and benzophenones in young men: Associations to reproductive hormones and semen quality are modified by mutations in the Filaggrin gene
Authors:	Joensen UN, Jørgensen N, Thyssen JP, Szecsi PB, Stender S, Petersen JH, Andersson AM, Frederiksen H
Publication year:	2018

		Score
Study design and conduct		Moderate
study type	case-control	
study year	2007-2009	
additional comments related to study design	Case-control study nested in a cross-sectional study of young men from the general population. The men were genotyped for filaggrin loss-of-function mutations. Cases were men hetero- or homozygous for filaggrin loss-of-function mutation.	
Study population		High
sampling method	other	
study size	65 cases, 130 controls	
age range	18.1-26.3 y (median: 18.9 y)	
sex	Male	
other population characteristics	No significant difference between case and control groups regarding age, BMI, smoking, alcohol intake, season for participation, time of day for participation	
quality of provided information on population characteristics	well defined	
additional comments related to study population	Study population is a subpopulation of a random, repesentative study population. Identified as cases / controls after recruitment.	
Exposure assessment		High
method(s) used for exposure assessment	biomonitoring, human samples	
validated methods used?	Yes	
Is the timing between exposure and outcomes assessment appropriate?	Yes	
If HBM:		
matrix and sample type	urine, timed spot	
validated biomarker measured?	Yes	
adjusted for urinary dilution?	yes, osmolality	
measured concentrations (median, range)	0.22 ng/mL (<LOD-192)	
% samples with BP <LOD/LOQ	42%	
LOD/LOQ for BP	0.14 ng/mL	
additional comment related to exposure assessment		

	Score
Outcome assessment	High

outcome(s) assessed	Associations to hormone levels (follicle-stimulating hormone (FSH), luteinising hormone (LH), sex hormone-binding globulin (SHBG), testosterone (T), estradiol (E), Inhibin B), hormone ratioes (calculated free testosterone (FT) based on total testosterone and SHBG and the ratio of FT/LH, T/E, T/LH and Inhibin B/FSH) and semen parameters (Semen volume, sperm concentration, Total sperm count [semen volume x sperm concentration], progressive motile spermatozoa, morphologically normal spermatozoa, total morphology	
quality of outcome assessment	standardised and validated outcome assessment	
additional comment related to outcome assessment		
Confounder control		High
is information available for confounders relevant to the scientific questions asked? (comment if needed)	Yes	
	Confounders relevant for the hormone levels: Smoking, BMI, time of day for blood sampling. Confounders relevant for the semen analysis: Smoking, time since last ejaculation (abstinence time), time from delivery of semen sample to analysis of semen parameters.	
are confounders clearly indicated?	Yes	
are confounders adequatly controlled for ?	Yes	
additional comment related to confounder control		
Statistical analysis		Low
methods used for investigating associations	Multiple reggresion analysis with BP levels entered in the model as a dichotomous variable (< LOD or > LOD) due to the low detection rate. Statistics were performed both for cases and controls separately and for the two groups combined.	
suitability of used methods?	Yes	
maximised use of data?	Yes	
only descriptive statistics or/and bivariate analysis	Yes	
appropriate control for confounders?	Yes	
unadjusted and adjusted estimates presented?	only adjusted estimates presented	
sensitivity tests and interaction analysis conducted?	Yes	
multiple testing issues	Yes	
additional comment related to the statistical analysis	Statistical methods used are sound but due to the relative small study size and low detection rate for BP the study may be underpowered for investigation of associations of BP to the listed outcomes. Multiple testing issues not sufficiently addressed but since no significant association was observed for BP it is not a problem for the interpretation.	
		Score
Reporting		High
key elements of M&M and results are reported in sufficient detail?	Yes	

a plausible mechanism for the association under investigation is provided?	Yes	
are the conclusion made justified by the data shown?	Yes	
additional comment related to the reporting of the study		
Key findings		
what are the key findings?	No associations were found between urinary concentrations of BP and any reproductive hormones or any of the semen parameters, neither in the filaggrin loss-of-function group or for controls.	
any secondary findings?	Associations between male reproductive health parameters and urinary levels of BPA and benzophenones such as BP-3, BP-1 and 4-HBP were observed in Filaggrin gene (FLG) mutation carriers but not in controls from the same study population. This difference between FLG mutation carriers and non-carriers was not explained solely by differences in exposure levels of the examined compounds as e.g. BPA and 4-HBP urinary levels did not differ between the two groups. It was hypothesised that effects of exposure to these compounds may be modulated in filaggrin mutation carriers by either different levels of co-exposures or by route of uptake, with a higher fraction of the uptake by dermal uptake.	
effect size in relation to biological relevance	In this study no effects of butyl paraben on hormones and semen parameters were observed	
	overall score:	Low

Study name:	Environmental factors and Male Infertility Study
Paper title:	Environmental exposure to parabens and sperm chromosome disomy
Authors:	Joanna Jurewicz, Michał Radwan, Bartosz Wielgomas, Anna Klimowska, Paweł Kałużny, Paweł Radwan, Lucjusz Jakubowski & Wojciech Hanke
Publication year:	2017

		Score
Study design and conduct		Low
study type	cross-sectional	
study year	2008-2011	
additional comments related to study design	Polish study	
Study population		Moderate
sampling method	convenience sample, non-representative	
study size	156	
age range	average 32 years (SD= 4.5 y)	
sex	Male	
other population characteristics	men with normal semen conc. 15-300 mill/ml, with biobanked samples available for analysis	
quality of provided information on population characteristics	well defined	
additional comments related to study population	Men attending fertility clinic for diagnostic reasons (but with normal semen concentration). They mention that they had sufficient sample from 195 men but do not explain why only 156 men are included in the analyses.	
Exposure assessment		Moderate
method(s) used for exposure assessment	biomonitoring, human samples	
validated methods used?	Yes	
Is the timing between exposure and outcomes assessment appropriate?	No	
If HBM:		
matrix and sample type	urine, random spot	
validated biomarker measured?	yes	
adjusted for urinary dilution?	yes, creatinine	
measured concentrations (median, range)	geometric mean 0.3 +/- SD 2.55 ng/mL	
% samples with BP <LOD/LOQ	91%	
LOD/LOQ for BP	0.5 ng/mL	
additional comment related to exposure assessment	Urine, saliva and semen sample provided on the same day. For effects on sperm disomy it might be more relevant to assess BP exposure during spermatogenesis (2-3 months before sampling of semen). LOD higher than in other studies resulting in low detection rate for BP in the samples	

Score

Outcome assessment		High
outcome(s) assessed	Sperm chromosome disomy (XX, YY, XY, 1313, 1818, 2121)	
quality of outcome assessment	standardised and validated outcome assessment	
additional comment related to outcome assessment		
Confounder control		High
is information available for confounders relevant to the scientific questions asked? (comment if needed)	Yes sexual abstinence (days), age (years), smoking (yes/no), past diseases (yes/no), sperm concentration (mill/ml), and motility (%).	
are confounders clearly indicated?	Yes	
are confounders adequatly controlled for ?	Yes	
additional comment related to confounder control		
Statistical analysis		Low
methods used for investigating associations	Negative binomial regression modelling	
suitability of used methods?	Yes	
maximised use of data?	Yes	
only descriptive statistics or/and bivariate analysis	No	
appropriate control for confounders?	Yes	
unadjusted and adjusted estimates presented?	only adjusted estimates presented	
sensitivity tests and interaction analysis conducted?	No	
multiple testing issues	Yes	
additional comment related to the statistical analysis	Multiple testing issues not adressed. Only 14 men with BP above LOD results in very low statistical power.	
		Score
Reporting		Moderate
key elements of M&M and results are reported in sufficient detail?	Yes	
a plausible mechanism for the association under investigation is provided?	No	
are the conclusion made justified by the data shown?	Yes	
additional comment related to the reporting of the study		
Key findings		
what are the key findings?	Urinary BP positively associated with XY disomy. However, 91% of samples with BP<LOD.	
any secondary findings?	In unadjusted data also associated with 2121 disomy, but not significant in adjusted data	
effect size in relation to biological relevance	Small: Coef. 0.23 95CI (0.003-0.045) P=0.045 for group >LOD vs < LOD (reference group)	
	overall score:	Low

Study name:	National Health and Nutrition Examination Survey (NHANES) 2007-2008
Paper title:	Relationship between urinary triclosan and paraben concentrations and serum thyroid measures in NHANES 2007-2008
Authors:	Koeppe *et al*.
Publication year:	2013

		Score
Study design and conduct		Moderate
study type	cross-sectional	
study year	2007-2008	
additional comments related to study design		
Study population		High
sampling method	random, representative sample	
study size	1831	
age range	>=12 years	
sex	male and female	
other population characteristics		
quality of provided information on population characteristics	well defined	
additional comments related to study population	exclusion criterias: history of thyroid disease, pregnancy, influential outlier values	
Exposure assessment		High
method(s) used for exposure assessment	biomonitoring, human samples	
validated methods used?	Yes	
Is the timing between exposure and outcomes assessment appropriate?	Yes	
If HBM:		
matrix and sample type	urine, random spot	
validated biomarker measured?	Yes	
adjusted for urinary dilution?	yes, creatinine	
measured concentrations (median, range)	Median ranging from 0.13 - 1.06 ug/g creatinine highest in adult females, lowest in adult males.	
% samples with BP <LOD/LOQ	35% and 37% in respectively adult and adolescent females and 73% and 66% in respectively adult and adolescent males	
LOD/LOQ for BP	0.2 ng/mL	
additional comment related to exposure assessment	Stratified on sex and adolescense (12-19 years) and adult (>=20 years)	

		Score
Outcome assessment		Moderate
outcome(s) assessed	Serum thyroid hormones: TSH, total and free T3, total and free T4, Thyroxine-binding globulin (TBG)	

quality of outcome assessment	standardised and validated outcome assessment	
additional comment related to outcome assessment	free T3 and T4 measured by direct immunoassay methods, which may be susceptable for interference from TBG	
Confounder control		High
is information available for confounders relevant to the scientific questions asked? (comment if needed)	Yes	
	yes. Confounders adjusted for: age, sex, BMI, urinary creatinine	
are confounders clearly indicated?	Yes	
are confounders adequatly controlled for ?	Yes	
additional comment related to confounder control	confounders also considered included: urinariy iodine, race, income, education, serum cotinine and alcohole intake.	
Statistical analysis		High
methods used for investigating associations	Data analysed stratified by age (adolescense; adult). Correlations between variables examined by Pearson correlations and ANOVA. Multiple linear regression models.	
suitability of used methods?	Yes	
maximised use of data?	Yes	
only descriptive statistics or/and bivariate analysis	No	
appropriate control for confounders?	Yes	
unadjusted and adjusted estimates presented?	only adjusted estimates presented	
sensitivity tests and interaction analysis conducted?	Yes	
multiple testing issues	Yes	
additional comment related to the statistical analysis	Addtional models performed with data analysis stratified on sex and adolescense (12-19 years) and adult (>=20 years). For BP, associations to thyroid hormones were modelled as above or below LOD for BP as overall >50% of the samples were below LOD. Statistical power may be an issue for BP due to the low detection rate.	
		Score
Reporting		Moderate
key elements of M&M and results are reported in sufficient detail?	Yes	
a plausible mechanism for the association under investigation is provided?	No	
are the conclusion made justified by the data shown?	Yes	
additional comment related to the reporting of the study	They suggest differences in exposure levels as a very plausible reason for why significant associations of BP with thyroid hormones only is seen for adult women (highest exposed group). Exposure to BP in general very low.	
Key findings		

what are the key findings?	BP was significantly negatively associated with fT3 serum levels in adult women (p=0.03) and also negatively associated with total T3 and total T4 although not significantly (p=0.09). Similar association to fT3 also seen for PP and EP.	
any secondary findings?	BP was positively associated with age and female gender and inversely associated with BMI. BP was only moderatly correlated to the other parabens.	
effect size in relation to biological relevance	Effect parameter of -0.02 for Ln-fT3 corresponds to an estimated 2% lower fT3 in adult women having detectable BP urinary levels. It is questionable if such a small change is of biological relevance. No association to TSH levels was observed.	
	overall score:	Moderate

Study name:	
Paper title:	Urinary concentrations of phenols and phthalate metabolites reflect extracellular vesicle microRNA expression in follicular fluid
Authors:	Martinez RM *et al*.
Publication year:	2019

		Score
Study design and conduct		Moderate
study type	Cross-sectional	
study year	2014-2016	
additional comments related to study design	Up to two urine samples collected in the same *In vitro* fertilisation (IVF) cycle: during stimulation and at day of oocyte retrival	
Study population		Moderate
sampling method	convenience sample, non-representative	
study size	130	
age range	19-38 year	
sex	females	
other population characteristics		
quality of provided information on population characteristics	well defined	
additional comments related to study population	Both fertile and infertile women included (fertile women were those who had conceived spontaneously in the past and underwent IVF for pre-gestational diagnosis of autosomal recessive diseases). Exclusion criteria:diagnosis of polycystic ovarian syndrome (PCOS), endometriosis, poor responders, had a male partner with severe male factor infertility.	
Exposure assessment		High
method(s) used for exposure assessment	biomonitoring, human samples	
validated methods used?	Yes	
Is the timing between exposure and outcomes assessment appropriate?	Probably	
If HBM:		
matrix and sample type	urine, random spot	
validated biomarker measured?	Yes	
adjusted for urinary dilution?	Yes, specific gravity	
measured concentrations (median, range)	BP SG adjusted median: 2.73 ng/mL (IQR: 0.44 - 14.2)	
% samples with BP <LOD/LOQ	5%	
LOD/LOQ for BP	0.1 ng/mL	
additional comment related to exposure assessment	Concentrations measured higher than most of the other studies	

		Score
Outcome assessment		Moderate
outcome(s) assessed	Expression of microRNAs in extracellular vesicles (EV-miRNAs) isolated from follicular fluid	

quality of outcome assessment		
additional comment related to outcome assessment	EV-miRNAs have been associated with *In vitro* Fertilisation (IVF) outcomes, but this is still a fairly new area	
Confounder control		Moderate
is information available for confounders relevant to the scientific questions asked? (comment if needed)	Yes	
	Yes, a priori selected covariates:age, BMI, pre-IVF fertility status (fertile vs infertile), and batch number	
are confounders clearly indicated?	Yes	
are confounders adequatly controlled for ?	No	
additional comment related to confounder control		
Statistical analysis		High
methods used for investigating associations	Regression analysis. To account for multiple-testing, they applied the Benjamini-Hochberg FDR "p. adjust" function in R	
suitability of used methods?	Yes	
maximised use of data?	Yes	
only descriptive statistics or/and bivariate analysis	No	
appropriate control for confounders?	Yes	
unadjusted and adjusted estimates presented?	both unadjusted and adjusted estimates presented	
sensitivity tests and interaction analysis conducted?	Yes	
multiple testing issues	No	
additional comment related to the statistical analysis	They perform multiple testing but address this by applying a post hoc test.	
		Score
Reporting		High
key elements of M&M and results are reported in sufficient detail?	Yes	
a plausible mechanism for the association under investigation is provided?	Yes	
are the conclusion made justified by the data shown?	Yes	
additional comment related to the reporting of the study		
Key findings		
what are the key findings?	BP was not found to be associated with altered expression of EV-miRNAs in follicular fluid	
any secondary findings?		
effect size in relation to biological relevance	N/A	
	overall score:	Moderate

Study name:	-
Paper title:	Urinary Concentrations of Parabens and Serum Hormone Levels, Semen Quality Parameters, and Sperm DNA Damage
Authors:	John D. Meeker, Tiffany Yang, Xiaoyun Ye, Antonia M. Calafat, and Russ Hauser
Publication year:	2011

		Score
Study design and conduct		Low
study type	cross-sectional	
study year	2000-2004	
additional comments related to study design		
Study population		Low
sampling method	convenience sample, non-representative	
study size	194	
age range	18-55	
sex	Male	
other population characteristics	exclusion criteria: postvasectomy	
quality of provided information on population characteristics	well defined	
additional comments related to study population	Male partners of couples seeking fertility treatment. No information is given on reproductive health medical history other than vasectomy. Might be relevant to exclude men with an obvious reason for subfertility (e.g. history of cryptorchidism or similar)	
Exposure assessment		High
method(s) used for exposure assessment	biomonitoring, human samples	
validated methods used?	Yes	
Is the timing between exposure and outcomes assessment appropriate?	Yes	
If HBM:		
matrix and sample type	urine, random spot	
validated biomarker measured?		
adjusted for urinary dilution?	yes, specific gravity	
measured concentrations (median, range)	Median BP (SG adjusted): <LOD (<LOD - 32 ng/mL)	
% samples with BP <LOD/LOQ	68%	
LOD/LOQ for BP	0.2 ng/mL	
additional comment related to exposure assessment	For 78 of the participants they have two urine samples on average 29 days apart. For some of the investigated sperm parameters it might be more relevant to time the exposure assessment 2-3 months prior to the collection of semen sample.	

	Score
Outcome assessment	High

outcome(s) assessed	Reproductive hormones, semen quality, sperm DNA damage (comet assay)	
quality of outcome assessment	standardised and validated outcome assessment	
additional comment related to outcome assessment		
Confounder control		Moderate
is information available for confounders relevant to the scientific questions asked? (comment if needed)	Yes Urinary dilution, Age, BMI, Abstinence, Race (white vs. other), smoking status, timing of the clinic visit (i.e., time of collection of urine/blood/semen samples) by season (winter vs. spring, summer, or fall) and by time of day (0900–1259 hours vs. 1300–1600 hours)	
are confounders clearly indicated?	Yes	
are confounders adequatly controlled for ?	No	
additional comment related to confounder control	Lacking information on history of reproductive health medical conditions (probably more relevant as exclusion criteria than confounder but also not mentioned as exclusion criteria).	
Statistical analysis		Moderate
methods used for investigating associations	Multivariable linear regression	
suitability of used methods?	Yes	
maximised use of data?	Yes	
only descriptive statistics or/and bivariate analysis	No	
appropriate control for confounders?	Yes	
unadjusted and adjusted estimates presented?	both unadjusted and adjusted estimates presented	
sensitivity tests and interaction analysis conducted?	Yes	
multiple testing issues	Yes	
additional comment related to the statistical analysis	Multiple testing issues with risk of Type 1 error. Low statistical power due to high percentage of samples with BP <LOD.	
		Score
Reporting		Moderate
key elements of M&M and results are reported in sufficient detail?	Yes	
a plausible mechanism for the association under investigation is provided?	No	
are the conclusion made justified by the data shown?	Yes	
additional comment related to the reporting of the study		
Key findings		
what are the key findings?	BP associated with increased sperm DNA damage. No associations with reproductive hormones and other semen parameters.	
any secondary findings?	BP concentrations in two consecutive urine sampels only correlated moderately (r = 0.46 Spearman)	

effect size in relation to biological relevance	Beta coefficients for sperm with DNA damage (%tail) were respectively 6.81 and 8.23 for men in the medium and high exposure group compared to men with BP below LOD. %tail may be associated to single strand DNA breaks.	
	overall score:	Low

Study name:	Environment and Reproductive Health (EARTH) Study
Paper title:	Preconception and Prenatal Urinary Concentrations of Phenols and Birth Size of Singleton Infants Born to Mothers and Fathers from the Environment and Reproductive Health (EARTH) Study
Authors:	Carmen Messerlian, Vicente Mustieles, Lidia Minguez-Alarcon, Jennifer B. Ford, Antonia M. Calafat, Irene Souter, Paige L. Williams, Russ Hauser
Publication year:	2018

		Score
Study design and conduct		High
study type	Prospective	
study year	2005 to 2016	
additional comments related to study design	Urine collected preconception and in each trimester during pregnancy	
Study population		Moderate
sampling method	Convenience sample, non-representative	
study size	346 mothers and 184 fathers (184 couples)	
age range		
sex	Females and couples	
other population characteristics	Infertile couples	
quality of provided information on population characteristics	Well defined	
additional comments related to study population	Infertile couples cannot be regarded as representative of the population. However, associations between exposure and birth size are likely not different in this group compared to the population in general	
Exposure assessment		High
method(s) used for exposure assessment	Biomonitoring, human samples	
validated methods used?	Yes	
Is the timing between exposure and outcomes assessment appropriate?	Yes	
If HBM:		
matrix and sample type	Urine, random spot	
validated biomarker measured?	Yes	
adjusted for urinary dilution?	Yes, specific gravity	
measured concentrations (median, range)	Median BP (SG-adjusted): paternal preconception: 0.22 ng/mL (IQR 25-75: 0.12-0.53 ng/mL); maternal preconception: 1.01 ng/mL (IQR 25-75: 0.26-5.3 ng/mL); and maternal prenatal: 0.59 (IQR 25-75: 0.2-2.1 ng/mL).	
% samples with BP <LOD/LOQ	Paternal preconception: 80% Maternal preconception: 50% Maternal prenatal: 55%	
LOD/LOQ for BP	0.1 ng/mL	
additional comment related to exposure assessment	Unclear how the reported SG-adjusted medians and lower boundery of the IQR can be higher than the limit of detection when more than 50% of the samples had BP levels below LOD	

	Score
Outcome assessment	Moderate

outcome(s) assessed	Child head circumference at birth and birth weight.	
quality of outcome assessment	Register or medical record non-confirmed	
additional comment related to outcome assessment	Outcomes were collected from delivery records (medical record, not confirmed).	
Confounder control		High
is information available for confounders relevant to the scientific questions asked? (comment if needed)	Yes Confounders adjusted for depended on the model. Adjustment seemed reasonable.	
are confounders clearly indicated?	Yes	
are confounders adequatly controlled for ?	Yes	
additional comment related to confounder control	Potential confounders selected a priori based on substantive knowledge using a directed acyclic graph (DAG). Confounders considered: age, BMI, race, education, smoking status, infertility diagnosis (female factor, male factor, unexplained) and type of fertility treatment for both parents as well as parity for the mother. For the offspring: gender, gestational age at birth, date of birth (for season) and mode of conception and delivery.	
Statistical analysis		Moderate
methods used for investigating associations	Multivariable linear regression models.	
suitability of used methods?	Yes	
maximised use of data?	Yes	
only descriptive statistics or/and bivariate analysis	No	
appropriate control for confounders?	Yes	
unadjusted and adjusted estimates presented?	Both unadjusted and adjusted estimates presented	
sensitivity tests and interaction analysis conducted?	No	
multiple testing issues	Yes	
additional comment related to the statistical analysis	Post hoc sensitivity analysis and interaction tests included - including interaction by sex. Associations of paternal preconception BP levels not modelled due to the low detection rate. However, a detection rate of 50% may also be an issue. It is not discussed how this is handled other than "Concentrations below the LOD were assigned the LOD divided by the square root of two". Low detection rate likely to led to lower statistical power	
		Score
Reporting		Moderate
key elements of M&M and results are reported in sufficient detail?	Yes	
a plausible mechanism for the association under investigation is provided?		
are the conclusion made justified by the data shown?	Yes	
additional comment related to the reporting of the study		

Key findings		
what are the key findings?	No associations between either preconception (mother/father) or prenatal BP levels (mother) and head circumference or weight of the child at birth.	
any secondary findings?	No	
effect size in relation to biological relevance	not relevant	
	overall score:	Moderate

Study name:	Environment and Reproductive Health (EARTH) Study
Paper title:	Urinary paraben concentrations and *in vitro* fertilisation outcomes among women from a fertility clinic
Authors:	Lidia Mínguez-Alarcón, Yu-Han Chiu, Carmen Messerlian, Paige L. Williams, Mary E. Sabatini, Thomas L. Toth, Jennifer B. Ford, Antonia M. Calafat, Russ Hauser and for the Earth Study Team
Publication year:	2016

		Score
Study design and conduct		High
study type	prospective	
study year	2004-2012	
additional comments related to study design	Up to two urine samples collected per participant. Urine collection timed in relation to menstrual cycle and day of oocyte retrieval	
Study population		Moderate
sampling method	convenience sample, non-representative	
study size	245	
age range	18-45	
sex	Female	
other population characteristics	Females undergoing IVF (cycles with donor eggs and cryo-thraw cycles excluded)	
quality of provided information on population characteristics	well defined	
additional comments related to study population	Not-representative for the general population as study participants are females in infertile couples. Women undergoing IVF are highly selected in relation to the outcomes studied, which may lead to Type II error. Medical or genetic cause of infertility present in the study population may hide any potential effects of exposure.	
Exposure assessment		High
method(s) used for exposure assessment	biomonitoring, human samples	
validated methods used?	Yes	
Is the timing between exposure and outcomes assessment appropriate?	Yes	
If HBM:		
matrix and sample type	urine, timed spot	
validated biomarker measured?	Yes	
adjusted for urinary dilution?	yes, specific gravity	
measured concentrations (median, range)	BP median (SG adjusted): 1.23ng/mL (<LOD - 137 ng/mL)	
% samples with BP <LOD/LOQ	27%	
LOD/LOQ for BP	0.2 ng/mL	
additional comment related to exposure assessment		

	Score
Outcome assessment	High

outcome(s) assessed	Total and mature oocyte counts, proportion of high quality embryos, fertilisation rates, and rates of implantation, clinical pregnancy and live births	
quality of outcome assessment	standardised and validated outcome assessment	
additional comment related to outcome assessment		
Confounder control		High
is information available for confounders relevant to the scientific questions asked? (comment if needed)	Yes Final models were adjusted for age (continuous), BMI (continuous), race (white vs nonwhite), smoking status (never vs ever), and infertility diagnosis (male factor, female factor, unexplained).	
are confounders clearly indicated?	Yes	
are confounders adequatly controlled for ?	Yes	
additional comment related to confounder control		
Statistical analysis		Moderate
methods used for investigating associations	Multivariable generalised linear mixed models with random intercepts were used to evaluate the association between urinary paraben (MP, PP, and BP) concentrations and IVF outcomes.	
suitability of used methods?	Yes	
maximised use of data?	Yes	
only descriptive statistics or/and bivariate analysis	No	
appropriate control for confounders?	Yes	
unadjusted and adjusted estimates presented?	both unadjusted and adjusted estimates presented	
sensitivity tests and interaction analysis conducted?	No	
multiple testing issues	No	
additional comment related to the statistical analysis	Although intertility diagnosis (male factor, female factor, unexplained) is included as a confounder in the models it may not be sufficient to make up for the study population potentially being biased regarding the outcomes investigated. Genetic, infectious, or developmental causes of e.g. poor oocyte maturation is likely to attenuate possible effect from exposures. It might have been relevant to perform sensitivity tests only including women from couples with a male-factor infertility diagnosis (for oocyte maturation at least).	
		Score
Reporting		Moderate
key elements of M&M and results are reported in sufficient detail?	Yes	
a plausible mechanism for the association under investigation is provided?	No	
are the conclusion made justified by the data shown?	Yes	
additional comment related to the reporting of the study	It is explained that parabens are considered to be estrogenic, but not how they are supposed to directly affect IVF-outcomes.	

Key findings		
what are the key findings?	Urinary paraben concentrations were not associated with IVF outcomes among women undergoing infertility treatments.	
any secondary findings?	No	
effect size in relation to biological relevance		
	overall score:	Moderate

Study name:	
Paper title:	Association between paraben exposure and menstrual cycle in female university students in Japan
Authors:	Nishihama Y, Yoshinaga J, Iida A, Konishi S, Imai H, Yoneyama M, Nakajima D, Shiraishi H
Publication year:	2016

		Score
Study design and conduct		Low
study type	cross-sectional	
study year	2012-2013	
additional comments related to study design		
Study population		Moderate
sampling method	convenience sample, representative	
study size	128	
age range	median age: 20 (min-max: 19-22)	
sex	female	
other population characteristics	university students, Japanese nationality	
quality of provided information on population characteristics	well defined	
additional comments related to study population	Moderate score due to questionable study power	
Exposure assessment		High
method(s) used for exposure assessment	biomonitoring, human samples	
validated methods used?	Yes	
Is the timing between exposure and outcomes assessment appropriate?	Yes	
If HBM:		
matrix and sample type	urine, random spot	
validated biomarker measured?	Yes	
adjusted for urinary dilution?	yes, specific gravity	
measured concentrations (median, range)	BP Median (min-max): 0.634 ng/mL (<LOD - 73.7)	
% samples with BP <LOD/LOQ	30%	
LOD/LOQ for BP	0.5 ng/mL	
additional comment related to exposure assessment		

		Score
Outcome assessment		Moderate
outcome(s) assessed	Menstrual cycle length and intra-individual variation in menstrual cycle length	
quality of outcome assessment	self-reported	
additional comment related to outcome assessment	Information was prospectively recorded in five months and for that reason recall bias was not an issue. (Moderate score mainly due to self-reported data)	

Confounder control		High
is information available for confounders relevant to the scientific questions asked? (comment if needed)	Yes yes. Confounders adjusted for in all models: BMI, age, age at menarche. Additionally adjusted for in some of the models: meat consumption, menstrual pain.	
are confounders clearly indicated?	Yes	
are confounders adequatly controlled for ?	Yes	
additional comment related to confounder control		
Statistical analysis		Moderate
methods used for investigating associations	Logistic regression analysis	
suitability of used methods?	Yes	
maximised use of data?	Yes	
only descriptive statistics or/and bivariate analysis	No	
appropriate control for confounders?	Yes	
unadjusted and adjusted estimates presented?	both unadjusted and adjusted estimates presented	
sensitivity tests and interaction analysis conducted?	Yes	
multiple testing issues	Yes	
additional comment related to the statistical analysis	The study size is small limiting the possibility of conducting regression analysis. Not clear why they categorise the outcome variable menstrual length into short, medium and long rather than using the actual individual average menstrual cycle length in days in a regression model.	
		Score
Reporting		Moderate
key elements of M&M and results are reported in sufficient detail?	Yes	
a plausible mechanism for the association under investigation is provided?	No	
are the conclusion made justified by the data shown?	Yes	
additional comment related to the reporting of the study	The authors state that a specific causal mechanism is unclear	
Key findings		
what are the key findings?	Higher levels of BP is significantly associated with shorter menstrual cycle length (OR= 0.83 (0.70-0.99)	
any secondary findings?	No association to variability of menstrual cycle length. ∑parabens (MP,PP,BP and EP) also adversely associated to menstrual cycle length	
effect size in relation to biological relevance	Difficult to interpret due to the way the result is reported	
	overall score:	Moderate

Study name:	
Paper title:	Paraben exposure and semen quality of Japanese male partners of subfertile couples
Authors:	Yukiko Nishihama, Hiroki Toshima, Jun Yoshinaga, Yoshifumi Mizumoto, Miyuki Yoneyama, Daisuke Nakajima, Hiroaki Shiraishi and Susumu Tokuoka
Publication year:	2017

		Score
Study design and conduct		Moderate
study type	cross-sectional	
study year	2010	
additional comments related to study design		
Study population		Low
sampling method	convenience sample, non-representative	
study size	42	
age range	29-58	
sex	Male	
other population characteristics		
quality of provided information on population characteristics	not sufficiently defined	
additional comments related to study population	Men in couples attending infertility consultation. Low score due to small study size	
Exposure assessment		High
method(s) used for exposure assessment	biomonitoring, human samples	
validated methods used?	Yes	
Is the timing between exposure and outcomes assessment appropriate?		
If HBM:		
matrix and sample type	urine, random spot	
validated biomarker measured?		
adjusted for urinary dilution?	yes, specific gravity	
measured concentrations (median, range)	Median (SG adjusted) 0.166 (LOD-29,9)	
% samples with BP <LOD/LOQ	12%	
LOD/LOQ for BP	0.009 ng/mL	
additional comment related to exposure assessment	Semen sample and spot urine collected the same day. Effects on spermatogenesis affecting the concentration in a semen sample may occur 2-3 months before the ejaculation. Lower limit of detection of the used method than in most other studies which may explain the higher detection rate for BP in adult men seen in this study	

		Score
Outcome assessment		High
outcome(s) assessed	Sperm concentration, Semen volume, Sperm motility	
quality of outcome assessment	standardised and validated outcome assessment	

additional comment related to outcome assessment		
Confounder control		Moderate
is information available for confounders relevant to the scientific questions asked? (comment if needed)	Yes age, BMI, and abstinence period along with urinary concentrations of some other chemicals (3-phenoxybenzoic acid (3-PBA), daidzein and mono-n-butyl phthalate (MBP)), current smoking, consumption frequency of fruits and coffee, whether the subject is equol producer and season of semen sampling.	
are confounders clearly indicated?	Yes	
are confounders adequatly controlled for ?		
additional comment related to confounder control	overadjustment may be an issue considering the small study size and the high number of confounders included	
Statistical analysis		Low
methods used for investigating associations	Multiple regression analysis and logistic regression analysis	
suitability of used methods?	Yes	
maximised use of data?	Yes	
only descriptive statistics or/and bivariate analysis	No	
appropriate control for confounders?	Yes	
unadjusted and adjusted estimates presented?	only adjusted estimates presented	
sensitivity tests and interaction analysis conducted?	No	
multiple testing issues	Yes	
additional comment related to the statistical analysis	No adjustment for multiple testing. Small study size and possibly overadjustment leading to poor statistical power is a concern and the main reason for a low score	
		Score
Reporting		High
key elements of M&M and results are reported in sufficient detail?	Yes	
a plausible mechanism for the association under investigation is provided?	No	
are the conclusion made justified by the data shown?	Yes	
additional comment related to the reporting of the study		
Key findings		
what are the key findings?	No associations observed between BP in urine and semen quality parameters	
any secondary findings?	No	
effect size in relation to biological relevance	not relevant	
	overall score:	Low

Study name:	EDEN mother-child cohort
Paper title:	Prenatal Exposure to Phenols and Growth in Boys
Authors:	Philippat C, Botton J, Calafat AM, Ye X, Charles MA, Slama R; EDEN Study Group.
Publication year:	2014

		Score
Study design and conduct		High
study type	prospective	
study year	2003-2006	
additional comments related to study design	prospective mother-child cohort with women recruited before the end of the 28th gestational week.	
Study population		High
sampling method	random, representative sample	
study size	520	
age range	from fetal life up to 3 years of age	
sex	males	
other population characteristics	Nationality: France	
quality of provided information on population characteristics	well defined	
additional comments related to study population	In this paper only mother-son pairs included	
Exposure assessment		High
method(s) used for exposure assessment	biomonitoring, human samples	
validated methods used?	Yes	
Is the timing between exposure and outcomes assessment appropriate?	Yes	
If HBM:		
matrix and sample type	urine, morning void	
validated biomarker measured?	Yes	
adjusted for urinary dilution?	yes, creatinine	
measured concentrations (median, range)	Median (5th; 95th pct): 2.0 µg/L (<LOD - 59)	
% samples with BP <LOD/LOQ	16%	
LOD/LOQ for BP	0.2 ng/mL	
additional comment related to exposure assessment	Urine samples collected in pregnancy week 22-29	

		Score
Outcome assessment		High
outcome(s) assessed	Growth rate in boys including the parameters: prenatal: biparietal diameter, head circumference, femoral length, postnatal: size, abdominal circumference, weight	
quality of outcome assessment	standardised and validated outcome assessment	
additional comment related to outcome assessment	Repeated measurements of several of the outcomes were performed	
Confounder control		High
is information available for	Yes	

confounders relevant to the scientific questions asked? (comment if needed)	Confounders adjusted for: maternal and paternal height, maternal pre-pregnancy weigth, maternal activity, smoking, maternal education level, recruitment centre, parity. Additional confounders included in some models: days since birth (for head circumference) and breast feeding duration (for postnatal growth)	
are confounders clearly indicated?	Yes	
are confounders adequatly controlled for ?	Yes	
additional comment related to confounder control		
Statistical analysis		High
methods used for investigating associations	Linear regression models	
suitability of used methods?	Yes	
maximised use of data?	Yes	
only descriptive statistics or/and bivariate analysis	No	
appropriate control for confounders?	Yes	
unadjusted and adjusted estimates presented?	both unadjusted and adjusted estimates presented	
sensitivity tests and interaction analysis conducted?	Yes	
multiple testing issues	Yes	
additional comment related to the statistical analysis	Crude estimates presented in supplementary material.	
		Score
Reporting		High
key elements of M&M and results are reported in sufficient detail?	Yes	
a plausible mechanism for the association under investigation is provided?	Yes	
are the conclusion made justified by the data shown?	Yes	
additional comment related to the reporting of the study	Mechanism discussed: *in vitro* parabens promote adipocyte differentiation in murine cells by an activation of the glucocorticoid receptor or the peroxisome proliferator-activated receptor gamma. An effect on these nuclear receptors may increase susceptibility to gain weight and might play a role in the positive associations observed of parabens with abdominal circumference at 36 months and with postnatal weight	
Key findings		
what are the key findings?	No significant association of maternal levels of BP during pregnancy and any of the growth parameters although borderline associated to weight at birth and weight at 3 yrs were seen	
any secondary findings?	Similar trends were seen for the other parabens. Triclosan was associated with reduced growth late in pregnancy and reduced head circumference at birth.	
effect size in relation to biological relevance	High exposed children were approximately 50g heavier at birth and 200g heavier at 3 yrs of age.	
	overall score:	High

Study name:	
Paper title:	Exposure to bisphenol A, chlorophenols, benzophenones, and parabens in relation to reproductive hormones in healthy women: A chemical mixture approach
Authors:	Pollack AZ, Mumford SL, Krall JR, Carmichael AE, Sjaarda LA, Perkins NJ, Kannan K, Schisterman EF.
Publication year:	2018

		Score
Study design and conduct		High
study type	longitudinal	
study year	2005-2007	
additional comments related to study design		
Study population		Moderate
sampling method	lack of detailed information on population selection	
study size	143	
age range	18-44 years	
sex	premenopausal women	
other population characteristics	Nationality: USA	
quality of provided information on population characteristics	well defined	
additional comments related to study population	women recruited at a university research centre but no clear indication of which women attended this research centre	
Exposure assessment		High
method(s) used for exposure assessment	biomonitoring, human samples	
validated methods used?	Yes	
Is the timing between exposure and outcomes assessment appropriate?	Yes	
If HBM:		
matrix and sample type	urine, timed spot	
validated biomarker measured?	Yes	
adjusted for urinary dilution?	yes, creatinine	
measured concentrations (median, range)	median (min-max): 0.5 µg/L (<LOD - 132.2)	
% samples with BP <LOD/LOQ	26%	
LOD/LOQ for BP	0.02 ng/mL	
additional comment related to exposure assessment	Several spot urine samples during two menstrual cycles were included. Urine collected at key menstrual phases. Four samples measured per participant with additional samples measured for anovulatory cycles	

		Score
Outcome assessment		High
outcome(s) assessed	Reproductive hormone levels; estradiol, progesterone, LH and FSH around ovulation	
quality of outcome assessment	standardised and validated outcome assessment	

additional comment related to outcome assessment	All hormone levels were analysed using a solid-phase competitve chemiluminescent enzymatic immunoassay. Blood sample collection timed in relation to ovulation.	
Confounder control		High
is information available for confounders relevant to the scientific questions asked? (comment if needed)	Yes Yes, confounders adjusted for: age, BMI, etnicity, urinary creatinine	
are confounders clearly indicated?	Yes	
are confounders adequatly controlled for ?	Yes	
additional comment related to confounder control		
Statistical analysis		High
methods used for investigating associations	Linear mixed models (single compound analysis) and principal component analysis (multiple compounds analysis)	
suitability of used methods?	Yes	
maximised use of data?	Yes	
only descriptive statistics or/and bivariate analysis	No	
appropriate control for confounders?	Yes	
unadjusted and adjusted estimates presented?	both unadjusted and adjusted estimates presented	
sensitivity tests and interaction analysis conducted?	Yes	
multiple testing issues	Yes	
additional comment related to the statistical analysis	The authors perform single compound analysis but also multiple compound analysis based on principal component analysis. Multiple testing issue acknowledged by the authors but not further addressed.	
		Score
Reporting		Moderate
key elements of M&M and results are reported in sufficient detail?	Yes	
a plausible mechanism for the association under investigation is provided?		
are the conclusion made justified by the data shown?	Yes	
additional comment related to the reporting of the study		
Key findings		
what are the key findings?	BP was not associated to any of the hormone levels in single compound analysis. Nor any of the other parabens.	
any secondary findings?	In multiple compound analysis, parabens (butyl, ethyl, methyl and propyl paraben) combined were positively associated to estradiol and progesterone and borderline positively associated to LH.	
effect size in relation to biological relevance	1% increase in the total level of parabens (butyl, ethyl, methyl and propyl) corresponded to a 0.21% increase in estradiol and 0.32% increase in progesterone after adjustment for confounders.	

overall score:	Moderate

Study name:	Longitudinal investigation of fertility and the environment (LIFE) study
Paper title:	Male urinary biomarkers of antimicrobial exposure and bi-directionalassociations with semen quality parameters
Authors:	Melissa M. Smarr, Masato Honda, Kurunthachalam Kannan, Zhen Chen, Sungduk Kim, Germaine M. Buck Louis
Publication year:	2018

		Score
Study design and conduct		Moderate
study type	prospective	
study year	2005-2009	
additional comments related to study design	Although data is from a prospective TTP study the data included in this study on exposure and semen quality associations represent a cross-sectional design. While associations are based on a single spot urine sample, it is not clear how much time passes from urine collection to the first sperm sample collection. In Buck Louis *et al* 2018 (same study population) it is reported that avarage time from urine collection to semen collection was 2 months.	
Study population		Moderate
sampling method	other	
study size	501 males of male-female couples pregnancy planners; 473 with urine and semen sample	
age range	reproductive age	
sex	males	
other population characteristics	Inclusion criteria were: male partners aged ≥18 years; in a committed relationship; no physician-diagnosed infertility; ability to communicate in English or Spanish; and couple off contraception for ≤2 months.	
quality of provided information on population characteristics	well defined	
additional comments related to study population	Recruitment through fishing and hunting license registries and marketing databases for these interests. Due to the recruitment strategy it is not clear if this is a representative study population	
Exposure assessment		Moderate
method(s) used for exposure assessment	biomonitoring, human samples	
validated methods used?	Yes	
Is the timing between exposure and outcomes assessment appropriate?		
If HBM:		
matrix and sample type	urine, random spot	
validated biomarker measured?		
adjusted for urinary dilution?	yes, creatinine	
measured concentrations (median, range)	BP median 0.03ng/mL (IQR: 0.01-0.17)	

% samples with BP <LOD/LOQ	> 25% (not reported but lower boundery of IQR seem to be at/below LOD)	
LOD/LOQ for BP	0.012ng/mL	
additional comment related to exposure assessment	The interval between exposure measurement in relation to measurement of sperm quality parmeters seem to vary (reported to be on average 2 months) and therefore uncertain if the timing is appropriate; hence the score moderate.	
		Score
Outcome assessment		High
outcome(s) assessed	Semen quality: Sperm conc., total count, motility, morphology, morphometry	
quality of outcome assessment	standardised and validated outcome assessment	
additional comment related to outcome assessment		
Confounder control		Moderate
is information available for confounders relevant to the scientific questions asked? (comment if needed)	Yes Confounders adjusted for were based on a priori selection of covariates: age (years), urinary creatinine (ng/mL), body mass index (kg/m2), and race (non-Hispanic White, non-Hispanic Black, Hispanic, other)	
are confounders clearly indicated?	Yes	
are confounders adequatly controlled for ?	No	
additional comment related to confounder control	An additional confounder which is relevant to adjust for when evaluating sperm concentration is abstinence time which they did not include. They state that men were requested to have two days abstinence time, however, in practice some variation in this is likely.	
Statistical analysis		Moderate
methods used for investigating associations	Linear mixed models (80% of the men provided two semen samples)	
suitability of used methods?	Yes	
maximised use of data?	Yes	
only descriptive statistics or/and bivariate analysis	No	
appropriate control for confounders?	No	
unadjusted and adjusted estimates presented?	only adjusted estimates presented	
sensitivity tests and interaction analysis conducted?	No	
multiple testing issues	Yes	
additional comment related to the statistical analysis	Not adjusted for abstinence time in models of sperm concentration and total count. Urinary concentrations of the measured chemicals were modelled individually for each semen parameter resulting in multiple testing. Unclear from the reporting how many % of the urine samples had a BP concentration below LOQ. A high number of samples below LOQ affects the statistical power.	
		Score
Reporting		High

key elements of M&M and results are reported in sufficient detail?	Yes	
a plausible mechanism for the association under investigation is provided?	No	
are the conclusion made justified by the data shown?	Yes	
additional comment related to the reporting of the study	Information on % samples below LOD/LOQ for BP is missing. In the same study population it has been reported that 67% of urine samples were below LOD for BP (Buck Louis *et al*. 2018)	
Key findings		
what are the key findings?	BP was inversely associated with sperm concentration and sperm motility. The association with sperm conc is: A SD increase in ln-BP and (beta 95%CI) = −6.96 (−12.8, −1.08)*10^6 cells/mL.	
any secondary findings?	BP is weakly associated with a lower % of sperm with micro heads (beta 95%CI) = -0,15(-0,28,-0,01)%	
effect size in relation to biological relevance	This reduction in sperm concentration may affect fecundity - especially for men who already have suboptimal sperm production	
	overall score:	Moderate

Study name:	Environment ande Reproductive Health Study
Paper title:	Urinary Paraben Concentrations and Ovarian Aging among Women from a Fertility Center
Authors:	Smith KW, Souter I, Dimitriadis I, Ehrlich S, Williams PL, Calafat AM, Hauser R
Publication year:	2013

		Score
Study design and conduct		Low
study type	prospective	
study year	2004-2010	
additional comments related to study design	Low score due to a large variation in timespan between collection of urine samples and outcomes measures (more than three years variation).	
Study population		Moderate
sampling method	convenience sample, non-representative	
study size	192 (ranging from 109 to 142 in each analysis)	
age range	21-47 years	
sex	females	
other population characteristics	females seeking infertility evaluation or treatment. Nationality: USA	
quality of provided information on population characteristics	not sufficiently defined	
additional comments related to study population		
Exposure assessment		Moderate
method(s) used for exposure assessment	biomonitoring, human samples	
validated methods used?	Yes	
Is the timing between exposure and outcomes assessment appropriate?	No	
If HBM:		
matrix and sample type	urine, random spot	
validated biomarker measured?	Yes	
adjusted for urinary dilution?	yes, specific gravity	
measured concentrations (median, range)	median (min-max): from 1.42 to 2.08 µg/L (<LOD - 177).	
% samples with BP <LOD/LOQ	20%	
LOD/LOQ for BP	0.02 ng/mL	
additional comment related to exposure assessment	Time period between collection of urine samples and outcome measures was up to three years (varying from 0-3 years).	

		Score
Outcome assessment		High
outcome(s) assessed	Markers of ovarian reserve (day 3): FSH levels, antral follicle count (AFC), ovarian volume (OV)	
quality of outcome assessment	standardised and validated outcome assessment	
additional comment related to outcome assessment	Serum levels of anti-Müllerian hormone could have been a relevant marker of ovarian reserve to also include	

Confounder control		Moderate
is information available for confounders relevant to the scientific questions asked? (comment if needed)	Yes	
	Yes, covariates considered for inclusion in the regression models included age in years at the time of the outcome measure and BMI at entry into the study	
are confounders clearly indicated?	Yes	
are confounders adequatly controlled for ?	No	
additional comment related to confounder control		
Statistical analysis		High
methods used for investigating associations	Multiple linear regression analysis and Poisson regression analysis.	
suitability of used methods?	Yes	
maximised use of data?	Yes	
only descriptive statistics or/and bivariate analysis	No	
appropriate control for confounders?	No	
unadjusted and adjusted estimates presented?	both unadjusted and adjusted estimates presented	
sensitivity tests and interaction analysis conducted?	Yes	
multiple testing issues	Yes	
additional comment related to the statistical analysis	Unadjusted (age-adjusted) estimates are given in Supplementary material. Differences in the length of period between time of exposure and outcome estimation was not taken into account in data analysis	
		Score
Reporting		High
key elements of M&M and results are reported in sufficient detail?	Yes	
a plausible mechanism for the association under investigation is provided?	Yes	
are the conclusion made justified by the data shown?	Yes	
additional comment related to the reporting of the study	Overall low score on this paper due to the time span (and variation in this) between exposure measurement in relation to timepoint of outcome measurement	
Key findings		
what are the key findings?	No clear association between urinary BP concentrations and markers of ovarian reserve (FSH, antral follicle count and ovarian volume)	
any secondary findings?	Propylparaben negatively associated to antral follicle count	
effect size in relation to biological relevance	N/A	
	overall score:	Low