



Drivers of divergent assessments of bisphenol-A hazards to semen quality by various European agencies, regulators and scientists

Andreas Kortenkamp^{a,*}, Olwenn Martin^b, Eleni Iacovidou^a, Martin Scholze^a

^a Brunel University London, Centre for Pollution Research and Policy, College of Health, Medicine and Life Sciences, Kingston Lane, Uxbridge, UB8 3PH, United Kingdom

^b University College London, Department of Arts and Science, Gower Street, London, WC1E 6BT, United Kingdom

ARTICLE INFO

Keywords:

Bisphenol-A
Semen quality
Reference doses
Tolerable daily intakes
Health-based guidance value

ABSTRACT

The downward revision of the bisphenol A (BPA) Health-based Guidance Value (HBGV) by the European Food Safety Authority (EFSA) has led to disagreements with other regulatory agencies, among them the German Federal Institute for Risk Assessment (BfR). The BfR has recently published an alternative Tolerable Daily Intake (TDI), 1000-times higher than the EFSA HBGV of 0.2 ng/kg/d. While the EFSA value is defined in relation to immunotoxicity, the BfR alternative TDI is based on declines in sperm counts resulting from exposures in adulthood. Earlier, we had used semen quality deteriorations to estimate a BPA Reference Dose (RfD) of 3 ng/kg/d for use in mixture risk assessments of male reproductive health. We derived this estimate from animal studies of gestational BPA exposures which both EFSA and BfR viewed as irrelevant for human hazard characterisations. Here, we identify factors that drive these diverging views. We find that the fragmented, endpoint-oriented study evaluation system used by EFSA and BfR, with its emphasis on data that can support dose-response analyses, has obscured the overall BPA effect pattern relevant to male reproductive effects. This has led to a disregard for the effects of gestational BPA exposures. We also identify problems with the study evaluation schemes used by EFSA and BfR which leads to the omission of entire streams of evidence from consideration. The main driver of the diverging views of EFSA and BfR is the refusal by BfR to accept immunotoxic effects as the basis for establishing an HBGV. We find that switching from immunotoxicity to declines in semen quality as the basis for deriving a BPA TDI by deterministic or probabilistic approaches produces values in the range of 2.4–6.6 ng/kg/d, closer to the present EFSA HBGV of 0.2 ng/kg/d than the BfR TDI of 200 ng/kg/d. The proposed alternative BfR value is the result of value judgements which erred on the side of disregarding evidence that could have supported a lower TDI. The choices made in terms of selecting key studies and methods for dose-response analyses produced a TDI that comes close to doses shown to produce effects on semen quality in animal studies and in human studies of adult BPA exposures.

1. Introduction

Exposure to bisphenol-A (BPA), a synthetic compound used in the production of polycarbonate plastics, epoxy resins and thermal papers, is widespread. There is concern about associated human health effects. Recently, the European Food Safety Authority (EFSA) has published a revised BPA Health-based Guidance Value (HBGV) based on immunotoxicity (EFSA 2023), 20,000-times lower than their previous temporary Tolerable Daily Intake (TDI). However, with the publication of an alternative TDI by the German Federal Institute for Risk Assessment (Bundesinstitut für Risikobewertung, BfR), the effects of BPA on male reproductive health have recently also come into focus. The BfR

alternative TDI is derived from effects of BPA on sperm counts (BfR 2023). It is 1000-fold higher than the EFSA HBGV.

Cumulative exposures to BPA are predicted to range from 87 ng/kg/d (median) to 257 ng/kg/d (95th percentile, P95) for adults in Europe, with dietary exposures contributing between 30 ng/kg/d (median) and 146 ng/kg/d (P95). A large fraction of the total adult exposures from food is from semi-skimmed milk (up to 20%), orange juice (ca. 5%), tinned tomatoes (<5%) and salmon (>5%) (Karrer et al., 2020).

The new HBGV of 0.2 ng/kg/d developed by the EFSA Panel on Food Contact Materials, Enzymes and Food Processing Aids is 20,000-fold lower than their previous temporary TDI. Based on an extensive review of relevant studies, EFSA found that BPA exposure was related to a broad spectrum of harmful effects, the most sensitive outcome (critical

* Corresponding author.

E-mail address: andreas.kortenkamp@brunel.ac.uk (A. Kortenkamp).

<https://doi.org/10.1016/j.ijheh.2023.114293>

Received 26 September 2023; Received in revised form 2 November 2023; Accepted 3 November 2023

Available online 15 November 2023

1438-4639/© 2023 The Authors. Published by Elsevier GmbH. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Abbreviations

| | |
|-------|---|
| AF | Assessment factor |
| AUC | Area under the curve |
| BfR | Bundesinstitut für Risikobewertung (German Federal Institute for Risk Assessment) |
| BMD | Benchmark dose |
| BMDL | Benchmark dose lower limit |
| BMR | Benchmark response |
| BPA | Bisphenol-A |
| EFSA | European Food Safety Authority |
| EMA | European Medicines Agency |
| HBGV | Health-based guidance value |
| HED | Human equivalent dose |
| HEDF | Human equivalent dose factor |
| LCL | Lower confidence limit |
| LOAEL | Lowest observed adverse effect level |
| NOAEL | No observed adverse effect level |
| RfD | Reference dose |
| TDI | Tolerable daily intake |

toxicity) being an increase in the number of a specific type of immune cell involved in inflammatory diseases and obesity (T-helper cells). The new HBGV is derived from a mouse study which demonstrated a rise in T helper cells (Th 17) at very low BPA exposures. As European populations are predicted to experience exposures 400 to 1,200 times above the new HBGV, pressure on further regulating BPA, e.g. by eliminating it from food contact materials, is likely to increase.

However, the strong downward correction of EFSA's temporary TDI has spawned criticism from the European Medicines Agency (EFSA and EMA, 2022) and the BfR (EFSA and BfR, 2022), among others. The disagreements between EFSA, EMA and BfR concern several aspects of EFSA's re-evaluation, but in essence, derive from both EMA's and BfR's refusal to accept effects on T helper cells as the basis for establishing a new HBGV.

The BfR has now countered the EFSA HBGV by deriving an alternative, 1000-fold higher TDI of 200 ng/kg/d (BfR 2023). The BfR value is based on the deterioration of sperm counts in animal studies as the critical effect and is the result of a probabilistic approach to hazard characterization according to principles elaborated by Chiu and Slob (2015) and WHO/IPSC (2018). Compared to the standard deterministic approach, in which uncertainty and variability are dealt with by combining default extrapolation factors, probabilistic approaches have the advantage of quantifying the degree of uncertainty inherent in estimations of "safe" exposures, but little guidance exists as to when to use them for deriving HBGVs. Current European exposures are predicted to be lower or slightly in excess of 200 ng/kg/d, and in consequence, the BfR TDI would likely not trigger any regulatory measures for the protection against risks from BPA exposures.

While EFSA developed their revised HBGV, we were finalising a systematic literature review to establish a BPA reference dose (RfD) for deteriorations of semen quality (Kortenkamp et al. 2022 a). This RfD was intended to support a mixture risk assessment focused on male reproductive health which included BPA and a range of other chemicals (Kortenkamp et al. 2022 b). As a reasonable estimate of BPA exposures without effects on semen quality, this value does not take account of other toxicities and therefore cannot meet the degree of protection expected from a normative value such as a HBGV. Our BPA RfD was derived specifically from animal studies of gestational (i.e., the period from conception to birth) and neonatal (i.e., first 28 days of life) BPA exposures and effects on semen quality. We chose 3 ng/kg/d as the midpoint of a range of possible values derived from relevant studies, using deterministic approaches. Our interest was not in defining a value

that can protect against the whole range of BPA toxicities, but to estimate an exposure no longer associated with declines in semen quality, in line with the goal of our mixture risk assessment.

In contrast to our evaluation of gestational BPA exposures and semen quality (Kortenkamp et al. 2022 a), EFSA (2023) came to the view that deteriorations of semen quality resulting from gestational or post-natal BPA exposures are unlikely to be of relevance to human health and therefore of no consequence for the derivation of a new HBGV. However, they concluded that BPA exposures in adulthood can lead to poor semen quality. Yet, the application of the same assessment factors chosen by EFSA for the derivation of their HBGV to the key semen quality study selected by them (Wang et al., 2016) produces an RfD of 1 ng/kg/d for diminished sperm motility, below our RfD of 3 ng/kg/d.

Another area of dispute between EFSA and BfR concerns the use of assessment factors to support species-to-species extrapolations. Due to interspecies differences in toxicokinetics, doses in animal experiments must be converted into so-called human equivalent doses. BfR have argued that the factors used by EFSA for extrapolations from mouse studies are too high, leading to excessively low extrapolated human doses. This dispute is of consequence as the key study to support EFSA's new HBGV is in mice.

Because of BfR's derivation of an alternative TDI, estimates for "safe" human exposures protecting against BPA effects on semen quality now range from 1 ng/kg/d to 200 ng/kg/d, a difference of more than two orders of magnitude. This prompted us to identify the factors that drive these divergent hazard characterisation outcomes. Accordingly, in this paper we focus solely on semen quality and will not consider the dispute between EFSA, EMA and BfR concerning BPA effects on the immune system.

The following questions guide our evaluation: (1) What are the factors that drive the diverging views on the relevance of BPA exposures during gestation or adulthood for deteriorations of semen quality? (2) What motivated the choice of species-to-species extrapolation factors, specifically for mouse-to-human extrapolations? (3) What guided the choice of critical studies for the estimation of RfD or TDIs, and therefore determined their numerical values? (4) How do the various estimates for "safe" human BPA exposures relate to the available epidemiological evidence of associations between BPA and deteriorations of semen quality?

2. Materials

We based our comparative assessment on the following publications: The EFSA re-evaluation of the risks to public health related to the presence of BPA in foodstuffs (EFSA 2023) and associated technical documents such as the accompanying hazard assessment protocol (EFSA 2017), the BfR statement on a HBGV for BPA of 19 April 2023 (BfR 2023), a statement detailing the differences in opinion of EFSA and BfR (EFSA BfR 2023) and the material presented in our systematic review of BPA and declining semen quality after gestational exposures (Kortenkamp et al. 2022 a).

We conducted probabilistic hazard characterizations according to the principles elaborated by Chiu and Slob (2015) and WHO/IPSC (2018) by using the APROBA PLUS tool, version 1.14, available from <https://www.rivm.nl/en/aproba-plus>. As recommended by EFSA, we used a body weight of 70 kg as the default value for the adult human, 0.5 kg for the rat and 0.022 kg for the mouse. Beausoleil et al. (2022) conducted an appraisal of the reproductive and developmental toxicity of Bisphenol S (BPS) after developmental exposures. Since many of the relevant BPS studies also investigated BPA, we used this paper as an additional basis for comparisons.

A note on terminology is required. We use the term "health-based guidance value" (HBGV) synonymous with "tolerable daily intake" (TDI) to denote regulatory values based on Points of Departure (PoD) for the first toxic effect that is observed as exposures are escalated, called "critical toxicity". The term "reference dose" (RfD) is reserved for

toxicities derived from PoDs for toxicities that materialize at higher doses than those associated with the critical toxicity.

3. Results

3.1. Evidence synthesis: The relevance of gestational exposures and exposures in adulthood for deriving a BPA reference dose

To assess whether the evidence from animal studies was sufficient to infer a causal link with effects in humans, EFSA scrutinised both their internal (is the work of good technical quality?) and external validity (are animal models and effects relevant to human health?). By application of the approaches developed in [NTP OHAT \(2015\)](#) for assessing internal validity, EFSA considered key evaluation elements related to controlling various forms of bias, endpoint by endpoint, in a tiered system. The aspects assessed included randomisation procedures during the allocation of animals to dose groups, comparability and consistency of experimental conditions across groups, complete data reporting, details of exposure characterization and outcome assessment and statistical methods, including power considerations (appropriate numbers of animals). BfR adopted the EFSA confidence rating scheme in broad terms, albeit not in detail. We show that this tiered system led to a disregard of evidence indicative of hazards from gestational BPA exposures.

EFSA employed three Tiers, with Tier 1 corresponding to high and Tier 3 to low reliability. Tier 2 occupies a middle position. Placements of studies in these Tiers were translated into study confidence ratings which formed the basis for an integration of the evidence into judgements about the likelihood of an effect occurring in humans. In this step, EFSA used the “hazard identification conclusion categories” of “very likely”, “likely”, “as likely as not”, “unlikely” and “not classifiable”. Only for endpoints classed as “very likely” or “likely” were hazard characterizations (dose-response analyses) performed ([EFSA 2017](#)). To provide a sound basis for dose-response analyses, EFSA only considered studies that included a vehicle-exposed control and at least 3 BPA doses. All other studies were excluded from the analysis. Studies placed in Tier 3 were not taken forward into hazard characterization (dose-response analysis) and were deleted from further consideration.

EFSA considered sperm counts, sperm morphology, sperm motility and sperm viability separately, study by study. Studies were further grouped according to the timing of BPA exposures into developmental exposures (gestational and neonatal), developmental and adult exposures, exposures during the growth phase of animals and exposures in adulthood.

EFSA judged associations between BPA exposures during development and effects on all four indicators of semen quality (i.e., count, morphology, motility and viability) as “unlikely”, because neither of the studies which they rated as reliable (Tier 1) detected any of these effects ([Camacho et al., 2019](#)). According to EFSA, all studies that demonstrated effects on semen quality after developmental exposures suffered from insufficiencies which led to a downgrading to Tier 3, the lowest degree of reliability (e.g., inadequacies regarding the blinding to dose group allocations or outcome measurement or relating to test system contamination through BPA-containing caging). Studies placed in Tier 3 were excluded from further consideration (for example [Hass et al., 2016](#), [Kalb et al., 2016](#), [Rahman et al., 2017](#); [Shi et al., 2018](#)). EFSA made similar judgements in relation to pre-natal and post-natal exposures in pups until adulthood and exposures during the growth phase (young animals). Examples of studies of gestational exposures EFSA left out due to fewer than 3 BPA doses include [Chatsantiprapa et al. \(2016\)](#), [Salian et al. \(2009\)](#) and [Yang et al. \(2015\)](#). However, they missed [Vilela et al. \(2014\)](#), a mouse study with 3 different BPA doses. Other papers, such as [Shi et al. \(2019\)](#) and [Ullah et al. \(2019\)](#) were not considered as they appeared outside EFSA’s evaluation period which ended in 2018.

There is a considerable body of work on BPA gestational exposures and deteriorations of semen quality in animal studies. In our systematic

review ([Kortenkamp et al. 2022 a](#)) we identified 26 relevant studies that analysed total sperm count, sperm concentration, motility, morphology or vitality as outcome measures and that appeared up to August 2021. We followed a tiered evaluation system similar to that of EFSA, but unlike EFSA, did not evaluate semen quality parameters separately. Furthermore, we did not restrict the analysis to studies with at least 3 different BPA doses but considered all the evidence. We added key appraisal elements explicitly relating to the use of phytoestrogen-free feed, inclusion of positive controls and control of BPA contamination (these aspects were lumped together in the exposure characterisation key element of EFSA’s scheme). Because quality of reporting does not necessarily equate with quality of conduct, absence of information did not automatically lead to a downgrading to Tier 3 in our approach. Accordingly, only when there was positive evidence of the absence of phytoestrogen-free feed or of insufficient control of contaminations did we place studies in Tier 3. Before embarking on quantitative assessments, we assessed the strength of evidence for effects after gestational BPA exposures qualitatively.

For synthesising the evidence, we utilised the framework in [Radke et al. \(2018\)](#) which we modified according to the approach detailed in [EFSA \(2017\)](#). In this system, the strength of evidence is categorised as “Robust” when there are sets of studies with a Tier 1 confidence rating, with consistent findings of adverse effects across multiple laboratories and species. With our approach, 11 of the 26 studies obtained a high confidence rating. Of these, 10 studies reported the effects of BPA on semen quality parameters in multiple strains of rats and mice. We, therefore, evaluated the overall strength of evidence for associations of BPA exposures during gestation with poor semen quality as “robust” ([Kortenkamp et al. 2022 a](#)). This category is similar to EFSA’s “likely” rating. We did not consider BPA effects on semen quality after adult exposures.

EFSA rated associations between BPA exposures in adulthood and effects on semen quality as “likely”, but only for declines in sperm motility and sperm viability, and not for other measures of semen quality. Declines in sperm motility and viability were observed in one mouse study ([Wang et al., 2016](#)) which EFSA considered as well-conducted and reliable (Tier 1). Declines in sperm counts were rated as unlikely, as these were not seen in [Wang et al. \(2016\)](#), nor in one study placed by EFSA in Tier 2 ([Yin et al., 2017](#)). EFSA took the Wang et al. study forward into hazard characterization and conducted detailed dose-response analyses.

BfR used a confidence rating scheme quite similar to that employed by EFSA, with endpoint-by-endpoint evaluations. Like EFSA, they did not consider studies with fewer than 3 different BPA doses and dismissed the relevance of developmental exposures, but without any review of the literature. In agreement with EFSA, BfR regarded BPA effects on semen quality after exposures in adulthood as relevant. In contrast to EFSA, BfR viewed effects on sperm motility as unlikely but rated BPA effects on sperm counts as critical for deriving a TDI. These differences are the result of contrasting views of the reliability of the [Wang et al. \(2016\)](#) study. Noting that Wang et al. did not provide information about background BPA contamination via water bottles, caging and diet, BfR placed that study in their Tier 3 and promptly excluded it from further consideration. This exposes some inconsistencies in EFSA’s confidence rating approach, as missing information about BPA contamination in some other studies has led EFSA to downgrade them to Tier 3. However, this was not applied to [Wang et al. \(2016\)](#).

BfR based their TDI derivation on two studies of BPA effects on sperm counts after adult exposures, those by [Liu et al. \(2013\)](#) and [Srivastava and Gupta \(2018\)](#), both placed in their Tier 2. EFSA missed evaluating [Liu et al. \(2013\)](#) but, noting insufficient information about blinding and deficiencies in the reporting of background contaminants, downgraded [Srivastava and Gupta \(2018\)](#) to Tier 3. The details provided by [Srivastava and Gupta \(2018\)](#) allow the inference that the animal cages (polypropylene) did not leach BPA, but it is indeed unclear whether the animal feed in their study was free of phytoestrogens. [Liu et al. \(2013\)](#)

provided evidence that the animal feed used in their study was phytoestrogen-free, but there is insufficient information about the nature of their caging and about blinding. Application of EFSA's criteria would have led to a downgrading to Tier 3.

The differences in these evaluations are summarized in Table 1.

3.2. Conversion factors for mouse-to-human extrapolations

Rodents usually eliminate polar substances much faster than humans. To attain comparable tissue concentrations, rodents therefore must receive correspondingly higher doses than humans. To quantify these toxicokinetic interspecies differences, the blood-concentration-time profiles of chemicals, so-called areas under the curve (AUC), are compared after administration of similar doses. Faster metabolism and elimination give smaller AUC values. An animal-to-human extrapolation factor is then derived by calculating the ratio of the AUC of the species in question and the AUC measured for humans (human equivalent dose factor, HEDF). By multiplying the animal-specific reference dose with the HEDF, a human equivalent dose (HED) is derived. The inverse of BPA-specific HEDFs can then be used as substitution for the toxicokinetic standard assessment factor (AF) for interspecies extrapolations which is set to 4 by default (EFSA 2012). Critical for deriving a HEDF for BPA are measurements of the serum concentrations of free, unconjugated BPA.

Here, we argue that EFSA's choice for deriving a HEDF, although erring on the side of caution, is not well supported by the empirical evidence, while BfR's approach is more realistic.

EFSA (2023) derived an HEDF of 0.0155 for mouse-to-human extrapolations (the inverse of 0.0155 gives 64.5, used as AF for the kinetic interspecies extrapolation) based on an AUC of 0.244 nmol h L⁻¹ in mice and 15.7 nmol h L⁻¹ in humans. EFSA took the AUC mouse value from a study by Doerge et al. (2011); the human AUC value was chosen by calculating the median of two studies in volunteers (Teeguarden et al., 2015; Thayer et al., 2015).

BfR argued that the Doerge et al. study is not suitable for estimating a mouse HEDF. Doerge et al. had administered a single dose of 100 µg/kg to mice by gavage, and this dose resulted in free BPA serum levels that were too low to be measurable after 2 h of administration. The measured time window (0.25–1 h), so BfR's argument goes, does not cover the period of enterohepatic recirculation, which is higher in rodents than in humans, and which drives up free BPA levels. This results in larger AUCs and correspondingly higher HEDFs. In BfR's opinion, the mouse AUC for free BPA from Doerge et al. is unrealistically low. This opinion is

supported by the 1000-fold higher AUCs for total BPA (AUC = 247 nmol h L⁻¹, extrapolated to infinity) reported by Doerge et al. Accordingly, BfR prefer basing HEDF estimations on two alternative studies, those by Taylor et al. (2011) and Sieli et al. (2011) which produced larger AUCs for mice of 42 and 4.4 nmol h L⁻¹, respectively (after adjustment to a dose of 100 µg/kg BW). These values are 172-fold and 18-fold higher than the AUC from Doerge et al. In Taylor et al. and Sieli et al. the ratios of AUC of total to free BPA are much lower than the value of 1000 reported by Doerge et al., 124 (at 100 mg/kg) and 109 (at 20 mg/kg), respectively. Accordingly, BfR regarded mouse HEDFs between 0.2 and 1.56 as more appropriate (BfR 2023).

Further disputes concern the question of whether the free BPA AUC in mice increases linearly with the BPA dose and therefore supports a simple adjustment to a BPA dose of 100 µg/kg to enable comparisons between studies that used different BPA doses. EFSA concluded that the AUCs for unconjugated BPA in the studies by Taylor et al. (2011) did not increase linearly with dose, as indicated by the AUCs reported as 38.72 ng h mL⁻¹ at 400 µg/kg and 2991 ng h mL⁻¹ at 100,000 µg/kg. The latter value is three-times lower than what would be expected from the lower dose of 400 µg/kg, assuming that a linear increase of AUC with dose is correct. However, BfR argues that the measurement of the last time point (24 h) for the 400 µg/kg BW dose is unreliable due to analytical problems. An overestimation of the AUC is the consequence. If this timepoint is omitted from the AUC calculation, and the concentration-time profile of the high dose rescaled to the low dose, both AUCs are very similar. This suggests that, contrary to the points made by EFSA, there is no evidence against the linearity assumption between administered dose and unconjugated serum BPA in adult female mice throughout the entire 24 h period after oral administration, over a large dose range. This is supported by the linear correlation between four intake doses and their unconjugated serum levels measured 24 h after administration (Taylor et al., 2011). The linearity assumption is critical as the BPA doses in Taylor et al. and Sieli et al. were 4-, 200- and 1000-fold higher than in Doerge et al. (2011).

In addition, EFSA considered the study by Sieli et al. (2011) to be unreliable due to irregular concentration-time profiles of the unconjugated serum BPA. They argued that the data does not allow a clear separation of the kinetics of absorption and distribution from the elimination process. However, from the reported data we find it difficult to recapitulate this point. The key data from the various toxicokinetic studies are summarized in Table 2.

In summary, it appears that EFSA erred on the side of caution by relying exclusively on the data in one study, that by Doerge et al., and

Table 1
Study ratings of gestational and adult BPA exposures and semen quality in laboratory animals.

| | Outcome | EFSA (2023) | | Kortenkamp et al. (2022) | | BfR (2023) | |
|--------------------------------|-------------------------------------|-------------|-------------------------------------|--------------------------|-------------------------------|------------|------------------------------|
| | | Tier | Comments | Tier | Comments | Tier | Comments |
| Developmental exposures | | | | | | | |
| NTP Clarity 2018 | no effect | 1 | | 1 | assessed Delclos et al., 2014 | na | not assessed due to focus on |
| Camacho et al., (2019) | no effect | 1 | | na | assessed Delclos et al., 2014 | na | adult exposures |
| Hass et al., (2016) | Decreased sperm number | 3 | Blinding, incomplete reporting | 1 | | na | |
| Kalb et al., (2016) | Decreased sperm number | 3 | Blinding, exposure characterisation | na | | na | |
| Rahman et al., (2017) | Decreased sperm number | 3 | Blinding, exposure characterisation | 1 | | na | |
| Shi et al., (2018) | Decreased sperm number | 3 | Blinding, exposure characterisation | 1 | | na | |
| Adult exposures | | | | | | | |
| Wang et al., (2016) | Decreased sperm motility, viability | 1 | Key study | na | not assessed due to focus | 3 | Exposure characterisation |
| Liu et al., (2013) | Decreased sperm number | na | | na | on gestational exposures | 2 | Key study |
| Srivastava and Gupta (2018) | Decreased sperm number | 3 | Blinding, exposure characterisation | na | | 2 | Key study |

Notes: na: not assessed.

Table 2
Toxicokinetic parameters in selected BPA mouse studies.

| Study | Species | Dose (µg/kg/d) | AUC free BPA (nmoles h L ⁻¹) | AUC total BPA (nmoles h L ⁻¹) | Ratio total/free | Comment |
|-----------------------|----------------|----------------|--|---|------------------|---|
| Sieli et al., (2011) | adult C57Bl/6J | 20,000 | 4.4 | 481 | 109.3 | AUC 0–24 h |
| Taylor et al., (2011) | adult CD-1 | 400 | 42 | na | | AUC 0 to infinity |
| | adult CD-1 | 100,000 | 13.1 | 1626 | 124 | AUC 0 to infinity |
| Doerge et al., (2011) | adult CD-1 | 100 | 0.244 | 247 | 1012 | AUC 0 to infinity, AUC free BPA recalculated by EFSA (2023) |

Notes: AUC values adjusted to a dose of 100 µg/kg/d; values from Sieli et al. (2011) and Taylor et al. (2011) converted to molar units.

disregarding the evidence from Taylor et al. and Sieli et al. which collectively support a larger mouse HEDF than the value used by EFSA.

3.3. Studies critical for the derivation of a reference dose or TDI in relation to semen quality deteriorations

We present both deterministic and probabilistic approaches to deriving RfDs for BPA effects on semen quality. The application of probabilistic approaches in hazard characterization is relatively new, and to our knowledge, no guidance exists about when to favour probabilistic over the more familiar deterministic approaches in deriving HBGVs. Probabilistic approaches have considerable advantages in terms of clear communication of the risks associated with chemical exposures (Chiu and Slob 2015). We argue that both with deterministic and probabilistic approaches, considerably lower TDI's can be justified than the value proposed by BfR.

3.4. Deterministic approaches

In the key study of declines of semen quality in adulthood identified by EFSA (2023), C57BL/6 mice were dosed orally for 8 weeks with BPA at 10, 50 and 250 µg/kg/d (Wang et al., 2016). Decreases in sperm motility were observed at all doses; sperm viability was affected at 250 µg/kg/d. The lowest-observed-adverse-effect-level (LOAEL) in this study is 10 µg/kg/d. Benchmark dose modelling conducted by EFSA produced a benchmark dose lower limit associated with a 20% effect (BMDL₂₀) of 3.41 µg/kg/d for declines in sperm motility, and 26.1 µg/kg/d for reduced sperm viability. In a deterministic approach similar to that used by EFSA for the derivation of the immunotoxicity HBGV, with a mouse HEDF of 0.0155 (equivalent to an inter-species AF of 64.5), the BMDL₂₀ for reduced sperm motility translates into a human equivalent dose

(HED) of 0.053 µg/kg/d. Application of an AF of 2 to allow extrapolation from subchronic to chronic exposures and further AFs for interspecies differences in toxicodynamics (2.5), between-human differences in toxicokinetics (3.2) and toxicodynamics (3.2) produces an RfD of 0.001 µg/kg/d (Table 3).

EFSA based their BMD calculation on a relatively large benchmark response (BMR) of 20%, which might be considered too high. We, therefore, made our own estimations using the LOAEL from Wang et al. (2016) in which a deterministic approach produced an RfD of 0.0005 µg/kg/d, slightly lower than the deterministic value of 0.001 µg/kg/d derived from EFSA's benchmark dose modelling (Table 3).

Liu et al. (2013), one of the two key studies chosen by BfR, dosed adult Wistar rats with BPA (oral route) at 2, 20 and 200 µg/kg/d for 60 days. At 200 µg/kg/d, significantly decreased sperm counts were observed; all other doses were without effects. Accordingly, the NOAEL from this study is 20 µg/kg/d which, in a deterministic approach, translates into an RfD of 0.065 µg/kg/d.

BfR used the data from Liu et al. for benchmark dose modelling. With frequentist model averaging, they calculated a BMDL₁₀ of 2 µg/kg/d which increased to 26 µg/kg/d when Bayesian model averaging was used (BfR 2023). Frequentist model averaging produced unstable estimates, as suggested by the large ratio between the upper and lower benchmark dose limits (BfR 2023). According to EFSA guidance (EFSA 2022) this precludes the use of the Liu et al. data for deriving TDIs or HBGVs, and therefore BfR did not take the BMDL₁₀ of 2 µg/kg/d forward for establishing a deterministic RfD. The Bayesian model averaging BMDL of 26 µg/kg/d translates into a deterministic RfD of 0.085 µg/kg/d (Table 3).

In the second key study selected by BfR, Srivastava and Gupta (2018) treated adult Wistar albino rats with 50, 500 and 1000 µg/kg/d by the oral route. Significantly decreased sperm counts were observed at 500

Table 3
Derivation of BPA reference doses for semen quality by deterministic approaches.

| Study data | | | | Assessment factors for deterministic approaches | | | | | | |
|------------|---------------------------|------------------------------------|---------------------------|---|--------------------------|--------------------------|--------------------------|------------------|------------|---------------|
| Species | Point of departure (type) | Point of departure value (µg/kg/d) | Reference | AF LOAEL to NOAEL | AF subchronic to chronic | AF inter-species kinetic | AF inter-species dynamic | AF intra-species | Overall AF | RfD (µg/kg/d) |
| Mouse | BMDL | 3.41 | Wang et al. 2016 | 1 | 2 | 64.5 | 2.5 | 10 | 3225 | 0.001 |
| Mouse | LOAEL | 10 | Wang et al. 2016 | 6 | 2 | 64.5 | 2.5 | 10 | 19350 | 0.0005 |
| Rat | NOAEL | 20 | Liu et al. 2013 | 1 | 2 | 6.14 | 2.5 | 10 | 307 | 0.065 |
| Rat | BMDL B ¹ | 26 | Liu et al. 2013 | 1 | 2 | 6.14 | 2.5 | 10 | 307 | 0.085 |
| Rat | NOAEL | 50 | Srivastava and Gupta 2018 | 1 | 2 | 6.14 | 2.5 | 10 | 307 | 0.163 |
| Rat | BMDL S ¹ | 0.7 | Srivastava and Gupta 2018 | 1 | 2 | 6.14 | 2.5 | 10 | 307 | 0.002 |
| Mouse | LOAEL | 0.5 | Shi et al. 2018 | 6 | 2 | 64.5 | 2.5 | 10 | 19350 | 0.00003 |
| Rat | NOAEL | 1.1 | Ullah et al. 2019 | 1 | 2 | 6.14 | 2.5 | 10 | 307 | 0.004 |

Notes:

¹ BMDL S - benchmark dose modelling with standard frequentist model averaging, BMDL B - benchmark dose modelling with Bayesian model averaging.

Values from BfR (2023)

The gestational studies (Shi et al. 2018, Ullah et al. 2019) are highlighted by darker shading.

µg/kg/d. Accordingly, the NOAEL is 50 µg/kg/d. By application of a deterministic approach, BfR calculated an RfD of 0.163 µg/kg/d (Table 3). Benchmark dose modelling using frequentist model averaging produced unstable estimates, and Bayesian model averaging was not possible.

In addition to the key studies selected by EFSA and BfR, we analysed two further studies of gestational exposures to BPA in rodents (Shi et al., 2018; Ullah et al., 2019). These data featured in our estimation of a BPA RfD for declines in semen quality for the purposes of a mixture risk assessment (Kortenkamp et al. 2022 a). In our evaluation, both studies were placed in Tier 1, but we noted the lack of information about phytoestrogen-free feeds. In a recent evaluation of BPS by Beausoleil et al. (2022) both studies were rated as high-quality key studies (Shi et al., 2018; Ullah et al., 2019 also investigated BPS). EFSA (2023) placed the work by Shi et al. (2018) in their Tier 3, because of lacking documentation about background contamination and blinding during outcome measures. During their evaluation of effects on testis histology, however, they gave Shi et al. (2018) a higher confidence rating (Tier 2). Ullah et al. (2019) was not evaluated by EFSA.

Shi et al. (2018) dosed pregnant CD-1 mice orally with 0.5, 20 and 50 µg/kg/d BPA from gestational day 11 to birth. Sperm counts were measured when the mice were 60 days old and were found to be significantly decreased at the two lower doses, but not at the highest dose. Accordingly, the LOAEL is 0.5 µg/kg/d, which by application of the usual AFs gives a deterministic RfD of 0.00003 µg/kg/d (Table 3).

Ullah et al. (2019) administered BPA in drinking water to pregnant Sprague-Dawley rats from conception to birth, at concentrations of 5, 25 and 50 µg/L. Significant reductions in sperm counts became apparent at 50 µg/L, and accordingly, the NOAEL from this study is 25 µg/L. Assuming a daily water consumption of 6 mL per 100 g of body weight, and a body weight of 400 g, the NOAEL translates into a dosage of 1.1 µg/kg/d. From this, a deterministic RfD of 0.004 µg/kg/d can be derived (Table 3).

3.5. Probabilistic approaches

Based on the BMDL₁₀ from Bayesian model averaging (26 µg/kg/d) derived from Liu et al. (2013), BfR conducted their own probabilistic calculations which yielded a RfD of 2 µg/kg/d (lower confidence limit, LCL, Table 4). In these calculations, BfR changed the default distribution parameters in the APROBA tool (WHO/IPSC, 2018), as follows:

Instead of choosing the APROBA settings for extrapolations from rat to human, they constructed their own BPA-specific log-normal distribution which they derived from all possible permutations between three rat AUC values (Doerge et al., 2011; Pottenger et al., 2000; Domoradzki

et al., 2003) and two human AUC values (Thayer et al., 2015; Teeguarden et al., 2015). This generated a pool of six HEDF ranging from 0.11 to 1.58 which produced the AF distribution for inter-species extrapolations in APROBA (P05 = 0.64; P95 = 6.14).

The second deviation from the default APROBA settings concerns the distribution for intraspecies response variations which BfR chose to set much lower (P05 = 0.1; P95 = 10) than the APROBA values (P05 = 2.24; P95 = 41.88) (Table 4).

The third deviation relates to the subchronic-to-chronic extrapolation, which was lowered from (P05 = 0.5; P95 = 8, APROBA default) to (P05 = 0.5; P95 = 2), and the remaining inter-species extrapolation for which a narrower distribution setting was used (P05 = 0.4; P95 = 2.5 instead of P05 = 0.3; P95 = 3) (Table 4).

To provide a counterpoint to their own calculations, BfR also applied the default WHO APROBA settings to the Bayesian model averaging BMDL₁₀ from Liu et al. (2013) (26 µg/kg/d) but left the kinetic inter-species AF unchanged. This produced a lower value of 0.14 µg/kg/d.

With the NOAEL of 50 µg/kg/d from Srivastava and Gupta (2018), the “in house” BfR probabilistic method produced a RfD of 3.1 µg/kg/d (LCL), and application of what BfR referred to as the WHO approach gave 0.2 µg/kg/d (LCL) (Table 4).

These values are critical for the derivation of BfR’s alternative TDI. BfR took the “WHO approach” value of 0.14 µg/kg/d derived from Liu et al. (2013) and the corresponding value of 0.2 µg/kg/d based on Srivastava and Gupta (2018) which they consolidated into an overall TDI of 0.2 µg/kg/d.

We applied the WHO/IPSC (2018) default APROBA settings to the other points of departure from Liu et al. (2013) and to Wang et al. (2016). The tool performs interspecies extrapolations by allometric scaling. For the rat, this gives a median AF of 4.4 (P05: 3.61; P95: 5.37), lower than the value of 6.14 preferred by EFSA in its deterministic approach (EFSA 2023). The corresponding values for the mouse are 11.24 (P05: 8.14, P95: 15.25), well below the AF of 64.5 used by EFSA. Compared with the BfR values from the Liu et al. study, these settings yielded somewhat lower estimates, as follows: The NOAEL of 20 µg/kg/d gave a RfD of 0.044 µg/kg/d and the BMDL₁₀ of 26 µg/kg/d from Bayesian model averaging produced a RfD of 0.076 µg/kg/d (all LCL).

With the BMDL₂₀ of 3.41 µg/kg/d from Wang et al. (2016) we obtained a RfD of 0.0066 µg/kg/d (LCL, Table 4), nearly 7-fold higher than the deterministic value of 0.001 µg/kg/d (Table 3), mainly due to the lower inter-species AF.

When applied to the NOAEL of 1.1 µg/kg/d from the gestational study by Ullah et al. (2019), a probabilistic RfD of 0.0024 µg/kg/d can

Table 4
Derivation of BPA reference doses for semen quality by probabilistic approaches.

| Source | Study data | | Parameters (assessment factors) for probabilistic approaches | | | | | | | | | | Reference doses | | | | | | | | |
|--------------------------|------------|--------------------------------------|--|---------------------------|----------------|------|-----------------------|-----|------|---------------|-------|------|-------------------------|-----|------|---------------|-------|--------|----------|----------|------|
| | Species | Point of departure (value) (µg/kg/d) | Reference | NOAEL to BMDL | | | Subchronic to chronic | | | Inter-species | | | Inter-species remaining | | | Intra-species | | | Lower CL | Upper CL | |
| | | | | P05 | P50 | P95 | P05 | P50 | P95 | P05 | P50 | P95 | P05 | P50 | P95 | P05 | P50 | P95 | | | |
| BfR (2023) settings | Rat | BMDL B ¹ | 26 | Liu et al. 2013 | not applicable | 0.5 | 1 | 2 | 0.64 | 1.98 | 6.14 | 0.4 | 1 | 2.5 | 0.1 | 1 | 10 | 2 | 954 | | |
| BfR, WHO APROBA settings | Rat | BMDL B ¹ | 26 | Liu et al. 2013 | not applicable | 0.5 | 2 | 8 | 0.64 | 1.98 | 6.14 | 0.3 | 1 | 3 | 2.24 | 9.69 | 41.88 | 0.14 | 39 | | |
| BfR (2023) settings | Rat | NOAEL | 50 | Srivastava and Gupta 2018 | 0.07 | 0.33 | 1.57 | 0.5 | 1 | 2 | 0.64 | 1.98 | 6.14 | 0.4 | 1 | 2.5 | 0.1 | 1 | 10 | 3.1 | 1872 |
| BfR, WHO APROBA settings | Rat | NOAEL | 50 | Srivastava and Gupta 2018 | 0.07 | 0.33 | 1.57 | 0.5 | 2 | 8 | 0.64 | 1.98 | 6.14 | 0.3 | 1 | 3 | 2.24 | 9.69 | 41.88 | 0.2 | 78 |
| WHO (2018) APROBA | Rat | NOAEL | 20 | Liu et al. 2013 | 0.07 | 0.33 | 1.57 | 0.5 | 2 | 8 | 3.61 | 4.4 | 5.37 | 0.3 | 1 | 3 | 2.24 | 9.69 | 41.88 | 0.044 | 11.3 |
| | Rat | BMDL B ¹ | 26 | Liu et al. 2013 | not applicable | 0.5 | 2 | 8 | 3.61 | 4.4 | 5.37 | 0.3 | 1 | 3 | 2.24 | 9.69 | 41.88 | 0.076 | 14 | | |
| | Mouse | BMDL | 3.41 | Wang et al. 2016 | not applicable | 0.5 | 2 | 8 | 8.14 | 11.24 | 15.25 | 0.3 | 1 | 3 | 2.24 | 9.69 | 41.88 | 0.0066 | 8.1 | | |
| | Rat | NOAEL | 1.1 | Ullah et al. 2019 | 0.07 | 0.33 | 1.57 | 0.5 | 2 | 8 | 3.61 | 4.4 | 5.37 | 0.3 | 1 | 3 | 2.24 | 9.69 | 41.88 | 0.0024 | 0.6 |

Notes:

¹ BMDL S - benchmark dose modelling with standard frequentist model averaging, BMDL B - benchmark dose modelling with Bayesian model averaging (BfR 2023). The only gestational study (Ullah et al. 2019) is highlighted by darker shading.

be obtained.

Taken together, deterministic approaches to estimating RfDs from studies of adult exposures give values of between 0.001 and 0.163 $\mu\text{g}/\text{kg}/\text{d}$, while RfDs from gestational exposure studies are somewhat lower, between 0.00003 and 0.004 $\mu\text{g}/\text{kg}/\text{d}$ (Table 3).

The corresponding probabilistic RfDs range from 0.007 to 3 $\mu\text{g}/\text{kg}/\text{d}$ (Table 4).

The alternative TDI proposed by BfR (0.2 $\mu\text{g}/\text{kg}/\text{d}$) comes rather close to the doses associated with effects reported in some of these gestational studies: the margin left between the BfR TDI and the LOAEL from Shi et al. (2018) is only 2.5; the margins for the points of departure (POD) from Ullah et al. (2019) and Wang et al. (2016) are 5.5 and 50, respectively.

3.6. Epidemiological studies

EFSA (2023) reviewed cohort studies of associations between BPA in spot urine samples and measures of fertility in males, such as fertilization success or semen quality and concluded that such associations are unlikely. BfR (2023) did not consider epidemiological evidence.

Most studies of BPA effects on semen quality are cross-sectional which complicates assessments of whether the health outcomes can be attributed to chemical exposures because exposure and health outcome are measured at the same time. Furthermore, many of these studies relied on spot urine samples which only reflect recent exposures, while spermatogenesis takes up to 3 months. Referring to these difficulties, EFSA (2023) stated: “Although the associations were not entirely consistent in terms of directionality, findings from some individual studies could be interpreted as being adverse for male reproductive function.” However, due to low confidence in BPA exposure assessments based on spot urine samples, EFSA placed all epidemiological studies in Tier 3, and consequently, none were taken forward into hazard characterization steps.

Our systematic review and quality rating of human studies of BPA and semen quality (Kortenkamp et al. 2022) was not primarily motivated by hazard characterisations but aimed at evaluating the strength of evidence qualitatively (“Is BPA associated with declines in semen quality in humans?”). We identified 8 studies with null findings and 8 studies that reported associations of declining semen parameters with BPA exposures. Of the 8 null studies, 3 achieved an overall confidence rating of “medium” while we evaluated the others either as “low” or “uninformative”. Among the 8 studies that found associations with BPA, 5 were “medium to high”, and 2 “medium” and one study was rated “low”. Due to shortcomings in exposure characterisation (spot urine samples) no study achieved a confidence rating of “high”.

We attributed the disparity between the 3 “medium” confidence null studies (Caporossi et al., 2020; Goldstone et al., 2015; Mendiola et al., 2010) and the “medium” and “medium to high” studies that reported associations (Adoamnei et al., 2018; Ji et al., 2018; Lassen et al., 2014; Li et al., 2011; Meeker et al., 2010; Pollard et al., 2019; Radwan et al., 2018) to differences in exposures: All the null studies examined populations with quite low BPA urinary levels. We concluded that the differing outcomes cannot be viewed as conflicting evidence (defined as unexplained positive and negative results in similarly exposed human populations) but that the mixed results are explained by differing exposure levels, where low BPA exposures will have precluded the detection of effects on semen quality.

In summary, we identified 7 independent studies with a “medium” or “medium to high” confidence rating that reported associations between BPA urinary levels and declining semen quality. According to the scheme developed by Radke et al. (2018), the overall strength of evidence can be evaluated as “robust”. We noted the absence of epidemiological studies of gestational BPA exposures and declines in semen quality.

We attempted to compare BPA exposures apparently no longer associated with deteriorations in semen quality with the range of RfD from animal studies. By focusing on studies among the general

population (excluding occupationally exposed cohorts and populations from fertility clinics) we pinpointed four studies as useful for such comparisons: Adoamnei et al. (2018), Pollard et al. (2019), Ji et al. (2018) and Lassen et al. (2014). In these studies, BPA urinary levels were categorised into ranges of urinary concentrations. We identified ranges no longer associated with statistically significant declines in semen quality and converted the corresponding urinary BPA concentrations into estimated daily intakes. This analysis showed that BPA exposures above a range of between 0.01 and 0.18 $\mu\text{g}/\text{kg}/\text{d}$ are associated with declines in semen quality (Kortenkamp et al. 2022 a). While the present EFSA HBGV of 0.2 $\text{ng}/\text{kg}/\text{d}$ and RfDs derived from the Wang et al. (2016) study of between 0.001 and 0.0066 $\mu\text{g}/\text{kg}/\text{d}$ are below this range, the BfR TDI of 0.2 $\mu\text{g}/\text{kg}/\text{d}$ exceeds these values and even falls into ranges where some studies (Adoamnei et al., 2018; Ji et al., 2018; Pollard et al., 2019) observed reduced semen quality (Fig. 1).

4. Discussion

Our analysis exposes the approaches used for evaluating the likelihood of effects arising from BPA exposures to human health as problematic. Driven by the need for dose-response analyses and hazard characterizations essential for deriving a HBGV or TDI, EFSA (2023) and BfR (2023) adopted a parcellated and fragmented endpoint-by-endpoint study evaluation scheme in which entire ranges of effects were lost from consideration. Striking is the omission of epidemiological evidence from further analysis. These data did not feature in uncertainty considerations. The same applies to the evidence of the effects of gestational exposures in animal studies. A true synthesis of all the evidence was not achieved with these approaches. While these omissions did not affect the revised EFSA HBGV (as it is derived from immunotoxicity), they somewhat compromise the scientific credibility of the alternative BfR TDI and expose the fact that many decision points in the entire process appear to be value judgements, not based on purely scientific considerations. An important question will be how these schemes can be improved to avoid such paradoxical situations and to clearly demarcate science from value judgements.

4.1. Study evaluation schemes and gestational versus adult BPA exposures

Considering the importance of the gestational period in determining semen quality, it is striking that neither EFSA nor BfR evaluated the evidence for semen quality deteriorations after gestational BPA administrations more closely.

Because of their possible impacts on germ stem cell populations which are established in fetal and neonatal life, gestational BPA exposures are critical. Only after this period can spermatogenesis begin. Disruption of these processes can have life-long, irreversible effects.

Several independent studies from multiple laboratories support the conclusion that BPA produces qualitative and quantitative alterations of spermatogenesis, by disrupting germ cell meiosis, apoptosis and other processes. These alterations materialize in terms of decreased number and motility of spermatozoa (Salian et al., 2009; Shi et al., 2018, 2019; Ullah et al., 2019; Vilela et al., 2014; Yang et al., 2015) as well as changes in testicular histology, especially of stages VII – VIII in the spermatogenic cycle of the seminiferous epithelium (Rahman et al., 2017; Salian et al., 2009; Shi et al., 2018, 2019; Ullah et al., 2019). There are further effects, including disruption of the epigenetic programming necessary for spermatogenesis, as evidenced by gene expression changes in DNA methyl transferases (Shi et al., 2019) and increased oxidative stress in testicular tissues (Ullah et al., 2019).

Physiologically, the consequences of BPA exposures during gestation manifest as a spectrum of inter-related effects. In recognition of this, it makes little sense to evaluate each component of this spectrum separately, one-by-one, in terms of sperm counts, sperm motility, hormonal changes, testis histology etc. to arrive at conclusions about the relevance

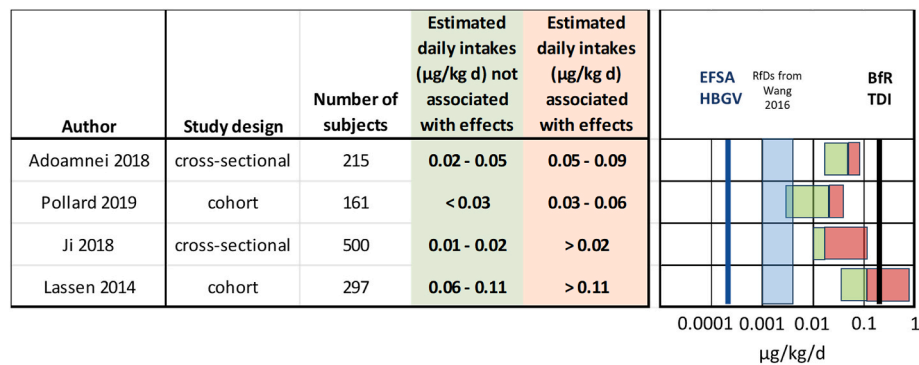


Fig. 1. Comparison of BPA daily intakes no longer associated with decrements in semen quality in epidemiological studies with various regulatory values. Shown are estimated daily BPA intakes derived from urinary BPA levels in epidemiological studies, using the procedure detailed in Kortenkamp et al. (2022). Values underlaid green are intakes not associated with deteriorations in semen quality (“epidemiological NOAELs”), those underlaid red are the first quantiles associated with significant effects (Adoamnei et al., 2018: decreases in sperm concentration and total sperm count; Pollard et al., 2019: morphological changes; Ji et al., 2018: sperm concentration; Lassen et al., 2014: sperm motility). The green bars on the intake grid (right) correspond to the “epidemiological NOAELs”, the red bars correspond to the intake ranges associated with deteriorations of semen quality. The vertical lines are the EFSA HBGV (blue) and the alternative BfR TDI (black). The range of RfD derived from Wang et al. (2016) is shown as a blue box. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

of effects for human hazard characterisation. The reductionist, fragmented, endpoint-oriented study evaluation system used by EFSA (2023) has obscured the overall effect pattern and has led to the omission of gestational BPA exposures from further consideration. BfR (2023) have uncritically adopted EFSA’s conclusions on this matter, without further assessment of the literature, and focused solely on the consequences of BPA exposures in adulthood.

In EFSA’s and BfR’s study quality rating systems, many academic studies were downgraded to Tier 3. Often, these decisions were based on deficits in the reporting of concealment methods (randomisation and blinding) and shortcomings in the characterisation of background contamination. The study rating applied by EFSA was not consistent across endpoints, as some studies downgraded to Tier 3 for certain effects, were placed in Tier 2 in others, despite having the same “deficits” (Shi et al., 2018 is an example: Tier 3 for sperm counts, Tier 2 for testes histology changes).

As the differences between EFSA and BfR in the quality rating of the studies by Wang et al. (2016) and Srivastava and Gupta (2018) show, the system does not necessarily produce consistent evaluations. The differences between our (Kortenkamp et al. 2022 a) and EFSA’s (2023) study quality ratings of some gestational BPA studies also show that the system is open to interpretation, complicated by the fact that the reporting of study details is often not clear and evidence for any existing bias frequently indirect. In any case, by excluding the Wang et al. study of sperm motility from further consideration, BfR set themselves on a path of considering only studies of sperm counts that reported effects at higher doses.

The diverging interests of academic researchers and risk assessors contribute to these problems. The primary motivation of many academic studies is in elucidating mechanisms, rather than “merely” providing data for hazard identification and dose-response analysis. The pressures of academic publishing are such that “descriptive” studies such as dose-response studies are viewed as being of limited originality, and accordingly are not likely to produce high impact papers. As a result, issues that take centre stage in regulatory testing, such as randomisation, blinding etc. are neglected in academic studies. On the other hand, authors of academic studies often do not fully appreciate how the impact of their work could be increased, if only they paid more attention to issues such as exposure characterisation, inclusion of positive controls, blinding etc. A solution might be to implement quality criteria for submissions to toxicology journals that take account of regulatory requirements, as elaborated by Martin et al. (2019).

4.2. Study selection

Several factors explain the rather high value of BfR’s alternative TDI. Apart from their refusal to accept immunotoxic effects as the basis for deriving a HBGV (not further discussed here), BfR’s choice of Liu et al. (2013) and Srivastava and Gupta (2018) as the key studies for semen quality deteriorations, rather than Wang et al. (2016), as in EFSA (2023), is critical. Their selection of benchmark dose modelling with Bayesian model averaging as the basis for a regulatory value and their use of rather low inter-species and intra-species AFs in probabilistic modelling also drove up their TDI. The use of such low factors is a value judgement, poorly supported by the available evidence (Martin et al., 2013). Conversely, EFSA’s adoption of Wang et al. (2016) as the key study and their choice of a rather high inter-species AF factor for mouse-to-human extrapolations led to correspondingly lower RfDs for semen quality deteriorations. Similarly, our preference for deriving an RfD from gestational BPA exposures (Kortenkamp et al. 2022 a) has produced a value considerably lower than BfR’s alternative TDI. The differences between our estimate and the values derived by EFSA (2023) for adult exposures from Wang et al. (2016) are far less pronounced.

BfR’s and EFSA’s conclusion that there is sufficient evidence from animal studies to assume a causal link between adult BPA exposures and deteriorations of semen quality is well founded. If this is accepted as the basis for further deliberations, the question becomes whether to choose Wang et al. (2016) or Liu et al. (2013) as the key studies for deriving a regulatory value. As already discussed, there are some deficiencies in both these studies, but overall, it is arguable that they are of sufficient quality to support the derivation of regulatory values. To meet the demand for being sufficiently protective, the deciding factor then must be that Wang et al. (2016) demonstrated effects at lower doses. An additional consideration should be that dose-response modelling is better supported by the data from Wang et al. because all chosen doses were associated with effects, while only the highest dose in Liu et al. produced observable effects.

4.3. Dose-response analyses

This latter factor is at the heart of some contradictions that arise from the use of the Liu et al. data by BfR in their benchmark dose (BMD) modelling. According to EFSA (2022) guidance, at least three doses associated with responses significantly different from each other are needed for reliable BMD estimations. Under certain circumstances, data with only two statistically significantly different responses can also be

considered adequate. Since only the largest of three doses produced an effect, the Liu et al. data conform with these requirements. However, the resulting low degree of definition and granularity in the dose-response data severely limits the choice of suitable regression models for deriving a BMD. Of the four general approaches available for estimating a BMD (i.e. application of a single dose-response model, a best-fitting model, frequentist model averaging, and Bayesian model averaging), EFSA (2022) recommends Bayesian model averaging. In general, BMDs and BMDLs derived from frequentist and Bayesian model averaging (if no informative priors are used) should produce similar values which in turn should be comparable to the corresponding NOAELs (EFSA, 2022). We assume that BfR used uninformative priors in Bayesian model averaging as recommended by the default setting in EFSA's web applications for statistical models (<https://r4eu.efsa.europa.eu/>) and therefore expected no marked differences in the BMDs and BMDLs derived from both model averaging approaches. However, the BMDLs which BfR obtained from the Liu et al. data differ considerably, with the Bayesian model averaging producing a value 13 times larger than that from frequentist model averaging. Furthermore, the BMD estimates which BfR derived from frequentist model averaging were quite unstable, as suggested by the high ratio between the upper and lower limits of BMDs (in excess of 50). In such cases, EFSA (2022) guidance advises the use of other points of departure for deriving a HBGV. This led BfR to base their TDI on the rather large BMDL produced by Bayesian model averaging (26 µg/kg/d). Although the ratio between the upper and lower BMD was acceptable in this case, this value is, rather unusually, larger than the corresponding NOAEL (20 µg/kg/d). These striking discrepancies were neither discussed nor explained by BfR. The other sensible option of using the NOAEL from Liu et al. as the basis for a TDI was not pursued by BfR. The Liu et al. NOAEL of 20 µg/kg/d translates into a deterministic RfD of 0.065 µg/kg/d and a probabilistic value of 0.044 µg/kg/d, approximately 2- to 3-times lower than BfR's value of 0.14 µg/kg/d (Tables 3 and 4). These differences were ignored by BfR (2023), but at the very least should have been the topic of uncertainty considerations.

4.4. Interspecies assessment factors

While the dispute between EFSA and BfR about the interspecies AF for the mouse is not material for BfR's suggested alternative TDI (the relevant studies are in rats), we share some of BfR's concerns regarding the large EFSA mouse AF of 64.5. In relying exclusively on the Doerge et al. (2011) data, EFSA chose a somewhat weak basis for their selection. The mouse interspecies AF derived from allometric scaling in the WHO APROBA tool is perhaps more appropriate. However, the AF chosen by BfR for rat-to-human extrapolations in the probabilistic evaluation appears to be unrealistically low. Derived from permutations of rat and human AUCs, the median of 1.98 of the proposed log-normal distribution implies that humans are on average only about two times more sensitive to BPA than rats, far less than suggested by allometric scaling in the APROBA tool. The ratio between the 5th (P05) and the 95th (P95) percentile of the log-normal distribution is an expression of the degree of uncertainty in extrapolating from the rat to humans. Therefore, the choice of the values for the tails of the distribution (P05 and P95) should be supported by data of sufficient quality. It is highly questionable whether only six (not independent) HEDF values can provide a sound basis for such estimations. Even more puzzling is BfR's choice of AF for sensitivity differences between humans, the intra-species AF. It is normally assumed that sensitive humans on average react to 10-times lower exposures than more resilient subjects, expressed by the median of the APROBA values (Table 4). The range chosen by BfR, with a median of 1, assumes that these differences do not exist. In selecting their values, BfR reference ECHA (2012) and EFSA (2012) guidance on the matter, but we were unable to locate these values in the references provided by BfR.

4.5. Conclusions

Switching from immunotoxicity to declines in semen quality after adult exposures as the critical toxicity for deriving a BPA HBGV would produce values 33-times higher (best estimate, derived from Wang et al. by using a probabilistic approach, Table 4) to 2.5 – 5-times higher (lowest estimates, from Wang et al. with deterministic approaches and the large EFSA mouse-to-human AF, Table 3) than the present EFSA HBGV of 0.2 ng/kg/d. Consideration of gestational exposures would give RfDs 12-times higher (best estimate, derived from Ullah et al., 2019 using a probabilistic method, Table 4). As detailed before (Kortenkamp et al. 2022 a), we maintain that a study of gestational BPA exposures is the most suitable basis for deriving a RfD for semen quality deteriorations, and that the Ullah et al. study is key. On the basis of that study, a value of 2–4 ng/kg/d is appropriate. The data provided by Shi et al. (2018) will lead to an even lower value.

Our deconstruction of the proposed alternative BfR TDI shows that this value is the result of a procedure in which less protective choices were made consistently at every possible turn: first, in selecting the key studies, then in settling for BMD modelling as a way of providing a PoD, ignoring a NOAEL as an alternative, and finally by adopting a low interspecies AF derived through a questionable procedure. The consequence is a value which comes very close to BPA doses empirically shown to produce semen quality deteriorations in animal studies of gestational and adult exposures. This TDI also exceeds exposures associated with poor semen quality in epidemiological studies.

4.6. Suggestions for improvements

To avoid losing sight of the entire effect patterns, especially when dealing with syndrome-like effects such as those relevant to male reproductive health, it will be necessary to adopt a system that takes account of **all** the evidence, initially without deleting from consideration data from Tier 3 studies. A way forward would be to split the evaluation process into a qualitative and a quantitative part, in which the primary aim should be to first evaluate the question “what is the strength of evidence linking BPA to the effect of interest”. In this step, **all** studies should be considered, initially independent of matters relevant to hazard characterisations such as number of doses used and other aspects. Only in the next, the quantitative step, should issues concerning dose-response analyses come to the fore. In this way, studies rated as less reliable from a hazard characterisation viewpoint are not entirely deleted from the analysis. Such a system might safeguard against the paradoxical situation that has arisen particularly with the BfR approach, where RfDs or HBGV derived from “high quality” studies come too close to the dose ranges associated with effects in animal studies, even though these may not be suitable for hazard characterisations. A summary of studies not further considered for dose-response analyses will be essential as a corrective, and this must be taken account of in uncertainty analyses.

Declaration of competing interest

Andreas Kortenkamp, Martin Scholze and Eleni Iacovidou declare they have no conflicts of interests. Olwenn V Martin reports a relationship with European Chemicals Agency that includes board membership and with the Food Packaging Forum that includes board membership and consulting or advisory services.

Acknowledgements

This work was conducted without external funding.

References

- Adoamnei, E., Mendiola, J., Vela-Soria, F., Fernández, M.F., Olea, N., Jørgensen, N., Swan, S.H., Torres-Cantero, A.M., 2018. Urinary bisphenol A concentrations are associated with reproductive parameters in young men. *Environ. Res.* 161, 122–128. <https://doi.org/10.1016/j.envres.2017.11.002>.
- Beausoleil, C., Le Magueresse-Battistoni, B., Vigüé, C., Babajko, S., Canivenc-Lavier, M. C., Chevalier, N., Emond, C., Habert, R., Picard-Hagen, N., Mhaouty-Kodja, S., 2022. Regulatory and academic studies to derive reference values for human health: the case of bisphenol S. *Environ. Res.* 204 <https://doi.org/10.1016/j.envres.2021.112233>.
- BfR, Bundesinstitut für Risikobewertung, 2023. Bisphenol A: BfR schlägt gesundheitsbasierten Richtwert vor, für eine vollständige Risikobewertung werden aktuelle Expositionsdaten benötigt. Stellungnahme Nr. 018/2023 des BfR vom 19. April 2023. <https://doi.org/10.17590/20230419-111937-0>.
- Camacho, L., Lewis, S.M., Vanlandingham, M.M., Olson, G.R., Davis, K.J., Patton, R., Twaddle, N.C., Doerge, D.R., Churchwell, M.I., Bryant, M.S., McLellen, F.M., Woodling, K., Felton, R.P., Maisha, M.P., Juliar, B.E., Gamboa da Costa, G., Delclos, K.B., 2019. NTP CLARITY-BPA report (2018). A two-year toxicology study of bisphenol A (BPA) in Sprague-Dawley rats: CLARITY-BPA core study results. *Food Chem. Toxicol.* 132, 110728.
- Caporossi, L., Alteri, A., Campo, G., Paci, E., Tranfo, G., Capanna, S., Papaleo, E., Pignini, D., Viganò, P., Papaleo, B., 2020. Cross sectional study on exposure to BPA and phthalates and semen parameters in men attending a fertility center. *Int. J. Environ. Res. Publ. Health* 17. <https://doi.org/10.3390/ijerph17020489>.
- Chatsantiprapa, K., Sophon, T., Sattayasai, J., 2016. Effects of continuous exposure to bisphenol A on male and female mice from. *Prenatally to Adulthood* 40, 61–69.
- Chiu, W.A., Slob, W., 2015. A unified probabilistic framework for dose-response assessment of human health effects. *Environ. Health Perspect.* 123 (12), 1241–1254. <https://doi.org/10.1289/ehp.1409385>.
- Doerge, D.R., Twaddle, N.C., Vanlandingham, M., Fisher, J.W., 2011. Pharmacokinetics of bisphenol A in neonatal and adult CD-1 mice: inter-species comparisons with Sprague-Dawley rats and rhesus monkeys. *Toxicol. Lett.* 207 (3), 298–305. <https://doi.org/10.1016/j.toxlet.2011.09.020>.
- Domoradzki, J.Y., Pottenger, L.H., Thornton, C.M., Hansen, S.C., Card, T.L., Markham, D. A., Dryzga, M.D., Shiotsuka, R.N., Waechter Jr., J.M., 2003. Metabolism and pharmacokinetics of bisphenol A (BPA) and the Embryo-fetal distribution of BPA and BPA-mono-glucuronide in CD sprague-dawley rats at three gestational stages. *Toxicol. Sci.* 76 (1), 21–34. <https://doi.org/10.1093/toxsci/kfg206>.
- EFSA, 2012, 2012. Scientific Committee; Guidance on selected default values to be used by the EFSA Scientific Committee, Scientific Panels and Units in the absence of actual measured data. *EFSA J.* 2579, 32. <https://doi.org/10.2903/j.efsa.2012.2579>, 10(3).
- EFSA, Gundert-Remy, U., Bodin, J., Bosetti, C., Fitzgerald, R., Hanberg, A., Hooijmans, C., Rooney, A.A., Rousselet, C., Loveren, H. Van, Barizzone, F., Croera, C., Pitzu, C., Castoldi, A.F., 2017. Bisphenol A (BPA) hazard assessment protocol. *EFSA Supporting Publications* 14 (12). <https://doi.org/10.2903/sp.efsa.2017.en-1354>.
- EFSA, 2022. EFSA Scientific Committee, More, S.J., Bampidis, V., Benford, D., Bragard, C., Halldorsson, T.I., Hernandez-Jerez, A.F., Bennekou, S.H., Koutsoumanis, K., Lambré, C., Mächera, K., Mennes, W., Mullins, E., Nielsen, S.S., Schrenk, D., Turck, D., Younes, M., Aerts, M., Edler, L., Sand, S., Wright, M., Binaglia, M., Bottex, B., Cortinas Abrahantes, J., Schlatter, J., 2022. Guidance on the use of the benchmark dose approach in risk assessment, 2022 EFSA J. 7584, 67. <https://doi.org/10.2903/j.efsa.2022.7584>, 20(10).
- EFSA, 2023. Re-evaluation of the risks to public health related to the presence of bisphenol A (BPA) in foodstuffs. Panel on Food Contact Material, Enzymes and Processing Aids (CEP). *EFSA J.* <https://doi.org/10.2903/j.efsa.2023.6857>.
- EFSA, BfR, 2022. Report on diverging opinions between EFSA and BfR on updated bisphenol A assessment. <https://www.efsa.europa.eu/sites/default/files/2023-04/bfr-efsa-art-30.pdf>.
- EFSA, EMA, 2022. Report on diverging opinions between EFSA and EMA on bisphenol-A final adopted. <https://www.ema.europa.eu/en/documents/report/report-divergent-opinion-between-efsa-ema-bisphenol-en.pdf>.
- Goldstone, A.E., Chen, Z., Perry, M.J., Kannan, K., Louis, G.M.B., 2015. Urinary bisphenol A and semen quality, the LIFE study. *Reprod. Toxicol.* 51, 7–13. <https://doi.org/10.1016/j.reprotox.2014.11.003>.
- Hass, U., Christiansen, S., Boberg, J., Rasmussen, M.G., Mandrup, K., Axelstad, M., 2016. Low-dose effect of developmental bisphenol A exposure on sperm count and behaviour in rats. *Andrology* 4, 594–607. <https://doi.org/10.1111/andr.12176>.
- Ji, H., Miao, M., Liang, H., Shi, H., Ruan, D., Li, Y., Wang, J., Yuan, W., 2018. Exposure of environmental Bisphenol A in relation to routine sperm parameters and sperm movement characteristics among fertile men. *Sci. Rep.* 8, 1–9. <https://doi.org/10.1038/s41598-018-35787-5>.
- Kalb, A.C., Kalb, A.L., Cardoso, T.F., Fernandes, C.G., Corcini, C.D., Junior, A.S.V., Martínez, P.E., 2016. Maternal transfer of bisphenol A during nursing causes sperm impairment in male offspring. *Arch. Environ. Contam. Toxicol.* 70 (4), 793–801. <https://doi.org/10.1007/s00244-015-0199-7>.
- Karrer, C., Andreassen, M., von Goetz, N., Sonnet, F., Sakhi, A.K., Hungerbühler, K., Dirven, H., Husoy, T., 2020. The EuroMix human biomonitoring study: source-to-dose modeling of cumulative and aggregate exposure for the bisphenols BPA, BPS, and BPF and comparison with measured urinary levels. *Environ. Int.* 136, 105397. <https://doi.org/10.1016/j.envint.2019.105397>.
- Kortenkamp, A., Martin, O., Ermiler, S., Baig, A., Scholze, M., 2022 a. Bisphenol A and declining semen quality: a systematic review to support the derivation of a reference dose for mixture risk assessments. *Int. J. Hyg Environ. Health* 241, 113942. <https://doi.org/10.1016/j.ijheh.2022.113942>.
- Kortenkamp, A., Scholze, M., Ermiler, S., Priskorn, L., Joergensen, N., Andersson, A.M., Frederiksen, H., 2022 b. Combined exposures to bisphenols, polychlorinated dioxins, paracetamol and phthalates as drivers of deteriorating semen quality. *Environ. Int.* 165, 107322.
- Lassen, T.H., Frederiksen, H., Jensen, T.K., Petersen, J.H., Joensen, U.N., Main, K.M., Skakkebaek, N.E., Juul, A., Jørgensen, N., Andersson, A.M., 2014. Urinary bisphenol A levels in young men: association with reproductive hormones and semen quality. *Environ. Health Perspect.* 122, 478–484. <https://doi.org/10.1289/ehp.1307309>.
- Li, D.K., Zhou, Z., Miao, M., He, Y., Wang, J., Ferber, J., Herrinton, L.J., Gao, E., Yuan, W., 2011. Urine bisphenol-A (BPA) level in relation to semen quality. *Fertil. Steril.* 95, 625–630.e4. <https://doi.org/10.1016/j.fertnstert.2010.09.026>.
- Liu, C., Duan, W., Li, R., Xu, S., Zhang, L., Chen, C., He, M., Lu, Y., Wu, H., Pi, H., Luo, X., Zhang, Y., Zhong, M., Yu, Z., Zhou, Z., 2013. Exposure to bisphenol A disrupts meiotic progression during spermatogenesis in adult rats through estrogen-like activity. *Cell Death Dis.* 4 (6) <https://doi.org/10.1038/cddis.2013.203>.
- Martin, O.V., Scholze, M., Kortenkamp, A., 2013. Dispelling urban myths about default uncertainty factors in chemical risk assessment – sufficient protection against mixture effects? *Environ. Health* 12, 53.
- Martin, O.V., Adams, J., Beasley, A., Belanger, S., Breton, R.L., Brock, T.C.M., Buonsante, V.M., Galay Burgos, M., Green, J., Guiney, P.D., Hall, T., Hanson, M., Harris, M.J., Henry, T.R., Huggett, D., Junghans, M., Laskowski, R., Maack, G., Moermond, C.T.A., Panter, G., Pease, A., Poulsen, V., Roberts, M., Rudén, C., Schlekot, C.E., Schoeters, I., Solomon, K.R., Staveley, J., Stubblefield, B., Sumpter, J. P., Warne, M.J., Wentsel, R., Wheeler, J.R., Wolff, B.A., Yamazaki, K., Zahner, H., Ågerstrand, M., 2019. Improving environmental risk assessments of chemicals: steps towards evidence-based ecotoxicology. *Environ. Int.* 128, 210–217. <https://doi.org/10.1016/j.envint.2019.04.053>.
- Meeker, J.D., Ehrlich, S., Toth, T.L., Wright, D.L., Calafat, A.M., Trisini, A.T., Ye, X., Hauser, R., 2010. Semen quality and sperm DNA damage in relation to urinary bisphenol A among men from an infertility clinic. *Reprod. Toxicol.* 30, 532–539. <https://doi.org/10.1016/j.reprotox.2010.07.005>.
- Mendiola, J., Jørgensen, N., Andersson, A.M., Calafat, A.M., Ye, X., Redmon, J.B., Drobnis, E.Z., Wang, C., Sparks, A., Thurston, S.W., Liu, F., Swan, S.H., 2010. Are environmental levels of bisphenol A associated with reproductive function in fertile men? *Environ. Health Perspect.* 118, 1286–1291. <https://doi.org/10.1289/ehp.1002037>.
- NTP, 2015. Handbook for Conducting a Literature-Based Health Assessment Using OHAT Approach for Systemic Review and Evidence Integration, pp. 1–98.
- Pollard, S.H., Cox, K.J., Blackburn, B.E., Wilkins, D.G., Carrell, D.T., Stanford, J.B., Porucznik, C.A., 2019. Male exposure to bisphenol A (BPA) and semen quality in the home observation of periconceptional exposures (HOPE) cohort. *Reprod. Toxicol.* 90, 82–87. <https://doi.org/10.1016/j.reprotox.2019.08.014>.
- Pottenger, L.H., Domoradzki, J.Y., Markham, D.A., Hansen, S.C., Cagen, S.Z., Waechter Jr., J.M., 2000. The relative bioavailability and metabolism of bisphenol A in rats is dependent upon the route of administration. *Toxicol. Sci.* 54 (1), 3–18. <https://doi.org/10.1093/toxsci/54.1.3>.
- Radke, E.G., Braun, J.M., Meeker, J.D., Cooper, G.S., 2018. Phthalate exposure and male reproductive outcomes: a systematic review of the human epidemiological evidence. *Environ. Int.* 121, 764–793. <https://doi.org/10.1016/j.envint.2018.07.029>.
- Radwan, M., Wielgomas, B., Dziejwirska, E., Radwan, P., Kaluźny, P., Klimowska, A., Hanke, W., Jurewicz, J., 2018. Urinary bisphenol A levels and male fertility. *Am. J. Men's Health* 12, 2144–2151. <https://doi.org/10.1177/1557988318799163>.
- Rahman, M.S., Kwon, W.S., Karmakar, P.C., Yoon, S.J., Ryu, B.Y., Pang, M.G., 2017. Gestational exposure to bisphenol A affects the function and proteome profile of F1 spermatozoa in adult mice. *Environ. Health Perspect.* 125, 238–245. <https://doi.org/10.1289/EHP378>.
- Salian, S., Doshi, T., Vanage, G., 2009. Perinatal exposure of rats to bisphenol A affects the fertility of male offspring. *Life Sci.* 85, 742–752.
- Shi, M., Sekulovski, N., MacLean, J.A., Hayashi, K., 2018. Prenatal exposure to bisphenol A analogues on male reproductive functions in mice. *Toxicol. Sci.* 163, 620–631. <https://doi.org/10.1093/toxsci/kfy061>.
- Shi, M., Whorton, A.E., Sekulovski, N., Maclean, J.A., Hayashi, K., 2019. Prenatal exposure to bisphenol A, E, and S induces transgenerational effects on male reproductive functions in mice. *Toxicol. Sci.* 172, 303–315. <https://doi.org/10.1093/toxsci/kfz207>.
- Sieli, P.T., Jašarević, E., Warzak, D.A., Mao, J., Ellersieck, M.R., Liao, C., Kannan, K., Collet, S.H., Toutain, P.L., Saal, F.S.V.S., Rosenfeld, C.S., 2011. Comparison of serum bisphenol A concentrations in mice exposed to bisphenol a through the diet versus oral bolus exposure. *Environ. Health Perspect.* 119 (9), 1260–1265. <https://doi.org/10.1289/ehp.1003385>.
- Srivastava, S., Gupta, P., 2018. Alteration in apoptotic rate of testicular cells and sperms following administration of Bisphenol A (BPA) in Wistar albino rats. *Environ. Sci. Pollut. Control Ser.* 25 (22), 21635–21643. <https://doi.org/10.1007/s11356-018-2229-2>.
- Taylor, J.A., vom Saal, F.S., Welshons, W.V., Drury, B., Rottinghaus, G., Hunt, P.A., Toutain, P.L., Laffont, C.M., VandeVoort, C.A., 2011. Similarity of bisphenol A pharmacokinetics in rhesus monkeys and mice: relevance for human exposure. *Environ. Health Perspect.* 119 (4), 422–430. <https://doi.org/10.1289/ehp.1002514>.
- Teeguarden, J.G., Twaddle, N.C., Churchwell, M.I., Yang, X., Fisher, J.W., Seryak, L.M., Doerge, D.R., 2015. 24-hour human urine and serum profiles of bisphenol A: evidence against sublingual absorption following ingestion in soup. *Toxicol. Appl. Pharmacol.* 288 (2), 131–142. <https://doi.org/10.1016/j.taap.2015.01.009>.
- Thayer, K.A., Doerge, D.R., Hunt, D., Schurman, S.H., Twaddle, N.C., Churchwell, M.I., Garantzios, S., Kissling, G.E., Easterling, M.R., Bucher, J.R., Birnbaum, L.S., 2015.

- Pharmacokinetics of bisphenol A in humans following a single oral administration. *Environ. Int.* 83, 107–115. <https://doi.org/10.1016/j.envint.2015.06.008>.
- Ullah, A., Pirzada, M., Jahan, S., Ullah, H., Razak, S., Rauf, N., 2019. Prenatal BPA and its Analogs BPB, BPF, and BPS Exposure and Reproductive axis Function in the Male Offspring of Sprague Dawley Rats. <https://doi.org/10.1177/0960327119862335>.
- Vilela, J., Hartmann, A., Silva, E.F., Cardoso, T., Corcini, C.D., Varela-Junior, A.S., Martinez, P.E., Colares, E.P., 2014. Sperm impairments in adult vesper mice (*Calomys cauda*) caused by *in utero* exposure to bisphenol A. *Andrologia* 46, 971–978. <https://doi.org/10.1111/and.12182>.
- Wang, H.F., Liu, M., Li, N., Luo, T., Zheng, L.P., Zeng, X.H., 2016. Bisphenol A impairs mature sperm functions by a CatSper-relevant mechanism. *Toxicol. Sci.* 152 (1), 145–154.
- WHO/IPSC (World Health Organisation & International Programme on Chemical Safety), 2018. Guidance Document on Evaluating and Expressing Uncertainty in Hazard Characterization, second ed. World Health Organization, Geneva <https://apps.who.int/iris/handle/10665/259858>.
- Yin, L., Dai, Y., Cui, Z., Jiang, X., Liu, W., Han, F., Lin, A., Cao, J., Liu, J., 2017. The regulation of cellular apoptosis by the ROS-triggered PERK/EIF2 α /chop pathway plays a vital role in bisphenol A-induced male reproductive toxicity. *Toxicol. Appl. Pharmacol.* 314, 98–108. <https://doi.org/10.1016/j.taap.2016.11.013>.
- Yang, Y., Hong, Y., Chae, S.A., 2015. Reduction in Semen Quality after Mixed Exposure to Bisphenol A and Isobutylparaben in Utero and during Lactation Periods, pp. 1–10. <https://doi.org/10.1177/0960327115608927>.