OPINION

Endocrine disruptors and the future of toxicology testing — lessons from CLARITY–BPA

Laura N. Vandenberg, Patricia A. Hunt and Andrea C. Gore

Abstract | Five years ago, an ambitious collaboration, the Consortium Linking Academic and Regulatory Insights on Toxicity of BPA (CLARITY–BPA; henceforth CLARITY), was launched by three US agencies. The goal was to provide a definitive evaluation of bisphenol A (BPA) and explain disparities between traditional regulatory studies and findings from independent investigators. BPA or vehicle-treated rats from an FDA facility were used in a guideline study and animals and/or tissues were provided to academic researchers for analysis. An interim summary released in February 2018 by the FDA concluded that currently authorized uses of BPA continue to be safe. We disagree. In this Perspectives, we summarize the goals, design and problems of CLARITY. We conclude that, despite its flaws, CLARITY provides important insight and, taken together, the data provide compelling evidence that low-dose BPA exposure induces marked adverse effects. Indeed, the greatest number of effects were observed at doses 20,000 times lower than the current 'safe' dose of BPA for humans.

Bisphenol A (BPA) is an endocrine-disrupting chemical (EDC) that is in widespread use in plastics, thermal receipts, food packaging, toys and many other applications[1]. A vast and ever-increasing number of peer-reviewed studies have reported adverse effects of low-dose BPA exposure, particularly during development[2–5]. However, traditional toxicity studies designed to identify hazards associated with BPA have reached very different conclusions about BPA's safety[6].

Five years ago, an ambitious collaborative project called the Consortium Linking Academic and Regulatory Insights on Toxicity of BPA (CLARITY–BPA; henceforth CLARITY) was launched by three US federal agencies: the FDA, the NIH National Institute of Environmental Health Sciences (NIEHS) and the National Toxicology Program (NTP)[7]. CLARITY had the goal of designing and implementing the definitive BPA study, which would provide insight into the risks posed by this common environmental contaminant, and illuminate the potential reasons for disparities between studies, and reconcile these outcomes.

Risk assessment traditionally has relied upon the results of guideline studies (BOX 1) to make determinations about the safety of chemicals and to calculate doses that are expected to be safe for human exposure[8]. Guideline studies follow validated protocols, agreed upon by international groups, and typically examine overt signs of toxicity (for example, changes in organ and body weight, pup survival and histopathology of selected target tissues[9]). By contrast, academic studies of EDCs are typically designed to test hypotheses about the effects of exposure on particular end points for which the laboratory has expertise. For BPA, discordance between findings from guideline and academic studies has made safety a contentious issue. Most guideline study results suggest marked effects of BPA on outcomes such as liver and kidney weight only at very high doses[10–12]; however, many experimental studies on BPA conducted by independent academic investigators have found evidence of adverse outcomes at much lower doses[3–5,13]. This discrepancy has sparked intense debate about the utility of current guideline studies for evaluating EDCs and whether these studies can be used to accurately predict 'safe' doses of exposure[14–17]. This point is particularly pertinent when considering exposure during development, when EDCs can induce subtle changes that might not have any immediate effect but that might have repercussions for adult health and fertility. On the basis of this idea, it has been suggested that a no observed adverse effect level (NOAEL) for EDCs might not exist[18].

CLARITY (BOX 1) was designed to directly address these issues by evaluating traditional measures of toxicity using a guideline study design conducted at a core facility, with animals or tissues from these animals provided to academic researchers for hypothesis testing about human disease. Given the scope of the mission, the price tag for the project was large (an estimated cost of US$15 million for the academic studies alone), but combining the forces of federal and university laboratories was a bold, innovative first step in moving regulatory testing into the 21st century. The individual academic studies comprising CLARITY are nearly all complete, and many of the data have been published[19–28]. With respect to the guideline studies, an interim report was posted online and the FDA put out a press release in February 2018 (REFS[29,30]). The finalized report with all results from the guideline study is now publicly available[31], and the FDA hosted a webinar in September 2018 to present the final conclusions reached in its evaluation of the core study results[32]. Whether the goal of the CLARITY project has been accomplished remains the subject of considerable debate. The true value of the endeavour, however, will probably not be the insight it provides about the effects and potential safety of BPA but, rather, the lessons that can be learned from this collaborative approach.

In this Perspectives, we provide a third-party synthesis of the CLARITY study from the perspective of three individuals not involved in the project but who have a long-standing interest in BPA and the challenges EDCs pose for traditional toxicology-based approaches to

Box 1 | **Nomenclature**

Studies

- CLARITY–BPA — Consortium Linking Academic and Regulatory Insights on Toxicity of Bisphenol A
- EE–NTP study — a prior guideline study on the toxicity and carcinogenesis of ethinyl estradiol, also conducted by the National Center for Toxicological Research (NCTR) and National Toxicology Program
- Pilot 90-day study — a subchronic study conducted at the NCTR in advance of CLARITY to work out experimental conditions and doses

Terminology pertaining to CLARITY

- Guideline study — a study conducted using validated protocols, agreed upon by international groups, that typically examines overt signs of toxicity
- Core study — in CLARITY, the core study was conducted following guideline protocols at the NCTR
- Academic studies — in CLARITY, the 14 investigators at US institutions who received animals or tissues prepared at the core institute for further testing of experimental hypotheses
- STOP-DOSE — the arm of CLARITY in which rats were exposed to bisphenol A (BPA) or vehicle daily beginning on gestational day 6 through postnatal day 21 (FIG. 1a)
- CONTINUOUS-DOSE — the arm of CLARITY in which rats were exposed to BPA, ethinyl estradiol or vehicle daily beginning on gestational day 6 through the rest of life (FIG. 1a)

Other terminology

- Gavage — the insertion of a tube down the oesophagus for administration of a test substance directly into the gastrointestinal tract
- LOAEL — the lowest observed adverse effect level, which is the lowest dose at which an adverse effect of a chemical can be observed
- NOAEL — the no observed adverse effect level, which is the highest dose at which no adverse effect of a chemical can be observed
- Tolerable daily intake — the 'safe' or allowable dose of a substance in humans

risk assessment. To do this, we examined data presented in the core study report[29] and the peer-reviewed academic studies published to date and, using these data, drew the conclusions presented here about the results of CLARITY. In our opinion, collaborative efforts provide the best means of improving existing approaches to risk assessment and devising new ones that will protect public health; however, we also assert that CLARITY did not live up to its full potential and that future collaborative work will benefit from a careful analysis of the challenges, flaws and findings of this project.

An uneasy alliance

To understand the complexities and challenges of the CLARITY endeavour, some familiarity with differences in the culture of research and divergent points of view of the scientists involved is needed. For example, FDA scientists rely on guideline studies to obtain data for regulatory purposes, but academic scientists use hypothesis-driven methods to understand whether and how biological processes are affected by chemical exposure. Although, ideally, a consensus between the approaches should be possible, differences in research culture made the CLARITY effort akin to expecting a group of folk and punk rock musicians to pick up their instruments and play together in harmony. As discussed below, some 'either-or' choices in study design could not be reconciled via compromise, leaving neither group entirely satisfied. For both groups, however, the driving force for the alliance was two important facts that are not in dispute: BPA is an EDC, and its widespread use in consumer and industrial products results in daily human exposure[7].

Overview of the study design

In designing the study, it was decided that rats would be housed and bred in the FDA's laboratory at the National Center for Toxicological Research (NCTR) and, as they were euthanized, tissues from animals would be coded and sent to academic researchers for blinded analyses. Animal model choice was, to some extent, dictated by the site of animal work, as previous use of the CD23/NctrBR Sprague-Dawley rat and the availability of a large colony of these rats at the NCTR made this a logical choice. Researchers also agreed that, as results were generated in academic laboratories, data would be submitted back to the NTP for storage before unblinding of treatment groups. Although agreement could be reached on these initial decisions, reaching consensus on other important features of the experimental design was more difficult.

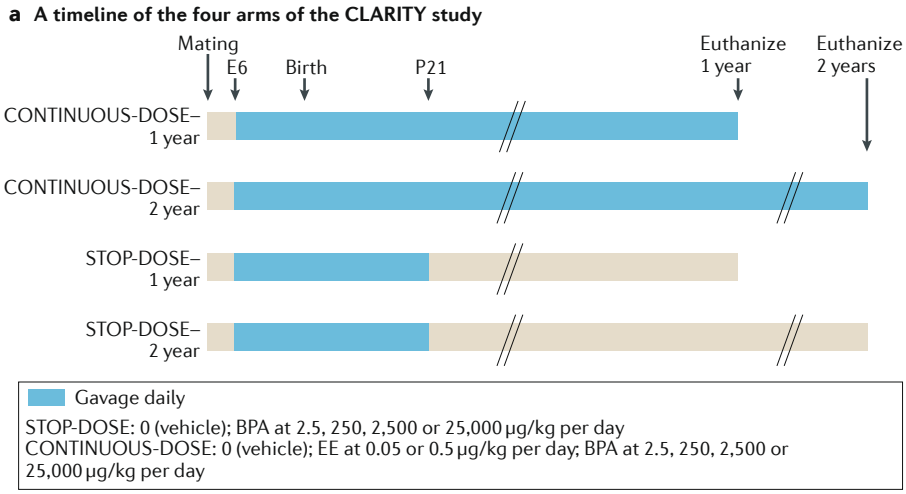
CLARITY was actually four intertwined studies (FIG. 1): two arms were CONTINUOUS-DOSE studies (BOX 1) and included animals exposed to vehicle; ethinyl estradiol (a positive control for oestrogen receptor agonists) at 0.05 or 0.5 μg/kg per day; or BPA at 2.5, 25, 250, 2,500 or 25,000 μg/kg per day from gestational day 6 throughout postnatal life (BOX 2). Half of these animals were necropsied at 1 year of age, and the other half at 2 years of age. The other two arms were STOP-DOSE studies (meaning exposure ceased at weaning, before puberty) and included animals exposed to vehicle or BPA at 2.5, 25, 250, 2,500 or 25,000 μg/kg per day from gestational day 6 through postnatal day 21. No ethinyl estradiol-treated animals were included in the STOP-DOSE experiments, leaving this arm of the study without positive controls. Again, half the animals were necropsied at 1 year of age and the other half at 2 years of age.

Animals from all four arms were used in the core study and by academic laboratories (FIG. 1a). A summary of select end points analysed is provided in FIG. 1b. Depending on the end points evaluated, some laboratories also requested and were provided with tissues collected at other ages. Sample sizes for each arm of the guideline studies were typically large (n = ~20–50 per sex), whereas sample sizes of tissues and/or animals provided for academic studies were smaller (for example, n = 4–12).

Reaching consensus on study design

In designing CLARITY, federal and extramurally funded researchers had to agree on methodology, which was not trivial. As even small environmental changes can affect phenotypic outcomes, all husbandry decisions, including but not limited to housing conditions (such as composition of the cage and water bottle, number of animals per cage, type of bedding, light cycle and feed), breeding conditions and weaning protocols were the subject of intense discussion. For BPA (and other EDC) research, the route of exposure is a particularly important consideration. The choice of gavage dosing was, from the NCTR point of view, practical — it allowed for precise delivery of chemicals, even to neonates[33]. Extensive research data, however, suggesting that gavage is a significant stressor[34], worried academic laboratories studying endocrine-sensitive end points.

Another difficult decision was the inclusion of an appropriate positive control, an issue that has previously been hotly debated, especially with respect to guideline studies[35,36]. Positive controls ensure that a response to a stimulus can be detected in a biological system. That is, if a BPA study



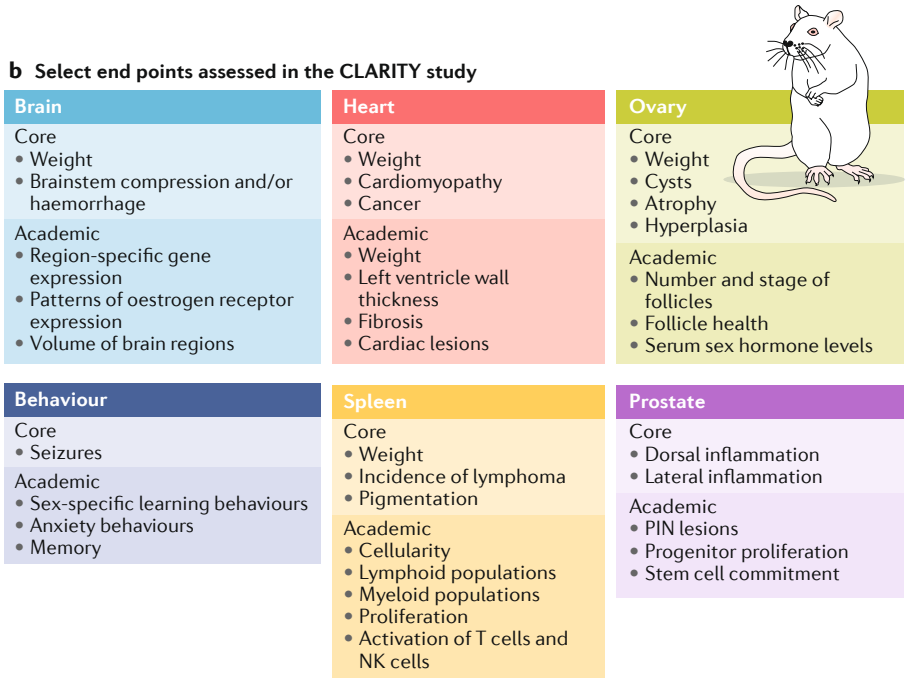


Fig. 1 | **The CLARITY study. a** | The study design of CLARITY, with the period of dosing indicated with blue shading. **b** | A subset of biological systems from CLARITY illustrates differences in the types of end point analysed in the core study conducted at the National Center for Toxicological Research (NCTR) and in academic laboratories. Following guideline protocols, the core study measured end points at necropsy (including organ weights, tumours and malformations). The academic investigators conducted further phenotyping of animals (for example, behavioural tests and response of the prostate to hormones) and other analyses involved in experimental hypothesis testing (such as number of stem cells and gene expression). BPA, bisphenol A; E6, embryonic (gestational) day 6; EE, ethinyl estradiol; P21, postnatal day 21; NK, natural killer; PIN, prostate intraepithelial neoplasia.

fails to detect effects, but effects are evident in positive controls, it can reasonably be concluded that the finding represents a 'true negative' (lack of an oestrogenic effect of BPA) rather than insensitivity of the test system (for example, due to the species, strain or end point evaluated or the presence of background contaminants). Indeed, the inclusion of low-dose positive controls is considered essential in experiments testing low doses of potential EDCs[37,38]. Arguments against inclusion of these controls in guideline studies have been the expense incurred and unnecessary repetition; furthermore, because many EDCs, including BPA, have more than one endocrine mode of action[39], the selection of appropriate positive controls is not straightforward.

Ethinyl estradiol, which is the active oestrogenic component found in oral contraceptives used by 100 million women worldwide[40], was selected as a positive control, at least for the CONTINUOUS-DOSE experiments in CLARITY, because it is a well-studied oestrogen receptor agonist and has been used as a positive control in previous studies. Indeed, despite its chameleon-like endocrine activities, the vast majority of BPA research has focused on its oestrogenic actions, making ethinyl estradiol a reasonable choice, at least for the evaluation of oestrogen-sensitive outcomes. In oral contraceptives, ethinyl estradiol is typically found in doses of 20–35 µg per day (equivalent to 0.29–0.5 µg/kg per day; BOX 2). Thus, the high ethinyl estradiol dose used in CLARITY is sufficient to block ovulation in women.

Confounding variables

A pilot study conducted in part to demonstrate the efficacy of the FDA–NTP–academic collaborative process (that is, a guideline 90-day subchronic study) uncovered two significant confounders that undermine the strength of any conclusions drawn on the basis of CLARITY data: environmental BPA contamination and animal stress.

During the design phase, researchers noted that a weakness of most prior studies, including guideline studies, was failure to evaluate internal doses of BPA associated with specific adverse effects[7]. Thus, a goal of CLARITY was to provide animals and tissues developed under conditions of analytical standardization often not included in academic research (for example, internal dosimetry and analytical quantification)[41]. Accordingly, the pilot 90-day study included serum measurement of both BPA and BPA metabolites in animals at postnatal days 4, 21 and 80 (REF.[41]). BPA that enters the body is metabolized by glucuronidases to BPA-glucuronide (the major metabolite) and by sulfatases to BPA-sulfate (a minor metabolite)[42]. Because only free BPA is found in consumer products, the detection of metabolites in serum demonstrates that BPA was in the animal long enough for these biological processes to occur.

In the 90-day pilot study, evaluation of two sets of controls (vehicle-only gavage and naive or untreated) revealed contamination with BPA. The presence of BPA metabolites in the control rats made it clear that contamination was not introduced during sample collection or processing but was the result of animal exposure[43]. The authors wrote that "the source of the unintended exposure leading to [BPA metabolites] in serum from vehicle and naive control groups was not ultimately identified" despite "rigorous evaluation of diet, water, cages,

Box 2 | **Putting doses in perspective**

CLARITY included rats exposed to bisphenol A (BPA) at 2.5, 25, 250, 2,500 and 25,000 μg/kg per day

- The lowest BPA dose in CLARITY (2.5 μg BPA/kg per day) is the most relevant to human intake. In Western countries, estimates of daily intake are 0.01–5 μg/kg per day for adults and 0.01–13 μg/kg per day for children; exposures are higher in Asian countries[1,63,64].
- BPA concentrations of 2.5 and 25 μg/kg per day are relevant to human safety. The tolerable daily intake calculated by US regulatory agencies is 50 μg/kg per day. Exposures below this level are assumed to be safe.
- The four lowest BPA doses (2.5, 25, 250 and 2,500 μg/kg per day) are relevant to toxicological safety. Prior guideline studies identified BPA 5,000 μg/kg per day as the dose at which no adverse effects occurred. Thus, no adverse effects should be observed at any of the four lowest BPA doses examined in CLARITY.

CLARITY included animals exposed to ethinyl estradiol at 0.05 and 0.5 μg/kg per day as positive oestrogenic controls

- The 0.5 μg/kg per day ethinyl estradiol dose is pharmacologically relevant to humans. Ethinyl estradiol is the active ingredient in many oral contraceptives. Typically, these pharmaceuticals are administered with doses of ethinyl estradiol ≤0.5 μg/kg per day.

bedding, vehicle and careful attention to dose certification and delivery"[43]. The authors proposed that faeces from a subset of animals administered a very high dose (BPA 300,000 μg/kg per day) could cause contamination of control cages, but low levels found on swabs of cages and the low volatility of BPA did not support this hypothesis[43].

Given that contamination was pervasive in the pilot study and could not reliably be eliminated, it is surprising that monitoring for contamination was not a major consideration in the CLARITY study. BPA metabolites were evaluated in only 46 vehicle-exposed controls at 1 year of age (CONTINUOUS-DOSE group, $n = 13$ per sex; STOP-DOSE group, $n = 10$ per sex), and detectable levels of BPA-glucuronide were found in 3 (7%; reported in[33], but not in the guideline study[31]). Although the levels of unconjugated BPA were not reported, it seems clear that contamination of animals with BPA was an uncontrolled variable in the CLARITY study and one that would act to diminish differences between control animals and animals receiving administered doses of BPA.

Data from control animals in the pilot study also provided evidence that animal stress might have diminished the effects of BPA. As noted above, academic researchers voiced concern during the design of CLARITY about the choice of gavage as an exposure method. In fact, the assessment of brains collected from the vehicle-gavage and naive and/or untreated animals in the 90-day pilot study provided evidence that gavage alone affected neuroendocrine development and diminished the effects of BPA exposure[44]. As with the contamination issue, these pilot study results were not translated into enhancements of the CLARITY study design. Despite evidence of a notable stress effect, all of the CLARITY animals were gavaged. Thus, in the absence of a naive (no gavage) control group, the effect of stress on study animals, and its potential interference with BPA exposure-induced effects, cannot be determined.

Stress and contamination are serious confounders, but, in the absence of appropriate controls, the effects of these variables on the results reported by CLARITY investigators are unclear. For example, we do not know whether background BPA contamination was sufficient to affect outcomes or change the dose or doses at which significant effects could be detected. In addition, it remains uncertain whether variability between studies with regards to the incidence of effects is evidence of differences in the sensitivity of specific end points or their sensitivity to stress. Finally, it is unknown whether diminished or absent effects in studies by individual CLARITY investigators who previously reported BPA effects are a reflection of differences in animal sensitivity (for example, with regard to strain or species), stress and/or contamination.

With these shortcomings in mind, we conducted an analysis of the CLARITY core data, publicly available online in the NTP final report[31], as well as the results of peer-reviewed academic studies that were published at the time of writing this Perspectives. By evaluating the data rather than relying on the interpretations of the authors, we have attempted to draw broader conclusions that can shed light on BPA's effects and collaborative study designs in general.

What does the core study tell us?

Guideline studies are typically not repeated. They are large, expensive studies evaluating multiple doses and end points, the latter 'validated' by international agencies, and reproducibility is assumed[9,45,46]. Our analysis of the guideline study presented in the NTP's February 2018 report[29] provides some evidence that the sensitivity of CLARITY might have been compromised, and our interpretation challenges assumptions about the reproducibility of guideline studies.

Prior guideline studies used to establish 'safe' doses for human exposure concluded that the NOAEL for BPA effects on adult animals, including rats (adult systemic toxicity), was 5,000 μg/kg per day; for reproductive and postnatal developmental end points, the NOAEL was determined to be 50,000 μg/kg per day[12]. These NOAEL doses were used to calculate a tolerable daily intake of 50 μg/kg per day by US regulatory agencies.

The first important point we would like to discuss regarding the core CLARITY guideline study is that it did reveal effects of BPA at the lowest doses examined. On the basis of the prior NOAEL doses, both the STOP-DOSE and CONTINUOUS-DOSE arms of CLARITY included at least one dose group at which effects on guideline end points were anticipated (for example, the highest dose). However, a summary of significant effects observed in the guideline studies (FIG. 2) shows that few outcomes were affected at these high doses of BPA. By contrast, several serious effects were observed at lower doses, including increases in the incidence of mammary adenocarcinoma (at 2.5 μg/kg per day in the STOP-DOSE group), inflammation of the dorsal and lateral lobes of the prostate (at 2.5 μg/kg per day in the CONTINUOUS-DOSE group), kidney nephropathy in females (at 2.5 μg/kg per day in the CONTINUOUS-DOSE group) and increased body weight in adult females (at 250 μg/kg per day in the CONTINUOUS-DOSE group). Despite these findings, the NTP and NCTR study authors concluded that BPA produced minimal effects that were distinguishable from background[31].

We do not believe there is a scientific basis for the FDA to dismiss adverse outcomes at low doses. For example, in the FDA guideline study, statistically significant effects of BPA were evident for several clinically relevant end points that were disregarded or referred to as 'sporadic' by the study authors. Statistically significant effects were dismissed if they were observed

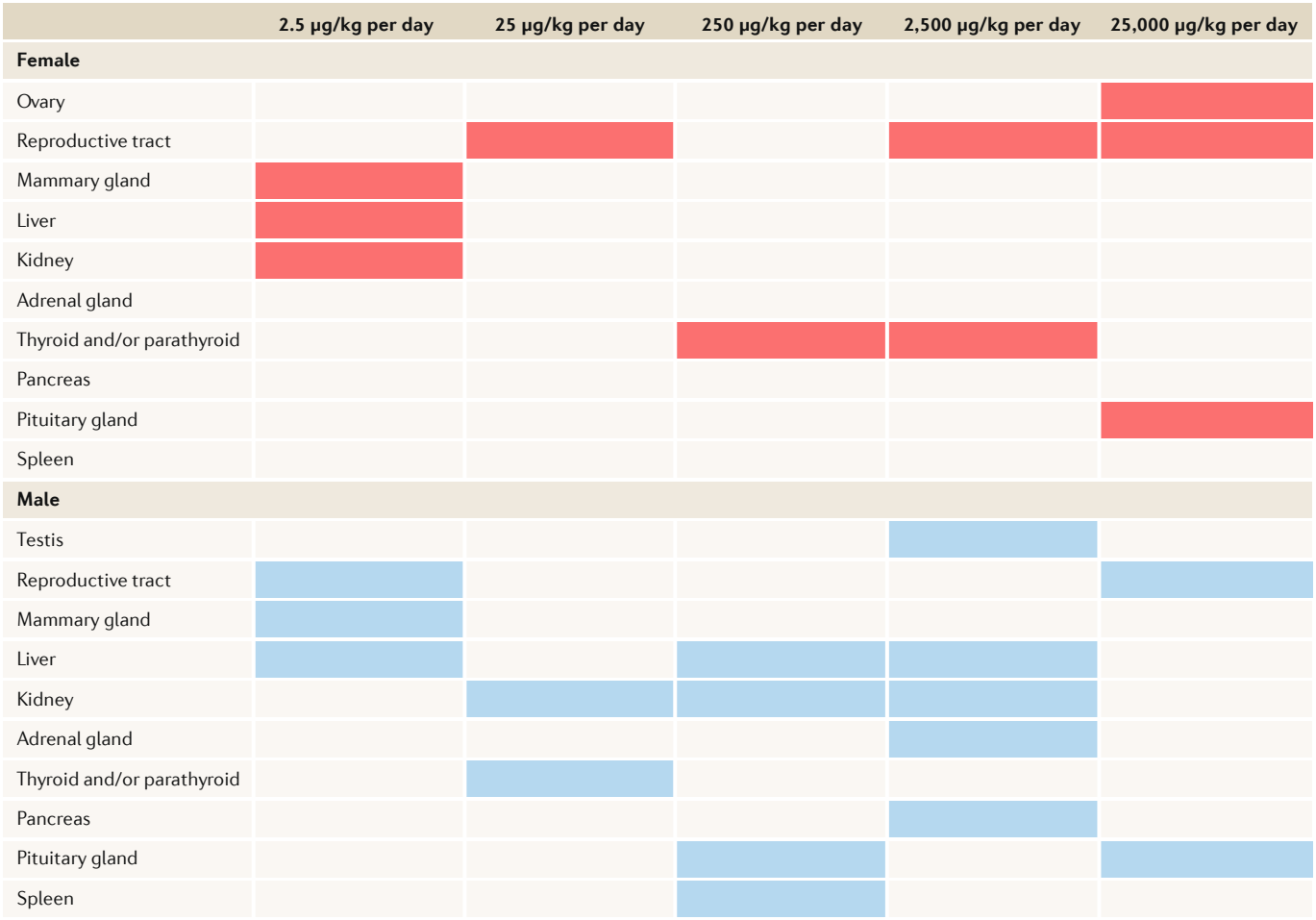


Fig. 2 | **Summary of CLARITY results for the five BPA doses: guideline study.** The shaded squares denote statistically significant effects; see Supplementary Table 3 for details. BPA, bisphenol A.

in only one of the study arms (STOP-DOSE versus CONTINUOUS-DOSE) or only at one dose[32]. These effects were described as 'not dose-responsive' and their biological relevance was called into question by investigators. The dismissal of effects on the basis of non-monotonicity, however, seems imprudent given the well-documented nonlinear effects of hormones and EDCs[47].

In their analysis, the FDA compared the most serious outcome — the incidence of mammary adenocarcinomas — not only with contemporaneous (vehicle-treated) negative controls but also to historical controls (that is, untreated controls from prior guideline studies conducted at the NCTR). The incidence of adenocarcinoma was significantly increased in the BPA 2.5 µg/kg per day STOP-DOSE group in comparison with contemporaneous controls ($P=0.016$; TABLE 1) but not historical controls from two prior guideline studies[48,49]. Importantly, these two prior studies differed in several important respects from the CLARITY study: the historical controls were not gavaged, suggesting a different background level of stress, and they were housed in polycarbonate cages, suggesting a different background level of BPA exposure. Thus, although it has been assumed that data from historical controls can be used to evaluate results in new studies, an important take-home message from CLARITY is that contemporaneous controls are essential to account for experimental drift. If contemporaneous controls are not used as the comparison group, the reproducibility of guideline studies is called into question.

The second point from the core study that we believe requires further discussion is that the CLARITY guideline data indicate that the rat model was insensitive to low doses of known oestrogens. As mentioned above, the high ethinyl estradiol dose used in CLARITY is sufficient to induce physiological changes in women. Thus, effects of the high dose of ethinyl estradiol (0.5 µg/kg per day) on some outcomes (such as onset of aberrant oestrous cycles and an increase in mammary gland adenocarcinoma) reported in the guideline study are not surprising (see TABLE 2 for summary). What is surprising is the lack of effects of ethinyl estradiol at a concentration of 0.5 µg/kg per day on other established oestrogen-sensitive outcomes, such as the timing of vaginal opening in females undergoing puberty, the weight of the testes and chronic inflammation in the prostate, among others.

A third and surprising point — and, in fact, a cause for concern — is the lack of reported effects on guideline end points in the animals treated with low-dose ethinyl estradiol (exposed to 0.05 µg/kg per day; TABLE 2), given that similar low-dose effects have been reported previously for non-guideline outcomes[50,51]. These results raise the spectre that the test system is not appropriately sensitive for the evaluation of low-dose effects of ethinyl estradiol or that the end points included in guideline studies are not sufficiently sensitive to detect low-dose effects that do exist. As ethinyl

Table 1 | **Outcomes for continuous-exposure ethinyl estradiol experiments**

Outcome	Subcategory	CLARITY		EE–NTP	
		Low-dose EE (0.05 µg/kg)	High-dose EE (0.5 µg/kg)	Low-dose EE (~0.1 µg/kg)	High-dose EE (~0.5 µg/kg)
Female					
Abnormal oestrous cycle	–	–	↑	–	–
Body weight	–	–	–	↓	↓
Mammary gland	Histopathology	–	↑	–	–
	Adenocarcinoma	–	↑	–	–
Pathological abnormalities	Kidney	↑	–	–	–
	Liver	↑	–	–	–
	Adrenal	–	↑	–	–
	Thyroid	↑	–	–	–
	Pituitary	–	↑	–	–
Male					
Body weight	–	–	–	↓	–
Mammary gland	Histopathology	–	–	↑	↑
	Adenocarcinoma	–	–	–	–
Pathological abnormalities	Kidney	–	↑	–	–
	Liver	↑	–	–	–
	Adrenal	↑	–	–	–
	Thyroid	↑	–	–	–
	Pituitary	–	↑	↑	↑

Arrows denote statistically significant changes and the direction of change relative to vehicle control. EE–NTP data are summarized from REF.[48]. See Supplementary Table 2 for end point details. BPA, bisphenol A; EE, ethinyl estradiol; NTP, National Toxicology Program.

estradiol has approximately 10,000-fold higher binding affinity for the oestrogen receptor than BPA[52], if binding affinity is predictive of biological activity, the lack of effects in the animals treated with ethinyl estradiol at 0.05 µg/kg per day suggests that it will be difficult to detect effects of BPA at doses ≤500 µg/kg per day. As discussed below, the inclusion of additional end points evaluated by academic researchers in the CLARITY study, including outcomes that are probably not mediated by the oestrogen receptor, provided a powerful and compelling approach to address these possibilities.

The fourth main point that we would like to discuss is that the CLARITY data challenge long-held assumptions about guideline studies. When a guideline study is being developed, it must be validated before it can be approved for use in regulatory toxicity testing. The validation process involves an assessment by multiple laboratories of 'known' chemicals (which are the positive controls) and coded 'unknowns' (which are the test chemicals) to evaluate the reliability and reproducibility of observed effects[9]. By using this approach, the outcomes that are measured in validated guideline studies have thus been considered reliable and reproducible, although they are rarely repeated and reproducibility has not been demonstrated[53].

We realized that the positive control (ethinyl estradiol treatment) groups in CLARITY provided an opportunity to test the long-held assumption of the reproducibility of data collected in guideline studies. Accordingly, we compared the ethinyl estradiol data from CLARITY with data from a previous guideline study on the toxicity and carcinogenesis of ethinyl estradiol that was also conducted by the NCTR and NTP[48] (referred to henceforth as the EE–NTP study; BOX 1).

The EE–NTP study investigators fed rats ethinyl estradiol in chow at concentrations of 2, 10 or 50 parts per billion (ppb), roughly equivalent to daily intakes of 0.1–0.2, 0.5–1 and 4–6 µg/kg per day, respectively. Thus, we assumed that the two lowest concentrations of ethinyl estradiol in the EE–NTP study (2 and 10 ppb) are the most similar to those evaluated in CLARITY. TABLE 1 provides a comparison of effects reported in 2-year-old rats from the EE–NTP study and from the CONTINUOUS-DOSE arm of CLARITY. We were surprised to find few similarities in affected outcomes between the two guideline studies. For example, a statistically significant increase in the incidence of mammary adenocarcinoma was observed in rats from the ethinyl estradiol 0.5 µg/kg per day group in CLARITY, but no increase was evident in the EE–NTP study, even after exposures approximately ten times higher[48] (TABLE 1).

Importantly, the treatment groups in the EE–NTP study differed in several important ways from the ethinyl estradiol treatment groups in CLARITY (TABLE 3). First, in the EE–NTP study, exposures to the parental generation started before mating. Second, rather than gavage, animals were exposed to ethinyl estradiol through food and, because ethinyl estradiol was incorporated into the chow, doses changed modestly during the course of the study on the basis of changes in consumption patterns. Third, rats in the EE–NTP study were housed in polycarbonate cages, suggesting that background exposure to BPA was likely, although it was not evaluated. Finally, only some outcomes, such as onset of abnormal oestrous cycles, histopathological abnormalities in organs such as the liver and pituitary and adenocarcinomas of the mammary gland, were evaluated in both studies. The differences in findings between the ethinyl estradiol groups in CLARITY and EE–NTP could be due to these methodological differences (TABLE 3), suggesting that consistent features such as housing (including avoiding low-level BPA leaching from polycarbonate cages) are essential to ensure the reproducibility of guideline studies. As background contamination of BPA and other EDCs is not typically controlled during guideline studies, this calls the reproducibility of guideline studies into question.

It is also plausible that the differences between the ethinyl estradiol results of CLARITY and EE–NTP reflect natural experimental drift over time. This experimental drift is a notable — although not widely acknowledged — feature of environmental research: even when the same strain of animal and identical protocols are used, changes in feed lots, animal husbandry and/or differences in reagents are inevitable, and any of these can affect experimental results.

Importantly, the mismatch in results between the ethinyl estradiol-treated rats in

Table 2 | **Summary of effects of ethinyl estradiol in CLARITY**

Outcome	Female		Male	
	Low dose EE (0.05 µg/kg)	High dose EE (0.5 µg/kg)	Low dose EE (0.05 µg/kg)	High dose EE (0.5 µg/kg)
Pup survival	↓	–	↑	–
Body weight	–	↑	–	–
Abnormal oestrous cycle	–	↑	NA	NA
Sperm abnormalities	NA	NA	–	–
Organ weight	–	↑ (5/12), ↓ (2/12)	–	–
Haematology and clinical chemistry	↑ (1/37)	↑ (1/39),↓ (3/39)	↑ (1/39), ↓ (1/39)	↑ (1/39)
Neoplastic lesions	–	↑ (1/3)	–	–
Non-neoplastic lesions	↑ (3/38)	↑ (18/38)	↑ (4/38)	↑ (2/38)

Numbers in parentheses denote the number of statistically significant changes out of the total end points analysed. Arrows denote the direction of change relative to vehicle control. See Supplementary Table 1 for detailed breakdown of end points and significant changes in CONTINUOUS-DOSE and STOP-DOSE group animals as well as the ages at which they were observed. NA, not applicable.

EE–NTP and the ethinyl estradiol-treated rats in CLARITY is disconcerting from the perspective of regulatory science: it suggests far less reproducibility between guideline studies than has been assumed[9]. In looking at these two studies, it is impossible to draw a definitive conclusion as to whether ethinyl estradiol induces mammary carcinoma in rats: the EE–NTP study concluded that there was "no evidence of carcinogenic activity" after continuous exposures to ethinyl estradiol, whereas the CLARITY study concluded that "there was a [statistically] significant increase in the incidence of mammary adenocarcinomas" in females exposed to ethinyl estradiol at 0.5 µg/kg per day. The failure to reproduce serious adverse outcomes (specifically, mammary adenocarcinoma) between the two guideline studies of ethinyl estradiol inevitably leads to the conclusion that the data used to calculate 'safe' levels of human exposure, which almost always come from guideline studies, are unreliable for this purpose; that is, if the positive control data cannot be reproduced, how reliable and reproducible are the guideline studies of BPA?

What do the extramural studies tell us?

By autumn 2018, at least 12 academic studies using CLARITY animals and/or tissues had been published[19–28,54,55] (FIG. 3). All but one of these studies report statistically significant BPA effects, although some are dismissed by the authors (see for example[20]).

The CLARITY academic studies support the conclusion that BPA induces effects at the lowest doses examined[56]. Of the published studies emerging from academic laboratories, among the most consistent effects are those of BPA on the brain, including alterations to the volume of sexually dimorphic structures and gene expression within specific brain regions. More modest effects were observed on neurobehaviours, ovarian follicle development, cardiac lesions and spleen myeloid populations. It is important to note that the lowest dose of BPA (2.5 µg/kg per day) elicited the greatest number of effects: alterations to gene expression and the size of certain brain regions; neurobehavioural disruptions; changes in the health, number and type of follicles in the prepubertal ovary; cardiovascular outcomes including heart weight, thickness of the left ventricle wall and incidence of cardiac lesions in females before puberty; alterations in myeloid populations in the spleen; and changes in the response to immune system challenges.

Taken together, these data suggest that low-dose BPA exposure induces subtle developmental changes that act to impair the endocrine, reproductive, neurobiological and immune system of adult rats. As discussed below, consistency among studies in finding effects in low-dose groups, coupled with low-dose findings from guideline data, reinforce the conclusion that current testing methods are inappropriate for EDCs.

We point out, however, that the remaining unpublished CLARITY academic studies might continue to provide insight. Publication of additional data on metabolic disease, obesity, brain, intestine, mammary cancer, uterine cancer, bladder, thyroid and periurethral gland function is anticipated from additional academic CLARITY studies[41]. There is always concern that failure to publish 'null data' (that is, no effect of BPA on particular end points) can lead to publication bias. In the case of CLARITY, this concern is allayed by the rules of participation; all participating academic laboratories signed an agreement to publish their data and agreed that any data not submitted in peer-reviewed publications would be published online by the NTP. Given the clear adverse outcomes described above from already published data, however, the remaining unpublished data would not negate the findings from CLARITY, even if no effects are evident.

The need for safety re-evaluation

Given the serious limitations and concerns raised in the previous section, and the vast investment of resources, CLARITY could be perceived as a costly failure. We argue, however, that the CLARITY approach demonstrates the power of combining guideline end points and in-depth analyses using state-of-the-art assessment of disease-relevant end points. Thus, several critically important conclusions can be drawn from this initiative. Lessons learned from CLARITY could and should be used to strengthen future collaborative efforts and drive a revolution in toxicity testing to better safeguard human and environmental health.

The CLARITY endeavour underscores the importance of validating the sensitivity and responsiveness of the experimental system and controlling for environmental contaminants. To this end, we argue that two design features are essential. First, appropriate positive controls must

Table 3 | **Study design for ethinyl estradiol experiments in CLARITY and EE–NTP studies**

Variable	CLARITY	EE–NTP
Animal strain	NCTR Sprague-Dawley	NCTR Sprague-Dawley
Responsible party for study	NCTR (FDA) laboratory	NCTR (FDA) laboratory
Route of exposure	Oral (gavage)	Oral (in chow)
Feed	5K96 (LabDiet)	5K96 (LabDiet)
Housing	Polysulfone cages	Polycarbonate cages
Bedding	Hardwood chip bedding	Hardwood chip bedding
Water	Tap, in glass bottles	Tap, bottles not specified

NCTR, National Center for Toxicological Research.

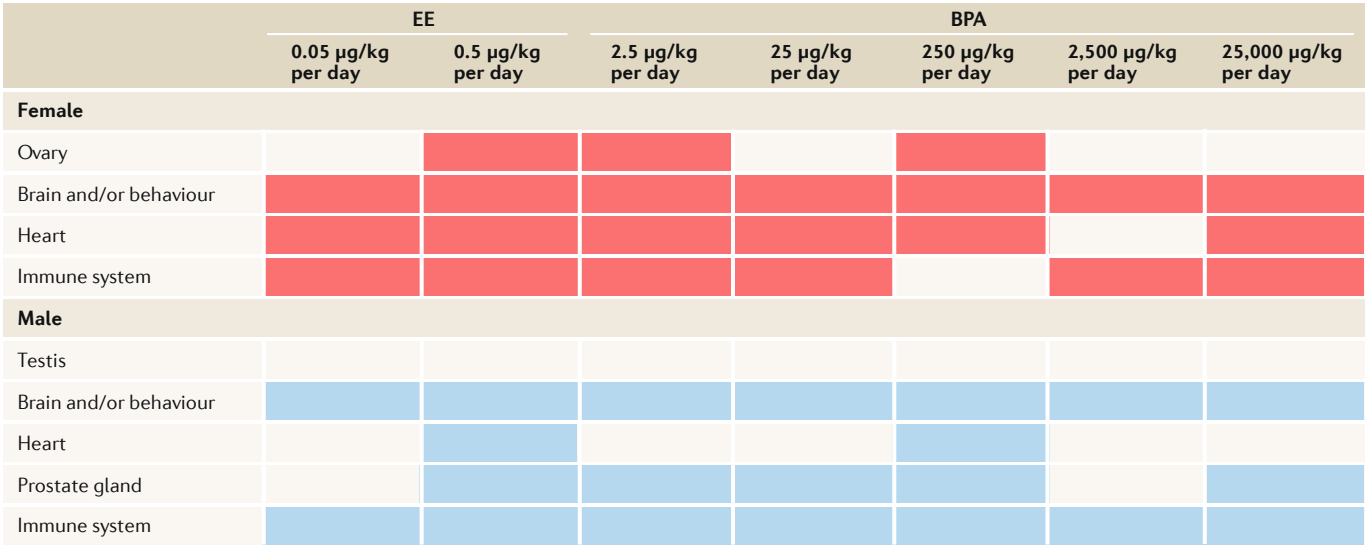


Fig. 3 | **Summary of CLARITY results for two ethinyl estradiol and five BPA treatment groups: academic studies.** The shaded squares denote statistically significant effects; see Supplementary Tables 4, 5 for details. BPA, bisphenol A; EE, ethinyl estradiol.

be run contemporaneously with test compounds, and second, biomonitoring must be conducted to detect inadvertent contamination. In addition, standard guideline approaches do not always provide an adequate assessment of health outcomes. The most important finding from CLARITY is that the combined data from all end points evaluated is greater than the sum of the findings from the guideline study alone. A conclusion from both the core and the academic studies is that BPA exposure induced statistically significant adverse effects at low doses (2.5 µg/kg per day) — far below the previously established NOAEL (5,000–50,000 µg/kg per day, depending upon the end point). Thus, in our opinion, an obvious conclusion from CLARITY is that the NOAEL for BPA needs to be revisited. Although the extent to which some outcomes can be considered adverse can be debated[57], the argument that cancerous lesions are adverse outcomes is irrefutable. Importantly, if regulatory agencies adopted a lowest observed adverse effects level (LOAEL) dose of 2.5 µg/kg per day for risk assessment, the 'safe' dose for human exposure (that is, the tolerable daily intake) would shift from 50 µg/kg per day to 0.0025 µg/kg per day (2.5 ng/kg per day) — which is 20,000 times lower than current standards.

Conclusions

The use of the sample pool of animals and tissues by multiple laboratories in CLARITY demonstrates the potential of collaborative research in improving our understanding of the actions and interactions of EDC contaminants and in the development of new approaches to assess the risks posed by this important class of environmental contaminants. What we need now is the completion of CLARITY through a full integration and analysis of results from both the core and academic studies. This approach is the only way that we can adequately capitalize upon this unique initiative and draw conclusions that were not previously possible on the basis of either guideline or academic studies alone[56].

Importantly, in this analysis, each organ needs to be evaluated so that the disease outcomes (including mammary cancer, obesity, erectile dysfunction, prostate cancer and diabetes mellitus) can be viewed in the context of the toxicity outcomes (such as organ weight) and mechanistic data (including gene expression and hormone receptor expression). As the purpose of CLARITY, similar to all hazard assessments, was to produce data that can be used to protect public health, it is imperative that experts be consulted for these final assessments and that the CLARITY results continue to be discussed in the context of the study strengths and weaknesses.

Although EDCs pose a notable challenge to toxicity testing, our own experience and expertise on effects of EDCs on the germ line, brain and/or behaviour and the mammary gland suggest that specific organs and end points could serve as canaries in the coalmine — that is, sensitive early warning systems of adverse effects for risk assessment. Importantly, evidence that some key findings in the germ line and brain are recapitulated in tractable and higher-throughput in vivo model systems such as worms[58,59] and zebrafish[60–62] offers promise of the development of sensitive, rapid screening approaches that will streamline testing.

In short, in our opinion, the importance of CLARITY lies in our ability to put the lessons learned from the endeavour into practice and build upon this important collaborative model. We anticipate that future studies, using a CLARITY-like approach to combine guideline end points with hypothesis-driven and mechanism-driven health outcomes, will prove useful in protecting public health from the most concerning environmental chemicals.

Laura N. Vandenberg[1], *Patricia A. Hunt*[2] *and Andrea C. Gore*[3]*

[1]*Department of Environmental Health Sciences, School of Public Health and Health Sciences, University of Massachusetts–Amherst, Amherst, MA, USA.*

[2]*Center for Reproductive Biology, School of Molecular Biosciences, Washington State University, Pullman, WA, USA.*

[3]*Division of Pharmacology and Toxicology, College of Pharmacy, The University of Texas at Austin, Austin, TX, USA.*

**e-mail:* andrea.gore@austin.utexas.edu

https://doi.org/10.1038/s41574-019-0173-y

1. Geens, T. et al. A review of dietary and non-dietary exposure to bisphenol-A. *Food Chem. Toxicol.* **50**, 3725–3740 (2012).
2. Rochester, J. R. Bisphenol A and human health: a review of the literature. *Reprod. Toxicol.* **42**, 132–155 (2013).
3. Vandenberg, L. N. et al. Low dose effects of bisphenol A: an integrated review of in vitro, laboratory animal and epidemiology studies. *Endocr. Disruptors* **1**, e25078 (2013).
4. Richter, C. et al. In vivo effects of bisphenol A in laboratory rodent studies. *Reprod. Toxicol.* **24**, 199–224 (2007).
5. Gore, A. C. et al. EDC-2: the Endocrine Society's second scientific statement on endocrine-disrupting chemicals. *Endocr. Rev.* **36**, E1–E150 (2015).

6. Vandenberg, L. N., Maffini, M. V., Sonnenschein, C., Rubin, B. S. & Soto, A. M. Bisphenol-A and the great divide: a review of controversies in the field of endocrine disruption. *Endocr. Rev.* **30**, 75–95 (2009).
7. Birnbaum, L. S. et al. Consortium-based science: The NIEHS's multipronged, collaborative approach to assessing the health effects of Bisphenol A. *Environ. Health Perspect.* **120**, 1640–1644 (2012).
8. Myers, J. P. et al. Why public health agencies cannot depend on good laboratory practices as a criterion for selecting data: the case of bisphenol A. *Environ. Health Perspect.* **117**, 309–315 (2009).
9. Tyl, R. W. Basic exploratory research versus guideline-compliant studies used for hazard evaluation and risk assessment: bisphenol A as a case study. *Environ. Health Perspect.* **117**, 1644–1651 (2009).
10. Stump, D. G. et al. Developmental neurotoxicity study of dietary bisphenol A in Sprague-Dawley rats. *Toxicol. Sci.* **115**, 167–182 (2010).
11. Tyl, R. W. et al. Two-generation reproductive toxicity study of dietary bisphenol A in CD-1 (Swiss) mice. *Toxicol. Sci.* **104**, 362–384 (2008).
12. Tyl, R. W. et al. Three-generation reproductive toxicity study of dietary bisphenol A in CD Sprague-Dawley rats. *Toxicol. Sci.* **68**, 121–146 (2002).
13. Peretz, J. et al. Bisphenol A and reproductive health: update of experimental and human evidence, 2007–2013. *Environ. Health Perspect.* **122**, 775–786 (2014).
14. Myers, J. P., Zoeller, R. T. & vom Saal, F. S. A clash of old and new scientific concepts in toxicity, with important implications for public health. *Environ. Health Perspect.* **117**, 1652–1655 (2009).
15. Schug, T. T. et al. Designing endocrine disruption out of the next generation of chemicals. *Green Chem.* **15**, 181–198 (2013).
16. Gore, A. C., Heindel, J. J. & Zoeller, R. T. Endocrine disruption for endocrinologists (and others). *Endocrinology* **147**, S1–S3 (2006).
17. Maffini, M. V. & Vandenberg, L. N. Closing the gap: improving additives safety evaluation to reflect human health concerns. *Environ. Risk Assess. Remediat.* **1**, 26–33 (2017).
18. Bergman, A. et al. Science and policy on endocrine disrupters must not be mixed: a reply to a "common sense" intervention by toxicology journal editors. *Environ. Health* **12**, 69 (2013).
19. Patel, S. et al. Bisphenol A exposure, ovarian follicle numbers, and female sex steroid hormone levels: results from a CLARITY-BPA study. *Endocrinology* **158**, 1727–1738 (2017).
20. Li, J. et al. CLARITY-BPA: effects of chronic bisphenol A exposure on the immune system. Part 2 — characterization of lymphoproliferative and immune effector responses by splenic leukocytes. *Toxicology* **396–397**, 54–67 (2018).
21. Li, J. et al. CLARITY-BPA: effects of chronic Bisphenol A exposure on the immune system. Part 1 — quantification of the relative number and proportion of leukocyte populations in the spleen and thymus. *Toxicology* **396–397**, 46–53 (2018).
22. Dere, E. et al. Effects of continuous bisphenol A exposure from early gestation on 90day old rat testes function and sperm molecular profiles: a CLARITY-BPA consortium study. *Toxicol. Appl. Pharmacol.* **347**, 1–9 (2018).
23. Arambula, S. E., Jima, D. & Patisaul, H. B. Prenatal bisphenol A (BPA) exposure alters the transcriptome of the neonate rat amygdala in a sex-specific manner: a CLARITY-BPA consortium study. *Neurotoxicology* **65**, 207–220 (2018).
24. Gear, R., Kendziorski, J. A. & Belcher, S. M. Effects of bisphenol A on incidence and severity of cardiac lesions in the NCTR-Sprague-Dawley rat: a CLARITY-BPA study. *Toxicol. Lett.* **275**, 123–135 (2017).
25. Arambula, S. E., Fuchs, J., Cao, J. & Patisaul, H. B. Effects of perinatal bisphenol A exposure on the volume of sexually-dimorphic nuclei of juvenile rats: a CLARITY-BPA consortium study. *Neurotoxicology* **63**, 33–42 (2017).
26. Johnson, S. A. et al. Effects of developmental exposure to bisphenol A on spatial navigational learning and memory in rats: a CLARITY-BPA study. *Horm. Behav.* **80**, 139–148 (2016).
27. Arambula, S. E., Belcher, S. M., Planchart, A., Turner, S. D. & Patisaul, H. B. Impact of low dose oral exposure to bisphenol A (BPA) on the neonatal rat hypothalamic and hippocampal transcriptome: a CLARITY-BPA consortium study. *Endocrinology* **157**, 3856–3872 (2016).
28. Rebuli, M. E. et al. Impact of low-dose oral exposure to bisphenol A (BPA) on juvenile and adult rat exploratory and anxiety behavior: a CLARITY-BPA consortium study. *Toxicol. Sci.* **148**, 341–354 (2015).
29. National Toxicology Program. Draft NTP research report on the CLARITY-BPA core study: a perinatal and chronic extended-dose-range study of bisphenol A in rats. *NIH.gov* https://ntp.niehs.nih.gov/ntp/about_ntp/rrprp/2018/april/rr09peerdraft.pdf (2018).
30. US Food & Drug Administration. Statement from Stephen Ostroff M.D., Deputy Commissioner for Foods andVeterinary Medicine, on National Toxicology Program draft report on bisphenol A. *FDA.gov* https://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/ucm598100.htm (2018).
31. National Toxicology Program. NTP research report on the CLARITY-BPA core study: a perinatal and chronic extended-dose-range study of bisphenol A in rats. https://ntp.niehs.nih.gov/ntp/results/pubs/rr/reports/rr09_508.pdf (2018).
32. Delclos, K. B. Bisphenol A: toxicology and pharmacokinetic data to inform on-going safety assessements. *FDA.gov* https://www.fda.gov/ScienceResearch/AboutScienceResearchatFDA/ucm621121.htm (2018).
33. Heindel, J. J. et al. NIEHS/FDA CLARITY-BPA research program update. *Reprod. Toxicol.* **58**, 33–44 (2015).
34. Vandenberg, L. N., Welshons, W. V., Vom Saal, F. S., Toutain, P. L. & Myers, J. P. Should oral gavage be abandoned in toxicity testing of endocrine disruptors? *Environ. Health* **13**, 46 (2014).
35. vom Saal, F. S. et al. The importance of appropriate controls, animal feed, and animal models in interpreting results from low-dose studies of bisphenol A. *Birth Defects Res. A* **73**, 140–145 (2005).
36. Vandenberg, L. N. et al. Regulatory decisions on endocrine disrupting chemicals should be based on the principles of endocrinology. *Reprod. Toxicol.* **38**, 1–15 (2013).
37. vom Saal, F. S. et al. Flawed experimental design reveals the need for guidelines requiring appropriate positive controls in endocrine disruption research. *Toxicol. Sci.* **115**, 612–613 (2010).
38. vom Saal, F. S. & Welshons, W. V. Large effects from small exposures. II. The importance of positive controls in low-dose research on bisphenol A. *Environ. Res.* **100**, 50–76 (2006).
39. Wetherill, Y. B. et al. In vitro molecular mechanisms of bisphenol A action. *Reprod. Toxicol.* **24**, 178–198 (2007).
40. Petitti, D. B. Clinical practice. Combination estrogen-progestin oral contraceptives. *N. Engl. J. Med.* **349**, 1443–1450 (2003).
41. Schug, T. T. et al. A new approach to synergize academic and guideline-compliant research: the CLARITY-BPA research program. *Reprod. Toxicol.* **40**, 35–40 (2013).
42. Vandenberg, L. N., Hunt, P. A., Myers, J. P. & Vom Saal, F. S. Human exposures to bisphenol A: mismatches between data and assumptions. *Rev. Environ. Health* **28**, 37–58 (2013).
43. Churchwell, M. I. et al. Comparison of life-stage-dependent internal dosimetry for bisphenol A, ethinyl estradiol, a reference estrogen, and endogenous estradiol to test an estrogenic mode of action in sprague dawley rats. *Toxicol. Sci.* **139**, 4–20 (2014).
44. Cao, J. et al. Prenatal bisphenol A exposure alters sex-specific estrogen receptor expression in the neonatal rat hypothalamus and amygdala. *Toxicol. Sci.* **133**, 157–173 (2013).
45. Tyl, R. W. In honor of the TeratologySociety's 50th anniversary: the role of Teratology Society members in the development and evolution of in vivo developmental toxicity test guidelines. *Birth Defects Res. C* **90**, 99–102 (2010).
46. Klimisch, H. J., Andreae, M. & Tillmann, U. A systematic approach for evaluating the quality of experimental toxicological and ecotoxicological data. *Regul. Toxicol. Pharmacol.* **25**, 1–5 (1997).
47. Vandenberg, L. N. et al. Hormones and endocrine-disrupting chemicals: low-dose effects and nonmonotonic dose responses. *Endocr. Rev.* **33**, 378–455 (2012).
48. National Toxicology Program. Toxicology and carcinogenesis study of ethinyl estradiol in Sprague-Dawley rats (feed study). *NIH.gov* https://ntp.niehs.nih.gov/go/tr548abs (2010).
49. National Toxicology Program. Toxicology and carcinogenesis studies of genistein in Sprague-Dawley rats (feed study). *NIH.gov* https://ntp.niehs.nih.gov/go/tr545abs (2008).
50. Catanese, M. C. & Vandenberg, L. N. Developmental estrogen exposures and disruptions to maternal behavior and brain: effects of ethinyl estradiol, a common positive control. *Horm. Behav.* **101**, 113–124 (2018).
51. Catanese, M. C. & Vandenberg, L. N. Low doses of 17 α-ethinyl estradiol alter the maternal brain and induce stereotypies in CD-1 mice exposed during pregnancy and lactation. *Reprod. Toxicol.* **73**, 20–29 (2017).
52. Shyu, C., Cavileer, T. D., Nagler, J. J. & Ytreberg, F. M. Computational estimation of rainbow trout estrogen receptor binding affinities for environmental estrogens. *Toxicol. Appl. Pharmacol.* **250**, 322–326 (2011).
53. Beronius, A., Hanberg, A., Zilliacus, J. & Ruden, C. Bridging the gap between academic research and regulatory health risk assessment of endocrine disrupting chemicals. *Curr. Opin. Pharmacol.* **19**, 99–104 (2014).
54. Cheong, A. et al. Gene expression and DNA methylation changes in the hypothalamus and hippocampus of adult rats developmentally exposed to bisphenol A or ethinyl estradiol: a CLARITY-BPA consortium study. *Epigenetics* **13**, 704–720 (2018).
55. Prins, G. S. et al. Evaluation of bisphenol A (BPA) exposures on prostate stem cell homeostasis and prostate cancer risk in the NCTR-Sprague-Dawley rat: an NIEHS/FDA CLARITY-BPA consortium study. *Environ. Health Perspect.* **126**, 117001 (2018).
56. Prins, G. S., Patisaul, H. B., Belcher, S. M. & Vandenberg, L. N. CLARITY-BPA academic laboratory studies identify consistent low-dose bisphenol A effects on multiple organ systems. *Basic Clin. Pharmacol. Toxicol.* https://doi.org/10.1111/bcpt.13125 (2018).
57. Woodruff, T. J. et al. Meeting report: moving upstream-evaluating adverse upstream end points for improved risk assessment and decision-making. *Environ. Health Perspect.* **116**, 1568–1575 (2008).
58. Allard, P. & Colaiacovo, M. P. Bisphenol A impairs the double-strand break repair machinery in the germline and causes chromosome abnormalities. *Proc. Natl Acad. Sci. USA* **107**, 20405–20410 (2010).
59. Lundby, Z., Camacho, J. & Allard, P. Fast functional germline and epigenetic assays in the nematode. *Caenorhabditis elegans. Methods Mol. Biol.* **1473**, 99–107 (2016).
60. Souder, J. P. & Gorelick, D. A. Assaying uptake of endocrine disruptor compounds in zebrafish embryos and larvae. *Comp. Biochem. Physiol. C* **208**, 105–113 (2018).
61. Moreman, J. et al. Acute toxicity, teratogenic, and estrogenic effects of bisphenol A and its alternative replacements bisphenol S, bisphenol F, and bisphenol AF in zebrafish embryo-larvae. *Environ. Sci. Technol.* **51**, 12796–12805 (2017).
62. Kinch, C. D., Ibhazehiebo, K., Jeong, J. H., Habibi, H. R. & Kurrasch, D. M. Low-dose exposure to bisphenol A and replacement bisphenol S induces precocious hypothalamic neurogenesis in embryonic zebrafish. *Proc. Natl Acad. Sci. USA* **112**, 1475–1480 (2015).
63. Corrales, J. et al. Global assessment of bisphenol A in the environment: review and analysis of its occurrence and bioaccumulation. *Dose Response* **13**, 1559325815598308 (2015).
64. Covaci, A. et al. Urinary BPA measurements in children and mothers from six European member states: overall results and determinants of exposure. *Environ. Res.* **141**, 77–85 (2015).

Acknowledgements
The authors gratefully acknowledge C. Lawson for assistance with figures and tables. The authors are supported by the NIH National Institute of Environmental Health Sciences grants K22ES025811 (L.N.V.), R56 ES013527 (P.A.H.) and RO1 ES023254 and R56 ES020662 (A.C.G.). The content of this manuscript is solely the responsibility of the authors and does not necessarily represent the official views of the NIH. The funders played no role in the writing of the report or in the decision to submit the article for publication.

Author contributions
The authors contributed equally to all aspects of the article.

Competing interests
L.N.V., P.A.H. and A.C.G. have received travel reimbursement from universities, governments, non-governmental organizations and industry to speak about endocrine-disrupting chemicals.

Publisher's note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Supplementary information
Supplementary information is available for this paper at https://doi.org/10.1038/s41574-019-0173-y.