Design, execution and interpretation flaws render invalid the FDA’s conclusion of bisphenol A safety based on core guideline data from CLARITY-BPA

Frederick S. vom Saal, Ph.D.
Curators’ Distinguished Professor
Division of Biological Sciences
University of Missouri-Columbia
Columbia MO 65211 USA

In 2012, three arms of the U.S. government launched an ambitious, ~$30 million program to resolve conflicting information available to chemical regulators around the world on the safety of bisphenol A (BPA), a widely used endocrine disrupting compound found in many consumer and industrial products. While standardized assays used in traditional ‘guideline’ studies showed few effects of BPA, especially at the low doses typical of human exposure, over a thousand studies by university-based epidemiologists and experimental scientists have revealed a wide-range of adverse effects, including at low doses of BPA.

Crucially, the ‘guideline’ studies use assays that had been developed during the 20th century (some as early as the 1930s), whereas the academic studies used methods largely stemming from NIH funded research over the past 20 years. The former are much less sensitive at detecting endocrine disrupting activity of chemicals such as BPA than the latter. But regulatory agencies are accustomed to working with guideline studies, and regularly reject research findings from experiments that use 21st century technology, even if they have been extensively replicated and published in the peer-reviewed scientific literature.

The program, “Consortium Linking Academic and Regulatory Insights on Toxicity of BPA” (CLARITY-BPA), was designed to compare in a single study these two approaches and resolve their differences. Because these disagreements emerge commonly in the
assessment of chemical safety, not just with BPA but also many other chemicals, the CLARITY approach, as initially envisioned, appeared to have great promise and value.

In the experiment, the US Food and Drug Administration (FDA) bred, raised and treated all animals, and executed standard guideline tests using Good Laboratory Practice (GLP) protocols; the FDA implemented this portion of the study without accepting input from the academic collaborators, which involved ignoring input from academic collaborators who raised serious concerns regarding experimental design issues (discussed below). The FDA distributed treated and control tissues to university research teams for analysis using advanced methodologies with which each research team had demonstrated expertise, including having published peer-reviewed findings on BPA and competing in a standard NIH grant review to participate in the project. The academic researchers were blind to sample identification until all data were loaded and locked in a federal data base, eliminating the possibility of bias in reporting results. The results of the two prongs of work are now being compared. Two key questions are: did the guideline studies reveal low-dose effects? Were the university experts able to replicate their own studies with animals raised and treated in the FDA facility under the controversial FDA protocol?

Not all the university studies are yet peer reviewed and published. Neither are the FDA’s findings, but the FDA in conjunction with the National Toxicology Program (NTP) released to the public the “core” guideline study results, as yet not peer-reviewed, in late February 2018. The release of these data was accompanied by a statement by the FDA’s deputy commissioner concluding that “the study found minimal effects, … and currently authorized uses of BPA continue to be safe for consumers” ¹.
Unfortunately, the released data and the university studies published to date reveal more about the role of false assumptions, flawed design and poor execution than they do about BPA safety. The collection of errors is extensive. Some of the problems with the FDA’s design, execution and interpretation of results in this experiment are listed here.

**Design Flaws**

1. The FDA was warned by academic researchers long before the experiment launched that the rat strain (CD-SD), which the FDA has used for decades, was relatively insensitive to estrogenic chemicals, such as BPA or the positive control estrogenic drug ethinylestradiol. Standard procedure in toxicological experiments is to choose the most sensitive animal model to protect the most sensitive humans. Choosing an insensitive model violated a core principle of toxicology.

2. A preliminary study conducted by the FDA and one of the academic collaborators (using CLARITY-BPA procedures) showed that the FDA’s stressful method of administering BPA, daily oral gavage, caused significant adverse effects in the brain of gavaged controls relative to non-gavaged controls. A common finding is that low dose BPA exposure during development eliminates sex differences in the brain and behavior, which is not unexpected since the background hormonal environment disrupted by BPA exposure is different in males and females. However, in the preliminary FDA-collaborator study, the physiological stress due to the gavage procedure eliminated the normal brain differences between males and females that were found in non-gavaged controls. Think about this. Every day in the life of these animals they were immobilized (restrained) and a tube was thrust down their throat to deliver the dose/placebo. The stress response this triggered could obscure aspects of the
endocrine disrupting effects that low doses of BPA might cause; how could BPA eliminate sex differences that had already been eliminated by the daily stress caused by gavage? But in the actual CLARITY study no un-manipulated (not restrained and not gavaged) controls were examined; there was thus no actual un-manipulated negative control group in CLARITY-BPA; all rats experienced daily restraint and gavage stress throughout their lives.

**Execution Flaws**

1. The FDA acknowledged an unknown source of contamination with BPA of negative control CD-SD rats in their preliminary CLARITY study 6, but the FDA deemed the contamination of controls to be unimportant. It is thus clear that procedures in the preliminary study caused contamination of controls, yet, in spite of harsh criticism 7, the full-scale CLARITY-BPA experiment went forward without testing for the possibility of contamination.

2. The academic researchers provided data to the FDA specifying how many animals their experiments would require to replicate earlier studies, based on studies they had conducted regarding how much variability was in the control and treatment groups, and the magnitude of difference they had previously observed between controls and BPA-exposed animals. This is called a power analysis, which is required for funding by NIH. However, the FDA failed to provide the number of animals needed in a number of cases, and for unknown reasons the variability among the provided samples was much higher than expected; two-day old rat pups provided to FvS for analysis of urogenital system morphology ranged in weight from 3.7-9.6 g, a >250% range in body weights, and the “randomly selected” subset of restraint / gavaged vehicle controls had significantly lower body weights relative to the core study vehicle control data. The high variability reduced the likelihood of finding
significant results and thus increased the likelihood of “false negative” findings that would be used to support the conclusion that BPA is harmless.

**Flaws in Interpretation of Results**

In spite of these flaws in the design and execution of the CLARITY-BPA experiment by FDA personnel, there were, in fact, a number of statistically significant findings. For example, despite the FDA’s statement of “minimal concern” with their findings, the newly released FDA core study data show significant increases in a number of adverse effects at low doses, acknowledged to be relevant to human BPA exposure. Specifically, the FDA ignored findings concerning mammary adenomas / adenocarcinoma at low doses, uterine stromal polyps, trends for uterine squamous metaplasia, depletion of corpora lutea, plus increases in prostatic inflammatory disease. These findings cannot be ignored.

The basis for this disconnect between the ‘core guideline’ study results and the FDA’s statement of ‘minimal concern’ involves an ongoing conflict between regulatory toxicologists and endocrinologists. The FDA toxicologists believe that since these adverse effects occurred at very low but not high doses of BPA, and thus did not show the classical monotonic dose-response relationship they expect for all toxic chemicals, these statistically significant findings could be ignored. However, low-dose stimulation and high-dose inhibition (i.e., non-monotonicity) of specific responses is a general characteristic of hormones, hormonal drugs such as Tamoxifen, and hormonally active chemicals such as BPA.

2. Rejecting over a century of progress in statistics, the FDA also assumes that if an outcome occurs in historical and / or true negative controls (which were not actually included
in the CLARITY-BPA study), such as mammary tumors even at a low incidence, then a statistically significant increase in the tumors at low doses can be ignored because they are within the range of possible outcomes in negative controls. The question is why would the FDA bother conducting statistical analyses if they are going to ignore the results?

Conclusions

Not surprising, given the list above, is that some prior findings from studies from academic collaborators that had been conducted with appropriate animal models, appropriated controls and had been executed by scientists with a high level of expertise, were not replicated. Not only were thousands of animals and tens of millions of public money not used optimally in CLARITY-BPA, the flaws identified here undermined the ability of CLARITY to resolve the important issue concerning the incorporation of 21st century science into the outdated ‘guideline’ approach used to regulate chemicals, which was what motivated the program.

A critical issue is that the FDA and other regulatory agencies have made decisions about chemical safety based only on studies that were conducted using ‘guideline’ assays conducted under GLP protocols, which require specific records to be kept due to evidence that industry-funded laboratories had been reporting data from experiments that had never actually been conducted. The flaws in the FDA portion of the CLARITY-BPA study, conducted with GLP record keeping protocols, demonstrate how GLP does not ensure that the design of the study was appropriate, that the study was executed properly, or that the interpretation of the results is unbiased 10.
Disagreements based on design, execution and interpretation of results from guideline and university studies, like those with CLARITY-BPA, are pervasive. If done correctly, the CLARITY model could be very useful for resolving these differences and could impact the approach regulatory agencies use to assess the hazards posed by all chemicals, not just BPA. In this regard, in retrospect it is unfortunate that BPA was chosen as the “test” chemical in the first attempt at resolving these conflicts, because BPA is one of the highest volume chemicals that is predicted to generate approximately 20-billion US dollars in revenue by 2020\textsuperscript{11}. The economic importance of BPA cannot be ignored when interpreting the FDA’s lack of concern with the issues raised here, which became obvious well before the release of the core data due to the aggressive insertion into the CLARITY-BPA decision-making process by risk assessors and managers from the FDA’s Center for Food Safety and Applied Nutrition (CFSAN), whose interest from our first interaction with them appeared to be defending the prior FDA decisions of BPA safety. In contrast, academic collaborators had initially assumed the collaboration would be between them and NCTR scientists.

The flaws in CLARITY-BPA lead us to conclude that any future attempt to use the CLARITY model should not involve any industry-associated corporation that has a financial interest in the outcome or any regulatory agency such as the FDA that has made repeated declarations over decades that the test chemical is safe, which the FDA has done regarding BPA. Reversing prior public declarations of safety for a regulatory agency is clearly problematic, as exemplified by the FDA’s misleading characterization of the CLARITY-BPA findings. For comparable efforts to be undertaken in the future, the flaws that eroded CLARITY’s value need to be understood and rectified. Importantly, these flaws render invalid the FDA’s assurance based on CLARITY data that BPA is harmless.
Literature Cited