Request for External Scientific Peer Review of the Scientific Basis of Microplastics Definition, Analytical Method, Monitoring & Reporting Order, and Health Effects Guidance Language

By Alan E Hubbard, PhD
Professor of Biostatistics
School of Public Health
Berkeley, CA
March 16, 2022

Introduction

Below I review the scientific basis for the draft *Microplastics in Drinking Water Policy Handbook*, adopted definition of `microplastics in drinking water`, proposed analytical methods to be used during required monitoring, and proposed health effects guidance language. As I noted in the response to the request, "I feel confident, given my expertise, experience and training, I can evaluate the quality of the statistical evidence for conclusion #4[1] based upon the design and analysis of studies considered in the evidence." The specific conclusion refers to the health-based guidance language being appropriate with respect to hazard knowledge and gaps. The recommendation by the SCCWRP was for the State Water Resources Board to "not adopt a notification level or other health-based level that would require water systems to inform consumers of contamination outside of their annual consumer confidence report or perform additional actions." Instead, the expert workgroup developed a brief health-based guidance language statement that is to be recommended for use by public utilities in informing consumers regarding findings of microplastics in drinking water in the annual consumer confidence reports, as included in the draft handbook.[2] The language of the focus of my review within the draft handbook is:

- "4.1. A principal research finding relevant to monitoring is that microplastics smaller than 10 μm in length have an increased likelihood of causing adverse health effects in mammals and should be prioritized for monitoring when possible. There is insufficient evidence at the time of writing this Policy to issue a notification level or other numerical guidance for microplastics."

- "4.1.1. Studies of rodents exposed to some types of microplastics through drinking water indicate potentially adverse effects, including on the reproductive system. However, more research is needed to understand potential human health implications and at what concentrations adverse effects may occur. Therefore, California is monitoring microplastics in drinking water to understand its occurrence and is supporting ongoing research."

Also relevant is the language in the document "Peer Review Request for Microplastics in Drinking Water_signed" which specified the goal of my review (Conclusion #4 in Appendix 2):

---

[1] See *Peer Review Request for Microplastics in Drinking Water_signed* document, page 9-10.
[2] Proposed rule-policy handbook_11-10-2021 document, section 4.1.

- "Peer reviewers should review the proposed health-based guidance language for its scientific basis and potential impacts to health and wellbeing (intentional and unintentional), including the validity of the underlying review of the science (Coffin *et al. Submitted*)."

The main issue for my review is whether the state of current research of the health impacts of exposure to microplastics in drinking water warrants a lack of numerical guidance or notification level for microplastic concentrations in drinking water and whether the language regarding health and wellbeing is consistent with the science.

Note, that I have no expertise in the general exposure science of microplastics (MP), including meaningful definitions, measurement, chemistry, and biologic response of MP. I am an expert in general biostatistics and more specifically their application to studies of human health, particularly population-based studies of health (epidemiology).

Evidence examined regarding health effects of human microplastic exposure in drinking water
The main document supporting the proposed approach comes from a submitted manuscript thoroughly reviewing the evidence and existing literature regarding microplastics in drinking water and health (Coffin et al., 2022), with additional information from a World Health Organization report (2019) , and an existing study on human health effects of occupational exposure to MP (Zarus et al., 2021). I reviewed several other publications provided to me at the review request.

I think a quote from Rahman et al., 2021 is an accurate summary of the state of scientific knowledge about the impacts of microplastics on human health -

"The ubiquitous presence of MPs in food products, water and air has led to their inevitable exposure to humans. However, the routes of exposure and the implications for human health have not yet explicitly been documented in the literature. Therefore, a scoping review of existing evidence and synthesis of the current knowledge on potential routes and effects of MPs exposure on human health and the mechanism of toxicity can serve as a foundation for future research."

I have found in the materials provided (as well as independent literature searches) no significant population epidemiological studies on direct human health impacts of MP's in drinking water. Though there are studies that measure the uptake and absorption of MP's in humans, there is little to no direct linking of the drinking water exposure of MP's and human health impacts. Though there are studies of occupational exposure of MP in air, most of the information on the health impacts in humans must be extrapolated from the limited exposure studies in people, and animal studies of controlled MP ingestion. Zarus et al., 2021 has a list of the relevant literature available for human exposure studies and experimental animal data, though only a subset of them are directly relevant for exposure through drinking water (e.g., exposures of humans through ingestion/drinking water and animal experiments with ingested MP's).

The most important evidence for the resulting recommendations comes from a submitted paper (Coffin et al., 2022), which conducted a screening process of the existing studies starting with 41 in vitro and 31 in vivo studies. Studies quality and relevance were determined largely based on reporting of dose-response relationships. They also made the sensible choice to only use oral exposure in vivo studies, which is based upon the lack of existing methods for extrapolating between in vitro and in vivo systems. They came to the conclusion that it was not possible to extrapolate a human–health-based threshold value for microplastic, largely due the relative quality and reliability of current data. Additionally, there exists no reliable methodology for extrapolating data from studies using monodisperse plastic particles, such as polystyrene spheres, to an environmentally relevant exposure of microplastics in drinking water.

The two-tiered process used to select studies for use in calculation of screening levels was described in great detail and appeared rigorous (I note that some of the criteria used to select require expertise outside of statistics). Screening levels were estimated for several physiological endpoints for 7 out of the 12 studies that passed the tier 2 screening, because 5 of the studies did not report adverse effects on both male and female reproductive systems. Then a standard of care methodology was used to estimate screening levels from the dose-response data reported for each outcome/study.

Methodology
BMD
The part of the report that is most relevant to my expertise was the method of translating the dose-response information available into a number relevant to human health that could be compared across different data. Specifically, the information provided from the expert panel for the studies deemed appropriate were used to estimate "points of departure" (e.g., benchmark doses [BMDs]). BMDs are doses (or concentrations) that elicit a predetermined change in response of an adverse effect based on a modelled dose-response curve. In short, a dose response curve is a function defined by, for outcome Y and exposure X, $E(Y|X=x)=m(x)$, or the conditional mean function of the physiological outcome given a particular level of the exposure, X. The BMD of endpoints, along with their respective upper and lower 95% confidence intervals were estimated based on modelled output using existing (US EPA) benchmark dose software (US EPA, 2012); these were compared and interpreted using the RIVM PROAST tool (Slob, 2018).

BMR
As stated in US EPA, 2012 the calculation of a BMD is directly determined by the selection of the benchmark response (BMR), the biologic response relative to controls that would be used to assess the magnitude of exposure associated with potential health risk. Selecting BMRs involves making judgments about the biological characteristics of the dataset about which the resulting BMD (and the reported

confidence limits) will be used. The US EPA (2012) report recommends, for continuous data, defining a BMR based on the level of change in the endpoint at which the effect is considered to become biologically significant (as determined by expert judgment or relevant guidance documents). They recommend, in the absence of any other idea of what level of response to consider adverse, a change in the mean equal to one control standard deviation (SD) from the control mean can be used; if warranted by statistical and biological considerations, a lower or higher increment of the control SD might be used. The authors calculations of screening levels used this last definition (a change in the predicted outcome via the fitted dose-response model if MP concentration increased by 1 sample SD (as measured in the controls) of the predicted outcome based upon the model or $BRM = m^{-1}(m(0)+1SD)$, or the dose at which the predicted outcome (by the estimated dose-response model) is equal to an increase of 1 SD (controls) above the estimated mean value of the outcome when exposure is 0 (or $m(0)$). Given my lack of expertise in choosing these values, I am assuming in review of the methodology, that the BMR have been chosen conservatively (that is chosen in a way that tends to favor lower screening levels).

The estimated BDM (benchmark dose) is thus a simple function of the estimated dose response curve, f(x) and the BMR, specifically BMD= m^{-1}(BMR), or the value of x such that the m(x)=BMR.   Thus, the choice of the BMR is critical to meaningful inferences about the BMD. Finally, the BMD and confidence limits are fed then into the following equation to derive potential screening levels:

Equation 1 $$Screening\ level\ (mg/L) = \frac{RfD\ (mg/kg-day)\ x\ RSC\ (unitless\ fraction)}{DWI\ (L/kg-day)}$$

where RfD (reference dose) is a function of the BMDL (lower confidence bound of BMD when available by estimation using the dose-response data) and DWI (drinking water intake).

Overall Algorithm for determining the Rfd for the screening level
 The estimate of the screening level for each study/outcome required several steps: 1) fitting a suite of parametric dose-response curves, 2) choosing the best fitting model for the dose-response using fit statistics (Akaike information criterion or AIC), and 3) using the chosen model to calculate the estimated BMD using the transformation discussed above, and 4) conducting a parametric bootstrap to estimate the sampling distribution of the estimated BMD by randomly generating data from an equivalent design based upon the model chosen and the study generating the data, 5) using this bootstrap distribution to derive a 95% confidence interval (CI) for the BMD, 6) conservatively selecting the BMD as the lower 95% CI limit of this confidence interval (BMDL) to feed into the equation 1 above. If the data in the study were not sufficient to compute the BMDL using the algorithm, then a conservative approach was used directly from observed data points in the study. Specifically, the lowest observable adverse effect level (LOAEL)

divided by 10 was used or the non-observable adverse effect level (NOAEL) depending on availability.

This methodology was applied one study/endpoint at a time to derive potential screening levels for MP. The final chosen screening level for MP was the minimum screening level for the combination of study and endponts in table 3 in Coffin et al., 2022. Though a sensitivity analysis was done to incorporate the uncertainty in RSC and DWI above, the screening level was based upon best estimates of these quantities (not the most extreme resulting in the lowest screening level) and without a correction factor for other sources of uncertainty (e.g., conservative reduction of the Rfd based upon inter-species variability, intra-species variability, and database deficiencies, which would divide the BMDL by 300 as stated in Coffin et al., 2022).

The specific study that provided data for the minimum BMDL was from Hou et al., 2021, a study of biological outcomes of rats (n=32) to exposure to 3 concentrations of 0.5 µm polystyrene MPs (and unexposed controls). I am assuming the authors of the Coffin paper were able to access the raw data from the Hou study, but they could also fit it with the mean outcome for the four different exposure groups and reported standard errors. The fitting of the dose response curve and the calculation of the lower 95% confidence bound are pertinent to my expertise in general statistical inference in finite samples. Since the results of the repeated application to the estimation/inference algorithm used by the authors is one piece of the evidence provided for the screening level recommendation, I will comment in more detail below on the validity of the approach.

Comments on the statistical methodology used to estimate the BMDL

The language provided in the Handbook[3] and quoted above is consistent with the scarcity of the evidence provided within and referenced by the Coffin, et al., 2021 report on the impact of human health of exposure to MP's. The Coffin report made a good faith effort to attempt to translate the few relevant toxicological animal studies available into an evidence-based screening level by using the limited dose-response information available and propagating the uncertainty of the estimation of these curves estimates in the calculation of the RfD, and by choosing the lowest estimated BMDL for any endpoint available across the available studies/endpoints. They also acknowledge that the MP exposures used in these studies (e.g., the polystyrene MP's in Hou et al., 2021) is a nonoptimal proxy for the complex chemistry of MP's likely in environmental human exposures. There is thus, a suite of non-statistical issues to consider in making the screening policy that must rely on expert judgement for translating the limited data into policy.

Given the inherent limitations of the data, the approach to estimate the BMD using the described methodology (US EPA, 2012) was proper. The general approach, so-called ensemble learning, is particularly well-suited to situations where there is no a priori reason to limit the estimation to a single dose-response model. Though the chosen model for each study/endpoint was based upon optimizing the AIC, a similar approach

---

[3] Proposed rule-policy handbook_11-10-2021 document, section 4.1

using cross-validated error to choose the best model has shown to be asymptotically equivalent to the so-called Oracle Selector, that is, the estimated model one would choose among candidates based upon knowing the true dose-response curve (van der Laan et al., 2007). Thus, one can make an argument that such a method, if combined with a large number of candidate dose-response curves, should asymptotically choose the best (among the list of models) for a particular endpoint. However, there are considerable limitations in interpreting the resulting estimated curve and functions of it, such as the estimated BMD. The procedure is fitting non-linear curves generally with very few design points (that is, very few doses per study used to expose the animals) and a limited number of animals per dose. Optimally, one would have a very large number of doses as well as a large number of rats exposed at each dose, conditions typically impractical for the type of studies on which (Coffin et al., 2022) relied, so that the asymptotic results would be relevant to the actual analyses. In this case properly characterizing the finite sample inference (e.g., getting consistent estimates of the BMD as defined relative to the BMR with confidence intervals with proper coverage) with such small data sets and so few design points is near impossible. In addition, because of the small sample sizes (small number of rats and doses), reliance of the inference on the central limit theorem is dubious and so the derived inference is based upon strong normality assumptions. Departures from normality in the true data-generating distribution can lead to bias in the CI's (improper coverage probabilities, possibly anti-conservative). Even though the BMD relies on estimating the dose response curve and the available data make estimation and inference regarding the dose-response curves and functions of it in a realistic statistical model (unknown specific dose-response model) problematic, there is no alternative procedure that would have guarantees of optimal performance. Thus, though there are robustness issues to the methodology used to provide estimates and inference, the approach is reasonable and appropriate given the limitations of the information available. I note that Coffin et al., 2022 are transparent about the limitations of the existing data available for MP ingestion exposure.

However, given the inherent limitations of making robust inferences about the impact on human health from the available relevant studies, the approach is much preferable to one that would use an arbitrary low dimensional dose-response model (model defined by very few parameters, e.g., exponential with two parameters) that could be fit to every study, versus only a more flexible model (exponential with 4 parameters) that might reduce bias, but would result in an unacceptable increase in variance. Thus, the method used does an appropriate bias-variance trade-off to try to glean as much information as possible, without overfitting, the dose-response data.

The only choice for deriving inference for the BMD that would not rely on strong assumptions would be to derive so-called bounds based on inequalities, such as Bernstein's inequality (Rosenblum and Van Der Laan, 2009). However, such approaches, in order to guarantee a minimum of the advertised coverage (95%) are highly conservative and in this situation would most likely result in lower bounds of 0 for most if not all of the mammalian studies used. Thus, given the nature of the data and

the parameter being estimated (BMD), they are impractical and would offer little to no guidance.

In short, I believe the statistical analyses underlying the screening value are appropriate for a situation where no perfect statistical approach is available.

Conclusions

In this report, I concentrated on the statistical evidence for the proposed screening value. As the authors of the report on which the screening values are based (Coffin et al., 2022) have noted, the evidence for making such recommendations is not extensive and the mammalian animal experiments conducted do not reflect the potential complexity of the chemistry of MP's in the environment. However, the authors attempted to use the animal data to inform the screening value and they did so using a conservative approach that balanced caution with providing practical information for guidance. To quote Coffin, et al, 2022: The screening value "…represents the most conservative estimate derived from the mammalian effects data considered in this study towards characterizing a potential human-health effect, which we suggest should only be used for helping to guide monitoring activities." No perfect methodology exists for the statistical inference problem of calculating the screening level from the limited existing data. However, given the limitations of the information available, the methods used are reasonable and justify the language in the draft handbook[4] quoted above. Finally, the provided guidance should be re-evaluated as more extensive microplastic exposure data, in the form of experimental animal, human exposure and human health impact data, become available.

---

[4] Proposed rule-policy handbook_11-10-2021 document, section 4.1.

References

Coffin, S., Bouwmeester, H., Brander, S., Damdimopoulou, P., Gouin, T., Hermabessiere, L., and Kahn, E. (2022). Development and Application of a Health-Based Framework for Informing Regulatory Action in Relation to Exposure of Microplastic Particles in California Drinking Water. Microplastics Nanoplastics *Submitted*.

Hou, J., Lei, Z., Cui, L., Hou, Y., Yang, L., An, R., Wang, Q., Li, S., Zhang, H., and Zhang, L. (2021). Polystyrene microplastics lead to pyroptosis and apoptosis of ovarian granulosa cells via NLRP3/Caspase-1 signaling pathway in rats. Ecotoxicol. Environ. Saf. *212*, 112012.

van der Laan, M.J., Polley, E.C., and Hubbard, A.E. (2007). Super learning. Stat Appl Genet Mol Biol *6*, Article25.

Rahman, A., Sarkar, A., Yadav, O.P., Achari, G., and Slobodnik, J. (2021). Potential human health risks due to environmental exposure to nano-and microplastics and knowledge gaps: A scoping review. Sci. Total Environ. *757*, 143872.

Rosenblum, M.A., and Van Der Laan, M.J. (2009). Confidence intervals for the population mean tailored to small sample sizes, with applications to survey sampling. Int. J. Biostat. *5*.

US EPA (2012). Benchmark Dose Technical Guidance.

Zarus, G.M., Muianga, C., Hunter, C.M., and Pappas, R.S. (2021). A review of data for quantifying human exposures to micro and nanoplastics and potential health risks. Sci. Total Environ. *756*, 144010.

(2019). Microplastics in Drinking-Water (World Health Organization).