Appendix F
ADDITIONAL DETAILS ON THE LIMITATIONS OF THE NTP REVIEW

FLUORIDEALERT.ORG
Fluoride Action Network

Submission to NAS on the revised draft NTP Monograph on the Systematic Review of Fluoride Exposure and Neurodevelopmental and Cognitive Health Effects[1]

October 19, 2020

Chris Neurath
Research Director
Fluoride Action Network (FAN)

Introduction

As the nominators for the NTP's systematic review of fluoride neurotoxicity we are pleased to submit comments to the NAS peer-review committee on the revised NTP monograph. We nominated fluoride for review in 2015 and it is noteworthy that since then the scientific evidence has rapidly expanded with the highest quality studies all being published since 2017. In just the months following the first draft NTP monograph there have been several more high-quality studies published that have been added to the revised monograph. The studies published in the past 5 years are important for their high quality and because many found adverse neurotoxic effects at low exposure levels, including the level (0.7 mg/L) used in artificial water fluoridation.

At the open meeting of this NAS committee in October 2019 we submitted comments on the first draft of the NTP monograph. We found that the objective aspects of systematic review were well conducted and followed OHAT guidelines but the more subjective aspects had some serious short-comings, including lack of transparency and lack of pre-specified methods in the protocol.

We found the main conclusion, that *fluoride poses a presumed hazard of developmental neurotoxicity*, to be well supported by the body of scientific evidence. However, we found that the section titled "Generalizability to the U.S. Population" had serious problems. It was, in essence, an informal risk assessment that incorporated simplistic exposure assessments and dose-response assessments to reach a conclusion about the confidence that fluoride at exposure levels in the US are likely to cause developmental neurotoxicity. The protocol, however, contained no mention of a Generalizability assessment which represents a fundamental violation of transparency and pre-specification. The Generalizability section appeared to be *ad hoc* and tacked on to the report.

[1] This is a modified excerpt of the submission that FAN submitted to the NAS.

This NAS committee also criticized the inclusion of the Generalizability section and recommended that NTP eliminate it and restrict itself to a hazard assessment and state clearly that the purpose of the NTP systematic review was not to weigh in on what a safe dose might be:

> Lastly, the discussion section of the monograph provides an informal assessment of the evidence with regard to exposure and concludes that adverse health effects are observed largely in association with exposures above those associated with water fluoridation. *The basis of that conclusion is not apparent and seems to contradict the earlier assertion that nearly all the studies are positive, including ones that evaluated groups exposed to lower concentrations*. More important, as noted above, *this discussion gives a false impression that NTP conducted a formal dose–response assessment. NTP should be clear that the monograph cannot be used to assess what concentrations of fluoride are safe.* [*emphasis added*; NAS 2020, p 5]

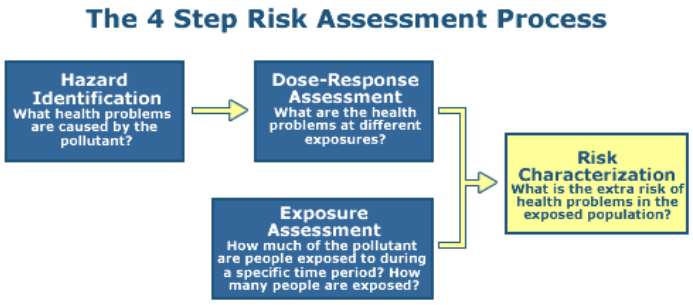
While we agree with the NAS's criticism of the Generalizability section we also believe the NTP's systematic review found sufficient evidence to answer the key question:

> *Is artificial water fluoridation in the US likely to be causing harm from developmental neurotoxicity?*

The NAS committee acknowledged this was a question of central interest.

To answer that question the NTP would have to use rigorous exposure assessment and dose-response assessment methods rather than *ad hoc* informal methods. The US EPA provides the following diagram (Figure 1) to clarify the distinction between the four steps of a risk assessment. It shows that to determine whether there is a risk at the exposure level due to artificial fluoridation both Dose-Response Assessment and Exposure Assessment are required. The EPA has extensive guidance for conducting valid assessments.

Figure 1.



Step 4: Risk Characterization is the last step of a human health risk assessment.
from https://www.epa.gov/risk/conducting-human-health-risk-assessment

Given that the revised monograph has addressed most of the suggestions for improvement offered by NAS, we believe there is now even stronger support for the presumed hazard conclusion. We urge NAS to now take a broader view and compare the fluoride neurotoxicity monograph to NTP's other monographs, as a way to check whether NTP has applied consistent standards for evaluating different chemicals, a primary goal of the OHAT systematic review process.

WEAKNESSES OF NTP'S REPORT

1. OHAT guidance on "unexplained inconsistency" ignored

Although OHAT provides no guidance on what constitutes consistency, it does clearly state that only "unexplained inconsistency" is grounds for downgrading evidence. The revised monograph Generalizability section argues that studies at exposures relevant to the US are inconsistent. However, on closer inspection, the inconsistency is only on effect magnitude or individual study dose-response relationships. All of these inconsistencies are explainable from differences in study design, populations, gender, exposure measures, exposure levels, and outcome measures. OHAT guidance says all of these reasons for differences are acceptable explanations and do not justify downgrading.

> Inconsistency that can be explained, such as variability in study populations, would not be eligible for a downgrade. Potential sources of inconsistency across studies are explored, including consideration of population or animal model (e.g., cohort, species, strain, sex, lifestage at exposure and assessment); exposure or treatment duration, level, or timing relative to outcome; study methodology (e.g., route of administration, methodology used to measure health outcome); conflict of interest, and statistical power and risk of bias. Generally, there is no downgrade when identified sources of inconsistency can be attributed to study design features such as differences in species, timing of exposure, or health outcome assessment. [NTP 2019 OHAT Handbook, p 53]

2. NTP used inappropriate methods for "Generalizability to the U.S."

While the evidence base and NTP's conclusion of presumed hazard has strengthened since the previous draft monograph, we also find that NTP has ignored the NAS recommendations to focus on hazard assessment and to avoid risk assessment and trying to identify a safe exposure level. The section titled "Generalizability to the U.S. Population" goes beyond Hazard Assessment into what is essentially risk assessment. Rather than omit the Generalizability section NTP has expanded it. Furthermore, we find that NTP has not used proper exposure assessment and dose-response assessment methods to underpin their revised Generalizability section. Their informal methods have tended to downgrade the evidence and understate the risk at low doses. The NTP has still not included any mention of a generalizability assessment in their protocol so there continues to be a lack of transparency and pre-specification. The only addition to their protocol is a description of a planned dose-response meta-analysis which would

presumably contribute to their generalizability section. However, a dose-response meta-analysis is not in itself a dose-response assessment. Ultimately, the planned dose-response meta-analysis was not even conducted for the studies with individual-participant data, which includes most of the strongest studies. Finally, the revised protocol addition describing dose-response meta-analysis was not released until September 16, 2020 and no public comment period was provided. This is a further deficiency of transparency and openness to public comment. If there had been opportunity to comment on the revised protocol we would have raised these concerns before the revisions to the systematic review were implemented.

3. Exposure assessment is simplistic and inadequate

The extent of NTP's exposure assessment to support its Generalizability section appears to be a single footnote with a link to a CDC website [NTP 2020 monograph, p 2]. Furthermore, the linked website does not contain any of the exposure information stated in the footnote but instead gives a general description of a confidential database managed by the CDC. The database is only accessible to approved CDC staff and state oral health and drinking water officials, not the public. The confidential database is called the Water Fluoridation Reporting System or WFRS.
https://www.cdc.gov/fluoridation/data-tools/reporting-system.html

There is thus no transparency in the exposure assessment. The public summaries of the WFRS data (CDC's My Water's Fluoride website) are not sufficient to support a valid exposure assessment of the US population and have been found to contain serious limitations and errors such as reporting water systems having artificial fluoridation at 1.2 mg/L as being at 0.7 mg/L.

4. Proper exposure assessment demonstrates that the NTP's presumed hazard conclusion applies directly to doses from artificial fluoridation

Even if the exposure information in the footnote could be verified and was reliable, it is insufficient for a valid exposure assessment. It is a summary of drinking water fluoride concentrations and only for public water systems and only for naturally occurring levels. Valid exposure assessments require dose information which requires information on the amount of water consumed in addition to its concentration. As described in more details below, the EPA has found that the 95th percentile consumer of water drinks more than twice as much as the average consumer. That finding applies to all ages. Therefore, the top 5% of consumers (millions of people in the US) when drinking water with a concentration of 0.7 mg/L will receive the same doses as the average person in a study where the drinking water concentration is 1.5 mg/L. Thus, studies finding that average exposures to 1.5 mg/L cause neurotoxic harm directly support a conclusion that millions of people in the US with artificial fluoridation at a concentration 0.7 mg/L will be harmed. The NTP has made the fundamental error of conflating concentration with dose and not

accounting for the wide range of doses that will occur for any given concentration of fluoride in drinking water.

5. Specific examples of downgraded evidence in revised NTP monograph

• **Excluding largest effect in the strongest study.** For the Bashash 2017 study, which was one of the strongest studies, and at exposures relevant to artificial water fluoridation, the NTP improperly focused on a minor secondary analysis and largely excluded consideration of the primary analyses, especially the primary analysis with the largest effect at the lowest exposure levels. The NTP focused on a comparison with the dichotomous exposure levels of <0.8 mg/L or ≥0.8 mg/L child urine F. There was only a small difference in IQ score between these two groups and it was not statistically significant. The small difference may be explained by the reduced information in the analysis and because child urine F was the exposure measure, not maternal urine F.

In contrast to this secondary analysis, the primary planned analyses of the study were largely excluded in narrative and meta-analysis portions of the NTP monograph. They were the multiple regression models between maternal urine F and GCI score for 4-year olds and WASI FSIQ score for 6-12-year olds. Both analyses found large statistically significant effects. The NTP further discriminated against the findings in GCI scores at 4 years old by treating it not as a measure of neurocognitive development but as an "other outcome". In meta-analyses and summaries of data the GCI analysis was thus excluded. Yet the GCI analysis found a linear dose-response relationship with no threshold, while the WASI FSIQ analysis found what may be a threshold at 0.8 mg/L. By excluding the GCI analysis, the NTP excluded the larger effect that occurred at lower doses. The GCI score should have been classified by NTP as a neurocognitive outcome rather than "other outcome". GCI is generally considered as a valid measure of neurocognitive development and has a strong correlation with several of the tests NTP did classify as tests of neurocognitive development. The abbreviation GCI stands for General Cognitive Index, which in itself should have helped NTP recognize it as a test of neurocognitive development.

• **Excluding strongest low-dose studies from dose-response meta-analysis.** The closest the revised NTP monograph gets to a proper dose-response assessment to support its generalizability section is a dose-response meta-analysis. There are several problems with it, however, the most serious being that they excluded the 10 studies with individual-participant data and instead relied on lower quality evidence from studies with only group-level analyses. The 10 individual-participant data studies included all of the highest quality studies and many of the studies at low doses, so this exclusion is especially problematical for conducting a valid and balanced dose-response assessment. It is important to note that this was a planned analysis in the revised protocol, so the decision to not conduct it is troubling. All other planned meta-analyses were conducted. Furthermore, the very brief reasons given for not conducting the analysis are unjustified and represent a double-standard. Here is what the NTP monograph said about the individual-participant dose-response meta-analysis:

> "A dose-response meta-analysis using the effect estimates reported in studies with individual-level exposure was considered. However, because of the small number of studies (n = 10), the various types of exposure metrics, and the different types of reported effect estimates that could not be combined, a dose-response meta-analysis of these studies could not be conducted." [NTP 2020 revised monograph p 253]

Taking each of the three stated reasons separately:

1.) "small number of studies (n = 10)" The claim that 10 studies are insufficient is contradicted by the NTP's own actions elsewhere in the monograph. The NTP conducted dose-response meta-analyses on studies without individual-level data when there were as few as 4 studies and meta-analyses on subgroups with as few as 2 studies.

2.) "various types of exposure metrics" Just as with "small number of studies", in other dose-response meta-analyses and meta-analyses the NTP has combined various types of exposure metrics. Elsewhere in the NTP monograph this issue is discussed and NTP concludes that for comparison purposes urine F levels can be considered equivalent to water F levels on a 1-to-1 ratio [NTP 2020 monograph p 72].

3.) "different types of reported effect estimates" Again, in other meta-analyses and dose-response meta-analyses the NTP combined studies with different effect measures. The revised NTP protocol considered a wide range of tests to be classifiable under the general domain "Leaning, Memory, Intelligence, Cognitive Development" [NTP 2020 protocol, Table 6, p 20]. Ten different specific tests were listed as examples that fit within this domain of outcomes.

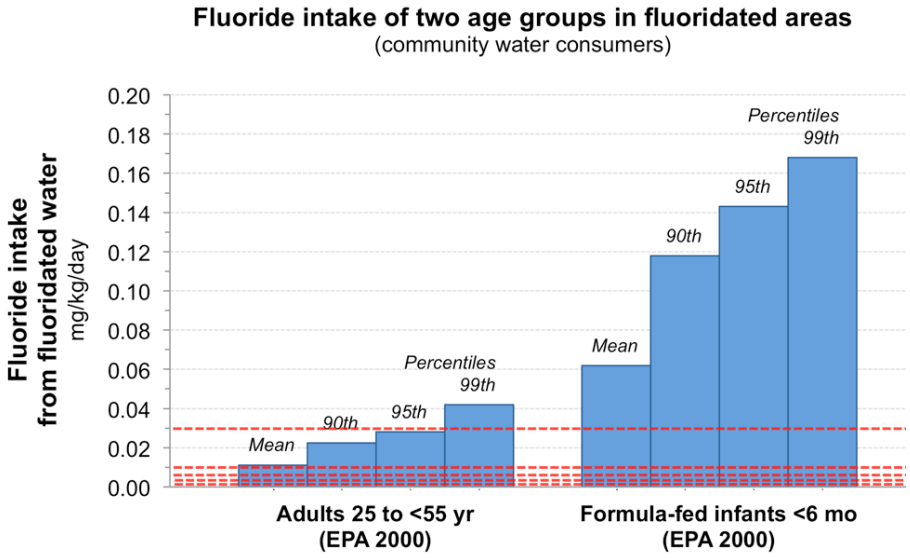
NTP has exhibited a clear double-standard when claiming they were unable to conduct a dose-response meta-analysis on the 10 studies with individual-level exposure data. Also, this represents a failure of the NTP to follow their revised protocol.

- **Unnecessary division of studies lowers power in dose-response meta-analyses.** The NTP's dose-response meta-analysis of *group-level studies* unnecessarily stratified by whether exposure was measured in urine F or water F [NTP 2020 monograph Table A5-3, p 254]. This stratification reduced the statistical power and produced lower confidence in pooled results for each subgroup of studies.

- **Simplistic exposure assessment underestimates hazard at doses relevant to US.** The NTP's simplistic exposure assessment assumed that only water F or urine F concentrations below 1.5 mg/L were applicable to the US. They further assumed that only water F or urine F concentrations of 0.7 mg/L are applicable to artificial fluoridation in the US. This perpetuates a fundamental error made by many fluoridation proponents that concentration is equivalent to dose. The US EPA has conducted rigorous exposure assessments of F from drinking water. They find that the 95th percentile of water

consumers, on a mL per kg body weight basis, consume about twice as much water and fluoride as the average consumer (see Figure 2). Therefore, the 95th percentile consumers drinking water with a concentration of 0.7 mg/L are receiving double the dose of the average water consumer. They thus receive the same dose as the average water consumer drinking water with a concentration of 1.5 mg/L. The consequence is that studies finding harm at water concentrations of 1.5 mg/L are relevant to the top 95th percentile water consumers. This subpopulation represents millions of people in the US and must be considered when generalizing from the results of epidemiological studies to the actual exposures in the US. This realistic exposure assessment alone is sufficient for the NTP to conclude that artificially fluoridated water at 0.7 mg/L is a presumed developmental neurotoxin for the 5% of the US population who consume the most water. This realistic exposure assessment also greatly expands the number of studies which should be considered relevant to exposures in the US and from artificial water fluoridation. Instead of 1.5 mg/L as the cut-off for relevance, the level should be 3.0 mg/L. Using this more appropriate cut-off, the NTP's dose-response meta-analysis shows that the group-level studies with water F as the exposure measure already show a pooled estimate that is statistically significant and large (SMD -0.27 equivalent to -4 IQ points for studies with mean water F below 2 mg/L) [NTP 2020 monograph Table A5-3, p 254].

Figure 2. Distribution of fluoride intake from fluoridated water, USA



Intakes based on EPA 2000 Water Intake Estimates in US, p. IV-3.

• **NTP's simplistic dose-response meta-analysis methods underestimated effects at low doses because they used the mean exposure while most studies had individual-level exposures that ranged well below the mean.** Furthermore, NTP dichotomized studies by whether the mean exposure was above or below a cut-off of 1.5 mg/L [NTP 2020 monograph Table A5-3, p 254]. The loss of information in taking the mean and then dichotomizing by the mean value is contrary to standard dose-response

assessment methodology. For example, the EPA currently prefers Benchmark Dose (BMD) methods be used for dose-response assessment. BMD methods account for the totality of the data and provide estimates of the dose likely to cause a specified degree of harm. That dose, called the BMD, is frequently lower than the mean dose.

To illustrate the difference between the information available in the complete individual-level data and just the mean or the dichotomized mean, we use simulated data from a hypothetical study, first plotted as a scattergram with all the data points (Figure 3) and then as a single data point showing the mean dose and mean response (Figure 4). These illustrations are then followed by the results of a BMD analysis of the same individual-level simulated data (Figure 5).

The mean exposure in this hypothetical study is 1.7 mg/L. That puts it over NTP's cut-off for relevance to exposures in the US. Yet the full exposure distribution and dose-response relationship as shown in the scattergram clearly shows it is relevant to exposures below 1.5 mg/L. This illustrates why NTP's informal dose-response analysis and generalizability discussion are invalid and will underestimate the confidence that exposures in the US and from artificial fluoridation will produce harm.

Figure 3. Hypothetical individual-level data:

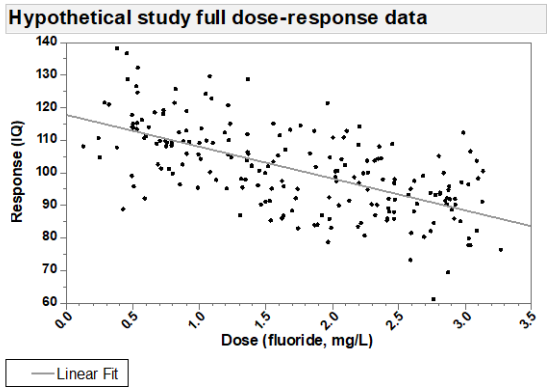


Figure 4. Here is what NTP has reduced it to:

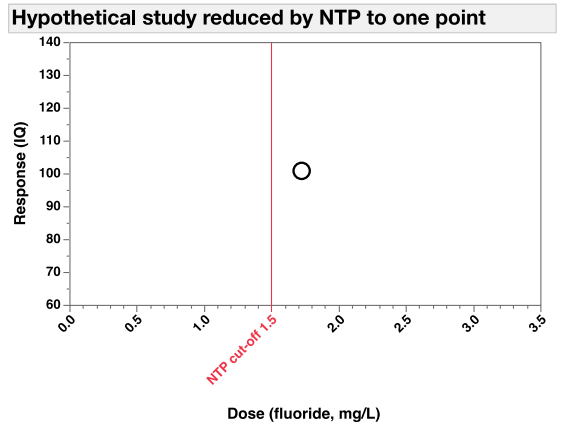
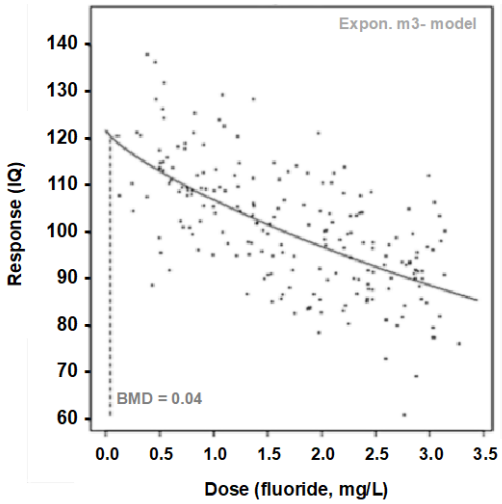


Figure 5. Hypothetical study Benchmark Dose (BMD) analysis:



Best fit model Exponential m3- using PROAST website of EFSA; BMR = -1 IQ point.

Benchmark Dose (BMD) analysis uses all available data from a study to estimate at what dose an adverse effect is predicted to occur. It was developed as an improvement over other methods of dose-response assessment and is EPA's preferred method when suitable data is available. This BMD analysis considered non-linear dose-response models.

This hypothetical study example, that NTP would have classified as too high a dose to be relevant to exposures below 1.5 mg/L, is found to provide clear evidence of an adverse effect well below 1.5 mg/L when analyzed with the BMD method.

6. Additional weaknesses of the revised NTP monograph

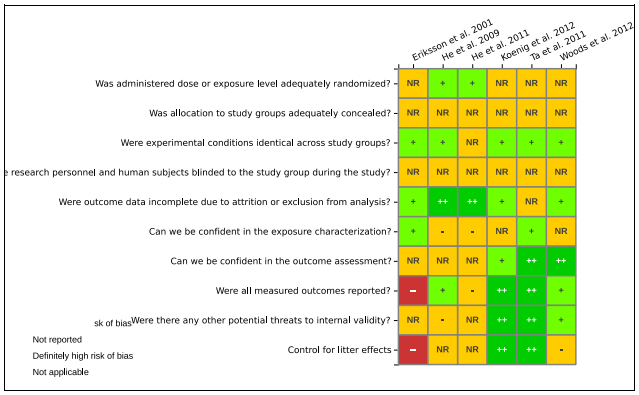
• **Improperly downgraded the animal evidence to "inadequate" despite the NTP 2016 review of the animal evidence concluding it was "low to moderate".** The NAS committee specifically chastised NTP for improperly downgrading the evidence based on the claim that sensory/motor effects might have played a role in the deficits in the learning and memory tests. Despite this clear rejection of that argument the NTP persisted and continues to use it in the revised monograph to downgrade the animal evidence. Furthermore, additional animal studies were identified since the NTP 2016 review, several of which were scored high quality. Therefore, it is difficult to understand how the revised NTP monograph can downgrade the overall body of animal evidence to "inadequate".

The NAS also raised concern in the other direction about some of the animal studies which the NAS suggested should have higher RoB scores. The concerns were mostly because of deficiencies in reporting, such as not reporting whether researchers were blinded to exposure status and whether litter effects had been controlled.

However, just as comparison to other OHAT systematic reviews for other chemicals provides perspective for the strength of evidence necessary to reach overall confidence conclusions, it is appropriate to consider other OHAT reviews and how they scored individual animal studies for RoB and how they assessed the overall animal evidence for confidence. The same NAS review used as an example above, for BDE-47, and employing OHAT methodology, demonstrates that the OHAT methodology applied to much weaker animal evidence than is available for fluoride was sufficient to give an animal evidence confidence rating of "moderate".

Figure 6 shows the total extent of animal studies of BDE-47 upon which a "moderate" confidence rating was concluded for effects on learning. Of the 6 animal studies, 5 found "some indication of effect on at least one measure of learning" for a consistency of 83% [NAS 2017, p 8]. But the quality of all 6 studies was low.

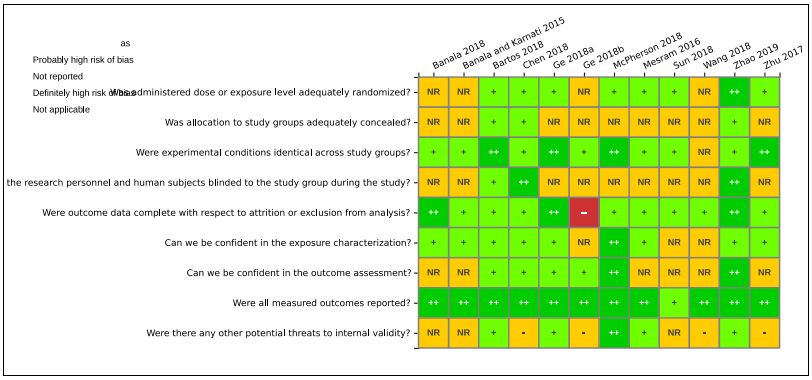
Figure 6. BDE-47 RoB heatmap, all animal studies



from NAS 2017 report: https://doi.org/10.17226/24758
HAWC visualization: https://hawcproject.org/summary/visual/353/

For comparison, Figure 7 is the RoB heatmap for just the newer fluoride animal studies.

Figure 7. Fluoride RoB heatmap, recent animal studies

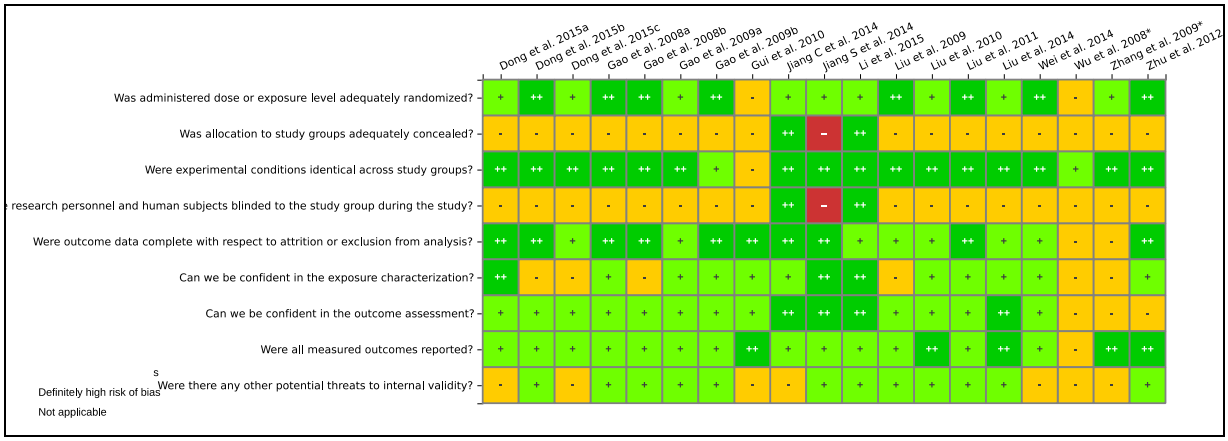


HAWC visualization: https://hawcproject.org/summary/visual/530/

Fluoride has 12 studies to just 6 for BDE-47. Five of the 12 fluoride studies would be rated lower RoB while none of the BDE-47 studies would rate lower RoB. The NTP protocol requires that to be rated "lower RoB" no more than two of the key RoB domains be rated "yellow" or "red".

These 12 animal studies are just those identified since the NTP 2016 systematic review of fluoride neurotoxicity in animals. The NTP 2016 review identified 19 additional earlier studies that used Morris Water Maze tests, considered the most applicable to learning and memory. Figure 8 is the RoB heatmap for these earlier studies.

Figure 8. Fluoride RoB heatmap, animal studies from NTP 2016



HAWC visualization: https://hawcproject.org/summary/visual/33/

Of the 19 additional animal studies, 8 would be rated as lower RoB using NTP's criteria. For fluoride the total number of animal studies is 31 with 13 rated lower RoB. The total number of BDE-47 is just 6 with none rated lower RoB.

Thus, the fluoride animal evidence is substantially greater in quantity and quality than BDE-47, yet NTP has rated confidence in fluoride "inadequate" while BDE-47 is rated "moderate".

With regards to consistency, almost all of the fluoride studies of learning and memory found statistically significant adverse effects. Even the much-touted McPherson 2018 study found a statistically significant adverse effect that was not acknowledged in its summary or abstract. Therefore, on consistency, fluoride is stronger than BDE-47 as well.

Finally, the doses (as measured in body tissues) found to cause adverse effects in the BDE-47 animal studies were hundreds of times higher than commonly occur in humans [Staskal 2007, EPA 2008]. In contrast, the doses of fluoride found to cause adverse effects in most of the animal studies were less than 20 times higher than commonly occur in humans, when taking into account pharmacokinetic differences. The OHAT Handbook

requires conversion between external and internal dosimetry in assessing relevance of animal studies to human exposures:

> "*Exposure*
> • *Route of administration in animal studies:* External dose comparisons used to reach level of concern conclusions need to consider internal dosimetry in animal models, which can vary based on route of administration, species, age, diet, and other cofactors." [NTP 2019 OHAT Handbook, p 58]

There is no justification for downgrading the fluoride animal data because of claimed irrelevant doses to humans, as the revised NTP monograph does [NTP 2020 monograph, p 58].

In conclusion, both NTP and the NAS are applying an extreme double-standard, with fluoride having to meet a much higher bar than BDE-47. Despite much greater quantity, quality, and consistency of the animal evidence for fluoride, NTP has rated it "inadequate" and BDE-47 "moderate". This violates an overarching goal of OHAT systematic reviews, which is to consistently apply the same standards for judging the hazard of different chemicals. The OHAT handbook says an objective of the OHAT evaluation process is to "ensure consistency across evaluations" [NTP 2019 OHAT Handbook, p 1]. BDE-47 was used as an example here, but similar comparisons can be made to many other chemicals the NTP concluded had "moderate" animal evidence.

- **The NTP monograph deviated from the OHAT guidelines in its section "Generalizability to the U.S. Population".** While OHAT methodology has little guidance on dose-response analysis, the revised monograph did not even follow what is available. The Generalizability section is essentially a dose-response and risk assessment evaluation although it did not follow OHAT methodology for such. OHAT methodology for dose-response and risk assessment is termed "Level of Concern Conclusions" (LoC) and is considered as a second type of conclusion beyond the initial "Hazard Identification Conclusion" [NTP 2019 OHAT handbook, p 3]. But the protocol for the NTP review does not include any Level of Concern assessment so there was no pre-specification of the Generalizability section. As stated above, the Generalizability section in the NTP monograph is not adequately documented or justified, even as a *post hoc* addition to the monograph. We believe this is a serious failure to follow the principles of transparency and pre-specification.

> "The National Toxicology Program (NTP) ... conducts evaluations to assess the evidence that environmental chemicals, physical substances, or mixtures (collectively referred to as "substances") cause adverse health effects and provides opinions on whether these substances may be of concern, given what is known about current human exposure levels. The opinions are referred to as NTP Level of Concern (LoC) conclusions." [NTP 2019 OHAT Handbook, p 1]

OHAT says LoC comes after Hazard Identification and requires an exposure assessment:

"*Exposure*

• *Human studies:* ... In OHAT's process, the applicability of a given exposure scenario for reaching a "level of concern" for a certain subpopulation is considered after hazard identification. For that subpopulation the health effect is interpreted in the context of what is known regarding the extent and nature of human exposure (Twombly 1998, Medlin 2003, Jahnke *et al.* 2005, Shelby 2005)." [NTP 2019 OHAT Handbook, p 58]

While OHAT offers little guidance on how NTP will conduct LoC determinations, it makes clear that an exposure assessment is required. The OHAT Handbook says NTP will "update" its LoC framework "to ensure integrated consideration of relevant and reliable evidence and to enhance transparency". The update is projected for completion in 2016-2017 but apparently has not yet been issued. Here is the currently available extent of guidance on LoC determinations:

• ***Level of Concern (LoC) Conclusions*** – For LoC conclusions OHAT integrates two categories of evidence: (1) health-outcome data from human, animal, and mechanistic studies to reach hazard identification conclusions and (2) information on the extent of exposure and pharmacokinetics. LoC conclusions are narrative (i.e., non-quantitative) conclusions that use a 5-point scale ranging from "negligible" to "serious" concern for exposure. As part of implementing systematic reviews the NTP will update its LoC framework to ensure integrated consideration of relevant and reliable evidence and to enhance transparency in describing how these conclusions are reached. These strategies will improve the LoC framework as a risk communication tool (expected completion in 2016-2017). The updated LoC framework will be included in a future version of the OHAT handbook." [NTP 2019 OHAT Handbook, p 3]

The OHAT Handbook specifies that the decision for whether NTP will conduct just a Hazard Assessment or also a Level of Concern determination should be made at the problem formulation stage, before a protocol is even written [NTP 2019 OHAT Handbook, p 10]. Neither the problem formulation nor protocol for NTP's review of fluoride neurotoxicity have ever mentioned a LoC determination as an objective.

• **Meta-analyses have inadequate documentation.** The meta-analyses are not adequately documented, especially the dose-response meta-analyses. The specific studies included at each dose should be provided in a table. Forest Plots should be provided for all dose-response meta-analyses. Bubble plots showing the dose-response curve with 95%CI for the dose-response meta-analyses should be provided.

No data underlying the meta-analyses are available at the HAWC project website, nor are any visualizations like Forest Plots available through HAWC. There is a downloadable Excel file named "meta-analysis data" but it is only column headings with no data. All data used in meta-analyses and dose-response meta-analyses should be provided in HAWC in downloadable data files as well as tables and visualizations.

7. Literature search update missed important recent study on adolescents

A study by Malin et al 2019 finding that fluoride exposure in adolescents was associated with disruptions in sleep patterns was not identified in the literature update [Malin 2019]. Apparently NTP's search criteria did not recognize sleep disruption as a form of developmental neurotoxicity. However, sleep is largely mediated by neurological functions and can impact neurological and psychological wellbeing in ways that may not be measured by intelligence tests. The Malin 2019 study is relevant because it was done with a NHANES sample of children age 16-19 years, that is nationally representative of the US. Exposure was measured through individual-level tap water samples. The study found a statistically significant doubling of odds of sleep-apnea symptoms for an increase of 0.5 mg/L in water F concentration. The authors suggested that fluoride may affect the pineal gland and melatonin production for which there is some animal study evidence. As a neuroendocrine organ in the brain, adverse effects on the pineal gland should be considered neurotoxic effects.

While this is the first study to ever examine sleep patterns in relationship to fluoride, it opens the possibility that fluoride neurotoxic effects might extend beyond prenatal and earlier childhood to adolescence. This could enlarge the portion of the population subject to neurotoxic harm from fluoride.

Conclusion

The scientific evidence is more than sufficient to justify a conclusion that fluoride is a presumed developmental neurotoxin in children. Compared to other chemicals reviewed by NTP and given a "presumed hazard" rating, there is greater quantity, quality, and consistency in the fluoride human studies.

It is extremely unlikely there could be any unidentified studies that could alter this conclusion. Likewise, the quantity and consistency of evidence mean it is extremely unlikely that any new studies could weaken this conclusion.

NTP's revisions have addressed NAS's recommendations and the revised monograph is substantially strengthened and more transparent as a result.

However, the Generalization section is weak and should be removed as recommended by NAS or redone using valid risk assessment methods. A valid risk assessment requires a valid exposure assessment and valid dose-response assessment. FAN has offered a risk assessment using methodologically rigorous and appropriate methods following EPA guidance. Our risk assessment finds that exposures to levels below 1.5 mg/L and below 0.7 mg/L both pose a high likelihood of neurotoxic harm to at least some proportion of children in the US population. Even without a formal risk assessment, recognition that about 5% of the population will receive twice the average dose because of greater than average water consumption provides sufficient support for this conclusion, when coupled with the NTP conclusion of presumed hazard above 1.5 mg/L.