like further details on how differences between reviewers were reconciled. For example, what were the kappa statistics or intraclass correlation coefficients on the title/abstract review prior to reconciliation? Did the degree of agreement warrant further training of the reviewers? The same concerns regarding reliability can be made for the data extraction.

Response: Disagree (no change)

- We have not found Kappa Statistics or other measures of inter-rater agreement to be useful in obtaining agreement in the screening process. Instead, we have implemented a process that emphasizes training, pilot testing, and group discussion to assure consistency of approach. As described in the *Methods* section, screening is conducted by two independent screeners at both the title/abstract and full-text steps. When conflicts arise, they are resolved through discussion between the two screeners and consultation with a senior team member, if necessary, to reach consensus. Our protocol also describes that training and pilot testing phases are conducted on small sets of studies to assure consistency of approach in applying the PECO criteria and inclusion/exclusion guidance. When questions arise, we address them as a group so that all screeners develop a consistent approach. The process emphasizes inclusion of studies if there is any uncertainty at the title/abstract stage. At the full-text stage, we confirm that studies indeed have original data and meet the PECO criteria, so there is little uncertainty at that step. Studies either have the relevant data or they do not. In addition, while cross-screener agreement within a project team is essential when each reference is screened by a single reviewer, the issue has a much smaller potential impact when two screeners review each study in duplicate, as in this systematic review.
- The review process for data extraction involves a quality control (QC) review rather than extraction in duplicate. Data extraction is conducted by a single extractor followed by QC review because we have not found added value or reliability with independent data extractions. The QC review is conducted for all data extracted into HAWC (https://hawcproject.org/assessment/405/), the web-based content management system for our systematic reviews.

H.11:

[redacted] **Comments:** The use of the SWIFT-Active Screener is well described and addresses the concerns in the prior review.

Response: No change requested

- No response necessary.

H.12:

[redacted] **Comments:** The supplemental search of the non-English language databases is appropriate. However, what is the rationale for saying that they were used primarily to identify null or no-effect studies? Does that mean that if a study was identified that showed an association it was not abstracted? Please be a bit more clear on this.

Response: Agree (change made)

- Although extraction of studies identified from the Chinese database searches was previously focused on no-effect studies, we have taken steps to translate and extract data from all non-English studies identified from the Chinese database

searches. Therefore, the statement about null or no-effect studies no longer applies and has been deleted.

2. Comment on whether the approach used to assess risk of bias was clearly described and appropriately applied to the set of studies designated as "low risk of bias."

H.13:

[redacted] **Comments:** The focus on confounding, exposure characterization and outcome assessment are, as indicated, the key components of evaluating observational research. The other parameter is whether the participants represent the population from which they are recruited, i.e. selection bias. In prospective cohort studies this is not an issue, as the population is really the combination of those exposed and non-exposed. For cross sectional studies, this is a bit trickier, as the participants may reflect a select group within the overall population. For studies based on national or regional registries, such as the Canadian studies, this is less of a problem, but for others there is the possibility of bias, and the direction of such bias is difficult to predict. As [redacted] looked at the studies, the vast majority do not address this issue, but [redacted] believe that it is worth a discussion or at least a mention that the possibility of selection bias is real.

Response: Agree (change made)

- We agree with [redacted] that selection bias is an important consideration in risk-of-bias evaluations. We have edited the following text in the *Methods* section to clarify that, in addition to the three key risk-of-bias questions, the answers to the other risk-of-bias questions were considered in assessing potential bias, including selection bias.

  "The other risk-of-bias questions, including selection of study participants, were also considered and were used to identify any other risk-of-bias concerns that may indicate serious issues with a study that could cause it to be considered high risk of bias."

- *Appendix E* includes a detailed summary of the population selection and the basis for the ratings for selection bias and exposure characterization. All 19 low risk-of-bias IQ-in-children studies and 9 other neurobehavioral studies in children were rated either *probably low risk of bias* or *definitely low risk of bias* due to selection bias. Generally speaking, these studies provide direct or indirect evidence that exposure groups were similar and were recruited within the same timeframe using the same methods with no differences in participation/response rates (i.e., either direct evidence of similar participation/response rates or no evidence of differences in participation/response rates). Differences in the subjects across exposure groups were noted and addressed in the analysis.

H.14:

[redacted] **Comments:** For confounding, please see [redacted] remarks above. [redacted] do think that biological sex needs to be considered an effect modifier as in other studies of neurotoxins and neurodevelopmental outcomes. Further, as indicated later in the monograph, at times the choice of confounders needs to be study and area specific, so this should also be mentioned in this section. Finally, for the arsenic variable, as [redacted] indicated above [redacted] really appreciate the efforts made in defining this. However, please justify the choice of

10µg/L as the cutpoint – while it is the WHO guideline it is quite possible that there are neurodevelopmental effects with concentrations under this level.

Response: Agree (change made)

- We agree that biological sex should be considered a potential effect modifier in addition to (not instead of) a potential confounder. Please see previous response for details on our rationale and how text was revised to address [redacted] comments on confounding.
- Regarding choice of important covariates being study- and area-specific, we consider what we currently state in the *Methods* section to address [redacted] suggestion*:*

  "Additional covariates considered important for this evaluation, depending on the study population and outcome, included…" and, "To be assigned a rating of probably low risk of bias for the key risk-of-bias question regarding the confounding domain, studies were not required to address every important covariate listed; however, studies were required to address the three key covariates for all studies, the potential for co-exposures, if applicable (e.g., arsenic and lead, both of which could affect cognitive function), and any other potential covariates considered important for the specific study population and outcome."
- As for the choice of 10 µg/L as the cutoff point, [redacted] is correct that we chose this based on the WHO guideline (WHO 2017). We agree that it is possible there may be neurodevelopmental effects at concentrations below 10 µg/L; however, we have no basis on which to select a lower cutoff point. Note that we had initially added a statement to the Sup02_2022_Prepublication_NTP_Monograph stating that "*arsenic may be associated with neurodevelopment effects at concentrations below 10 µg/L*" in response to this reviewer's comment; however, as we were unable to support this statement with a reference, it has been removed.

H.15:

[redacted] **Comments:** Exposure characterization: This is well described. As [redacted] mention above, missing is a discussion regarding the toxicokinetics of fluoride, to allow the reader to make decisions on how good the spot urine samples are in reflecting cumulative exposure. [redacted] understand that there is a high correlation between the spot urine samples and 24 hour collections (with and without correction for dilution) but this still does not give day to day, week to week, or season to season variation.

Response: Agree (change made)

- As described in a previous comment, we have added a brief discussion on fluoride toxicokinetics at the beginning of the *Exposure* section of the *Risk-of-bias Considerations for Human Studies* section.
- With respect to variations in fluoride exposures over time, we agree that additional study of these variations would be interesting; however, our assumption is that individual exposure to fluoride is relatively consistent because it reflects personal preferences and habits (e.g., daily water consumption, tea consumption, dental product use).

H.16:

[redacted] **Comments:** A further concern with exposure assessment brought up in the previous review concerns the issue of clustering with regard to exposure. The authors of the monograph do a very nice job of addressing this issue as it was raised in the prior review, but pointing to the sensitivity analyses. [redacted] only concern remaining is that this is mentioned up front when the exposure characterization is discussed in the methods.

Response: Agree (change made)

- To address this suggestion, we have provided an additional sentence in the *Methods* section where risk-of-bias considerations for exposure are discussed.

  "Ideally, these studies would still need to consider and adjust for area-level clustering; however, these concerns are captured in evaluations of other potential threats to internal validity."

H.17:

[redacted] **Comments:** Finally, some measure of agreement between the raters on their bias assessment would be a good addition.

Response: Disagree (no change)

- While we appreciate this comment, we have not found measures of inter-rater agreement (e.g., kappa statistics) to be useful in this process and instead have implemented a process that emphasizes pilot testing to develop a consistent approach and group discussion when there are questions in the rating. In addition, to further support consistency, a senior member of the team served as one of the risk-of-bias assessors for all of the studies. In addition, while cross-reviewer agreement within a project team is essential when each reference is assessed by a single reviewer, the issue has a much smaller impact when two screeners review each study in duplicate, as in the current systematic review. We consider that the most important issue for consistency is to reach collective agreement, and the final risk-of-bias ratings reflect that agreement.

3. Comment on assessment of the human studies with regard to:

a) How findings from individual studies designated as "low risk of bias" were interpreted.

H.18:

[redacted] **Comments:** In general, studies designated as "low risk of bias" were interpreted correctly. [redacted] have a few suggestions as to how to clarify many of the points made.

While the results are generally consistent (table 6) it would be useful to present the results based on the exposure metric used. For example, studies using fluoride concentrations in "high" and "low" areas could be grouped together to illustrate the change in IQ points. Additionally, the actual IQ test used could also be used to group studies within exposure metric. There are clear differences in the scoring for the Raven and the WASI/WPPSI, for example and these are hard to tell from the presentation.

Response: Disagree (edited for clarity)

- We considered several ways to organize the table and each way has its benefits and drawbacks. There are limitations to a static table, which is why we are increasing our use of interactive tools and platforms to visualize data. For the purpose of this document, we consider the current organization to be most clear and appropriate for providing a quick summary of study characteristics and key findings per study. We have edited the paragraph that precedes Table 6 to clarify that the Table 6 organization is by country and then by year.
- Note that we considered [redacted] suggestion to group studies using fluoride concentrations in "low" and "high" areas together to illustrate the change in IQ scores. While an association is consistently observed when comparing low to high fluoride areas, comparing changes in IQ scores across these studies is challenging due to the variability in the exposure levels that are considered "low" and "high." There are no consistent definitions of "low" and "high" that apply across all cases. For this reason, we do not find this suggested organizational structure for Table 6 to be a more effective presentation of the data. We also considered [redacted] suggestion to group studies by IQ test; however, as the Raven's tests were almost exclusively conducted in China, India, and Iran, the current organization by country, to a large extent, also organizes the studies by IQ test. Therefore, we find the current structure accommodating for focusing on results by IQ test.

H.19:

[redacted] **Comments:** At times, associations are presented as different when other covariates are controlled. [redacted] presume that these assessments were made by inspection of the results in the studies, but should either be backed up with statistical testing or admitted that they were made by inspection. For example, in table 6 the study by Rocha-Amador, et al states that the estimated associations between fluoride and the full scale IQ (WISC) were smaller when arsenic was controlled, the estimated betas are given, but there is no indication whether the differences are statistically different.

Response: Disagree (edited for clarity)

- When study authors present associations between fluoride exposure and IQ that differ when other covariates are included, we reported the results as described by the study authors. We did not perform additional testing to support the author's reporting of results as this is beyond the scope of the assessment.
- The statement that [redacted] notes for Rocha-Amador et al. (2007) and the association with arsenic was misinterpreted. The purpose of the statement was to note that the association between arsenic exposure and children's IQ was smaller in magnitude than the association between fluoride exposure and children's IQ, not that the association with fluoride was smaller after controlling for arsenic. The revised text in Table 6 of the Sup02_2022_Prepublication_NTP_Monograph reads as follows:

  "Significant associations between log-transformed fluoride and IQ scores (full-scale IQ adjusted ß's of −10.2 [water] and −16.9 [urine]; CIs not reported); arsenic also present, but the association between log-transformed arsenic and IQ scores was smaller (full-scale IQ adjusted ß's of −6.15 [water] and −5.72 [urine]; CIs not reported)"

H.20:

[redacted] **Comments:** Please note when the result is not statistically significant and likely due to small sample sizes (e.g. discussion of the Green et al paper on page 37). Also for that paper, the results seem to be different by biological sex, an example of effect modification that would be expected for a neurotoxin.

Response: Agree (change made)

- We added the qualifier "not significant" for the results in girls. However, since the scope of this section is to present the observed IQ effects in children, we refrain from suggesting reasons for non-significance, such as sample size. In each study, there are a multitude of factors that could yield nonsignificant results, in addition to lack of power. The study-specific risk-of-bias evaluations describe study details (including sample size) and aspects that could impact the ability to detect an association. With respect to biological sex as an effect modifier, we consider our revised terminology in response to a previous comment to address [redacted] concern.

H.21:

[redacted] **Comments:** The results also need to be interpreted based on age of test administration. Some higher order functions do not develop until later ages and thus cannot be tested well in younger children. Also, as with other neurotoxins, deficits can occur at a variety of ages, and either persist or not. So the age at assessment becomes an important variable in the interpretation of findings and should be accounted for in the discussion.

Response: Disagree (edited for clarity)

- The available data are not provided in a way that allows for evaluating deficits occurring at a variety of ages and whether the deficits persist or not. Although some studies provide the results by specific ages, these are mainly high risk-of-bias studies conducted in areas with high fluorosis rates, and the tests were generally conducted in children 8–12 years old. The following text was added to the *Discussion* section as a limitation of the evidence base:

  "The database does not allow for comparison of ages and possible changes at different developmental stages in children to assess if there is a delay in development or if associations persist."

- We have already considered age at test administration in the risk-of-bias evaluation of individual studies in two different ways: (1) whether the test used to measure neurodevelopment or cognition was age-appropriate and (2) when a study included a range of ages, whether age was assessed as a potential confounder (for the reasons noted by [redacted]).

H.22:

[redacted] **Comments:** When discussing the variations in associations by genetic polymorphisms, it would be useful to discuss the function of the gene, especially the function related to neurodevelopment or the developing brain.

Response: Disagree (no change)

- Although information on the possible interaction of fluoride with genetic polymorphisms is an active area of investigation, only two studies were available as of the cutoff date for this systematic review. Our intent was to simply point this out as an emerging area of research rather than speculate about potential mechanisms of fluoride action, which would require much further study and a deeper understanding.

H.23:

[redacted] **Comments:** As indicated above, please be very careful in discussing dose response relationships, especially when these may be non-linear.

Response: Agree (no change)

- We agree that discussion of dose-response relationships should be done carefully, and we re-reviewed all of the dose-response text to address this concern. The Sup02_2022_Prepublication_NTP_Monograph summarizes the findings of the qualitative analysis of children's IQ studies that evaluated lower fluoride exposures without reporting on the evidence for dose response (available in full in the 2019 draft NTP Monograph). The Sup02_2022_Prepublication_NTP_Monograph refers the reader to the revised meta-analysis document as it provides a quantitative assessment of dose response to further inform this discussion.

b) How the overall set of confounders across the body of evidence from children's IQ studies was considered and presented.

H.24:

[redacted] **Comments:** Please see the discussion of confounding above. [redacted] do appreciate Figure 6 which describes the confounders measured in the low risk of bias studies, stratified by rating for confounding. In the three studies in which the RoB rating for confounding was high, however, it appears that such confounding may influence the results to some degree. It would be useful to have an assessment of the direction and magnitude of bias introduced by not clearly defining and controlling for key confounders, even if that discussion is somewhat speculative.

Response: Agree (no change)

- An assessment of the potential magnitude and direction of bias in the low risk-of-bias studies, as requested by [redacted], was included in *Appendix E* in the Sup02_2022_Prepublication_NTP_Monograph (previously *Appendix 4* of the Sup03_2021_draft_NTP_Monograph, the version of the monograph reviewed by [redacted]).

c) How the confidence rating in the body of evidence was developed and supported.

H.25:

[redacted] **Comments:** In general, the confidence rating in the body of evidence for this outcome is supported. However, several concerns necessitate a refinement of this confidence rating.

[redacted] agree with the prior review in that conclusions can only be made above the WHO drinking water limit for fluoride. It seems as though there is a lack of dose response curve estimation for lower levels of exposure, so an inference cannot be made over the entire range of exposure. Indeed, it is this lower dose range that is of interest for the US population.

Response: Agree (change made)

- As [redacted] notes, earlier versions of this monograph examined the evidence for dose response across the range of exposures represented in the human body of evidence, both from a qualitative and quantitative perspective. The current monograph intentionally does not dwell on this question, as the conclusions from individual included studies about dose response for cognitive neurodevelopmental associations at the lower fluoride exposure levels are somewhat conflicting. The uncertainty of the evidence at these lower levels is cited as one of the limitations of the evidence base. Given that the revised meta-analysis specifically addresses this question and incorporates newer literature, we have decided to revise these considerations in the Sup02_2022_Prepublication_NTP_Monograph to focus on the data on which we base our confidence statement, and to acknowledge the need for further studies at lower exposure levels. The following text has been added to the abstract and summary:

  "This review finds, with moderate confidence, that higher fluoride exposure (e.g., represented by populations whose total fluoride exposure approximates or exceeds the World Health Organization Guidelines for Drinking-water Quality of 1.5 mg/L of fluoride) is consistently associated with lower IQ in children. More studies are needed to fully understand the potential for lower fluoride exposure to affect children's IQ."

H.26:

[redacted] **Comments:** Because the urine and serum biomarkers of fluoride represent relatively recent exposure, it is difficult to infer that the associations are from cumulative exposure without laying out the assumptions, i.e. long term residential history, similar habits of toothpaste use, etc.

Response: Agree (change made)

- Text was added to address the best measures for assessing long-term fluoride exposure (see quote below). Although urine and serum reflect recent exposures, they represent total fluoride exposure. The indicators and assumptions for long-term exposure in the cross-sectional studies are laid out in the *Overall Findings* section for IQ in children and the results are described in *Results by Study Design – Cross-sectional Studies* section where we address the assumptions for prior exposure, one of the factors that we considered in establishing the confidence level as moderate.

  "There is general consensus that the best measures of long-term fluoride exposure are bone and/or tooth measurements, and other than measures of dental fluorosis, these were not performed in any of the studies reviewed in this document."

H.27:

4. ***NTP concludes a rating of moderate confidence in the body of evidence for lower IQ in children associated with fluoride exposure***.

☐ Agree
X Agree in principle with the exception(s) listed below:
Please see point a above. The exception would be that there is low confidence of the association for levels of exposure in the lower dose range.
☐ Do not agree because:

Response: Agree (change made)

- We provided our response to this point above.

II. **Fluoride exposure and non-IQ neurodevelopmental or cognitive effects in children**

H.28:

1. Comment on whether the approach described in the methods to search for and select human studies on neurodevelopmental or cognitive function effects associated with fluoride exposure was appropriate for evaluating potential effects of fluoride exposure on non-IQ neurodevelopmental or cognitive effects in children.

[redacted] **Comments:** Please see [redacted] comments in section III.1 above. The search terms used are encompassing of neurodevelopmental outcomes in children.

Response: No change requested

- No response necessary.

H.29:

2. Comment on whether the approach used to assess risk of bias for studies in children on non-IQ neurodevelopmental or cognitive effects was clearly described and appropriately applied.

[redacted] **Comments:** Please see [redacted] points in III 2 above. Further, for these outcomes, one would definitely need to stratify the results based on age of the child, as some of these skills develop differently. For example, children age 6-8 years are very different from neonates. Also, please note that the thinking regarding assessment of confounding would be outcome specific as some variables, e.g. SES, may not be applicable to some skills.

Response: Agree (change made)

- See our previous response to [redacted] comment on selection bias. In short, we have added clarifying language in the *Methods* section to indicate that selection bias was consider in determining the overall risk-of-bias status of each study (response above includes a quote of the revised monograph text).
- Furthermore, we agree that confounding is outcome-specific, but SES along with sex and age were identified as key covariates for all studies. This means that SES would need to be considered in any human study of fluoride and cognitive neurodevelopmental health effects; however, if there was reason to believe that SES (or age or sex) was not a potential confounder or risk-of-bias concern for a given study, then that would have been taken into consideration when determining the risk-of-bias rating for confounding. The risk-of-bias rating rationale would have described the reason that SES was not considered a concern for a particular study.

3. Comment on assessment of the human studies with regard to:

H.30:

a) How findings from individual "low risk of bias" studies were interpreted.

[redacted] **Comments:** Many of [redacted] comments in section III 3a are also applicable here.

As noted above, please note that the assessment of confounding needs to be outcome (and likely age) specific. For example, measures of socioeconomic status may not be confounders for outcomes measured in neonates (the Li study did not control for anything) but may be proxy measures for variables such as maternal smoking, that was not measured or controlled and which could be a confounder.

Response: Agree (change made)

- See section A1 where we addressed this comment when it was previously raised.

H.31:

[redacted] **Comments:** For the studies that measured multiple outcomes, there would need to be some adjustment for multiple testing, using either a conservative Bonferroni correction or some other method. This is particularly important here as the behavioral outcomes, for example, are correlated.

Response: Disagree (no change)

- *Appendix E* in the Sup02_2022_Prepublication_NTP_Monograph (previously *Appendix 4* of the Sup03_2021_draft_NTP_Monograph) includes considerations of adjustment methods (including use of the Benjamini–Hochberg false discovery rate) when information was provided by the study authors. We disagree that adjustment for multiple testing is necessary in our risk-of-bias assessment where studies are estimating an effect of exposure on an outcome. Adjustment for multiple comparisons is only necessary when a study is doing strict hypothesis testing (Rothman 1990).

H.32:

[redacted] **Comments:** (minor) Please note that often the GCI on the MSCA is considered a measure of IQ, so perhaps the study of Bashash et al (2017) could be considered in the IQ studies.

Response: Disagree (no change)

- The MSCA measures developmental abilities in children using tasks that assess verbal, quantitative, perceptual, memory and motor skills. Children can earn an IQ-like score (the General Cognitive Index; GCI) based on summed performance across tasks. We agree that the GCI can be considered as a measure of IQ; however, we considered it appropriate to categorize this test with other tests of cognitive function in the *Other Neurodevelopmental or Cognitive Effects in Children* section. Moreover, the *IQ in Children* section includes Bashash et al. (2017) for its results from the Wechsler Abbreviated Scale of Intelligence, which is typically considered an IQ test. Categorizing the MSCA results in the *Other Neurodevelopmental or Cognitive Effects in Children* section allowed us to avoid double-counting the Bashash et al. (2017) study in the *IQ in Children* section.

- Note that adding GCI to the *IQ in Children* section rather than the section on other neurodevelopmental outcomes may add to the strength of evidence, but it would not change the moderate confidence determination in the monograph. Furthermore, the revised meta-analysis includes sensitivity analyses with GCI scores from Bashash et al. (2017) and a second study that reported findings from the GCI portion of the MSCA.

H.33:

[redacted] **Comments:** Some of the associations are really quite large, e.g. adjusted betas of -19 in the study of Valedez Jimenez et al 2017, especially for the Bayley Scale. Such associations are either suspect or are not adjusted for the concentration of fluoride appropriately (maybe it is a log unit change). This needs to be clarified.

Response: Agree (change made)

- We have clarified in the tables and text that the associations are per log10-mg/L increase in fluoride exposure. The revised text in *Results in Infants* section of the Sup02_2022_Prepublication_NTP_Monograph reads as follows:

  "In infants 3 to 15 months of age, the Mental Development Index (MDI)—which measures functions including hand-eye coordination, manipulation, understanding of object relations, imitation, and early language development—was significantly inversely associated with maternal urinary fluoride in both the first and second trimesters (adjusted ß̂s per log10-mg/L increase = −19.05 with standard error of 8.9 for first trimester [p-value = 0.04] and −19.34 with standard error of 7.46 for second trimester [p-value = 0.013]) (Valdez Jimenez et al. 2017)."

H.34:

[redacted] **Comments:** Please clarify what a construction task is (page 56). Do you mean a fine motor copy task?

Response: Agree (change made)

- We revised the text to characterize the task more accurately as a visuoconstructional score from the Rey-Osterrieth Complex Figure Test. The revised sentence reads as follows:

  "Another study using visuoconstructional and memory scores from the Rey-Osterrieth Complex Figure Test in children 6–11 years old observed significantly lower scores with increasing concurrent child single spot urinary fluoride even after adjusting for age (partial correlation coefficients, per log-mg/L increase = −0.29 and −0.27 for copy [p-value <0.001] and immediate recall [p-value <0.001], respectively [CIs not reported]) (Rocha-Amador et al. 2009)."

H.35:

[redacted] **Comments:** Also on page 56 and highlighted in blue: this is unclear. Even though urinary arsenic is not associated with scores on these tasks, it could still very well be a confounder of the relationships between fluoride and the test scores.

Response: Agree (change made)

- As we discuss in *Appendix E* in the Sup02_2022_Prepublication_NTP_Monograph (previously *Appendix 4* of the Sup03_2021_draft_NTP_Monograph), although the model in Rocha-Amador et al. (2009) did not adjust for arsenic, arsenic in the F-As group was not associated with either outcome; therefore, arsenic as a co-exposure is not considered a major concern in this study. We revised text to mention the results adjusted for both fluoride and arsenic, as follows:

  "Another study using visuoconstructional and memory scores from the Rey-Osterrieth Complex Figure Test in children 6–11 years old observed significantly lower scores with increasing concurrent child single spot urinary fluoride even after adjusting for age (partial correlation coefficients, per log-mg/L increase = −0.29 and −0.27 for copy [p-value <0.001] and immediate recall [p-value <0.001], respectively [CIs not reported]) (Rocha-Amador et al. 2009). Although these children were also exposed to arsenic, the presence of arsenic could not explain the changes because, in the area with natural contamination by fluoride and arsenic (F–As), the test scores were not significantly associated with urinary arsenic levels (partial correlation coefficients, per log-mg/L increase = −0.05 and 0.02 for copy and immediate recall, respectively [CIs not reported]). The test scores were only marginally increased from fluoride alone when both fluoride and arsenic were included simultaneously in the model (partial correlation coefficients, per log-mg/L increase = −0.32 and −0.34 for copy and immediate recall, respectively [CIs not reported]) (Rocha-Amador et al. 2009)."

H.36:

[redacted] **Comments:** Also please address the issue that children with behavior problems may be more apt to, for example, drink excessive amounts of water or swallow toothpaste. This would be indicative of reverse causation.

Response: Disagree (no change)

- While polydipsia has been associated with clinical psychoses, we have failed to find reports of excessive consumption of water or toothpaste associated with the types of behaviors addressed in the studies examining fluoride exposure and other cognitive or neurodevelopmental conditions.

H.37:

b) How the confidence rating in the body of evidence was developed and supported.

[redacted] **Comments:** [redacted] fully agree with the low confidence rating for this body of evidence. The issues that [redacted] have highlighted above would only lend more support to the low confidence.

Response: No change requested

- No response necessary.

H.38:

4. ***The NTP concludes a rating of low confidence in the body of evidence for decreases in measures of other neurodevelopmental or cognitive effects in children associated with fluoride exposure***.

X Agree
☐ Agree in principle with the exception(s) listed below:
☐ Do not agree because:

Response: No change requested

- [redacted] agreed with the low confidence rating.

III. **Fluoride exposure and cognitive effects in adults**

H.39:

1. Comment on whether the approach described in the methods to search for and select human studies on neurodevelopmental or cognitive function effects associated with fluoride exposure was appropriate for evaluating potential effects of fluoride exposure on cognitive effects in adults.

[redacted] **Comments:** Please see the comments above.

Response: No response necessary

- Comments were addressed where previously made by [redacted].

H.40:

2. Comment on whether the approach used to assess risk of bias for studies in adults on cognitive effects was clearly described and appropriately applied.

[redacted] **Comments:** Please see the comments above. [redacted] one additional comment here is that the results from China (Li et al 2016) perhaps indicate that the critical time of exposure is at earlier ages, since the exposure was residing in low and high fluorosis-endemic areas of China.

Response: Agree (no change)

- While we agree with [redacted] that earlier exposures could be an important factor in this study, there is insufficient information provided in the study to assess critical timing of exposure for cognitive impairments in adults.

3. Comment on assessment of the human studies with regard to:

H.41:

a) How findings from individual studies were interpreted.

[redacted] **Comments:** The studies were interpreted appropriately.

Response: No change requested

- No response necessary.

H.42:

b) How the confidence rating in the body of evidence was developed and supported.

[redacted] **Comments:** [redacted] fully support the confidence rating of low for this body of evidence.

Response: No change requested

- No response necessary.

H.43:

4. ***The NTP concludes a rating of low confidence in the body of evidence for changes in cognitive effects in adults with fluoride exposure.***

X Agree
☐ Agree in principle with the exception(s) listed below:
☐ Do not agree because:

Response: No change requested

- [redacted] agreed with the low confidence rating.

C. Studies in non-human animals

H.44:

The NTP agrees with the comments of the NASEM committee (NASEM 2020, 2021) concerning the overall poor quality of the experimental animal database on fluoride exposure and neurodevelopmental effects, with many studies suffering from major reporting deficiencies. As indicated above, the monograph focuses on the large human epidemiology database because it directly addresses the question of whether fluoride affects human neurodevelopment. Therefore, based on the recommendations of the NASEM committee, the experimental animal section and risk of bias details have been removed from this monograph and ***the NTP concludes that the scientific evidence from experimental animal data are inadequate to inform whether fluoride exposure is associated with cognitive effects (including cognitive neurodevelopmental effects) in humans***.

X Agree
☐ Agree in principle with the exception(s) listed below:
☐ Do not agree because:

Response: No change requested

- [redacted] agreed with the inadequate designation.

References

Cowell WJ, Wright RJ. 2017. Sex-Specific Effects of Combined Exposure to Chemical and Non-chemical Stressors on Neuroendocrine Development: a Review of Recent Findings and Putative Mechanisms. Curr Envir Health Rpt. 4: 415–425. https://doi.org/10.1007/s40572-017-0165-9

D'Amico D, Parrott MD, Greenwood CE, Ferland G, Gaudreau P, Belleville S, Laurin D, Anderson ND, Kergoat MJ, Morais JA, Presse N, Fiocco AJ. 2020. Sex differences in the relationship between dietary pattern adherence and cognitive function among older adults: findings from the NuAge study. Nutr J. 19(58) . https://doi.org/10.1186/s12937-020-00575-3

Gochfeld M. 2017. Sex Differences in Human and Animal Toxicology. Toxicol Pathol. 45(1):172-189. doi: 10.1177/0192623316677327. PMID: 27895264; PMCID: PMC5371029.

Keller KL, Kling SMR, Fuchs B, Pearce AL, Reigh NA, Masterson T, Hickok K. 2019. A Biopsychosocial Model of Sex Differences in Children's Eating Behaviors. Nutrients. 11(3): 682. https://doi.org/10.3390/nu11030682

Lash TL, VanderWeele TJ, Haneuse S, Rothman KJ. Modern Epidemiology, 4th edition. Wolters Kluwer, 2021.

Rothman KJ. 1990. No adjustments are needed for multiple comparisons. Epidemiology. 1(1):43-6. PMID: 2081237

Thomas DB, Basu N, Martinez-Mier EA, Sánchez BN, Zhang Z, Liu Y, Parajuli RP, Peterson K, Mercado-Garcia A, Bashash M, Hernández-Avila M, Hu H, Téllex-Rojo MM. 2016. Urinary and plasma fluoride levels in pregnant women from Mexico City. Env Res. 150: 489-495.

In November 2021, [redacted] received: 1) the 2021 *Draft NTP Monograph on the State of the Science concerning Fluoride Exposure and Neurodevelopmental and Cognitive Health Effects: A Systematic Review,* 2) a copy of the NASEM Committee's comments on the 2020 draft NTP Monograph with NIEHS/DNTP responses (draft version of Sup01_Monograph), and 3) the [redacted] instructions. The instructions consisted of a preface, charge, instructions for the review, and a series of specific peer-review questions grouped by the following three topics: General Comments, Human Studies, and Studies in Non-Human Animals.

[redacted] were asked to provide their substantive scientific and technical comments and suggestions within the [redacted] form. In addition, they were asked whether they "Agree", "Agree in principle", or "Do not agree" with each NTP conclusion on confidence in a body of evidence.

The [redacted] instructions and specific peer-review questions are reproduced in the pages that follow in black text. [redacted] comments and responses to each question are also provided in black text starting with the words "[redacted] **comments**" in bold font. The NIEHS/DNTP responses have been inserted in blue text following each of the comments beginning with the word "**Response**" in bold font. Formatting has been applied to aid in reading.

The prepublication 2022 NTP Monograph reflects changes made after consideration of the comments from the [redacted] along with all other input received through April of 2022. The prepublication 2022 NTP Monograph was subsequently sent to [redacted] for additional comments. A revised "track changes" version of the monograph was developed in September 2022 titled the "DocMon_Track_Changes_2022_NTP_Monograph." The following bullets describe how edits are documented in the track changes version of the monograph in response to [redacted] comments and [redacted] comments:

- [redacted] For comments related to DocG_Monograph, DocH_Monograph, DocI_Monograph, DocJ_Monograph, and DocK_Monograph:
    - Edits are marked with a comment bubble in the DocMon_Track_Changes_2022_NTP_Monograph that identifies the text in question and briefly describes any revisions.
    - The comment bubble contains the exact text of the [redacted] Comment.
    - The comment bubble also provides a reference to the specific response to comments document with the detailed NIEHS/DNTP response (e.g., comments made in response to this [redacted] would be marked "*see DocI_Monograph for detailed response*").
- [redacted] For comments DocA1_Monograph, DocA2_Monograph, DocB1_Monograph; DocB2_Monograph, and DocC_Monograph through DocF_Monograph:
    - Edits are marked in track changes format in the DocMon_Track_Changes_2022_NTP_Monograph.
    - A comment bubble has been added to the text in question containing the exact text of the [redacted] Comment.
    - The comment bubble also provides a reference to the specific response to comments document with the detailed NIEHS/DNTP response.

Preliminary comments on the draft NTP monograph prepared by the peer review [redacted] are noted below.
These preliminary comments are not binding and should not be construed to represent NTP determination or policy.

National Toxicology Program
NTP Monograph Letter Peer-Review Panel
Draft NTP Monograph on the State of the Science Concerning Fluoride Exposure and Neurodevelopmental and Cognitive Health Effects: A Systematic Review

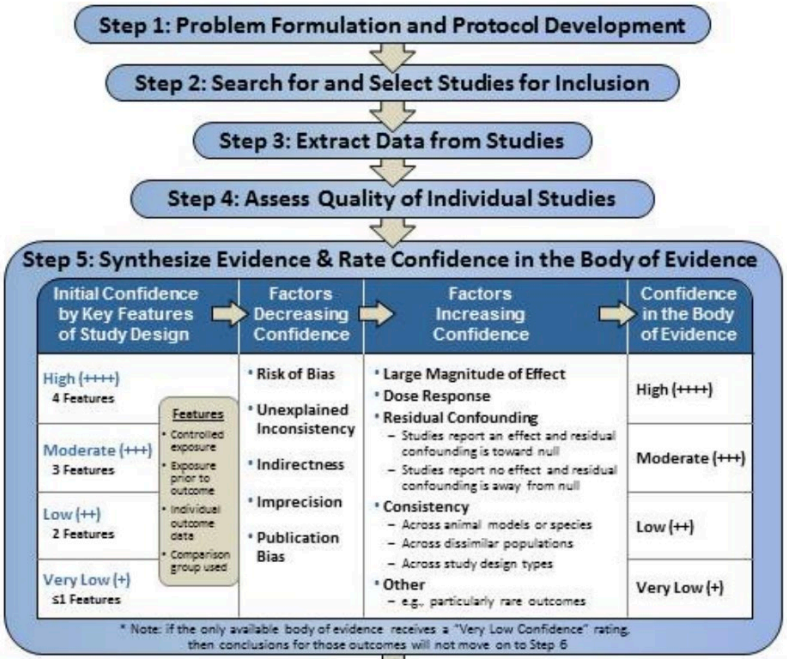
National Institute of Environmental Health Sciences
Research Triangle Park, NC

January 18, 2022

Fluoride State of the Science Document Review Form

[redacted]

Preface:
The objective of this evaluation was to conduct a systematic review of the published literature regarding the potential for exposure to fluoride to affect neurodevelopment and cognition in humans. The evaluation presented in the draft *NTP Monograph on the State of the Science Concerning Fluoride Exposure and Neurodevelopmental and Cognitive Health Effects* represents a comprehensive and current assessment. The methods used are from the Handbook for Conducting a Literature-Based Health Assessment Using OHAT Approach for Systematic Review and Evidence Integration, which presents a seven-step framework for systematic review and evidence integration. Please note: this evaluation stops at step 5 of the systematic review process and does not proceed to step 6 to translate the confidence rating for the body of evidence into a level of evidence for health effects (see Figure 2 from the handbook).



Charge:

(1) Comment on the technical accuracy and whether the draft *NTP Monograph on State of the Science Concerning Fluoride Exposure and Neurodevelopmental and Cognitive Health Effects* is clearly stated, and objectively presented.

(2) Determine whether the scientific evidence supports the NTP's confidence ratings for the bodies of evidence regarding neurodevelopmental and cognitive health effects associated with exposure to fluoride.

Instructions for Review:
All materials for this review are available in the Electronic Council Book (ECB). You will receive the specific URL and a password for accessing the ECB.

This evaluation identified 159 human studies relevant for assessing neurological health effects of exposure to fluoride; however, many studies included only secondary outcomes (e.g., 55 studies of thyroid hormones that were investigated as a potential mechanism). The scientific evidence in children and adults was evaluated separately to address potential differences in the health impact of fluoride exposure during development versus adulthood. Several studies evaluated learning and memory (n = 8 studies) or other cognitive developmental effects (e.g., total neurobehavioral scores and total mental capacity index in children, cognitive impairment in adults; n = 14 studies). Sixty-six human studies investigated IQ in children. Nineteen of the 66 IQ studies were determined to have low potential for bias and therefore, were categorized as "low risk of bias". Please give special attention to our assessment of these 19 studies.

- The 19 studies are available as PDFs and organized alphabetically in a folder on the ECB.
- All other studies are provided in the Health Assessment Workspace Collaborative, or HAWC database under the "studies list" tab, also organized alphabetically. You will also be provided a username and password for HAWC that will give you [redacted] permissions to access the PDFs in HAWC along with visualizations and other study information for this project at the following link (https://hawcproject.org/study/assessment/405/).

Please provide your substantive scientific and technical comments and suggestions within this [redacted] form. Identify and provide the rationale or scientific support for proposed changes or suggestions where possible.

If necessary, you can also provide additional editorial comments and recommendations for improving the report outside your specific charge questions (this form) within the draft report itself. Please note that only those comments included on the [redacted] form will be considered part of NTP's peer review report.

A. General Comments

I. Please comment on whether the scientific information presented in the draft monograph, including presentation of data in tables and figures, is technically correct, and clearly and objectively presented. Please suggest any improvements.

I.1: [redacted] **Comments:** The scientific information presented appears technically correct and objectively presented. A few suggestions are noted below to improve clarity. The background section could be reorganized for clarity and flow. It might be beneficial to begin the abstract and background with the pervasive use of fluoride in drinking water followed by a brief statement of the benefits. The benefits of fluoride in water has not been articulated. The benefits only need a sentence or two. The background appears to be more of a justification for the report rather than a true background of the evidence leading to the study/report.

Response: Agree (change made)

- We agree with the suggestion to reorganize the *Introduction* section. In response, we have moved text from the first paragraph of the *Introduction* closer to the end of the section. As such, the uses of and exposure sources to fluoride are now the first topics covered. We briefly discuss the benefits of fluoride but have not emphasized it or mentioned it in the *Objective* or *Specific Aims* as this topic is not the focus of the monograph. There is also no attempt in the monograph to compare hazards with benefits.

I.2: [redacted] **Comments:** Might consider beginning the background with the PHS recommendations.

Response: Disagree (no change)

- We have chosen not to highlight fluoridation of drinking water as the monograph focuses on the question of whether fluoride from all sources can affect neurodevelopmental outcomes and is written to avoid giving the mistaken impression that this systematic review is focused only on drinking water. While drinking water provides the majority of fluoride exposure in many of the studies, total exposure can vary widely even in optimally fluoridated areas based on personal habits in the use of dental products and consumption of beverages such as black tea that can contain fluoride.

I.3: [redacted] **Comments:** The abstract and background also need to be consistent in terms of presentation of human and animal studies. This consistent ordering of the studies (human, animal, mechanistic – for example) descriptions would improve flow and readability. Given the final conclusion of the animal studies section, is it possible to omit the non-human studies component?

Response: Disagree (no change)

- The ordering of topics in the various monograph sections has been determined after considering options and feedback from all reviewers. As a whole, we consider the current organization of topics in the monograph as appropriate to best support the ultimate rating of moderate confidence for effects of fluoride on children's IQ.

- With respect to the inclusion of the animal section, we consider it to be a valuable addition to the monograph even though the details have been largely relegated to earlier drafts that were reviewed by the NASEM Committee. The animal section provides an update to the 2016 NTP animal systematic review, identifies the studies that were conducted by the DNTP to address deficiencies in the 2016 NTP animal systematic review, and reiterates the lack of consistent evidence from this body of literature to support human findings.

I.4: [redacted] **Comments:** The term 'neurodevelopment' includes cognition, so if you would like to focus on cognition, you could simply state 'neurocognition.' Neurodevelopment is typically used as an umbrella term for all neurodevelopment, including cognition and motor function.

Response: Disagree (no change)

- We chose to use the terms "neurodevelopment" and "cognition" because the children's literature includes studies on both cognition and behavior.

I.5: [redacted] **Comments:** As currently written, the objective is not clearly stated. *Potential rewrite*: The objective of this report to assess the relationship between fluoride exposure and neurocognitive effects in humans and animals. To accomplish this objective, a systematic review of the literature was undertaken and mechanistic data was considered.

Response: Disagree (no change)

- We understand that the suggested refined objective may better reflect the ultimate emphasis of the monograph based on the data that were found; however, the systematic review was more comprehensive in scope and we consider it to be better represented by the current wording. Furthermore, the current wording is consistent with the published protocol.

I.6: [redacted] **Comments:** Why is the meta-analysis not included?

Response: No change requested

- The decision to pursue a narrative evidence synthesis rather than a meta-analysis was made while preparing the 2019 draft NTP Monograph because our goal of generating a document to support a hazard assessment did not require a quantitative estimate of hazard (e.g., numeric estimate of IQ points lost per mg F/L of drinking water or urine). However, as outlined in a new table that provides a timeline of draft monographs and important decision points (Table B-1 in *Appendix B* of the Sup02_2022_Prepublication_NTP_Monograph), comments received from the NASEM Committee that reviewed the 2019 draft NTP Monograph (NTP, 2019) recommended that we perform a meta-analysis and indicated that the outcome would be critical to reaching a hazard conclusion. We therefore performed a meta-analysis, which included a *dose-response meta-analysis*, and included the meta-analysis in the revised Sup04_2020_draft_NTP_Monograph (NTP, 2020). In its review of that Sup04_2020_draft_NTP_Monograph, the NASEM Committee again stated that the

document fell short of supporting our hazard call, and the Committee also had additional recommendations to improve the meta-analysis.

- After reflecting on the NASEM Committee comments on the Sup04_2020_draft_NTP_Monograph, we decided to remove the evidence integration step from the systematic review of the literature and instead issue the report (after further independent peer review) as a document outlining the state of the science on the association between fluoride exposure and deficits in neurodevelopment and cognition. We then decided to revise and submit the meta-analysis as a separate peer-reviewed publication because it was no longer required to support the "presumed" hazard call which was reached in the 2019 monograph and Sup04_2020_draft_NTP_Monograph. The meta-analysis, including the *dose-response meta-analysis*, was performed only on the studies addressing fluoride exposure in relation to deficits in children's IQ. The separate meta-analysis considers comments from the NASEM Committee in its revisions.

I.7: [redacted] **Comments:** Why limit to thyroid function as an effect/mechanism?

Response: No change requested

- Hypothyroidism and prematurity are among the few well-established risk factors for delayed or deficient neurodevelopment in children (for example, see review by Prezioso et al. [2018]). Many of the better-quality human studies controlled for gestational age at birth, and there is a growing body of literature on the interaction between fluoride exposure and low iodine levels in relation to children's IQ. This is why iodine was considered an important co-exposure in our risk-of-bias assessments (e.g., Goodman et al., 2022).

I.8: [redacted] **Comments:** Figure 1: [redacted] don't see where confounding or co-exposure is included.

Response: No change requested

- Confounding and co-exposures are part of the risk-of-bias assessment so are not individually listed in Figure 1. Details on confounding and co-exposures first appear in the *Quality Assessment of Individual Studies* section.

II. Please identify any information that should be added or deleted.

I.9: [redacted] **Comments:** Thyroid function isn't mentioned until the specific aims. It should be included in background along with other possible mechanisms, if known. It is unclear why thyroid function is being evaluated as the only mechanistic pathway. A figure or illustration depicting the theoretical pathway would be helpful.

Response: Agree (change made)

- We have added a footnote to the *Introduction* section of the Sup02_2022_Prepublication_NTP_Monograph to explain the focus on potential thyroid effects. The footnote reads:

  "The current review has evaluated the fluoride literature with an eye toward potential thyroid effects because a large literature base has accumulated

examining the interaction of fluoride with iodine uptake by the thyroid gland and consequential effects on synthesis of thyroid hormones, which are recognized to play significant roles in neurodevelopment in utero and during early childhood. This literature, along with a detailed proposed mechanism of action, was recently reviewed by Waugh (2019)."

I.10: [redacted] **Comments:** A brief discussion of serum fluoride needs to be included – similar to the urinary fluoride description (page 16).

Response: Agree (change made)

- We included a statement concerning serum fluoride in the *Exposure* section of the Sup02_2022_Prepublication_NTP_Monograph to explain why serum fluoride levels are considered a poor measure of long-term fluoride exposure. The new statement reads, *"Fluoride ion is rapidly absorbed from the gastrointestinal tract and is rapidly cleared from serum by distribution into calcified tissues and urinary excretion (IPCS 2002)."*

I.11: [redacted] **Comments:** Table 6 could include the following: 1) statistical methods; 2) confounders, particularly exposure to other known neurotoxicants, and how they were measured; 3) might rename 'Assessment timing' to age of participants or just combine the information with the location/subject's column

Response: Agree (edited for clarity)

- Although additional information could be added to Table 6, the information requested by [redacted] is already in *Appendix E* in the Sup02_2022_Prepublication_NTP_Monograph (previously *Appendix 4* of the Sup03_2021_draft_NTP_Monograph) for all the studies presented in Tables 6 and 7. Therefore, to address this comment, text has been added to the paragraphs that introduce Tables 6 and 7 to point to *Appendix E* for this additional information by study. We considered [redacted] suggestion to rename the 'assessment timing' column to 'age of participants'; however, we have retained the current column header as the information provided in this column is the age of participants at which outcome was assessed. The current header is the most concise way to communicate this.

B. Human studies

I. **Fluoride exposure and children's IQ**

1. Comment on whether the approach described in the methods to search for and select human studies on neurodevelopmental or cognitive function effects associated with fluoride exposure was appropriate for evaluating potential effects of fluoride exposure on measures of IQ in children.

I.12: [redacted] **Comments:** The approach described was appropriate. It is not clear when child and adult studies were separated from the main search or if each search was done independently (child and adult). It appears that it was only 'human studies.' [redacted] wonder how the search would change, if at all, if search terms for the target population was included? It should be clearly stated how and each population (child and adult) were separated.

Response: No change requested

- All life stages were relevant to the assessment according to our PECO statement (Population: "*Humans without restriction as to age or sex, geographic location, or life stage at exposure or outcome assessment*"). The same search was designed to identify child and adult studies, and the search did not include terms related to life stage (see response under B.III.1 for further explanation as to why this approach is thought to effectively capture relevant studies from all life stages). Although the process for deciding which bodies of evidence to synthesize and whether to separate groups of studies by health effects or age was described in the protocol, specific decisions were made based on the results of the literature search and selection. The groupings by age and the separation of child and adult studies were done after study selection and during the initial evaluation of the studies to determine what information was available. The initial evaluation sorted studies into children and adult studies to see if there was enough information to group the literature in a similar way as had been done for the 2016 NTP animal systematic review. As there was determined to be sufficient data, the decision was made to evaluate children separately from adults. The monograph explains that children and adults were evaluated separately to address potential differences in the health impact of fluoride exposure during development versus adulthood.

I.13: [redacted] **Comments:** The rationale for date selection needs to be more clearly articulated. The specific dates are included in the appendix, perhaps they could be included in the main text for clarity in the methods.

Response: Agree (change made)

- In an effort to provide further clarity on the progression of this multiyear assessment, we have developed a new table (Table B-1 in *Appendix B*) that provides a timeline of key activities contributing to the Sup02_2022_Prepublication_NTP_Monograph, including information relevant to the timing of the literature searches. For example, the expanded literature search to include non-English databases took place in May 2020 in response to the NASEM Committee's peer review report on the 2019 draft NTP Monograph.

2. Comment on whether the approach used to assess risk of bias was clearly described and appropriately applied to the set of studies designated as "low risk of bias."

I.14: [redacted] **Comments:** The approach to assess risk of bias was clearly described. A brief discussion is needed about critical confounders, including a biological exposure measure for tobacco use or exposure, such as serum cotinine, and parental IQ for the child studies If there are unique confounders for child and adult studies, this needs to be articulated. It currently appears that there are no unique confounders for child and adult.

Response: Agree (change made)

- [redacted] is correct that there are no unique confounders for children and adults. As noted in the monograph, the potential confounders that were considered important for all studies, populations, and outcomes were age, sex, and socioeconomic status regardless of whether the subjects were children or

adults. However, we realize that, as written in the Sup03_2021_draft_NTP_Monograph, it may be interpreted that age and sex confounders were only applied to children. Text has been updated in the Sup02_2022_Prepublication_NTP_Monograph to clarify that age and sex are important potential confounders regardless of life stage. For all other potential confounders considered in the evaluation, their importance was dependent on the study population and outcome being evaluated, and no specific potential confounder was unique to either children or adults.

- Smoking was considered an important confounder in adult studies evaluating Alzheimer's disease, but smoking was only considered a major concern if there were reasons to believe that it would be a source of bias.
- We agree with [redacted] that parental IQ is an important potential confounder in the studies of children. Because parental IQ, educational attainment, and other measures of socioeconomic status (SES) all likely share a common cause, the latter two covariates were considered to be potential proxy measures of parental IQ. Therefore, parental IQ was considered indirectly addressed if a study accounted for parental education and/or socioeconomic status. For clarification, we added a footnote to Figure 6 that lists all measures related to SES that were considered in the low risk-of-bias IQ-in-children studies.

3. Comment on assessment of the human studies with regard to:

a) How findings from individual studies designated as "low risk of bias" were interpreted.

I.15: [redacted] **Comments:** Findings from low-risk studies were interpreted well. They were interpreted objectively.

Response: No change requested

- No response necessary.

b) How the overall set of confounders across the body of evidence from children's IQ studies was considered and presented.

I.16: [redacted] **Comments:** The overall set of confounders has been thoughtfully considered and presented. Figure 6 is very comprehensive. Are there any unique confounders for the age groups (child and adult)?

Response: Agree (change made)

- This repeats a more extensive comment made previously on question B.I.2; see above for a more detailed response.

c) How the confidence rating in the body of evidence was developed and supported.

I.17:

[redacted] **Comments:** NTP used the GRADE system for rating confidence in the body of evidence. GRADE is a published method for reaching confidence. NTP also elaborated on factors they considered for potential downgrading and upgrading of research. Figure 1 outlines the process. It might be beneficial to include a 'scale' of factors that result in a score of high, moderate, low or very low in Figure 1, if applicable.

Response: Disagree (no change)

- As [redacted] points out, Figure 1 outlines the GRADE-based method, and the accompanying text elaborates on the factors considered for potential upgrading or downgrading of the confidence in the bodies of evidence. Given the complexity of the possible upgrade and downgrade decisions across the nine factors, we outline the process in Figure 1 rather than trying to predict all the combinations of factors that might result in different ratings of high, moderate, or low.

I.18:

4. ***NTP concludes a rating of moderate confidence in the body of evidence for lower IQ in children associated with fluoride exposure***.

☒ Agree
☐ Agree in principle with the exception(s) listed below:
☐ Do not agree because:

Response: No change requested

- Reviewer agreed with the moderate confidence rating.

II. **Fluoride exposure and non-IQ neurodevelopmental or cognitive effects in children**

1. Comment on whether the approach described in the methods to search for and select human studies on neurodevelopmental or cognitive function effects associated with fluoride exposure was appropriate for evaluating potential effects of fluoride exposure on non-IQ neurodevelopmental or cognitive effects in children.

I.19: [redacted] **Comments:** The approach described to search and select human studies on neurodevelopmental or cognitive function effects potentially associated with fluoride exposure was well-designed and executed. It should be stated if there were any literature review or data extraction methods for child and adult populations.

Response: Agree (no change)

- We agree that if literature review or data extraction methods had differed for child and adult populations, they would need to be clearly stated; however, in the case of this systematic review, the methods were not different. The systematic review protocol and monograph thoroughly describe the methods for screening (literature review) and data extraction and neither document indicates that methods would differ for children and adults. Study selection and data extraction methods were applied consistently across studies of both child and adult populations. Table 2 of the systematic review provides the inclusion and exclusion criteria used to determine study eligibility and states that there are no restrictions on age or life stage at exposure or outcome assessment, while not drawing any distinctions between child and adult studies. *Appendix 2* of the systematic review protocol lists data extraction elements and also does not draw any distinctions for studies in children versus adults.

I.20: [redacted] **Comments:** Page 13: Should consider adding team member initials to their roles in the review.

Response: Agree (change made)

- A front matter section titled *About This Review* has been added to the Sup02_2022_Prepublication_NTP_Monograph that lists the names of all team members along with a description of tasks to which they contributed (e.g., conducted literature screening, conducted data extraction).

I.21: [redacted] **Comments:** Page 13: there is a statement about studies 'evaluating only goiters or thyroid size were not extracted." If so, shouldn't they be part of the exclusion criteria? Similarly, data was not extracted from in vitro studies. This clarification is needed only because it appears that this report includes methods on data extraction for the meta-analysis that is in progress. For a reader, this description isn't necessary to understand the current report, but understand if these methods are needed.

Response: Disagree (edited for clarity)

- We have taken steps to increase clarity in the Sup02_2022_Prepublication_NTP_Monograph regarding the exclusion of topics for full evaluation. For example, details have been added to the *Data Extraction* methods discussion to further clarify why data on specific endpoints were not considered informative to the systematic review and did not undergo full data extraction or study quality evaluation (see below).

  "Data for primary and secondary outcomes, as well as thyroid hormone level data, were extracted from human studies. Studies evaluating only goiters or thyroid size were not extracted because they do not provide specific information on thyroid hormone levels that would inform whether a thyroid-mediated mechanism was involved in fluoride-associated changes in neurodevelopment."

  "Thyroid data were not extracted for animal studies due to inconsistency in the available data in humans."

  "In vitro studies were evaluated, although data were not extracted from these studies as none of the findings were considered informative with respect to biological plausibility."

- Note that the decision not to extract data on goiters was reached after studies went through the study selection process (where we apply inclusion and exclusion criteria to studies identified from the literature search). When this happens, it is standard practice to explain the reasoning in the systematic review methods, not to amend the protocol with this level of detail.

- The decision on thyroid data was reached by technical experts during the review because changes in thyroid size would not inform whether a thyroid-mediated mechanism was involved in fluoride-associated changes in neurodevelopment. The protocol did not include a level of detail on thyroid-related studies to specify preferred or less informative thyroid-related data. [redacted] makes a valid point that, in hindsight, the protocol could have specified that studies only reporting thyroid size or goiters would be excluded. Similarly, the consideration of mechanistic studies followed a tiered or phased approach to identify pockets of data that might support critical analysis with preference given to fluoride exposures of 20 ppm or less (deemed by technical experts to be most relevant to human exposures) and also to identify commonly reported thyroid-mediated mechanisms. The decision was also reached by technical experts during the

review that full data extraction of in vitro studies was not necessary to assess the biological plausibility of the human and animal results.

2. Comment on whether the approach used to assess risk of bias for studies in children on non-IQ neurodevelopmental or cognitive effects was clearly described and appropriately applied.

I.22: [redacted] **Comments:** Add a brief section on serum fluoride levels. Urinary fluoride levels is fully described, but serum has been omitted.

Response: Agree (change made)

- This repeats a more extensive comment made previously on question A.II.; see above for a more detailed response.

I.23: [redacted] **Comments:** One key feature for confidence rating is 'comparison group used.' This needs to be discussed further since fluoride exposure may be pervasive in water supplies. If so, in studies including a comparison group, include the comparison and how it was determined. Cross-sectional studies using biomarkers as continuous variables can be very strong.

Response: Disagree (edited for clarity)

- Tables 6, 7, and 8 already include data on exposure levels in comparison groups. Additionally, *Appendix E* in the Sup02_2022_Prepublication_NTP_Monograph (previously *Appendix 4* of the Sup03_2021_draft_NTP_Monograph) discusses in detail each low risk-of-bias study and indicates when biomarker measures were used.
- The comparisons in the epidemiological studies are between populations that had a range of fluoride exposures that could be compared with similar populations with lower or no fluoride exposures. To further distinguish between the comparison group and the group(s) exposed to higher levels of fluoride, we have added the word "higher" to specify "higher fluoride exposure," as appropriate, in several places throughout the Sup02_2022_Prepublication_NTP_Monograph. For example, we added the word "higher" to the sentence below from the *Results by Study Design – Cross-sectional Study Variations* section.

  "Overall, the cross-sectional studies consistently provide evidence that higher fluoride exposure is associated with lower IQ scores in children."

3. Comment on assessment of the human studies with regard to:

I.24:

a) How findings from individual "low risk of bias" studies were interpreted.

[redacted] **Comments:** Well done!

Response: No change requested

- No response necessary.

b) How the confidence rating in the body of evidence was developed and supported.

I.25: [redacted] **Comments:** Has the OHAT been published? If so, it should be referenced. Since it's a critical tool in this review, it needs to be further described. What other QA tools are available and why weren't they used? Were the Cochrane Review recommendations for assessment of the risk of bias in research studies followed?

Response: Agree (edited for clarity)

- We agree that the OHAT risk-of-bias tool should be referenced, and we have added this reference to both the protocol and the Sup02_2022_Prepublication_NTP_Monograph as https://ntp.niehs.nih.gov/go/riskbias. The risk-of-bias tool was reviewed by an expert panel as part of the development of the OHAT methods and is publicly posted on the NTP web pages.
- We disagree that the tool needs to be further described in the Sup02_2022_Prepublication_NTP_Monograph because it is described in detail in the protocol, which is appropriately referenced in the *Methods* section.
- The OHAT risk-of-bias tool was selected for this systematic review because it uses a parallel approach to assessing study quality across different study designs for both human and animal research, thus enabling synthesis and development of the confidence ratings to meet the objectives. It is the only tool that is designed to assess studies of environmental exposures, studies of varying study designs that were necessary for this systematic review, and studies in both humans and experimental animals. As described in the tool and the protocol for this systematic review, the OHAT risk-of-bias tool is based on Cochrane and AHRQ methods; therefore, the Cochrane Review recommendations for assessment of risk of bias of human studies were followed.

I.26:

4. ***The NTP concludes a rating of low confidence in the body of evidence for decreases in measures of other neurodevelopmental or cognitive effects in children associated with fluoride exposure***.

☒ Agree
☐ Agree in principle with the exception(s) listed below:
☐ Do not agree because:

Response: No change requested

- [redacted] agreed with the low confidence rating.

III. **Fluoride exposure and cognitive effects in adults**

I.27:

1. Comment on whether the approach described in the methods to search for and select human studies on neurodevelopmental or cognitive function effects associated with fluoride exposure was appropriate for evaluating potential effects of fluoride exposure on cognitive effects in adults.

[redacted] **Comments:** Well described – since it appears to be the same for the child studies. Search terms does not include "child," "pediatrics," or "adult," or other terms to

separate out the child and adult studies. When were these terms added or were they added in the search?

Response: No change requested

- The search terms "child", "pediatrics", and "adult" were not included in the literature search. It was unnecessary to include these or other terms related to life stage because relevant studies of all life stages were captured with the existing search strategy. The search strategy included a set of exposure terms (e.g., "fluoride") and a set of health outcome terms (e.g., "neurodevelopment") as detailed in the appendices to the monograph. All life stages were relevant to the assessment according to our PECO statement (Population: "*Humans without restriction as to age or sex, geographic location, or life stage at exposure or outcome assessment*"), and we are confident that all relevant child and adult studies were identified by searching for relevant exposure and outcome terms only (i.e., all fluoride and neurodevelopmental studies would be identified across all life stages). Moreover, we are confident that the absence of search terms related to life stage would not result in missing studies with relevant exposures and relevant outcomes.

I.28:

2. Comment on whether the approach used to assess risk of bias for studies in adults on cognitive effects was clearly described and appropriately applied.

[redacted] **Comments:** Since it is similar to the methods used for child studies.

Response: No change requested

- No response necessary; [redacted] only notes that similar methods were used for studies in children.

3. Comment on assessment of the human studies with regard to:

I.29:

a) How findings from individual studies were interpreted.

[redacted] **Comments:** Not sure of this question – how is it different from the question in the 'child section'? Adult studies were interpreted well.

Response: No change requested

- No response necessary.

I.30:

b) How the confidence rating in the body of evidence was developed and supported.

[redacted] **Comments:** Similar response to the 'child section' above. The confidence in the adult studies was interpreted well.

Response: No change requested

- No response necessary.

I.31:

4. ***The NTP concludes a rating of low confidence in the body of evidence for changes in cognitive effects in adults with fluoride exposure***.

☒ Agree
☐ Agree in principle with the exception(s) listed below:
☐ Do not agree because:

Response: No change requested

- [redacted] agreed with the low confidence rating.

C. Studies in non-human animals

I.32:

The NTP agrees with the comments of the NASEM committee (NASEM 2020, 2021) concerning the overall poor quality of the experimental animal database on fluoride exposure and neurodevelopmental effects, with many studies suffering from major reporting deficiencies. As indicated above, the monograph focuses on the large human epidemiology database because it directly addresses the question of whether fluoride affects human neurodevelopment. Therefore, based on the recommendations of the NASEM committee, the experimental animal section and risk of bias details have been removed from this monograph and ***the NTP concludes that the scientific evidence from experimental animal data are inadequate to inform whether fluoride exposure is associated with cognitive effects (including cognitive neurodevelopmental effects) in humans***.

☒ Agree
☐ Agree in principle with the exception(s) listed below:
☐ Do not agree because:

Response: No change requested

- [redacted] agreed with the inadequate designation.

I.33:

[redacted] **Comments:** If this is the conclusion of the review, [redacted] question the inclusion of non-human studies in this monograph.

Response: Disagree (no change)

- As discussed earlier, we contend that the animal studies section is a valuable part of the review because it provides a brief update to the 2016 NTP animal systematic review, identifies studies conducted by the DNTP to address deficiencies noted in the 2016 NTP animal systematic review, and reiterates the lack of consistent evidence from this body of literature to support human findings.

References:

Goodman CV, Hall M, Green R, Chevrier J, Ayotte P, Matinez-Mier EA, McGuckin T, Krzeczkowski J, Flora D, Hornung R, Lanphear B, Till C. (2022). Iodine Status Modifies the Association between

Fluoride Exposure in Pregnancy and Preschool Boys' Intelligence. Nutrients: 14(14):2920. https://doi.org/10.3390/nu14142920.

Prezioso G., Giannini C., and Chiarelli F. (2018). Effect of thyroid hormones on neurons and neurodevelopment. *Horm Res Paediatr*: 90:73-81. https://doi.org/10.1159/000492129.

In November 2021, [redacted] received: 1) the 2021 *Draft NTP Monograph on the State of the Science concerning Fluoride Exposure and Neurodevelopmental and Cognitive Health Effects: A Systematic Review,* 2) a copy of the NASEM Committee's comments on the 2020 draft NTP Monograph with NIEHS/DNTP responses (draft version of Sup01_Monograph), and 3) the [redacted] instructions. The instructions consisted of a preface, charge, instructions for the review, and a series of specific peer-review questions grouped by the following three topics: General Comments, Human Studies, and Studies in Non-Human Animals.

[redacted] were asked to provide their substantive scientific and technical comments and suggestions within the [redacted] form. In addition, they were asked whether they "Agree", "Agree in principle", or "Do not agree" with each NTP conclusion on confidence in a body of evidence.

The [redacted] instructions and specific peer-review questions are reproduced in the pages that follow in black text. [redacted] comments and responses to each question are also provided in black text starting with the words "[redacted] **comments**" in bold font. The NIEHS/DNTP responses have been inserted in blue text following each of the comments beginning with the word "**Response**" in bold font. Formatting has been applied to aid in reading.

The prepublication 2022 NTP Monograph reflects changes made after consideration of the comments from the [redacted] along with all other input received through April of 2022. The prepublication 2022 NTP Monograph was subsequently sent to [redacted] for additional comments. A revised "track changes" version of the monograph was developed in September 2022 titled the "DocMon_Track_Changes_2022_NTP_Monograph." The following bullets describe how edits are documented in the track changes version of the monograph in response to [redacted] comments and [redacted] comments:

- [redacted] For comments related to DocG_Monograph, DocH_Monograph, DocI_Monograph, DocJ_Monograph, and DocK_Monograph:
    - Edits are marked with a comment bubble in the DocMon_Track_Changes_2022_NTP_Monograph that identifies the text in question and briefly describes any revisions.
    - The comment bubble contains the exact text of the [redacted] Comment.
    - The comment bubble also provides a reference to the specific response to comments document with the detailed NIEHS/DNTP response (e.g., comments made in response to this [redacted] would be marked "*see DocJ_Monograph for detailed response*").
- [redacted] For comments DocA1_Monograph, DocA2_Monograph, DocB1_Monograph; DocB2_Monograph, and DocC_Monograph through DocF_Monograph:
    - Edits are marked in track changes format in the DocMon_Track_Changes_2022_NTP_Monograph.
    - A comment bubble has been added to the text in question containing the exact text of the [redacted] Comment.
    - The comment bubble also provides a reference to the specific response to comments document with the detailed NIEHS/DNTP response.

Preliminary comments on the draft NTP monograph prepared by the peer review [redacted] are noted below.
These preliminary comments are not binding and should not be construed to represent NTP determination or policy.

National Toxicology Program
NTP Monograph Letter Peer-Review Panel
Draft NTP Monograph on the State of the Science Concerning Fluoride Exposure and Neurodevelopmental and Cognitive Health Effects: A Systematic Review

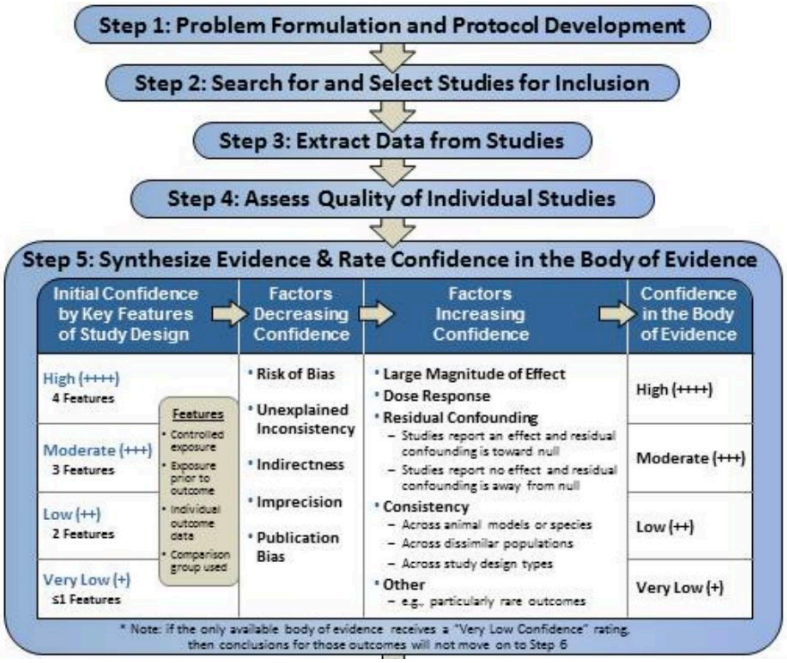
National Institute of Environmental Health Sciences
Research Triangle Park, NC

December 22, 2021

Fluoride State of the Science Document Review Form

[redacted]

Preface*:*
The objective of this evaluation was to conduct a systematic review of the published literature regarding the potential for exposure to fluoride to affect neurodevelopment and cognition in humans. The evaluation presented in the draft *NTP Monograph on the State of the Science Concerning Fluoride Exposure and Neurodevelopmental and Cognitive Health Effects* represents a comprehensive and current assessment. The methods used are from the Handbook for Conducting a Literature-Based Health Assessment Using OHAT Approach for Systematic Review and Evidence Integration, which presents a seven-step framework for systematic review and evidence integration. Please note: this evaluation stops at step 5 of the systematic review process and does not proceed to step 6 to translate the confidence rating for the body of evidence into a level of evidence for health effects (see Figure 2 from the handbook).



Charge:

(1) Comment on the technical accuracy and whether the draft *NTP Monograph on State of the Science Concerning Fluoride Exposure and Neurodevelopmental and Cognitive Health Effects* is clearly stated, and objectively presented.

(2) Determine whether the scientific evidence supports the NTP's confidence ratings for the bodies of evidence regarding neurodevelopmental and cognitive health effects associated with exposure to fluoride.

Instructions for Review:
All materials for this review are available in the Electronic Council Book (ECB). You will receive the specific URL and a password for accessing the ECB.

This evaluation identified 159 human studies relevant for assessing neurological health effects of exposure to fluoride; however, many studies included only secondary outcomes (e.g., 55 studies of thyroid hormones that were investigated as a potential mechanism). The scientific evidence in children and adults was evaluated separately to address potential differences in the health impact of fluoride exposure during development versus adulthood. Several studies evaluated learning and memory (n = 8 studies) or other cognitive developmental effects (e.g., total neurobehavioral scores and total mental capacity index in children, cognitive impairment in adults; n = 14 studies). Sixty-six human studies investigated IQ in children. Nineteen of the 66 IQ studies were determined to have low potential for bias and therefore, were categorized as "low risk of bias". Please give special attention to our assessment of these 19 studies.

- The 19 studies are available as PDFs and organized alphabetically in a folder on the ECB.
- All other studies are provided in the Health Assessment Workspace Collaborative, or HAWC database under the "studies list" tab, also organized alphabetically. You will also be provided a username and password for HAWC that will give you [redacted] permissions to access the PDFs in HAWC along with visualizations and other study information for this project at the following link (https://hawcproject.org/study/assessment/405/).

Please provide your substantive scientific and technical comments and suggestions within this [redacted] form. Identify and provide the rationale or scientific support for proposed changes or suggestions where possible.

If necessary, you can also provide additional editorial comments and recommendations for improving the report outside your specific charge questions (this form) within the draft report itself. Please note that only those comments included on the [redacted] form will be considered part of NTP's peer review report.

A. General Comments

1. Please comment on whether the scientific information presented in the draft monograph, including presentation of data in tables and figures, is technically correct, and clearly and objectively presented. Please suggest any improvements.
2. Please identify any information that should be added or deleted.

J.1: [redacted] **Comments:** Congratulations on a thorough and comprehensive systematic review – not only is the review itself impressive, but the HAWK system, your online portal, and all of your processes for assessing COI and training [redacted] were equally impressive.

Overall, the review is well organized, clearly written, and transparently documented. Below are a series of comments and questions, that if considered, may improve the review.

Response: provided below

- We appreciate [redacted] feedback and have provided responses to the series of comments in [redacted] table below where the issues were raised.

Section, page #	Comment
Objective and Specific Aims; page 5	**J.2:** Because this review has an extensive history that could be difficult for a reader to follow (i.e., the original 2016 review, and drafts from 2019, 2020, and the current draft), it would be helpful to develop a table or flowchart that documents that history. For example, you may consider noting the purpose/research question, findings, and noteworthy differences from previous/subsequent versions. See comments below, but the literature search section, in particular, was a little difficult to follow - and having the "big picture" of the review in a table or flowchart to refer to, would better allow the reader to follow all of the searches conducted, and how the differ, yet fit together to contribute to the present document. **Response: Agree (change made)** - In an effort to provide further clarity on the progression of this multiyear assessment, we have developed a new table (Table B-1 in *Appendix B* of the Sup02_2022_Prepublication_NTP_Monograph) that provides a timeline of key activities contributing to the Sup02_2022_Prepublication_NTP_Monograph, including literature searches that were utilized for the various drafts that underwent different peer reviews.
Objective and Specific Aims; page 5	**J.3:** It is not clear why the "hazard assessment step" was removed from the methodology. Is it because the authors deemed the step not possible based on available evidence? Or is it because the hazard assessment step will occur separately, taking into consideration both the review and the results of meta-analysis?

	Response: Agree (edited for clarity) ○ The *Preface* of the monograph clearly describes why the hazard assessment step was removed from the Sup02_2022_Prepublication_NTP_Monograph. Additionally, we developed a new table (Table B-1 in *Appendix B* of the Sup02_2022_Prepublication_NTP_Monograph) that provides a timeline of key activities contributing to the Sup02_2022_Prepublication_NTP_Monograph, including when the hazard assessment step was removed and that it was removed in response to the NASEM Committee's review report of the Sup04_2020_draft_NTP_Monograph. ○ In brief, the NASEM Committee's comments indicated they did not believe that the Sup04_2020_draft_NTP_Monograph presented a clear and convincing assessment to support its hazard conclusions. Although many of the comments offered by the NASEM Committee are addressed in the current document, we chose to delete the hazard assessment step and instead express our level of confidence in the evidence of an association between fluoride exposure and effects on cognitive neurodevelopment as our contribution to the larger ongoing discussion on the safe use of fluoride for oral health.
Methods, page 7	**J.4:** Would it be possible to define what is meant by "Categories focused on were those with more robust data at levels of fluoride more relevant to human exposure"? Should this information be documented as part of the PECO? Was this an inclusion criteria, or just used in prioritizing or weighting the evidence in drawing conclusions? **Response: Agree (edited for clarity)** ○ The sentence describing the use of categories with more robust data is a description of the approach used to collect, prioritize, and consider the available mechanistic data following the organization of the PECO statements. These were not inclusion criteria or an additional factor that could have been added to the PECO. Although the process for deciding which groupings of health effects to synthesize and whether to synthesize all groupings of health effects was described in the protocol, the specific decisions were made based on the results of the literature search and selection. This approach is specifically outlined in the *PECO Statements* section to describe how the in vitro/mechanistic data were evaluated and considered because it often differs from how human or animal data are assessed. We have edited the cited text for clarity and it now reads as follows: *"To prioritize and consider available mechanistic data, the categories focused on were those with more robust data at levels of fluoride more relevant to human exposure."*

Methods, page 8	**J.5:** The literature search section was somewhat confusing to follow, though, given the complexity of updating reviews, etc it is understandable why multiple searches were conducted. See previous comment regarding the various iterations of this review, historically, and how a table or flowchart may help the reader understand the progression of this review, and thus, better follow the searches that were carried out. For example, you may consider adding sub-headings within this section to distinguish which searches were run to capture which types of studies. **Response: Agree (edited for clarity)** ○ In response to [redacted] earlier comment on organization, and in an effort to provide further clarity on the progression of this multiyear assessment, we have developed a new table (Table B-1 in *Appendix B* of the Sup02_2022_Prepublication_NTP_Monograph) that provides a timeline of key activities contributing to the Sup02_2022_Prepublication_NTP_Monograph, including information relevant to the timing of the literature searches. For example, the expanded literature search to include non-English databases took place in May 2020 in response to the NASEM Committee's peer review report on the 2019 draft NTP Monograph.
Methods, page 9	**J.6:** Is there any plan to update the literature search run on May 1, 2020? Given that the search is now 1.5 years old, and this seems to be a topic with emerging evidence, it would be beneficial to update the search to ensure all relevant studies have been captured. **Response: Agree (No change)** ○ We performed an updated literature search in November 2021. There were a number of newer relevant publications identified, including several in Chinese journals. These newer publications (n = 7) are included as part of the meta-analysis, which is being prepared as a separate report for publication. We determined that, while the newer publications may slightly affect the quantitative results of the meta-analysis and dose-response meta-analysis, their findings are largely consistent with the literature reviewed in the current monograph and do not materially affect the level of confidence we have in the database. Because inclusion of these new studies in the monograph would necessitate further peer review, we have chosen not to include them.
Methods, page 10	**J.7:** This is the first time the "Flouride Action" website is mentioned (and the actual hyperlink appears in the subsequent section). It may be helpful to the reader to provide some rationale for why this website was specifically targeted. **Response: Agree (change made)** ○ We added new text to introduce the Fluoride Action Network and to clarify that the site was used as another resource because it is known to index fluoride publications. The new text appears as follows:

	"Fluoride Action Network website (http://fluoridealert.org/)—a site used as another resource to identify potentially relevant studies because it is known to index fluoride publications…" **J.8:** In addition, can it be assumed that any non-English paper that met criteria, regardless of outcome, would have been included? While it is understandable why the confirmatory search was done, it could be perceived as biased to only search for and include papers with null findings. **Response: Agree (change made)** ○ In terms of Chinese databases, we conducted the literature search independent of study findings, but we initially gave translation priority to studies that appeared to show no association. Although this was done to address potential publication bias, we agree that this was not appropriate and therefore have taken additional steps to translate and extract data from all relevant non-English studies identified from the Chinese database searches, including those that were not previously translated. Furthermore, the statements about null or no-effect studies have been deleted from the Sup02_2022_Prepublication_NTP_Monograph. ○ In addition, we updated the text in the *Literature Search* section to reflect that the search of Chinese databases was conducted to identify studies that may have been missed in previous searches because non-English language studies are not always indexed in the main databases used for this systematic review.
Methods, page 11	**J.9:** [redacted] can appreciate the use of machine-learning software to prioritize articles for screening. And the authors have done a nice job in describing and evaluating the algorithm employed when stopping at 98% - estimating that 2-4 studies may have been missed. However, given the high-profile nature of this review, and some level of uncertainty in the prediction algorithms of the tool, it may have been beneficial to manually screen the entire set of search results (2-4 studies is not an insignificant number when considering the total # of included articles). Use of machine-learning is helpful in that it can prioritize and identify sooner most included articles; however, when conducting systematic reviews used in large scale public health decision-making, it may be worth screening 100% of search results to ensure that all potentially relevant studies have been included. **Response: Disagree (no change)** ○ By using SWIFT Active Screener software to screen the initial literature search results, we avoided the need to manually screen over 13,000 abstracts. As outlined in the Sup02_2022_Prepublication_NTP_Monograph and systematic review protocol (https://ntp.niehs.nih.gov/go/785076), in addition to the screening of bibliographical databases, several additional methods to identify relevant literature were also employed. These included publicly

	posting the literature search results and asking peer reviewers at each stage whether they were aware of any additional relevant articles, screening the reference lists of reviews and included papers for possible articles, and conducting updated literature searches as outlined in response to a previous comment by the reviewer. The use of SWIFT Active Screener was estimated to result in the potential to miss one or two relevant human studies with primary neurodevelopmental or cognitive outcomes. The savings in time and impact were weighed against the potential impact of missing 1 or 2 studies relative to the nearly 100 human epidemiological studies identified with primary neurodevelopmental or cognitive outcomes, and this tradeoff was deemed to be acceptable.
Methods, page 13	**J.10:** If studies evaluating only goiters or thyroid size were not extracted, then why include them in the review altogether. Would it be more accurate to have amended the protocol to exclude these as outcomes of interest? **Response: Disagree (no change)** ○ The decision not to extract data on goiters was reached by technical experts during the review because changes in thyroid size would not inform whether a thyroid-mediated mechanism was involved in fluoride-associated changes in neurodevelopment. The protocol did not specify preferred or less informative thyroid-related data. [redacted] makes a valid point that, in hindsight, the protocol could have specified that studies only reporting thyroid size or goiters could have been excluded. Given that this decision was reached during the assessment, it is common practice to provide reasoning in the systematic review for these types of decisions, not to amend the protocol with these details.
Methods, page 15	**J.11:** Given that all included study designs were observational in nature, risk of bias due to confounding is a serious consideration. This [redacted] appreciates the thorough discussion of key confounders considered in risk of bias assessments but has concerns that even in studies rated as "low risk of bias," there remain serious concerns about the potential for confounding. This is especially important when considering an outcome like IQ, for which concerns are often raised about the specificity of the outcome, and its relationship with other constructs, such as SES, education, and race. **Response: Disagree (no change)** ○ We agree that risk of bias due to confounding is a serious consideration in any risk-of-bias assessment of observational studies but, due to both the comprehensive risk-of-bias assessment of each individual study and the assessment of potential confounding across studies, we disagree that there remain serious concerns about potential confounding among the low risk-of-bias studies. As described in the protocol, for a study to be considered low risk of bias for confounding, there had to be direct or indirect evidence that the key covariates (age, sex, and SES) and any other covariates considered important for the study's specific study

	population and/or outcome were sufficiently considered in terms of confounding. For example, studies of populations in China, India, and Mexico, where there is concern about exposures to high fluoride and high arsenic, were required to address arsenic. Figure 6 shows that 16 of 19 low risk-of-bias studies addressed each of the three key covariates and other important covariates, meeting the requirements for low risk of bias due to confounding. Looking across the body of literature, we observed considerable variation in covariates addressed across the 19 low risk-of-bias studies. When considering the impact of potential confounding on the consistency of results, no trends were discernable that would suggest that bias due to confounding has impacted or would explain the consistency in findings across the body of evidence that higher fluoride exposure is associated with lower IQ in children. ○ If a key covariate or other important covariate was not addressed in a study, we would also consider the most likely direction and magnitude of the potential bias. If the bias was likely to be toward the null, that may increase our confidence in the reported direction of the association. *Appendix E* (*Details for Low Risk-of-bias Studies*) includes detailed assessments of and justifications for each risk-of-bias rating, including considerations for the direction and magnitude of potential bias. ○ [redacted] identifies SES, education, and race. SES was a key covariate, education was considered as a measure or proxy of SES (see footnote to Figure 6), and race/ethnicity is listed in the protocol as a potentially important confounder; however, every study was conducted outside of the United States and there was no direct or indirect evidence to indicate that confounding by race/ethnicity was a concern. **J.12:** Is child sex a true confounder, in that its related both to the exposure and the outcome? (i.e., is there data to suggest that fluoride exposure differs based on sex?) **Response: No change requested** ○ We consider biological sex to be an important covariate and potential confounder for several reasons: (1) sex has historically been considered an important potential confounder in the literature (see Table 6 in Sup02_2022_Prepublication_NTP_Monograph) (Lash et al. 2021; Gochfeld 2017); (2) sex is an important risk factor for neurodevelopmental and cognitive outcomes (Cowell and Wright 2017); and (3) sex-related dietary ingestion and dietary differences are realistic in observational studies (D'Amico et al. 2020; Keller et al. 2019).
Methods, page 17	**J.13:** The paragraph describing RoB procedures could be moved up (prior to the PECO sections; as currently placed it gets a little lost and/or could be misperceived as relating only to outcome assessment).

	Response: Disagree (no change) ○ The systematic review process involves several steps and stages, and there is a general order by which these stages take place. The risk-of-bias discussion is located in an area of the *Methods* section that corresponds to the appropriate stage of the systematic review process, as is standard in publications of these types of reviews. Moving the risk-of-bias methods before the PECO (and therefore the literature search and screening methods) would create a misperception that the literature screening was influenced by study quality.
Methods, page 19	**J.14:** It is not clear why the meta-analysis portion of this review is being prepared as a separate report. **Response: Agree (edited for clarity)** ○ The decision to pursue a narrative evidence synthesis rather than a meta-analysis was made while preparing the 2019 draft NTP Monograph because our goal of generating a document to support a hazard assessment did not require a quantitative estimate of hazard (e.g., numeric estimate of IQ points lost per mg F/L of drinking water or urine). However, as outlined in a new table that provides a timeline of draft monographs and important decision points (Table B-1 in *Appendix B* of the Sup02_2022_Prepublication_NTP_Monograph), comments received from the NASEM Committee that reviewed the 2019 draft NTP Monograph (NTP, 2019) recommended that we perform a meta-analysis and indicated that the outcome would be critical to reaching a hazard conclusion. We therefore prepared a meta-analysis and included both the meta-analysis and dose-response meta-analysis in the revised Sup04_2020_draft_NTP_Monograph (NTP, 2020). In its review of that 2020 draft NTP Monograph, the NASEM Committee again stated that the document fell short of supporting our hazard call, and the Committee also had additional recommendations to improve the meta-analysis. After reflecting on the NASEM Committee comments on the Sup04_2020_draft_NTP_Monograph, we decided to remove the evidence integration step from the systematic review of the literature and instead issue the report (after further independent peer review) as a document outlining the state of the science on the association between fluoride exposure and deficits in neurodevelopment and cognition. This change is outlined in the *Preface* to the Sup02_2022_Prepublication_NTP_Monograph. Removing the evidence integration step from the systematic review precluded a determination of an overall hazard call. We then decided to revise and submit the meta-analysis as a separate peer-reviewed publication because it was no longer needed in an evaluation of confidence in the database of human evidence. An additional consideration was that the meta-analysis and dose-response analysis were performed only on the studies addressing fluoride exposure in relation to deficits in children's IQ, rather than on

	other neurological outcomes in children or cognition in adults. The separate meta-analysis considers comments from the NASEM Committee in its revisions.
Methods, page 20	**J.15:** [redacted] does not agree with the premise that all human studies are direct; it seems that certain measures of fluoride exposure have concerns with directness (i.e., endemic geographical region, job title). **Response: Disagree (no change)** ○ [redacted] cites two examples of fluoride exposure "endemic geographical region" and "job title" as potential concerns with directness. However, these examples are both direct evidence for this systematic review as defined in Table 1 the human PECO (Population, Exposure, Comparator and Outcome) Statement. Direct evidence comes from research that directly assesses exposures that are the focus of a given systematic review when described in populations that are also within the focus of a systematic review. As listed below, the PECO Statement in Table 1 specifies the population of interest as "humans without restriction" and exposure includes "job title" and "water levels" that cover groundwater exposure from endemic geographical regions. *"**Population**: Humans without restriction as to age or sex, geographic location, or life stage at exposure or outcome assessment* ***Exposure:** Exposure to fluoride based on administered dose or concentration, biomonitoring data (e.g., urine, blood, other specimens), environmental measures (e.g., air, water levels), or job title or residence…"* ○ [redacted] is using a definition of the population and exposures of interest that differ from the PECO statement for this review. [redacted] example would only apply when the specific question of a review is directed toward a narrow subpopulation and that would be stated in an alternate PECO. For example, if the review had been to evaluate the evidence on the association between occupational exposures to fluoride through mining and cognitive effects, then there would be direct and indirect human evidence. Direct evidence would include studies of miners with inhalation exposure or other occupational exposures determined by "job title" or other metrics); indirect evidence might include studies of oral exposures through water or dietary sources. The objective of this systematic review is to evaluate the evidence concerning the association between any fluoride exposure and neurodevelopmental and cognitive effects; therefore, all human studies are direct evidence.
Results, Figure 2, page 23	**J.16:** Identifying 15 references through other sources seems somewhat high. Was there a need to adjust the original search strategy to capture those references?

	Response: Agree (change made) ○ We have added text to clarify why the references identified by other sources were not captured in the database searches. In brief, 11 of the 15 references identified through other sources were not indexed in the bibliographic databases searched and therefore were not captured by the database searches. Many of the studies initially identified by other sources were non-English-language studies, and we recognized that additional targeted search strategies were required to identify non-English-language studies for this review. The supplemental search of Chinese databases was designed and conducted to address these challenges. Upon further review, we have clarified that four of the references in question were captured in the Chinese database searches, and we have made this correction to the text and study flow diagram. We were unable to identify the remaining 11 studies in any database searches. Regarding the impact of these 11 studies on the systematic review, only 1 of the 11 studies was a low risk-of-bias IQ study in children, and this study was included in the 19 low risk-of-bias studies on which the moderate confidence in the IQ-in-children body of evidence is based. The omission of this single study would not impact the moderate confidence rating. Of the remaining 10 studies, 7 were high risk-of-bias IQ-in-children studies and 1 was a high risk-of-bias adult study. The omission of the 7 (out of 53) high risk-of-bias IQ-in-children studies or the 1 (out of 8) high risk-of-bias adult studies would not impact any confidence conclusions in the monograph. Similarly, the two experimental animal studies would not impact the evaluation as the animal evidence was considered inadequate. The following new text appears as a footnote in the *Literature Search Results* section of the monograph: *"These 11 studies (9 human and 2 animal studies) were not identified through the electronic database searches, as they were not indexed in any of the electronic databases searched. Note that the supplemental search of non-English-language databases was designed in part to identify non-English-language studies that are not indexed in traditional bibliographic databases such as PubMed. It was successful in this goal, as multiple studies that were initially only identified through "other sources" were subsequently captured in the supplemental Chinese database search, leaving only 11 as identified through other sources."*
Results	**J.17:** In general, it would be helpful to the reader to describe the included study designs earlier in the respective results sections (i.e., along with the total #s of included articles, describing the # of cross-sectional, prospective cohort, etc is important). **Response: Agree (change made)** ○ We have added descriptions of the cohort, case-control, cross-sectional, and case report/case series study designs based on the NRC Report on

	Environmental Epidemiology (NRC 1997) as footnotes to Table 4 in the prepublication 2022 NTP Monograph, as follows: *"[c]Cohort studies are observational, studies in humans that examine a cohort prospectively or retrospectively over time.* *[d]Case-control studies are observational studies in humans that compare exposures of individuals who have a specific health effect or disease with exposures of controls who do not have the effect or disease. Controls generally come from the same population from which the cases were derived.* *[e]Cross-sectional studies are observational studies in humans that examine the relationship between exposures and outcomes or health effects assessed contemporaneously. Cross-sectional studies include population surveys with individual data (e.g., NHANES) and surveys with aggregate data (i.e., ecological studies).* *[f]A case report (or case study) is a descriptive study of a single individual or small group in which the study of an association between an observed effect and a specific environmental exposure is based on clinical evaluations and histories of the individual(s). A case series study in environmental epidemiology is designed to share health-related events on a collection of case reports on subjects with the same or similar health outcome(s) and environmental exposure(s)."* ○ We also added information on counts of studies per study design to the *Overview of Studies* subsections of the *IQ in Children*, *Other Neurodevelopmental or Cognitive Effects in Children*, and *Cognitive Effects in Adults* sections as indicated below. *"Nineteen studies (3 longitudinal prospective cohort and 16 cross-sectional studies) with low potential for bias evaluated the association between fluoride exposure and IQ in children (see Quality Assessment of Individual Studies section for methods on determining which studies pose low risk of bias)."* *"Nine low risk-of-bias studies (three prospective cohort and six cross-sectional studies) evaluated the association between fluoride exposure and cognitive neurodevelopmental effects other than IQ in children."* *"Two low risk-of-bias cross-sectional studies evaluated the association between fluoride exposure and cognitive effect in adults (Jacqmin et al. 1994; Li et al. 2016)."* **J.18:** It would also be of interest to expand Fig 3, or create a similar figure, to capture the ages at which fluoride exposure was measured. **Response: Disagree (no change)** ○ Although we agree that an expansion of Figure 3 could be interesting, the purpose of Figure 3 was to visualize the number of relevant studies identified in order to evaluate the outcome categories for pockets of data. For Figure 3, the studies were labeled as child or adult in order to

	evaluate if there were sufficient data to evaluate child and adult studies separately, as was done for the 2016 NTP animal evaluation.

B. Human studies

I. Fluoride exposure and children's IQ

J.19:

1. Comment on whether the approach described in the methods to search for and select human studies on neurodevelopmental or cognitive function effects associated with fluoride exposure was appropriate for evaluating potential effects of fluoride exposure on measures of IQ in children.

[redacted] **Comments:** Yes, the approach used to search for and select studies was appropriate.

Response: No change requested

- No response necessary.

J.20:

2. Comment on whether the approach used to assess risk of bias was clearly described and appropriately applied to the set of studies designated as "low risk of bias."

[redacted] **Comments:** In general, it is difficult to understand how cross-sectional studies that adjusted for few, or no, confounders, employed somewhat indirect measures of fluoride exposure (or did not fully capture all sources of exposures to fluoride), or had concerns related to selection bias, were designated as "low risk of bias." If, for example, some confounders were accounted for in the design or analysis, other than statistical adjustment, it may be worth noting that on Table 6 (otherwise, it appears that many papers accounted for no confounders).

For example, Xiang, 2003a did not statistically adjust for any confounders. They did report some findings in relation to some of the confounders, but not to the extent that [redacted] would perceive them to have been fully accounted for.

For Xiang, 2011, the paper is published in a somewhat abbreviated format, appearing almost as a conference report or short correspondence. The journal does appear to be peer-reviewed currently, but it may be worth confirming that papers from 2011 were in fact peer-reviewed.

For Till, 2020, it does not appear that the authors applied inclusion/exclusion criteria related to the length of time subjects lived in the geographical areas tested. Therefore, it is difficult to know if exposure was accurately estimated. In addition, the authors did not confirm that formula preparation was done with tap or bottled water, but rather they used a proxy (maternal report of drinking tap/bottled); and it is unclear whether maternal drinking behaviors match formula preparation methods. Finally, the study measured exposure from 0-6mo, and did not account of fluoride exposure that occurred over the course of follow-up to age 3-4y (i.e., teeth brushing, supplemental intake, dietary intake). Therefore, it seems that there are some serious concerns related to potential exposure misclassification in this study.

Response: Disagree (edited for clarity)

- We appreciate [redacted] concern regarding cross-sectional studies; however, we disagree with the assertion that the low risk-of-bias cross-sectional IQ studies

have serious concerns related to confounding, exposure assessment, or selection bias that would preclude them from their designation as "low risk of bias." As described below and in detail in *Appendix E*, using the criteria in our protocol, we determined that these are well-conducted studies with minimal risk-of-bias concerns. The subsequent bullets in this response detail the strengths of these studies regarding their study design and low potential for bias due to confounding, exposure misclassification, and selection bias. In addition, we address the study-specific concerns raised by [redacted].

- **Confounding:** Due to both the comprehensive risk-of-bias assessment of each individual study and the assessment of potential confounding across studies, we disagree that there remain serious concerns about potential confounding among the low risk-of-bias studies. [redacted] is correct that Table 6 reports only the covariates that were adjusted for statistically. However, as is recommended by [redacted], Figure 6 does indicate when a covariate was adjusted for statistically and/or was not a concern for confounding in a particular study. As described in the protocol, for a study to be considered low risk of bias for confounding, there had to be direct or indirect evidence that the key covariates (age, sex, and SES), and any other covariates considered important for the specific study population and/or outcome, were sufficiently considered in terms of confounding. Examples of what it means for a covariate to be sufficiently considered in terms of confounding are described in a revised footnote to Figure 6 and include: it (the covariate) was statistically adjusted for in the final model, it was included in the model but not the final model because it did not substantially change the effect estimate, it was reported to have the same distribution in both the exposed and unexposed groups, and it was reported to not be associated with the exposure or outcome in that specific study population (thereby eliminating it as a potential confounder). Figure 6 shows that 14 of 16 low risk-of-bias cross-sectional studies addressed each of the three key covariates and other important covariates, meeting the requirements for low risk of bias due to confounding. Looking across the body of literature, we observed considerable variation in the covariates addressed across the 19 low risk of bias studies (16 cross-sectional and 3 prospective cohort studies). When considering the impact of potential confounding on the consistency of results, no trends were discernable that would suggest that bias due to confounding has impacted or would explain the consistency in findings across the body of evidence that higher fluoride exposure is associated with lower IQ in children.

  If a key covariate was not addressed in a study, we would also consider the most likely direction and magnitude of the potential bias. If the bias was likely to be toward the null, it may increase our confidence in the reported direction of the association (Xiang et al. [2003] is an example of this because it did not directly address potential co-exposure to arsenic; see further discussion of this study below). Detailed assessments of and justifications for risk-of-bias ratings for the low risk-of-bias studies, including considerations for likely direction and magnitude of bias, are provided in *Appendix E* (*Details for Low Risk-of-bias Studies*).

- **Exposure (characterization and considering potential misclassification):** Fifteen of the 16 low risk-of-bias cross-sectional studies that assessed the association

between fluoride exposure and IQ provide direct or indirect evidence that exposure was consistently assessed using acceptable methods and used individual, direct exposure data based on urine or water measures with appropriate analyses. For each study, a detailed summary of the exposure characterization, the risk-of-bias rating, and the basis for the rating for exposure characterization are provided in *Appendix E*, which includes discussion of any potential exposure misclassification and the potential impact on direction and magnitude of effect size. As we detail in *Appendix E* and summarize in the *Exposure Characterization in IQ Studies* section, there were few, if any, risk-of-bias concerns regarding exposure characterization in these studies. Thirteen of the 16 cross-sectional studies utilized an exposure measure (i.e., urine or serum) that would capture all sources of exposure to fluoride. Only one of the 16 cross-sectional studies had potential for bias due to exposure misclassification, which is discussed in detail in the *Exposure Characterization in IQ Studies* section and *Appendix E*. In this study (Seraj et al. 2012), a statistically significant association between water fluoride and IQ was reported. We determined that the potential exposure misclassification would bias the results toward the null, indicating that the true association may be greater than what was observed in this study.

- **Exposure (whether exposure preceded outcome):** Note that we acknowledge in the *Results by Study Design – Cross-sectional Studies* section that, as a general study design, cross-sectional studies often do not provide sufficient information to ensure that exposure preceded outcome. However, we do not judge studies simply by study type. Each study is assessed individually for multiple factors, including if the research design and conduct inform whether exposure preceded outcome assessment, as is the case for the low risk-of-bias cross-sectional IQ studies (see below):

  "In some cases, cross-sectional studies do provide indicators of prior exposure (e.g., prevalence of dental fluorosis, limiting study populations to subjects who lived in the same area for long periods of time). Evidence that exposure occurred prior to the outcome of interest increases the confidence in results and any potential association reported in these studies. Of the 16 low risk-of-bias cross-sectional studies, 12 established that exposure preceded the outcome assessment…"

- **Selection:** [redacted] also raised the concern about potential selection bias for cross-sectional studies. We agree with [redacted] that selection bias is an important consideration in risk-of-bias evaluations. As described previously, *Appendix E* includes a detailed summary of population selection for each low risk-of-bias studies and the basis for the ratings for selection bias and exposure characterization. All 16 low risk-of-bias cross-sectional IQ studies were rated either *definitely low risk of bias* or *probably low risk of bias* due to selection bias. In addition, we edited the following text in the *Methods* section to clarify that, in addition to the three key risk-of-bias questions, the answers to the other risk-of-bias questions were considered in assessing potential bias, including selection bias.

  "The other risk-of-bias questions, including selection of study participants, were also considered and were used to identify any other risk-of-bias concerns that may

indicate serious issues with a study that could cause it to be considered high risk of bias."

- **Individual studies cited by [redacted]:**
  - Xiang et al. (2003): This study was considered low risk of bias for confounding. [redacted] is correct that the study did not statistically adjust for the three key covariates; however, as described in *Appendix E*, the key covariates were considered not a concern for confounding because authors noted that these factors were similar between the two compared villages. We did note that there was potential co-exposure to arsenic, which would likely bias the observed association toward the null due to the reporting of higher arsenic levels in the control area.
  - Xiang et al. (2011): The journal *Fluoride* says it "contains peer-reviewed scientific reports on agricultural, analytical, biochemical, biological, chemical, clinical, dental, ecological, environmental, industrial, medical, metabolic, pharmacological, synergistic, toxicological, and veterinary aspects of inorganic and organic fluorides." Therefore, we do not have reason to believe that the manuscript was not peer-reviewed.
  - Till et al. (2020): [redacted] concerns for this study fall under potential exposure misclassification. We agree that this analysis was not designed to account for fluoride exposure postweaning. We also agree with the possibility of exposure misclassification. However, there is no evidence to suggest that the potential exposure misclassification is differential based on whether a participant lived in a fluoridated or non-fluoridated area. Therefore, as described in *Appendix E*, the possibility of exposure misclassification is non-differential and is likely similar in all participants, which would likely bias the association toward the null.

3. Comment on assessment of the human studies with regard to:
   a) How findings from individual studies designated as "low risk of bias" were interpreted.
   b) How the overall set of confounders across the body of evidence from children's IQ studies was considered and presented.
   c) How the confidence rating in the body of evidence was developed and supported.

J.21: [redacted] **Comments:** See above for comments on risk of bias ratings, and concerns related to confounding and/or residual confounding.
Page 39, last full paragraph includes the sentence, "Despite these few variations, the overall evidence of an effect on IQ is apparent." This [redacted] suggests editing the word "effect" to "association" or "correlation," given that the included studies are all observational.

Response: Agree (change made)

- Edits have been made throughout the Sup02_2022_Prepublication_NTP_Monograph to use the terms 'effect,' 'association,' and 'correlation' consistently and most appropriately. For example, the sentence referenced by [redacted] has been revised and reads as follows:

"Despite these few variations, the overall evidence of an association with lower IQ is apparent."

J.22: Page 40, "Gender considerations": Is there some biological plausibility that there would be sex differences in the relationship between fluoride exposure and neurocognitive outcomes. The term "susceptibility" is used several times, but it is unclear what that means. It seems to imply a biological reason, but it is unclear whether mechanistic evidence is supportive of that (or if gender differences actually represent some sort of residual confounding).

Response: Agree (change made)

- Note that this response refers to sex considerations because we updated the language from "gender" to "sex" in the monograph in response to a comment from [redacted]. There are several reasons we considered potential sex differences in this systematic review: (1) sex has historically been considered an important potential confounder in the literature (see Table 6 in Sup02_2022_Prepublication_NTP_Monograph) (Lash et al. 2021; Gochfeld 2017); (2) sex is an important risk factor for neurodevelopmental and cognitive outcomes (Cowell and Wright 2017); and (3) potential sex-related ingestion and dietary differences are realistic in cross-sectional studies (D'Amico et al. 2020; Keller et al. 2019).
- We have added the following text to the *Sex Considerations* section to address [redacted] comment, as follows:

  "Recent literature suggests that adverse neurodevelopmental effects of early-life exposure to fluoride may differ depending on timing of exposure and sex of the exposed. In a review of the human and animal literature, Green et al. (2020) concluded that, compared with females, male offspring appear to be more sensitive to prenatal but not postnatal exposure to fluoride, with several potential sex-specific mechanisms."

J.23: Page 48-49, **Assessment** of Risk of Bias: While the studies noted as "low risk of bias" are certainly lower risk than the studies noted as "high risk of bias," it appears that the evidence base is still subject to a number of important risks, particularly related to confounding and exposure classification (i.e., are they "low risk" or "lower risk"?).

Response: Agree (edited for clarity)

- [redacted] raises a valid point on clear terminology and word choice for the terms "low" and "lower" as well as "high" and "higher" to describe risk of bias. Word choice was carefully considered and reflects input from technical experts and [redacted]. In particular, use of the term "lower" risk of bias may raise the question, "lower than what?" Given this input, the decision was made to use a clear definition of "low risk-of-bias studies" and "high risk-of-bias studies" and to describe in detail how these terms are used early on in the document in the *Quality Assessment of Individual Studies* and *Risk-of-bias Considerations for Human Studies* sections. To clarify the definition of "high risk-of-bias studies," the quoted text below was added to the *Risk-of-bias Considerations for Human* Studies section. In addition, the detailed assessments of and justifications for risk-

of-bias ratings for the key studies are provided in *Appendix E* (*Details for Low Risk-of-bias Studies*). It is also important to note that the confidence rating of moderate for the association between higher fluoride exposures and lower children's IQ reflects assessment of risk of bias across the body of evidence as one of multiple specific factors evaluated in determining the confidence rating.

"Studies could also be considered high risk of bias if rated probably high risk of bias for one key risk-of-bias question along with other concerns, including potential for selection bias and concerns with statistical methods."

J.24: Page 48-49, Assessment of Unexplained Inconsistencies: While there is some consistency in findings suggesting that increased exposure to fluoride is associated with lower IQ, many studies reported mixed results (generally reporting a mix of inverse associations and null findings). How were these mixed findings taken into consideration when evaluating unexplained inconsistencies?

Response: Agree (change made)

1) We revised the text in the *Confidence Assessment of Findings on IQ in Children* section regarding unexplained inconsistencies. The 'Unexplained inconsistencies' bullet now reads as follows:

"Unexplained inconsistencies: The direction of the association is consistent in the majority of studies, and there was no downgrade for this factor. Eighteen of the 19 low risk-of-bias studies reported associations between higher fluoride levels and lower IQ scores in children. These studies were conducted in 5 different countries on more than 7,000 children from 15 different study populations. There is consistency in the direction of the association across prospective and cross-sectional study designs. There is also consistency in the direction of the association across studies using different fluoride exposure measures, including urinary and drinking water fluoride. The one study that did not observe an association did not provide results in a comparable manner and therefore this body of evidence is not considered to have unexplained inconsistencies."

- As we further explain in the *Summary of Key Findings for Low Risk-of-bias Children's IQ Studies* section of the Sup02_2022_Prepublication_NTP_Monograph, *"Although some studies that conducted multiple analyses observed within-study variations in results (e.g., differences between subsets of IQ tests), these variations were unique to individual studies and did not detract from the overall consistency in the findings that higher fluoride is associated with lower IQ scores."*

J.25:

4. ***NTP concludes a rating of moderate confidence in the body of evidence for lower IQ in children associated with fluoride exposure***.

☐ Agree:

X Agree in principle with the exception(s) listed below: [redacted] agree in principle with the direction of association concluded by NTP, but am uncertain that a rating of moderate is appropriate for a body of evidence comprised of mostly cross-sectional studies, that have not considered the full range of key confounders, or may have some concerns with exposure classification.

☐ Do not agree because:

Response: Disagree (no change)

- Study type (e.g., cross-sectional, cohort, case-control) should not serve as a proxy for assessing level of confidence in a body of evidence. Instead, NTP's framework for developing a confidence rating for a body of evidence starts with an initial confidence rating that is determined by the ability of the studies to address causality as reflected in the confidence that exposure preceded and was associated with the outcome (Rooney et al. 2014). This ability, in turn, is based on four key study design features (controlled exposure, exposure prior to outcome, individual outcome data, and comparison group) (https://ntp.niehs.nih.gov/go/ohathandbook). To meet the criteria for an initial confidence rating of moderate, studies must have three of the four key features. Among the 19 low risk-of-bias studies that form the basis of this body of evidence, 15 studies (3 prospective cohort studies and 12 cross-sectional studies) have 3 of the 4 features (individual outcome data, comparison group, and exposure prior to outcome) and so support an initial confidence rating of moderate. More specifically, the 12 cross-sectional studies provide sufficient details to establish that exposure preceded the outcome assessment (e.g., by providing prevalence of dental fluorosis, limiting study populations to subjects who lived in the same fluorosis area for long periods of time), in addition to having individual outcome data and a comparison group. Although cross-sectional studies can have limitations in ensuring that exposure preceded outcome, that is not the case with the cross-sectional studies that contributed to the determination of moderate confidence in an association between fluoride exposure and lower IQ in children. The three prospective cohort studies also provide individual outcome data, include a comparison group, and demonstrate that exposure preceded outcome, and so support initial confidence rating of moderate. Finally, the consistency of results across the body of evidence, including both study designs, and after consideration of all of the GRADE-based factors that may increase or decrease confidence, support the final confidence rating of moderate.
- Please see previous responses in Sections A and B.I.2.of this document that explain why we disagree that serious concerns remain about potential confounding and exposure misclassification among the low risk-of-bias studies that would impact our confidence in the literature.

II. Fluoride exposure and non-IQ neurodevelopmental or cognitive effects in children

J.26:

1. Comment on whether the approach described in the methods to search for and select human studies on neurodevelopmental or cognitive function effects associated with fluoride exposure was appropriate for evaluating potential effects of fluoride exposure on non-IQ neurodevelopmental or cognitive effects in children.

[redacted] **Comments**: Yes, the approach used to search for and select studies was appropriate.

Response: No change requested

- No response necessary.

J.27:

2. Comment on whether the approach used to assess risk of bias for studies in children on non-IQ neurodevelopmental or cognitive effects was clearly described and appropriately applied.

[redacted] **Comments:** In general, this body of evidence has fewer apparent concerns with confounding or residual confounding. However, there are likely some concerns with exposure assessment/classification, given that some of the longitudinal studies assessed maternal fluoride status and neurocognitive outcomes later in childhood, without accounting for fluoride exposure of the child during the period of follow-up.

Response: Disagree (no change)

- We agree that, ideally, studies would account for a child's lifetime exposure; however, when considering risk of bias for exposure misclassification, in order for timing of exposure to impact the risk-of-bias rating, the exposure assessment would have to take place at a time that would not be appropriate for the outcome assessed (e.g., measurement of exposure after outcome). Evaluating exposure during a specific life stage prior to the outcome assessment does not indicate any misclassification for that specific life stage.

J.28:

3. Comment on assessment of the human studies with regard to:
   a) How findings from individual "low risk of bias" studies were interpreted.
   b) How the confidence rating in the body of evidence was developed and supported.

[redacted] **Comments:** The findings were interpreted correcting and a confidence rating of low seems an appropriate assessment.

Response: No change requested

- No response necessary.

J.29:

4. ***The NTP concludes a rating of low confidence in the body of evidence for decreases in measures of other neurodevelopmental or cognitive effects in children associated with fluoride exposure***.

X Agree
☐ Agree in principle with the exception(s) listed below:
☐ Do not agree because:

Response: No change requested

- [redacted] agreed with the low confidence rating.

III. Fluoride exposure and cognitive effects in adults

J.30:

1. Comment on whether the approach described in the methods to search for and select human studies on neurodevelopmental or cognitive function effects associated with fluoride exposure was appropriate for evaluating potential effects of fluoride exposure on cognitive effects in adults.

[redacted] **Comments:** Yes, the approach used to search for and select studies was appropriate.

Response: No change requested

- No response necessary.

J.31:

2. Comment on whether the approach used to assess risk of bias for studies in adults on cognitive effects was clearly described and appropriately applied.

[redacted] **Comments:** Yes, the approach used to assess risk of bias was clearly described and generally appropriate.

Though, it is unclear whether these studies adequately captured a critical fluoride exposure window likely to impact neurocognitive health (i.e., does fluoride exposure in older adulthood impact neurocognitive health?) For example, lifelong fluoride exposure, and/or fluoride exposure at different lifestages that may be more critical to neurocognitive development, were not captured in these cross-sectional studies. Thus, it raises questions as to whether these cross-sectional studies are truly "low risk of bias," or are "lower" risk of bias than others.

Response: Agree (edited for clarity)

- We agree with [redacted] that none of the studies evaluated differential fluoride exposures in adults with adequate adjustment for earlier life exposures. The available body of evidence does not provide sufficient information to draw a conclusion on the critical period of exposure assessment/classification. While we agree that questions about critical periods of exposure and duration of exposure are important to understanding relative hazards from fluoride to neurocognitive health, these would not be addressed in the risk-of-bias evaluation of individual studies and are instead a limitation of the evidence base. If there had been more data and greater confidence in the body of evidence for studies in adults, the ability of the studies to address questions of lifelong exposure or critical exposure windows would have been added to the *Discussion* section. We consider the approach for the risk-of-bias evaluation to be appropriate for assessing the quality of the studies and conclude that an overall assessment of low confidence in an association between higher fluoride exposures and cognitive effects in adults is appropriate based on the body of evidence. In addition, the following text was added to the *Limitations of the Evidence Base* subsection of the *Discussion* of the Sup02_2022_Prepublication_NTP_Monograph to acknowledge the lack of studies to inform these questions.

  "No studies are available to evaluate lifelong exposure in adults, or fluoride exposure over a child's lifetime and neurodevelopmental or cognitive changes over time."

J.32:

3. Comment on assessment of the human studies with regard to:
   a) How findings from individual studies were interpreted.
   b) How the confidence rating in the body of evidence was developed and supported.

[redacted] **Comments:** The findings were interpreted correcting and a confidence rating of low and that the evidence is inadequate seems an appropriate assessment.

Response: No change requested

- No response necessary.

J.33:

4. ***The NTP concludes a rating of low confidence in the body of evidence for changes in cognitive effects in adults with fluoride exposure***.

X Agree
☐ Agree in principle with the exception(s) listed below:
☐ Do not agree because:

Response: No change requested

- [redacted] agreed with the low confidence rating.

C. Studies in non-human animals

J.34:

The NTP agrees with the comments of the NASEM committee (NASEM 2020, 2021) concerning the overall poor quality of the experimental animal database on fluoride exposure and neurodevelopmental effects, with many studies suffering from major reporting deficiencies. As indicated above, the monograph focuses on the large human epidemiology database because it directly addresses the question of whether fluoride affects human neurodevelopment. Therefore, based on the recommendations of the NASEM committee, the experimental animal section and risk of bias details have been removed from this monograph and ***the NTP concludes that the scientific evidence from experimental animal data are inadequate to inform whether fluoride exposure is associated with cognitive effects (including cognitive neurodevelopmental effects) in humans***.

X Agree
☐ Agree in principle with the exception(s) listed below:
☐ Do not agree because:

Response: No change requested

- [redacted] agreed with the inadequate designation.

References

Cowell WJ, Wright RJ. 2017. Sex-Specific Effects of Combined Exposure to Chemical and Non-chemical Stressors on Neuroendocrine Development: a Review of Recent Findings and Putative Mechanisms. Curr Envir Health Rpt. 4: 415–425. https://doi.org/10.1007/s40572-017-0165-9

D'Amico D, Parrott MD, Greenwood CE, Ferland G, Gaudreau P, Belleville S, Laurin D, Anderson ND, Kergoat MJ, Morais JA, Presse N, Fiocco AJ. 2020. Sex differences in the relationship between dietary pattern adherence and cognitive function among older adults: findings from the NuAge study. Nutr J. 19(58) . https://doi.org/10.1186/s12937-020-00575-3

Gochfeld M. 2017. Sex Differences in Human and Animal Toxicology. Toxicol Pathol. 45(1):172-189. doi: 10.1177/0192623316677327. PMID: 27895264; PMCID: PMC5371029.

Keller KL, Kling SMR, Fuchs B, Pearce AL, Reigh NA, Masterson T, Hickok K. 2019. A Biopsychosocial Model of Sex Differences in Children's Eating Behaviors. *Nutrients*. 11(3): 682. https://doi.org/10.3390/nu11030682

Lash TL, VanderWeele TJ, Haneuse S, Rothman KJ. Modern Epidemiology, 4th edition. Wolters Kluwer, 2021.

National Research Council (NRC). 1997. Environmental Epidemiology: Volume 2: Use of the Gray Literature and Other Data in Environmental Epidemiology. Washington (DC): National Academies Press (US); 1997. 2, Environmental-Epidemiology Studies: Their Design and Conduct. Available from: https://www.ncbi.nlm.nih.gov/books/NBK233644/

In November 2021, [redacted] received: 1) the 2021 *Draft NTP Monograph on the State of the Science concerning Fluoride Exposure and Neurodevelopmental and Cognitive Health Effects: A Systematic Review,* 2) a copy of the NASEM Committee's comments on the 2020 draft NTP Monograph with NIEHS/DNTP responses (draft version of Sup01_Monograph), and 3) the [redacted] instructions. The instructions consisted of a preface, charge, instructions for the review, and a series of specific peer-review questions grouped by the following three topics: General Comments, Human Studies, and Studies in Non-Human Animals.

[redacted] were asked to provide their substantive scientific and technical comments and suggestions within the [redacted] form. In addition, they were asked whether they "Agree", "Agree in principle", or "Do not agree" with each NTP conclusion on confidence in a body of evidence.

The [redacted] instructions and specific peer-review questions are reproduced in the pages that follow in black text. [redacted] comments and responses to each question are also provided in black text starting with the words "[redacted] **comments**" in bold font. The NIEHS/DNTP responses have been inserted in blue text following each of the comments beginning with the word "**Response**" in bold font. Formatting has been applied to aid in reading.

The prepublication 2022 NTP Monograph reflects changes made after consideration of the comments from the [redacted] along with all other input received through April of 2022. The prepublication 2022 NTP Monograph was subsequently sent to [redacted] for additional comments. A revised "track changes" version of the monograph was developed in September 2022 titled the "DocMon_Track_Changes_2022_NTP_Monograph." The following bullets describe how edits are documented in the track changes version of the monograph in response to [redacted] comments and [redacted] comments:

- [redacted] For comments related to DocG_Monograph, DocH_Monograph, DocI_Monograph, DocJ_Monograph, and DocK_Monograph:
    - Edits are marked with a comment bubble in the DocMon_Track_Changes_2022_NTP_Monograph that identifies the text in question and briefly describes any revisions.
    - The comment bubble contains the exact text of the [redacted] Comment.
    - The comment bubble also provides a reference to the specific response to comments document with the detailed NIEHS/DNTP response (e.g., comments made in response to this [redacted] would be marked "*see DocK_Monograph for detailed response*").
- [redacted] For comments DocA1_Monograph, DocA2_Monograph, DocB1_Monograph; DocB2_Monograph, and DocC_Monograph through DocF_Monograph:
    - Edits are marked in track changes format in the DocMon_Track_Changes_2022_NTP_Monograph.
    - A comment bubble has been added to the text in question containing the exact text of the [redacted] Comment.
    - The comment bubble also provides a reference to the specific response to comments document with the detailed NIEHS/DNTP response.

Preliminary comments on the draft NTP monograph prepared by the peer revie [redacted] are noted below. These preliminary comments are not binding and should not be construed to represent NTP determination or policy.

National Toxicology Program
NTP Monograph Letter Peer-Review Panel
Draft NTP Monograph on the State of the Science Concerning Fluoride Exposure and Neurodevelopmental and Cognitive Health Effects: A Systematic Review

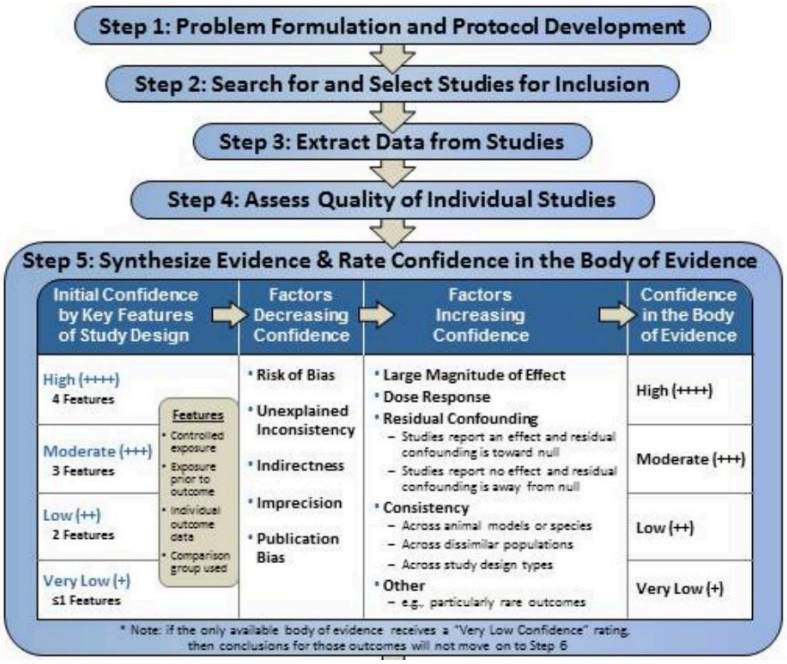
National Institute of Environmental Health Sciences
Research Triangle Park, NC

February 11, 2022

Fluoride State of the Science Document Review Form

[redacted]

Preface:
The objective of this evaluation was to conduct a systematic review of the published literature regarding the potential for exposure to fluoride to affect neurodevelopment and cognition in humans. The evaluation presented in the draft *NTP Monograph on the State of the Science Concerning Fluoride Exposure and Neurodevelopmental and Cognitive Health Effects* represents a comprehensive and current assessment. The methods used are from the Handbook for Conducting a Literature-Based Health Assessment Using OHAT Approach for Systematic Review and Evidence Integration, which presents a seven-step framework for systematic review and evidence integration. Please note: this evaluation stops at step 5 of the systematic review process and does not proceed to step 6 to translate the confidence rating for the body of evidence into a level of evidence for health effects (see Figure 2 from the handbook).



Charge:

(1) Comment on the technical accuracy and whether the draft *NTP Monograph on State of the Science Concerning Fluoride Exposure and Neurodevelopmental and Cognitive Health Effects* is clearly stated, and objectively presented.

(2) Determine whether the scientific evidence supports the NTP's confidence ratings for the bodies of evidence regarding neurodevelopmental and cognitive health effects associated with exposure to fluoride.

Instructions for Review:
All materials for this review are available in the Electronic Council Book (ECB). You will receive the specific URL and a password for accessing the ECB.

This evaluation identified 159 human studies relevant for assessing neurological health effects of exposure to fluoride; however, many studies included only secondary outcomes (e.g., 55 studies of thyroid hormones that were investigated as a potential mechanism). The scientific evidence in children and adults was evaluated separately to address potential differences in the health impact of fluoride exposure during development versus adulthood. Several studies evaluated learning and memory (n = 8 studies) or other cognitive developmental effects (e.g., total neurobehavioral scores and total mental capacity index in children, cognitive impairment in adults; n = 14 studies). Sixty-six human studies investigated IQ in children. Nineteen of the 66 IQ studies were determined to have low potential for bias and therefore, were categorized as "low risk of bias". Please give special attention to our assessment of these 19 studies.

- The 19 studies are available as PDFs and organized alphabetically in a folder on the ECB.
- All other studies are provided in the Health Assessment Workspace Collaborative, or HAWC database under the "studies list" tab, also organized alphabetically. You will also be provided a username and password for HAWC that will give you [redacted] permissions to access the PDFs in HAWC along with visualizations and other study information for this project at the following link (https://hawcproject.org/study/assessment/405/).

Please provide your substantive scientific and technical comments and suggestions within this [redacted] form. Identify and provide the rationale or scientific support for proposed changes or suggestions where possible.

If necessary, you can also provide additional editorial comments and recommendations for improving the report outside your specific charge questions (this form) within the draft report itself. Please note that only those comments included on the [redacted] form will be considered part of NTP's peer review report.

A. General Comments

K.1:

1. Please comment on whether the scientific information presented in the draft monograph, including presentation of data in tables and figures, is technically correct, and clearly and objectively presented. Please suggest any improvements.

   [redacted] **Comments:** The data was clearly presented and put together. In particular, the tables and figures are helpful. A few particular suggestions for the tables are mentioned in sections below.

   Response: No change requested

   - No response necessary because [redacted] feedback on tables is addressed where detailed suggestions are presented below.

2. Please identify any information that should be added or deleted.

   K.2: [redacted] **Comments:** It might be useful to have reminder, or reference back to the section in the text where the risk of bias information for human and animal studies is described in the methods (page 18), prior to presentation of the low risk of bias results for humans (page 28) and animals (page 67).

   Response: Agree (change made)

   - At the beginning of the *Low Risk-of-bias IQ Studies* section, we added the parenthetical text in the quote below to refer readers back to the *Methods* section that describes the risk-of-bias assessment for human studies; however, we determined that a similar reference back to risk-of-bias methods would be less helpful for the *Animal Learning and Memory Data* section, as the animal section does not discuss animal studies in terms of risk-of-bias status.

     "Nineteen studies (3 longitudinal prospective cohort and 16 cross-sectional studies) with low potential for bias evaluated the association between fluoride exposure and IQ in children (see Quality Assessment of Individual Studies section for methods on determining which studies pose low risk of bias)."

   K.3: [redacted] **Comments:** Additionally, it might be helpful to identify a limited set of confounder as required for evaluation. For example, those included in Figure 6 do not include all described in Table 6, and in fact [redacted] not present an important one: parental educational attainment.

   Response: Agree (edited for clarity)

   - Age, sex, and socioeconomic status (SES) are identified as the limited set of key covariates/potential confounders in the *Risk-of-bias Considerations for Human Studies* section. Each of these covariates had to be addressed in any human study of fluoride and cognitive neurodevelopmental health effects to be considered as low risk of bias for confounding. Other covariates may be considered important potential confounders depending on the specific study population and/or outcome assessed. We note that maternal education is listed in this section as a measure of SES. To provide further clarity that parental education is captured under SES in Figure 6, we added a footnote to Figure 6 that states, "*Covariates considered measures of SES include SES scaled scores, household/family income, child education, caretaker/parental education, and occupation/employment*."

B. Human studies

I. **Fluoride exposure and children's IQ**

K.4:

1. Comment on whether the approach described in the methods to search for and select human studies on neurodevelopmental or cognitive function effects associated with fluoride exposure was appropriate for evaluating potential effects of fluoride exposure on measures of IQ in children.

[redacted] **Comments:** This monograph is clearly written and nicely uses tables and figures to display the search criteria and key information points. Furthermore, the level of detail in the methods provides an excellent path forward for understanding exact terms and criteria implemented.

Response: No change requested

- No response necessary.

K.5:

2. Comment on whether the approach used to assess risk of bias was clearly described and appropriately applied to the set of studies designated as "low risk of bias."

[redacted] **Comments:** As stated above, the methods section for this monograph is exemplary. Application of the criteria was clear and appropriate.

Response: No change requested

- No response necessary.

K.6:

3. Comment on assessment of the human studies with regard to:
How findings from individual studies designated as "low risk of bias" were interpreted.
How the overall set of confounders across the body of evidence from children's IQ studies was considered and presented.

[redacted] **Comments:** See comment below on parental IQ

Response: No change requested

- No response necessary because the question of parental IQ is addressed under [redacted] comment for question 4 below.

K.7:

How the confidence rating in the body of evidence was developed and supported.

4. ***NTP concludes a rating of moderate confidence in the body of evidence for lower IQ in children associated with fluoride exposure***.

☐ Agree
☒ Agree in principle with the exception(s) listed below:
☐ Do not agree because:

[redacted] **Comments:** [redacted] am concerned that only one study in the low risk of bias category included parent IQ as a potential confounder. Given the known heritability of IQ, and established connections between socio-economic status (SES) and performance

testing, and SES and educational attainment, substantial confounding may be present. Of note, Figure 6 does not include parental educational attainment, which may be a proxy for an IQ related measure (or those via inherited variation) when a direct measure of IQ was not collected, though is mentioned in Table 6. Additionally, it is notable that many of the low risk of bias studies are cross-sectional and provide limited information regarding temporality and timing of exposure.

Response: Agree (edited for clarity)

- [redacted] made several comments related to this question, and for the first, we agree that parental IQ is important. We also agree that educational attainment (and SES) may be proxy measures of parental IQ. Therefore, parental IQ was considered indirectly addressed if a study accounted for parental educational attainment and/or SES. Figure 6 does not specifically include parental educational attainment because it was considered as a measure of SES. For clarification, we added a footnote to Figure 6 of the Sup02_2022_Prepublication_NTP_Monograph that lists the covariates identified in the studies included in Figure 6 that were considered measures of SES as follows:

  "Covariates considered measures of SES include SES scaled scores, household/family income, child education, caretaker/parental education, and occupation/employment."

- We disagree that many of the low risk-of-bias cross-sectional studies provide limited information regarding temporality and timing of exposure for determining the initial confidence rating. Most of the low risk-of-bias cross-sectional studies (12 of 16) did provide indicators of prior exposure (e.g., by providing prevalence of dental fluorosis, limiting study populations to subjects who lived in the same area for long periods of time). Evidence that exposure occurred prior to the outcome of interest increases the confidence in results (see Figure 1 in Sup02_2022_Prepublication_NTP_Monograph) and any potential association reported in these studies.

II. **Fluoride exposure and non-IQ neurodevelopmental or cognitive effects in children**

K.8:

1. Comment on whether the approach described in the methods to search for and select human studies on neurodevelopmental or cognitive function effects associated with fluoride exposure was appropriate for evaluating potential effects of fluoride exposure on non-IQ neurodevelopmental or cognitive effects in children.

[redacted] **Comments:** The methods were clearly described.

Response: No change requested

- No response necessary.

K.9:

2. Comment on whether the approach used to assess risk of bias for studies in children on non-IQ neurodevelopmental or cognitive effects was clearly described and appropriately applied.

[redacted] **Comments:** The approach for risk of bias was clearly described. The report might benefit from additional explanation of performance-based vs. reporter-based metrics of non-IQ outcomes, relative clinical importance, and interpretation.

Response: Agree (no change)

- *Appendix E* in the Sup02_2022_Prepublication_NTP_Monograph describes study-specific considerations for the risk-of-bias evaluation, including whether outcomes were assessed based on test performance or reporting, and the basis for the risk-of-bias rating. The following two excerpts from *Appendix E* illustrate how reporter-based and performance-based metrics of non-IQ outcomes were considered, respectively, in the risk-of-bias ratings and explanations.

  Excerpt 1

  "Outcome:

  Rating: Probably high risk of bias (–)

  Summary: The primary outcome variable, diagnosis of a learning disability by a health professional, was based on a single item from a household survey asked to all respondents: "Do you have a learning disability?" Answer options were: "yes," "no," "don't know," or the participant refused to answer. For Cycle 2, those who indicated having a learning disability were also asked what kind, with the answer options of: "ADD," "ADHD," "dyslexia," or "other." This question was omitted in Cycle 3, and the reason for omission was not described. Parents or guardians answered all questions for children aged 3–11 years, while children 12 years and older answered questions themselves. The self-reporting of a learning disability did not appear to have been confirmed by medical records or a health professional (– for methods based on self-report of diagnosis by a health care professional; also, in Cycle 3, no specific disabilities were described). Blinding was not a concern as spot urine samples were sent to a separate lab, and self-reports would not have knowledge of their urine or tap water exposure level (+ for blinding). Overall rating = –.

  Basis for rating: Probably high risk of bias based on indirect evidence that the outcome was measured using an insensitive method in the study population."

  Excerpt 2

  "Outcome:

  Rating: Definitely low risk of bias (++)

  Summary: Neurodevelopment was assessed with the Bayley Scales of Infant Development II (BSDI-II) that was noted to be reliable and valid for evaluating children from 3 months to 5 years of age. The average age of children assessed was 8 months, with a range of 3–15 months) (++ for methods). The study report stated that a trained psychologist who was blinded about the mother's fluoride exposure evaluated the infants at home (++ for blinding). Overall rating for methods and blinding = ++.

  Basis for rating: Definitely low risk of bias based on direct evidence that the outcome was assessed using instruments that were valid and reliable in the study

population, and that the outcome assessor was blind to participants' fluoride exposure."

3. Comment on assessment of the human studies with regard to:
How findings from individual "low risk of bias" studies were interpreted.

How the confidence rating in the body of evidence was developed and supported.

K.10:

4. ***The NTP concludes a rating of low confidence in the body of evidence for decreases in measures of other neurodevelopmental or cognitive effects in children associated with fluoride exposure***.

☒ Agree
☐ Agree in principle with the exception(s) listed below:
☐ Do not agree because:

Response: No change requested

- [redacted] agreed with the low confidence rating.

Note: [redacted] only provided comments on the questions above. [redacted] indicated that they had reviewed the Sup03_2021_draft_NTP_Monograph and provided comments under Question A. "General Comments" and Sections I (Fluoride exposure and children's IQ) and II (Fluoride exposure and non-IQ neurodevelopmental or cognitive effects in children) of Question B. "Human Studies". However, they did not have time to provide comments on the remaining sections.

III. **Fluoride exposure and cognitive effects in adults**

1. Comment on whether the approach described in the methods to search for and select human studies on neurodevelopmental or cognitive function effects associated with fluoride exposure was appropriate for evaluating potential effects of fluoride exposure on cognitive effects in adults.

2. Comment on whether the approach used to assess risk of bias for studies in adults on cognitive effects was clearly described and appropriately applied.

3. Comment on assessment of the human studies with regard to:
How findings from individual studies were interpreted.

How the confidence rating in the body of evidence was developed and supported.

4. ***The NTP concludes a rating of low confidence in the body of evidence for changes in cognitive effects in adults with fluoride exposure***.

☐ Agree
☐ Agree in principle with the exception(s) listed below:
☐ Do not agree because:

C. Studies in non-human animals

The NTP agrees with the comments of the NASEM committee (NASEM 2020, 2021) concerning the overall poor quality of the experimental animal database on fluoride exposure and

neurodevelopmental effects, with many studies suffering from major reporting deficiencies. As indicated above, the monograph focuses on the large human epidemiology database because it directly addresses the question of whether fluoride affects human neurodevelopment. Therefore, based on the recommendations of the NASEM committee, the experimental animal section and risk of bias details have been removed from this monograph and ***the NTP concludes that the scientific evidence from experimental animal data are inadequate to inform whether fluoride exposure is associated with cognitive effects (including cognitive neurodevelopmental effects) in humans***.

- ☐ Agree
- ☐ Agree in principle with the exception(s) listed below:
- ☐ Do not agree because:

Title: Association between fluoride exposure and children's intelligence: A systematic review and meta-analysis

Authors:

Kyla W. Taylor[1], PhD (corresponding author)

Sorina E. Eftim[2], PhD

Christopher A. Sibrizzi[2], MPH

Robyn B. Blain[2], PhD

Kristen Magnuson[2], MESM

Pamela A. Hartman[2], MEM

Andrew A. Rooney[1], PhD

John R. Bucher[1], PhD

Affiliations: [1]Division of the National Toxicology Program, National Institute of Environmental Health Sciences, National Institutes of Health, Research Triangle Park, NC, USA; [2]ICF, Fairfax, VA, USA.

Corresponding author contact:

Kyla W. Taylor, PhD

530 Davis Drive, Morrisville NC 27719

Email: kyla.taylor@nih.gov

Phone: 919-530-0061

Abstract word count: 325

Manuscript word count: 3,751

Abstract

IMPORTANCE Water and water-based beverages are the main source of systemic fluoride intake; however, an individual's total exposure to fluoride also reflects contributions from other sources such as food, dental products, industrial emissions, and some pharmaceuticals. Previous meta-analyses suggest that exposure to fluoride adversely affects children's intelligence.

OBJECTIVE To perform a systematic review and meta-analysis to investigate associations between fluoride exposure and children's intelligence.

DATA SOURCES BIOSIS, EMBASE, PsychINFO, PubMed, Scopus, Web of Science, CNKI, and Wanfang databases were searched for relevant literature published up to November 2021.

STUDY SELECTION Inclusion criteria were assessment of cognitive outcomes, fluoride exposure, and statistical data on effect size.

DATA EXTRACTION AND SYNTHESIS Meta-analysis of Observational Studies in Epidemiology (MOOSE) reporting guidelines were followed for data extraction. The quality of individual studies was evaluated for risk of bias using a standardized tool. Pooled standardized mean differences (SMDs) and regression coefficients were estimated with random-effects models.

MAIN OUTCOMES AND MEASURES Children's intelligence levels reflected by intelligence quotient (IQ) scores.

RESULTS The meta-analysis of 55 studies (N = 18,845 children) with group-level exposures found that, when compared to children exposed to lower fluoride levels, children exposed to higher fluoride levels had lower mean IQ scores (pooled SMD: −0.46; 95% CI: −0.55, −0.37; p-value < 0.001). There was a dose-response relationship between group-level fluoride exposure measures and mean children's IQ. The meta-analysis of studies that reported individual-level measures of fluoride and children's IQ scores found a decrease of 1.81 points (95% CI: −2.80, −0.81; p-value < 0.001) per 1-mg/L increase in urinary

Commented [I1]: See Doc01_Meta-analysis, 1.B., page 1

Commented [I2]: See Doc02_Meta-analysis, 2.B., page 1

Commented [I3]: See Doc08_Meta-analysis, 8.G., page 5 and 6

Commented [I4]: See Doc01_Meta-analysis, 1.C., page 1

Commented [I5]: See Doc08_Meta-analysis, 8.K., page 7 and 8

Commented [I6]: See Doc01_Meta-analysis, 1.D., page 2

fluoride. Overall, the direction of the association was robust to stratification by study quality (high *vs*. low risk of bias), sex, age group, outcome assessment, study location, exposure timing, and exposure metric.

CONCLUSIONS AND RELEVANCE This meta-analysis confirms results of previous meta-analyses and extends them by including newer, more precise studies with individual-level exposure measures. The consistency of the data supports an inverse association between fluoride exposure and children's IQ.

Commented [I7]: See Doc08_Meta-analysis, 8.Q., page 10

Commented [I8]: See Doc06a_Meta-analysis, 6a.A., page 1

Introduction

Fluoride from natural sources occurs in some community water systems and, in the United States and some other countries, fluoride is added to public drinking water systems for the prevention of tooth decay. Water and water-based beverages are the main source of systemic fluoride intake; however, an individual's total exposure also reflects contributions from fluoride in other sources such as food, dental products, industrial emissions, and some pharmaceuticals.[1] Accumulating evidence suggests that fluoride exposure may affect brain development. A 2006 report from the National Research Council (NRC) concluded that high levels of naturally occurring fluoride in drinking water may be of concern for neurotoxic effects.[2] This report was largely based on studies from endemic fluorosis areas in China that had limitations in study design or methods (e.g., high risk of bias). Following the NRC review, more evidence has emerged in studies from India, Iran, Pakistan, New Zealand, Spain, and Canada (**Figure 1**). Two previous meta-analyses[3, 4] found an association between high fluoride exposure and lower children's IQ; however, many of the studies in these meta-analyses lacked the information necessary to evaluate study quality and all used group-level estimates of fluoride exposure. Since the most recent meta-analysis,[4] eleven new studies on exposure to fluoride and children's IQ have been published, including three prospective North American birth cohort studies[5-7] that used individual-level measures of maternal and children's urinary fluoride.

To incorporate this newer evidence, and to complement a larger systematic review[8] that concluded there is moderate confidence in the evidence of an inverse association between fluoride exposure and children's IQ, we conducted a meta-analysis of studies that provided group-and individual-level fluoride exposure measurements in relation to children's IQ scores.

Methods

The search, selection, extraction, and risk-of-bias evaluation of studies for this meta-analysis were part of a larger systematic review.[8] Brief methods are outlined below with detailed methods available in the protocol[9] and the **Supplemental Materials**.

Commented [I9]: See Doc03_Meta-analysis, 3.B. (page 1) and 3.C. (page 1, 2, and 3)

Commented [I10]: See Doc08_Meta-analysis, 8.F., page 5

Commented [I11]: See Doc01_Meta-analysis, 1.F and 1.G., page 3

Commented [I12]: See Doc01_Meta-analysis, 1.F., page 3

Commented [I13]: See Doc02_Meta-analysis, 2.C., page 1 and 2

Commented [I14]: See Doc02_Meta-analysis, 2.Q., page 5 and 6

Commented [I15]: See Doc05_Meta-analysis, 5.H., page 8 and 9

Commented [I16]: See Doc05_Meta-analysis, 5.H., page 8 and 9

Commented [I17]: See Doc02_Meta-analysis, 2.A. (page 1) and 2.Q. (page 5 and 6)

Systematic literature review

Literature searches were conducted in BIOSIS, EMBASE, PsychINFO, PubMed, Scopus, Web of Science, CNKI, and Wanfang databases through November 2021, without language restrictions. Search strategies are available in the protocol.[9]

Commented [I18]: See Doc01_Meta-analysis, 2.D., page 2

Study selection

To be eligible for inclusion, individual study publications had to satisfy review eligibility criteria outlined in the protocol.[9] References retrieved from the literature search were independently screened by two reviewers by title and abstract followed by full-text review. Studies that estimated the association between exposure to fluoride (based on environmental measures or biomonitoring data, reported as either individual-level or group-level measurements) and a quantitative measure of children's intelligence were included. Studies that did not report quantitative effect estimates (mean outcome measures or regression coefficients), measures of variability (95% confidence intervals [CIs], standard errors [SEs], or standard deviations [SDs]), or numbers of participants were excluded. Studies with missing measures of variability but with reported p-values for differences were included, and SDs were calculated using the approach in the Cochrane Handbook for Systematic Reviews.[10] To avoid sample overrepresentation, if the same cohort was followed at multiple timepoints resulting in multiple study publications,[11, 12] only the study publication that included the largest number of participants was included in this meta-analysis (see **eTable 1**).

Commented [I19]: See Doc02_Meta-analysis, 2.E., page 2

Commented [I20]: See Doc05_Meta-analysis, 5.L., page 10

Data extraction

Data were collected from included studies by one extractor and verified by a second extractor. Data were extracted in Health Assessment Workspace Collaborative (HAWC), an open source, web-based application for data extraction elements listed in the protocol. Data extraction results for included studies are publicly available and downloadable (https://hawcproject.org/assessment/405/).

Quality assessment: Risk-of-bias evaluation

Quality of individual studies, also called "risk of bias," was assessed using the National Toxicology Program's Office of Health Assessment and Translation approach.[13] Studies were independently evaluated by two trained assessors who answered risk-of-bias questions following prespecified criteria detailed in the protocol.[9] Risk-of-bias questions concerning confounding, exposure characterization, and outcome assessment were considered key. If not addressed appropriately, these questions were thought to have the greatest potential impact on the results.[9] The other risk-of-bias questions were used to identify other concerns that may indicate serious risk-of-bias issues (e.g., selection bias, statistical analysis). No study was excluded from the meta-analysis based on concerns for risk of bias; however, subgroup analyses were conducted with and without high risk-of-bias studies (i.e., studies rated "probably high" risk of bias for at least two key risk-of-bias questions or "definitely high" for any single question) to assess their impact on the results.

Commented [I21]: See Doc01_Meta-analysis, 1.H., page 3

Commented [I22]: See Doc02_Meta-analysis, 2.F., page 2 and 3

Commented [I23]: See Doc06a_Meta-analysis, 6a.B., page 2

Statistical analysis

We conducted the following analyses, planned *a priori* in the protocol: (1) *a mean-effects meta-analysis*, (2) *a dose-response mean-effects meta-analysis*, and (3) *a regression slopes meta-analysis.* We also conducted several subgroup and sensitivity analyses.

The *mean-effects meta-analysis* included studies that reported mean IQ scores and group-level exposures for at least one exposed and one reference group. The effect estimates in the primary *mean-effects meta-analysis* were the standardized mean differences (SMDs) for heteroscedastic population variances.[14-16] The SMDs were calculated from the difference in mean IQ scores between an exposed group and a reference group. If mean IQ scores were reported for multiple exposure groups within a single study, the highest exposure group was considered the exposed group and the lowest exposure group was considered the reference group. A sensitivity analysis was performed to evaluate the impact of all exposure groups combined compared to a reference group (see additional details on the approach, effect estimation, and study selection in the **Supplemental Materials**). Predefined subgroup analyses

Commented [I24]: See Doc06b_Meta-analysis, 6b.N., page 12 and 13

Commented [EAM25]: See Doc01_Meta-analysis, 1.J., page 4.
Note: Current language reflects revisions to the earlier version of document.

were stratified by risk of bias (high or low), study location (e.g., country), outcome assessment, exposure matrix (e.g., urinary fluoride or water fluoride concentrations), sex, and age group. To further evaluate potential sources of heterogeneity, we conducted meta-regression analyses using mean age in years (from the age range reported in each study) and year of publication in each study.

To determine whether the data support an exposure-response relationship, we conducted a *dose-response mean-effects meta-analysis.* This analysis included studies from the *mean-effects meta-analysis* that reported fluoride exposure levels and used a one-step approach as described in the protocol.[9, 17, 18] This approach uses linear mixed models to analyze all available mean effect estimates for the reference group and one or more exposure group and estimates a pooled dose-response curve using a restricted maximum likelihood estimation method. Model comparison was based on the maximum likelihood Akaike information criterion (AIC).[19] We also examined whether there was a dose-response relationship at lower exposure levels that corresponded with the U.S. Environmental Protection Agency drinking water standards[20] and World Health Organization drinking water guidelines[21] (details provided in the **Supplemental Materials**).

The *regression slopes meta-analysis* included studies that reported regression slopes to estimate associations between individual-level fluoride exposure and children's IQ. The primary regression slopes meta-analysis used regression slopes from models that adjusted for potential confounders. If results from multiple models were reported within a single study, either the most adjusted results or the main model results as presented by the study authors were selected. The study outcomes were evaluated with respect to a 1-mg/L unit increase in water or urinary fluoride, or 1-mg/day fluoride intake.

Data from individual studies were pooled using a random-effects model.[22] Heterogeneity was assessed by Cochran's Q test[23] and the I^2 statistic.[24] Forest plots were used to display results and to examine possible heterogeneity between studies. Potential publication bias was assessed by developing funnel plots and performing Egger regression on the estimates of effect size.[25-27] If publication bias was

Commented [I26]: See Doc02_Meta-analysis, 2.H., page 3

Commented [I27]: See Doc01_Meta-analysis, 1.E. (page 2) and 1.K. (page 4)

Commented [I28]: See Doc02_Meta-analysis, 2.I., page 3 and 4

Commented [I29]: See Doc06a_Meta-analysis, 6b.R., page 15 and 16

Commented [I30]: See Doc05_Meta-analysis, 5.J., page 9 and 10

present, trim-and-fill methods[28, 29] were used to estimate the number of missing studies and to predict the impact of the hypothetical "missing" studies on the pooled effect estimate. Subgroup analyses were performed to investigate sources of heterogeneity. Subgroup analyses were stratified by risk of bias (high or low), study location (e.g., country), outcome assessment, exposure matrix (e.g., urinary fluoride or water fluoride concentrations), pre- or post-natal exposure, and sex.

Statistical analyses were performed using the software STATA version 17.0[30] with the *combine, meta esize*, *meta set*, *meta summarize*, *drmeta*, *meta funnel*, *meta bias*, *meta trimfill* and *metareg* packages.[31]

Results

Study sample

Results of the study identification process are provided in **eFigure 1.** Characteristics of the 60 publications included in the meta-analysis are shown in **Table 1** (see **eTable 1** for list of excluded publications). A total of 55 publications reported mean IQ scores for group-level exposures. Eleven publications reported regression slopes for individual-level exposures based on urinary or water fluoride concentrations.[5-7, 11, 12, 32-37] Additional details on study characteristics are provided in the **Supplemental Materials**. Results from risk-of-bias evaluations are presented in **eFigure 2a** and **eFigure 2b**. Study-specific effect estimates used in the meta-analysis are presented in **eTable 2**.

Mean-effects meta-analysis

The meta-analysis of 55 studies (45 high risk-of-bias studies and 10 low risk-of-bias studies) that provided mean IQ scores shows that, when compared to children exposed to lower levels of fluoride, children exposed to higher fluoride levels had statistically significantly lower IQ scores (random-effects pooled SMD, −0.46; 95% CI: −0.55, −0.37; p-value < 0.001) (**Table 2, Figure 2**). There was evidence of high heterogeneity (I^2 = 87%, p-value < 0.001; **Table 2**) and publication bias (funnel plot and Egger's p-value < 0.001, Begg's p-value = 0.031; **eFigures 3 and 4**). Adjusting for possible publication bias

Commented [I31]: See Doc05_Meta-analysis, 5.I., page 9

Commented [I32]: See Doc06b_Meta-analysis_6b.V., page 18 and 19

Commented [I33]: See Doc02_Meta-analysis, 2.N., page 4

Commented [I34]: See Doc02_Meta-Analysis, 2.J. (page 4) and Doc08_Meta-analysis, 8.M. (page 8 and 9)
Note: Changes in study numbers from reviewer text reflects updated literature search.

Commented [I35]: See Doc02_Meta-analysis, 2.K., page 4

through trim-and-fill analysis suggested the imputation of seven additional studies to the right side, with an adjusted pooled SMD of – 0.36 (95% CI: −0.46, −0.26) (**eFigures 5 and 6**). The pattern of results across the 55 studies was consistent; 52 (95%) reported an inverse association with SMDs ranging from −5.34 (95% CI: −6.34, −4.34) to −0.04 (95% CI: −0.45, 0.36) (**Figure 2**). The(95% CI: −0.19, 0.21),[6] 0.01 (95% CI: −0.19, 0.22),[38] and 0.13 (95% CI: −0.16, 0.42).[5] Three studies[39, 40, 41 [translated in Li et al. 2008b]] lacked clear descriptions of their intelligence assessment methods; however, sensitivity analyses did not reveal substantial changes in the pooled SMD estimate when these studies were excluded or when a study[43] that reported the cognitive subset of evaluations using Bayley and McCarthy tests was included (**eTable 3**).

Among the low risk-of-bias studies (n = 10), [5, 6, 11, 32, 33, 36, 44-47] the random-effects pooled SMD was −0.22 (95% CI: −0.39, −0.05; p-value = 0.011) with high heterogeneity (I^2 = 83%) (**Table 2 and eFigure 7**). There was no evidence of publication bias (funnel plot and Egger's p-value = 0.93; **eFigures 8 and 9**). Among the high risk-of-bias studies (n = 45), the random-effects pooled SMD was −0.52 (95% CI: −0.63, −0.42; p-value < 0.001) with high heterogeneity (I^2 = 86%,) (**Table 2 and eFigure 7**). There was evidence of publication bias among the high risk-of-bias studies (funnel plot and Egger's p-value < 0.001; **eFigures 8 and 9**); adjusting for possible publication bias through trim-and-fill analysis supports the results with an adjusted pooled SMD estimate of −0.37 (95% CI: −0.48, −0.25) (**eFigures 10 and 11**). Subgroup analyses by sex, age group, study location, outcome assessment type, and exposure assessment type further support the consistent and robust pattern of an inverse association between fluoride exposure and children's IQ (**Table 2, eFigures 12-16**). The subgroup and meta-regression analyses did not explain a large amount of the overall heterogeneity; however, the degree of heterogeneity was lower We also examined whether there was a dose-response relationship at lower exposure levels that corresponded with the U.S. Environmental Protection Agency drinking water standards[20] and World restricted to Iran (I^2=56%), children ages 10 and older (I^2=68%), and girls (I^2=76%) (see **Supplemental Materials**).

Commented [I36]: See Doc06b_Meta-analysis, 6b.C., page 2 and 3

Commented [I37]: See Doc05_Meta-analysis, 5.J., page 9 and 10

Commented [I38]: See Doc01_Meta-analysis, 1.I., page 3

Commented [EAM39]: See Doc02_Meta-analysis, 2.L., page 4.
Note: Changes in study numbers from review text reflects updated literature search.

Commented [I40]: See Doc06b_Meta-analysis, 6b.G., page 7and 8

The sensitivity analysis to evaluate the impact of combining all exposed groups and comparing them to the reference group did not appreciably change the effect estimates (**eTable 3**). Sensitivity analyses that removed an outlier study[39] or a study with an unspecified IQ test[41 [translated in Li et al. 2008b]] also did not appreciably change the effect estimates (**eTable 3**).

Commented [I41]: See Doc08_Meta-analysis, 8.H., page 6

Dose-response mean-effects meta-analysis

The *dose-response mean-effects meta-analysis* combining data from 29 studies with group-level fluoride measurements in drinking water (23 high risk-of-bias and 6 low risk-of-bias studies) and 18 studies with group-level mean urinary fluoride levels (9 high risk-of-bias and 9 low risk-of-bias studies) show statistically significantly lower children's IQ scores with increasing fluoride exposures. Based on the linear models, the decrease in mean SMD between exposed and reference groups is −0.15 (95% CI: −0.20, −0.11; p-value < 0.001) for drinking water fluoride levels and −0.16 (95% CI: −0.24, −0.08; p-value < 0.001) for urinary fluoride levels (**eTable 4**). Based on the AIC and likelihood ratio tests, the best model fit was achieved when quadratic or restricted cubic spline exposure levels were added to the linear models for drinking water (**eFigure 17**); the linear model was the best fit for urinary fluoride (**eFigure 18**). Given the small difference in AICs between the different models, and for ease of interpretability, the linear model results were chosen for the purposes of discussion, although results from all models are presented (**eTable 4**). The direction of the associations did not change when the exposed groups were restricted to <4 mg/L or <2 mg/L fluoride in drinking water or fluoride in urine (**eTable 4** and **eTable 5**).

Commented [I42]: See Doc08_Meta-analysis, 8.I., page 6 and 7

Regression slopes meta-analysis

The *regression slopes meta-analysis* includes ten studies with individual-level exposure measures (1 high risk-of-bias and 9 low risk-of-bias studies) (**Table 1**). Each of these studies reported urinary fluoride levels,[5-7, 11, 12, 32-37] two reported fluoride intake,[6, 7] and two reported water fluoride levels.[6, 11] Two studies[7, 12] are not included in the primary meta-analysis they had overlapping populations with already-included studies[6, 11 respectively] (see **Supplemental Materials**). Similarly, three studies reporting scores

based on Bayley assessments[43, 48, 49] were only included in sensitivity analyses (see **Supplemental Materials**).

The overall pooled effect estimate from the nine studies with individual-level urinary fluoride measures shows that a 1-mg/L increase in urinary fluoride is associated with a statistically significant lower IQ score of 1.81 points (95% CI: −2.80, −0.81; p-value < 0.001) with evidence of heterogeneity (I^2 = 77%, p-value < 0.001; **Table 3, eFigure 19**) and no indications of publication bias (**eFigures 20 and 21**). When restricted to only low risk-of-bias studies, the decrease in IQ score was 1.33 points (95% CI: −2.09, −0.57; p-value < 0.001). There was evidence of moderate heterogeneity (I^2 = 46%, p-value < 0.072; **Table 3, eFigure 22**) and no indications of publication bias. The results for fluoride intake and water fluoride levels are available in **Supplemental Materials**.

Subgroup analyses by risk of bias, sex, country, exposure type, outcome assessment type, and pre- or post-natal exposure further support the consistent and robust pattern of an inverse association between fluoride exposure and children's IQ (**Table 3, eFigures 22–27**). The observed heterogeneity in the overall effect estimate was explained by the subgroup analyses, with no significant heterogeneity remaining in analyses of low-risk-of bias studies, by sex, by country, by assessment type, and by exposure timing (**Table 3**). The sensitivity analyses including reporting scores based on Bayley assessments[43, 48, 49] showed no substantial changes in the pooled effect estimates (**eTable 6**).

Discussion

The results of this meta-analysis support a statistically significant association between higher fluoride exposure and lower children's IQ. The direction of the association was robust to stratification by risk of bias, sex, age group, timing of exposure, study location, outcome assessment type, and exposure assessment type. There is also evidence of a dose-response relationship. Although the estimated decreases in IQ may seem small, research on other neurotoxicants has shown that subtle shifts in IQ at the population level can have a profound impact on the number of people who fall within the high and low

Commented [EAM43]: See Doc02_Meta-analysis, 2.M., page 4.
Note: Changes in study numbers from reviewer text reflects updated literature search.

Commented [I44]: See Doc02_Meta-analysis, 2.O., page 5

Commented [I45]: See Doc06b_Meta-analysis, 6b.G., page 7 and 8

Commented [I46]: See Doc02_Meta-analysis, 2.P., page 5

ranges of the population's IQ distribution.[50-54] For example, a 5-point decrease in a population's IQ would nearly double the number of people classified as intellectually disabled.[55]

The results of the *mean-effects meta-analysis* are consistent with two previous meta-analyses that, when comparing children exposed to lower fluoride levels, reported statistically significantly lower IQ scores in children exposed to higher fluoride levels ($p < 0.001$) (**Table 2**). However, this meta-analysis included more recently published studies that were considered low risk of bias and studies with different exposure assessment types. We also found a statistically significant dose-response between lower children's IQ with increasing fluoride exposures as measured in both drinking water (p-value < 0.001) and urine (p-value < 0.001). Associations appeared to be non-linear for drinking water and linear for urine. The Duan et al.[4] meta-analysis reported a significant non-linear dose-response relationship above 3 ppm [3 mg/L] in water. A more recent literature review[56] did not comment on the shape of the dose-response curve; however, based on the three publications from Mexico and Canada,[5-7] the author concluded that the association between maternal urinary fluoride and children's neurotoxicity appeared to be "dose dependent."

Whereas the previously published meta-analyses only included group-level exposures, the *regression slopes meta-analysis* included nine studies with individual urinary fluoride measures, a more precise exposure measure. It also included recent North American prospective cohort studies[5-7] with maternal urinary fluoride levels comparable to those found in the United States.[57] In contrast to urinary fluoride measures, drinking water measures capture only a portion of a person's total exposure to fluoride. Consequently, relying on drinking water levels alone likely underestimates an individual's total exposure to fluoride. For community water systems that add fluoride, the Public Health Service recommends a fluoride concentration of 0.7 mg/L; however, it is important to note that there are regions of the United States where public systems and private wells contain natural fluoride concentrations of more than 2 mg/L.[58] In April 2020, the Centers for Disease Control and Prevention (CDC) estimated that community water systems supplying water with ≥2 mg/L naturally occurring fluoride served 0.31% of the U.S.

Commented [I47]: See Doc08_Meta-analysis, 8.P., page 9 and 10

Commented [I48]: See Doc08_Meta-analysis, 8.N., page 9

Commented [I49]: See Doc01_Meta-analysis, 1.N. and 1.O., page 7

Commented [I50]: See Doc01_Meta-analysis, 1.M., page 6 and 7

population (~1 million people).[59] For the purposes of reducing dental fluorosis, the CDC recommends that parents use an alternative source of water for children aged 8 years and younger and for bottle-fed infants if their primary drinking water contains greater than 2 mg/L of fluoride.[60]

Commented [I51]: See Doc03_Meta-analysis, 3.B. (page 1 and 2) and 3.C. (page 2 and 3)

Commented [I52]: See Doc01_Meta-analysis, 1.Q., page 8

Commented [I53]: See Doc02_Meta-analysis, 2.T., page 6

Strengths and Limitations

Strengths of this meta-analysis include a large body of literature and predefined systematic search and screening process, a risk-of-bias assessment of individual studies, a variety of intelligence assessment methods and exposure matrices, varying exposure levels from multiple study locations, prespecified subgroup analyses, and use of both group-level and individual-level exposure data. The direction of the association is consistent across different analytical approaches and subgroup analyses.

There are also limitations to consider. Most of the studies included in the *mean-effects* and *dose-response mean effects meta-analyses* were considered to have study design and/or methodological limitations. For example, all but three studies were cross-sectional in design. However, among the low risk-of-bias cross-sectional studies, most provided information to suggest that exposure preceded the outcome (e.g., including only children who had lived in the area since birth, or children that had dental fluorosis). In addition, subgroup analyses suggest that the association between higher fluoride exposure and lower IQ was consistent even when restricted to low risk-of-bias studies (see **Table 2** and **eFigure 7** for additional details). Although we conducted subgroup analyses by sex, only 1 of the 14 studies that reported IQ scores separately for boys and girls analyzed fluoride exposure for each sex separately.[6] This is essential for evaluating whether a differential change in IQ by sex may be related to higher susceptibility or higher exposure in that sex. With a couple exceptions, the subgroup analyses in the *mean-effects meta-analysis* did not explain a large amount of the overall heterogeneity. However, the heterogeneity in the *regression slopes meta-analysis* was explained by subgroup analyses. This suggests that the aggregate nature of the *mean-effects meta-analysis* might not be sufficiently sensitive to capture potential sources of heterogeneity, as seen possible when using studies with individual-level data in the *regression slopes meta-analysis*. However, the large number of studies included in the *mean-effects meta-*

Commented [I54]: See Doc02_Meta-analysis, 2.U., page 7

Commented [I55]: See Doc08_Meta-analysis, 8.T., page 11

Commented [I56]: See Doc03_Meta-analysis, 3.E., page 4

Commented [I57]: See Doc06b_Meta-analysis, 6b.O., page 13 and 14

analysis and the consistency in the direction of the association across the analyses make this is less of a concern.

Another limitation of the *mean-effects meta-analyses* is that exposure values are assumed to be the same for each child in an exposure group, either because the study used a community-level water fluoride measure or a median, mean, or midpoint in water or urine as the exposure value. Fluoride exposure may vary considerably depending on individual behaviors and is best captured by individual-level measures of total exposure, such as urinary fluoride measures. Because drinking water measures capture only some of a person's total exposure to fluoride, it is reasonable to assume that some children in the meta-analysis had higher exposure to fluoride and those children may have skewed the mean IQ deficits of the entire group. Urinary fluoride levels include all ingested fluoride and are considered a valid measure to estimate total fluoride exposure.[61, 62] When compared with 24-hour urine samples, spot urine samples are more prone to the influence of timing of exposure (e.g., when water was last consumed, when teeth were last brushed) and can also be affected by differences in dilution. However, correlations between urinary fluoride concentrations from 24-hour samples and spot samples adjusted for urinary dilution have been described,[63] and with one exception[35] all studies in the *regression slopes meta-analysis*, accounted for dilution.

There is inconsistency in which model is the best fit at lower exposure levels (**eTable 4** and **eTable 5**) leading to uncertainty in the shape of the dose-response curve at these levels. More individual-level data would increase our certainty in the shape of the dose-response curve at these lower exposure levels. There are also several limitations to the existing approaches for evaluating potential for publication bias. The funnel plot asymmetry is a subjective assessment and is recommended only when at least 10 studies are included in the meta-analysis.[64] Furthermore, the Egger regression test and Begg's rank tests[25-27] may suffer from inflated type I power and limited power in certain situations.[65] Finally, the small number of studies reporting slopes for association with individual-level exposure data limits the power of the *regression slopes meta-analysis*.

Commented [I58]: See Doc06b_Meta-analysis, 6b.G., page 7 and 8

Commented [I59]: See Doc08_Meta-analysis, 8.E., page 4 and 5

Commented [I60]: See Doc02_Meta-analysis, 2.S., page 6

Commented [I61]: See Doc01_Meta-analysis, 1.M., page 6 and 7

Commented [I62]: See Doc06a_Meta-analysis, 6a.G., page 5

Commented [I63]: See Doc08_Meta-analysis, 8.D., page 3 and 4

Commented [I64]: See Doc03_Meta-analysis, 3.C., page 2 and 3

Commented [I65]: See Doc05_Meta-analysis, 5.I., page 9

This meta-analysis complements a larger systematic review[8] that concluded moderate confidence in the body of evidence that fluoride exposure is associated with lower IQ in children. Confidence would be increased with additional prospective cohort studies with individual urinary fluoride measures. Studies conducted in the United States, which as of the writing of this manuscript were not available, would also be valuable.

Commented [EAM66]: See Doc08_Meta-analysis, 8.B., page 2.

Conclusions

This meta-analysis extends the findings of our larger systematic review that concluded, with moderate confidence, that higher fluoride exposure is associated with lower children's IQ. These findings are consistent with prior meta-analyses and demonstrate that the direction of the association is robust to stratifications by risk of bias, sex, age group, outcome assessment, study location, exposure timing, and exposure measurement (including both drinking water and urinary fluoride). Therefore, the consistency of the data supports an inverse association between fluoride exposure and children's IQ.

Commented [I67]: See Doc06a_Meta-analysis, 6a.C., page 2

Acknowledgements

This work was supported by the Intramural Research Program (ES103316, ES103317) at the National Institute of Environmental Health Sciences (NIEHS), National Institutes of Health, and was performed for NIEHS under contract GS00Q14OADU417 (Order No. HHSN273201600015U). Data sets and Stata code are available upon reasonable request from study authors. We appreciate the helpful comments and input on the draft manuscript provided by Suril Mehta, DrPH, MPH; Kelly Ferguson, PhD, MPH; Allen Wilcox, MD, PhD; and Alison Mogsinger-Reif, PhD. We would like to thank Jonathan Cohen, PhD, who performed independent verification of the dose-response meta-analysis; Cynthia Lin, PhD, and Alexandra Goldstone, MPH, for independent quality control of the data; and Jeremy Frye, MSLS, for conducting the literature search and reference management.

References

1. U.S. Environmental Protection Agency (US EPA). *Fluoride: Exposure and relative source contribution analysis*. U.S. Environmental Protection Agency; 2010. 820-R-10-015. Accessed 19 August 2019. http://www.epa.gov/dwstandardsregulations/fluoride-risk-assessment-and-relative-source-contribution

2. National Research Council (NRC). *Committee on fluoride in drinking water, board on environmental studies and toxicology. Fluoride in drinking water: A scientific review of EPA's standards*. National Research Council; 2006. Accessed 19 August 2019. http://www.nap.edu/catalog/11571/fluoride-in-drinking-water-a-scientific-review-of-epas-standards

3. Choi AL, Sun G, Zhang Y, Grandjean P. Developmental fluoride neurotoxicity: A systematic review and meta-analysis. *Environ Health Perspect*. 2012;120:1362-1368.

4. Duan Q, Jiao J, Chen X, Wang X. Association between water fluoride and the level of children's intelligence: A dose-response meta-analysis. *Public Health*. 2018;154:87-97. doi:https://doi.org/10.1016/j.puhe.2017.08.013

5. Bashash M, Thomas D, Hu H, et al. Prenatal fluoride exposure and cognitive outcomes in children at 4 and 6-12 years of age in Mexico. *Environ Health Perspect*. 2017;125(9):1-12. doi:https://doi.org/10.1289/ehp655

6. Green R, Lanphear B, Hornung R, et al. Association between maternal fluoride exposure during pregnancy and IQ scores in offspring in Canada. *JAMA Pediatr*. 2019:E1-E9.

7. Till C, Green R, Flora D, et al. Fluoride exposure from infant formula and child IQ in a Canadian birth cohort. *Environ Int*. 2020;134:105315. doi:https://doi.org/10.1016/j.envint.2019.105315

8. National Toxicology Program (NTP). *Revised draft NTP Monograph on the systematic review of fluoride exposure and neurodevelopmental and cognitive health effects*. National Toxicology Program; 2020. https://www.nationalacademies.org/event/10-19-2020/docs/DDA97C9170D1A255D69C004CEB77B698E8D005011EFB

9. National Toxicology Program (NTP). *Protocol for systematic review of effects of fluoride exposure on neurodevelopment*. U.S. Department of Health and Human Services, Public Health Service, National Institutes of Health; 2020. https://ntp.niehs.nih.gov/ntp/ohat/fluoride/ntpprotocol_revised20200916_508.pdf

10. Higgins JT, Thomas J, Chandler J, et al., eds. *Cochrane Handbook for systematic reviews of interventions version 6.0 (updated July 2019)*. Cochrane; 2019.

11. Yu X, Chen J, Li Y, et al. Threshold effects of moderately excessive fluoride exposure on children's health: A potential association between dental fluorosis and loss of excellent intelligence. *Environ Int*. 2018;118:116-124. doi:https://doi.org/10.1016/j.envint.2018.05.042

12. Wang M, Liu L, Li H, et al. Thyroid function, intelligence, and low-moderate fluoride exposure among Chinese school-age children. *Environ Int*. 2020b;134:105229. doi:https://doi.org/10.1016/j.envint.2019.105229

13. National Toxicology Program (NTP). *OHAT risk of bias rating tool for human and animal studies*. U.S. Department of Health and Human Services, Public Health Service, National Institutes of Health; 2015.

14. Bonett DG. Confidence intervals for standardized linear contrasts of means. *Psychol Methods*. 2008;13(2):99-109. doi:https://doi.org/10.1037/1082-989x.13.2.99

15. Hedges LV, Olkin I. *Statistical methods for meta-analysis*. Academic Press; 1985.

16. Rosenthal R. Parametric measures of effect size. In: Cooper H, Hedges LV, eds. *The Handbook of Research Synthesis*. Russell Sage Foundation; 1994.

17. Crippa A, Thomas I, Orsini N. A pointwise approach to dose-response meta-analysis of aggregated data. *Int J Stat Med Res*. 2018;7(2):25-32.

18. Crippa A, Discacciati A, Bottai M, Spiegelman D, Orsini N. One-stage dose–response meta-analysis for aggregated data. *Stat Methods Med Res*. 2019;28(5):1579-1596. doi:https://doi.org/10.1177/0962280218773122

19. Müller S, Scealy JL, Welsh AH. Model selection in linear mixed models. *Statist Sci*. 2013;28(2):135-167. doi:https://doi.org/10.1214/12-STS410

20. U.S. Environmental Protection Agency (US EPA). *New fluoride risk assessment and relative source contribution documents*. Office of Water; 2011. EPA-822-F-11-001. https://www.epa.gov/sites/default/files/2019-03/documents/fluoride-risk-assess-factsheet.pdf

21. World Health Organization (WHO). *Guidelines for drinking-water quality. Fourth edition*. 2011. https://apps.who.int/iris/bitstream/handle/10665/44584/9789241548151_eng.pdf?sequence=1&isAllowed=y&ua=1

22. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials*. 1986;7(3):177-88. doi:https://doi.org/10.1016/0197-2456(86)90046-2

23. Cochran WG. The combination of estimates from different experiments. *Biometrics*. 1954;10(1):101-129. doi:https://doi.org/10.2307/3001666

24. Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ*. 2003;327(7414):557-560. doi:https://doi.org/10.1136/bmj.327.7414.557

25. Begg CB, Mazumdar M. Operating characteristics of a rank correlation test for publication bias. *Biometrics*. 1994;50(4):1088-1101. doi:https://doi.org/10.2307/2533446

26. Egger M, Smith G, Schneider M, Minder C, eds. *Systematic reviews in health care: Meta-analysis in context*. BMJ Publishing Group; 2008.

27. Egger M, Smith GD, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. *BMJ*. 1997;315(7109):629-634. doi:https://doi.org/10.1136/bmj.315.7109.629

28. Duval S, Tweedie R. A nonparametric "trim and fill" method of accounting for publication bias in meta-analysis. *J Am Stat Assoc*. 2000;95(449):89-98. doi:https://doi.org/10.1080/01621459.2000.10473905

29. Duval S, Tweedie R. Trim and fill: A simple funnel-plot–based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*. 2000;56(2):455-463. doi:https://doi.org/10.1111/j.0006-341X.2000.00455.x

30. StataCorp. *Stata Statistical Software: Release 17*. StataCorp LP; 2021.

31. Palmer TM, Sterne JAC, eds. *Meta-analysis in Stata: An updated collection from the Stata Journal*. 2nd ed. Stata Press; 2016.

32. Ding Y, Sun H, Han H, et al. The relationships between low levels of urine fluoride on children's intelligence, dental fluorosis in endemic fluorosis areas in Hulunbuir, Inner Mongolia, China. *J Hazard Mater*. 2011;186:1942-1946.

33. Zhang S, Zhang X, Liu H, et al. Modifying effect of COMT gene polymorphism and a predictive role for proteomics analysis in children's intelligence in endemic fluorosis area in Tianjin, China. *Toxicol Sci*. 2015b;144:238-245.

34. Cui Y, Zhang B, Ma J, et al. Dopamine receptor D2 gene polymorphism, urine fluoride, and intelligence impairment of children in China: A school-based cross-sectional study. *Ecotoxicol Environ Saf*. 2018;165:270-277. doi:https://doi.org/10.1016/j.ecoenv.2018.09.018

35. Saeed M, Rehman MYA, Farooqi A, Malik RN. Arsenic and fluoride co-exposure through drinking water and their impacts on intelligence and oxidative stress among rural school-aged children of Lahore and Kasur districts, Pakistan. *Environ Geochem Health*. Nov 9 2021;doi:https://dx.doi.org/10.1007/s10653-021-01141-4

36. Xu K, An N, Huang H, et al. Fluoride exposure and intelligence in school-age children: evidence from different windows of exposure susceptibility. *BMC Public Health*. Nov 4 2020;20(1):1657. doi:https://dx.doi.org/10.1186/s12889-020-09765-4

37. Zhao L, Yu C, Lv J, et al. Fluoride exposure, dopamine relative gene polymorphism and intelligence: A cross-sectional study in China. *Ecotoxicol Environ Safety*. Feb 2021;209:111826. doi:https://dx.doi.org/10.1016/j.ecoenv.2020.111826

38. Broadbent JM, Thomson WM, Moffitt TE, Poulton R. Community water fluoridation and intelligence response. *Am J Public Health*. 2015;105:3-4.

39. Khan SA, Singh RK, Navit S, et al. Relationship between dental fluorosis and intelligence quotient of school going children in and around Lucknow District: A cross-sectional study. *J Clin Diagn Res*. 2015;9(11):10-15. doi:https://doi.org/10.7860/JCDR/2015/15518.6726

40. Lin F, Ai H, Zhao H, Lin J, Jhiang J, Maimaiti. [High fluoride and low iodine environment and subclinical cretinism in Xinjiang]. *Endem Dis Bull*. 1991;6(2):62-67.

41. Li Y, Li X, Wei S. [Effects of high fluoride intake on child mental work capacity: Preliminary investigation into the mechanisms involved]. *J West Chin University Med Sci*. 1994;25(2):188-191.

42. Li Y, Li X, Wei S. Effects of high fluoride intake on child mental work capacity: Preliminary investigation into the mechanisms involved. *Fluoride*. 2008b;41:331-335.

43. Ibarluzea J, Gallastegi M, Santa-Marina L, et al. Prenatal exposure to fluoride and neuropsychological development in early childhood: 1-to 4 years old children. *Environ Res*. Oct 8 2021:112181. doi:https://dx.doi.org/10.1016/j.envres.2021.112181

44. Xiang Q, Liang Y, Chen L, et al. Effect of fluoride in drinking water on children's intelligence. *Fluoride*. 2003a;36:84-94.

45. Seraj B, Shahrabi M, Shadfar M, et al. Effect of high water fluoride concentration on the intellectual development of children in Makoo, Iran. *J Dent*. 2012;9:221-229.

46. Trivedi M, Sangai N, Patel R, Payak M, Vyas S. Assessment of groundwater quality with special reference to fluoride and its impact on IQ of schoolchildren in six villages of the Mundra Region, Kachchh, Gujurat, India. *Fluoride*. 2012;45(4):377-383.

47. Cui Y, Yu J, Zhang B, Guo B, Gao T, Liu H. The relationships between thyroid-stimulating hormone and/or dopamine levels in peripheral blood and IQ in children with different urinary iodine concentrations. *Neurosci Lett*. 2020;729:134981. doi:https://doi.org/10.1016/j.neulet.2020.134981

48. Valdez Jimenez L, Lopez Guzman OD, Cervantes Flores M, et al. In utero exposure to fluoride and cognitive development delay in infants. *Neurotoxicology*. Mar 2017;59:65-70. doi:https://dx.doi.org/10.1016/j.neuro.2016.12.011

49. Cantoral A, Tellez-Rojo MM, Malin AJ, et al. Dietary fluoride intake during pregnancy and neurodevelopment in toddlers: A prospective study in the progress cohort. *Neurotoxicology*. Dec 2021;87:86-93. doi:https://dx.doi.org/10.1016/j.neuro.2021.08.015

50. Bellinger DC. Interpretation of small effect sizes in occupational and environmental neurotoxicology: individual versus population risk. *Neurotoxicology*. 2007;28(2):245-51. doi:https://dx.doi.org/10.1016/j.neuro.2006.05.009

51. Needleman HL. What can the study of lead teach us about other toxicants? *Environ Health Perspect*. 1990;86:183-9. doi:https://dx.doi.org/10.1289/ehp.9086183

52. Rose G. Sick individuals and sick populations. *Int J Epidemiol*. 1985;14(1):32-8. doi:https://dx.doi.org/10.1093/ije/14.1.32

53. Rose G. Sick individuals and sick populations. *Int J Epidemiol*. 2001;30(3):427-32; discussion 433-4. doi:https://dx.doi.org/10.1093/ije/30.3.427

54. Weiss B. Vulnerability of children and the developing brain to neurotoxic hazards. *Environ Health Perspect*. 2000;108 Suppl 3(Suppl 3):375-81. doi:https://dx.doi.org/10.1289/ehp.00108s3375

55. Braun JM. Early-life exposure to EDCs: Role in childhood obesity and neurodevelopment. *Nature Rev Endocrin*. 2016;13(3):161-173. doi:https://dx.doi.org/10.1038/nrendo.2016.186

56. Grandjean P. Developmental fluoride neurotoxicity: an updated review. *Environ Health*. 2019;18(1):110. doi:https://doi.org/10.1186/s12940-019-0551-x

57. Abduweli Uyghurturk D, Goin DE, Martinez-Mier EA, Woodruff TJ, DenBesten PK. Maternal and fetal exposures to fluoride during mid-gestation among pregnant women in northern California. *Environ Health*. 2020;19(1):38. doi:https://doi.org/10.1186/s12940-020-00581-2

58. McMahon PB, Brown CJ, Johnson TD, Belitz K, Lindsey BD. Fluoride occurrence in United States groundwater. *Sci Total Environ*. 2020;732:139217. doi:https://doi.org/10.1016/j.scitotenv.2020.139217

59. CDC Division of Oral Health. *Personal communication*. Centers for Disease Control and Prevention; 2020. September 3, 2020.

60. Centers for Disease Control and Prevention (CDC). *Community water fluoridation FAQs: Infant formula*. Centers for Disease Control and Prevention; 2015. Accessed 10 February 2022. https://www.cdc.gov/fluoridation/faqs/infant-formula.html

61. Villa A, Anabalon M, Zohouri V, Maguire A, Franco AM, Rugg-Gunn A. Relationships between fluoride intake, urinary fluoride excretion and fluoride retention in children and adults: An analysis of available data. *Caries Res*. 2010;44(1):60-8. doi:https://dx.doi.org/10.1159/000279325

62. Watanabe M, Kono K, Orita Y, Usuda K, Takahashi Y, Yoshida Y. Influence of dietary fluoride intake on urinary fluoride concentration and evaluation of corrected levels in spot urine. *Fluoride*. 1995;28(2):61-70.

63. Zohouri FV, Swinbank CM, Maguire A, Moynihan PJ. Is the fluoride/creatinine ratio of a spot urine sample indicative of 24-h urinary fluoride? *Community Dent Oxral Epidemiol*. Apr 2006;34(2):130-8. doi:https://dx.doi.org/10.1111/j.1600-0528.2006.00269.x

64. Higgins JPT, Chandler TJ, Cumpston M, Li T, Page MJ, Welch VA, eds. *Cochrane Handbook for Systematic Reviews of Interventions version 6.2 (updated February 2021)*. Cochrane; 2021.

65. Lin L, Chu H. Quantifying publication bias in meta-analysis. *Biometrics*. 2018;74(3):785-794. doi:https://doi.org/10.1111/biom.12817

66. Ren D, Li K, Liu D. [A study of the intellectual ability of 8-14 year-old children in high fluoride, low iodine areas]. *Chin J Control Endem Dis*. 1989;4(4):251.

67. Ren D, Li K, Liu D. A study of the intellectual ability of 8-14 year-old children in high fluoride, low iodine areas. *Fluoride*. 2008;41:319-320.

68. Chen YX, Han FL, Zhoua ZL, et al. [Research on the intellectual development of children in high fluoride areas]. *Chin J Control Endem Dis*. 1991;6(Suppl):99-100.

69. Chen YX, Han FL, Zhoua ZL, et al. Research on the intellectual development of children in high fluoride areas. *Fluoride*. 2008;41:120-124.

70. Guo XC, Wang RY, Cheng CF, et al. [A preliminary investigation of the IQs of 7-13 year-old children from an area with coal burning-related fluoride poisoning]. *Chin J Epidemiol*. 1991;10(2):98-100.

71. Guo XC, Wang RY, Cheng CF, et al. A preliminary investigation of the IQs of 7-13 year-old children from an area with coal burning-related fluoride poisoning. *Fluoride*. 2008a;41:125-128.

72. Sun M, Li S, Wang Y, Li F. [Measurement of intelligence by drawing test among the children in the endemic area of Al-F combined toxicosis]. *J Guiyang Med Coll*. 1991;16(3):204-206.

73. An J, Mei S, Liu A, Fu Y, Wang C. [Effect of high level of fluoride on children's intelligence]. *Chin J Control Endem Dis*. 1992;7(2):93-94.

74. Xu Y, Lu C, Zhang X. [The effect of fluorine on the level of intelligence in children]. *Endem Dis Bull*. 1994;9(2):83-84.

75. Li XS, Zhi JL, Gao RO. Effect of fluoride exposure on the intelligence of children. *Fluoride*. 1995;28:189-192.

76. Wang G, Yang D, Jia F, Wang H. [A study of the IQ levels of four- to seven-year-old children in high fluoride areas]. *Endem Dis Bull*. 1996;11(1):60-62.

77. Wang G, Yang D, Jia F, Wang H. A study of the IQ levels of four- to seven-year-old children in high fluoride areas. *Fluoride*. 2008b;41:340-343.

78. Yao L, Zhou J, Wang S, Cui K, Lin F. Analysis on TSH and intelligence level of children with dental Fluorosis in a high fluoride area. *Lit Inf Prev Med*. 1996;2(1):26-27.

79. Zhao LB, Liang GH, Zhang DN, Wu XR. Effect of a high fluoride water supply on children's intelligence. *Fluoride*. 1996;29:190-192.

80. Yao Y. Comparable analysis on the physical and mental development of children in endemic fluorosis area with water improvement and without water improvement. *Lit Inf Prev Med*. 1997;3(1):42-43.

81. Zhang J, Yao H, Chen Y. [The effect of high levels of arsenic and fluoride on the development of children's intelligence]. *Chin J Public Health*. 1998;17(2):119.

82. Lu Y, Sun ZR, Wu LN, Wang X, Lu W, Liu SS. Effect of high-fluoride water on intelligence in children. *Fluoride*. 2000;33:74-78.

83. Hong FG, Cao YX, Yang D, Wang H. [Research on the effects of fluoride on child intellectual development under different environmental conditions]. *Chin Prim Health Care*. 2001;15(3):56-57.

84. Hong FG, Cao YX, Yang D, Wang H. Research on the effects of fluoride on child intellectual development under different environmental conditions. *Fluoride*. 2008;41:156-160.

85. Hong F, Wang Hui Yang Dong Zhang Z. Investigation on the intelligence and metabolism of iodine and fluoride in children with high iodine and fluoride. *Chin J Endem Dis Control*. 2001b:12-14.

86. Wang X, Wang L, Hu P, Guo X, Luo X. [Effects of high iodine and high fluorine on children's intelligence and thyroid function]. *Chin J Endemiol*. 2001;20(4):288-290.

87. Li YP, Jing XY, Chen D, Lin L, Wang ZJ. [Effects of endemic fluoride poisoning on the intellectual development of children in Baotou]. *Chin J Public Health Manag*. 2003;19(4):337-338.

88. Li YP, Jing XY, Chen D, Lin L, Wang ZJ. Effects of endemic fluoride poisoning on the intellectual development of children in Baotou. *Fluoride*. 2008c;41:161-164.

89. Wang SX, Wang ZH, Cheng XT, et al. [Investigation and evaluation on intelligence and growth of children in endemic fluorosis and arsenism areas]. *Chin J Endemiol*. 2005;24:179-182.

90. Seraj B, Shahrabi M, Falahzade M, Falahzade F, Akhondi N. Effect of high fluoride concentration in drinking water on children's intelligence. *J Dent Med*. 2006;19(2):80-86.

91. Wang Z, Wang S, Zhang X, Li J, Zheng X, Hu C. Investigation of children's growth and development under long-term fluoride exposure. *Chin J Control Endem Dis*. 2006;21(4):239-241.

92. Fan Z, Dai H, Bai A, Li P, Li T, Li G. The effect of high fluoride exposure in children's intelligence. *J Environ Health*. 2007;24(10):802-803.

93. Trivedi MH, Verma RJ, Chinoy NJ, Patel RS, Sathawara NG. Effect of high fluoride water on intelligence of school children in India. *Fluoride*. 2007;40:178-183.

94. Wang SX, Wang ZH, Cheng XT, et al. Arsenic and fluoride exposure in drinking water: Children's IQ and growth in Shanyin County, Shanxi Province, China. *Environ Health Perspect*. 2007;115:643-647.

95. Li F, Chen X, Huang R, Xie Y. The impact of endemic fluorosis caused by the burning of coal on the development of intelligence in children. *J Environ Health*. 2009;26(4):838-840.

96. Li X, Hou G, Yu B, Yuan C, Liu Y, L Z. [Investigation and analysis of children's IQ and dental fluorosis in high fluoride area]. *Chin J Pest Control*. 2010;26(3):230-231.

97. Eswar P, Nagesh L, Devaraj CG. Intelligent quotients of 12-14 year old school children in a high and low fluoride village in India. *Fluoride*. 2011;44:168-172.

98. Kang J, Cheng Y, Wu K, Lin S, He G, Jin Y. [Effect of exposure to fluoride and arsenic in drinking water of Hangjinhouqi on children's intelligence]. *Chin School Health*. 2011:679-681.

99. Poureslami HR, Horri A, Garrusi B. A comparative study of the IQ of children age 7-9 in a high and a low fluoride water city in Iran. *Fluoride*. 2011;44:163-167.

100. Shivaprakash PK, Ohri K, Noorani H. Relation between dental fluorosis and intelligence quotient in school children of Bagalkot District. *J Indian Soc Pedod Prev Dent*. 2011;29:117-120.

101. Wang S, Zhu X. Investigation and analysis of intelligence level of children in high fluoride area. *Chin J Endem Dis Control*. 2012b;27(1):67-68.

102. Bai A, Li Y, Fan Z, Li X, Li P. [Intelligence and growth development of children in coal-burning-borne arsenism and fluorosis areas: An investigation study]. *Chin J Endemiol*. 2014;33(2):160-163.

103. Karimzade S, Aghaei M, Mahvi AH. Investigation of intelligence quotient in 9-12 year-old children exposed to high- and low-drinking water fluoride in West Azerbaijan Province, Iran. *Fluoride*. 2014;47:9-14.

104. Sebastian ST, Sunitha S. A cross-sectional study to assess the intelligence quotient (IQ) of school going children aged 10-12 years in villages of Mysore District, India with different fluoride levels. *J Indian Soc Pedod Prev Dent*. 2015;33:307-311.

105. Zhang P, Cheng L. Effect of coal-burning endemic fluorosis on children's physical development and intellectual level. *Chin J Endem Dis Control*. 2015c;30(6):458-459.

106. Das K, Mondal NK. Dental fluorosis and urinary fluoride concentration as a reflection of fluoride exposure and its impact on IQ level and BMI of children of Laxmisagar, Simlapal Block of Bankura District, W.B., India. *Environ Monit Assess*. 2016;188:218.

107. Mondal D, Dutta G, Gupta S. Inferring the fluoride hydrogeochemistry and effect of consuming fluoride-contaminated drinking water on human health in some endemic areas of Birbhum District, West Bengal. *Environ Geochem Health*. 2016;38:557-576.

108. Zhao Y, Cui Y, Yu J, et al. Study on the relationship between water-borne high iodine and thyroid hormone and children's intelligence level. *J Environ Health*. 2018:6-9.

109. Wang J, Yu M, Yang L, Yang X, Deng B. The effect of coal-burning fluoride exposure on children's intelligence and physical development. *J Environ Health*. 2020c;37(11):971-974. doi:https://dx.doi.org/10.16241/j.cnki.1001-5914.2020.11.007

110. Guo B, Yu J, Cui Y, et al. DBH gene polymorphism, iodine and fluoride and their interactions and their interaction with children's intelligence. *J Environ Hygiene*. 2021;11(02):134-140.

111. Lou D, Luo Y, Liu J, et al. Refinement impairments of verbal-performance intelligent quotient in children exposed to fluoride produced by coal burning. *Biol Trace Elem Res*. Feb 2021;199(2):482-489. doi:https://dx.doi.org/10.1007/s12011-020-02174-z

112. Wang R, He N, Wang Y, Hou G, Zhang P-J. Investigation and analysis of children's dental fluorosis and IQ level in high fluoride district of Hengshui City. *Med Anim Control*. 2021;37(8):796-800. doi:https://dx.doi.org/10.7629/yxdwfz202108023

Commented [I68]: See Doc07_Meta-analysis, 7.D., page 3

List of Tables and Figures

Tables and Figures

Table 1. Characteristics of Studies Included in the Meta-analysis

Reference[a] *Study Design*	Study Location	Age Range (Years)	Fluoride Exposure: Assessment (Metric, Exposure Groups)	Fluoride Exposure: Levels	Intelligence Assessment	Overall RoB Rating	Confounders Considered
Ren et al. (1989)[66] [translated in Ren et al. 2008][me, o] *Cross-sectional*	China	8–14	No fluoride measurement Low iodine village/high fluoride and low iodine village	Not specified	Wechsler Intelligence Scale for Children	High	Sex; iodine
Chen et al. (1991)[68] [translated in Chen et al. 2008][me, w] *Cross-sectional*	China	7–14	Drinking water Nonendemic/endemic fluorosis village	0.89 mg/L (nonendemic) 4.55 mg/L (endemic)	Chinese Standardized Raven Test	High	Age; sex
Guo et al. (1991)[70] [translated in Guo et al. 2008a][me, o] *Cross-sectional*	China	7–13	Serum Reference area using wood/coal burning-related fluoride endemic area	0.1044 ± 0.0652 mg/L (reference) 0.1483 ± 0.0473 mg/L (endemic)	Chinese Binet Intelligence Test	High	Age; sex; SES
Lin et al. (1991)[40me, o] *Cross-sectional*	China	7–14	Urine, drinking water Reference area with iodine supplementation/high fluoride and low iodine village	Urine: 1.6 mg/L (reference area with iodine supplementation) 2.56 mg/L (high fluoride, low iodine village) Water: 0.34 mg/L (low iodine village) 0.88 mg/L (high fluoride, low iodine village)	Combined Raven's Test for Rural China	High	SES
Sun et al. (1991)[72me, o] *Cross-sectional*	China	6.5–12	No fluoride measurement Nonendemic/endemic (aluminum-fluoride endemic toxicosis)	Fluorosis: 98.36% (endemic)	Japan's Shigeo Kobayashi's 50-point scoring method	High	Age
An et al. (1992)[73me, w] *Cross-sectional*	China	7–16	Drinking water Nonhigh/high fluoride area	0.6−1.0 mg/L (nonhigh) 2.1−3.2 mg/L (secondary high) 5.2−7.6 mg/L (high) 2.1−7.6 mg/L (combined high)	Wechsler Intelligence Scale for Children-Revised	High	Age; race; SES

Commented [I69]: See Doc08_Meta-analysis, 8.K., page 7 and 8

Commented [I70]: See Doc03_Meta-analysis, 3.B., page 1 and 2

Commented [I71]: See Doc02_Meta-analysis, 2.V., page 7

Commented [I72]: See Doc01_Meta-analysis, 1.C., page 1

Reference[a] *Study Design*	Study Location	Age Range (Years)	Fluoride Exposure		Intelligence Assessment	Overall RoB Rating	Confounders Considered
			Assessment (Metric, Exposure Groups)	Levels			
Li et al. (1994)[41] [translated in Li et al. 2008b][me, o] *Cross-sectional*	China	12–13	Grain (cooked by burning high-fluoride coal) Reference group (no dental fluorosis)/high fluoride group I (no dental fluorosis)/high fluoride group II (dental fluorosis present)/high fluoride group III (dental fluorosis present)	0.5 mg/kg (reference group) 4.7 mg/kg (group I) 5.2 mg/kg (group II) 31.6 mg/kg (group III)	Proofing test	High	Age; sex; SES
Xu et al. (1994)[74me, w*] *Cross-sectional*	China	8–14	Drinking water Reference region/low- and high-fluoride regions[b]	0.8 mg/L (reference region) 0.38 mg/L (low fluoride) 1.8 mg/L (high fluoride)	Binet-Simon Scale	High	–
Li et al. (1995)[75me, o, u] *Cross-sectional*	China	8–13	Urine, dental fluorosis index (DFI) Nonfluorosis/fluorosis area due to soot from coal burning	1.02 mg/L; DFI: <0.4 (nonfluorosis) 1.81 mg/L; DFI: 0.8 (slight fluorosis) 2.01 mg/L; DFI: 2.5 (medium fluorosis) 2.69 mg/L; DFI: 3.2 (severe fluorosis)	China Rui Wen Scaler for Rural Areas	High	Sex
Wang et al. (1996)[76] [translated in Wang et al. 2008b][me, o, w] *Cross-sectional*	China	4–7	Drinking water (well) Low/high fluoride region Fluoride exposure from drinking water, contaminated food, and coal burning	0.58−1.0 mg/L (low) >1.0−8.6 mg/L (high)	Wechsler Preschool and Primary Scale of Intelligence	High	Age; sex
Yao et al. (1996)[78me, w] *Cross-sectional*	China	8–12	Drinking water Nonendemic/endemic fluorosis area	1 mg/L (nonendemic) 2 mg/L (slightly endemic) 11 mg/L (severely endemic)	Raven Test – Associative Atlas	High	Iodine; SES
Zhao et al. (1996)[79me, w] *Cross-sectional*	China	7–14	Drinking water Low fluoride village (Xinghua)/high fluoride village (Sima)	0.91 mg/L (low) 4.12 mg/L (high)	China Rui Wen Scaler for Rural Areas	High	Age; SES
Yao (1997)[80me, w*] *Cross-sectional*	China	7–12	Drinking water Nonfluorosis area/fluorosis area with water improvements/fluorosis area without water improvements	0.4 mg/L (nonfluorosis area) 0.33 mg/L (fluorosis area with water improvement) 2 mg/L (fluorosis area without water improvement)	Raven's Standard Progressive Matrices (China's Rural Version)	High	Iodine; SES
Zhang et al. (1998)[81me, o] *Cross-sectional*	China	4–10	Drinking water Reference/high fluoride group (all observation groups included arsenic exposure)	0.58 mg/L (reference) 0.8 mg/L (high fluoride)	Shigeo Kobayashi 50-pt. test	High	Age; arsenic

Reference[a] *Study Design*	Study Location	Age Range (Years)	Fluoride Exposure		Intelligence Assessment	Overall RoB Rating	Confounders Considered
			Assessment (Metric, Exposure Groups)	Levels			
Lu et al. (2000)[82me, w, u] *Cross-sectional*	China	10–12	Urine, drinking water Low/high fluoride area	Urine: 1.43 ± 0.64 mg/L (low) 4.99 ± 2.57 mg/L (high) Water: 0.37 ± 0.04 mg/L (low) 3.15 ± 0.61 mg/L (high)	Chinese Combined Raven Test-C2	High	SES
Hong et al. (2001)[83] [translated in Hong et al. 2008][me, w*] *Cross-sectional*	China	8–14	Drinking water Reference/high fluoride[b]	0.75 mg/L (reference) 2.90 mg/L (high fluoride)	Chinese Standardized Raven Test	High	Iodine; SES; demographics
Hong et al. (2001b)[85me, o] *Cross-sectional*	China	8–14	Urine, drinking water Nonendemic/endemic fluorosis areas (high fluoride, high iodine)	Urine: 0.796 ± 0.53 mg/L (nonendemic) 2.09 ± 1.03 mg/L (endemic) Water: 0.48 mg/L (nonendemic) 2.81 mg/L (endemic)	Combined Raven's Test for Rural China	High	–
Wang et al. (2001)[86me, o] *Cross-sectional*	China	8–12	Urine, drinking water Reference point (low fluoride, low iodine)/investigative point (high fluoride, high iodine)	Urine: 0.82 mg/L (low fluoride, low iodine) 3.08 mg/L (high fluoride, high iodine) Water: 0.5 mg/L (low fluoride, low iodine) 2.97 mg/L (high fluoride, high iodine)	Combined Raven's Test for Rural China	High	–
Li et al. (2003)[87] [translated in Li et al. 2008c][me] *Cross-sectional*	China	6–13	No fluoride measurement Reference/endemic fluorosis areas	Not specified	Chinese Standardized Raven Test	High	–

Reference[a] *Study Design*	Study Location	Age Range (Years)	Fluoride Exposure		Intelligence Assessment	Overall RoB Rating	Confounders Considered
			Assessment (Metric, Exposure Groups)	Levels			
Xiang et al. (2003a)[44me, w*, u] *Cross-sectional*	China	8–13	Urine, drinking water Nonendemic/endemic fluorosis areas	Urine: 1.11 ± 0.39 mg/L (reference) 3.47 ± 1.95 mg/L (high fluoride) Water: 0.36 ± 0.15 mg/L (nonendemic) 0.75 ± 0.14 mg/L (endemic fluorosis area group A) 1.53 ± 0.27 mg/L (endemic fluorosis area group B) 2.46 ± 0.3 mg/L (endemic fluorosis area group C) 3.28 ± 0.25 mg/L (endemic fluorosis area group D) 4.16 ± 0.22 mg/L (endemic fluorosis area group E) 2.47 ± 0.79 mg/L (high fluoride)	Combined Raven's Test for Rural China	Low	Age; sex; iodine; lead; SES
Wang et al. (2005)[89me, w, u] *Cross-sectional*	China	8–12	Urine, drinking water Reference/high fluoride group[c]	Urine: 1.51 mg/L(reference) 5.09 mg/L (high fluoride group) Water: 0.48 mg/L (reference) 8.31 mg/L (high fluoride group)	Chinese Combined Raven Test-C2	High	SES
Seraj et al. (2006)[90me, w *Cross-sectional*	Iran	7–11	Drinking water Low/high fluoride area	0.4 ppm (low) 2.5 ppm (high)	Raven Test	High	Sex
Wang et al. (2006)[91me, w, u] *Cross-sectional*	China	8–12	Urine, drinking water Reference/high (area severely affected by fluorosis)	Urine: 1.51 ± 1.66 mg/L (reference) 5.50 ± 2.40 mg/L (high) Water: 0.73 ± 0.28 mg/L (reference) 5.54 ± 3.88 mg/L (high)	Combined Raven's Test for Rural China	High	–
Fan et al. (2007)[92me, w, u] *Cross-sectional*	China	7–14	Urine, drinking water Low/high fluoride area	Urine: 1.78 ± 0.46 mg/L (low) 2.89 ± 1.97 mg/L (high) Water: 1.03 mg/L (low) 3.15 mg/L (high)	Chinese Combined Raven Test-C2	High	–
Trivedi et al. (2007)[93me, w, u] *Cross-sectional*	India	12–13	Urine, drinking water Low/high fluoride area	Urine: 2.30 ± 0.28 mg/L (low) 6.13 ± 0.67 mg/L (high) Water: 2.01 ± 0.009 mg/L (low) 5.55 ± 0.41 mg/L (high)	questionnaire prepared by Professor JH Shah	High	Age; sex

Reference[a] *Study Design*	Study Location	Age Range (Years)	Fluoride Exposure		Intelligence Assessment	Overall RoB Rating	Confounders Considered
			Assessment (Metric, Exposure Groups)	Levels			
Wang et al. (2007)[94me, o, w, u] *Cross-sectional*	China	8–12	Urine, drinking water Low fluoride, low arsenic/high fluoride, low arsenic area	Urine: 1.5 ± 1.6 mg/L (low fluoride, low arsenic) 5.1 ± 2.0 mg/L (high fluoride, low arsenic) Water: 0.5 ± 0.2 mg/L (low fluoride, low arsenic) 8.3 ± 1.9 mg/L (high fluoride, low arsenic)	Combined Raven's Test for Rural China	High	Age; sex; arsenic; SES
Li et al.(2009)[95me, o, u*] *Cross-sectional*	China	8–12	Urine Endemic fluorosis region caused by coal burning (reference/mild/medium/severe) Degree of dental fluorosis (normal/suspected/very mild/mild/medium/severe)	0.962 ± 0.517 mg/L (reference) 1.235 ± 0.426 mg/L (mild) 1.670 ± 0.663 mg/L (medium) 2.336 ± 1.128 mg/L (severe) 0.867 ± 0.233 mg/L (normal) 1.094 ± 0.355 mg/L (suspected) 1.173 ± 0.480 mg/L (very mild) 1.637 ± 0.682 mg/L (mild) 2.005 ± 0.796 mg/L (medium) 2.662 ± 1.093 mg/L (severe)	Combined Raven's Test for Rural China	High	Age; sex
Li et al. (2010)[96me] *Cross-sectional*	China	7–10	No fluoride measurement Nondental fluorosis children/dental fluorosis children	Not specified	Combined Raven's Test for Rural China	High	Sex
Ding et al. (2011)[32me, u*, rs] *Cross-sectional*	China	7–14	Dental fluorosis (normal/ questionable/very mild/mild/moderate) Urine Mean urinary fluoride levels (10 groups)	0.80 ± 0.55 mg/L (normal) 1.13 ± 0.73 mg/L (questionable) 1.11 ± 0.74 mg/L (very mild) 1.31 ± 0.78 mg/L (mild) 1.46 ± 0.79 mg/L (moderate) 0.26 mg/L (group 1) 0.45 mg/L (group 2) 0.56 mg/L (group 3) 0.66 mg/L (group 4) 0.75 mg/L (group 5) 0.89 mg/L (group 6) 1.08 mg/L (group 7) 1.33 mg/L (group 8) 1.74 mg/L (group 9) 2.96 mg/L (group 10) 0.10–3.55 mg/L	Combined Raven's Test for Rural China	Low	Age; arsenic; iodine; lead; SES; demographics

Reference[a] *Study Design*	Study Location	Age Range (Years)	Fluoride Exposure		Intelligence Assessment	Overall RoB Rating	Confounders Considered
			Assessment (Metric, Exposure Groups)	Levels			
Eswar et al.(2011)[97me, w] *Cross-sectional*	India	12–14	Drinking water Low/high fluoride villages	0.29 mg/L (low) 2.45 mg/L (high)	Standard Progressive Matrices	High	Age; sex
Kang et al. (2011)[98me, o] *Cross-sectional*	China	6–12	Drinking water Reference/high fluoride areas (both areas with high arsenic exposure)	1.24 ± 0.74 mg/L (all children) <1.2 mg/L (reference) ≥1.2 mg/L (high fluoride)	Chinese Combined Raven Test-C2	High	Age; sex
Poureslami et al. (2011)[99me, w] *Cross-sectional*	Iran	7–9	Drinking water Reference/endemic dental fluorosis city	0.41 mg/L (reference) 2.38 mg/L (endemic)	Persian version of Raven's Matrices Test	High	Sex
Shivaprakash et al. (2011)[100me, w] *Cross-sectional*	India	7–11	Drinking water No fluorosis/fluorosis severity groups (mild/moderate/severe)/all fluorosis	<0.5 ppm (no fluorosis) 2.5–3.5 ppm (mild) 2.5–3.5 ppm (moderate) 2.5–3.5 ppm (severe) 2.5–3.5 ppm (all)	Raven's Colored Progressive Matrices	High	Health factors; SES
Seraj et al. (2012)[45me, w] *Cross-sectional*	Iran	6–11	Drinking water Normal/medium/high fluoride levels	0.8 ± 0.3 mg/L (normal) 3.1 ± 0.9 mg/L (medium) 5.2 ± 1.1 mg/L (high)	Raven's Colored Progressive Matrices	Low	Age; sex; SES
Trivedi et al. (2012)[46me, w, u] *Cross-sectional*	India	12–13	Urine, ground water Low/high fluoride area	Urine: 0.42 ± 0.23 mg/L (low) 2.69 ± 0.92 mg/L (high) Water: 0.84 ± 0.38 mg/L (low) 2.3 ± 0.87 mg/L (high)	Questionnaire prepared by Professor JH Shah	Low	Sex; SES
Wang et al. (2012b)[101me] *Cross-sectional*	China	Primary school age	No fluoride measurement Reference/high fluoride areas	Not specified	Combined Raven's Test for Rural China	High	–
Bai et al. (2014)[102me, o] *Cross-sectional*	China	8–12	Urine Coal-burning-borne fluorosis areas (reference/lightly-affected/seriously-affected)	0.54 mg/L (reference) 0.81 mg/L (lightly-affected area) 1.96 mg/L (seriously-affected area)	Chinese Combined Raven Test-C2	High	SES
Karimzade et al. (2014)[103me, w] *Cross-sectional*	Iran	9–12	Drinking water Low/high fluoride area	0.25 mg/L (low) 3.94 mg/L (high)	Iranian version of the Raymond B Cattell test	High	Sex

Reference[a] *Study Design*	Study Location	Age Range (Years)	Fluoride Exposure		Intelligence Assessment	Overall RoB Rating	Confounders Considered
			Assessment (Metric, Exposure Groups)	Levels			
Broadbent et al. (2015)[38me, w*] *Prospective Cohort*	New Zealand	7–13	Drinking water Area without community water fluoridation (low)/area with community water fluoridation (high) Fluoride tablet use (never/ever) Fluoride toothpaste use (never/sometimes/always)	Water: 0.0–0.3 mg/L (low) 0.7–1.0 mg/L (high) Tablet use: 0 mg (never used) 0.5 mg (ever used) Range not specified for fluoride toothpaste use (always/sometimes/never)	Wechsler Intelligence Scale for Children-Revised	High	Sex; SES; low birth weight; breastfeeding
Khan et al. (2015)[39me] *Cross-sectional*	India	6–11	Drinking water Low fluoride areas (Tiwariganj)/high fluoride areas (Unnao) Fluorosis grades (normal/very mild/mild/moderate/severe)	0.19 mg/L (Tiwariganj) 2.41 mg/L (Unnao) Ranges not specified by fluorosis grades	Raven's Colored Progressive Matrices	High	Health factors; SES
Sebastian and Sunitha (2015)[104me, w*] *Cross-sectional*	India	10–12	Drinking water Low/normal/high fluoride villages	0.40 mg/L (low) 1.2 mg/L (normal) 2.0 mg/L (high)	Raven's Colored Progressive Matrices	High	Age; sex; SES
Zhang et al.(2015b)[33me, w*, u, rs] *Cross-sectional*	China	10–12	Urine, drinking water, serum Reference/high fluoride areas	Urine: 1.10 ± 0.67 mg/L (reference) 2.40 ± 1.01 mg/L (high) Water: 0.63 (0.58–0.68) mg/L (reference) 1.40 (1.23–1.57) mg/L (high) Serum: 0.06 ± 0.03 (reference) 0.18 ± 0.11 (high)	Combined Raven's Test for Rural China	Low	Age; sex; arsenic; iodine; drinking water fluoride; SES; thyroid hormone levels; COMT genotype
Zhang et al. (2015c)[105me, o] *Cross-sectional*	China	7−13	Urine Coal-burning endemic fluorosis area Reference (no dental fluorosis)/mild dental fluorosis/moderate dental fluorosis/critically ill dental fluorosis	0.83 ± 0.71 mg/L (reference) 1.54 ± 0.57 mg/L (mildly ill) 2.41 ± 0.76 mg/L (moderately ill) 3.32 ± 1.02 mg/L (critically ill)	Combined Raven's Test for Rural China	High	–

Reference[a] *Study Design*	Study Location	Age Range (Years)	Fluoride Exposure		Intelligence Assessment	Overall RoB Rating	Confounders Considered
			Assessment (Metric, Exposure Groups)	Levels			
Das and Mondal (2016)[106me, u] *Cross-sectional*	India	6–18	Urine, drinking water intake, dental fluorosis (normal/questionable/very mild/mild/moderate/severe)	Urine: 2.91 ± 1.76 mg/L (normal) 2.50 ± 2.39 mg/L (questionable) 2.58 ± 1.31 mg/L (very mild) 2.95 ± 1.44 mg/L (mild) 4.82 ± 3.57 mg/L (moderate) 3.81 ± 2.51 mg/L (severe) Water: 0.069 ± 0.021 mg/kg-d (normal) 0.064 ± 0.004 mg/kg-d (questionable) 0.060 ± 0.036 mg/kg-d (very mild) 0.060 ± 0.030 mg/kg-d (mild) 0.099 ± 0.063 mg/kg-d (moderate) 0.093 ± 0.040 mg/kg-d (severe)	Combined Raven's Test for Rural China	High	–
Mondal et al. (2016)[107me, w] *Cross-sectional*	India	10–14	Drinking water Low/high fluoride areas	Not reported (low) 0.33–18.08 mg/L (high)	Raven Standard Theoretical Intelligence Test	High	SES
Bashash et al. (2017)[5me, u, rs] *Prospective Cohort*	Mexico	6−12	Maternal urine Reference/high fluoride (based on children urinary fluoride)	<0.80 mg/L (reference) ≥0.80 mg/L (high)	Wechsler Abbreviated Scale of Intelligence	Low	Age; sex; weight at birth; parity; gestational age; maternal characteristics (smoking history, marital status, age at delivery, IQ, education, cohort)
Cui et al. (2018)[34rs] *Cross-sectional*	China	7–12	Urine	Boys: 1.3 (0.9−1.7)[d] mg/L Girls: 1.2 (0.9−1.6)[d] mg/L	Combined Raven's Test for Rural China	Low	Age; maternal education; smoking in family member; stress; anger; dopamine receptor-2 polymorphism

Reference[a] *Study Design*	Study Location	Age Range (Years)	Fluoride Exposure		Intelligence Assessment	Overall RoB Rating	Confounders Considered
			Assessment (Metric, Exposure Groups)	Levels			
Yu et al. (2018)[11me, w, u*, rs] *Cross-sectional*	China	7–13	Maternal urine Low/medium/high fluoride ranges Drinking water Normal/high fluoride	Urine: 0.01–1.60 mg/L (low) 1.60–2.50 mg/L (medium) 2.50–5.54 mg/L (high) Water: ≤1 mg/L (normal) >1 mg/L (high) Overall: 0.01−5.54 mg/L (urine) 0.20−3.90 mg/L (water)	Combined Raven's Test for Rural China	Low	Age; sex; health factors; SES
Zhao et al. (2018)[108me, o] *Cross-sectional*	China	7–12	Urine Reference/exposed areas All areas with iodine exposure	≤2.16 mg/L (reference) >2.16 mg/L (exposed)	Combined Raven's Test for Rural China	High	–
Green et al. (2019)[6me, w*, u*, rs] *Prospective Cohort*	Canada	3−4	Maternal urine, drinking water, maternal fluoride intake Nonfluoridated/fluoridated area	Urine: 0.40 ± 0.27 mg/L (nonfluoridated) 0.69 ± 0.42 mg/L (fluoridated) Water: 0.13 ± 0.06 mg/L (nonfluoridated) 0.59 ± 0.08 mg/L (fluoridated) Intake: 0.30 ± 0.26 mg/day (nonfluoridated) 0.93 ± 0.43 mg/day (fluoridated) Overall: 0.51 ± 0.36 mg/L (urine) 0.54 ± 0.44 mg/day (intake) 0.31 ± 0.23 mg/L (water)	Wechsler Primary and Preschool Scale of Intelligence-III	Low	Sex; city; maternal education; race/ethnicity; HOME score; prenatal secondhand smoke exposure
Cui et al. (2020)[47me, u] *Cross-sectional*	China	7–12	Urine Low/medium/high fluoride levels	<1.6 mg/L (low) 1.6–2.5 mg/L (medium) >=2.5 mg/L (high)	Combined Raven's Test	Low	Sex; arsenic; iodine

Reference[a] *Study Design*	Study Location	Age Range (Years)	Fluoride Exposure		Intelligence Assessment	Overall RoB Rating	Confounders Considered
			Assessment (Metric, Exposure Groups)	Levels			
Till et al. (2020)[7rs] *Prospective Cohort*	Canada	3−4	Residence, maternal urine, drinking water, infant fluoride intake from formula Nonfluoridated/fluoridated areas	Urine: 0.38−0.42 mg/L (nonfluoridated) 0.64−0.70 mg/L (fluoridated) Water: 0.13 mg/L (nonfluoridated) 0.58 mg/L (fluoridated) Intake: 0.02−0.08 mg/day (nonfluoridated) 0.12−0.34 mg/day (fluoridated)	Wechsler Primary and Preschool Scale of Intelligence-III	Low	Age; sex; maternal education; maternal race; HOME total score; secondhand smoke status in the child's house
Wang et al. (2020c)[109me, o] *Cross-sectional*	China	7−12	Urine Coal-burning endemic fluorosis area Nonendemic/endemic fluorosis regions	0.461 ± 0.210 mg/L (nonendemic) 0.689 ± 0.502 mg/L (endemic)	Combined Raven's Test for Rural China	High	Age; sex
Xu et al. (2020)[36me, u*, rs] *Cross-sectional*	China	7−13	Urine Reference/high prenatal exposure only/high childhood exposure only/both prenatal and childhood exposure group	0.82 ± 0.30 mg/L (reference) 0.98 ± 0.29 mg/L (high prenatal exposure only) 2.05 ± 0.58 mg/L (high childhood exposure only) 2.13 ± 0.59 mg/L (both prenatal and childhood exposure group)	Combined Raven's Test for Rural China	Low	Age; sex; gestational weeks; maternal education level; paternal education level; children's BMI
Guo et al., (2021)[110me] *Cross-sectional*	China	7−12	Urine Reference/exposed areas (also with iodine exposure)	1.16 mg/L (reference) 1.29 mg/L (iodine area 1) 2.01 mg/L (iodine area 2)	Combined Raven's Test for Rural China	High	–
Lou et al. (2021)[111me, o] *Cross-sectional*	China	8−12	Coal-burning endemic fluorosis area No fluoride measurement Nondental fluorosis children/dental fluorosis children	Not specified	Wechsler Intelligence Scale for Children-Revised in China (WISC-CR)	High	–
Saeed et al. (2021)[35me, o, rs] *Cross-sectional*	Pakistan	5−16	Urine, drinking water Reference/high fluoride areas Co-exposure with arsenic	Urine: 0.24 ± 0.15 mg/L (reference) 3.27 ± 2.60 mg/L (high) Water: 0.15 ± 0.13 mg/L (reference) 5.64 ± 3.52 mg/L (high)	Wechsler scale of intelligence (WISC-IV)	High	Age; sex; parental education; dental fluorosis

Reference[a] *Study Design*	Study Location	Age Range (Years)	Fluoride Exposure		Intelligence Assessment	Overall RoB Rating	Confounders Considered
			Assessment (Metric, Exposure Groups)	Levels			
Wang et al. (2021)[112] me, w *Cross-sectional*	China	9−11	Drinking water Reference/high fluoride areas	1.0 ± 0.07 mg/L (reference) 2.8 ± 0.06 mg/L (high fluoride)	Combined Raven's Test	High	Age; sex
Zhao et al. (2021)[37rs] *Cross-sectional*	China	6−11	Urine Nonendemic/endemic fluorosis areas	1.03 (0.72, 1.47) mg/L	Combined Raven's Test for Rural China	Low	Age; sex; BMI; paternal educational level; maternal educational level; household income; abnormal birth; maternal age at delivery

Notes:
COMT = catechol-O-methyltransferase; RoB = risk of bias; SES = socioeconomic status; HOME = Home Observation for Measurement of the Environment
[a]An "me" superscript indicates that the studies included in the mean-effects meta-analysis; an "o" superscript indicates a study included in "other" exposures *mean-effects meta-analysis* (see **Table 2** footnote); a "w" superscript indicates studies included in the mean-effects dose-response meta-analysis using fluoride in water; a "u" superscript indicates studies included in the mean-effects dose-response meta-analysis using fluoride in urine; "*" indicates studies included in the mean-effects dose-response meta-analysis at levels < 1.5 mg/L; an "rs" superscript indicates studies included in the regression slopes meta-analysis.
[b]Additional exposure regions including iodine levels were not included in the analysis.
[c]Additional exposure regions including arsenic levels were not included in the analysis.
[d]Median (q1−q3).

Table 2. Pooled SMDs and 95% CIs for Children's IQ Score and Exposures to Fluoride

Analysis	Number of Studies	SMD (95% CI)	Heterogeneity	
			p-value	I^2
Overall Effect	55	−0.46 (−0.55, −0.37)	<0.001	87%
Subgroup Analyses				
Risk of Bias				
Low	10	−0.22 (−0.39, −0.05)	<0.001	83%
High	45	−0.52 (−0.63, −0.42)	<0.001	86%
Sex				
Males	14	−0.62 (−0.81, −0.42)	<0.001	78%
Females	13	−0.53 (−0.72, −0.34)	<0.001	74%
Age Group				
<10 years[a]	13	−0.41 (−0.60, −0.22)	<0.001	80%
≥10 years	16	−0.55 (−0.70, −0.40)	<0.001	68%
Country				
China	39	−0.43 (−0.52, −0.34)	<0.001	85%
India	8	−0.99 (−1.55, −0.43)	<0.001	93%
Iran	4	−0.68 (−0.99, −0.38)	0.077	56%
Canada	1	0.01 (−0.19, 0.21)	NA	NA
Mexico	1	0.13 (−0.16, 0.42)	NA	NA
New Zealand	1	0.01 (−0.19, 0.22)	NA	NA
Pakistan	1	−0.25 (−0.65, 0.16)	NA	NA
Assessment Type				
CRT-RC tests	29	−0.36 (−0.46, −0.27)	<0.001	82%
Non-CRT-RC tests	26	−0.60 (−0.78, −0.42)	<0.001	89%
Raven's tests	10	−0.76 (−1.10, −0.43)	<0.001	91%
Other tests	16	−0.52 (−0.74, −0.29)	<0.001	89%
Exposure Type				
Water fluoride	32	−0.37 (−0.46, −0.27)	<0.001	82%
Dental fluorosis	7	−0.99 (−1.57, −0.41)	<0.001	96%
Other exposures[b]	16	−0.54 (−0.71, −0.37)	<0.001	81%
Previous Meta-analyses				
Duan et al. (2018)[4]	26	−0.52 (−0.62, −0.42)	<0.001	69%
Choi et al. (2012)[3]	27	−0.45 (−0.56, −0.34)	<0.001	80%

Notes: CI = confidence interval; CRT-RC = Combined Raven's Test–The Rural edition in China; NA = not applicable; SMD = standardized weighted mean difference

[a]An et al. (1992)[73] and Li et al. (2010)[96] include 10-year-old children in the <10 age group (7−10 years reported).

[b]Includes iodine [40, 66 [translated in Ren et al. 2008], 85, 86, 108]; arsenic[35, 81, 94]; aluminum[72]; and non-drinking water fluoride (i.e., fluoride from coal burning[41 [translated in Li et al. 2008b], 70 [translated in Guo et al. 2008a], 75, 76 [translated in Wang et al. 2008b], 89, 95, 102, 105, 109, 111]).

[c] p-value for differences between the estimates based on CRT-RC tests *vs.* non-CRT-RC tests.

[d] p-value for differences between the estimates based on CRT-RC tests, Raven's test and other tests. Note that non-CRT-RC test include Raven's tests and other tests.

Commented [l73]: See Doc06a_Meta-analysis, 6a.E., page 3 and 4

Commented [l74]: See Doc02_Meta-analysis, 2.X., page 7 and 8

Table 3. Pooled Regression Slopes and 95% CIs for Children's IQ Score and Exposures to Fluoride

Analysis	Number of Studies	Beta (95% CI)	Heterogeneity	
			p-value	I^2
Overall Effect				
Full-scale IQ	9	−1.81 (−2.80, −0.81)	<0.001	77%
Subgroup Analyses				
Risk of Bias				
Low	8	−1.33 (−2.09, −0.57)	0.072	46%
High	1	−3.45 (−4.44, −2.46)	NA	NA
Sex				
Males	2	−2.23 (−5.45, 0.99)	0.092	65%
Females	2	−0.27 (-3.64, 3.10)	0.145	53%
Country				
Canada	1	−1.95 (−5.18, 1.28)	NA	NA
China	6	−1.06 (−1.70, −0.42)	0.191	33%
Mexico	1	−5.00 (−8.53, −1.47)	NA	NA
Pakistan	1	−3.45 (−4.44, −2.46)	NA	NA
Assessment Type				
CRT-RC tests	6	−1.06 (−1.70, −0.42)	0.191	33%
Non-CRT-RC tests	3	−3.43 (−4.35, −2.52)	0.457	0%
Exposure Type				
Urinary fluoride	9	−1.81 (−2.80, −0.81)	<0.001	77%
Intake	2	−3.87 (−7.15, −0.59)	0.737	0%
Water fluoride	2	−4.77 (−9.09, −0.45)	0.707	0%
Exposure timing				
Pre-natal exposure	3	−3.08 (−5.43, −0.72)	0.351	5%
Post-natal exposure	7	−1.84 (-2.97, -0.72)	<0.001	78%

Notes: CI = confidence interval; NA = not applicable

Commented [I75]: See Doc02_Meta-analysis, 2.H., page 3

Commented [I76]: See Doc06a_Meta-analysis, 6a.I., page 6

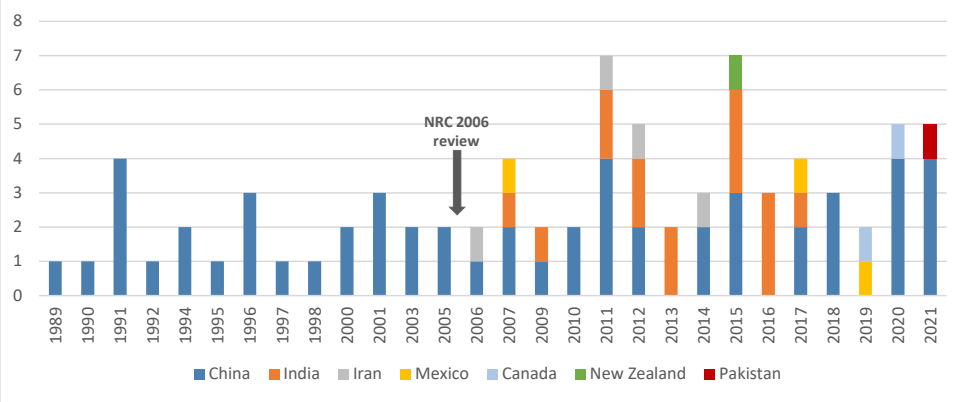


Figure 1. Number of Studies of Fluoride Exposure and IQ in Children by Country and Year of Publication

Note: Figure includes 80 epidemiological studies that were identified during the larger systematic review and the November 2021 literature search update that evaluated the effects of fluoride exposure on children's IQ.

Commented [EAM77]: See Doc02_Meta-analysis, 2.W, page 7

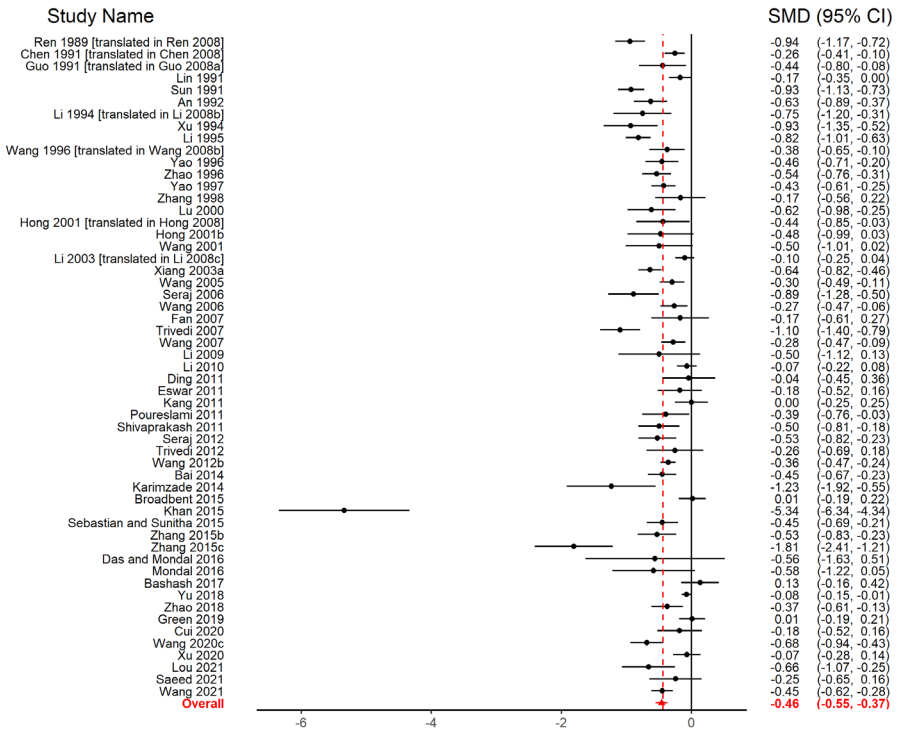


Figure 2. Association Between Fluoride Exposure and IQ Scores in Children

Note: Forest plot for random-effects meta-analysis of the association between fluoride exposure and child's IQ scores. Effect size is expressed as the standardized weighted mean difference for heteroscedastic population variances (SMD). The random-effects pooled SMD is shown as a solid triangle. Horizontal lines represent 95% CIs for the study-specific SMDs.

Commented [I78]: See Doc06b_Meta-analysis, 6b.R., page 15 and 16

SUPPLEMENTAL MATERIALS

List of Tables

List of Figures

Additional Detail on Methods

Systematic Literature Review

Literature searches were conducted in the following databases: BIOSIS, EMBASE, PsychINFO, PubMed, Scopus, Web of Science, CNKI, and Wanfang. Search strategies tailored for each database are available in the protocol (https://ntp.niehs.nih.gov/go/785076). The last search was performed on May 1, 2020. The identification of studies for the meta-analysis was part of a larger systematic review.[1]

Study Selection

In order to be eligible for inclusion in the systematic literature review, individual study publications (referred to in this paper as "studies") had to satisfy eligibility criteria outlined in the protocol (i.e., address PECO statement in Table 1 and specific exclusion criteria in Table 2, https://ntp.niehs.nih.gov/go/785076).

Commented [I1]: See Doc02_Meta-analysis, 2.E., page 2

The following exclusions were made:

(1) Case studies and case reports.

(2) Articles without original data (e.g., reviews, editorials, commentaries). Reference lists from these materials, however, were reviewed to identify potentially relevant studies not identified from the database searches. New studies identified were assessed for eligibility for inclusion.

(3) Conference abstracts or reports and dissertations.

References retrieved from the literature search were independently screened by two trained screeners at the title and abstract level to determine whether a reference met the evidence selection criteria. Studies that were not excluded during the title and abstract screening were further screened for inclusion with a full-text review by two independent reviewers. Translation assistance was obtained to assess the relevance of non-English studies. Following full-text review, the remaining studies were "included" and used for the evaluation.

Results of the study identification process are provided in **eFigure 1**.

Statistical Analysis

Mean-effects meta-analysis

A sensitivity analysis was performed to evaluate the impact of using any exposed group compared to the reference group. This was accomplished by using the approach outlined in the Cochrane Handbook for Systematic Reviews[2] which combines the data from all available exposure groups (n, mean, and standard deviation [SD]). Subgroup analyses were stratified by risk of bias (high or low), outcome assessment, exposure matrix (e.g., urine or water), pre- or post-natal exposures, outcome, gender, and age group. If results were not reported by gender or age-specific subgroups (<10, ≥10 years), they were calculated (if possible) by combining smaller subgroups. If SDs were not reported, but mean effects, sample sizes (n values), and p-values for differences between groups were available, SDs were calculated using the SE and t-statistic (assuming equal variances). To avoid sample overrepresentation, if the same cohort was followed at multiple timepoints resulting in multiple study publications (e.g., Yu et al.[3] and Wang et al.[4]), only the study publication that included the largest number of participants was included in the meta-analysis (see **eTable 1** for list of excluded studies and rationales). For studies with overlapping populations (i.e., multiple study publications that used the same cohort), results from one study publication were selected considering the following factors: most appropriate exposure metric, exposure

range, exposure period, number of subjects, and statistical adjustment for potential confounders (see **eTable 2** for study-specific effect estimates used in the meta-analysis).

Dose-response meta-analysis

To determine whether the data support an exposure-response relationship, we conducted a *dose-response mean-effects meta-analysis.* This analysis included studies from the *mean-effects meta-analysis* that reported fluoride exposure levels; we excluded studies for which there was evidence that co-exposures to arsenic or iodine might be differential (see **eTable 2**).

The *dose-response meta-analysis* was conducted using a one-step approach developed in the protocol (https://ntp.niehs.nih.gov/go/78500.76).[5, 6] The approach uses linear mixed models to analyze all available mean effect estimates for the reference group and one or more of the non-reference exposure groups. For each study, the median or mean fluoride level for each exposure group was assigned to its corresponding effect estimate. If median or mean levels by exposure group were not provided, the midpoint of the upper and lower boundaries in every exposure category was assigned as the average level. If the upper boundary for the highest exposure group was not reported, the boundary was assumed to have the same amplitude as the nearest exposure category. For each study, the SMDs and corresponding SEs were used to compare the differences in mean IQ between the exposed and reference groups. The corresponding SMD for the reference group was set to zero for this analysis. The SMDs and corresponding variances were used to estimate a pooled dose-response curve using a restricted maximum likelihood estimation method. To examine a potential nonlinear relationship between exposure to fluoride and children's IQ levels, quadratic terms and restricted cubic splines were created, and a potential departure from a linear trend was assessed by testing the coefficient of the quadratic term and a second spline equal to zero. Models were compared and the best model fit was determined based on the maximum likelihood Akaike information criterion (AIC).[7] The AIC is a goodness-of-fit measure that adjusts for the number of parameters in the model, and lower AIC values indicate better fitting models. Models using a pooled dose-response curve using a restricted maximum likelihood estimation method and a maximum likelihood method were also reported (**eTable4** and **eTable 5**, respectively).

Commented [I2]: See Doc06a_Meta-analysis, 6a.D., page 6 and 7

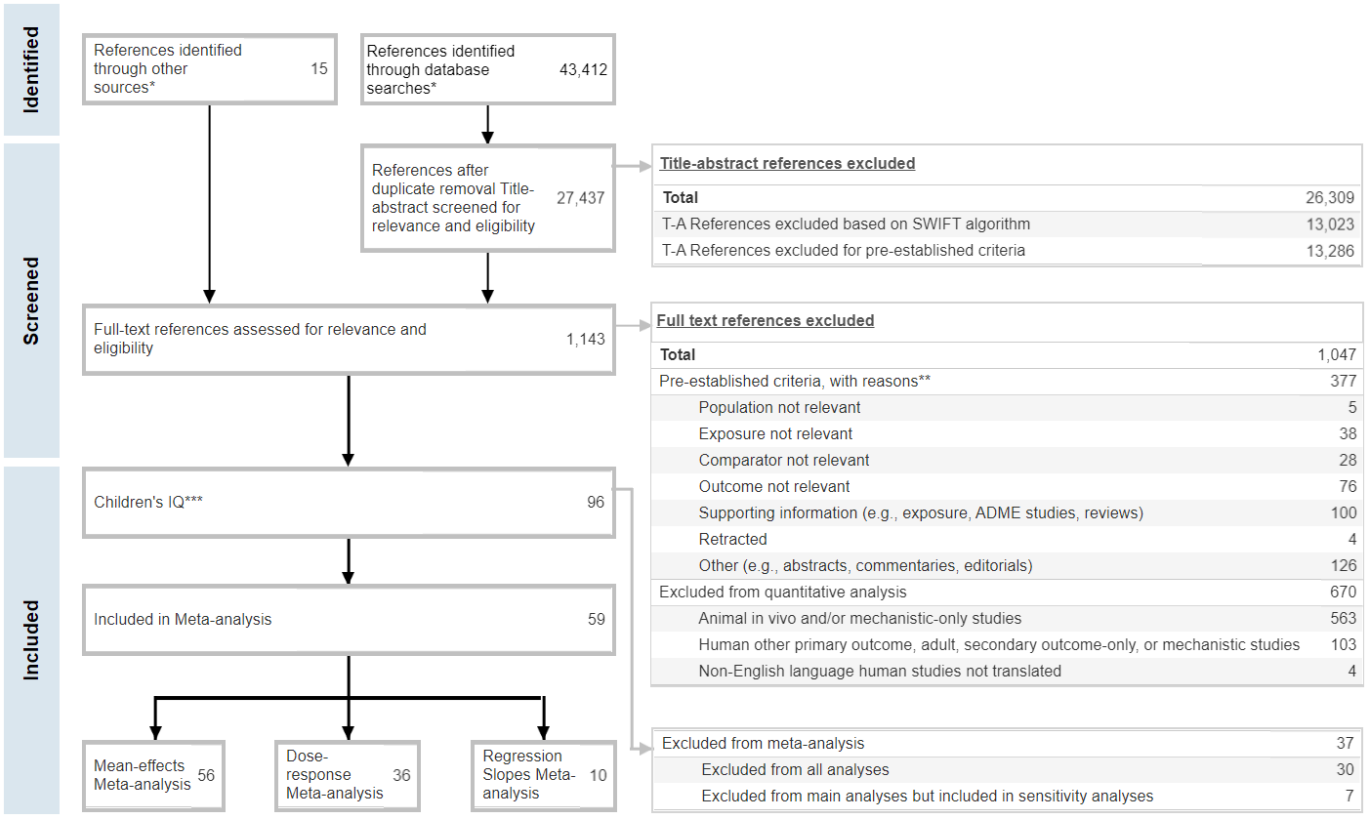
To examine whether there were effects at lower levels of exposure, we conducted sub-group analyses for both drinking water and urinary fluoride measures. Analyses were restricted to <4 mg/L, the EPA's current enforceable drinking water standard for fluoride; <2 mg/L, the EPA's non-enforceable secondary standard for fluoride in drinking water;[8] and <1.5 mg/L, the WHO's guideline for fluoride in drinking water.[9]

Results

Study Sample

Results of the study identification process are provided in **eFigure 1**. Characteristics of the 55 studies that compared mean IQ scores between groups of children with different levels of fluoride exposure are shown in **Table 1** of the main publication (see **eTable 1** for list of excluded publications). Study-specific effect estimates used in the meta-analyses are presented in **eTable 2**. One study per country was conducted in New Zealand, Mexico, Pakistan, and Canada; 4 studies were conducted in Iran, 8 studies were conducted in India, and the remaining 39 studies were performed in China (see **Table 1** of the main publication). Nine study populations were exposed to fluoride from coal burning[10 [translated in Guo et al. 2008a], 12 [translated in Li et al. 2008b], 14-16,17-19]; otherwise, it is assumed that study populations were exposed to fluoride primarily through drinking water. Measures of fluoride exposure included water fluoride (n = 32 studies), dental fluorosis (n = 7), and other non-drinking water sources of exposure to fluoride (e.g., fluoride exposure from coal burning [n = 16]). Fourteen studies presented results for boys and 13 studies reported results for girls; children < 10 years old and children ≥ 10 years old were examined in 13 and 16 studies, respectively (**Table 2**). The Combined Raven's Test for Rural China (CRT-RC) was used to measure

children's IQ in 29 studies. Other measures of IQ included the Wechsler intelligence tests,[20 [translated in Ren et al. 2008], 22 [translated in Wang et al. 2008b], 24, 25] Binet IQ test[10 [translated in Guo et al. 2008a], 26], Raven's Standard Progressive Matrices test,[27-36] Raymond B Cattell test,[37] Japan IQ test,[38, 39] Index of Mental Capacity,[12 [translated in Li et al. 2008b]] and other tests using a doctor-prepared questionnaire.[40, 41] There were 10 low risk-of-bias studies and 45 high risk-of-bias studies (https://hawcproject.org/summary/visual/assessment/405/Figure-X-Meta-analysis-RoB/).



eFigure 1. Prisma Flow Diagram of Study Inclusion

*This information was part of a larger systematic review effort resulting in many studies in the search strategy and PRISMA that were not considered for meta-analysis.

**Studies may have been excluded for more than one reason. The first one identified by the screener was recorded.

*** For the purpose of this PRISMA figure, the Children's IQ count includes three publications[42-44] based on subsamples (i.e., 50–60 children) of a larger Yu et al.[3] cohort. These three publications are not included in the meta-analysis and are not displayed in Figure 1 in the main publication.

eTable 1. List of Excluded Studies from Mean-effects Meta-analysis

Commented [l3]: See Doc05_Meta-analysis, 5.E. (pages 6 and 7), 5.G. (page 8), and 5.M. (page 11)

Reference, Country	Reason for Exclusion
Qin et al. (1990)[45] [translated in Qin et al. 2008], China	Missing mean or SD of outcome measure
Yang et al. (1994)[47] [translated in Yang et al. 2008], China	Overlapping population with Wang et al. (2001)[49]; Table 2 in Yang et al. (1994)[47] seemed incomplete
Wang et al. (2005b)[50] [translated in Wang et al. 2008a], China	Missing mean or SD of outcome measure
Rocha-Amador et al. (2007)[52], Mexico	Missing mean or SD of outcome measure
Liu et al. (2000)[53] [translated in Liu et al. 2008] , China	Overlapping population with Lu et al. (2000)[55]
Sudhir et al. (2009)[56], India	Missing mean or SD of outcome measure
He and Zhang (2010)[57], China	Missing mean or SD of outcome measure
Xiang et al. (2011)[58], China	Overlapping population with Xiang et al. (2003a)[59]
Saxena et al. (2012)[60], India	Missing mean or SD of outcome measure
Wang et al. (2012)[61], China	Overlapping population with Xiang et al. (2003a)[59]
Nagarajappa et al. (2013)[62], India	Seguin Foam Board test; due to the test measuring eye-hand coordination and cognitive-perceptual abilities
Pratap et al.(2013)[63], India	Missing mean or SD of outcome measure
Asawa et al. (2014)[64], India	Seguin Foam Board test; due to the test measuring eye-hand coordination and cognitive-perceptual abilities
Wei et al. (2014)[65], China	Missing mean or SD of outcome measure
Choi et al. (2015)[66], China	Cognitive functions other than IQ
Kundu et al. (2015)[67], India	Unusual IQ scores based on Raven's Standardized Progressive Matrices Test; used only for sensitivity analysis for the *mean-effects meta-analysis*
Aravind et al. (2016)[68], India	Unusually low IQ scores Raven's Standardized Progressive Matrices Test; used only for sensitivity analysis for the *mean-effects meta-analysis*
Jin et al.(2016)[69], China	Cognitive functions other than IQ; potential overlap with Zhang et al. (2015c)[70]
Kumar et al. (2016)[71], India	Seguin Foam Board test; due to the test measuring eye-hand coordination and cognitive-perceptual abilities
Jin et al.(2017)[72], China	Overlap with Jin et al. (2016)[69]; unusual IQ scores reported as percentiles
Razdan et al. (2017)[73], India	Unusually low IQ scores based on Raven's Standardized Progressive Matrices Test; used only for sensitivity analysis for the *mean-effects meta-analysis*
Valdez Jiménez et al. (2017)[74], Mexico	Bayley tests; used only for sensitivity analysis for the *regression slopes meta-analysis*
Wang et al. (2017)[75], China	Overlapping population with Xiang et al. (2003a)[59]

Reference, Country	Reason for Exclusion
Cui et al. (2018)[76], China	Missing mean or SD of outcome measure; used in *regression slopes meta-analysis*
Luo et al. (2018)[77], China	Overlapping population with Lou et al. (2021)[19]
Naik et al. (2018)[78], India	Missing sample sizes by exposure groups. Missing mean and SD for IQ scores
Sharma et al.(2018)[79], India	Missing mean and SD for IQ scores
Soto-Barreras et al. (2019)[80], Mexico	Missing mean or SD of outcome measure
Zhao et al. (2019)[43], China	Overlapping population with Yu et al. (2018)[3], but smaller sample size
Zhou et al. (2019)[44], China	Overlapping population with Yu et al. (2018)[3], but smaller sample size
Till et al.(2020)[81], Canada	Missing mean or SD of outcome measure; used in *regression slopes meta-analysis*
Wang et al. (2020b)[4], China	Missing mean or SD of outcome measure; used in sensitivity analysis for the *regression slopes meta-analysis*
Zhao et al. (2020)[42], China	Overlapping population with Yu et al. (2018)[3], but smaller sample size
Aggeborn and Öhman (2021)[82], Sweden	Cognitive functions other than IQ; cognitive test not specified
Cantoral et al. (2021)[83], Mexico	Bayley tests; used only for sensitivity analysis for the *regression slopes meta-analysis*
Farmus et al. (2021)[84], Canada	Same data as Till et al.(2020)[81]
Guo et al. (2021)[85], China	Overlapping population with Zhao et al. (2018),[86] but smaller sample size; excluded from overall *mean-effects meta-analysis* but used in mean-effects subgroup meta-analysis by age group
Ibarluzea et al. (2021)[87], Spain	Bayley and McCarthy tests; used only for sensitivity analysis for the *mean-effects meta-analysis*, *dose-response meta-analysis*, and *regression slopes meta-analysis*
Wang et al. (2021b)[88], China	Overlapping population with Wang et al. (2021)[89]; cognitive functions other than IQ
Yu et al. (2021)[90], China	Overlapping population with Yu et al. (2018)[3], but smaller sample size
Zhao et al. (2021)[91], China	Missing mean or SD of outcome measure; used in *regression slopes meta-analysis*
Zhou et al. (2021)[92], China	Overlapping population with Yu et al. (2018)[3], but smaller sample size

eTable 2. Study Characteristics and Study-specific Effect Estimates Included in the Meta-analyses and Sensitivity Analyses

Commented [I4]: See Doc05_Meta-analysis, 5.E. (pages 6 and 7) and 5.K. (page 10)

Reference[a] *Study Design*	Study Location	Age Range (Years)	Assessment (Metric, Exposure Groups)	Fluoride Exposure Levels	Mean-effects Meta-analysis N, Mean (SD) [Reference] [Exposed]	Dose-response Mean-effects Meta-analysis N, Mean (SD) [Reference] [Exposed]	Regression Slopes Meta-analysis Slope (SE) or 95% CI per Unit Change Fluoride	Source
Ren et al. (1989)[20] [translated in Ren et al. 2008][me, o] *Cross-sectional*	China	8–14	No fluoride measurement Low iodine village/high fluoride and low iodine village	Not specified	169, 85.00 (22.30) 160, 64.80 (20.40)			Subjects, Methods, Results section
Chen et al. (1991)[93] [translated in Chen et al. 2008][me, w] *Cross-sectional*	China	7–14	Drinking water Nonendemic/endemic fluorosis village	0.89 mg/L (nonendemic) 4.55 mg/L (endemic)	320, 104.03 (14.96) 320, 100.24 (14.52)	320, 104.03 (14.96) 320, 100.24 (14.52)		Results section, Table 1
Guo et al. (1991)[10] [translated in Guo et al. 2008a][me, o] *Cross-sectional*	China	7–13	Serum Reference area using wood/coal burning-related fluoride endemic area	0.1044 ± 0.0652 mg/L (reference) 0.1483 ± 0.0473 mg/L (endemic)	61, 81.39 (10.26) 60, 76.71 (10.85)			Calculated by ICF
Lin et al. (1991)[95me, o] *Cross-sectional*	China	7–14	Urine, drinking water Reference area with iodine supplementation/high fluoride and low iodine village	Urine: 1.6 mg/L (reference area with iodine supplementation) 2.56 mg/L (high fluoride, low iodine village) Water: 0.34 mg/L (low iodine village) 0.88 mg/L (high fluoride, low iodine village)	256, 78.00 (40.07) 250, 71.00 (40.07)			Calculated by ICF
Sun et al. (1991)[38me, o] *Cross-sectional*	China	6.5–12	No fluoride measurement Nonendemic area/endemic (aluminum-fluoride endemic toxicosis)	Fluorosis: 98.36% (endemic)	224, 82.68 (10.91) 196, 72.35 (11.36)			Calculated by ICF
An et al. (1992)[24me, w] *Cross-sectional*	China	7–16	Drinking water Nonhigh/high fluoride area	0.6−1.0 mg/L (nonhigh) 2.1−3.2 mg/L (secondary high) 5.2−7.6 mg/L (high) 2.1−7.6 mg/L (combined high)	121, 84.00 (12.10) 121, 75.90 (13.60)	121, 84.00 (12.10) 56, 76.10 (13.90) 65, 75.60 (13.30)		Table 1, Table 2

Reference[a] *Study Design*	Study Location	Age Range (Years)	Assessment (Metric, Exposure Groups)	Fluoride Exposure Levels	Mean-effects Meta-analysis N, Mean (SD) [Reference] [Exposed]	Dose-response Mean-effects Meta-analysis N, Mean (SD) [Reference] [Exposed]	Regression Slopes Meta-analysis Slope (SE) or 95% CI per Unit Change Fluoride	Source
Li et al. (1994)[12] [translated in Li et al. 2008b][me, o] *Cross-sectional*	China	12–13	Grain (cooked by burning high-fluoride coal) Reference group (no dental fluorosis)/high fluoride group I (no dental fluorosis)/high fluoride group II (dental fluorosis present)/high fluoride group III (dental fluorosis present)	0.5 mg/kg (reference group) 4.7 mg/kg (group I) 5.2 mg/kg (group II) 31.6 mg/kg (group III)	49, 267.20 (39.50) 36, 240.00 (30.80)			Table 1
Xu et al. (1994)[26me, w] *Cross-sectional*	China	8–14	Drinking water Reference region/low- and high-fluoride regions[b]	0.8 mg/L (reference region) 0.38 mg/L (low fluoride) 1.8 mg/L (high fluoride)	32, 83.83 (9.10) 97, 79.25 (2.25)	32, 83.83 (9.10) 21, 80.21 (8.27) 97, 79.25 (2.25)		Chart 1
Li et al. (1995)[14me, o, u] *Cross-sectional*	China	8–13	Urine, dental fluorosis index (DFI) Nonfluorosis/fluorosis area due to soot from coal burning	1.02 mg/L; DFI: <0.4 (nonfluorosis) 1.81 mg/L; DFI: 0.8 (slight fluorosis) 2.01 mg/L; DFI: 2.5 (medium fluorosis) 2.69 mg/L; DFI: 3.2 (severe fluorosis)	226, 89.90 (10.40) 230, 80.30 (12.90)	226, 89.90 (10.40) 227, 89.70 (12.70) 224, 79.70 (12.70) 230, 80.30 (12.90)		Table 2
Wang et al. (1996)[22] [translated in Wang et al. 2008b][me, o, w] *Cross-sectional*	China	4–7	Drinking water (well) Low/high fluoride regions Fluoride exposure from drinking water, contaminated food, and coal burning	0.58−1.0 mg/L (low) >1.0−8.6 mg/L (high)	83, 101.23 (15.84) 147, 95.64 (14.34)	83, 101.23 (15.84) 147, 95.64 (14.34)		Table 1
Yao et al. (1996)[28me, w] *Cross-sectional*	China	8–12	Drinking water Nonendemic/endemic fluorosis areas	1 mg/L (nonendemic) 2 mg/L (slightly endemic) 11 mg/L (severely endemic)	270, 98.46 (13.21) 78, 92.53 (12.34)	270, 98.46 (13.21) 188, 94.89 (11.15) 78, 92.53 (12.34)		Table 2
Zhao et al. (1996)[96me, w] *Cross-sectional*	China	7–14	Drinking water Low fluoride village (Xinghua)/high fluoride village (Sima)	0.91 mg/L (low) 4.12 mg/L (high)	160, 105.21 (14.99) 160, 97.69 (13.00)	160, 105.21 (14.99) 160, 97.69 (13.00)		Table 1

Reference[a] *Study Design*	Study Location	Age Range (Years)	Assessment (Metric, Exposure Groups)	Fluoride Exposure Levels	Mean-effects Meta-analysis N, Mean (SD) [Reference] [Exposed]	Dose-response Mean-effects Meta-analysis N, Mean (SD) [Reference] [Exposed]	Regression Slopes Meta-analysis Slope (SE) or 95% CI per Unit Change Fluoride	Source
Yao (1997)[27me, w*] *Cross-sectional*	China	7–12	Drinking water Nonfluorosis/fluorosis area with water improvements/fluorosis area without water improvements	0.4 mg/L (nonfluorosis area) 0.33 mg/L (fluorosis area with water improvement) 2 mg/L (fluorosis area without water improvement)	314, 99.98 (12.21) 183, 94.89 (11.15)	314, 99.98 (12.21) 326, 97.83 (11.27) 183, 94.89 (11.15)		Section 2.1 Intelligence Tests, page 2
Zhang et al. (1998)[39me, o] *Cross-sectional*	China	4–10	Drinking water Reference/high fluoride group (all observation groups included arsenic exposure)	0.58 mg/L (reference) 0.8 mg/L (high fluoride)	52, 87.69 (11.04) 51, 85.62 (13.23)			Table 1
Lu et al. (2000)[55me, w, u] *Cross-sectional*	China	10–12	Urine, drinking water Low/high fluoride area	Urine: 1.43 ± 0.64 mg/L (low) 4.99 ± 2.57 mg/L (high) Water: 0.37 ± 0.04 mg/L (low) 3.15 ± 0.61 mg/L (high)	58, 103.05 (13.86) 60, 92.27 (20.45)	58, 103.05 (13.86) 60, 92.27 (20.45)		Table 1
Hong et al. (2001)[97] [translated in Hong et al. 2008] [me, w] *Cross-sectional*	China	8–14	Drinking water Reference/high fluoride[b]	0.75 mg/L (reference) 2.90 mg/L (high fluoride)	32, 82.79 (8.98) 85, 80.58 (2.28)	32, 82.79 (8.98) 85, 80.58 (2.28)		Table 2
Hong et al. (2001b)[99me, o] *Cross-sectional*	China	8–14	Urine, drinking water Nonendemic/endemic fluorosis areas (high fluoride, high iodine)	Urine: 0.796 ± 0.53 mg/L (nonendemic) 2.09 ± 1.03 mg/L (endemic) Water: 0.48 mg/L (nonendemic) 2.81 mg/L (endemic)	30, 80.66 (11.93) 31, 75.89 (7.74)			Table 3, Table 4
Wang et al. (2001)[49,me, o] *Cross-sectional*	China	8–12	Urine, drinking water Reference point (low fluoride, low iodine)/investigative point (high fluoride, high iodine)	Urine: 0.82 mg/L (low fluoride, low iodine) 3.08 mg/L (high fluoride, high iodine) Water: 0.5 mg/L (low fluoride, low iodine) 2.97 mg/L (high fluoride, high iodine)	30, 81.67 (11.97) 30, 76.67 (7.75)			Table 2

Reference[a] *Study Design*	Study Location	Age Range (Years)	Assessment (Metric, Exposure Groups)	Fluoride Exposure Levels	Mean-effects Meta-analysis N, Mean (SD) [Reference] [Exposed]	Dose-response Mean-effects Meta-analysis N, Mean (SD) [Reference] [Exposed]	Regression Slopes Meta-analysis Slope (SE) or 95% CI per Unit Change Fluoride	Source
Li et al. (2003)[100] [translated in Li et al. 2008c][me] *Cross-sectional*	China	6–13	No fluoride measurement Reference/endemic fluorosis areas	Not specified	236, 93.78 (14.30) 720, 92.07 (17.12)			Table 1
Xiang et al. (2003a)[59,me, w*, u] *Cross-sectional*	China	8–13	Urine, drinking water Nonendemic/endemic fluorosis areas	Urine: 1.11 ± 0.39 mg/L (reference) 3.47 ± 1.95 mg/L (high fluoride) Water: 0.36 ± 0.15 mg/L (nonendemic) 0.75 ± 0.14 mg/L (endemic fluorosis area group A) 1.53 ± 0.27 mg/L (endemic fluorosis area group B) 2.46 ± 0.3 mg/L (endemic fluorosis area group C) 3.28 ± 0.25 mg/L (endemic fluorosis area group D) 4.16 ± 0.22 mg/L (endemic fluorosis area group E) 2.47 ± 0.79 mg/L (high fluoride)	290, 100.41 (13.21) 222, 92.02 (13.00)	290, 100.41 (13.21) 9, 99.56 (14.13) 42, 95.21 (12.22) 111, 92.19 (12.98) 52, 89.88 (11.98) 8, 78.38 (12.68)		Table 6, Table 8
Wang et al. (2005)[102,me, w, u] *Cross-sectional*	China	8–12	Urine, drinking water Reference/high fluoride group[c]	Urine: 1.51 mg/L(reference) 5.09 mg/L (high fluoride group) Water: 0.48 mg/L (reference) 8.31 mg/L (high fluoride group)	196, 112.36 (14.87) 253, 107.83 (15.45)	196, 112.36 (14.87) 253, 107.83 (15.45)		Table 1
Seraj et al. (2006)[29,me, w] *Cross-sectional*	Iran	7–11	Drinking water Low/high fluoride area	0.4 ppm (low) 2.5 ppm (high)	85, 98.90 (12.90) 41, 87.90 (11.00)	85, 98.90 (12.90) 41, 87.90 (11.00)		Methodology, Findings section (Text under Table 2)
Wang et al. (2006)[103,me, w, u] *Cross-sectional*	China	8–12	Urine, drinking water Reference/high (area severely affected by fluorosis)	Urine: 1.51 ± 1.66 mg/L (reference) 5.50 ± 2.40 mg/L (high) Water: 0.73 ± 0.28 mg/L (reference) 5.54 ± 3.88 mg/L (high)	166, 111.55 (15.19) 202, 107.46 (15.38)	166, 111.55 (15.19) 202, 107.46 (15.38)		Table 2

Reference[a] *Study Design*	Study Location	Age Range (Years)	Assessment (Metric, Exposure Groups)	Fluoride Exposure Levels	Mean-effects Meta-analysis N, Mean (SD) [Reference] [Exposed]	Dose-response Mean-effects Meta-analysis N, Mean (SD) [Reference] [Exposed]	Regression Slopes Meta-analysis Slope (SE) or 95% CI per Unit Change Fluoride	Source
Fan et al. (2007)[104,me, w, u] *Cross-sectional*	China	7–14	Urine, drinking water Low/high fluoride area	Urine: 1.78 ± 0.46 mg/L (low) 2.89 ± 1.97 mg/L (high) Water: 1.03 mg/L (low) 3.15 mg/L (high)	37, 98.41 (14.75) 42, 96.11 (12.00)	37, 98.41 (14.75) 42, 96.11 (12.00)		Table 1
Trivedi et al. (2007)[41,me, w, u] *Cross-sectional*	India	12–13	Urine, drinking water Low/high fluoride area	Urine: 2.30 ± 0.28 mg/L (low) 6.13 ± 0.67 mg/L (high) Water: 2.01 ± 0.009 mg/L (low) 5.55 ± 0.41 mg/L (high)	101, 104.44 (12.36) 89, 91.72 (10.66)	101, 104.44 (12.36) 89, 91.72 (10.66)		Table 2
Wang et al. (2007)[105,me, o, u, w] *Cross-sectional*	China	8–12	Urine, drinking water Low fluoride, low arsenic/high fluoride, low arsenic area	Urine: 1.5 ± 1.6 mg/L (low fluoride, low arsenic) 5.1 ± 2.0 mg/L (high fluoride, low arsenic) Water: 0.5 ± 0.2 mg/L (low fluoride, low arsenic) 8.3 ± 1.9 mg/L (high fluoride, low arsenic)	196, 104.80 (14.70) 253, 100.50 (15.80)	196, 104.80 (14.70) 253, 100.50 (15.80)		Table 2, Table 3
Li et al. (2009)[15,me, o, u*] *Cross-sectional*	China	8–12	Urine Endemic fluorosis region caused by coal burning (reference/mild/medium /severe) Degree of dental fluorosis (normal/suspected/ very mild/mild/medium/ severe)	0.962 ± 0.517 mg/L (reference) 1.235 ± 0.426 mg/L (mild) 1.670 ± 0.663 mg/L (medium) 2.336 ± 1.128 mg/L (severe) 0.867 ± 0.233 mg/L (normal) 1.094 ± 0.355 mg/L (suspected) 1.173 ± 0.480 mg/L (very mild) 1.637 ± 0.682 mg/L (mild) 2.005 ± 0.796 mg/L (medium) 2.662 ± 1.093 mg/L (severe)	20, 102.70 (17.61) 20, 93.85 (18.11)	20, 102.70 (17.61) 20, 97.30 (18.56) 20, 93.90 (17.60) 20, 93.85 (18.11)		Table 1
Li et al. (2010)[106,me] *Cross-sectional*	China	7–10	No fluoride measurement Nondental fluorosis children/dental fluorosis children	Not specified	329, 97.36 (18.24) 347, 98.73 (21.07)			Table 3

Reference[a] *Study Design*	Study Location	Age Range (Years)	Assessment (Metric, Exposure Groups)	Fluoride Exposure Levels	Mean-effects Meta-analysis N, Mean (SD) [Reference] [Exposed]	Dose-response Mean-effects Meta-analysis N, Mean (SD) [Reference] [Exposed]	Regression Slopes Meta-analysis Slope (SE) or 95% CI per Unit Change Fluoride	Source
Ding et al. (2011)[107,me, u*, rs] *Cross-sectional*	China	7–14	Dental fluorosis (normal/ questionable/very mild/mild/ moderate) Urine Mean urinary fluoride levels (10 groups)	0.80 ± 0.55 mg/L (normal) 1.13 ± 0.73 mg/L (questionable) 1.11 ± 0.74 mg/L (very mild) 1.31 ± 0.78 mg/L (mild) 1.46 ± 0.79 mg/L (moderate) 0.26 mg/L (group 1) 0.45 mg/L (group 2) 0.56 mg/L (group 3) 0.66 mg/L (group 4) 0.75 mg/L (group 5) 0.89 mg/L (group 6) 1.08 mg/L (group 7) 1.33 mg/L (group 8) 1.74 mg/L (group 9) 2.96 mg/L (group 10) Range: 0.10−3.55 mg/L	136, 104.07 (12.30) 28, 103.54 (13.59)	136, 104.07 (12.30) 54, 103.00 (16.10) 74, 102.11 (15.05) 39, 106.03 (12.33) 28, 103.54 (13.59)	−0.59 (−1.09, −0.08) per 1 mg/L urinary F	Table 2, Section 3 Results and discussion (under Fig. 2)
Eswar et al. (2011)[31,me, w] *Cross-sectional*	India	12–14	Drinking water Low/high fluoride villages	0.29 mg/L (low) 2.45 mg/L (high)	65, 88.80 (15.30) 68, 86.30 (12.80)	65, 88.80 (15.30) 68, 86.30 (12.80)		Table 1
Kang et al. (2011)[108me, o] *Cross-sectional*	China	6–12	Drinking water Reference/high fluoride areas (both areas high arsenic exposure)	1.24 ± 0.74 mg/L (all children) <1.2 mg/L (reference) ≥1.2 mg/L (high fluoride)	90, 96.8 (12.7) 178, 96.8 (16.3)			Table 1. Section 2.1
Poureslami et al. (2011)[32,me, w] *Cross-sectional*	Iran	7–9	Drinking water Reference/endemic dental fluorosis city	0.41 mg/L (reference) 2.38 mg/L (endemic)	60, 97.80 (15.95) 59, 91.37 (16.63)	60, 97.80 (15.95) 59, 91.37 (16.63)		Table 3, Results section (under Table 3)
Shivaprakash et al. (2011)[33,me, w] *Cross-sectional*	India	7–11	Drinking water No fluorosis/fluorosis severity groups (mild/moderate/ severe)/all fluorosis	<0.5 ppm (no fluorosis) 2.5–3.5 ppm (mild) 2.5–3.5 ppm (moderate) 2.5–3.5 ppm (severe) 2.5–3.5 ppm (all)	80, 76.36 (20.84) 80, 66.63 (18.09)	80, 76.36 (20.84) 80, 66.63 (18.09)		Table 1

Reference[a] *Study Design*	Study Location	Age Range (Years)	Assessment (Metric, Exposure Groups)	Fluoride Exposure Levels	Mean-effects Meta-analysis N, Mean (SD) [Reference] [Exposed]	Dose-response Mean-effects Meta-analysis N, Mean (SD) [Reference] [Exposed]	Regression Slopes Meta-analysis Slope (SE) or 95% CI per Unit Change Fluoride	Source
Seraj et al. (2012)[30,me, w] *Cross-sectional*	Iran	6–11	Drinking water Normal/medium/high fluoride levels	0.8 ± 0.3 mg/L (normal) 3.1 ± 0.9 mg/L (medium) 5.2 ± 1.1 mg/L (high)	91, 97.77 (18.91) 96, 88.58 (16.01)	91, 97.77 (18.91) 106, 89.03 (12.99) 96, 88.58 (16.01)		Table 2
Trivedi et al. (2012)[40,me, w, u] *Cross-sectional*	India	12–13	Urine, ground water Low/high fluoride area	Urine: 0.42 ± 0.23 mg/L (low) 2.69 ± 0.92 mg/L (high) Water: 0.84 ± 0.38 mg/L (low) 2.3 ± 0.87 mg/L (high)	50, 97.17 (17.96) 34, 92.58 (18.25)	50, 97.17 (17.96) 34, 92.58 (18.25)		Table 3, Results section (above Table 3)
Wang et al. (2012b)[109me] *Cross-sectional*	China	Primary school age	No fluoride measurement Reference/high fluoride areas	Not specified	455, 98.36 (14.56) 800, 92.21 (18.45)			Table 1
Bai et al. (2014)[16,me, o] *Cross-sectional*	China	8–12	Urine Coal-burning-borne fluorosis areas (reference/lightly-affected/seriously-affected)	0.54 mg/L (reference) 0.81 mg/L (lightly-affected area) 1.96 mg/L (seriously-affected area)	164, 107.92 (13.62) 162, 101.22 (15.97)			Table 2
Karimzade et al. (2014)[37,me, w] *Cross-sectional*	Iran	9–12	Drinking water Low/high fluoride area	0.25 mg/L (low) 3.94 mg/L (high)	20, 104.25 (20.75) 19, 81.21 (16.17)	20, 104.25 (20.75) 19, 81.21 (16.17)		Table 1

Reference[a] *Study Design*	Study Location	Age Range (Years)	Assessment (Metric, Exposure Groups)	Fluoride Exposure Levels	Mean-effects Meta-analysis N, Mean (SD) [Reference] [Exposed]	Dose-response Mean-effects Meta-analysis N, Mean (SD) [Reference] [Exposed]	Regression Slopes Meta-analysis Slope (SE) or 95% CI per Unit Change Fluoride	Source
Broadbent *et al.* (2015)[25,me, w*] *Prospective Cohort*	New Zealand	7–13	Drinking water Area without community water fluoridation (low)/area with community water fluoridation (high) Fluoride tablet use (never/ever) Fluoride toothpaste use (never/sometimes/always)	Water: 0.0–0.3 mg/L (low) 0.7–1.0 mg/L (high) Tablet use: 0 mg (never used) 0.5 mg (ever used) Range not specified for fluoride toothpaste use (always/sometimes/never)	99, 99.80 (14.50) 891, 100.00 (15.10)	99, 99.80 (14.50) 891, 100.00 (15.10)		Table 1
Khan et al. (2015)[34,me] *Cross-sectional*	India	6–11	Drinking water Low fluoride areas (Tiwariganj)/high fluoride areas (Unnao) Fluorosis grades (normal/very mild/mild/moderate/severe)	0.19 mg/L (Tiwariganj) 2.41 mg/L (Unnao) Ranges not specified by fluorosis grades	241, 110.10 (9.00) 5, 62.40 (2.40)			Table/Fig-5
Kundu et al. (2015)[67,sa] *Cross-sectional*	India	8−12	Drinking water Low fluoride areas/high fluoride areas	Not specified	100, 85.80 (18.85) 100, 76.20 (19.10)			Table 2
Sebastian and Sunitha (2015)[35,me, w*] *Cross-sectional*	India	10–12	Drinking water Low/normal/high fluoride villages	0.40 mg/L (low) 1.2 mg/L (normal) 2.0 mg/L (high)	135, 86.37 (13.58) 135, 80.49 (12.67)	135, 86.37 (13.58) 135, 88.60 (14.01) 135, 80.49 (12.67)		Table 1, Table 2
Zhang et al. (2015b)[110,me, w*, u, rs] *Cross-sectional*	China	10–12	Urine, drinking water, serum Reference/high fluoride areas	Urine: 1.10 ± 0.67 mg/L (reference) 2.40 ± 1.01 mg/L (high) Water: 0.63 (0.58–0.68) mg/L (reference) 1.40 (1.23–1.57) mg/L (high) Serum: 0.06 ± 0.03 (reference) 0.18 ± 0.11 serum (high)	96, 109.42 (13.30) 84, 102.33 (13.46)	96, 109.42 (13.30) 84, 102.33 (13.46)	−2.42 (−4.59, −0.24) per 1 mg/L urinary F	Table 1, Table 3

Reference[a] *Study Design*	Study Location	Age Range (Years)	Assessment (Metric, Exposure Groups)	Fluoride Exposure Levels	Mean-effects Meta-analysis N, Mean (SD) [Reference] [Exposed]	Dose-response Mean-effects Meta-analysis N, Mean (SD) [Reference] [Exposed]	Regression Slopes Meta-analysis Slope (SE) or 95% CI per Unit Change Fluoride	Source
Zhang et al. (2015c)[70me, o] *Cross-sectional*	China	7−13	Urine Coal-burning endemic fluorosis area Reference (no dental fluorosis)/mild dental fluorosis/middle dental fluorosis/critically ill dental fluorosis	0.83 ± 0.71 mg/L (reference) 1.54 ± 0.57 mg/L (mildly ill) 2.41 ± 0.76 mg/L (moderately ill) 3.32 ± 1.02 mg/L (critically ill)	30, 110.34 (11.52) (reference) 30, 90.52 (10.37) (critically ill)			Table 1, Table 3
Aravind et al. (2016)[68,sa] *Cross-sectional*	India	10−12	Drinking water Low/high fluoride levels	<1.2 ppm (low) >2 ppm (high)	96, 41.03 (16.36) 96, 31.59 (16.81)			Table 1
Das and Mondal (2016)[111,me, u] *Cross-sectional*	India	6–18	Urine, drinking water intake Dental fluorosis (normal/questionable/very mild/ mild/ moderate/severe)	Urine: 2.91 ± 1.76 mg/L (normal) 2.50 ± 2.39 mg/L (questionable) 2.58 ± 1.31 mg/L (very mild) 2.95 ± 1.44 mg/L (mild) 4.82 ± 3.57 mg/L (moderate) 3.81 ± 2.51 mg/L (severe) Water: 0.069 ± 0.021 mg/kg-d (normal) 0.064 ± 0.004 mg/kg-d (questionable) 0.060 ± 0.036 mg/kg-d (very mild) 0.060 ± 0.030 mg/kg-d (mild) 0.099 ± 0.063 mg/kg-d (moderate) 0.093 ± 0.040 mg/kg-d (severe)	4, 108.30 (53.20) 23, 85.91 (37.68)	4, 108.30 (53.20) 17, 103.18 (33.35) 27, 107.70 (27.92) 35, 92.83 (26.90) 43, 84.51 (35.16) 23, 85.91 (37.68)		Table 3
Mondal et al. (2016)[36,me, w] *Cross-sectional*	India	10–14	Drinking water Low/high fluoride areas	Not reported (low) 0.33–18.08 mg/L (high)	22, 26.41(10.46) 18, 21.17 (6.77)	22, 26.41 (10.46) 18, 21.17 (6.77)		Table 9

Reference[a] *Study Design*	Study Location	Age Range (Years)	Assessment (Metric, Exposure Groups)	Fluoride Exposure Levels	Mean-effects Meta-analysis N, Mean (SD) [Reference] [Exposed]	Dose-response Mean-effects Meta-analysis N, Mean (SD) [Reference] [Exposed]	Regression Slopes Meta-analysis Slope (SE) or 95% CI per Unit Change Fluoride	Source
Bashash et al. (2017)[112,me, u, rs] *Prospective Cohort*	Mexico	6−12	Maternal urine Reference/high fluoride levels (based on children urinary fluoride)	<0.80 mg/L (reference) ≥0.80 mg/L (high)	77, 95.37 (10.31) 112, 96.80 (11.16)	77, 95.37 (10.31) 112, 96.80 (11.16)	−2.50 (−4.12, −0.59) per 0.5 mg/L maternal urinary F	Abstract, Table 3
Razdan et al. (2017)[73,sa] *Cross-sectional*	India	12–14	Drinking water Low/high fluoride levels	0.6 ppm (low) 4.99 ppm (high)	69, 38.61 (6.34) 75, 13.95 (5.14)			Table 2
Valdez Jiménez et al. (2017)[74sa] *Prospective Cohort*	Mexico	Infancy	Maternal urine, drinking water	Urine: 1.9 ± 1.0 mg/L (1st trimester) 2.0 ± 1.1 mg/L (2nd trimester) 2.7 ± 1.1 mg/L (3rd trimester) Water: 2.6 ± 1.1 mg/L (1st trimester) 3.1 ± 1.1 mg/L (2nd trimester) 3.7 ± 1.0 mg/L (3rd trimester)			Bayley MDI: −19.05 (8.9) per 1 log10 mg/L maternal urinary F (1st trimester) −19.34 (7.46) per 1 log10 mg/L maternal urinary F (2nd trimester)	Table 2, Table 4
Cui et al. (2018)[76,rs] *Cross-sectional*	China	7–12	Urine	Boys: 1.3 (0.9−1.7)[d] mg/L Girls: 1.2 (0.9−1.6)[d] mg/L			−2.47 (−4.93, −0.01) per 1 log urinary F	Table 2
Yu et al. (2018)[3,me, w, u*, rs] *Cross-sectional*	China	7–13	Maternal urine Low/medium/high fluoride ranges Drinking water Normal/high fluoride	Urine: 0.01–1.60 mg/L (low) 1.60–2.50 mg/L (medium) 2.50–5.54 mg/L (high) Water: ≤1 mg/L (normal) >1 mg/L (high) Overall: 0.01−5.54 mg/L (urine) 0.20−3.90 mg/L (water)	1636, 107.40 (13.00) 1250, 106.40 (12.30)	1636, 107.40 (13.00) 1250, 106.40 (12.30)	0.36 (−0.29, 1.01) per 0.5 mg/L maternal urinary F	Table 1, Table 3
Zhao et al. (2018)[86me, o] *Cross-sectional*	China	7–12	Urine Reference/exposed areas All areas with iodine exposure	≤2.16 mg/L (reference) >2.16 mg/L (exposed)	199, 114.52 (12.72) 100, 109.59 (14.24)			Table 4

Reference[a] *Study Design*	Study Location	Age Range (Years)	Assessment (Metric, Exposure Groups)	Fluoride Exposure Levels	Mean-effects Meta-analysis N, Mean (SD) [Reference] [Exposed]	Dose-response Mean-effects Meta-analysis N, Mean (SD) [Reference] [Exposed]	Regression Slopes Meta-analysis Slope (SE) or 95% CI per Unit Change Fluoride	Source
Green et al. (2019)[113,me, w*, u*, rs] *Prospective Cohort*	Canada	3−4	Maternal urine, drinking water, maternal fluoride intake Nonfluoridated/ fluoridated area	Urine: 0.40 ± 0.27 mg/L (nonfluoridated) 0.69 ± 0.42 mg/L (fluoridated) Water: 0.13 ± 0.06 mg/L (nonfluoridated) 0.59 ± 0.08 mg/L (fluoridated) Intake: 0.30 ± 0.26 mg/day (nonfluoridated) 0.93 ± 0.43 mg/day (fluoridated) Overall: 0.51 ± 0.36 mg/L (urine) 0.54 ± 0.44 mg/day (intake) 0.31 ± 0.23 mg/L (water)	238, 108.07 (13.31) 162, 108.21 (13.72)	238, 108.07 (13.31) 162, 108.21 (13.72)	−1.95 (−5.19, 1.28) per 1 mg/L maternal urinary F −5.29 (−10.39, −0.19) per 1 mg/L water F −3.66 (−7.16, 0.15) per 1 mg maternal F intake	Table 2, text page 945, eTable 4
Cui et al. (2020)[114,me, u] *Cross-sectional*	China	7–12	Urine Low/medium/high fluoride levels	<1.6 mg/L (low) 1.6–2.5 mg/L (medium) >=2.5 mg/L (high)	396, 112.16 (11.50) 36, 110.00 (14.92)	396, 112.16 (11.50) 66, 112.05 (12.01) 36, 110.00 (14.92)		Table 1

Reference[a] *Study Design*	Study Location	Age Range (Years)	Assessment (Metric, Exposure Groups)	Fluoride Exposure Levels	Mean-effects Meta-analysis N, Mean (SD) [Reference] [Exposed]	Dose-response Mean-effects Meta-analysis N, Mean (SD) [Reference] [Exposed]	Regression Slopes Meta-analysis Slope (SE) or 95% CI per Unit Change Fluoride	Source
Till et al. (2020)[81,rs] *Prospective Cohort*	Canada	3−4	Residence, maternal urine, drinking water, infant fluoride intake from formula Nonfluoridated areas/fluoridated	Urine: 0.38−0.42 mg/L (nonfluoridated) 0.64−0.70 mg/L (fluoridated) Water: 0.13 mg/L (nonfluoridated) 0.58 mg/L (fluoridated) Intake: 0.02−0.08 mg/day (nonfluoridated) 0.12−0.34 mg/day (fluoridated)			−2.69 (−7.38, 2.01) per 0.5 mg/day infant F intake (formula)	Table 2
Wang et al. (2020b)[4,sa] *Cross-sectional*	China	7−13	Urine, drinking water	Urine: 0.01−5.54 mg/L Water: 0.20−3.90 mg/L			−1.214 (−1.987, −0.442) per 1 mg/L urinary F −1.037 (−2.040, −0.035) per 1 mg/L urinary F (males) −1.379 (−2.628, −0.129) per 1 mg/L urinary F (females); −1.587 (−2.607, −0.568) per 1 mg/L water F −1.422 (−2.792, −0.053) per 1 mg/L water F (males) −1.649 (−3.201, −0.097) per 1 mg/L water F (females)	Table 4
Wang et al. (2020c)[18me, o] *Cross-sectional*	China	7−12	Urine Coal-burning endemic fluorosis area Nonendemic/endemic fluorosis regions	0.461 ± 0.210 mg/L (nonendemic) 0.689 ± 0.502 mg/L (endemic)	100, 97 (20.3) 170, 82.5 (21.7)			Section 2.1, Table 2

Reference[a] *Study Design*	Study Location	Age Range (Years)	Assessment (Metric, Exposure Groups)	Fluoride Exposure Levels	Mean-effects Meta-analysis N, Mean (SD) [Reference] [Exposed]	Dose-response Mean-effects Meta-analysis N, Mean (SD) [Reference] [Exposed]	Regression Slopes Meta-analysis Slope (SE) or 95% CI per Unit Change Fluoride	Source
Xu et al. (2020)[115] me, u*, rs *Cross-sectional*	China	7−13	Urine Reference/high prenatal exposure only/high childhood exposure only/both prenatal and childhood exposure group	0.82 ± 0.30 mg/L (reference) 0.98 ± 0.29 mg/L (high prenatal exposure only) 2.05 ± 0.58 mg/L (high childhood exposure only) 2.13 ± 0.59 mg/L (both prenatal and childhood exposure group)	228, 123.92 (12.50) 141, 123.04 (11.24)	228, 123.92 (12.50) 107, 119.76 (11.28) 157, 124.65 (10.88) 141, 123.04 (11.24)	−0.055 (−1.626, 1.517) per 1 mg/L urinary F 2.785 (−0.832, 6.403) per 1 mg/L urinary F (<1.7 mg/L) −4.965 (−9.198, −0.732) per 1 mg/L urinary F (≥1.7 mg/L) 4.054 (−3.169, 11.277) per 1 mg/L prenatal urinary F (<1.7 mg/L) −3.929 (−9.396, 1.538) per 1 mg/L prenatal urinary F (≥1.7 mg/L) 3.146 (−1.138, 7.430) per 1 mg/L postnatal urinary F (<1.7 mg/L) −6.595 (−13.323, 0.133) per 1 mg/L postnatal urinary F (≥1.7 mg/L)	Table 1, Table 3, author correspondence
Cantoral et al. (2021)[83sa] *Prospective Cohort*	Mexico	1−2	Maternal fluoride intake	1.12 ± 0.54 mg/day			Bayley III cognitive scores: −1.14 (−3.26, 0.99) per 0.5 mg/L maternal F intake 0.07 (−2.37, 2.51) per 0.5 mg/L maternal F intake (females) −3.50 (−6.58, −0.42) per 0.5 mg/L maternal F intake (males)	Table 3, Table 4

Reference[a] *Study Design*	Study Location	Age Range (Years)	Assessment (Metric, Exposure Groups)	Fluoride Exposure Levels	Mean-effects Meta-analysis N, Mean (SD) [Reference] [Exposed]	Dose-response Mean-effects Meta-analysis N, Mean (SD) [Reference] [Exposed]	Regression Slopes Meta-analysis Slope (SE) or 95% CI per Unit Change Fluoride	Source
Guo et al., (2021)[85me] *Cross-sectional*	China	7−12	Urine Reference/exposed areas (all areas with iodine exposure)	1.16 mg/L (reference) 1.29 mg/L (iodine area 1) 2.01 mg/L (iodine area 2)	7−9 years: 71, 116.71 (12.16) 35, 118.11 (12.8) 22, 113.95 (12.26) 10−12 years: 79, 109.86 (12.05) 48, 110.83 (10.58) 44, 105.39 (13.6)			Table 2, Table 3
Ibarluzea et al. (2021)[87sa] *Prospective Cohort*	Spain	1, 4	Maternal urine Nonfluorinated/ fluoridated communities	Urine: 0.38 ± 0.27 mg/L (nonfluorinated) 0.70 ± 0.41 mg/L (fluoridated) Water: <0.1 mg/L (nonfluorinated) 0.81 ± 0.15 mg/L (fluoridated)	Bayley MDI scores: 153, 97.696 (14.91) 160, 100.395 (15.411) McCarthy GCI scores: 123, 98.666 (15.531) 124, 101.473 (15.423)	Bayley MDI scores: 153, 97.696 (14.91) 160, 100.395 (15.411) McCarthy GCI scores: 123, 98.666 (15.531) 124, 101.473 (15.423)	Bayley MDI scores: 4.67 (−1.78, 11.13) per 1 mg/L maternal urinary F 7.86 (−1.68, 17.40) per 1 mg/L maternal urinary F (males) 1.77 (−7.32, 10.87) per 1 mg/L maternal urinary F (females) McCarthy GCI scores: −2.16 (−8.56, 4.23) per 1 mg/L maternal urinary F −1.79 (−11.85, 8.27) per 1 mg/L maternal urinary F (males) −3.60 (−12.07, 4.86) per 1 mg/L maternal urinary F (females)	Section 2.2, author correspondence
Lou et al. (2021)[19me, o] *Cross-sectional*	China	8−12	Coal-burning endemic fluorosis area No fluoride measurement Nondental fluorosis children/dental fluorosis children	Not specified	44, 96.64 (11.70) 55, 88.51 (12.77)			Table 4

Reference[a] *Study Design*	Study Location	Age Range (Years)	Assessment (Metric, Exposure Groups)	Fluoride Exposure Levels	Mean-effects Meta-analysis N, Mean (SD) [Reference] [Exposed]	Dose-response Mean-effects Meta-analysis N, Mean (SD) [Reference] [Exposed]	Regression Slopes Meta-analysis Slope (SE) or 95% CI per Unit Change Fluoride	Source
Saeed et al. (2021)[116me, o, rs] *Cross-sectional*	Pakistan	5−16	Urine, drinking water Reference/high fluoride areas Co-exposure with arsenic	Urine: 0.24 ± 0.15 mg/L (reference) 3.27 ± 2.60 mg/L (high fluoride) Water: 0.15 ± 0.13 mg/L (reference) 5.64 ± 3.52 mg/L (high fluoride)	30, 100.93 (13.10) 118, 97.26 (15.39)		−3.54 (0.50) per 1 mg/L urinary F	Table 1, Table 3
Wang et al. (2021)[89me, w] *Cross-sectional*	China	9−11	Drinking water Reference/high fluoride areas	1.0 ± 0.07 mg/L (reference) 2.8 ± 0.06 mg/L (high fluoride)	303, 109.0 (14.4) 275, 102.1 (16.3)	303, 109.0 (14.4) 275, 102.1 (16.3)		Section 2.1, Table 2
Zhao et al. (2021)[91rs] *Cross-sectional*	China	6−11	Urine Nonendemic/endemic fluorosis areas	1.03 (0.72, 1.47) mg/L			−5.957 (−9.712, −2.202) per 1 log urinary F	Section 3.1, Table 3

Notes:

SD = standard deviation; SE = standard error; MDI = Mental Development Index; GCI = General Cognitive Index

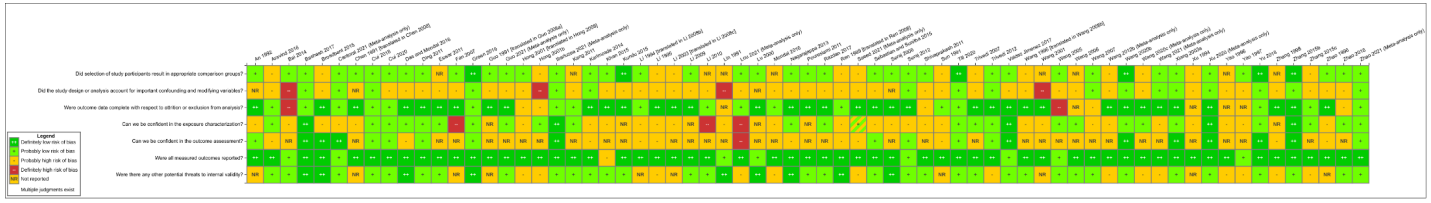
[a]An "me" superscript indicates that the studies included in the *mean-effects meta-analysis*; an "o" superscript indicates a study included in "other" exposures mean-effects analysis (see Table 2 footnote in the main publication); a "w" superscript indicates studies included in the *mean-effects dose-response meta-analysis* using fluoride in water; a "u" superscript indicates studies included in the *mean-effects dose-response meta-analysis* using fluoride in urine; "*" indicates studies included in the *mean-effects dose-response meta-analysis* at levels < 1.5 mg/L; an "rs" superscript indicates studies included in the *regression slopes meta-analysis*.

[b]Additional exposure regions including iodine levels were not included in the analysis.

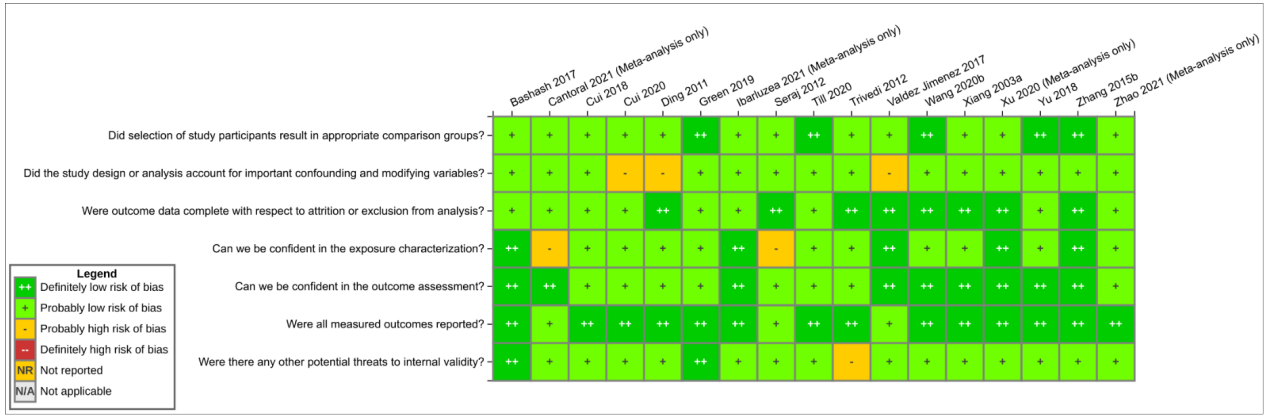
[c]Additional exposure regions including arsenic levels were not included in the analysis.

[d]Median (q1−q3).

(a) All Studies



(b) Low Risk-of-bias Studies



eFigure 2. Results from Risk-of-bias Evaluations for Studies Included in the Meta-analyses and Sensitivity Analyses[a]

Panel (a) presents risk-of-bias results for all studies. An interactive version of eFigure 2(a) is available here: https://hawcproject.org/summary/visual/assessment/405/eFigure-2-Meta-analysis-RoB/. Panel (b) presents risk-of-bias results for low risk-of-bias studies only. An interactive version of eFigure 2(b) is available here: https://hawcproject.org/summary/visual/assessment/405/eFigure-2b-Meta-analysis-RoB-low-RoB-studies/.

The following studies are included in the *mean-effects meta-analysis* and *mean-effects dose-response meta-analysis*: Bashash et al. (2017),[112] Cui et al. (2020),[114] Ding et al. (2011),[107] Green et al. (2019),[113] Seraj et al. (2012),[30] Trivedi et al. (2012),[40] Xiang et al. (2003a),[59] Xu et al. (2020),[115] Yu et al. (2018),[3] and Zhang et al.(2015b).[110]

The following studies are included in the *regression slopes meta-analysis*: Bashash et al. (2017),[112] Cui et al. (2018),[76] Ding et al. (2011),[107] Green et al. (2019),[113] Till et al. (2020),[81] Xu et al. (2020),[115] Yu et al. (2018),[3] Zhang et al.(2015b),[110] and Zhao et al. (2021).[91]

Four studies are only included in sensitivity analyses. All four of these studies are included in sensitivity analyses for the *regression slopes meta-analysis* and include Cantoral et al. (2021),[83] Ibarluzea et al. (2021),[87] Valdez Jiménez et al. (2017),[74] and Wang et al. (2020b).[4] Ibarluzea et al. (2021)[87] is also included in sensitivity analyses for the *mean-effects meta-analysis* and *mean-effects dose-response meta-analysis*.

Mean-effects Meta-analysis

in fluoridated *vs*. non-fluoridated areas in Canada,[113] or in New Zealand.[25] No other studies included in the main *mean-effects meta-analysis* made comparisons between fluoridated *vs*. non-fluoridated areas. In both studies, levels of fluoride in water were low, even in communities with fluoridated drinking water, likely limiting the power to detect an effect.

In Bashash et al.,[112] the SMD compares mean IQ scores in children with urinary fluoride levels below *vs*. above 0.80 mg/L in Mexico.[112] Unlike other studies in the *mean-effects meta-analysis* which compared mean IQ scores between fluoridated *vs*. non-fluoridated areas, or areas with high *vs.* low fluoride exposures (see **eTable 2**), the Bashash et al.[112] study was not designed to measure fluoride exposure by geographical area. However, since the mean IQ scores were provided in the manuscript for children with urinary fluoride levels below *vs*. above 0.80 mg/L, we included them in this analysis. It's worth noting that there was no significant difference when comparing MUF levels between the groups of children with urinary fluoride levels above or below 0.80 mg/L, however when children's IQs were regressed against MUF, a statistically significant inverse association was found.

Commented [I5]: See Doc06a_Meta-analysis, 6a.M., page 8 and 9

Commented [I6]: See Doc06b_Meta-analysis, 6b.C., page 2 and 3

Meta-regression results

The results of the meta-regression models indicate that year of publication and mean age of study children did not explain a large degree of heterogeneity as neither were significant predictors of the relationship between fluoride and children's intelligence, and the residual I^2 remained high (85% and 87%, respectively). Year of publication (SMD = 0.01, 95% CI: -0.01, 0.02) and mean age (SMD = -0.04, 95% CI: -0.13, 0.04) explained relatively little between-study variance (adjusted R^2 of 12% and 5%, respectively). When both year of publication and mean age were included in the model, there were no notable improvements to the amount of between-study variance explained (adjusted R^2 = 13%) or percent residual variation due to heterogeneity (residual I^2 = 85%).

Excluding the outlier study[34] resulted in a slightly lower heterogeneity for the overall effect estimate (I^2=84%) and for the India-specific effect estimate (I^2=69%). The meta-regression indicates that mean age is a significant predictor of the effect (SMD = -0.06, 95% CI: −0.12, −0.01, p-value =0.025), explaining 9% of the between-study variance. Year of publication (SMD = 0.01, 95% CI: 0.001, 0.02, p-value=0.028) explained a larger degree of between-study variance (R^2 = 19 %).

Commented [I7]: See Doc06b_Meta-analysis, 6b.G., page 7 and 8

Commented [I8]: See Doc08_Meta-analysis, 8.E., page 4 and 5

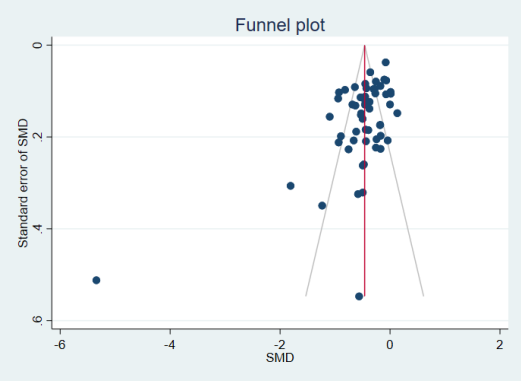
Mean-effects meta-analysis sensitivity analyses

eTable 3. Sensitivity Analyses for Mean-effects Meta-analysis: Pooled SMDs and 95% CIs for Children's IQ Score and Exposures to Fluoride

Commented [l9]: See Doc05_Meta-analysis, 5.D., page 4 and 5

Analysis	Number of Studies	SMD (95% CI)	Heterogeneity	
			p-value	I^2
Excluding Khan et al. (2015)[34]				
	54	−0.43 (−0.51, −0.34)	<0.001	84%
Excluding Lin et al. (1991)[95]				
	54	−0.47 (−0.56, −0.37)	<0.001	87%
Excluding Li et al. (1994)[12] [translated in Li et al. 2008b]				
	54	−0.46 (−0.55, −0.36)	<0.001	87%
Excluding Trivedi et al. (2012)[40]				
	54	−0.46 (−0.56, −0.37)	<0.001	87%
Low risk of bias studies, excluding Trivedi et al. (2012)[40]				
	9	−0.22 (−0.40, −0.04)	<0.001	85%
Including Ibarluzea et al. (2021),[87] Bayley MDI score				
	56	−0.45 (−0.54, −0.36)	<0.001	88%
Including Ibarluzea et al. (2021),[87] McCarthy GCI score				
	56	−0.45 (−0.54, −0.36)	<0.001	87%
Including Aravind et al. (2016),[68] Kundu et al. (2015),[67] Razdan et al. (2017)[73]				
	58	−0.52 (−0.62, −0.42)	<0.001	93%
Including Aravind et al. (2016),[68] Kundu et al. (2015),[67] Razdan et al. (2017)[73], Ibarluzea et al. (2021),[87] Bayley MDI score				
	59	−0.51 (−0.61, −0.40)	<0.001	91%
Including Aravind et al. (2016),[68] Kundu et al. (2015),[67] Razdan et al. (2017)[73], Ibarluzea et al. (2021),[87] McCarthy GCI score				
	59	−0.51 (−0.61, −0.40)	<0.001	91%
Any exposure group				
	55	−0.44 (−0.54, −0.34)	<0.001	91%

Notes:
CI = confidence interval; SMD = standardized weighted mean difference; MDI = Mental Development Index; GCI = General Cognitive Index.



eFigure 3. Funnel Plot of Included Studies

This funnel plot shows individual studies included in the analysis according to random-effect standardized weighted mean difference (SMD) estimates (x-axis) and the standard error (SE) of each study-specific SMD (y-axis). The solid vertical line indicates the pooled SMD estimate for all studies combined and the dashed lines indicated pseudo 95% confidence limits around the pooled SMD estimate.

```
Regression-based Egger test for small-study effects
Random-effects model
Method: DerSimonian-Laird

H0: beta1 = 0; no small-study effects
           beta1 =     -3.20
      SE of beta1 =     0.576
                z =     -5.55
       Prob > |z| =    0.0000

Begg's test for small-study effects

Kendall's score =   -299.00
    SE of score =   137.750
                z =     -2.18
     Prob > |z| =    0.0305
```

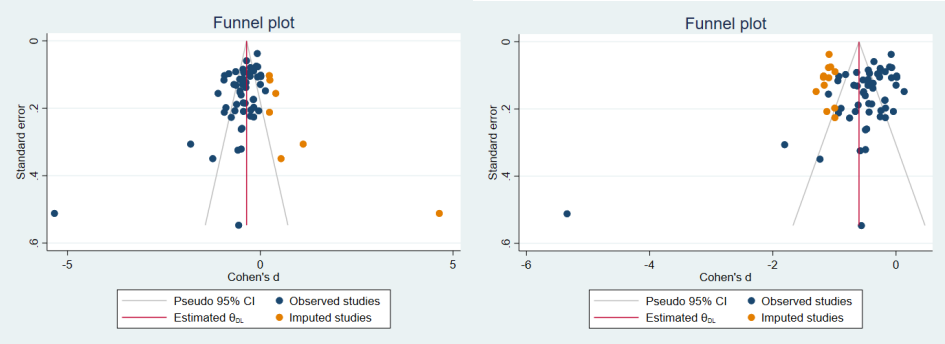
eFigure 4. Test for Publication Bias

```
Nonparametric trim-and-fill analysis of publication bias
Run estimator, imputing on the right

Iteration                           Number of studies =     62
  Model: Random-effects                    observed =     55
 Method: DerSimonian-Laird                 imputed =      7

Pooling
  Model: Random-effects
 Method: DerSimonian-Laird

------------------------------------------------------------
             Studies |        Cohen's d    [95% conf. interval]
---------------------+--------------------------------------
            Observed |           -0.461      -0.554      -0.368
  Observed + Imputed |           -0.357      -0.459      -0.255
------------------------------------------------------------
```

```
Nonparametric trim-and-fill analysis of publication bias
Linear estimator, imputing on the left

Iteration                           Number of studies =     67
  Model: Random-effects                    observed =     55
 Method: DerSimonian-Laird                 imputed =      12

Pooling
  Model: Random-effects
 Method: DerSimonian-Laird

------------------------------------------------------------
             Studies |        Cohen's d    [95% conf. interval]
---------------------+--------------------------------------
            Observed |           -0.461      -0.554      -0.368
  Observed + Imputed |           -0.601      -0.713      -0.489
------------------------------------------------------------
```

eFigure 5. Trim-and-fill Analysis

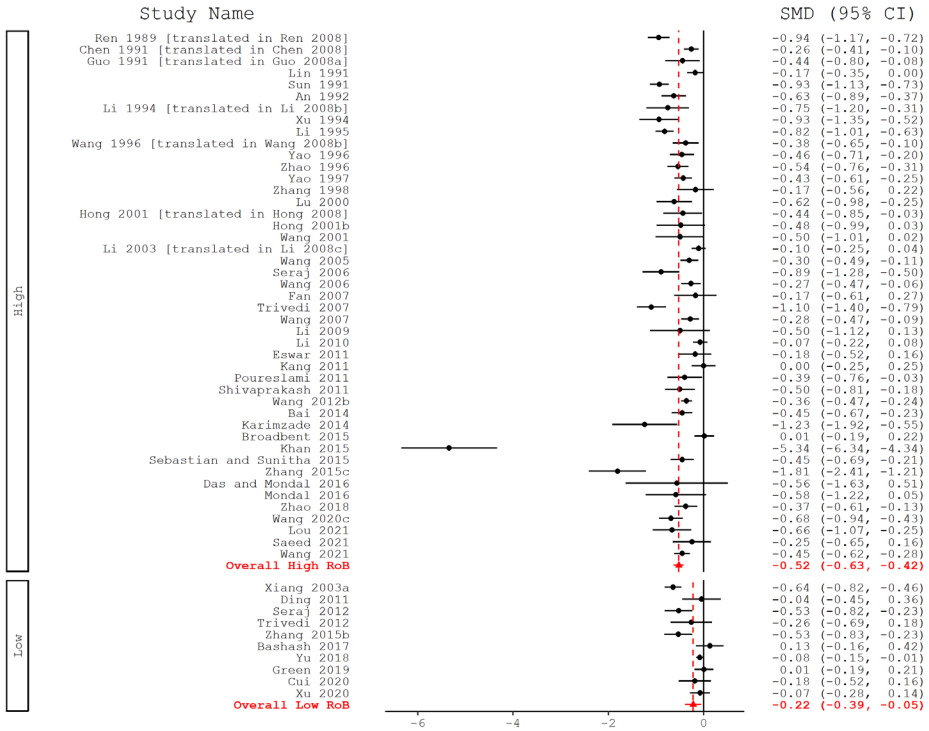
Left panel shows the random-effects pooled SMD after filling in to the right using a run estimator (the linear estimator to the right showed no change in pooled SMD); right panel shows random-effects pooled SMD after filling in to the left using a linear estimator (the run estimator to the left showed no change in pooled SMD).



eFigure 6. Filled-in Funnel Plots to Eliminate Publication Bias

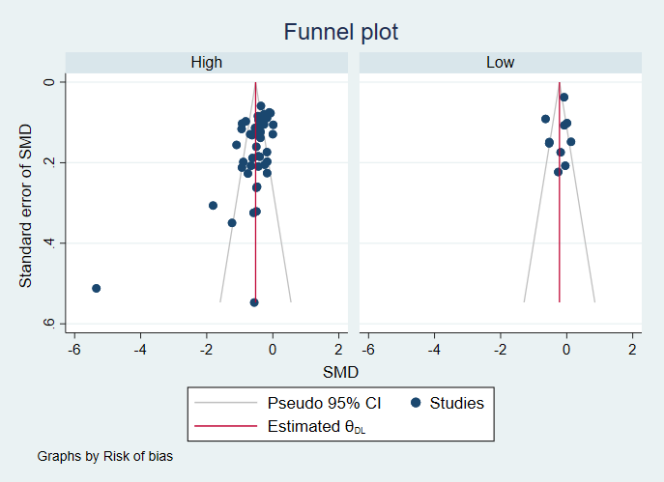
Left panel shows the funnel plot filled in to the right using a run estimator (the linear estimator to the right showed no change in pooled SMD); right panel shows the funnel plot filled in to the left using a linear estimator (the run estimator to the left showed no change in pooled SMD).

Risk-of-bias Subgroup Analysis



eFigure 7. Association Between Fluoride Exposure and IQ Scores in Children: Effect by Risk of Bias

Commented [EAM10]: See Doc06b_Meta-analysis, 6b.W., page 19 through 21.



eFigure 8. Funnel Plot by Risk-of-bias Evaluation

Commented [EAM11]: See Doc06b_Meta-analysis, 6b.W., page 19 through 21.

```
Regression-based Egger test for small-study effects
Random-effects model
Method: DerSimonian-Laird                    High RoB

H0: beta1 = 0; no small-study effects
            beta1 =     -3.41
      SE of beta1 =     0.618
                z =     -5.52
       Prob > |z| =    0.0000

. *meta bias if rob==1, begg rob==1
. meta bias if rob==2, egger random(dl) nometashow

Regression-based Egger test for small-study effects
Random-effects model
Method: DerSimonian-Laird                    Low RoB

H0: beta1 = 0; no small-study effects
            beta1 =     -0.17
      SE of beta1 =     1.835
                z =     -0.09
       Prob > |z| =    0.9275
```

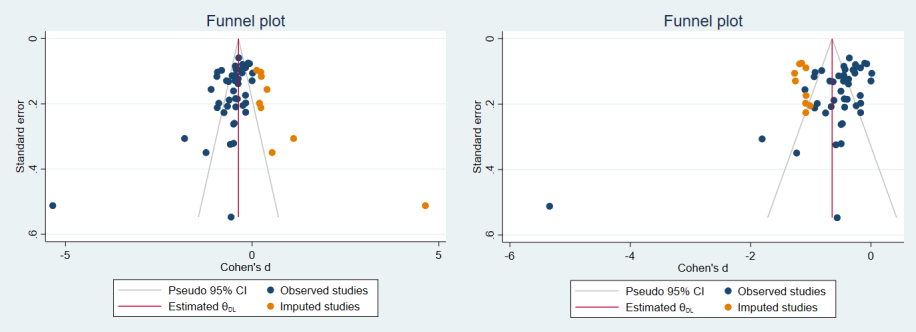
eFigure 9. Test for Publication Bias by Risk of Bias

```
Nonparametric trim-and-fill analysis of publication bias
Run estimator, imputing on the right

Iteration                           Number of studies =     54
  Model: Random-effects                     observed =     45
 Method: DerSimonian-Laird                  imputed =      9

Pooling
  Model: Random-effects
 Method: DerSimonian-Laird

------------------------------------------------------------
             Studies |        Cohen's d    [95% conf. interval]
---------------------+--------------------------------------
            Observed |           -0.521      -0.625      -0.416
  Observed + Imputed |           -0.365      -0.484      -0.246
------------------------------------------------------------
```

```
Nonparametric trim-and-fill analysis of publication bias
Linear estimator, imputing on the left

Iteration                           Number of studies =     54
  Model: Random-effects                     observed =     45
 Method: DerSimonian-Laird                  imputed =      9

Pooling
  Model: Random-effects
 Method: DerSimonian-Laird

------------------------------------------------------------
             Studies |        Cohen's d    [95% conf. interval]
---------------------+--------------------------------------
            Observed |           -0.521      -0.625      -0.416
  Observed + Imputed |           -0.646      -0.765      -0.526
------------------------------------------------------------
```

eFigure 10. Trim-and-fill Analysis for High Risk-of-bias Studies

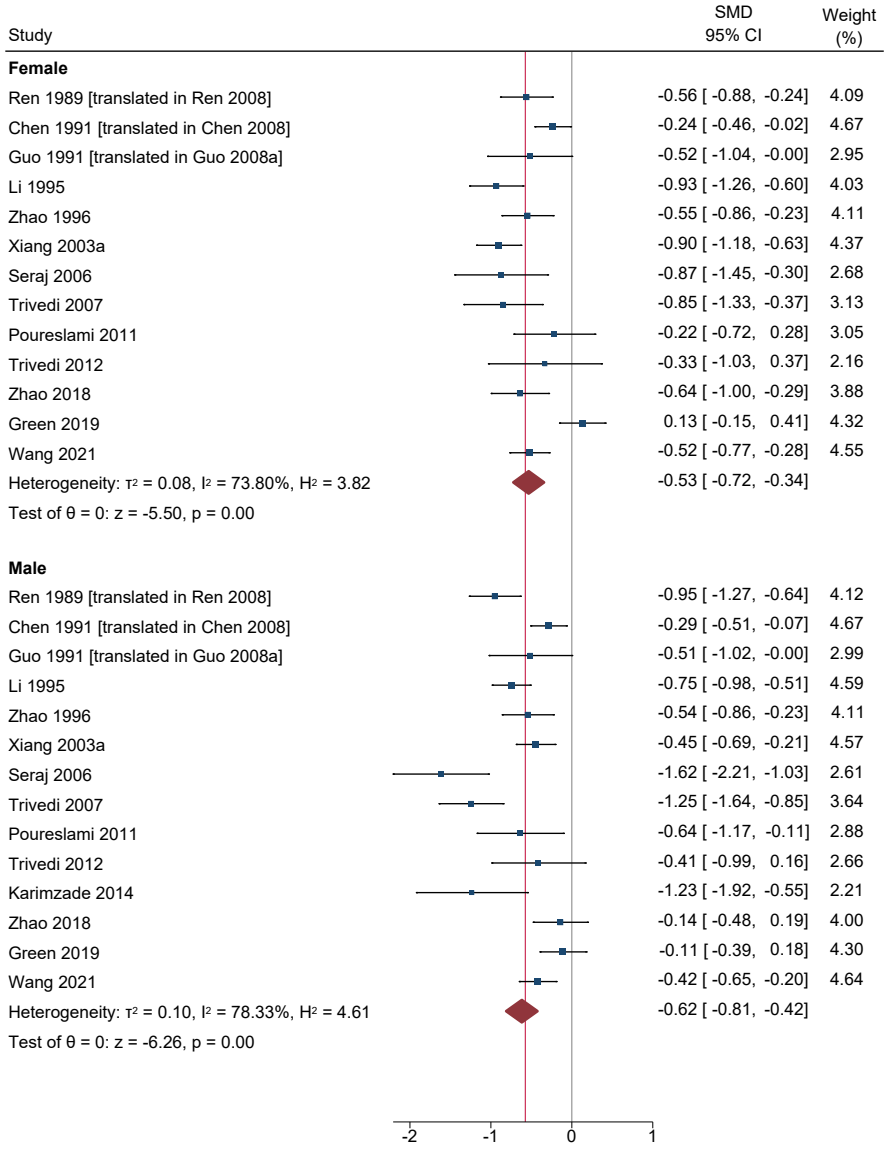
Filling in to the right using a linear estimator or to the left using a run estimator showed no change in the pooled SMD.



eFigure 11. Filled-in Funnel Plots for High Risk-of-bias Studies

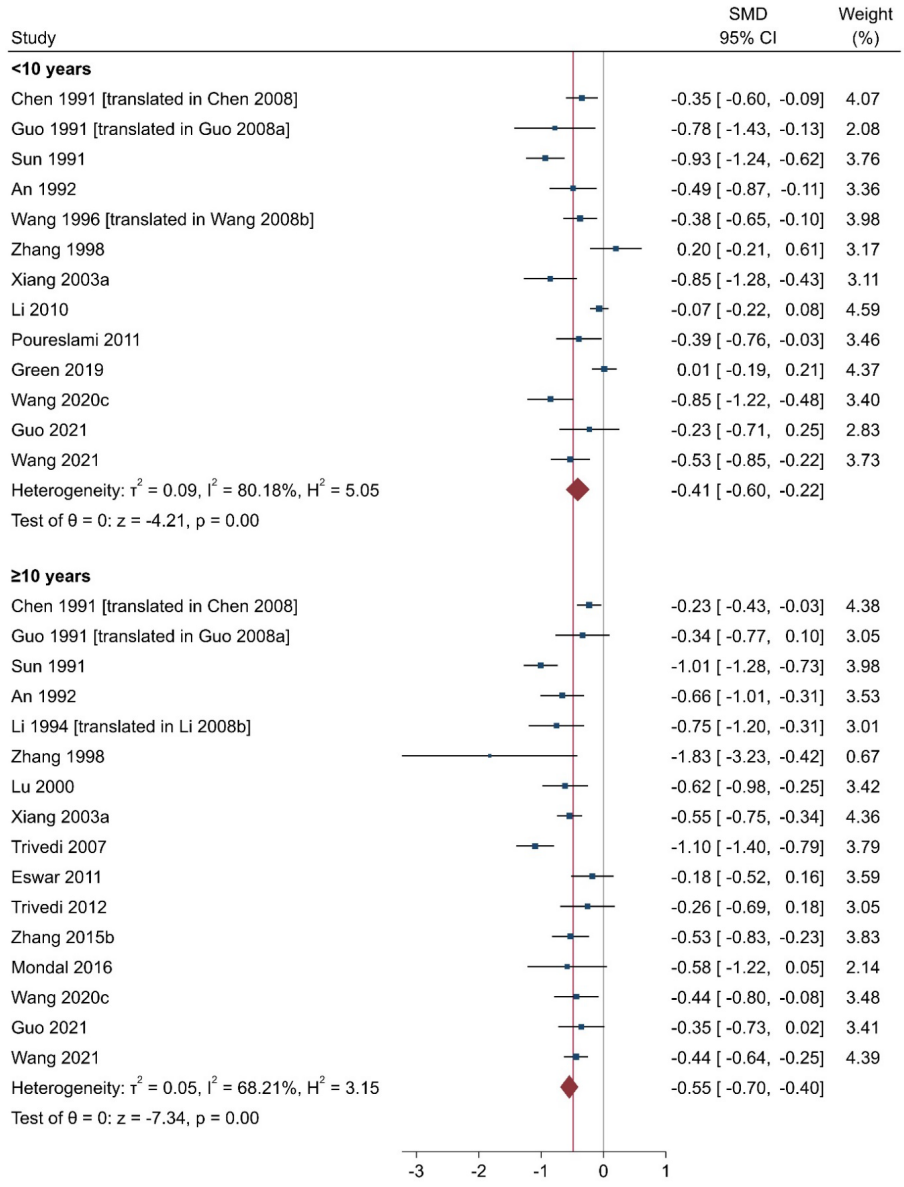
Left panel shows the random-effects pooled SMD after filling in to the right using a run estimator (the linear estimator to the right showed no change in the pooled SMD); right panel shows random-effects pooled SMD after filling in to the left using a linear estimator (the run estimator to the left showed no change in the pooled SMD).

Sex Subgroup Analysis



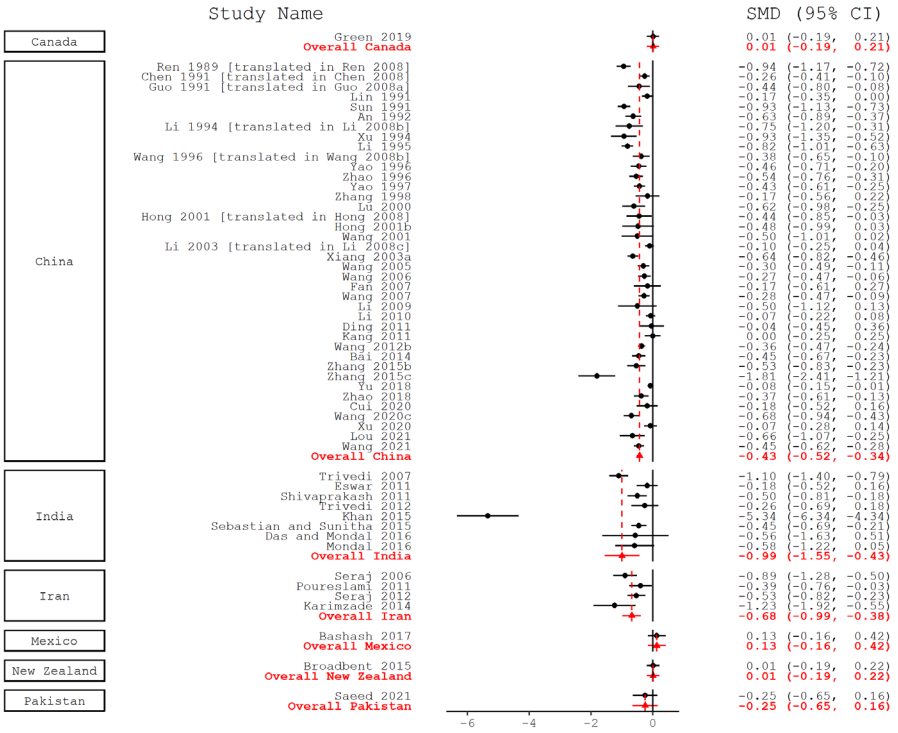
eFigure 12. Association Between Fluoride Exposure and IQ Scores in Children: Effect by Sex

Age Group Subgroup Analysis



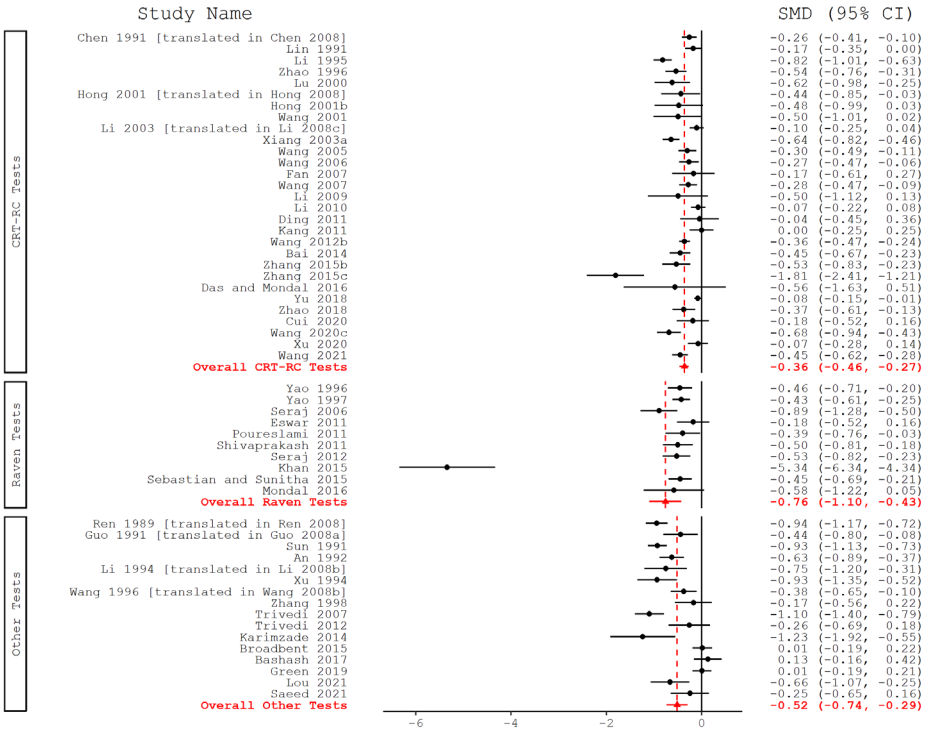
eFigure 13. Association Between Fluoride Exposure and IQ Scores in Children: Effect by Age Group

Country Subgroup Analysis



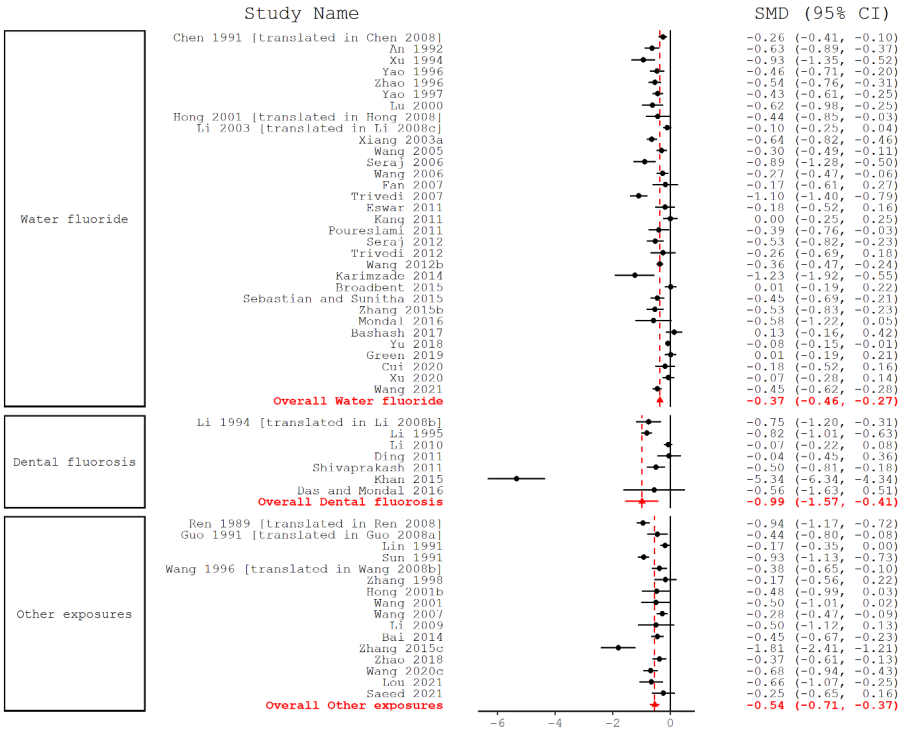
eFigure 14. Association Between Fluoride Exposure and IQ Scores in Children: Effect by Country

Assessment Type Subgroup Analysis



eFigure 15. Association Between Fluoride Exposure and IQ Scores in Children: Effect by Assessment Type

Exposure Type Subgroup Analysis



eFigure 16. Association Between Fluoride Exposure and IQ Scores in Children: Effect by Exposure Type

Exposure types include water, dental fluorosis, and other exposures (iodine, arsenic, aluminum, and fluoride from coal burning).

Dose-Response Meta-analysis Using Mean Effect Estimates

When analyses were restricted to exposed groups with <4 mg/L (i.e., 0 to <4 mg/L) fluoride in drinking water (n = 21 publications [6 low and 15 high risk-of-bias studies]), there was a statistically significant inverse association between fluoride exposure and children's IQ (SMD: −0.22; 95% CI: −0.27, −0.17; p-value < 0.001) (**eTable 4**). When restricted to <2 mg/L (i.e., 0 to <2 mg/L) in drinking water (n = 7 publications [3 low and 4 high risk-of-bias studies]), the magnitude of the effect estimate did not substantially change (SMD: −0.15; 95% CI: −0.41, 0.12; p-value = 0.274). However, when restricted to exposed groups with <1.5 mg/L (i.e., 0 to <1.5 mg/L) in drinking water (n = 7 publications [3 low and 4 high risk-of-bias studies]), there was no longer an association between fluoride in drinking water and children's IQ (SMD: 0.05; 95% CI: −0.36, 0.45; p-value = 0.816). When analyses were further restricted to low risk-of-bias publications at <4 mg/L, <2 mg/L, and <1.5 mg/L, the associations remained in the same direction and were larger in magnitude compared to when data from both low and high risk-of-bias studies were combined (**eTable 4** and **eTable 5**).

When analyses were restricted to exposed groups with <4 mg/L urinary fluoride (n = 13 publications [9 low and 4 high risk-of-bias studies]), there was a statistically significant inverse association between children's urinary fluoride exposure and IQ (SMD: −0.17; 95% CI: −0.30, −0.05; p-value = 0.005) (**eTable 4**). When restricted to <2 mg/L urinary fluoride (n = 7 publications [5 low and 2 high risk-of-bias studies]), there was an inverse association (SMD: −0.06; 95% CI: −0.14, 0.01; p-value = 0.094). When restricted to exposed groups with <1.5 mg/L urinary fluoride (n = 5 publications [4 low and 1 high risk-of-bias studies]), there was an inverse association (SMD: −0.09; 95% CI: −0.16, −0.01; p-value = 0.026). When analyses were further restricted to low risk-of-bias publications, the associations at <2 mg/L and <1.5 mg/L became smaller in magnitude and were statistically significant at <1.5 mg/L (p-value = 0.472 and p-value = 0.028, respectively) (**eTable 4**). Similar results were observed when the maximum likelihood estimation method was used (**eTable 5**).

Commented [I12]: See Doc06a_Meta-analysis, 6a.L., page 7 and 8

Commented [I13]: See Doc06a_Meta-analysis, 6a.L., page 7 and 8

Commented [I14]: See Doc06a_Meta-analysis, 6a.L., page 7 and 8

Commented [I15]: See Doc01_Meta-analysis, 1.P. (page 8) and Doc06b_Meta-analysis, 6b.EE. (page 24 and 25)

eTable 4. Dose-Response Meta-analysis Using Mean Effects—Model Selection[a]

Commented [I16]: See Doc01_Meta-analysis, 1.K., Page 4 and 5

Exposure Analysis	Parameters	Fluoride Exposure			
		All data	<4 mg/L	<2 mg/L	<1.5 mg/L
Water Fluoride – All Studies					
No. Studies/No. Observations Number of Children		29/39 11,656	21/27 8,723	7/9 2,971	7/7 2,832
Linear Model[b]	Beta (95% CI) p-value AIC	−0.15 (−0.20, −0.11) p < 0.001 AIC = 53.8	−0.22 (−0.27, −0.17) p < 0.001 AIC = 16.1	−0.15 (−0.41, 0.12) p = 0.274 AIC = 11.8	0.05 (−0.36, 0.45) p = 0.816 AIC = 8.2
Quadratic Model[c]	Beta (95% CI); p-value Beta (95% CI); p-value AIC p-value*	−0.27 (−0.34, −0.21); p < 0.001 0.02 (0.01, 0.03); p < 0.001 AIC = 48.8 p* < 0.001	−0.12 (−0.35, 0.11); p = 0.318 −0.04 (−0.10, 0.03); p = 0.280 AIC = 21.2 p* = 0.012	0.79 (−0.01, 1.58); p = 0.052 −0.56 (−0.97, −0.16); p = 0.006 AIC = 12.5 p* = 0.007	0.30 (−0.53, 1.14); p = 0.477 −0.23 (−1.01, 0.55); p = 0.561 AIC = 11.3 p* = 0.04
Restricted Cubic Splines Model[d]	Beta (95% CI); p-value Beta (95% CI); p-value AIC p-value*	−0.29 (−0.39, −0.20); p < 0.001 0.48 (0.18, 0.78); p = 0.002 AIC = 42.3 p* < 0.001	−0.14 (−0.34, 0.06), p = 0.162 −0.23 (−0.66, 0.20), p = 0.295 AIC = 16.9 p* = 0.009	1.15 (0.07, 2.22) p = 0.037 −1.20 (−2.03, −0.36) p = 0.005 AIC = 10.5 p* = 0.010	0.49 (−0.50, 1.47) p = 0.334 −0.69 (−2.40, 1.02) p = 0.428 AIC = 10.2 p* = 0.05
Water Fluoride – Low Risk-of-bias Studies					
No. Studies/No. Observations Number of Children		6/11 4,355	6/9 4,251	3/4 921	3/3 879
Linear model	Beta (95% CI) p-value AIC	−0.19 (−0.34, −0.05) p = 0.009 AIC = 10.3	−0.22 (−0.36, −0.07) p = 0.003 AIC = 3.9	−0.34 (−0.72, 0.03) p = 0.070 AIC = 4.5	−0.32 (−0.91, 0.26) p = 0.276 AIC = 4.1

Exposure Analysis	Parameters	Fluoride Exposure			
		All data	<4 mg/L	<2 mg/L	<1.5 mg/L
Urinary Fluoride – All Studies					
No. Studies/No. Observations Number of Children		18/32 8,502	13/26 6,885	7/11 4,654	5/8 3,992
Linear Model[b]	Beta (95% CI) p-value AIC	−0.16 (−0.24, −0.08) p < 0.001 AIC = 73.8	−0.17 (−0.30, −0.05) p = 0.005 AIC = 68.0	−0.06 (−0.14, 0.01) p = 0.094 AIC = 1.2	−0.09 (−0.16, −0.01) p = 0.026 AIC= 2.8
Quadratic Model[c]	Beta (95% CI); p-value Beta (95% CI); p-value AIC p-value*	−0.10 (−0.31, 0.11); p = 0.360 −0.01 (−0.05, 0.02); p = 0.496 AIC = 84.3 p* = 0.14	0.07 (−0.23, 0.38); p = 0.645 −0.07 (−0.16, 0.01); p = 0.071 AIC = 75.8 p* = 0.08	−0.22 (−0.65, 0.20); p = 0.303 0.08 (−0.13, 0.30); p = 0.456 AIC = 9.2 p* = 0.42	0.65 (−1.46, 2.76); p = 0.548 −0.66 (−2.11, 0.80); p = 0.379 AIC = 8.3 p* = 0.10
Restricted Cubic Splines Model[d]	Beta (95% CI); p-value Beta (95% CI); p-value AIC p-value*	−0.12 (−0.28, 0.04); p = 0.150 −0.10 (−0.43, 0.23); p = 0.545 AIC = 79.6 p* = 0.13	−0.03 (−0.22, 0.16); p = 0.741 −0.24 (−0.47, −0.002); p = 0.048 AIC = 73.3 p* = 0.07	−0.14 (−0.32, 0.04); p = 0.130 0.13 (−0.17, 0.43); p = 0.395 AIC = 8.5 p* = 0.37	−0.52 (−1.65, 0.62); p = 0.371 0.63 (−1.32, 2.59); p = 0.524 AIC = 6.7 p* = 0.07
Urinary Fluoride – Sensitivity analysis including Ibarluzea et al. **(2021)[87] Bayley MDI scores**					
No. Studies/No. Observations Number of Children		19/33 8,815	14/27 7,445	8/12 4,967	6/9 4,305
Linear model	Beta (95% CI) p-value AIC	−0.15 (−0.23, −0.07) p < 0.001 AIC = 75.0	−0.15 (−0.28, −0.03) p = 0.015 AIC = 69.0	−0.04 (−0.14, 0.05) p = 0.371 AIC = 1.7	−0.08 (−0.15, −0.003) p = 0.043 AIC = 3.6
Urinary Fluoride – Sensitivity analysis including Ibarluzea et al. **(2021)[87] McCarthy GCI scores**					
No. Studies/No. Observations Number of Children		19/33 8,749	14/27 7,445	8/12 4,901	6/9 4,239

Exposure Analysis	Parameters	Fluoride Exposure			
		All data	<4 mg/L	<2 mg/L	<1.5 mg/L
Linear model	Beta (95% CI) p-value AIC	−0.15 (−0.23, −0.07) p < 0.001 AIC = 74.5	−0.16 (−0.28, −0.04) p = 0.011 AIC = 68.6	−0.05 (−0.14, 0.04) p = 0.259 AIC = 1.3	−0.08 (−0.16, −0.01) p = 0.036 AIC = 3.0
Urinary Fluoride – Low Risk-of-bias Studies					
No. Studies/No. Observations Number of Children		9/15 5,713	9/15 5,713	5/8 4,141	4/7 3,952
Linear model	Beta (95% CI) p-value AIC	−0.10 (−0.21, 0.01) p = 0.082 AIC = 5.9	−0.10 (−0.21, −0.01) p = 0.082 AIC = 5.9	−0.05 (−0.17, 0.08) p = 0.472 AIC = 2.8	−0.08 (−0.16, −0.01) p = 0.028 AIC = 2.5

Notes:

AIC = Akaike information criterion; SMD = standardized mean difference; p = p-value for effect estimate; p* = p-value for likelihood ratio tests; MDI = Mental Development Index; GCI = General Cognitive Index

[a]Parameter estimates are changes in SMDs (beta [95% CI]) based on the restricted maximum likelihood models; model fit is represented by the maximum likelihood AIC.

[b]The estimates represent change in SMD for the linear model and AIC, respectively.

[c]The estimates represent change in SMD for the linear term, change in SMD for quadratic term, AIC, and p-values for likelihood ratio test versus linear model, respectively. Potential departure from a linear trend was assessed by testing the coefficient of the quadratic term equal to zero.

[d]The estimates represent change in SMD for the first spline term, change in SMD for the second spline term, AIC, and p-value for likelihood ratio test vs linear model, respectively. Potential departure from a linear trend was assessed by testing the coefficient of the second spline equal to zero.

eTable 5. Dose-response Meta-analysis Using Mean Effects: Maximum Likelihood Models[a]

Commented [I17]: See Doc01_Meta-analysis, 1.K., page 4 and 5

Exposure Analysis	Parameters	Fluoride Exposure			
		All data	<4 mg/L	<2 mg/L	<1.5 mg/L
Water Fluoride – All Studies					
No. Studies/No. Observations Number of Children		29/39 11,656	21/27 8,723	7/9 2,971	7/7 2,832
Linear Model[b]	Beta (95% CI) p-value AIC	−0.15 (−0.20, −0.11) p < 0.001 AIC = 47.9	−0.22 (−0.27, −0.17) p < 0.001 AIC = 10.5	−0.15 (−0.39, 0.08) p = 0.202 AIC = 9.6	0.02 (−0.33, 0.36) p = 0.928 AIC = 6.7
Quadratic Model[c]	Beta (95% CI); p-value Beta (95% CI); p-value AIC p-value*	−0.26 (−0.32, −0.20); p < 0.001 0.02 (0.01, 0.03); p < 0.001 AIC = 33.0 p* < 0.001	−0.11 (−0.33, 0.11); p = 0.332 −0.04 (−0.10, 0.02); p = 0.229 AIC= 10.2 p* = 0.012	0.64 (0.04, 1.24); p = 0.036 −0.49 (−0.81, −0.16); p = 0.003 AIC = 8.2 p* = 0.007	0.34 (−0.37, 1.04); p = 0.349 −0.26 (−0.88, 0.35); p = 0.405 AIC = 8.5 p* = 0.04
Restricted Cubic Splines Model[d]	Beta (95% CI); p-value Beta (95% CI); p-value AIC p-value*	−0.29 (−0.38, −0.21); p < 0.001 0.48 (0.20, 0.78); p = 0.001 AIC = 33.9 p* < 0.001	−0.13 (−0.32, 0.05); p = 0.162 −0.24 (−0.65, 0.16); p = 0.233 AIC= 9.7 p* = 0.009	0.27 (−0.09, 0.62); p = 0.140 −0.44 (−0.83, −0.04); p = 0.029 AIC = 8.9 p* = 0.010	0.26 (−0.26, 0.79); p = 0.321 −0.49 (−1.54, 0.56); p = 0.363 AIC = 8.7 p*= 0.05
Water Fluoride – Low Risk-of-bias Studies					
No. Studies/No. Observations Number of Children		6/11 4,355	6/9 4,251	3/4 921	3/3 879
Linear model	Beta (95% CI) p-value AIC	−0.19 (−0.31, −0.06) p = 0.003 AIC = 6.7	−0.21 (−0.33, −0.09) p = 0.001 AIC = 0.3	−0.35 (−0.63, −0.07) p = 0.015 AIC = 2.7	−0.34 (−0.80, 0.12) p = 0.153 AIC = 3.3

Exposure Analysis	Parameters	Fluoride Exposure			
		All data	<4 mg/L	<2 mg/L	<1.5 mg/L
Urinary Fluoride – All Studies					
No. Studies/No. Observations Number of Children		18/32 8,502	13/26 6,885	7/11 4,654	5/8 3,992
Linear Model[b]	Beta (95% CI) p-value AIC	−0.16 (−0.23, −0.08) p < 0.001 AIC = 69.2	−0.17 (−0.29, −0.06) p = 0.004 AIC = 64.2	−0.07 (−0.13, 0.003) p = 0.060 AIC = −3.7	−0.12 (−0.36, 0.12) p = 0.325 AIC = 0.8
Quadratic Model[c]	Beta (95% CI); p-value Beta (95% CI); p-value AIC p-value*	−0.19 (−0.44, 0.06); p = 0.131 0.01 (−0.02, 0.05); p = 0.462 AIC = 73.0 p*= 0.14	0.08 (−0.21, 0.37); p = 0.587 −0.08 (−0.16, 0.0004); p = 0.051 AIC = 67.2 p* = 0.08	−0.23 (−0.62, 0.17); p = 0.267 0.08 (−0.12, 0.29); p = 0.423 AIC = 1.7 p* = 0.42	−0.11 (−1.45, 1.23); p = 0.868 0.02 (−0.74, 0.77); p = 0.967 AIC = 4.1 p* = 0.10
Restricted Cubic Splines Model[d]	Beta (95% CI); p-value Beta (95% CI); p-value AIC p-value*	−0.12 (−0.28, 0.04); p = 0.138 −0.10 (−0.41, 0.21); p = 0.524 AIC = 72.9 p*= 0.13	−0.03 (−0.21, 0.15); p = 0.775 −0.24 (−0.47, −0.02); p = 0.034 AIC = 66.8 p* = 0.07	−0.13 (−0.29, 0.03); p = 0.107 0.12 (−0.14, 0.38); p = 0.366 AIC = 1.5 p* = 0.37	−0.26 (−0.72, 0.20); p = 0.270 0.36 (−0.58, 1.29); p = 0.453 AIC = 3.5 p* = 0.07
Urinary Fluoride – Sensitivity analysis including Ibarluzea et al. **(2021)[87] Bayley MDI scores**					
No. Studies/No. Observations Number of Children		19/33 8,815	14/27 7,445	8/12 4,967	6/9 4,305
Linear model	Beta (95% CI) p-value AIC	−0.15 (−0.23, −0.07) p < 0.001 AIC = 70.3	−0.16 (−0.28, −0.04) p = 0.010 AIC = 65.2	−0.06 (−0.13, 0.01) p = 0.086 AIC = −3.2	−0.08 (−0.15 −0.003) p = 0.043 AIC = −1.2
Urinary Fluoride – Sensitivity analysis including Ibarluzea et al. **(2021)[87] GCI scores**					
No. Studies/No. Observations Number of Children		19/33 8,749	14/27 7,445	8/12 4,901	6/9 4,239

Exposure Analysis	Parameters	Fluoride Exposure			
		All data	**<4 mg/L**	**<2 mg/L**	**<1.5 mg/L**
Linear model	Beta (95% CI) p-value AIC	−0.15 (−0.23, −0.07) p < 0.001 AIC = 69.8	−0.16 (−0.28, −0.04) p = 0.008 AIC = 64.9	−0.04 (−0.20, 0.13) p = 0.653 AIC = −0.9	−0.08 (−0.16, −0.01) p = 0.036 AIC = −1.7
Urinary Fluoride – Low Risk-of-bias Studies					
No. Studies/No. Observations Number of Children		9/15 5,713	9/15 5,713	5/8 4,141	4/7 3,952
Linear model	Beta (95% CI) p-value AIC	−0.10 (−0.20, 0.004) p = 0.059 AIC = 2.0	−0.10 (−0.20, 0.004) p = 0.059 AIC = 2.0	−0.07 (−0.14, 0.01) p = 0.073 AIC = −1.8	−0.08 (−0.16, −0.01) p = 0.028 AIC = −2.2

Notes:

AIC = Akaike information criterion; SMD = standardized mean difference; p = p-value for effect estimate; p* = p-value for likelihood ratio tests; MDI = Mental Development Index; GCI = General Cognitive Index

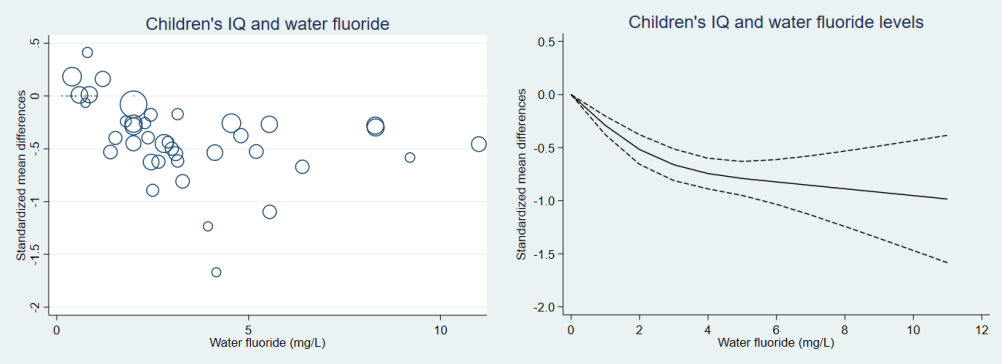
[a]Parameter estimates are changes in SMDs (beta [95% CI]) based on the maximum likelihood models; model fit is represented by the maximum likelihood AIC.

[b]The estimates represent change in SMD for the linear model and AIC, respectively.

[c]The estimates represent change in SMD for the linear term, change in SMD for quadratic term, AIC, and p-values for likelihood ratio test versus linear model, respectively. Potential departure from a linear trend was assessed by testing the coefficient of the quadratic term equal to zero

[d]The estimates represent change in SMD for the first spline term, change in SMD for the second spline term, AIC, and p-value for likelihood ratio test vs linear model, respectively. Potential departure from a linear trend was assessed by testing the coefficient of the second spline equal to zero.

Water Fluoride Exposure



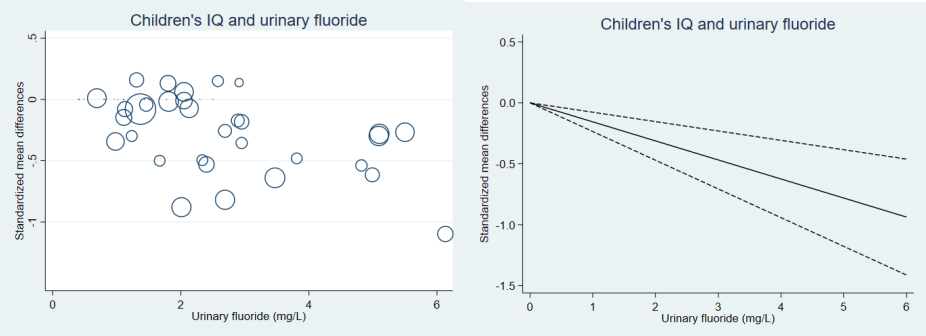
eFigure 17. Pooled Dose-Response Association Between Fluoride in Water and Standardized Mean Differences in Children's IQ

Left panel: circles indicate standardized weighted mean differences (SMDs) in individual studies; size of bubbles is proportional to precision (inverse of variance) of the standardized mean differences. Right panel: Water fluoride levels were modeled with restricted cubic splines terms in a random-effects model (solid line). Dashed lines represent the 95 % confidence intervals for the spline model. Please see **eTable 2** for characteristics of the studies included in the *dose-response meta-analysis* (studies with water fluoride exposure and at least two exposure levels).

Commented [I18]: See Doc08_Meta-analysis., 8.R., page 10 and 11

Commented [I19]: See Doc03_Meta-analysis, 3.F., page 4

Urinary Fluoride Exposure



eFigure 18. Pooled Dose-Response Association Between Fluoride in Urine and Standardized Mean Differences in Children's IQ

Left panel: Circles indicate standardized weighted mean differences in individual studies; size of bubbles is proportional to precision (inverse of variance) of the standardized mean differences. Right panel: Urinary fluoride levels were modeled with a linear random-effects model (solid line). Dashed lines represent the 95 % confidence intervals for the linear model. Please see **eTable 2** for characteristics of the studies included in the *dose-response meta-analysis* (studies with urinary fluoride exposure and at least two exposure levels).

Commented [I20]: See Doc03_Meta-analysis, 3.F., page 4

Regression Slopes Meta-analysis

Studies with overlapping populations

Yu et al.[3] and Wang et al.[4] used the same study cohort of children recruited in 2015 from the rural areas of Tianjin City, China. Since Wang et al.[4] (n = 571) used a subset of the original study sample from Yu et al.[3] (n = 2,886), only results from Yu et al.[3] were included in the meta-analysis. A sensitivity analysis was performed to evaluate the impact of using the effect estimate from Wang et al.[4] rather than the pooled effect estimate from Yu et al.[3]. Green et al.[113] and Till et al.[81] used the same Maternal-Infant Research on Environmental Chemicals (MIREC) cohort that reported drinking tap water in 10 Canadian cities, with the studies overlapping for 398 mother-child pairs. Both studies reported effect estimates for maternal urinary fluoride (MUF) and water fluoride concentrations. In the Green et al.[113] study, 512 mother-child pairs had MUF data compared to 398 pairs in Till et al.[81]. Water fluoride levels were available for 420 pairs in Green et al.[113] compared to 398 pairs in Till et al.[81]. Both studies reported effect estimates adjusted for maternal education, maternal race, child's sex, HOME total score, and secondhand smoke status in the child's home. In addition, Till et al.[81] adjusted for child's age at IQ testing (the age range for all children was 3–4 years old). Because of the larger sample size and because covariate adjustments were similar, results from Green et al.[113] were included in the main analysis. However, because of the more adjusted estimates from Till et al.[81] compared to Green et al.[113], a sensitivity analysis was performed using the water fluoride result for formula-fed children and the MUF result from Till et al.[81]. For fluoride from intake, the estimates from both studies were used since they represent total fluoride intake from Green et al.[113] and infant fluoride intake from formula Till et al.[81].

Three studies were excluded with reported slopes because the exposure was measured at the community level.[25, 30, 35] Only one study[116] included in this meta-analysis was considered high risk of bias. For Bashash et al.[112], Yu et al.[3] and Till et al.[81], units of exposure were transformed from 0.5 mg/L to 1 mg/L. Cui et al.[76], and Zhao et al. (2021)[91] reported associations between IQ and log-transformed exposure, and units of exposure were transformed from 1 log mg/L to 1 mg/L[117]. Yu et al.[3] reported estimates from piecewise linear regression models and provided three ranges for urinary fluoride exposure (low 0.01–1.60 mg/L, medium 1.60–2.50 mg/L, high 2.50–5.54 mg/L) and two ranges for water fluoride (low 0.20–3.40 mg/L and high 3.40–3.90 mg/L). Since these piecewise effect estimates are likely correlated, the study-specific pooled effect estimates were used for urine and water fluoride exposures for the overall effect meta-analysis. A sensitivity analysis was performed to evaluate the impact of using pooled estimates rather than piecewise estimates from Yu et al.[3].

For studies reporting multiple measures of fluoride exposure, the results associated with measured or estimated individual-level exposures, biomarker levels (such as urinary fluoride), or fluoride intake levels were prioritized over water fluoride concentrations (see protocol; https://ntp.niehs.nih.gov/go/785076); however, subgroup analyses by exposure metric (urinary fluoride, fluoride intake, and water fluoride) were also performed.

Regression slopes meta-analysis sensitivity analyses

Information about demographic variables was not always accessible, making it difficult to study the impact of potential confounders on effect estimates. Sensitivity analyses for the regression slopes explored the impact of using unadjusted estimates, and results were not significantly impacted (**eTable 6**). Also, most of the estimates used in the *mean-effects meta-analyses* come from studies that used fluoride concentrations at the community level to represent exposure. Therefore, unless community-level clustering is accounted for in the analysis, the standard errors of the difference in means between exposed and reference groups are likely to be biased. This is less of an issue in studies using individual-level exposures (e.g., the *regression slopes meta-analysis*). However, most studies lacked adjustment for clustering,[3, 76, 110] or for complex sampling strategies.[3, 110] Therefore, we performed sensitivity analyses to

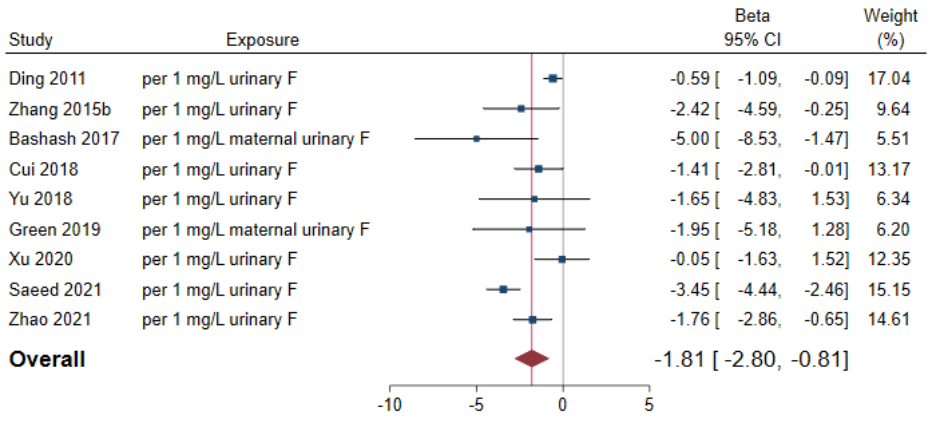
assess the impact of such issues and there were minimal changes in the pooled slopes (**eTable 6**). In the *regression slopes meta-analysis*, from the Green et al.[113] and Bashash et al.[112] studies, we used the estimates reported from the models using the clustering variable (city or cohort, respectively) as a fixed effect. However, the sensitivity analysis using the regression slopes from the corresponding models with random effects from the Green et al.[113] and Bashash et al.[112] studies,[118, 119] showed that a 1-mg/L increase in urinary fluoride was associated with a statistically significant lower IQ score of 1.80 points (95% CI: −2.80, −0.81). This suggests that clustering is not a significant issue in the results of our *regression slopes meta-analysis*.

eTable 6. Regression Slopes Meta-analysis

Analysis	Number of Studies	Beta (95% CI)	Heterogeneity	
			p-value	I^2
Overall Estimate				
Full-scale IQ	9	−1.81 (−2.80, −0.81)	<0.001	77%
Sensitivity Analyses				
Using the piecewise estimates from Yu et al. (2018)[3]				
Full-scale IQ	11	−1.68 (−2.65, −0.71)	<0.001	79%
Using effect estimates from Wang et al. (2020b)[4] rather than Yu et al. (2018)[3]				
Full-scale IQ	9	−1.70 (−2.55, −0.85)	<0.001	77%
Using Till et al. *(2020)[81] rather than Green et al. (2019)[113] estimates*				
Full-scale IQ	9	−1.83 (−2.80, −0.86)	<0.001	77%
Using estimates from random effect models for Green et al. (2019)[113] and Bashash et al. (2017)[112]				
Full-scale IQ	9	−1.80 (−2.80, −0.80)	<0.001	76%
Males	2	−2.39 (−5.89, 1.10)	0.070	69%
Females	2	−0.53 (−3.43, 2.37)	0.186	43%
Excluding Cui et al.[76]				
Full-scale IQ	8	−1.89 (−3.03, −0.74)	<0.001	80%
Excluding Yu et al. (2018)[3] and Zhang et al. (2015b)[110]				
Full-scale IQ	7	−1.76 (−2.90, −0.62)	<0.001	82%
Using unadjusted estimates from Bashash et al. (2017),[112] Cui et al. (2018),[76] Green et al. (2019)[113], *Yu et al. (2018)[3]*				
Full-scale IQ	9	−1.82 (−2.81, −0.83)	<0.001	76%
Using Verbal or Performance IQ scores from Green et al. (2019)[113]				
Verbal IQ	9	−1.78 (−2.78, −0.79)	<0.001	77%
Performance IQ	9	−1.77 (−2.77, −0.77)	<0.001	77%
Using Bashash et al. (2017)[112] McCarthy GCI scores, Valdez Jimenez et al. (2017)[74] (Bayley MDI scores), Cantoral et al. (2021)[83] (Bayley III cognitive scores), Ibarluzea et al. (2021)[87] (Bayley MDI scores).				
Urinary fluoride	11	−1.78 (−2.78, −0.78)	<0.001	75%
Intake	3	−3.28 (−5.87, −0.68)	0.799	0%
Water fluoride	2	−4.77 (−9.09, −0.45)	0.707	0%
Using Bashash et al. (2017)[112] McCarthy GCI scores, Valdez Jimenez et al. (2017)[74] (Bayley MDI scores), Cantoral et al. (2021)[83] (Bayley III cognitive scores), Ibarluzea et al. (2021)[87] (McCarthy GCI scores).				
Urinary fluoride	11	−1.90 (−2.86, −0.94)	<0.001	73%
Intake	3	−3.28 (−5.87, −0.68)	0.799	0%
Water fluoride	2	−4.77 (−9.09, −0.45)	0.707	0%

Notes:
CI = confidence interval; GCI = General Cognitive Index; MDI = Mental Development Index.

Commented [l21]: See Doc05_Meta-analysis, 5.C. (page 3 and 4), 5.D. (page 4 through 6) and 5.F. (page 7 and 8). See Doc06b_Meta-analysis, 6b.W., page 19 through 21.

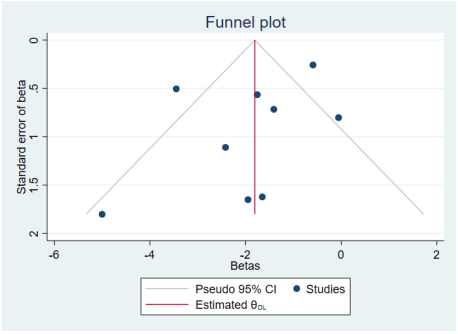


eFigure 19. Association Between Individual-level Fluoride Exposure and IQ Scores in Children: Overall Analysis

Estimates (betas) for individual studies are shown with solid boxes representing the weight, and the pooled estimate is shown as a solid diamond. Horizontal lines represent 95% Cis for the study-specific betas.

Commented [l22]: See Doc06b_Meta-analysis, 6b.C., page 2 and 3

Commented [l23]: See Doc06a_Meta-analysis, 6a.M., page 8 and 9



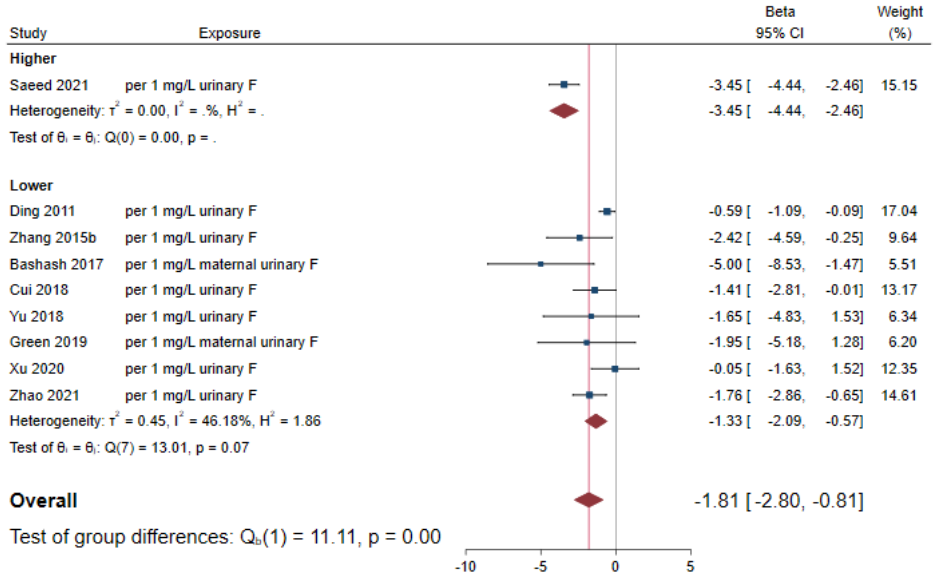
eFigure 20. Funnel Plot for Studies with Individual-level Exposures

```
Regression-based Egger test for small-study effects
Random-effects model
Method: DerSimonian-Laird

H0: beta1 = 0; no small-study effects
            beta1 =     -1.06
      SE of beta1 =     1.066
                z =     -1.00
       Prob > |z| =    0.3192
```

eFigure 21. Test for Publication Bias for Studies with Individual-level Exposures

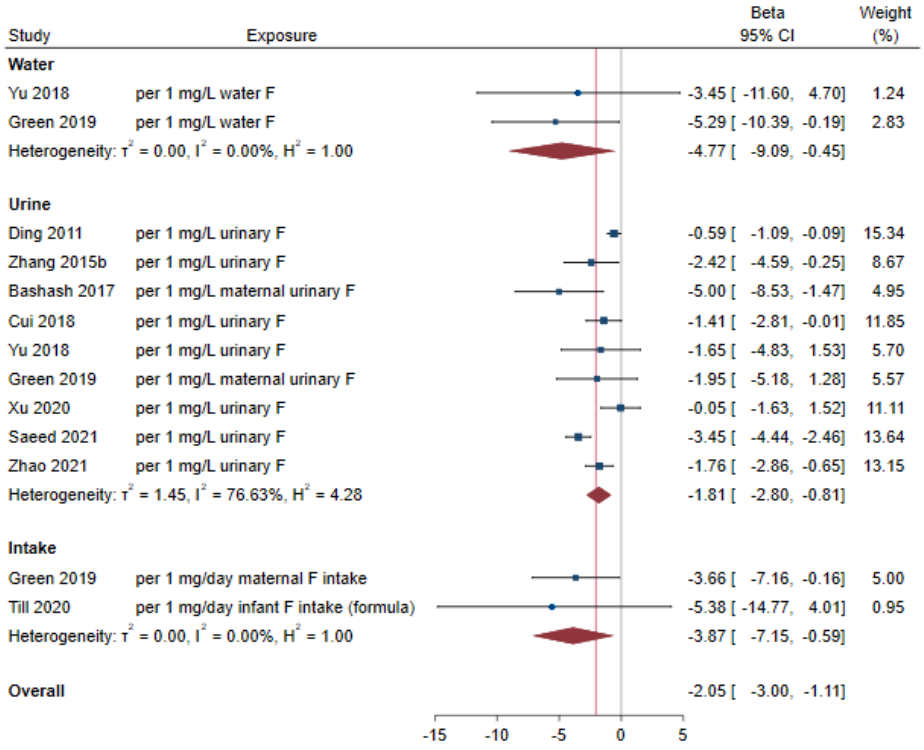
Subgroup Analyses

Risk-of-bias Subgroup Analysis



eFigure 22. Association Between Individual-level Fluoride Exposure and IQ Scores in Children: Effect by Risk of Bias

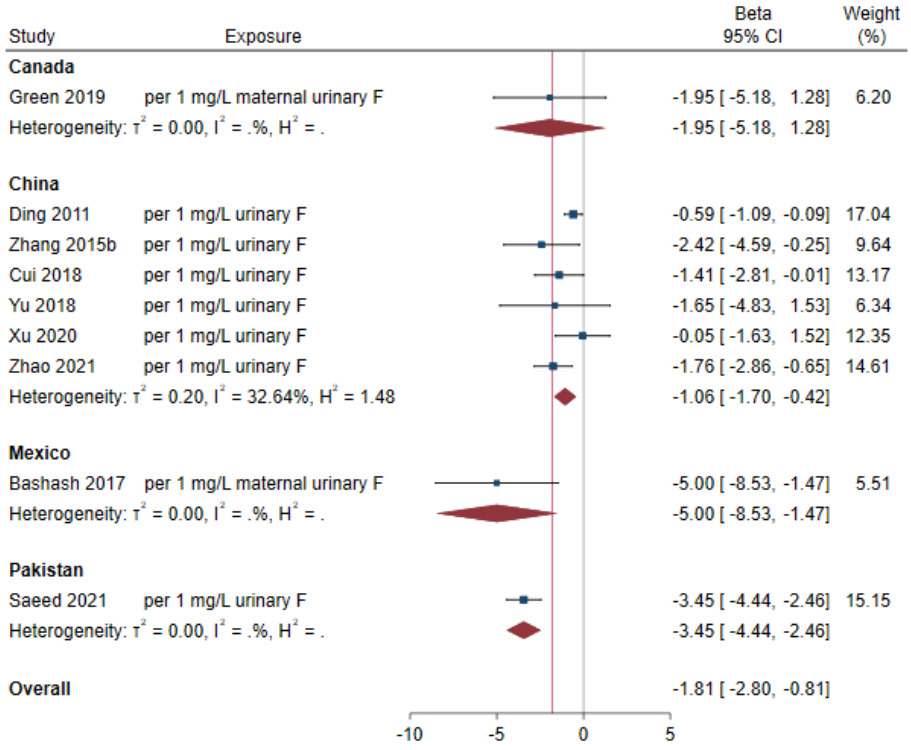
Exposure Type Subgroup Analysis



eFigure 23. Association Between Individual-level Fluoride Exposure and IQ Scores in Children: Effect by Exposure Type

Commented [l24]: See Doc06b_Meta-analysis, 6b.C., page 2 and 3

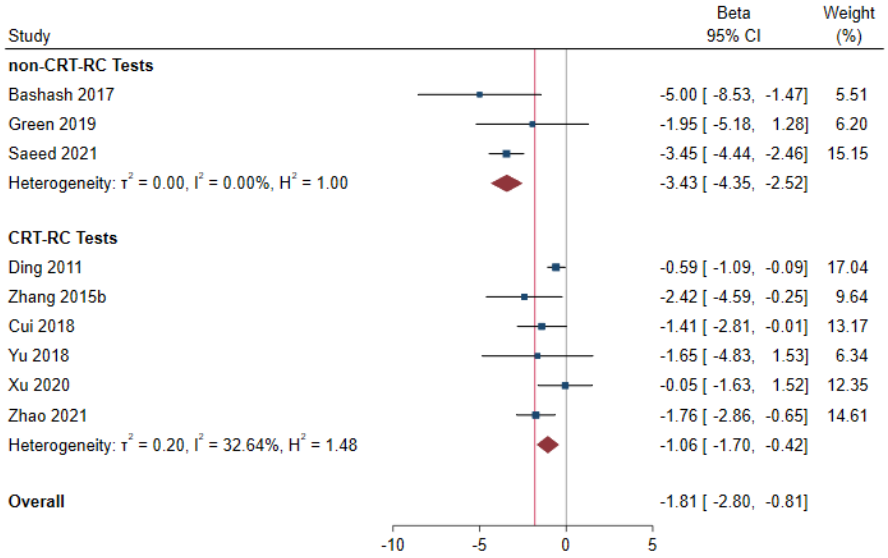
Country Subgroup Analysis



eFigure 24. Association Between Individual-level Fluoride Exposure and IQ Scores in Children: Effect by Country

Note: The analyses for publication bias for studies from China, Canada, and Mexico rely on a very small number of studies each and are not shown.

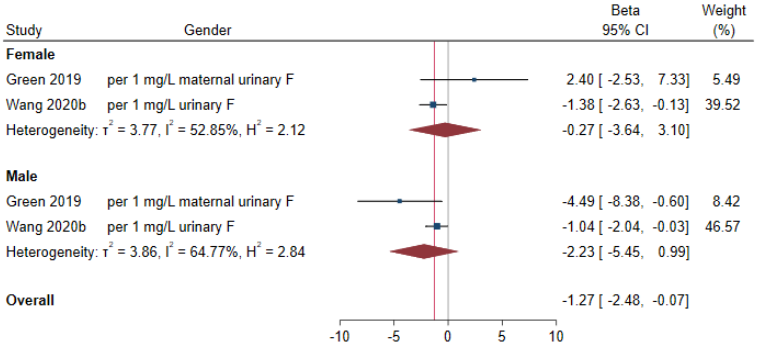
Assessment Type Subgroup Analysis



eFigure 25. Association Between Individual-level Fluoride Exposure and IQ Scores in Children: Effect by Assessment Type

Note: The analyses for publication bias for CRT-RC studies and non-CRT-RC studies include only six and three studies, respectively, and are not shown.

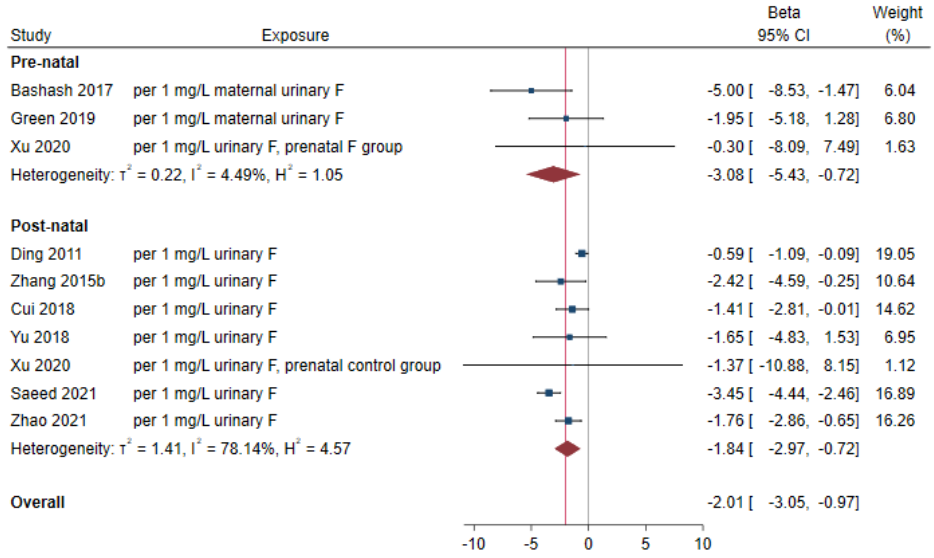
Sex Subgroup Analysis



eFigure 26. Association Between Individual-level Fluoride Exposure and IQ Scores in Children: Effect by Sex

Note: The analysis for publication bias by gender relies on two studies each and are not shown.

Pre-natal vs post-natal exposure Subgroup Analysis



eFigure 27. Association Between Individual-level Fluoride Exposure and IQ Scores in Children: Effect by Prenatal vs. Postnatal Exposure

References

1. National Toxicology Program (NTP). *Revised draft NTP Monograph on the systematic review of fluoride exposure and neurodevelopmental and cognitive health effects*. National Toxicology Program; 2020. https://www.nationalacademies.org/event/10-19-2020/docs/DDA97C9170D1A255D69C004CEB77B698E8D005011EFB
2. Higgins JT, Thomas J, Chandler J, et al., eds. *Cochrane Handbook for systematic reviews of interventions version 6.0 (updated July 2019)*. Cochrane; 2019.
3. Yu X, Chen J, Li Y, et al. Threshold effects of moderately excessive fluoride exposure on children's health: A potential association between dental fluorosis and loss of excellent intelligence. *Environ Int*. 2018;118:116-124. doi:https://doi.org/10.1016/j.envint.2018.05.042
4. Wang M, Liu L, Li H, et al. Thyroid function, intelligence, and low-moderate fluoride exposure among Chinese school-age children. *Environ Int*. 2020b;134:105229. doi:https://doi.org/10.1016/j.envint.2019.105229
5. Crippa A, Thomas I, Orsini N. A pointwise approach to dose-response meta-analysis of aggregated data. *Int J Stat Med Res*. 2018;7(2):25-32.
6. Crippa A, Discacciati A, Bottai M, Spiegelman D, Orsini N. One-stage dose–response meta-analysis for aggregated data. *Stat Methods Med Res*. 2019;28(5):1579-1596. doi:https://doi.org/10.1177/0962280218773122
7. Müller S, Scealy JL, Welsh AH. Model selection in linear mixed models. *Statist Sci*. 2013;28(2):135-167. doi:https://doi.org/10.1214/12-STS410
8. U.S. Environmental Protection Agency (US EPA). *New fluoride risk assessment and relative source contribution documents*. Office of Water; 2011. EPA-822-F-11-001. https://www.epa.gov/sites/default/files/2019-03/documents/fluoride-risk-assess-factsheet.pdf
9. World Health Organization (WHO). *Guidelines for drinking-water quality. Fourth edition*. 2011. https://apps.who.int/iris/bitstream/handle/10665/44584/9789241548151_eng.pdf?sequence=1&isAllowed=y&ua=1
10. Guo XC, Wang RY, Cheng CF, et al. [A preliminary investigation of the IQs of 7-13 year-old children from an area with coal burning-related fluoride poisoning]. *Chin J Epidemiol*. 1991;10(2):98-100.
11. Guo XC, Wang RY, Cheng CF, et al. A preliminary investigation of the IQs of 7-13 year-old children from an area with coal burning-related fluoride poisoning. *Fluoride*. 2008a;41:125-128.
12. Li Y, Li X, Wei S. [Effects of high fluoride intake on child mental work capacity: Preliminary investigation into the mechanisms involved]. *J West Chin University Med Sci*. 1994;25(2):188-191.
13. Li Y, Li X, Wei S. Effects of high fluoride intake on child mental work capacity: Preliminary investigation into the mechanisms involved. *Fluoride*. 2008b;41:331-335.
14. Li XS, Zhi JL, Gao RO. Effect of fluoride exposure on the intelligence of children. *Fluoride*. 1995;28:189-192.
15. Li F, Chen X, Huang R, Xie Y. The impact of endemic fluorosis caused by the burning of coal on the development of intelligence in children. *J Environ Health*. 2009;26(4):838-840.
16. Bai A, Li Y, Fan Z, Li X, Li P. [Intelligence and growth development of children in coal-burning-borne arsenism and fluorosis areas: An investigation study]. *Chin J Endemiol*. 2014;33(2):160-163.

17. Zhang P, Cheng L. Effects of coal-burning endemic fluorosis on children's physical development and intelligence. *Chin J Endem Dis Control*. 2015:458-459.

18. Wang J, Yu M, Yang L, Yang X, Deng B. The effect of coal-burning fluoride exposure on children's intelligence and physical development. *J Environ Health*. 2020c;37(11):971-974. doi:https://dx.doi.org/10.16241/j.cnki.1001-5914.2020.11.007

19. Lou D, Luo Y, Liu J, et al. Refinement impairments of verbal-performance intelligent quotient in children exposed to fluoride produced by coal burning. *Biol Trace Elem Res*. Feb 2021;199(2):482-489. doi:https://dx.doi.org/10.1007/s12011-020-02174-z

20. Ren D, Li K, Liu D. [A study of the intellectual ability of 8-14 year-old children in high fluoride, low iodine areas]. *Chin J Control Endem Dis*. 1989;4(4):251.

21. Ren D, Li K, Liu D. A study of the intellectual ability of 8-14 year-old children in high fluoride, low iodine areas. *Fluoride*. 2008;41:319-320.

22. Wang G, Yang D, Jia F, Wang H. [A study of the IQ levels of four- to seven-year-old children in high fluoride areas]. *Endem Dis Bull*. 1996;11(1):60-62.

23. Wang G, Yang D, Jia F, Wang H. A study of the IQ levels of four- to seven-year-old children in high fluoride areas. *Fluoride*. 2008b;41:340-343.

24. An J, Mei S, Liu A, Fu Y, Wang C. [Effect of high level of fluoride on children's intelligence]. *Chin J Control Endem Dis*. 1992;7(2):93-94.

25. Broadbent JM, Thomson WM, Moffitt TE, Poulton R. Community water fluoridation and intelligence response. *Am J Public Health*. 2015;105:3-4.

26. Xu Y, Lu C, Zhang X. [The effect of fluorine on the level of intelligence in children]. *Endem Dis Bull*. 1994;9(2):83-84.

27. Yao Y. Comparable analysis on the physical and mental development of children in endemic fluorosis area with water improvement and without water improvement. *Lit Inf Prev Med*. 1997;3(1):42-43.

28. Yao L, Zhou J, Wang S, Cui K, Lin F. Analysis on TSH and intelligence level of children with dental Fluorosis in a high fluoride area. *Lit Inf Prev Med*. 1996;2(1):26-27.

29. Seraj B, Shahrabi M, Falahzade M, Falahzade F, Akhondi N. Effect of high fluoride concentration in drinking water on children's intelligence. *J Dent Med*. 2006;19(2):80-86.

30. Seraj B, Shahrabi M, Shadfar M, et al. Effect of high water fluoride concentration on the intellectual development of children in Makoo, Iran. *J Dent*. 2012;9:221-229.

31. Eswar P, Nagesh L, Devaraj CG. Intelligent quotients of 12-14 year old school children in a high and low fluoride village in India. *Fluoride*. 2011;44:168-172.

32. Poureslami HR, Horri A, Garrusi B. A comparative study of the IQ of children age 7-9 in a high and a low fluoride water city in Iran. *Fluoride*. 2011;44:163-167.

33. Shivaprakash PK, Ohri K, Noorani H. Relation between dental fluorosis and intelligence quotient in school children of Bagalkot District. *J Indian Soc Pedod Prev Dent*. 2011;29:117-120.

34. Khan SA, Singh RK, Navit S, et al. Relationship between dental fluorosis and intelligence quotient of school going children in and around Lucknow District: A cross-sectional study. *J Clin Diagn Res*. 2015;9(11):10-15. doi:https://doi.org/10.7860/JCDR/2015/15518.6726

35. Sebastian ST, Sunitha S. A cross-sectional study to assess the intelligence quotient (IQ) of school going children aged 10-12 years in villages of Mysore District, India with different fluoride levels. *J Indian Soc Pedod Prev Dent*. 2015;33:307-311.

36. Mondal D, Dutta G, Gupta S. Inferring the fluoride hydrogeochemistry and effect of consuming fluoride-contaminated drinking water on human health in some endemic areas of Birbhum District, West Bengal. *Environ Geochem Health*. 2016;38:557-576.

37. Karimzade S, Aghaei M, Mahvi AH. Investigation of intelligence quotient in 9-12 year-old children exposed to high- and low-drinking water fluoride in West Azerbaijan Province, Iran. *Fluoride*. 2014;47:9-14.

38. Sun M, Li S, Wang Y, Li F. [Measurement of intelligence by drawing test among the children in the endemic area of Al-F combined toxicosis]. *J Guiyang Med Coll*. 1991;16(3):204-206.

39. Zhang J, Yao H, Chen Y. [The effect of high levels of arsenic and fluoride on the development of children's intelligence]. *Chin J Public Health*. 1998;17(2):119.

40. Trivedi M, Sangai N, Patel R, Payak M, Vyas S. Assessment of groundwater quality with special reference to fluoride and its impact on IQ of schoolchildren in six villages of the Mundra Region, Kachchh, Gujurat, India. *Fluoride*. 2012;45(4):377-383.

41. Trivedi MH, Verma RJ, Chinoy NJ, Patel RS, Sathawara NG. Effect of high fluoride water on intelligence of school children in India. *Fluoride*. 2007;40:178-183.

42. Zhao Q, Tian Z, Zhou G, et al. SIRT1-dependent mitochondrial biogenesis supports therapeutic effects of resveratrol against neurodevelopment damage by fluoride. *Theranostics*. 2020;10:4822-4838. doi:https://doi.org/10.7150/thno.42387

43. Zhao Q, Niu Q, Chen J, et al. Roles of mitochondrial fission inhibition in developmental fluoride neurotoxicity: Mechanisms of action in vitro and associations with cognition in rats and children. *Arch Toxicol*. Jan 18 2019;93(3):709-726. doi:https://doi.org/10.1007/s00204-019-02390-0

44. Zhou G, Tang S, Yang L, et al. Effects of long-term fluoride exposure on cognitive ability and the underlying mechanisms: Role of autophagy and its association with apoptosis. *Toxicol Appl Pharmacol*. 2019;378:114608. doi:https://doi.org/10.1016/j.taap.2019.114608

45. Qin L, Huo S, Chen R, Chang Y, Zhao M. [Using the Raven's Standard Progressive Matrices to determine the effects of the level of fluoride in drinking water on the intellectual ability of school-age children]. *Chin J Endem Dis*. 1990;5(4):203-204.

46. Qin LS, Huo SY, Chen RL, Chang YZ, Zhao MY. Using the Raven's Standard Progressive Matrices to determine the effects of the level of fluoride in drinking water on the intellectual ability of school-age children. *Fluoride*. 2008;41:115-119.

47. Yang Y, Wang X, X. G, Hu P. [The effects of high levels of fluoride and iodine on child intellectual ability and the metabolism of fluoride and iodine]. *Chin J Epidemiol*. 1994;15(4):296-298.

48. Yang Y, Wang X, Guo X, Hu P. The effects of high levels of fluoride and iodine on child intellectual ability and the metabolism of fluoride and iodine. *Fluoride*. 2008;41:336-339.

49. Wang X, Wang L, Hu P, Guo X, Luo X. [Effects of high iodine and high fluorine on children's intelligence and thyroid function]. *Chin J Endemiol*. 2001;20(4):288-290.

50. Wang SY, Zhang HX, Fan W, et al. [The effects of endemic fluoride poisoning caused by coal burning on the physical development and intelligence of children]. *J Appl Clin Ped*. 2005b;20(9):897-899.

51. Wang SY, Zhang HX, Fan W, et al. The effects of endemic fluoride poisoning caused by coal burning on the physical development and intelligence of children. *Fluoride*. 2008a;41:344-348.

52. Rocha-Amador D, Navarro ME, Carrizales L, Morales R, Calderon J. Decreased intelligence in children and exposure to fluoride and arsenic in drinking water. *Cad Saude Publica*. 2007;23(Suppl 4):S579-587.

53. Liu SL, Lu Y, Sun ZR, et al. [Report on the intellectual ability of children living in high-fluoride water areas]. *Chin J Control Endem Dis*. 2000;15(4):231-232.

54. Liu SL, Lu Y, Sun ZR, et al. Report on the intellectual ability of children living in high-fluoride water areas. *Fluoride*. 2008;41:144-147.

55. Lu Y, Sun ZR, Wu LN, Wang X, Lu W, Liu SS. Effect of high-fluoride water on intelligence in children. *Fluoride*. 2000;33:74-78.

56. Sudhir KM, Chandu GN, Prashant GM, Subba Reddy VV. Effect of fluoride exposure on intelligence quotient (IQ) among 13-15 year old school children of known endemic area of fluorosis, Nalgonda District, Andhra Pradesh. *J Indian Assoc Public Health Dent*. 2009;2009(13):88-94.

57. He MX, Zhang CN. [Investigation of children's intelligence quotient and dental fluorosis in drinking water-type of endemic fluorosis area in Pucheng County, Shaanxi Province before and after drinking water change]. *Chin J Endemiol*. 2010;29:547-548.

58. Xiang Q, Liang Y, Chen B, Chen L. Analysis of children's serum fluoride levels in relation to intelligence scores in a high and low fluoride water village in China. *Fluoride*. 2011;44:191-194.

59. Xiang Q, Liang Y, Chen L, et al. Effect of fluoride in drinking water on children's intelligence. *Fluoride*. 2003a;36:84-94.

60. Saxena S, Sahay A, Goel P. Effect of fluoride exposure on the intelligence of school children in Madhya Pradesh, India. *J Neurosci Rural Pract*. 2012;3:144-149.

61. Wang G, Gao M, Zhang M, Yang M, Xiang Q. Correlation between total fluoride intake and children's IQ. *J Southeast Univ Med Ed*. 2012:743-746.

62. Nagarajappa R, Pujara P, Sharda AJ, et al. Comparative assessment of intelligence quotient among children living in high and low fluoride areas of Kutch, India: A pilot study. *Iran J Public Health*. 2013;42:813-818.

63. Pratap SV, Singh CD, Sandeep T, et al. A correlation between serum vitamin, acetylcholinesterase activity and IQ in children with excessive endemic fluoride exposure in Rajasthan, India. *Int Res J Medical Sci*. 2013;1(3):12-16.

64. Asawa K, Pujara P, Thakkar JP, et al. Assessment of intelligence quotient among schoolchildren of fishermen community of Kutch, Gujarat, India. *Int Marit Health*. 2014;65(2):73-8. doi:https://doi.org/10.5603/imh.2014.0017

65. Wei N, Li Y, Deng J, Xu S, Guan Z. [The effects of comprehensive control measures on intelligence of school-age children in coal-burning-borne endemic fluorosis areas]. *Chin J Endemiol*. 2014;33(3):320-322.

66. Choi AL, Zhang Y, Sun G, et al. Association of lifetime exposure to fluoride and cognitive functions in Chinese children: A pilot study. *Neurotoxicol Teratol*. 2015;47:96-101.

67. Kundu H, Basavaraj P, Singla A, Gupta R, Singh K, Jain S. Effect of fluoride in drinking water on children's intelligence in high and low fluoride areas of Delhi. *J Indian Assoc Public Health Dent*. Apr 2015;13(2):116-121. doi:https://doi.org/10.4103/2319-5932.159043

68. Aravind A, Dhanya RS, Narayan A, Sam G, Adarsh VJ, Kiran M. Effect of fluoridated water on intelligence in 10-12-year-old school children. *J Int Soc Prev Community Dent*. Dec 2016;6(Suppl 3):S237-S242. doi:https://doi.org/10.4103/2231-0762.197204

69. Jin T, Han T, Wei Y, Wu Y, Wang Z, Zhang H. Investigation on working memory level of children aged 8-12 years in coal-burning fluorosis area. *J Environ Health*. 2016:409-411.

70. Zhang P, Cheng L. Effect of coal-burning endemic fluorosis on children's physical development and intellectual level. *Chin J Endem Dis Control*. 2015c;30(6):458-459.

71. Kumar R, Gupta VK, Kumar A, Shetty R, Singh M, Pandey V. Evaluation of correlation of dental fluorosis and cognitive status. *Ann Int Med Dental Res*. 2016;2(4):155.

72. Jin T, Wang Z, Wei Y, Wu Y, Han T, Zhang H. Investigation on intelligence level of children aged 8-12 years old in coal-burning fluorosis area. *J Environ Health*. 2017:229-231.

73. Razdan P, Patthi B, Kumar JK, Agnihotri N, Chaudhari P, Prasad M. Effect of fluoride concentration in drinking water on intelligence quotient of 12-14-year-old children in Mathura District: A cross-sectional study. *J Int Soc Prev Community Dent*. 2017;7:252-258. doi:https://doi.org/10.4103/jispcd.JISPCD_201_17

74. Valdez Jimenez L, Lopez Guzman OD, Cervantes Flores M, et al. In utero exposure to fluoride and cognitive development delay in infants. *Neurotoxicology*. Mar 2017;59:65-70. doi:https://dx.doi.org/10.1016/j.neuro.2016.12.011

75. Wang G, Zhang M, Wang Q, et al. Investigation on the relationship between serum fluoride content and IQ of children before and after reducing fluoride to water. *Capital Pub Health*. 2017:274-277.

76. Cui Y, Zhang B, Ma J, et al. Dopamine receptor D2 gene polymorphism, urine fluoride, and intelligence impairment of children in China: A school-based cross-sectional study. *Ecotoxicol Environ Saf*. 2018;165:270-277. doi:https://doi.org/10.1016/j.ecoenv.2018.09.018

77. Luo Y, Ma R, Liu Z, Guan Z, Lou D, Zheng D. Intelligence investigation and forensic significance of children in coal-burning fluorosis area. *Chin J Forensic Med*. 2018:590-593.

78. Naik SP, Bankur PK, Sathe S, Haris PMM, Kadar N, Satyanarayan A. Impact of fluoridated water on intelligence quotient levels of school children: A exploratory study. *Int J Oral Care Res*. 2018;6(1):63-66.

79. Sharma P, Bhardwaj AK, Singh M, Kumar D, Sharma A, Grover A. Does fluorosis affect the intelligence profile of children? A cross sectional analysis of school children of district Una, Himachal Pradesh, India. *Int J Community Med Public Health*. 2018;5(3):1047-53. doi:http://dx.doi.org/10.18203/2394-6040.ijcmph20180759

80. Soto-Barreras U, Escalante-Villalobos KY, Holguin-Loya B, et al. Effect of fluoride in drinking water on dental caries and IQ in children. *Fluoride*. 2019;52:474-482.

81. Till C, Green R, Flora D, et al. Fluoride exposure from infant formula and child IQ in a Canadian birth cohort. *Environ Int*. 2020;134:105315. doi:https://doi.org/10.1016/j.envint.2019.105315

82. Aggeborn L, Öhman M. The effects of fluoride in drinking water. *J Polit Econ*. 2020;129(2):465-491. doi:https://dx.doi.org/10.1086/711915

83. Cantoral A, Tellez-Rojo MM, Malin AJ, et al. Dietary fluoride intake during pregnancy and neurodevelopment in toddlers: A prospective study in the progress cohort. *Neurotoxicology*. Dec 2021;87:86-93. doi:https://dx.doi.org/10.1016/j.neuro.2021.08.015

84. Farmus L, Till C, Green R, et al. Critical windows of fluoride neurotoxicity in Canadian children. *Environ Res*. Sep 2021;200:111315. doi:https://doi.org/10.1016/j.envres.2021.111315

85. Guo B, Yu J, Cui Y, et al. DBH gene polymorphism, iodine and fluoride and their interactions and their interaction with children's intelligence. *J Environ Hygiene*. 2021;11(02):134-140.

86. Zhao Y, Cui Y, Yu J, et al. Study on the relationship between water-borne high iodine and thyroid hormone and children's intelligence level. *J Environ Health*. 2018:6-9.

87. Ibarluzea J, Gallastegi M, Santa-Marina L, et al. Prenatal exposure to fluoride and neuropsychological development in early childhood: 1-to 4 years old children. *Environ Res*. Oct 8 2021:112181. doi:https://dx.doi.org/10.1016/j.envres.2021.112181

88. Wang R, He N, Hou G, Zhang P, Jin W, Wang Y. Hengshui Survey and analysis of the mental health status of children aged 8-12 years in the high-fluoride area of Shanghai. *Med Anim Control*. 2021b;37(06):583-586.

89. Wang R, He N, Wang Y, Hou G, Zhang P-J. Investigation and analysis of children's dental fluorosis and IQ level in high fluoride district of Hengshui City. *Med Anim Control*. 2021;37(8):796-800. doi:https://dx.doi.org/10.7629/yxdwfz202108023

90. Yu X, Xia L, Zhang S, et al. Fluoride exposure and children's intelligence: Gene-environment interaction based on SNP-set, gene and pathway analysis, using a case-control design based on a cross-sectional study. *Environ Int*. 2021;155:106681. doi:https://doi.org/10.1016/j.envint.2021.106681

91. Zhao L, Yu C, Lv J, et al. Fluoride exposure, dopamine relative gene polymorphism and intelligence: A cross-sectional study in China. *Ecotoxicol Environ Safety*. Feb 2021;209:111826. doi:https://dx.doi.org/10.1016/j.ecoenv.2020.111826

92. Zhou G, Zhao Q, Luo C, et al. Low-moderate fluoride exposure and intelligence among Chinese school-aged children: Role of circulating mtDNA content. *Sci Total Environ*. Sep 10 2021;786:147330. doi:https://dx.doi.org/10.1016/j.scitotenv.2021.147330

93. Chen YX, Han FL, Zhoua ZL, et al. [Research on the intellectual development of children in high fluoride areas]. *Chin J Control Endem Dis*. 1991;6(Suppl):99-100.

94. Chen YX, Han FL, Zhoua ZL, et al. Research on the intellectual development of children in high fluoride areas. *Fluoride*. 2008;41:120-124.

95. Lin F, Ai H, Zhao H, Lin J, Jhiang J, Maimaiti. [High fluoride and low iodine environment and subclinical cretinism in Xinjiang]. *Endem Dis Bull*. 1991;6(2):62-67.

96. Zhao LB, Liang GH, Zhang DN, Wu XR. Effect of a high fluoride water supply on children's intelligence. *Fluoride*. 1996;29:190-192.

97. Hong FG, Cao YX, Yang D, Wang H. [Research on the effects of fluoride on child intellectual development under different environmental conditions]. *Chin Prim Health Care*. 2001;15(3):56-57.

98. Hong FG, Cao YX, Yang D, Wang H. Research on the effects of fluoride on child intellectual development under different environmental conditions. *Fluoride*. 2008;41:156-160.

99. Hong F, Wang Hui Yang Dong Zhang Z. Investigation on the intelligence and metabolism of iodine and fluoride in children with high iodine and fluoride. *Chin J Endem Dis Control*. 2001b:12-14.

100. Li YP, Jing XY, Chen D, Lin L, Wang ZJ. [Effects of endemic fluoride poisoning on the intellectual development of children in Baotou]. *Chin J Public Health Manag*. 2003;19(4):337-338.

101. Li YP, Jing XY, Chen D, Lin L, Wang ZJ. Effects of endemic fluoride poisoning on the intellectual development of children in Baotou. *Fluoride*. 2008c;41:161-164.

102. Wang SX, Wang ZH, Cheng XT, et al. [Investigation and evaluation on intelligence and growth of children in endemic fluorosis and arsenism areas]. *Chin J Endemiol*. 2005;24:179-182.

103. Wang Z, Wang S, Zhang X, Li J, Zheng X, Hu C. Investigation of children's growth and development under long-term fluoride exposure. *Chin J Control Endem Dis*. 2006;21(4):239-241.

104. Fan Z, Dai H, Bai A, Li P, Li T, Li G. The effect of high fluoride exposure in children's intelligence. *J Environ Health*. 2007;24(10):802-803.

105. Wang SX, Wang ZH, Cheng XT, et al. Arsenic and fluoride exposure in drinking water: Children's IQ and growth in Shanyin County, Shanxi Province, China. *Environ Health Perspect*. 2007;115:643-647.

106. Li X, Hou G, Yu B, Yuan C, Liu Y, L Z. [Investigation and analysis of children's IQ and dental fluorosis in high fluoride area]. *Chin J Pest Control*. 2010;26(3):230-231.

107. Ding Y, Sun H, Han H, et al. The relationships between low levels of urine fluoride on children's intelligence, dental fluorosis in endemic fluorosis areas in Hulunbuir, Inner Mongolia, China. *J Hazard Mater*. 2011;186:1942-1946.

108. Kang J, Cheng Y, Wu K, Lin S, He G, Jin Y. [Effect of exposure to fluoride and arsenic in drinking water of Hangjinhouqi on children's intelligence]. *Chin School Health*. 2011:679-681.

109. Wang S, Zhu X. Investigation and analysis of intelligence level of children in high fluoride area. *Chin J Endem Dis Control*. 2012b;27(1):67-68.

110. Zhang S, Zhang X, Liu H, et al. Modifying effect of COMT gene polymorphism and a predictive role for proteomics analysis in children's intelligence in endemic fluorosis area in Tianjin, China. *Toxicol Sci*. 2015b;144:238-245.

111. Das K, Mondal NK. Dental fluorosis and urinary fluoride concentration as a reflection of fluoride exposure and its impact on IQ level and BMI of children of Laxmisagar, Simlapal Block of Bankura District, W.B., India. *Environ Monit Assess*. 2016;188:218.

112. Bashash M, Thomas D, Hu H, et al. Prenatal fluoride exposure and cognitive outcomes in children at 4 and 6-12 years of age in Mexico. *Environ Health Perspect*. 2017;125(9):1-12. doi:https://doi.org/10.1289/ehp655

113. Green R, Lanphear B, Hornung R, et al. Association between maternal fluoride exposure during pregnancy and IQ scores in offspring in Canada. *JAMA Pediatr*. 2019:E1-E9.

114. Cui Y, Yu J, Zhang B, Guo B, Gao T, Liu H. The relationships between thyroid-stimulating hormone and/or dopamine levels in peripheral blood and IQ in children with different urinary iodine concentrations. *Neurosci Lett*. 2020;729:134981. doi:https://doi.org/10.1016/j.neulet.2020.134981

115. Xu K, An N, Huang H, et al. Fluoride exposure and intelligence in school-age children: evidence from different windows of exposure susceptibility. *BMC Public Health*. Nov 4 2020;20(1):1657. doi:https://dx.doi.org/10.1186/s12889-020-09765-4

116. Saeed M, Rehman MYA, Farooqi A, Malik RN. Arsenic and fluoride co-exposure through drinking water and their impacts on intelligence and oxidative stress among rural school-aged children of Lahore and Kasur districts, Pakistan. *Environ Geochem Health*. Nov 9 2021;doi:https://dx.doi.org/10.1007/s10653-021-01141-4

117. Rodriguez-Barranco M, Tobias A, Redondo D, Molina-Portillo E, Sanchez M-J. Standardizing effect size from linear regression models with log-transformed variables for meta-analysis. *BMC Med Res Methodol*. 2017;17(44)doi:https://doi.org/10.1186/s12874-017-0322-8

118. Bashash M. Personal communication from Morteza Bashash to Sorina Eftim regarding estimates from models with cohort effect considered using a random effect model. 2021.

119. Till C. Personal communication from Christine Till to Sorina Eftim regarding estimates from multi-level models predicting child FSIQ from MUF using city as a random effect. 2021.

Peer Review of the draft Meta-analysis Manuscript to Evaluate the Association between Fluoride Exposure and Children's Intelligence

[redacted] received a draft version of the manuscript as well as a copy of the NASEM Committee comments on the meta-analysis and the NIEHS/DNTP responses (draft version of Sup01_Meta-analysis). [redacted] provided comments in track changes on the draft manuscript in Microsoft Word. The full comments have been reproduced below verbatim along with the specific text referred to by as quotes under a heading for the specific section of the document (e.g., "Abstract section"). Note that the yellow highlighting was in the document as provided. Formatting has been applied to aid in reading. Responses have been added in blue text following each of the comments beginning with the word "**Response**" in bold font.

[redacted]
Date: July 27, 2021

1.A: ***Overall Comments:*** *This is such impressive work and glad you have put it into what [redacted] sure will be a high impact paper. [redacted] attaching some comments. [redacted] tried to highlight a couple of places where [redacted] thought your tox language needed to be modified to be more easily palatable to the clinical audience you're targeting, particularly the dose-response results. Those tended to be pretty confusing. [redacted] also had a lot of questions reading the abstract, prior to reading the paper. These are all editorial and [redacted] think your analysis is robust and your conclusions are in line with what the data are showing. Congratulations on great work.*

Response: Agree (no change requested)

- We appreciate [redacted] comments about this work, especially that the analyses are robust and the conclusions are in line with what the data are showing.

1.B: ***Abstract section:*** *"To perform a systematic review and meta-analysis to investigate associations between fluoride exposure and children's intelligence."*

[redacted] **Comment:** Maybe it's because [redacted] know what's coming but [redacted] find myself wanting to know in this sentence when fluoride exposure occurred for the studies included.

Response: Disagree (no change)

- The additional detail is not necessary in the abstract to convey the scope of the manuscript and there is a word limit for the abstract. The full description of fluoride exposure is included in the *Methods* section wherein the timing of exposure is described as "*pre- or post-natal exposure.*"

1.C: ***Abstract section:*** *"Inclusion criteria were assessment of cognitive outcomes, fluoride exposure, and statistical data on effect size."*

[redacted] **Comment:** Again, reading abstract before the article, but [redacted] am wondering if the exposure was all drinking water, or you combined studies with info on exposure biomarkers and drinking water levels?

Response: Disagree (no change)

- The exposures to fluoride included, but were not limited to, drinking water. There were biomonitoring measures (e.g., urinary fluoride) and other environmental measures (e.g.,

coal burning, areas endemic for dental fluorosis). We do not consider the additional detail in the abstract necessary to convey the scope of the manuscript, and due to the strict word limit for the abstract, the full description of fluoride exposures considered is included in the *Methods* section and in *Table 1. Characteristics of Studies Included in the Meta-analysis* (excerpt provided below).

Excerpt of Table 1. Characteristics of Studies Included in the Meta-analysis

Reference[a] *Study Design*	Study Location	Age Range (Years)	Fluoride Exposure		Intelligence Assessment	Overall RoB Rating	Confounders Considered
			Assessment (Metric, Exposure Groups)	Levels			
Ren et al. (1989)[66] [translated in Ren et al. 2008][me, o] *Cross-sectional*	China	8–14	No fluoride measurement Low iodine village/high fluoride and low iodine village	Not specified	Wechsler Intelligence Scale for Children	High	Sex; iodine
Chen et al. (1991)[68] [translated in Chen et al. 2008][me, w] *Cross-sectional*	China	7–14	Drinking water Nonendemic/endemic fluorosis village	0.89 mg/L (nonendemic) 4.55 mg/L (endemic)	Chinese Standardized Raven Test	High	Age; sex
Guo et al. (1991)[70] [translated in Guo et al. 2008a][me, o] *Cross-sectional*	China	7–13	Serum Reference area using wood/coal burning-related fluoride endemic area	0.1044 ± 0.0652 mg/L (reference) 0.1483 ± 0.0473 mg/L (endemic)	Chinese Binet Intelligence Test	High	Age; sex; SES

1.D: ***Abstract section:*** *"The meta-analysis of forty-six studies with group-level exposures found that children exposed to higher fluoride levels had lower IQ scores (SMD, −0.49; 95%CI, −0.60, −0.38; p-value<0.001)."*

[redacted] **Comment:** What does the SMD correspond to for this effect estimate?

Response: Agree (no change requested)

- This SMD compares groups of children living in areas with "high" fluoride exposure with children living in areas with "low" fluoride exposure. After updating the literature search, we revised the sentence to say:

  "The meta-analysis of 55 studies (N = 18,845 children) with group-level exposures found that, when compared to children exposed to lower fluoride levels, children exposed to higher fluoride levels had lower mean IQ scores (pooled SMD: −0.46; 95% CI: −0.55, −0.37; p-value < 0.001)."

1.E: ***Abstract section:*** *"When analyses were restricted to studies with groups exposed to ≤2 mg/L fluoride in drinking water, the mean SMD in children's IQ scores remained lower (SMD, -0.27; 95% CI: -0.36, -0.17; p-value<0.001)."*

[redacted] **Comment:** Can you briefly state why the cutoffs were chosen for this analysis?

Response: Agree (change made)

- We have removed this sentence from the abstract but have added the following rationale for the cutoffs to the *Methods* section:

  "We also examined whether there was a dose-response relationship at lower exposure levels that corresponded with the U.S. Environmental Protection Agency drinking water standards[20] and World Health Organization drinking water guidelines[21] (details provided in the Supplemental Materials)."

1.F: ***Introduction section:*** *(the yellow text highlighting was added by [redacted]) "However, many of the studies lacked details and key information necessary to evaluate study quality (e.g., measurement of covariates and other neurotoxic co-exposures)."*

[redacted] **Comment:** Both of the meta-analyses?

Response: Agree (change made)

- Yes, this statement refers to both previous meta-analyses (Choi et al. 2012 and Duan et al. 2018). We have revised the text in the *Introduction* section as follows:

  "Two previous meta-analyses[3, 4] found an association between high fluoride exposure and lower children's IQ; however, many of the studies in these meta-analyses lacked the information necessary to evaluate study quality and all used group-level estimates of fluoride exposure."2

1.G: [redacted] ***Comment:*** *You have references formatted in a few different ways so just double check throughout before submitting*

Response: Agree (change made)

- We have reformatted the references using superscript notation to align with the journal requirements.
- The changes are reflected throughout the document and an example from the previous comment is repeated below to show the formatting:

  "Two previous meta-analyses[3, 4] found an association between high fluoride exposure and lower children's IQ…"

1.H: ***Methods section:*** *(the yellow text highlighting was added by [redacted]) "Quality of individual studies, also called "risk of bias", was assessed using NTP's HAT approach."*

[redacted] **Comment:** Spell out?

Response: Agree (change made)

- This sentence has been revised in the *Methods* section as follows:

  "Quality of individual studies, also called "risk of bias," was assessed using the National Toxicology Program's Office of Health Assessment and Translation approach."

1.I: ***Results section:*** *(the yellow text highlighting was added by [redacted]) "Three studies[OBJ] unclear descriptions of their intelligence assessment methods, and sensitivity analyses did not reveal substantial changes in the pooled SMD estimate with or without these studies (−0.57; 95% CI: −0.69, −0.45)."*

[redacted] **Comment:** Something funny going on here

Response: Agree (change made)

- This formatting issue was a glitch with the citations and a broken link to the references which has been fixed.
- This sentence has been revised in the *Results* section as follows:

"Three studies[39, 40, 41 [translated in Li et al. 2008b]] lacked clear descriptions of their intelligence assessment methods; however, sensitivity analyses did not reveal substantial changes in the pooled SMD estimate when these studies were excluded or when a study[43] that reported the cognitive subset of evaluations using Bayley and McCarthy tests was included (eTable 3)."

1.J: ***Results section:*** *"For studies that had more than one exposed group (n = 17), a sensitivity analysis was performed to evaluate the impact of using all exposed groups combined compared to the reference group."*

[redacted] **Comment:** Made this a new paragraph.

Response: Agree (change made)

- We have taken [redacted] suggestion to make the text a new paragraph.

1.K: ***Results section:*** *"When the analyses were restricted to studies with the exposed group <1.5 mg/L fluoride in drinking water (n = 9; 2 lower risk-of-bias and 7 higher risk-of-bias studies) there was a non-significant positive association between fluoride exposure and children's IQ (SMD, 0.32; 95% CI: −0.57, 1.20). When restricted to studies with the exposed group <1.5 mg/L urinary fluoride (n = 4; 2 lower risk-of-bias and 2 higher risk-of-bias studies), the association was negative (SMD, −0.13; 95% CI: −0.29, 0.03; p-value=0.111]. When including groups exposed to < 2 mg/L urinary fluoride (n = 6; 3 lower risk-of-bias studies and 3 higher risk-of-bias studies), the association did not change substantially (−0.09; 95% CI: −0.22, 0.03; p-value=0.143). However, when including groups exposed to <2 mg/L fluoride in drinking water (n = 9; 2 lower risk-of-bias and 7 higher risk-of-bias studies), the association remained significant (SMD, −0.27; 95% CI: −0.36, −0.17; p-value<0.001) (eTable 4)."*

[redacted] **Comment:** This paragraph was tough for me to understand. First, [redacted] uncertain about why the cutoffs were chosen. Second, ordering should maybe start with <2? Going from all exposure levels to slightly lower exposure levels, to lowest exposure levels (<1.5). [redacted] also suggest grouping together the drinking water and urinary results in two separate paragraphs (including the overall linear results at the beginning of each paragraph). Last, [redacted] might be interested in comparing the number of low/high risk of bias studies across each grouping but it's too hard to follow, so maybe you could have a table of this info and summarize in a sentence or two here?

Response: Agree (change made)

- As noted in a previous comment, we have added the following rationale for the cutoffs to the *Methods* section:

  "We also examined whether there was a dose-response relationship at lower exposure levels that corresponded with the U.S. Environmental Protection Agency drinking water standards[20] and World Health Organization drinking water guidelines[21] (details provided in the Supplemental Materials)."

- Due to word count restrictions, we have limited the discussion of the results in the main manuscript to the linear model results. Therefore, the need for two separate paragraphs describing the drinking water and urinary results became unnecessary. We included additional results in the *Results* section of the supplemental materials. New tables

suggested by [redacted] have been added to the supplemental materials (eTable 4 and eTable 5; excerpts provided below). As suggested by [redacted], these tables were reordered to go from the least restrictive (all data) to most restrictive (<1.5 mg/L) exposure levels. These tables provide the overall linear results separately for drinking water and urinary results, the numbers of low and high risk-of-bias studies across each group, and the results when restricted to only the low risk-of-bias studies.

Excerpt of eTable 4. Dose-Response Meta-analysis Using Mean Effects—Model Selection

Exposure Analysis	Parameters	Fluoride Exposure			
		All data	<4 mg/L	<2 mg/L	<1.5 mg/L
Water Fluoride – All Studies					
No. Studies/No. Observations Number of Children		29/39 11,656	21/27 8,723	7/9 2,971	7/7 2,832
Linear Model[b]	Beta (95% CI) p-value AIC	−0.15 (−0.20, −0.11) p < 0.001 AIC = 53.8	−0.22 (−0.27, −0.17) p < 0.001 AIC = 16.1	−0.15 (−0.41, 0.12) p = 0.274 AIC = 11.8	0.05 (−0.36, 0.45) p = 0.816 AIC = 8.2
Quadratic Model[c]	Beta (95% CI); p-value Beta (95% CI); p-value AIC p-value*	−0.27 (−0.34, −0.21); p < 0.001 0.02 (0.01, 0.03); p < 0.001 AIC = 48.8 p* < 0.001	−0.12 (−0.35, 0.11); p = 0.318 −0.04 (−0.10, 0.03); p = 0.280 AIC = 21.2 p* = 0.012	0.79 (−0.01, 1.58); p = 0.052 −0.56 (−0.97, −0.16); p = 0.006 AIC = 12.5 p* = 0.007	0.30 (−0.53, 1.14); p = 0.477 −0.23 (−1.01, 0.55); p = 0.561 AIC = 11.3 p* = 0.04
Restricted Cubic Splines Model[d]	Beta (95% CI); p-value Beta (95% CI); p-value AIC p-value*	−0.29 (−0.39, −0.20); p < 0.001 0.48 (0.18, 0.78); p = 0.002 AIC = 42.3 p* < 0.001	−0.14 (−0.34, 0.06), p = 0.162 −0.23 (−0.66, 0.20), p = 0.295 AIC = 16.9 p* = 0.009	1.15 (0.07, 2.22) p = 0.037 −1.20 (−2.03, −0.36) p = 0.005 AIC = 10.5 p* = 0.010	0.49 (−0.50, 1.47) p = 0.334 −0.69 (−2.40, 1.02) p = 0.428 AIC = 10.2 p* = 0.05
Water Fluoride – Low Risk-of-bias Studies					
No. Studies/No. Observations Number of Children		6/11 4,355	6/9 4,251	3/4 921	3/3 879
Linear model	Beta (95% CI) p-value AIC	−0.19 (−0.34, −0.05) p = 0.009 AIC = 10.3	−0.22 (−0.36, −0.07) p = 0.003 AIC = 3.9	−0.34 (−0.72, 0.03) p = 0.070 AIC = 4.5	−0.32 (−0.91, 0.26) p = 0.276 AIC = 4.1

Excerpt of eTable 5. Dose-response Meta-analysis Using Mean Effects: Maximum Likelihood Models

Exposure Analysis	Parameters	Fluoride Exposure			
		All data	<4 mg/L	<2 mg/L	<1.5 mg/L
Urinary Fluoride – All Studies					
No. Studies/No. Observations Number of Children		18/32 8,502	13/26 6,885	7/11 4,654	5/8 3,992
Linear Model[b]	Beta (95% CI) p-value AIC	−0.16 (−0.23, −0.08) p < 0.001 AIC = 69.2	−0.17 (−0.29, −0.06) p = 0.004 AIC = 64.2	−0.07 (−0.13, 0.003) p = 0.060 AIC = −3.7	−0.12 (−0.36, 0.12) p = 0.325 AIC = 0.8
Quadratic Model[c]	Beta (95% CI); p-value Beta (95% CI); p-value AIC p-value*	−0.19 (−0.44, 0.06); p = 0.131 0.01 (−0.02, 0.05); p = 0.462 AIC = 73.0 p*= 0.14	0.08 (−0.21, 0.37); p = 0.587 −0.08 (−0.16, 0.0004); p = 0.051 AIC = 67.2 p* = 0.08	−0.23 (−0.62, 0.17); p = 0.267 0.08 (−0.12, 0.29); p = 0.423 AIC = 1.7 p* = 0.42	−0.11 (−1.45, 1.23); p = 0.868 0.02 (−0.74, 0.77); p = 0.967 AIC = 4.1 p* = 0.10
Restricted Cubic Splines Model[d]	Beta (95% CI); p-value Beta (95% CI); p-value AIC p-value*	−0.12 (−0.28, 0.04); p = 0.138 −0.10 (−0.41, 0.21); p = 0.524 AIC = 72.9 p*= 0.13	−0.03 (−0.21, 0.15); p = 0.775 −0.24 (−0.47, −0.02); p = 0.034 AIC = 66.8 p* = 0.07	−0.13 (−0.29, 0.03); p = 0.107 0.12 (−0.14, 0.38); p = 0.366 AIC = 1.5 p* = 0.37	−0.26 (−0.72, 0.20); p = 0.270 0.36 (−0.58, 1.29); p = 0.453 AIC = 3.5 p* = 0.07
Urinary Fluoride – Low Risk-of-bias Studies					
No. Studies/No. Observations Number of Children		9/15 5,713	9/15 5,713	5/8 4,141	4/7 3,952
Linear model	Beta (95% CI) p-value AIC	−0.10 (−0.20, 0.004) p = 0.059 AIC = 2.0	−0.10 (−0.20, 0.004) p = 0.059 AIC = 2.0	−0.07 (−0.14, 0.01) p = 0.073 AIC = −1.8	−0.08 (−0.16, −0.01) p = 0.028 AIC = −2.2

1.L: ***Discussion section:*** *"However, the associations did not remain significant when exposure was restricted to <1.5 mg/L, the current WHO safe water guideline."*

[redacted] **Comment:** There it is! Was looking for that info in the abstract and the methods/results. [redacted] question though—why use these cutoffs for the urinary levels as well? Do they directly translate?

Response: Agree (no change requested)

- These cutoffs are useful for comparison across different exposure measures. Drinking water levels roughly translate to urinary levels, but there is variation depending on the level of fluoride in the drinking water as well as individual behaviors. There is literature suggesting that among people living in areas with high levels of fluoride in drinking water, 1 mg/L in drinking water fluoride approximates 1 mg/L in urinary fluoride (e.g., Kumar et al. 2017); however, there is also literature suggesting that, at lower drinking water fluoride levels, drinking water only represents a portion of a person's total exposure to fluoride (EPA 2010), which includes exposure from other sources like dental products, foods, and beverages. Therefore, relying on drinking water levels may underestimate exposure.

  References

  S Kumar, S Lata, J Yadav, and JP Yadav (2017) Relationship between water, urine and serum fluoride and fluorosis in school children of Jhajjar District, Haryana, India. Appl Water Sci 7, 3377–3384. https://doi.org/10.1007/s13201-016-0492-2

  EPA (2010) Fluoride: Exposure and Relative Source Contribution Analysis. 890-R-10-015. US Environmental Protection Agency. Office of Water. Washington, D.C. Available at https://www.epa.gov/sites/default/files/2019-03/documents/fluoride-exposure-relative-report.pdf

1.M: ***Discussion section:*** *"While the results of our meta-analyses were consistent with two previous meta-analyses (Table 2), they differed in several ways. Our meta-analyses included more recently published studies that are lower risk of bias, and studies with different exposure assessment types. …If children with higher exposures had a greater IQ deficiency than children with lower exposures, the highly exposed children may have driven the mean IQ deficits of the entire group. Therefore, it is important to keep in mind that fluoride concentrations in drinking water alone do not reflect the magnitude of fluoride exposures to children who consume excessive amounts of fluoridated toothpaste or to formula-fed babies who consume powdered formula that is reconstituted with fluoridated water."*

[redacted] **Comment:** Not following the point you're trying to make in these sentences. It seems like first you were trying to make a point about how this meta-analysis is better than the previous, but then you're commenting on a problem with exposure assessment more generally. Seems like these should be separate discussion paragraphs.

Response: Agree (change made)

- We agree that the original paragraph sounds disjointed. Therefore, we revised the text and separated these topics into different paragraphs in the *Discussion* section as shown below:

  "Whereas the previously published meta-analyses only included group-level exposures, the regression slopes meta-analysis included nine studies with individual urinary fluoride measures, a more precise exposure measure. It also included recent North American

prospective cohort studies[5-7] with maternal urinary fluoride levels comparable to those found in the United States.[57]"

In a later paragraph in the Discussion section, we say:

"Fluoride exposure may vary considerably depending on individual behaviors and is best captured by individual-level measures of total exposure, such as urinary fluoride measures. Because drinking water measures capture only some of a person's total exposure to fluoride, it is reasonable to assume that some children in the meta-analysis had higher exposure to fluoride and those children may have skewed the mean IQ deficits of the entire group. Urinary fluoride levels include all ingested fluoride and are considered a valid measure to estimate total fluoride exposure.[61, 62]"

1.N: ***Discussion section:*** *"Consistent with previous literature, our dose-response meta-analysis shows statistically significantly lower children's IQ with increasing fluoride exposure."*

[redacted] **Comment:** Individual studies or the meta-analyses?

Response: Agree (change made)

- This was referring to one meta-analysis and another literature review, but we have since removed "consistent with previous literature" and directly cite the literature we are referring to in the *Discussion* section as follows:

  "The Duan et al.[4] meta-analysis reported a significant non-linear dose-response relationship above 3 ppm [3 mg/L] in water. A more recent literature review[56] did not comment on the shape of the dose-response curve; however, based on the three publications from Mexico and Canada,[5-7] the author concluded that the association between maternal urinary fluoride and children's neurotoxicity appeared to be "dose dependent."

1.O: ***Discussion section:*** *"Consistent with previous literature, our dose-response meta-analysis shows statistically significantly lower children's IQ with increasing fluoride exposure. Duan et al (2018) suggested a significant non-linear dose-response relationship above 3ppm [3 mg/L] in water [5]."*

[redacted] **Comment:** [redacted] having a hard time interpreting what this means. Relationship was non-linear, but there was an effect when levels were above 3 mg/L in the water (but not when levels were lower than that)? [redacted] not sure [redacted] would say this is "consistent with the literature" since it is just one other study? Maybe [redacted] just am not used to this dose-response language but consider that your other readers may not be either.

Response: Agree (change made)

- We have revised this part of the *Discussion* section to no longer say "consistent with the literature."

1.P: ***Discussion section****: "Our dose-response meta-analysis also revealed a significant dose-response relationship at <2mg/L fluoride in drinking water, levels that occur naturally in some U.S. drinking water systems."*

[redacted] **Comment:** [redacted] think you could consider language like "when levels were restricted to those below <2mg/L". [redacted] find it confusing because it's almost like you're comparing <2mg/L to some other level, which isn't what you're doing

Response: Agree (change made)

- We have removed this sentence from the manuscript and revised text in the supplemental materials to clarify that the dose-response analyses were restricted to (1) <4 mg/L, (2) <2 mg/L, and (3) <1.5 mg/L as follows:

  "When analyses were restricted to exposed groups with <4 mg/L (i.e., 0 to <4 mg/L) fluoride in drinking water (n = 21 publications [6 low and 15 high risk-of-bias studies]), there was a statistically significant inverse association between fluoride exposure and children's IQ (SMD: −0.22; 95% CI: −0.27, −0.17; p-value < 0.001) (eTable 4). When restricted to <2 mg/L (i.e., 0 to <2 mg/L) in drinking water (n = 7 publications [3 low and 4 high risk-of-bias studies]), the magnitude of the effect estimate did not substantially change (SMD: −0.15; 95% CI: −0.41, 0.12; p-value = 0.274). However, when restricted to exposed groups with <1.5 mg/L (i.e., 0 to <1.5 mg/L) in drinking water (n = 7 publications [3 low and 4 high risk-of-bias studies]), there was no longer an association between fluoride in drinking water and children's IQ (SMD: 0.05; 95% CI: −0.36, 0.45; p-value = 0.816)."

1.Q: ***Discussion section:*** *"Our dose-response meta-analysis also revealed a significant dose-response relationship at <2mg/L fluoride in drinking water, levels that occur naturally in some U.S. drinking water systems. As of April 2020, the CDC estimated that 0.59% of persons living in the United States (~ 1.9 million people) were served by community water systems (CWS) containing ≥ 1.5 mg/L naturally occurring fluoride and 0.31% (~1 million people) were served by CWS containing ≥ 2 mg/L (https://www.cdc.gov/fluoridation/data-tools/reporting-system.html)."*

[redacted] **Comment:** Not sure the point you're trying to make here. Shouldn't you be comparing CDC estimates to where levels are <2mg/L? [redacted] think these statistics are really important for contextualizing results but not sure this is the best result to compare them to?

Response: Agree (change made)

- We agree that our point was not clear. The point we wanted to make was that, even though the recommended level of artificially fluoridated water in the United States is 0.7 mg/L, some people may still be exposed to higher levels of naturally occurring fluoride in their drinking water. The revised text reads as follows:

  "For community water systems that add fluoride, the Public Health Service recommends a fluoride concentration of 0.7 mg/L; however, it is important to note that there are regions of the United States where public systems and private wells contain natural fluoride concentrations of more than 2 mg/L.[58] In April 2020, the Centers for Disease Control and Prevention (CDC) estimated that community water systems supplying water with ≥2 mg/L naturally occurring fluoride served 0.31% of the U.S. population (~1 million people).[59] For the purposes of reducing dental fluorosis, the CDC recommends that parents use an alternative source of water for children aged 8 years and younger and for bottle-fed infants if their primary drinking water contains greater than 2 mg/L of fluoride.[60]"

1.R: ***Discussion section:*** *"However, because all studies were considered lower risk of bias along with the moderate statistical heterogeneity and robustness our findings suggest that the small number of studies is unlikely to have influenced the meta-analysis findings."*

[redacted] **Comment:** Reword this sentence kind of run on

Response: Agree (change made)

- We agree with [redacted] comment; however, in the process of revising the manuscript, we have removed that sentence from the *Discussion* section and the issue no longer applies.

1.S: ***Conclusions section:*** *"The association remained statistically significant when restricted to <2 mg/L fluoride in drinking water (p-value<0.001), levels that occur naturally in some U.S. community water systems."*

[redacted] **Comment:** For more impact you could clarify this here with some statistic

Response: Agree (change made)

- We have removed this sentence from the *Conclusions* section because it was no longer accurate after the literature update.

Peer Review of the draft Meta-analysis Manuscript to Evaluate the Association between Fluoride Exposure and Children's Intelligence

[redacted] received a draft version of the manuscript as well as a copy of the NASEM Committee comments on the meta-analysis and the NIEHS/DNTP responses (draft version of Sup01_Meta-analysis). [redacted] provided comments in track changes on the draft manuscript in Microsoft Word. The full comments have been reproduced below verbatim along with the specific text referred to by [redacted] as quotes under a heading for the specific section of the document (e.g., "Abstract section"). Note that the red formatted text was in the document as provided. Formatting has been applied to aid in reading. Responses have been added in blue text following each of the comments beginning with the word "**Response**" in bold font.

[redacted]
Date: July 1, 2021

2.A: [redacted] **Comment:** Is the paper being submitted alongside a companion SR to discuss non-meta issues with fluoride and IQ?

Response: Agree (no change requested)

- Yes, the NTP Monograph on the systematic review of fluoride exposure and cognitive neurodevelopmental health effects is being published first and is referred to and cited in the *Methods* section as follows:

  "The search, selection, extraction, and risk-of-bias evaluation of studies for this meta-analysis were part of a larger systematic review.[8]"

2.B: Abstract section

[redacted] **suggested text:** [redacted] inserted text (shown here in red font) as follows: "*To perform a systematic review and meta-analysis to investigate epidemiological? associations between fluoride exposure and children's intelligence*."

[redacted] **Comment:** Even though this is for [NIEHS/DNTP removed journal name], [redacted] still think somewhere you should specify these are only epi studies

Response: Disagree (no change)

- We consider the objective in the *Abstract* to already imply that the meta-analysis only includes epidemiological studies with the word "children's" (i.e., "to investigate associations between fluoride exposure and children's intelligence"). In addition, details such as study eligibility are provided in the meta-analysis protocol (see Appendix 6 to the systematic review protocol located here: https://ntp.niehs.nih.gov/go/785076) and the *Methods* section of the manuscript.

2.C: Introduction section

[redacted] **suggested text:** [redacted] inserted text (shown here in red font) adding one sentence as follows**:** *This analysis was used to inform a larger systematic review on fluoride exposure and neurodevelopment*.

[redacted] **Comment:** [redacted] added this (to the Introduction) because doing a meta in isolation is bound to raise flags with reviewers, so best to mention this before the methods.

Response: Agree (change made)

- New text was added to the *Introduction* section to say that the meta-analysis complements "a larger systematic review" as follows:

  "To incorporate this newer evidence, and to complement a larger systematic review[8] that concluded there is moderate confidence in the evidence of an inverse association between fluoride exposure and children's IQ, we conducted a meta-analysis of studies that provided group-and individual-level fluoride exposure measurements in relation to children's IQ scores."

2.D: **Methods section:** "Literature searches were conducted in BIOSAS, EMBASE, PsychINFO, PubMed, Scopus, Web of Science, CNKI, and Wanfang databases through May 1, 2020 without language restrictions. Search strategies are available in the protocol.[8]"

[redacted] **Comment:** This also may raise flags if you are submitting 1+ year later. Worthwhile to be explicit for the early cutoff date (i.e., "the cutoff date chosen as part of our larger SR").

Response: Agree (change made)

- The literature search was updated in November 2021; therefore, the *Methods* section contains revised text:

  "Literature searches were conducted in BIOSIS, EMBASE, PsychINFO, PubMed, Scopus, Web of Science, CNKI, and Wanfang databases through November 2021, without language restrictions."

2.E: [redacted] **Comment:** You may benefit from upfront defining your inclusion criteria in a Supp table to stave off queries of age ranges, neurocognitive tests, etc.

Response: Disagree (no change)

- We define inclusion criteria in the protocol, which is referenced in the *Methods* sections of the manuscript and the supplemental materials, respectively, as follows:

  "To be eligible for inclusion, individual study publications had to satisfy review eligibility criteria outlined in the protocol.[9]"

  "In order to be eligible for inclusion in the systematic literature review, individual study publications (referred to in this paper as "studies") had to satisfy eligibility criteria outlined in the protocol (i.e., address PECO statement in Table 1 and specific exclusion criteria in Table 2, https://ntp.niehs.nih.gov/go/785076*)."*

2.F: **Methods section:** "The other risk-of-bias questions were also considered and were used to identify any other concerns that may indicate serious risk-of-bias issues (e.g., statistical analysis)."

[redacted] **Comment:** Maybe also mention as another example "selection bias", another critical domain in these studies

Response: Agree (change made)

- We have revised this sentence in the *Methods* section as follows:

"The other risk-of-bias questions were used to identify other concerns that may indicate serious risk-of-bias issues (e.g., selection bias, statistical analysis)."

2.G: **Methods section:** "No study was excluded from the meta-analysis based on concerns for risk of bias; however, subgroup analyses were conducted with and without higher risk-of-bias studies (i.e., studies rated probably high risk of bias for at least two key risk-of-bias questions or definitely high for any single question) to assess their impact on the results."

[redacted] **Comment:** Did you consider the magnitude and direction of the risk of bias? If your high RoB studies are high risk of bias for non-diff exp and outcome misclassification and all of your results bias toward the null, it may impact the interpretation of your results

Response: Agree (no change requested)

- Yes, this information has been considered and is available in Appendix E to the prepublication 2022 NTP Monograph.

2.H: **Methods section:** "Subgroup analyses were stratified by risk of bias (higher or lower), study location (e.g., country), outcome assessment, exposure matrix (e.g., urine or water), pre- or post-natal exposure, gender-specific groups, and age-specific subgroups."

[redacted] **Comment:** Are there others apart from urinary biomarkers and water exposure? Also, [redacted] maybe categorize as "urinary F, water F concentrations" or "biomonitoring, environmental sampling"

Response: Agree (change made)

- We have revised this sentence in the *Methods* section as follows:

  "Predefined subgroup analyses were stratified by risk of bias (high or low), study location (e.g., country), outcome assessment, exposure matrix (e.g., urinary fluoride or water fluoride concentrations), sex, and age group."

- Other exposure types were also considered, such as fluoride intake (see excerpt of Table 3 below).

  Excerpt of Table 3. Pooled Regression Slopes and 95% CIs for Children's IQ Score and Exposures to Fluoride

Analysis	Number of Studies	Beta (95% CI)	Heterogeneity	
			p-value	I^2
Overall Effect				
Full-scale IQ	9	−1.81 (−2.80, −0.81)	<0.001	77%
Subgroup Analyses				
Exposure Type				
Urinary fluoride	9	−1.81 (−2.80, −0.81)	<0.001	77%
Intake	2	−3.87 (−7.15, −0.59)	0.737	0%
Water fluoride	2	−4.77 (−9.09, −0.45)	0.707	0%

2.I: **Methods section:** "The study outcomes were evaluated with respect to a 1-mg/L unit increase in exposure."

[redacted] **Comment:** Is a 1 mg/L unit increase the same for urinary F (biomarker) and water measures (envir monitoring)?

Response: Agree (change made)

- We have revised this sentence in the *Methods* section as follows:

  "The study outcomes were evaluated with respect to a 1-mg/L unit increase in water or urinary fluoride, or 1-mg/day fluoride intake."

2.J: **Results section:** "The meta-analysis of 46 studies (37 lower risk of bias studies and 9 higher risk of bias studies) that provided mean IQ scores showed that children exposed to higher fluoride levels had statistically significantly lower IQ scores (random-effects pooled SMD, −0.49; 95% CI: −0.60, −0.38; p-value<0.001) (Figure 2)."

[redacted] **Comment:** That is ½ an IQ point. That's a big deal.

Response: Agree (no change requested)

2.K: **Results section:** "There was evidence of high heterogeneity (I^2 = 89%, p-value < 0.001, Table 2) and publication bias (funnel plot and Egger's p-value < 0.001, Begg's p = 0.04, eFigures 2 and 3)."

[redacted] **Comment:** Given the high level of heterogeneity, should you mention that this is further justification to conduct subgroup analyses (in addition to your justification for only reporting RE models)? Just a suggestion.

Response: Disagree (no change)

- We do not consider this to be necessary given that the protocol and *Methods* section describe that prespecified subgroup analyses were performed to investigate sources of heterogeneity.

2.L: **Results section:** "Among the higher risk-of-bias studies (n = 37), the random-effects pooled SMD was −0.55 (95% CI: −0.68, −0.43) with high heterogeneity (I^2 = 84%, p-value < 0.001) (Table 2 and eFigure 6)."

[redacted] **Comment:** Just a note that seeing consistency in your metas between your high and low RoB studies adds to your justification of an association

Response: Agree (no change requested)

2.M: **Results section:** "The overall pooled effect estimate from the six studies with individual-level urinary fluoride measures shows that a 1-mg/L increase in urinary fluoride is associated with a statistically significant lower IQ score of 1.58 points (95% CI: −2.63, −0.53; p-value=0.003)."

[redacted] **Comment:** Wow; is this after accounting for potential confounders? If so, that is substantial

Response: Agree (no change requested)

- Correct, this represents the pooled effect estimate using each study's adjusted regression coefficient.

2.N: **Results section**: "Adjusting for possible publication bias through trim-and-fill analysis supports the conclusion that a 1-mg/L increase in urinary fluoride was associated with lower IQ, with an adjusted pooled effect estimate of −0.87 (95% CI: −1.93, 0.19) (eFigure 19)."

[redacted] **Comment:** Report your p-value

Response: Agree (change made)

- We added p-values throughout the *Results* section.

2.O: **Results section:** "A 1-mg/L increase in fluoride intake and water fluoride are also significantly associated with a lower IQ score of 3.87 points (95% CI: −7.15, −0.59; p-value=0.021) and 4.77 points (95% CI: −9.10, −0.45; p-value=0.031), respectively (Table 3); however, the results for both metrics are based on a small sample of studies (n=2 for each measure) and should be interpreted with caution."

[redacted] **Comment:** Is an N of 2 even worth reporting, or including as a main result?

Response: Agree (change made)

- We have replaced the above sentence with the following:

  "The results for fluoride intake and water fluoride levels are available in Supplemental Materials."

2.P: [redacted] **Comment:** This section (Discussion) is great, but it is missing a robust discussion of the biological plausibility or proposed mechanism of action. A ton of implications that [redacted] think deserves its own pub.

Response: Disagree (no change)

- Potential biological mechanisms are covered in the state of the science report. However, currently, the data on mechanisms are too limited and heterogeneous to make a determination of biological plausibility and therefore we do not think it is appropriate to include this in the *Discussion*. However, we do agree this is an important area for continuing study and deserves a separate analysis and publication expanding on the potential limitations and promising research on mechanisms.

2.Q: **Discussion section:** "The results of our three meta-analyses support an inverse association between fluoride exposure and children's IQ. Results were robust to stratification by risk of bias, gender, age group, timing of exposure, study location, outcome assessment type, and exposure assessment type. The association remained statistically significant when the exposed group was restricted to <2 mg/L fluoride in drinking water (p-value<0.001), levels that occur naturally in some U.S. community water systems. However, the associations did not remain significant when exposure was restricted to <1.5 mg/L, the current WHO safe water guideline."

[redacted] **Comment:** Somewhere in this paragraph consider adding a sentence that this meta is used to inform the larger SR, and the meta is a piece of the larger equation

Response: Disagree (no change)

- We have included that this meta-analysis is part of the larger systematic review in the *Introduction* and *Methods* sections, respectively, as follows:

"To incorporate this newer evidence, and to complement a larger systematic review[8] that concluded there is moderate confidence in the evidence of an inverse association between fluoride exposure and children's IQ, we conducted a meta-analysis of studies that provided group-and individual-level fluoride exposure measurements in relation to children's IQ scores."

"The search, selection, extraction, and risk-of-bias evaluation of studies for this meta-analysis were part of a larger systematic review.[8]"

2.R: **Discussion section:** "Individual levels are a more precise measure of exposure compared to group-level measures; however, drinking water levels comprise only a portion of a person's total exposure to fluoride."

[redacted] **Comment:** Do you mean "household concentrations" or something else?

Response: Agree (change made)

- "Drinking water levels" in the above sentence referred to individual exposures to drinking water. However, during our revisions, this sentence was removed.

2.S: **Discussion section**

[redacted] **suggested text:** [redacted] inserted text (shown here in red font) at the end of the following sentence: "*Consequently, it is reasonable to assume that some children in our mean effects meta-analyses had higher exposure to fluoride from other common sources (e.g., dental products, foods and beverages)*; *though these are generally considered negligible relative to water*."

[redacted] **Comment:** [redacted] almost positive the relative source contribution of water compared to other sources is really disparately large

Response: Agree (change made)

- It has been estimated that other sources make up about 30% of total fluoride exposure.
- We have revised this sentence in the *Discussion* section as follows:

  "Because drinking water measures capture only some of a person's total exposure to fluoride, it is reasonable to assume that some children in the meta-analysis had higher exposure to fluoride and those children may have skewed the mean IQ deficits of the entire group."

2.T: [redacted] **Comment:** [redacted] know a major concern for EPA and other groups is infants whose sole source of consumption is formula from reconstituted tap water.

Response: Agree (change made)

- We have revised the *Discussion* section to include the following:

  "For the purposes of reducing dental fluorosis, the CDC recommends that parents use an alternative source of water for children aged 8 years and younger and for bottle-fed infants if their primary drinking water contains greater than 2 mg/L of fluoride."

2.U: **Discussion section:** "There are also several limitations to consider. Studies included in our meta-analyses also had various intrinsic limitations."

[redacted] **Comment**: Such as?

Response: Agree (change made)

- We have revised this sentence in the *Discussion* section as follows:

  "Most of the studies included in the mean-effects and dose-response mean effects meta-analyses were considered to have study design and/or methodological limitations. For example, all but three studies were cross-sectional in design."

2.V: [redacted] **Comment:** Table 1: You need to include study design in table 1

Response: Agree (change made)

- We have revised Table 1 to include study design in the first column (excerpt provided below).

  Excerpt of Table 1. Characteristics of Studies Included in the Meta-analysis

Reference[a] *Study Design*	Study Location	Age Range (Years)	Fluoride Exposure		Intelligence Assessment	Overall RoB Rating	Confounders Considered
			Assessment (Metric, Exposure Groups)	Levels			
Ren et al. (1989)[66] [translated in Ren et al. 2008][me, o] *Cross-sectional*	China	8–14	No fluoride measurement Low iodine village/high fluoride and low iodine village	Not specified	Wechsler Intelligence Scale for Children	High	Sex; iodine
Chen et al. (1991)[68] [translated in Chen et al. 2008][me, w] *Cross-sectional*	China	7–14	Drinking water Nonendemic/endemic fluorosis village	0.89 mg/L (nonendemic) 4.55 mg/L (endemic)	Chinese Standardized Raven Test	High	Age; sex
Guo et al. (1991)[70] [translated in Guo et al. 2008a][me, o] *Cross-sectional*	China	7–13	Serum Reference area using wood/coal burning-related fluoride endemic area	0.1044 ± 0.0652 mg/L (reference) 0.1483 ± 0.0473 mg/L (endemic)	Chinese Binet Intelligence Test	High	Age; sex; SES

2.W: [redacted] **Comment:** Figure 1: Add a Y axis (even if it is in the title)

Response: Agree (change made)

- We have revised Figure 1 to include the y-axis description in the title as follows:

  "Figure 1. Number of Studies of Fluoride Exposure and IQ in Children by Country and Year of Publication".

2.X: [redacted] **Comment:** Figure 2: Do you have space to add additional columns to increase the informativeness of this forest plot? Adding in the Ns and study designs would be helpful, but most importantly the exposure assessment used.

Response: Disagree (no change)

- We have kept Figure 2 as is for readability, but the subgroup analysis stratified by exposure assessment type is included in Table 2 (excerpt provided below).

Excerpt of Table 2. Pooled SMDs and 95% CIs for Children's IQ Score and Exposures to Fluoride

Analysis	Number of Studies	SMD (95% CI)	Heterogeneity	
			p-value	I^2
Overall Effect	55	−0.46 (−0.55, −0.37)	<0.001	87%
Subgroup Analyses				
Risk of Bias				
Low	10	−0.22 (−0.39, −0.05)	<0.001	83%
High	45	−0.52 (−0.63, −0.42)	<0.001	86%
Sex				
Males	14	−0.62 (−0.81, −0.42)	<0.001	78%
Females	13	−0.53 (−0.72, −0.34)	<0.001	74%
Age Group				
<10 years[a]	13	−0.41 (−0.60, −0.22)	<0.001	80%
≥10 years	16	−0.55 (−0.70, −0.40)	<0.001	68%
Country				
China	39	−0.43 (−0.52, −0.34)	<0.001	85%
India	8	−0.99 (−1.55, −0.43)	<0.001	93%
Iran	4	−0.68 (−0.99, −0.38)	0.077	56%
Canada	1	0.01 (−0.19, 0.21)	NA	NA
Mexico	1	0.13 (−0.16, 0.42)	NA	NA
New Zealand	1	0.01 (−0.19, 0.22)	NA	NA
Pakistan	1	−0.25 (−0.65, 0.16)	NA	NA
Assessment Type				
CRT-RC tests	29	−0.36 (−0.46, −0.27)	<0.001	82%
Non-CRT-RC tests	26	−0.60 (−0.78, −0.42)	<0.001	89%
Raven's tests	10	−0.76 (−1.10, −0.43)	<0.001	91%
Other tests	16	−0.52 (−0.74, −0.29)	<0.001	89%
Exposure Type				
Water fluoride	32	−0.37 (−0.46, −0.27)	<0.001	82%
Dental fluorosis	7	−0.99 (−1.57, −0.41)	<0.001	96%
Other exposures[b]	16	−0.54 (−0.71, −0.37)	<0.001	81%
Previous Meta-analyses				
Duan et al. (2018)[4]	26	−0.52 (−0.62, −0.42)	<0.001	69%
Choi et al. (2012)[3]	27	−0.45 (−0.56, −0.34)	<0.001	80%

Peer Review of the draft Meta-analysis Manuscript to Evaluate the Association between Fluoride Exposure and Children's Intelligence

[redacted] received a draft version of the manuscript as well as a copy of the NASEM Committee comments on the meta-analysis and the NIEHS/DNTP responses (draft version of Sup01_Meta-analysis). The full [redacted] comments have been reproduced below verbatim. Formatting has been applied to aid in reading. Responses have been added in blue text following each of the comments beginning with the word "**Response**" in bold font.

[redacted]
Date: September 14, 2021

3.A: [redacted] **Comment:** [redacted] this review represents an enormous effort. Meta-analyses are not my specialty but, by all evidence, what you have done is state-of-the-art. [redacted] do have some thoughts on how the paper might be framed for a clinical journal.

Response: Agree (no change requested)

- We appreciate the comment that this meta-analysis is state of the art.

3.B: [redacted] **Comment:** [redacted] realize the NTP defines its role in its reports strictly and narrowly–but if this paper is intended for a medical journal, then readers will expect some context. The paper could benefit from a brief section in the introduction on things like the sources of fluoride (ground water vs water supplement vs diet vs dental treatment), the drinking-water levels considered "optimum," the levels associated with toxicity (dental fluorosis), and an idea of the range of levels found in human populations. (This is probably not a complete list – [redacted] not an expert in this area.)

Response: Agree (change made)

- We agree that context for the findings is important and have added (1) information on the sources of fluoride to the *Introduction* section; (2) drinking water levels considered optimal as recommended by the U.S. Public Health Service to the *Discussion* section (this also addresses the levels associated with dental fluorosis because the optimal level is meant to provide enough fluoride to prevent tooth decay in children and adults while limiting the risk of dental fluorosis); and (3) the degree of exposures to high levels of naturally occurring fluoride in the United States to the *Discussion* section. As for the range of levels found in human populations, this varies widely depending on the geographic location of the population, the source of the exposure, and individual behaviors. Therefore, we considered it best to provide the exposure levels for each individual study population as reported by the study authors, which are available in Table 1 (excerpt provided below).
- We added text to the *Introduction* section as follows:

  "Fluoride from natural sources occurs in some community water systems and, in the United States and some other countries, fluoride is added to public drinking water systems for the prevention of tooth decay. Water and water-based beverages are the main source of systemic fluoride intake; however, an individual's total exposure also reflects contributions from fluoride in other sources such as food, dental products, industrial emissions, and some pharmaceuticals."

- We added text to the *Discussion* section as follows:

  "For community water systems that add fluoride, the Public Health Service recommends a fluoride concentration of 0.7 mg/L; however, it is important to note that there are regions of the United States where public systems and private wells contain natural fluoride concentrations of more than 2 mg/L.[58] In April 2020, the Centers for Disease Control and Prevention (CDC) estimated that community water systems supplying water with ≥2 mg/L naturally occurring fluoride served 0.31% of the U.S. population (~1 million people).[59]"

  Excerpt of Table 1. Characteristics of Studies Included in the Meta-analysis

Reference[a] *Study Design*	Study Location	Age Range (Years)	Fluoride Exposure: Assessment (Metric, Exposure Groups)	Fluoride Exposure: Levels	Intelligence Assessment	Overall RoB Rating	Confounders Considered
Broadbent et al. (2015)[38me, w*] *Prospective Cohort*	New Zealand	7–13	Drinking water Area without community water fluoridation (low)/area with community water fluoridation (high) Fluoride tablet use (never/ever) Fluoride toothpaste use (never/sometimes/always)	Water: 0.0–0.3 mg/L (low) 0.7–1.0 mg/L (high) Tablet use: 0 mg (never used) 0.5 mg (ever used) Range not specified for fluoride toothpaste use (always/sometimes/never)	Wechsler Intelligence Scale for Children-Revised	High	Sex; SES; low birth weight; breastfeeding
Cui et al. (2018)[34rs] *Cross-sectional*	China	7–12	Urine	Boys: 1.3 (0.9–1.7)[d] mg/L Girls: 1.2 (0.9–1.6)[d] mg/L	Combined Raven's Test for Rural China	Low	Age; maternal education; smoking in family member; stress; anger; dopamine receptor-2 polymorphism
Green et al. (2019)[6me, w*, u*, rs] *Prospective Cohort*	Canada	3–4	Maternal urine, drinking water, maternal fluoride intake Nonfluoridated/fluoridated area	Urine: 0.40 ± 0.27 mg/L (nonfluoridated) 0.69 ± 0.42 mg/L (fluoridated) Water: 0.13 ± 0.06 mg/L (nonfluoridated) 0.59 ± 0.08 mg/L (fluoridated) Intake: 0.30 ± 0.26 mg/day (nonfluoridated) 0.93 ± 0.43 mg/day (fluoridated) Overall: 0.51 ± 0.36 mg/L (urine) 0.54 ± 0.44 mg/day (intake) 0.31 ± 0.23 mg/L (water)	Wechsler Primary and Preschool Scale of Intelligence-III	Low	Sex; city; maternal education; race/ethnicity; HOME score; prenatal secondhand smoke exposure

3.C: [redacted] **Comment:** Is there a threshold effect? You report that "when restricted to exposed groups with <1.5 mg/L in drinking water…, there was no longer an association between fluoride in drinking water and children's IQ (SMD: 0.01; 95% CI: -0.37, 0.39; p-value=0.972)." This seems crucially important. The US Public Health Service recommends an optimum drinking-water concentration of 0.7 mg/L. According to an NHANES paper, 95% of US kids are exposed to drinking water with less than 1 mg/L fluoride. If all of this is correct, then the range at which the effects you find might actually occur are rare in the US, and perhaps other places as well. That needs to be emphasized.

Response: Agree (change made)

- Although [redacted] refers to a quote, the text has been paraphrased by [redacted], and the actual text in the version of the manuscript reviewed by [redacted] is as follows:

  "When the analyses were restricted to studies with the exposed group <1.5 mg/L fluoride in drinking water … there was a non-significant positive association between fluoride exposure and children's IQ (SMD, 0.32; 95% CI: −0.57, 1.20)."

- We agree that readers may consider the question of threshold and shape of the dose-response curve at low doses based on the results of the meta-analysis. We revised our discussion of the shape of the dose-response curve at low doses in the *Discussion* section of the manuscript as follows:

"There is inconsistency in which model is the best fit at lower exposure levels (eTable 4 and eTable 5) leading to uncertainty in the shape of the dose-response curve at these levels. More individual-level data would increase our certainty in the shape of the dose-response curve at these lower exposure levels."

- [redacted] links uncertainty in the shape of the dose-response curve at lower doses to a potential threshold and then focuses on drinking water concentrations only and the number of U.S. people with drinking water below 1 mg/L fluoride. However, as we stated in response to the previous question, there are multiple sources of fluoride that contribute to total exposure. We added a sentence to the *Introduction* section of the manuscript to emphasize the importance of total fluoride exposure and additional context to the *Discussion* section on the number of people in the United States served by water systems >2 mg/L fluoride as described below.

  Text was added to the *Introduction* section as follows:

  "Water and water-based beverages are the main source of systemic fluoride intake; however, an individual's total exposure also reflects contributions from fluoride in other sources such as food, dental products, industrial emissions, and some pharmaceuticals."

  Text was added to the *Discussion* section as follows:

  "For community water systems that add fluoride, the Public Health Service recommends a fluoride concentration of 0.7 mg/L; however, it is important to note that there are regions of the United States where public systems and private wells contain natural fluoride concentrations of more than 2 mg/L.[58] In April 2020, the Centers for Disease Control and Prevention (CDC) estimated that community water systems supplying water with ≥2 mg/L naturally occurring fluoride served 0.31% of the U.S. population (~1 million people).[59]"

3.D: [redacted] **Comment:** You know that people in certain quarters fear the government is "poisoning" water with fluoride. You are wading into that territory when you publish this in a medical journal, and you should provide as clear a picture of the practical implications as you can. The paper might benefit from adding a coauthor who is an expert in the clinical and public health context of fluoride research, and who could help connect this intensive statistical analysis to its public health setting.

Response: Disagree (no change)

- Yes, we are fully aware of the controversial nature of this topic. We are primarily interested in providing an accurate as possible analysis of the relevant literature with a transparent listing of strengths and limitations of the database. The issue of water fluoridation is not emphasized as we found no studies in the literature that were specifically designed or powered to examine this practice in relation to children's neurological development. In addition, we fully agree that the practical implications of this research are potentially wide ranging. However, given the additional analyses and scope of considerations involved, we consider the implications in the public health setting to be deserving of a more comprehensive risk-benefit analysis that is beyond the scope of this effort.

3.E: [redacted] **Comment:** The effect size appears markedly stronger in boys. [redacted] could not tell if this result might depend on the studies based on group measures rather than individual measures. If group measures are contributing, the result could be due to boys in hot climates drinking more water and thus having higher exposure. If the result is not persistent in studies that rest on individual exposure measures, then the interpretation could lean towards a biological vulnerability of boys. This seems like an important distinction to explore.

Response: Agree (change made)

- We have added text to the *Discussion* section to acknowledge this topic as a limitation as follows:

  "Although we conducted subgroup analyses by sex, only 1 of the 14 studies that reported IQ scores separately for boys and girls analyzed fluoride exposure for each sex separately.[6] This is essential for evaluating whether a differential change in IQ by sex may be related to higher susceptibility or higher exposure in that sex."

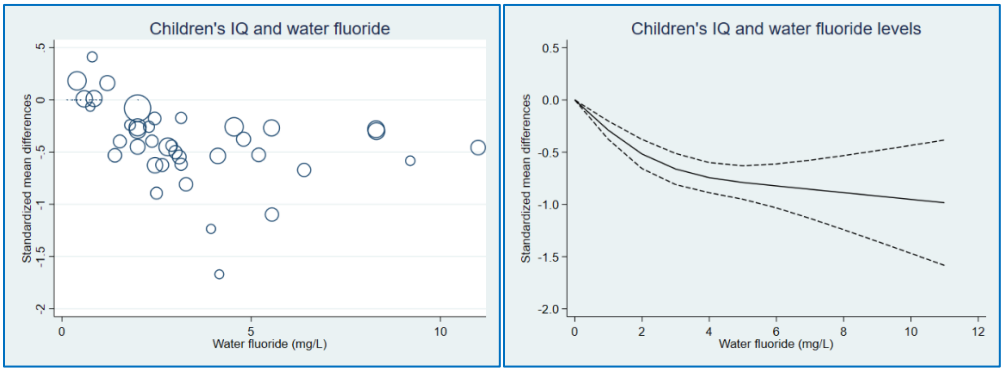
- This topic is also addressed more fully in the prepublication 2022 NTP Monograph which, now that it has undergone exhaustive peer review, will be cited in the manuscript.

3.F: [redacted] **Comment:** Some minor comments.

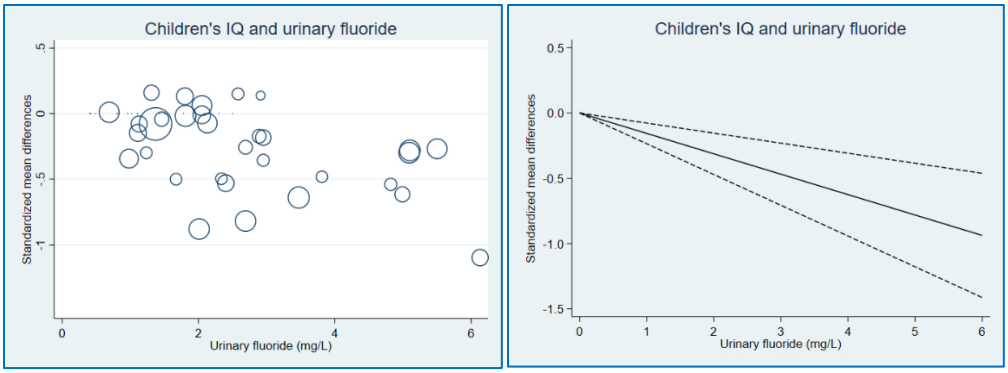
[redacted] **Comment:** [redacted] would suggest that the Y axis on the two panels in Supplementary Figure 16 be made on the same scale, so that it is easier to move between the two panels.

Response: Agree (change made)

- [redacted] is referring to eFigure 26. We agree with the commenter's suggestion on scale and also found this change could be applied to eFigure 25. Therefore, we have updated eFigure 25 and eFigure 26 (which are eFigure 17 and eFigure 18 in the current draft supplemental materials) so that the y-axes on the two panels use the same scale.



eFigure 17. Pooled Dose-Response Association Between Fluoride in Water and Standardized Mean Differences in Children's IQ



eFigure 18. Pooled Dose-Response Association Between Fluoride in Urine and Standardized Mean Differences in Children's IQ

3.G: [redacted] **Comment:** With regard to outcome, there should be some mention of the usual standard deviation of IQ in the population, to give a better idea of how large a one-point difference might be.

Response: Disagree (no change)

- The standard deviations of measured IQs are specific to study population. Since the meta-analyses we perform pool the results of many different study populations together, and range between mean-effects, dose-response, and regression slopes, we consider it to be misleading to provide a "usual standard deviation IQ."

3.H: [redacted] **Comment:** Medical journals typically do not allow footnotes.

Response: Agree (change made)

- All footnotes have been removed from the manuscript.

3.I: [redacted] **Comment:** [redacted] assume the words IQ "score" and IQ "point" are equivalent, but as [redacted] first read the abstract, [redacted] couldn't be sure. Consistent use might avoid any possible confusion.

Response: Agree (change made)

- When compared against appropriate population norms, the IQ score has a point value. Thus, decreases in IQ score can be expressed as numeric points. However, not all studies report scores in this manner, with some reporting only raw IQ test scores. We have reviewed text in the manuscript to make sure that all references to changes in IQ scores or points correctly reflect the underlying information.

3.J: [redacted] **Comment:** Finally, [redacted] found the constant "thanks" to [redacted] in each draft response to be distracting. Better to get to the point.

Response: Agree (change made)

- We have toned down the "thanks" in our responses to the NASEM Committee comments, which is reflected in Sup01_Meta-analysis and Sup01_Monograph.

3.K: [redacted] **Comment:** [redacted] hope this is useful. Thanks for the opportunity to have a look at this important piece of work. [[redacted]]

Response: Agree (no change requested)

Peer Review of the draft Meta-analysis Manuscript to Evaluate the Association between Fluoride Exposure and Children's Intelligence

[redacted] received a draft version of the manuscript as well as a copy of the NASEM Committee comments on the meta-analysis and the NIEHS/DNTP responses (draft version of Sup01_Meta-analysis). The full [redacted] comments have been reproduced below verbatim. Formatting has been applied to aid in reading. A response has been added in blue text following the comments beginning with the word "**Response**" in bold font.

[redacted]
Date: September 17, 2021

4.A: [redacted] **Comment**: [redacted] gone over the paper in detail, and this is excellent work. [redacted] genuinely don't have any concerns or suggestions. [redacted] think the analysis itself is excellent, and you thoroughly addressed comments.

Response: Agree (no change requested)

- We appreciate [redacted] comments that [redacted] does not have any concerns or suggestions and that we have thoroughly addressed the NASEM Committee comments.

In February 2022, the [redacted] provided comments to NIEHS/DNTP on the 2021 *Draft NTP Monograph on the State of the Science concerning Fluoride Exposure and Neurodevelopmental and Cognitive Health Effects: A Systematic Review* and a draft manuscript on a meta-analysis of fluoride exposure and IQ in children. This document contains a subset of the overall [redacted] comments related to the meta-analysis manuscript along with the NIEHS/DNTP responses. The meta-analysis-related comments from the [redacted] are reproduced here in black text, and the NIEHS/DNTP responses have been inserted in blue text following each of the comments beginning with the word "**Response**" in bold font.

February 1, 2022

[redacted]

Feedback to NTP/NIEHS regarding:

1. Fluoride state of the science document
2. Fluoride and IQ Meta-analysis manuscript

5.A: Issue: Keeping findings in context

As NASEM noted in their review of the 2019 Draft Monograph, "the context into which the monograph falls calls for much more carefully developed and articulated communication on this issue." [redacted] fully concurs with this recommendation and with NASEM's 2019 assessment that "NTP needs to state clearly that the monograph is not designed to be informative with respect to decisions about the concentrations of fluoride that are used for water fluoridation."

NTP stated in the revised draft of the monograph that the evidence of "effects on cognitive neurodevelopment are inconsistent, and therefore unclear" at the levels typically found in drinking water in the US. NASEM agreed with this assessment, stating that "[m]uch of the evidence presented in the report comes from studies that involve relatively high fluoride concentrations. Little or no conclusive information can be garnered from the revised monograph about the effects of fluoride at low exposure concentrations (less than 1.5 mg/L)."

[redacted] is extremely concerned that the revised 2021 NTP report and the meta-analysis omit this important context that was previously included. Without clarification, readers may interpret that exposure to fluoride at any concentration is associated with lower IQ, a conclusion that is not borne out by the available science or the findings of the systematic review.

Recommendation:

- [redacted] requests NTP include a statement in the systematic review abstract and fulltext, as well as the meta-analysis, like that found in the 2020 draft monograph: "When focusing on findings from studies with exposures in ranges typically found in drinking water in the United States (0.7 mg/L for optimally fluoridated community water systems) that can be evaluated for dose response, effects on cognitive neurodevelopment are inconsistent, and therefore unclear."

  Response: Disagree (no change)

  - We remain sensitive to the need to provide context concerning fluoride exposures in the United Stated from fluoridated water, and we have included the PHS recommendations for optimal water fluoridation in the meta-analysis manuscript. However, we also stress that the subject of our fluoride monograph and meta-analysis is total fluoride exposures from all sources. The 2022 update of the meta-analysis includes a number of new non-U.S. studies that further inform the relationship between IQ deficits in children and exposures to fluoride that were not available for inclusion in the 2020 draft NTP Monograph. These studies provide additional information to sharpen the dose-response mean-effects estimates and improve the *regression slopes meta-analysis*. Although the clarity of effects at lower fluoride exposures is improving, there are no studies on the potential association between fluoride exposures and IQ in children in the United States, and no nationally representative urinary fluoride levels are available, making it difficult to make more specific

statements about the relevance of our meta-analysis findings to the U.S. population.

Note: [redacted] comments on the animal studies for the prepublication 2022 NTP Monograph are not reproduced here as they are not relevant to the meta-analysis. See DocA1_Monograph for the monograph-relevant comments and responses.

5.B: Issue: Limitations section

In its response letter, NASEM requested adding clarifying information in the manuscript. NTP itemized items in the state-of-the-science manuscript on limitations of the evidence based and the systematic review. However, these limitations do not address the following issues comprehensively:

Note: [redacted] comments on the protocol and literature search (numbered as "1" and "2" in the original comments) for the prepublication 2022 NTP Monograph are not reproduced here as they are not directly relevant to the meta-analysis. To avoid confusion, the number "3" was removed from following comment. See DocA1_Monograph for the monograph-relevant comments and responses.

5.C: Some included studies with complex sample designs did not report if they used population weights to generate estimates.

Recommendation: In addition to listing this as a limitation, NTP should identify these studies in the body of the report.

Response: Agree (change made)

- We have addressed these issues in the meta-analysis. We specifically mentioned the studies in the *Results* section of supplemental materials. In addition, we performed a new sensitivity analysis excluding results from the studies that did not account for complex sampling strategies (Yu et al. (2018), Zhang et al. (2015b)). Based on this analysis, the pooled effect estimate did not change appreciably (see excerpt of eTable 6 below).

Excerpt of eTable 6. Regression Slopes Meta-analysis

Analysis	Number of Studies	Beta (95% CI)	Heterogeneity	
			p-value	I^2
Overall Estimate				
Full-scale IQ	9	−1.81 (−2.80, −0.81)	<0.001	77%
Sensitivity Analysis				
Excluding Yu et al. (2018)[3] and Zhang et al. (2015b)[110]				
Full-scale IQ	7	−1.76 (−2.90, −0.62)	<0.001	82%

- Additionally, our risk-of-bias assessment carefully considered accounting for sampling strategy or clustering in determining study-specific potential for bias. Our analyses stratify results by risk-of-bias status to evaluate the potential impact on the overall effect estimates from studies that have high potential for bias versus studies that have low potential for bias.

5.D: Clustering: NASEM identified that in some population studies, participants living in the same communities were assigned the same measure of fluoride exposure without considering the effect in the data analysis. These correlation may artificially increase the statistical power.

Recommendation: Limitations should note the studies where clustering was a potential threat and specifically whether the investigators addressed this.

Response: Agree (change made)

- Based on the NASEM Committee's comment, we revised text in Appendix E of the prepublication NTP 2022 Monograph (previously Appendix 4 in the 2020 draft NTP Monograph) to clearly specify which low risk-of-bias studies addressed clustering as a feature of the study design or statistical analysis. When clustering was not accounted for, we describe the expected impact that this may have on the study's results.

  We have performed several additional sensitivity analyses to address the NASEM Committee's comments on clustering (further described below). The new results are presented in eTable 3 and eTable 6 of the supplemental materials (excerpts provided below).

  For example, we added a sensitivity analysis excluding Trivedi et al. (2012) from the *mean-effects meta-analysis* (both the overall effect analysis and the low risk-of-bias subgroup analysis) to assess the impact of clustering. Excluding Trivedi et al. (2012) did not change the results appreciably. The results of this new sensitivity analysis compared to the main overall effect estimate are shown below in the excerpt of eTable 3.

Excerpt of eTable 3. Sensitivity Analyses for Mean-effects Meta-analysis: Pooled SMDs and 95% CIs for Children's IQ Score and Exposures to Fluoride

Analysis	Number of Studies	Beta (95% CI)	Heterogeneity	
			p-value	I^2
Overall Estimates				
Overall effect	55	−0.46 (−0.55, −0.37)	<0.001	87%
Low risk of bias	10	−0.22 (−0.39, −0.05)	<0.001	83%
Sensitivity Analyses excluding Trivedi et al. (2012)[40]				
Overall effect	54	−0.46 (−0.56, −0.37)	<0.001	87%
Low risk of bias	9	−0.22 (−0.40, −0.04)	<0.001	85%

- As suggested by the NASEM Committee, lack of accounting for clustering has little impact on studies with individual-level exposure measures (e.g., urinary fluoride levels) that also account for important confounders capturing the cluster (city) effect. For example, the minimal impact of clustering is illustrated by Bashash et al. (2017) who accounted for clustering at the cohort level by using cohort as a fixed effect in the models. In addition, the models accounted for many important confounders, which are also likely to reflect the cohort effect. The unadjusted and adjusted effect estimates were similar (β [95% CI] = −2.37 [−4.45, −0.29] and −2.50 [−4.12, −0.59], respectively).
- In the case of Green et al. (2019), we contacted the study authors and received the results from models using city as a random intercept. The overall adjusted effect estimates with city as a fixed effect and with city as a random effect were not significantly different from each other (β [95% CI] = −1.95 [−5.19, 1.28] and −2.20 [−5.39, 0.98], respectively).
- To address the NASEM Committee's concerns about clustering, we performed two new sensitivity analyses—one using the unadjusted effect estimates from Bashash et al. (2017), Cui et al. (2018), Green et al. (2019), and Yu et al. (2018) and another using the estimates from the random effect models from Bashash et al. (2017) and Green et al. (2019). The additional sensitivity analyses had minimal impact on the overall results of the meta-analysis (see excerpt of eTable 6 below).

Excerpt of eTable 6. Regression Slopes Meta-analysis

Analysis	Number of Studies	Beta (95% CI)	Heterogeneity	
			p-value	I^2
Overall Estimate				
Full-scale IQ	9	−1.81 (−2.80, −0.81)	<0.001	77%
Sensitivity Analyses				
Using estimates from random effect models for Green et al (2019)[113] and Bashash et al. (2017)[112]				
Full-scale IQ	9	−1.80 (−2.80, −0.80)	<0.001	76%
Using unadjusted estimates from Bashash et al. (2017),[112] Cui et al. (2018),[76] Green et al. (2019)[113], Yu et al. (2018)[3]				
Full-scale IQ	9	−1.82 (−2.81, −0.83)	<0.001	76%

Note: [redacted] comment on contacting authors of studies considered in the prepublication 2022 NTP Monograph with reporting quality questions as part of the risk-of-bias assessment are not reproduced here as they are not directly relevant to the meta-analysis. See DocA1_Monograph for the monograph-relevant comments and responses.

Meta-Analysis: The meta-analysis, originally requested by NASEM to obtain measures of association and sensitivity analysis across selected studies was removed to be published separately.

Note: [redacted] comment on the association between the prepublication 2022 NTP Monograph and meta-analysis is not reproduced here as it is not directly relevant to the meta-analysis itself. See DocA1_Monograph for the monograph-relevant comments and responses.

[redacted] concluded their comment with the statement that:
[NTP] should address NASEM's critiques of the September 2020 draft (abstracted below):

5.E:

a. Provide sufficient information about each study to allow the reader to understand why particular outcomes/results were selected (data transparency)

Response: Agree (change made)

- The NASEM Committee suggested the addition of a table providing more information on each study that "would allow readers to identify which result from each study was used and support a better understanding of why NTP selected the results that it did for inclusion in the meta-analysis." We found the suggestion helpful and have newly included eTable 2 (Study Characteristics and Study-specific Effect Estimates Included in the Meta-analyses and Sensitivity Analyses; excerpt below) to clarify study details including selected effect estimates used from each study (i.e., means, standard deviations, sample sizes, regression slopes with 95% confidence intervals, and

exposure levels). The source of the results (e.g., table, figure) from each study publication is also listed. eTable 1 (excerpt below) provides details on excluded studies and studies with overlapping populations.

Excerpt of eTable 2. Study Characteristics and Study-specific Effect Estimates Included in the Meta-analyses and Sensitivity Analyses

Reference[a] *Study Design*	Study Location	Age Range (Years)	Assessment (Metric, Exposure Groups)	Fluoride Exposure Levels	Mean-effects Meta-analysis N, Mean (SD) [Reference] [Exposed]	Dose-response Mean-effects Meta-analysis N, Mean (SD) [Reference] [Exposed]	Regression Slopes Meta-analysis Slope (SE) or 95% CI per Unit Change Fluoride	Source
Bashash et al. (2017)[112,me, u, rs] *Prospective Cohort*	Mexico	6–12	Maternal urine Reference/high fluoride levels (based on children urinary fluoride)	<0.80 mg/L (reference) ≥0.80 mg/L (high)	77, 95.37 (10.31) 112, 96.80 (11.16)	77, 95.37 (10.31) 112, 96.80 (11.16)	−2.50 (−4.12, −0.59) per 0.5 mg/L maternal urinary F	Abstract, Table 3
Razdan et al. (2017)[73,sa] *Cross-sectional*	India	12–14	Drinking water Low/high fluoride levels	0.6 ppm (low) 4.99 ppm (high)	69, 38.61 (6.34) 75, 13.95 (5.14)			Table 2
Valdez Jiménez et al. (2017)[74sa] *Prospective Cohort*	Mexico	Infancy	Maternal urine, drinking water	Urine: 1.9 ± 1.0 mg/L (1st trimester) 2.0 ± 1.1 mg/L (2nd trimester) 2.7 ± 1.1 mg/L (3rd trimester) Water: 2.6 ± 1.1 mg/L (1st trimester) 3.1 ± 1.1 mg/L (2nd trimester) 3.7 ± 1.0 mg/L (3rd trimester)			Bayley MDI: −19.05 (8.9) per 1 log10 mg/L maternal urinary F (1st trimester) −19.34 (7.46) per 1 log10 mg/L maternal urinary F (2nd trimester)	Table 2, Table 4
Cui et al. (2018)[76,rs] *Cross-sectional*	China	7–12	Urine	Boys: 1.3 (0.9–1.7)[d] mg/L Girls: 1.2 (0.9–1.6)[d] mg/L			−2.47 (−4.93, −0.01) per 1 log urinary F	Table 2
Yu et al. (2018)[3,me, w, u*, rs] *Cross-sectional*	China	7–13	Maternal urine Low/medium/high fluoride ranges Drinking water Normal/high fluoride	Urine: 0.01–1.60 mg/L (low) 1.60–2.50 mg/L (medium) 2.50–5.54 mg/L (high) Water: ≤1 mg/L (normal) >1 mg/L (high) Overall: 0.01–5.54 mg/L (urine) 0.20–3.90 mg/L (water)	1636, 107.40 (13.00) 1250, 106.40 (12.30)	1636, 107.40 (13.00) 1250, 106.40 (12.30)	0.36 (−0.29, 1.01) per 0.5 mg/L maternal urinary F	Table 1, Table 3

Excerpt of eTable 1. List of Excluded Studies from Mean-effects Meta-analysis

Reference, Country	Reason for Exclusion
Qin et al. (1990)[45] [translated in Qin et al. 2008], China	Missing mean or SD of outcome measure
Yang et al. (1994)[47] [translated in Yang et al. 2008], China	Overlapping population with Wang et al. (2001)[49]; Table 2 in Yang et al. (1994)[47] seemed incomplete
Wang et al. (2005b)[50] [translated in Wang et al. 2008a], China	Missing mean or SD of outcome measure
Rocha-Amador et al. (2007)[52], Mexico	Missing mean or SD of outcome measure
Liu et al. (2000)[53] [translated in Liu et al. 2008] , China	Overlapping population with Lu et al. (2000)[55]
Sudhir et al. (2009)[56], India	Missing mean or SD of outcome measure
He and Zhang (2010)[57], China	Missing mean or SD of outcome measure
Xiang et al. (2011)[58], China	Overlapping population with Xiang et al. (2003a)[59]
Saxena et al. (2012)[60], India	Missing mean or SD of outcome measure
Wang et al. (2012)[61], China	Overlapping population with Xiang et al. (2003a)[59]

5.F:

b. Conduct additional sub-group analyses (study design, attention to concerns about blinding, complex sampling designs, and statistical analyses that account for clustered sampling designs)

Response: Agree (change made)

- We have performed several additional sensitivity analyses to address the NASEM Committee's comments on blinding, complex sampling designs, and clustering. The results are presented in eTable 6 (excerpt

below). One analysis excluded Cui et al. (2018) to respond to the Committee's concerns about blinding. To address the NASEM Committee's comments about complex sampling designs, we conducted a sensitivity analysis excluding Yu et al. (2018) and Zhang et al. (2015b). To address the Committee's concerns about clustering, we performed two sensitivity analyses—one using the unadjusted effect estimates and one using the estimates from the random effect models from Bashash et al. (2017) and Green et al. (2019). The additional sensitivity analyses had minimal impact on the overall results of the meta-analysis (see excerpt of eTable 6 below).

Excerpt of eTable 6. Regression Slopes Meta-analysis

Analysis	Number of Studies	Beta (95% CI)	Heterogeneity	
			p-value	I^2
Overall Estimate				
Full-scale IQ	9	−1.81 (−2.80, −0.81)	<0.001	77%
Sensitivity Analyses				
Using estimates from random effect models for Green et al. (2019)[113] and Bashash et al. (2017)[112]				
Full-scale IQ	9	−1.80 (−2.80, −0.80)	<0.001	76%
Males	2	−2.39 (−5.89, 1.10)	0.070	69%
Females	2	−0.53 (−3.43, 2.37)	0.186	43%
Excluding Yu et al. (2018)[3] and Zhang et al. (2015b)[110]				
Full-scale IQ	7	−1.76 (−2.90, −0.62)	<0.001	82%
Using unadjusted estimates from Bashash et al. (2017),[112] Cui et al. (2018),[76] Green et al. (2019)[113], Yu et al. (2018)[3]				
Full-scale IQ	9	−1.82 (−2.81, −0.83)	<0.001	76%

5.G:

c. Revisit the inclusion of data from overlapping studies

Response: Agree (change made)

- The NASEM Committee identified one set of overlapping populations—Xiang et al. (2003) and Xiang et al. (2011)—and suggested review of all of the analyses to ensure that overlapping publications are not included in any meta-analyses. We have removed the Xiang et al. (2011) assessment of IQ associated with serum fluoride levels from the meta-analyses. We have also confirmed that there are no overlapping publications used in the same meta-analysis. As stated previously, eTable 1 (excerpt above) provides details on studies with overlapping populations.

5.H:

d. Describe the meta-analysis methods in a single location for ease of reading

Response: Agree (change made)

- The separation of the meta-analysis from the NTP Monograph supports greater clarity in the presentation of methods for the meta-analysis versus the overall systematic review methods for the NTP Monograph. The peer-reviewed protocol contains the complete methodological details for the meta-analysis in one location. The *Methods* section of the meta-analysis manuscript also has improved clarity as it is now solely focused on the meta-analysis.

5.I:

e. Acknowledge weaknesses in the subjective way publication bias was assessed

Response: Agree (change made)

- We agree with the NASEM Committee's overall comment that, "NTP provides a reasonably thorough and appropriate evaluation of publication bias." The NASEM Committee recommended NTP consider "adjusting for possible publication bias" rather than "eliminating publication bias" when referring to results of fill-and-trim analyses. We accepted the recommendation, addressed Committee comments, and to provide additional clarity, we have added a brief discussion of the existing approaches for evaluating potential for publication bias to the *Methods* section of the meta-analysis manuscript, as follows:

  "Potential publication bias was assessed by developing funnel plots and performing Egger regression on the estimates of effect size.[25-27] If publication bias was present, trim-and-fill methods[28, 29] were used to estimate the number of missing studies and to predict the impact of the hypothetical "missing" studies on the pooled effect estimate."

- We agree that the limitations of the tests used to evaluate publication bias should be mentioned, and we have added the following to the *Discussion* section as follows:

  "There are also several limitations to the existing approaches for evaluating potential for publication bias. The funnel plot asymmetry is a subjective assessment and is recommended only when at least 10 studies are included in the meta-analysis.[64] Furthermore, the Egger regression test and Begg's rank tests[25-27] may suffer from inflated type I power and limited power in certain situations.[65]"

5.J:

f. Assess heterogeneity multiple ways

Response: Agree (change made)

- The NASEM Committee had several comments on heterogeneity and noted that NTP primarily used the Cochran's Q test to assess

heterogeneity. The Committee did not suggest assessing heterogeneity in multiple ways but noted that "heterogeneity can also be assessed by providing a count or percentage of the number of studies to the right or left of the null value. Some would consider that a much simpler, more intuitive, and perhaps more useful way of assessing heterogeneity, especially in light of the marked differences between the studies in design, study populations, exposure and outcome assessment methods, and statistical analyses. Although that approach should not be used as the sole basis of conclusions, it can be a useful first step in exploring why heterogeneity might exist."

- The meta-analysis manuscript now includes clear references to the studies with effect estimates to the right of the null in the *Results* section of the manuscript as follows:

  "The three studies with a non-negative association reported SMD estimates of 0.01 (95% CI: −0.19, 0.21),[6] 0.01 (95% CI: −0.19, 0.22),[38] and 0.13 (95% CI: −0.16, 0.42).[5]"

- In the *Methods* section, we provide details on how heterogeneity was assessed as follows:

  "Heterogeneity was assessed by Cochran's Q test[23] and the I2 statistic.[24] Forest plots were used to display results and to examine possible heterogeneity between studies."

5.K:

g. Provide the rationale for selecting individual outcomes from a single study when multiple outcomes were present

Response: Agree (change made)

- We reviewed the analyses to ensure that a consistent approach matching the data criteria outlined in the meta-analysis protocol was applied to all studies. Results were selected considering the most appropriate exposure metric, exposure range, exposure period, number of subjects, and statistical adjustment for potential confounders. See excerpt of eTable 2 referenced in our response to comment "a" above for study-specific effect estimates used in the meta-analysis.

5.L:

h. Revisit decisions made to exclude particular study results

Response: Agree (change made)

- The NASEM Committee recommended that NTP review the process to exclude study results from the meta-analysis. In response, we reviewed the analyses to ensure that a consistent approach matching the data criteria outlined in the meta-analysis protocol was applied to all studies.

For reasons why particular outcomes/results were selected, see our responses to comments "a" and "g" above.

- We also revised the meta-analysis to include standardized mean differences (SMDs) from Green et al. (2019). We agree with the Committee that Ding et al. (2011) and Zhang et al. (2015) were correctly included in both the *mean-effects* and *regression slopes meta-analyses*.

5.M: Issue: New evidence

Two studies (Ibarluzea et al., 2021 and Aggeborn & Ohman, 2021) published in 2021 were not included in the systematic review or meta-analysis. These studies have comparable methods to other included studies.

Recommendation: The Ibarluzea and Aggeborn & Oehman studies should be evaluated and included when assessing the evidence, similar to the 15 additional studies from the Chinese databases. [redacted] also recommends NTP include a comparison between Ibarluzea et al., 2021, and Green et al., 2019, because both studies investigate fluoride exposures at levels used for water fluoridation.

Response: Agree (change made)

- We have updated the literature search for the meta-analysis through November 1, 2021, using appropriate methods to identify critically assessed and relevant new publications. After integrating the results, the conclusions of the meta-analysis were essentially unchanged.
- In updating the literature search for the meta-analysis, 10 new studies (including Ibarluzea et al. 2021) were added to the evidence database. These new studies were published in the past 2 years and their addition left the findings of the analysis essentially unchanged. Our meta-analysis now includes 60 studies of children's cognition and fluoride exposure, 13 of which are high quality.
- Aggeborn and Ohman (2021) had been previously reviewed when it was a 2017 non-peer-reviewed white paper but was excluded because it was not peer-reviewed. The study was excluded from the meta-analysis because it assessed cognitive functions other than IQ and the cognitive tests were not specified (see supplemental materials, eTable 1).

This document contains the complete first set of comments provided by the [redacted] in January 2022 in its original format and the NIEHS/DNTP responses to those comments. Note that the yellow highlighting as well as the purple and red formatted text were in the document as provided. The NIEHS/DNTP responses have been inserted in blue text following each of the comments beginning with the word "**Response**" in bold font. Formatting has been applied to aid in reading.

6a.A: From Abstract

RESULTS The meta-analysis of 46 studies (N = 15,538 children) with group-level exposures found that children exposed to higher fluoride levels had lower mean IQ scores (pooled SMD: −0.49; 95% CI: −0.60, −0.38; p-value < 0.001). Results were robust to stratification by study quality (high *vs*. low risk of bias), gender, age group, outcome assessment, study location, exposure timing, and exposure metric. There was a dose-response relationship between mean children's IQ and group-level fluoride exposure measures. [1] The meta-analysis of the association between individual-level measures of fluoride and children's IQ found a decrease of 1.58 IQ points (95% CI: −2.63, −0.53; p-value = 0.003) per 1-mg/L increase in urinary fluoride. **CONCLUSIONS AND RELEVANCE** Our meta-analysis confirms results of previous meta-analyses and extends them by including newer, more precise studies with individual-level exposure measures. The data support a consistent inverse association between fluoride exposure and children's IQ.

[1] This dose-response statement is not consistent with the level of fidelity of the data presented/available and infers there are negative health effects attributable to fluoride. This is a critical concern that applies to the highlighted statements below.

Response: Disagree (no change)

- The highlighted text accurately describes the available data, analysis, and results. The language in the abstract and throughout the manuscript objectively and fairly describes the data, including strengths and limitations.
- In its 2021 report on the 2020 draft NTP Monograph, the NASEM Committee agreed with our statements on consistency: "As noted in the revised monograph, 44 of the 46 studies represented in that figure had effect estimates to the left of zero—results that indicate an association between higher fluoride exposures and lower IQ. Those results highlight the marked consistency in the current epidemiologic literature on fluoride and childhood IQ."
- Please note that the last sentence that was highlighted was subsequently changed as follows:

  "The consistency of the data supports an inverse association between fluoride exposure and children's IQ."

6a.B: FROM Manuscript

No study was excluded from the meta-analysis based on concerns for risk of bias; however, subgroup analyses were conducted with and without high risk-of-bias studies (i.e., studies rated "probably high" risk of bias for at least two key risk-of-bias questions or "definitely high" for any single key question) to assess their impact on the results.

Response: No change requested

- This text was highlighted but was not accompanied by a comment or request for revision. We assume that the text was highlighted to imply that this approach is a flaw. Excluding studies from systematic reviews or systematic reviews with meta-analyses is not considered a best practice in the systematic review community (Higgins et al. 2021). As a well-documented systematic review and meta-analysis, this evaluation follows a protocol where inclusion and exclusion criteria were defined a priori. As the non-highlighted text above clearly states, a subgroup analysis was conducted with and without the high risk-of-bias studies.

  Reference: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). 2021. Cochrane Handbook for Systematic Reviews of Interventions version 6.2 (updated February 2021). Cochrane. Available from www.training.cochrane.org/handbook

6a.C: Conclusions

Our meta-analysis confirms and extends prior meta-analyses that reported associations between higher fluoride exposures and lower IQ levels of children. The results were robust to stratifications by risk of bias, gender, age group, outcome assessment, study location, exposure timing, and exposure type (including both drinking water and urinary fluoride). Therefore, the data support a consistent inverse association between fluoride exposure and children's IQ.

Response: No change requested

- This text was highlighted but was not accompanied by a comment or request for revision. We are unaware of the reviewer's thoughts on this highlighted text but will note that the sentences in the conclusion are factual statements describing the data.

6a.D: From Supplemental Documentation

If median or mean levels by exposure group were not provided, the midpoint of the upper and lower boundaries in every exposure category was assigned as the average level. If the upper boundary for the highest exposure group was not reported, the boundary was assumed to have the same amplitude as the nearest exposure category

Response: No change requested

- Again, these sentences were highlighted but were not accompanied by a comment or request for revision. We are unaware of the reviewer's thoughts on this highlighted text but will note that this method is common practice in dose-response analyses in determining exposure levels for each data point (Boffetta et al. 2020) and is described in our peer-reviewed protocol.

Reference: Boffetta, P., Zunarelli, C., & Borron, C. (2020). Dose-Response Analysis of Exposure to Arsenic in Drinking Water and Risk of Skin Lesions: A Systematic Review of the Literature. Dose-response: a publication of International Hormesis Society, 18(4), 1559325820957823. https://doi.org/10.1177/1559325820957823

From NTP 2020 revision NTP Protocol: Systematic Review of Effects of Fluoride Exposure on Neurodevelopment (nih.gov)

Note: A comment related to the protocol for the NTP Monograph (see https://ntp.niehs.nih.gov/go/785076) is not reproduced here as it is not directly relevant to the meta-analysis.

6a.E: Additional concerns

- Measure assessment of "intelligence" was different in different studies (the scores/scales for different countries, different tools and the interpretation of the "mean" of disparate classification systems. Examples: Wechsler Abbreviated Scale of Intelligence vs. "Wechsler Intelligence Scale for Children-Revised" vs. "Combined Raven's Test for Rural China" or "Wechsler Primary and Preschool Scale of Intelligence-III"

  Response: Disagree (no change)
  - We view the use of these different tests in studies of different study populations as the proper approach and consider whether the test is appropriate for a given population as part of risk-of-bias assessment. As per our protocol, for a "definitely" or "probably low risk-of-bias" rating for outcome assessment, it is required that studies use an intelligence test that is appropriate to the population being studied. The consistency of the direction of the association across a diverse range of tests supports the conclusions of our meta-analysis.
  - The difference in tests is also a reason we used the standardized mean difference (SMD) as the unit of measure in our meta-analysis. The SMD is commonly used in meta-analysis when the studies all assess the same outcome (e.g., intelligence) but measure it in a variety of ways (e.g., WISC-R, Combined Raven's Test for Rural China, etc.). It is necessary to standardize the results of the studies to a uniform scale before they can be combined (Higgins et al. 2021).
  - In addition, this comment fails to acknowledge that we also conducted a subgroup analysis stratified by assessment type. The results of this subgroup analysis show that the direction of the association is robust to stratification by assessment type and that assessment type does not explain the observed heterogeneity. The results of this subgroup analysis compared to the main overall effect estimate are shown below.

    Excerpt of Table 2. Pooled SMDs and 95% CIs for Children's IQ Score and Exposures to Fluoride

Analysis	Number of Studies	SMD (95% CI)	Heterogeneity	
			p-value	I^2
Overall Effect				
Overall Effect	55	−0.46 (−0.55, −0.37)	<0.001	87%
Subgroup Analyses				
Assessment Type				
CRT-RC tests	29	−0.36 (−0.46, −0.27)	<0.001	82%
Non-CRT-RC tests	26	−0.60 (−0.78, −0.42)	<0.001	89%
Raven's tests	10	−0.76 (−1.10, −0.43)	<0.001	91%
Other tests	16	−0.52 (−0.74, −0.29)	<0.001	89%

Table 2 Notes: CI = confidence interval; CRT-RC = Combined Raven's Test–The Rural edition in China; NA = not applicable; SMD = standardized weighted mean difference

Reference: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). 2021. Cochrane Handbook for Systematic Reviews of Interventions version 6.2 (updated February 2021). Cochrane. Available from www.training.cochrane.org/handbook

6a.F:

- Definition of "high"/"low" fluoride levels were different across studies and not defined by the author. One newer study Bashash et al. (2017) defined "Low" as <0.80 mg/L and "High" ≥0.80 mg/L but without upper limit and the difference between Low/High in this example could be as small as 1/100th

Response: Disagree (no change)

- It would be inappropriate for us to define high and low fluoride levels for the purpose of this meta-analysis**.** Our approach is consistent with all previous fluoride meta-analyses (Choi et al. 2015, Duan et al. 2018, Miranda et al. 2021). Table 1 transparently reports the high and low fluoride levels as presented in each individual study.
- In its peer review of the first draft of the meta-analysis, which was included in the 2020 draft NTP Monograph, the NASEM Committee agreed with this method for the *mean-effects meta-analysis*: "The overall approach appears to be sound in comparing mean IQ scores for the most and least highly exposed to fluoride even though the absolute fluoride concentrations are not comparable among studies.", and "The committee found the meta-analysis to be a valuable addition to the monograph and acknowledges the tremendous amount of work that was required. The meta-analysis applied standard, broadly accepted methods, and the data shown in Figure A5-1 and the related evaluations are especially informative (NTP 2020a, p. 235)."
- In addition, this comment fails to acknowledge that we also conducted a *regression slopes meta-analysis* that used studies reporting continuous data estimating associations between individual-level fluoride exposure and children's

IQ. In this analysis, differences across studies with respect to what study authors might consider high or low fluoride levels are irrelevant.

6a.G:

- Quantified dose of exposure not presented. Urinary spot testing not good surrogate and no correlation to quantifiable exposure given relatively rapid clearance of Fl from the body.

  Response: Disagree (no change)

  - We understand the concerns regarding urinary fluoride levels. However, fluoride levels measured during pregnancy and in children include all ingested fluoride and are considered a valid measure to estimate total fluoride exposure (Villa *et al.* 2010, Watanabe *et al.* 1995).
  - We acknowledge that the type and timing of urinary sample collection is important to consider, and we have considered these factors in our analysis as described in the prepublication 2022 NTP Monograph. When compared to 24-hour urine samples, spot urine samples are more prone to the influence of timing of exposure and can also be affected by differences in dilution; however, many studies attempted to account for dilution using either urinary creatinine or specific gravity. Good correlations between 24-hour samples and urinary fluoride concentrations from spot samples adjusted for urinary dilution have been described in the literature (e.g., Zohouri et al. 2006). Both 24-hour samples and spot urine samples adjusted for dilution are considered acceptable, with 24-hour samples considered the more accurate measure of fluoride. If authors made appropriate efforts to reduce the concern for bias (e.g., accounting for dilution), studies that used this metric were generally considered to have probably low risk of bias for exposure.
  - However, we have added the following sentence to the *Strengths and Limitations* section of the meta-analysis to acknowledge this concern:

    "When compared with 24-hour urine samples, spot urine samples are more prone to the influence of timing of exposure (e.g., when water was last consumed, when teeth were last brushed) and can also be affected by differences in dilution."

6a.H:

- The lack of a direct measure of dose and thus, exposure is a significant design limitation and the strength and specificity of the conclusions are out of proportion given the limitations; the statements/conclusions of the manuscript overstate what can be fairly concluded from the studies.

  Response: Disagree (no change)

  - The conclusions of our meta-analysis are consistent with two prior meta-analyses of studies using group-level exposures and extend these analyses with a

confirmatory *regression slopes meta-analysis* that uses individual-level exposure and outcome assessments. In the *Discussion* section, we clearly address the limitations of a *mean-effects meta-analysis,* including the way in which exposure is measured.

6a.I:

- The results could be used to recommend improvements to future studies but the lack of an individual fluoride exposure variable and dose measurement precludes the conclusions asserted in this paper. This weakness could be responsible for complete misclassification of many of the data points.

  Response: Disagree (no change)

  - This meta-analysis does not lack individual fluoride exposure variables. Our *regression slopes meta-analysis* includes 11 studies with individual-level exposure measures (with 10 high quality publications) from 6 different study populations. Each of these studies reported individual urinary fluoride levels, with two also reporting fluoride intake and two also reporting water fluoride levels.
  - As we mentioned in our response to a previous comment, urinary fluoride in children is a valid measure to estimate total fluoride exposure. In addition, the consistency of the results from the *regression slopes meta-analysis* stratified by exposure type (Table 3 excerpt provided below) suggest that the results cannot be explained by a "complete misclassification of many of the data points."

    Excerpt of Table 1. Pooled Regression Slopes and 95% CIs for Children's IQ Score and Exposures to Fluoride

Analysis	Number of Studies	Beta (95% CI)	Heterogeneity	
			p-value	I^2
Overall Effect				
Full-scale IQ	9	−1.81 (−2.80, −0.81)	<0.001	77%
Subgroup Analyses				
Exposure Type				
Urinary fluoride	9	−1.81 (−2.80, −0.81)	<0.001	77%
Intake	2	−3.87 (−7.15, −0.59)	0.737	0%
Water fluoride	2	−4.77 (−9.09, −0.45)	0.707	0%

6a.J:

- The strength and specificity of the conclusions are out of proportion and overstated given the significant limitations of the available data from these studies

  Response: Disagree (no change)

  - The statements made in the meta-analysis are measured and representative of the data.

6a.K: Additional Background:

From Dose Response Assessment - an overview | ScienceDirect Topics,

Dose–Response Assessment
Dose–response assessment characterizes the quantitative relationship between exposure (usually determined in toxicity studies) and the occurrence of adverse health effects. Typically applied or administered dose, rather than effective tissue dose, is used to develop the dose–response relationship.

- These are important points that support the premise that there really is no measure or attempt to measure "dose" of exposure. An environmental, naturally occurring metal (Fl^-) merely being in the environment does not constitute an exposure of any particular magnitude. This is missing.

  Response: Disagree (no change)
  - We are somewhat unclear on the points being raised. Concerning a "typical" dose- -response relationship, the comment above is correct that most dose-response relationships are based on estimates of applied or administered dose; however, this is also commonly considered a practical limitation of the method. We explain in the manuscript that drinking water measures are indirect measures of exposure and that internal measures such as those reflected by urinary fluoride data are preferred. We also disagree that fluoride is classified as a metal.

6a.L: Also, [redacted] read of eTable 4. Dose-Response Meta-analysis Using Mean Effects – Model Selection for Water Fluoride does *not* represent dose response, as [redacted] see it. For example: Linear Model: the P value for <2 mg/L and <1.5 mg/L are not significant and it makes sense then, that if the "All data" p= <0.001 is disproportionately influenced by the <4 mg/L exposure. It is also not clear from this table whether these numbers represent <4 but >= 2mg/L and <2 but >=1.5, etc. Are these mutually exclusive categories? Needs clarification

Response: Disagree (edited for clarity)
- We disagree that statistical significance is necessary to indicate a dose-response relationship. Data should be evaluated in their full context for epidemiological studies, and statistical significance is only one consideration (EPA 2020). We report p-values and consider them as an important, but not exclusive contribution to the overall data interpretation.
- However, we have taken the suggestion to clarify that the exposure categories are not mutually exclusive and have added the range of exposure for each group when they are first mentioned in the supplemental materials as follows:

  <4 mg/L (i.e., 0 to <4 mg/L) fluoride in drinking water

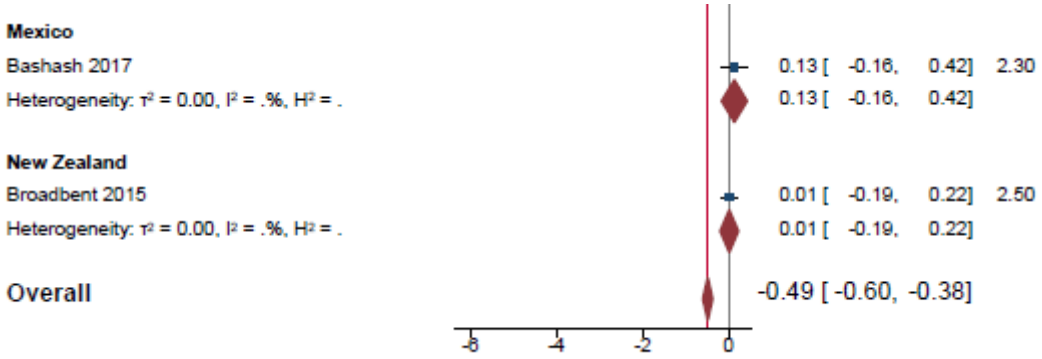
<2 mg/L (i.e., 0 to <2 mg/L) fluoride in drinking water

<1.5 mg/L (i.e., 0 to <1.5 mg/L) fluoride in drinking water

Reference: U.S. EPA. ORD Staff Handbook for Developing IRIS Assessments (Public Comment Draft, Nov 2020). U.S. EPA Office of Research and Development, Washington, DC, EPA/600/R-20/137, 2020.

6a.M: Children's Urinary Fluoride – All Studies

Also, worth noting: the newer studies included in the analysis… Mexico and New Zealand Country subgroup analysis are both at **Zero** or **above zero**, eFigure 13. Association Between Fluoride Exposure and IQ Scores in Children: Effect by Age Group



Response: No change requested

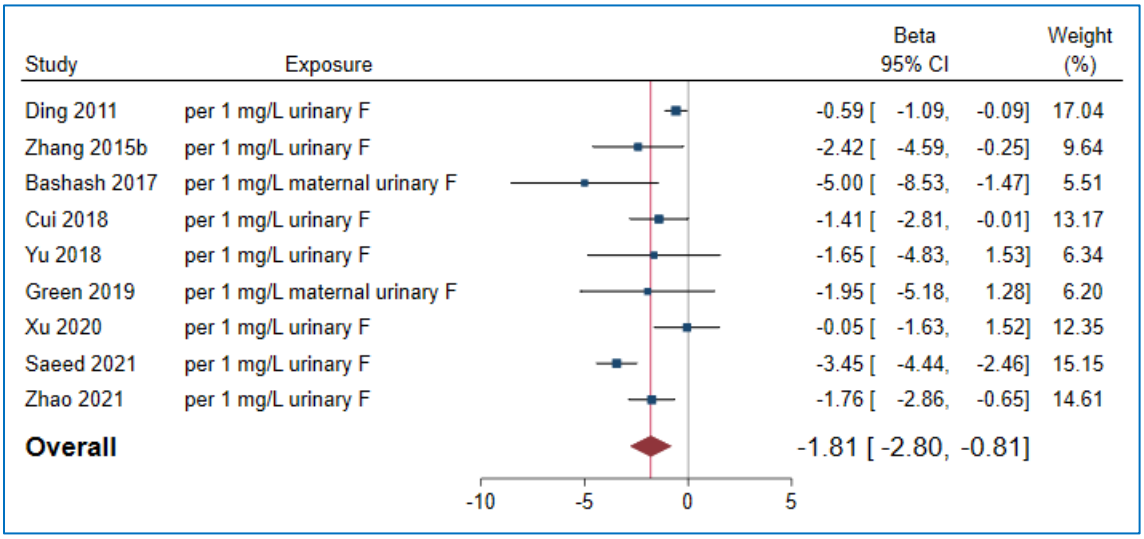
- This is an excellent example of DNTP considering all data irrespective of direction of effect. In fact, we point out these non-negative effect estimates in the *Results* section of the supplemental materials:

  "The three studies with non-negative associations reported SMD estimates of 0.01 (95% CI: −0.19, 0.21),[113] 0.01 (95% CI: −0.19, 0.22),[25] and 0.13 (95% CI: −0.16, 0.42).[112] Two of the three studies with non-negative SMDs compare mean IQs in children living in fluoridated vs. non-fluoridated areas in Canada,[113] or in New Zealand.[25] No other studies included in the main mean-effects meta-analysis made comparisons between fluoridated vs. non-fluoridated areas. In both studies, levels of fluoride in water were low, even in communities with fluoridated drinking water, likely limiting the power to detect an effect.

  In Bashash et al.,[112] the SMD compares mean IQ scores in children with urinary fluoride levels below vs. above 0.80 mg/L in Mexico.[112] Unlike other studies in the mean-effects meta-analysis which compared mean IQ scores between fluoridated vs. non-fluoridated areas, or areas with high vs. low fluoride exposures (see eTable 2), the Bashash et al.[112] study was not designed to measure fluoride exposure by geographical area. However, since the mean IQ scores were provided in the manuscript for children with urinary fluoride levels below vs. above 0.80 mg/L, we included them in this analysis. It's worth

noting that there was no significant difference when comparing MUF levels between the groups of children with urinary fluoride levels above or below 0.80 mg/L, however when children's IQs were regressed against MUF, a statistically significant inverse association was found."

- Note: In our November 2021 update of the literature, we also included the INMA cohort study (Ibarluzea et al., 2021) that found positive associations between fluoride exposure and cognitive effects in boys.



eFigure 19. Association Between Individual-level Fluoride Exposure and IQ Scores in Children: Overall Analysis

eFigure 19 note: Estimates (betas) for individual studies are shown with solid boxes representing the weight, and the pooled estimate is shown as a solid diamond. Horizontal lines represent 95% CIs for the study-specific betas.

This document contains the complete second set of comments provided by [redacted] in February 2022 in its original format and the NIEHS/DNTP responses to those comments. The NIEHS/DNTP responses have been inserted in blue text following each of the comments beginning with the word "**Response**" in bold font. Formatting has been applied to aid in reading.

6b.A: [redacted] **Critique of the NTP meta-analysis manuscript**

Summary:

The group of studies included in this meta-analysis had three significant issues identified by the manuscript authors that weaken its results: publication bias, high heterogeneity, and lack of uniformity in measuring and reporting the primary studies' outcome or IQ measure. Overshadowing these problems is the inappropriate use of a meta-analysis for observational studies when randomized clinical trials are not available. The intervention and control arms among a group of similar randomized trials are comparable because randomization tends to balance the arms with respect to both known and unknown confounders, but this is not true of observational studies. In the present meta-analysis, those categorized as consuming higher levels of fluoride are compared to those consuming lower levels. The fluoride exposure is not randomized and may be dictated by national or regional governments. Two potential consequences are spurious associations between fluoride and IQ and differential results by country. [redacted] begin with these two consequences.

Response: Disagree (no change)

- First, it is important to make clear that this meta-analysis was conducted at the strong recommendation of the NASEM Committee's peer review of the 2019 draft NTP Monograph. The 2019 draft NTP Monograph evaluated a large number of human observational studies but did not include a meta-analysis. The NASEM Committee's peer review report stated that the "committee strongly recommends that NTP reconsider its decision not to perform a meta-analysis."
- Second, the NASEM Committee agreed with the methods used in the meta-analysis. In its peer review of the first draft of the meta-analysis, which was included in the 2020 draft NTP Monograph, the NASEM Committee stated: "The critical information regarding comparison of study results comes from the new meta-analysis, which seeks to extract and integrate comparable findings from selected studies as discussed further below. The overall approach appears to be sound in comparing mean IQ scores for the most and least highly exposed to fluoride even though the absolute fluoride concentrations are not comparable among studies.", and "The committee found the meta-analysis to be a valuable addition to the monograph and acknowledges the tremendous amount of work that was required. The meta-analysis applied standard, broadly accepted methods, and the data shown in Figure A5-1 and the related evaluations are especially informative (NTP 2020a, p. 235)."
- Finally, consideration of the use and value of data from observational studies relative to randomized clinical trials (RCTs) in meta-analyses has been empirically studied, and Cochrane analyses have repeatedly shown that there is little evidence for significant effect estimate differences between observational studies and RCTs regardless of specific observational study design, heterogeneity, or inclusion of studies of pharmacological interventions (Anglemyer et al. 2014; Benson and Hartz 2000; Schwingshackl et al. 2021).

6b.B: In the present meta-analysis, the majority of studies are from one country: China. There are plausible mechanisms that might create the appearance of an association between fluoridation and IQ scores. For example, people in rural communities may exhibit lower IQ test scores than those in urban areas; they may also be more likely to drink tap water or well water as opposed to bottled water or other beverages. Thus, they might consume more fluoride. This would induce a non-causal correlation between fluoridation and IQ scores. [redacted] are presenting this scenario not as a fact, but to suggest that there are plausible explanations for a spurious correlation between fluoride and IQs in observational studies.

Response: Disagree (no change)

- The point of a risk-of-bias assessment is to evaluate whether the design or conduct of a study compromised the credibility of the link between exposure and outcome (Higgins and Green 2011, IOM 2011, Viswanathan et al. 2012). The concern in this comment appears to be related to potential bias due to confounding in individual studies. This issue is addressed in the meta-analysis through a rigorous assessment of risk of bias, which included an extensive evaluation of potential bias due to confounding in each individual study, addressing situations exactly like the example presented in the comment. (See eFigure 2a and 2b for risk-of-bias summaries, links to assessments of individual studies, and Appendix E of prepublication 2022 NTP Monograph for more detail.)
- We would also like to note that Chinese studies provide the opportunity to compare the cognitive abilities of children in villages of similar size, SES, and other relevant characteristics where drinking water sources differ widely in their level of naturally occurring fluoride. These variations in fluoride levels can be larger than those found in the other areas, including most of the United States, and therefore provide greater power to detect an effect.

6b.C: Given that most of the studies in this meta-analysis are in China, whose environmental policies could explain a spurious association, [redacted] might expect to see different results in countries with policies more aligned with those of western nations. In fact, that is exactly what [redacted] see in this meta-analysis. Broadbent (2015) and Green (2019) are studies in New Zealand and Canada, arguably the two countries most comparable to the United States. Figure 2 of the manuscript shows narrow confidence intervals centered on no effect in these two studies. This is consistent with the idea that the apparent association between fluoride and IQs may not be causal.

Response: Disagree (no change)

- It's not clear what the comments refer to when citing environmental policies that would explain a spurious effect. However, a spurious association is unlikely given the included studies span broad geographical regions and time periods (1989–2021) and cover a range of methods for outcome and exposure assessment (including different exposure metrics). In addition, potential confounders and co-exposures to other possible neurotoxicants were extensively considered in the risk-of-bias assessment and evaluation of each study.
- [redacted] is correct that, in the *mean-effects meta-analysis*, the SMDs for children's urinary fluoride (CUF) and children's IQ in Broadbent et al. (2015) and Green et al. (2019) were non-

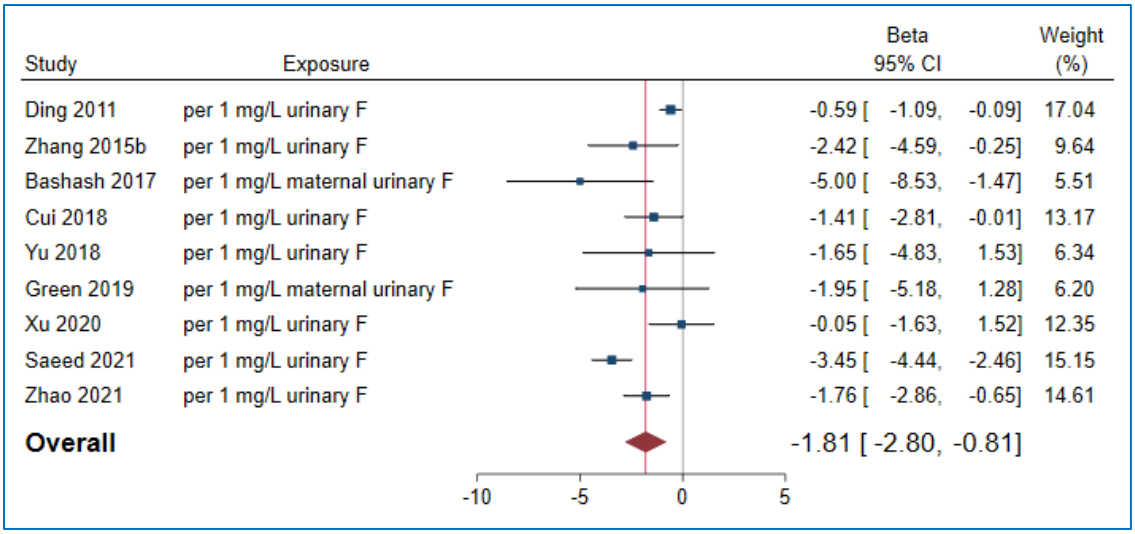
negative; the SMD for Bashash et al. was also non-negative. We clearly describe these non-negative effect estimates in the *Results* section of the manuscript:

"The three studies with a non-negative association reported SMD estimates of 0.01 (95% CI: −0.19, 0.21),[6] 0.01 (95% CI: −0.19, 0.22),[38] and 0.13 (95% CI: −0.16, 0.42).[5]"

- In both Broadbent et al. (2015) and Green et al. (2019), levels of fluoride in water were low, even in communities with fluoridated drinking water. So, when using group-level exposure data (as opposed to individual-level exposure data), as was done in the *mean-effects meta-analysis*, the power to detect an effect may be limited. We note that [redacted] comment ignores the results of the *regression slopes meta-analysis*, which used individual-level maternal urinary fluoride (MUF) for the Canadian (Green et al. 2019) and Mexican (Bashash et al. 2017) studies (MUF levels were comparable in these two studies [Till et al. 2018]) and found an inverse association between MUF and children's IQ as shown in eFigure 19 (provide below; see Bashash et al. 2017 and Green et al. 2019). Green et al. (2019) also reported a statistically significant inverse association between maternal water fluoride levels and children's IQ as shown in eFigure 23 (provided below).
- In response to a comment from the NASEM Committee, we added text to the supplemental materials to identify likely reasons why results from the three studies differed from results of the other studies, as follows:

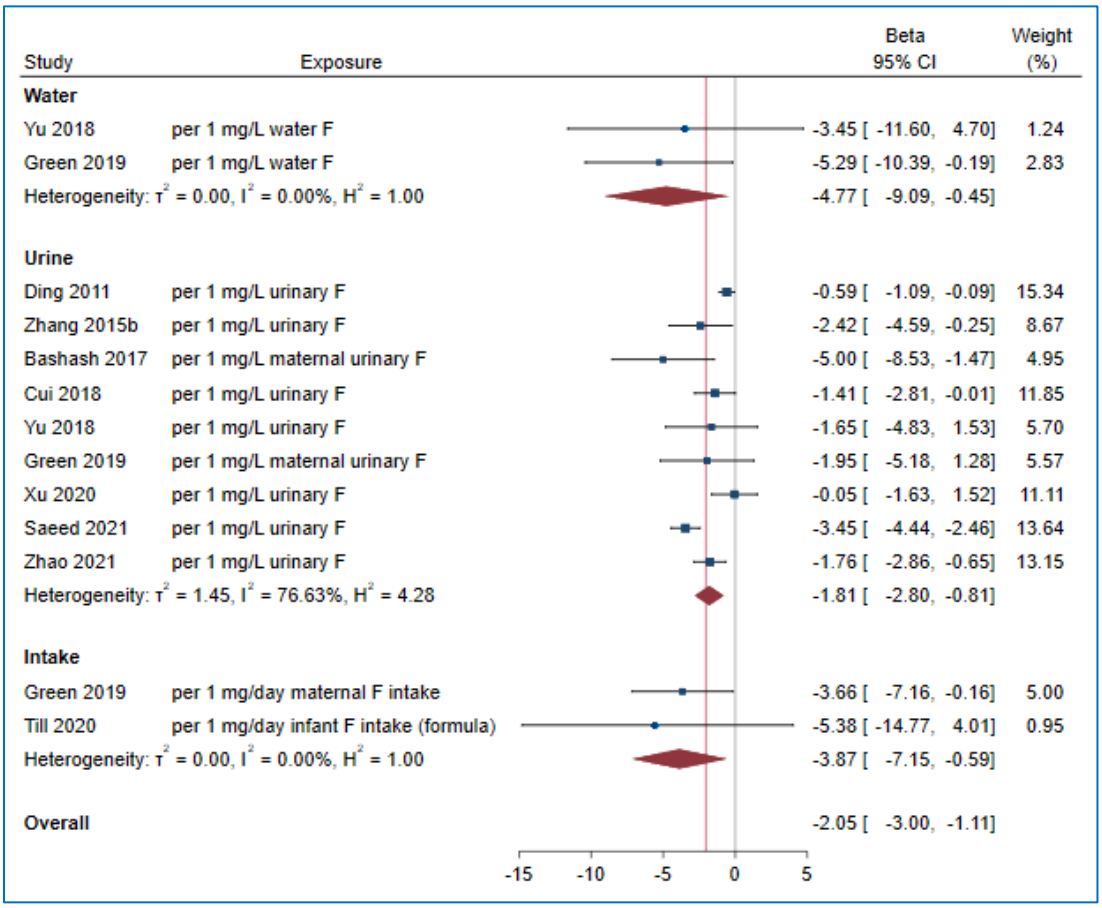
  "The three studies with non-negative associations reported SMD estimates of 0.01 (95% CI: −0.19, 0.21),[113] 0.01 (95% CI: −0.19, 0.22),[25] and 0.13 (95% CI: −0.16, 0.42).[112] Two of the three studies with non-negative SMDs compare mean IQs in children living in fluoridated vs. non-fluoridated areas in Canada,[113] or in New Zealand.[25] No other studies included in the main mean-effects meta-analysis made comparisons between fluoridated vs. non-fluoridated areas. In both studies, levels of fluoride in water were low, even in communities with fluoridated drinking water, likely limiting the power to detect an effect.

  In Bashash et al.,[112] the SMD compares mean IQ scores in children with urinary fluoride levels below vs. above 0.80 mg/L in Mexico.[112] Unlike other studies in the mean-effects meta-analysis which compared mean IQ scores between fluoridated vs. non-fluoridated areas, or areas with high vs. low fluoride exposures (see eTable 2), the Bashash et al.[112] study was not designed to measure fluoride exposure by geographical area. However, since the mean IQ scores were provided in the manuscript for children with urinary fluoride levels below vs. above 0.80 mg/L, we included them in this analysis. It's worth noting that there was no significant difference when comparing MUF levels between the groups of children with urinary fluoride levels above or below 0.80 mg/L, however when children's IQs were regressed against MUF, a statistically significant inverse association was found."



eFigure 1. Association Between Individual-level Fluoride Exposure and IQ Scores in Children: Overall Analysis.

eFigure 19 note: Estimates (betas) for individual studies are shown with solid boxes representing the weight, and the pooled estimate is shown as a solid diamond. Horizontal lines represent 95% Cis for the study-specific betas.



eFigure 23. Association Between Individual-level Fluoride Exposure and IQ Scores in Children: Effect by Exposure Type

6b.D: A meta-analysis of randomized controlled trials with similar results bolsters the evidence of an intervention effect, but a meta-analysis of observational studies, all subject to the same biases, increases the probability of a misleading result. The p-value will become smaller by virtue of increased sample size, but not because of any true cause and effect relationship. The bottom line is that potentially confounding effects on IQ are not randomly assigned, and that makes tenuous any conclusion of a causal effect of fluoride on IQs.

Response: Disagree (no change)

- Unfortunately, as is the case with most studies of potentially harmful exposures, there are no randomized controlled trials assessing the association between exposure to fluoride and children's intelligence (likely due to ethical concerns about randomizing pregnant women and/or children to fluoride). Therefore, observational studies are the best source of available information.
- This comment also repeats the RCT argument relative to observational studies. See prior responses regarding support of the value of observational studies in the public health, systematic review, and environmental epidemiological communities. Other public health conclusions and practices have long been supported by observational studies. For example, the evidence showing that community water fluoridation protects against tooth decay was largely based on observational or "association" studies, most of which were conducted prior to the introduction of fluoridated toothpaste in the early 1970s (Iheozor-Ejiofor et al. 2015).
- The assumption that all observational studies in a meta-analysis suffer from the same biases is unfounded. As mentioned in an earlier response to comment, risk of bias was systematically assessed for each individual study. Multiple potential sources of bias (including confounding bias, selection bias, exposure characterization, and outcome assessment) were extensively evaluated for each individual study, and results of those assessments are presented in Appendix E of the prepublication 2022 NTP Monograph.

6b.E: Next, [redacted] consider the weaknesses identified by the authors themselves. To their credit, the authors attempted to assess the impact and ameliorate the consequences of these weaknesses through analytical approaches such as funnel plots and Egger's test to detect publication bias, trim and fill methods to correct for publication bias, the I^2 and Q statistic for detection of heterogeneity, and subgroup and sensitivity analyses to try to explain the heterogeneity. Unfortunately, these approaches cannot correct for the use of inappropriate study design in a meta-analysis.

Response: Disagree (no change)

- We disagree that meta-analyses of observational studies are not appropriate. See prior responses regarding support of this in the public health, systematic review, and environmental epidemiological communities.
- Furthermore, we conducted the meta-analysis in response to the NASEM Committee's peer review of the 2019 draft NTP Monograph, which stated that the "committee strongly recommends that NTP reconsider its decision not to perform a meta-analysis."

  Additionally, as mentioned in a previous response to comment, the NASEM Committee supported our approach and described the information presented in the meta-analysis as valuable and informative.

- Note: A quick search[1] for meta-analyses of observational studies in PubMed alone identifies over 20,000 studies, which indicates how prevalent meta-analyses of observational studies are in the scientific literature.

6b.F: As recommended in the Cochran Handbook of Systematic Reviews, concerns such as high heterogeneity should preclude the use of the meta-analysis in the first place. A meta-analysis can be performed with data from as few as two studies and therefore it is best to consider only clinically and methodologically similar and sound studies. When results from studies with substantial differences in design, exposure, outcome measures, and risk of bias are combined, the effects of any exposure are more likely to be overestimated. This result was evident in the manuscript when comparing adjusted and non-adjusted findings, and findings from high vs. low risk-of-bias (RoB) studies. For example, the primary effect estimate of differences in children's IQ (Standardized Mean Difference, SMD) shifted from a medium effect size for all studies combined (-0.49) to a small effect size among the low risk of bias studies (-0.24). At the very least, adjusted results and findings only from those studies with a low risk of bias should be emphasized, and the reader should be given a clear interpretation of what the SMD values reflect.

Response: Disagree (no change)

- The Cochrane Handbook (Higgins et al. 2021) does not say that high heterogeneity should preclude the use of meta-analysis, as is suggested in the comment. In fact, Cochrane Section 10.10.3 (Deeks et al. 2021) says that sources of heterogeneity should be explored using prespecified subgroup analyses. Therefore, the meta-analysis followed the Cochrane Handbook recommendations as reflected in the protocol (which underwent peer review) that identified prespecified potential sources of heterogeneity for later analyses. These prespecified potential sources of heterogeneity were then appropriately explored in the subgroup analyses.
- Note that, in addition to recommending NTP conduct a meta-analysis (see response to previous comment), the NASEM Committee, in their 2020 peer review report on the 2019 draft NTP Monograph, stated that a properly conducted meta-analysis can account for heterogeneity in exposure measurements and other aspects of study design.
- Furthermore, in its peer review of the 2020 draft NTP Monograph, the NASEM Committee supported the subgroup analyses NTP used in this evaluation, finding them informative and directly responsive to some of the Committee's previous concerns. They also recommended additional subgroup and sensitivity analyses that were subsequently added to the manuscript:

  "As part of its meta-analysis, NTP presents several subgroup and sensitivity analyses. The committee finds them very informative; several are directly responsive to some of the committee's previous concerns. However, **NTP should also include subgroup or sensitivity**

[1]The following search string was used for the "quick" search because it identifies a high percentage of appropriate studies: (("meta-analysis"[Publication Type] AND ("meta-analysis of observational studies"[tiab] OR "meta-analyses of observational studies"[tiab] OR "observational studies as topic"[MeSH Terms] OR "Observational Studies"[title] OR "Observational Study"[title] OR "Cohort Studies"[Mesh] OR "Cohort Study"[Title] OR "Cohort Studies"[Title] OR "Case-Control Studies"[Mesh] OR "Case-Control Study"[Title] OR "Case-Control Studies"[Title] OR "Cross-Sectional Studies"[Mesh] OR "Cross-Sectional Study"[Title] OR "Cross-Sectional Studies"[Title] OR "Ecological Study"[Title] OR "Ecological Studies"[Title] OR "Interrupted Time Series Analysis"[Mesh] OR "Time Series Analysis"[Title] OR "Time Series Analyses"[Title] OR "Time Series Study"[Title] OR "Time Series Studies"[Title])) NOT "Randomized Controlled Trial"[publication type]).

analyses that respond to the committee's concerns about blinding, complex sampling designs, and statistical analyses that account for clustered study designs.… The additional subgroup or sensitivity analyses noted could help to alleviate some of the committee's current concerns."

- As stated in the protocol, when available, we used the adjusted effect estimates in the meta-analyses. Also, in contrast to what the comment implies, the adjusted versus non-adjusted sensitivity analysis found no difference in results [Adjusted β (95% CI)= −1.81 (−2.80, −0.81); unadjusted β (95% CI) = −1.81 (−2.81, −0.83)]. As the comment points out, the direction of the association was consistent across study quality from the high risk-of-bias to the low risk-of-bias studies, and the effect estimate was smaller among the low risk-of-bias studies. This may be due to lower levels of exposure and/or smaller differences in exposure between "high" and "low" exposure groups among the low risk-of-bias studies. The comment fails to note that there are other stratified estimates that would be considered to underestimate the effect estimate in both the *mean-effects* and *regression slopes meta-analyses*.

6b.G: Additional interpretation and explanation of the subgroup analyses are also needed. Using the results of sub-group analyses to investigate and explain heterogeneity does not accomplish that goal. In most subgroups, there seem to be subgroup effects implying interactions between SMD and investigated factors such as gender, country, and risk of bias (although between sub-groups p-values are not supplied to determine whether these interactions were significant or not). There was also significant unexplained heterogeneity among studies that needs to be investigated further.

Response: Agree (change made)

- We agree that the manuscript benefited from additional discussion of the results of the subgroup analyses. To be responsive to [redacted] comment, we added the following new text in the *Results* section for the *mean-effects meta-analysis*:

  "The subgroup and meta-regression analyses did not explain a large amount of the overall heterogeneity; however, the degree of heterogeneity was lower for studies restricted to Iran (I^2=56%), children ages 10 and older (I^2=68%), and girls (I^2=76%)".

- In the *Results* section for the *regression slopes meta-analysis*, we added the following new text:

  "The observed heterogeneity in the overall effect estimate was explained by the subgroup analyses, with no significant heterogeneity remaining in analyses of low-risk-of bias studies, by sex, by country, by assessment type, and by exposure timing (Table 3)."

- In the *Discussion* section, we added the following new text with further interpretations of the subgroup analyses:

  "With a couple exceptions, the subgroup analyses in the mean-effects meta-analysis did not explain a large amount of the overall heterogeneity. However, the heterogeneity in the regression slopes meta-analysis was explained by subgroup analyses. This suggests that the aggregate nature of the mean-effects meta-analysis might not be sufficiently sensitive to capture potential sources of heterogeneity, as seen possible when using studies with individual-level data in the regression slopes meta-analysis. However, the large number of studies included in the mean-effects meta-analysis and the consistency in the direction of the association across the analyses make this is less of a concern."

- As recommended in the comment, we also further investigated potential sources of heterogeneity by conducting a meta-regression analysis using mean age in years and year of publication in each study. In the supplemental materials we added:

  "The results of the meta-regression models indicate that year of publication and mean age of study children did not explain a large degree of heterogeneity as neither were significant predictors of the relationship between fluoride and children's intelligence, and the residual I^2 remained high (85% and 87%, respectively). Year of publication (SMD = 0.01, 95% CI: -0.01, 0.02) and mean age (SMD = -0.04, 95% CI: -0.13, 0.04) explained relatively little between-study variance (adjusted R^2 of 12% and 5%, respectively). When both year of publication and mean age were included in the model, there were no notable improvements to the amount of between-study variance explained (adjusted R^2 = 13%) or percent residual variation due to heterogeneity (residual I^2 = 85%).

  Excluding the outlier study34 resulted in a slightly lower heterogeneity for the overall effect estimate (I^2=84%) and for the India-specific effect estimate (I^2=69%). The meta-regression indicates that mean age is a significant predictor of the effect (SMD = -0.06, 95% CI: −0.12, −0.01, p-value =0.025), explaining 9% of the between-study variance. Year of publication (SMD = 0.01, 95% CI: 0.001, 0.02, p-value=0.028) explained a larger degree of between-study variance (R^2 = 19 %)."

6b.H: These inconsistencies create uncertainty regarding the validity and significance of the exposure effect estimate for each subgroup. There were fewer than ten studies in many subgroups, thereby reducing their ability to identify statistically significant differences.

Response: Disagree (no change)

- We disagree with [redacted] that there are inconsistencies that would create uncertainty regarding the validity and significance of the exposure effect estimate for each subgroup.
- Also, as previously mentioned, the purpose of the subgroup analyses was to explore sources of potential heterogeneity, not to detect differences between the groups or "interactions between SMD and investigated factors." However, except for certain countries and for studies with other sources of fluoride exposure, all the subgroup analyses of the *mean-effects meta-analysis* included at least 10 studies. In addition, all the subgroup analyses with fewer than 10 studies but more than 1 study (subgroups: India, Iran, and dental fluorosis) reported statistically significant estimates, as shown in the excerpt of Table 2 below.
- As mentioned previously, the NASEM Committee agreed with our use of the prespecified subgroup analyses to investigate sources of heterogeneity, finding them informative and directly responsive to some of the Committee's previous concerns. They also recommended additional subgroup and sensitivity analyses that were subsequently added to the manuscript.

Excerpt of Table 2. Pooled SMDs and 95% CIs for Children's IQ Scores and Exposures to Fluoride.

Analysis	Number of Studies	SMD (95% CI)	Heterogeneity	
			p-value	I^2
Subgroup Analyses				
Country				
China	39	−0.43 (−0.52, −0.34)	<0.001	85%
India	8	−0.99 (−1.55, −0.43)	<0.001	93%
Iran	4	−0.68 (−0.99, −0.38)	0.077	56%
Canada	1	0.01 (−0.19, 0.21)	NA	NA
Mexico	1	0.13 (−0.16, 0.42)	NA	NA
New Zealand	1	0.01 (−0.19, 0.22)	NA	NA
Pakistan	1	−0.25 (−0.65, 0.16)	NA	NA
Assessment Type				
CRT-RC tests	29	−0.36 (−0.46, −0.27)	<0.001	82%
Non-CRT-RC tests	26	−0.60 (−0.78, −0.42)	<0.001	89%
Raven's tests	10	−0.76 (−1.10, −0.43)	<0.001	91%
Other tests	16	−0.52 (−0.74, −0.29)	<0.001	89%
Exposure Type				
Water fluoride	32	−0.37 (−0.46, −0.27)	<0.001	82%
Dental fluorosis	7	−0.99 (−1.57, −0.41)	<0.001	96%
Other exposures[b]	16	−0.54 (−0.71, −0.37)	<0.001	81%

6b.I: Furthermore, the "dose-response" relationship assessments yielded conflicting conclusions, ranging from "non-linear" for fluoride water study to "linear" for urine studies, to "no effect" for other exposure groups, lacking biologic plausibility and casting additional doubt on the overall assessment.

Response: Disagree (no change)

- We disagree that there are conflicting conclusions. The direction of the observed association was consistent across both the water and urine *dose-response meta-analyses*. There are, however, differences in which model was the best fit for the data. Given the heterogeneity and the fact that the individual studies contributing to the water and urine *dose-response meta-analyses* were different, differences in model fit are expected.

6b.J: In summary, while the results of this meta-analysis imply a statistical link between fluoride exposure and IQ, they should be interpreted and communicated with great caution due to the potential for bias from observational studies, the lack of an underlying biologic or scientific plausibility, numerous methodological and statistical issues, and the potential for detriment to the public's health caused by the effect on public perception and policy caused by improperly attributing a putative adverse health effect to an intervention with significant known benefits.

Response: Agree (no change)

- We agree that the results of this analysis require careful and clear communication, which is why we are working closely with the NIEHS Office of Communications to draft relevant communications. We agree that public perceptions around exposures to fluoride are very important and think that this meta-analysis and the prepublication 2022 NTP Monograph should be used to inform a careful analysis of data concerning the potential risks as well as benefits of fluoride. We have provided detailed responses to [redacted] critique concerning risk of bias from observational studies elsewhere in a previous response. We discuss biological plausibility of the studies included in this meta-analysis in the prepublication 2022 NTP Monograph.

6b.K: The authors made laudable attempts to mitigate the impact of these problems, but no statistical approach can solve all of the problems caused by the inappropriate choice of meta-analysis. As indicated in the Cochrane Handbook, results from the investigation of high heterogeneity studies that is designed after heterogeneity is identified can at best lead to hypotheses generation and to support proposals for additional studies. They should be interpreted with caution and should generally not be listed among the conclusions of a review (Cochran Handbook, Section 10.10.3). They should certainly not be used as the rationale for changing public policy.

Response: Disagree (no change requested)

- Section 10.10.3 of the Cochrane Handbook has been misrepresented in the above comment. The sentence from the Cochrane Handbook immediately before the one referenced in the comment states: "Reliable conclusions can only be drawn from analyses that are truly pre-specified before inspecting the studies' results, and even these conclusions should be interpreted with caution." We again point out that all the analyses investigating potential sources of heterogeneity were planned a priori as reflected in the protocol or were added at the recommendation of the NASEM Committee or other peer reviewers.

6b.L: **A more scientifically justifiable conclusion for this review is that extensive, rigorous, and reproducible research in both animals and humans is needed to address the important question of causal influences of fluoride on human cognition.**

Response: Disagree (no change)

- It's always easy to call for more research and we agree that targeted research can certainly add clarity to the existing data—particularly at lower exposure levels. However, hundreds of human and animal studies have been published on this topic. Although these comments are on a previous draft of the meta-analysis, we would like to point out that a recent update of the literature identified 10 new studies that were subsequently added to the database (and are included in the current draft). These new studies were published in the past 2 years and their addition left the findings of the analysis essentially unchanged. Our meta-analysis now includes 60 studies of children's cognition and fluoride exposure, 13 of which are high quality. Many high-quality meta-analyses have been based on fewer studies and the current meta-analysis includes more than double the number of studies of any previous meta-analysis of fluoride.

6b.M: Introduction:

The manuscript described three different types of meta-analyses, using Standardized Mean Difference (SMD) as the effect estimate for each study's outcome (IQ), and assessed and addressed issues related to heterogeneity and publication bias.

First, *mean effect meta-analysis* of group-level fluoride measurement studies (n=46) was conducted to investigate putative associations between fluoride exposure and a child's IQ, with the conclusion that there was "an inverse association between fluoride exposure and children's IQ" (pooled SMD for all studies: −0.49; 95% CI: −0.60, −0.38; p-value < 0.001). However, there was evidence of high heterogeneity (I^2 = 89%, p-value < 0.001) and publication bias (funnel plot and Egger's p-value < 0.001, Begg's p = 0.04), both of which militate against the use of meta-analysis.

Second, *dose-response meta-analysis* of group-level fluoride measurement studies (n=46) was conducted to assess dose-response relationships between fluoride and IQ, with the conclusion that "associations for drinking water appeared to be non-linear and associations for urine appeared to be linear." Heterogeneity and publication bias issues were not reported in this section of the manuscript.

Third, *meta-analysis of regression slopes* for the individual-level urine studies (n=6) was conducted to assess study outcomes with respect to a 1-mg/L unit increase in urinary fluoride. There was moderate heterogeneity (I^2=48%, p=0.09) and indication of publication bias. The manuscript concluded that, after adjustment for publication bias using a trim and fill approach, "a 1-mg/L increase in urinary fluoride was associated with lower IQ, with an adjusted pooled effect estimate of −0.87 (95% CI: −1.93, 0.19; p-value = 0.302)". A p-value of 0.302 indicates that chance may be a reasonable explanation for this finding. The critiques of meta-analyses that follow are categorized by the major issues mentioned above.

Response: Disagree (no change)

- As described in responses to earlier comments, we disagree that evidence of heterogeneity and publication bias militate against the use of meta-analysis. We conducted the meta-analysis in response to the NASEM Committee's peer review of the 2019 draft NTP Monograph. The NASEM Committee urged us not to avoid conducting a meta-analysis because of heterogeneity: "The committee strongly recommends that NTP reconsider its decision not to perform a meta-analysis and, if it still decides not to do a meta-analysis, that it provide a more thorough and convincing justification for its decision…A properly conducted meta-analysis can account for heterogeneity in exposure measurements and other aspects of study design, so it is not clear why heterogeneity was listed as a reason for not performing one."
- The *dose-response meta-analysis* used the same studies that were used in the *mean-effects meta-analysis*. The *mean-effects meta-analysis* already describes heterogeneity and publication bias issues. Therefore, it would be redundant to describe them again.
- After updating the *regression slopes meta-analysis* with new studies from the updated literature search, there was no longer evidence of publication bias, so the quoted text has been removed from the manuscript.

6b.N: Issues related to using SMD as an effect estimate

From the manuscript:

> "The effect estimates in the primary mean-effects meta-analysis were the standardized mean differences (SMDs) for heteroscedastic population variances."

Comment:

The overall treatment effect [in terms of SMD] can be difficult to interpret as it is reported in units of standard deviation rather than in units of any of the measurement scales used in review (Egger et al., 2008). Why would the true effect of fluoride (assuming there is one) depend on the standard deviation? If the reason for using the SMD instead of the more interpretable difference in IQs is that different tests were used to assess intelligence, it is an indication that combining such disparate studies may be inappropriate. *"There is a price for standardization—the SMD does not have any meaningful units. Instead, it can only indicate whether there is any statistical significance of pooled results." (Mickael and Merja, 2021).* Nonetheless, interpreting SMD values of 0.2, 0.5, and 0.8 as small, medium, and large effect sizes, respectively, is a widely accepted role of thumb (Cohen, 1988). Accordingly, using these terms throughout will help clarify the meaning of this estimate. There is also concern that the inclusion of studies with both high RoB and large sample size leads to overstated estimates of the effect sizes.

Response: Disagree (no change)

- The difference in tests is an appropriate reason to use the SMD as the unit of measure in our meta-analysis. The SMD is commonly used in meta-analysis when the studies all assess the same outcome (e.g., intelligence) but measure it in a variety of ways (e.g., WISC-R, Combined Raven's Test for Rural China, etc.). It is necessary to standardize the results of the studies to a uniform scale before they can be combined (Higgins et al. 2021). To address the concern of combining studies that used different tests to assess intelligence, we conducted subgroup analyses that stratified by type of IQ assessment. We also acknowledge limitations of the *mean-effects meta-analysis* in our discussion.
- In addition, in its peer review of the 2019 draft NTP Monograph, the NASEM Committee supported the use of SMDs in this meta-analysis as the Committee recommended that NTP update the Choi et al. (2012) meta-analysis (which used SMDs) with more recent papers. The protocol, which clearly describes these methods, was also peer reviewed.
- Also, as previously mentioned, in the peer review of the 2020 draft NTP Monograph, the NASEM Committee agreed with the methods used in the meta-analysis: "The overall approach appears to be sound in comparing mean IQ scores for the most and least highly exposed to fluoride even though the absolute fluoride concentrations are not comparable among studies", and "The meta-analysis applied standard, broadly accepted methods, and the data shown in Figure A5-1 and the related evaluations are especially informative (NTP 2020a, p 235)."
- We appreciate the suggestion regarding interpretation of the SMDs; however, because the standard deviations of measured IQs are specific to the study population from which they are measured, and the meta-analyses pools the results of many different study populations, we did not translate the pooled SMD into IQ points nor did we characterize them as small, medium, or large. In addition, the Cochrane guidance (Cochrane Section 12.6.2) states that

"…some methodologists believe that such interpretations are problematic because patient importance of a finding is context-dependent and not amenable to generic statements." Also, the SMD interpretations based on cutoffs mentioned by [redacted] are values used in social sciences research (as cited in the 1988 Cohen book "Statistical Power for the Behavioral Sciences") and the utility of those values in analyzing observational environmental health studies has not been demonstrated.

- The concern about combining results from high and low risk-of-bias studies was addressed by the subgroup analyses stratified by risk of bias. As the comment points out, the effect estimate was smaller among the low risk-of-bias studies. This may be due to lower levels of exposure and/or smaller differences in exposure between "high" and "low" exposure groups among the low risk-of-bias studies. The comment fails to note that there are other stratified estimates that would be considered to underestimate, rather than overestimate, the pooled effect estimates in both the *mean-effects* and *regression slopes meta-analyses*. As for studies with large sample sizes, we performed the meta-analyses using random effects models which account for study-specific sample sizes.
- Finally, we would like to note that this comment completely ignores that this manuscript was not restricted to an SMD meta-analysis. The manuscript also includes a *regression slopes meta-analysis* (which has not been previously done in the fluoride and IQ literature). The *regression slopes meta-analysis* does not have the same limitations as the SMD analysis. It uses individual-level exposure data, and the regression coefficient can be directly interpreted as the expected change in IQ points in the study population per 1-mg/L increase in urinary fluoride.

High heterogeneity among studies

6b.O:

From the manuscript:

> The heterogeneity among the group-level fluoride measurement studies was high for all studies (n=46) as well as for the low risk of bias studies (n=9).

Comments:

- This can be seen not only in statistics such as I^2 and the p-value for heterogeneity, but in the figures as well. For example, if there were homogeneity of effects, then approximately 5% of SMDs should be outside of the dotted lines in supplemental eFigure 3. Instead, more than a third of them are outside the dotted lines. A similar phenomenon can be seen in supplemental eFigure 8, even for the studies with low risk of bias. This indicates an unacceptably large level of heterogeneity, such that mean-effect and dose-response meta-analyses should not be conducted in first place. *"High heterogeneity can potentially lead to misleading and non-generalizable results and may indicate that meta-analysis is contra-indicated. A group of studies needs to be similar enough clinically and methodologically to be pooled in a meta-analysis before considering their statistical heterogeneity." (Cochrane handbook, Section 9.5).*

Response: Disagree (no change)

- [redacted] refers to eFigure 3, which is a funnel plot of the included studies in the *mean-effects meta-analysis*. The funnel plots are not used to illustrate or evaluate

homogeneity as implied by [redacted], but to evaluate the potential for publication bias. To evaluate heterogeneity, we performed and reported results of statistical tests for heterogeneity, while also transparently discussing limitations of such tests in the *Discussion* section.

- We disagree with [redacted] that the *mean-effects* and *dose-response meta-analyses* should not have been conducted. As previously explained, high heterogeneity is not a valid rationale for not conducting a meta-analysis. Again, the NASEM Committee agreed with the methods used in the meta-analysis in its peer review. Additionally, in meta-analyses of observational studies, especially those using SMDs as effect measures, high levels of heterogeneity are to be expected. Our protocol outlined the study inclusion criteria which were carefully evaluated to ensure that the appropriate studies were included in the meta-analyses. The protocol also outlined the subgroup analyses that were to be performed to investigate potential sources of heterogeneity.
- The select quote from Section 9.5 misrepresents the totality of the Cochrane guidance, particularly on heterogeneity. Section 10 of Cochrane is "Analyzing data and undertaking a meta-analysis" and section 10.10.3 is on heterogeneity, where Cochrane recommends exploring heterogeneity by conducting subgroup analyses. We transparently presented the heterogeneity results and investigated potential sources of heterogeneity.
- In the *Discussion* section, we clearly outline the limitations of the *mean-effects meta-analysis* and the unexplained heterogeneity. We also added new text to point out that:

  "…the aggregate nature of the mean-effects meta-analysis might not be sufficiently sensitive to capture potential sources of heterogeneity, as seen possible when using studies with individual-level data in the regression slopes meta-analysis."

6b.P:

- Miranda et al. (2021) performed a similar meta-analysis with their results, pointing to an association between fluoride and IQ. However, due to the high heterogeneity among the existing studies they concluded that current evidence is inadequate to support such a conclusion, even at high fluoride levels.

Response: Disagree (no change)

- The Miranda et al. (2021) meta-analysis was not similar to our meta-analysis, which was different in both scope and methodological approach. For example, the systematic review by Miranda et al. (2021) had very limited inclusion criteria, which did not allow for studies using individual-level fluoride exposure measurements to be included. Their analysis only included cross-sectional studies, while our meta-analysis also included prospective cohort studies. In addition, their analysis was much smaller (n = 10 studies) than our meta-analysis (n = 60 studies) and had a very different methodological approach, as it was limited to studies from which crude (unadjusted) odds ratios could be calculated. Finally, Miranda et al. (2021) was limited to one analysis that found a strong association between high fluoride exposure and decreased IQ (unadjusted OR = 3.88, 95% CI 2.41−6.23; $p < 0.00001$). Our analysis found consistent results across different analysis types (*mean-effects, dose-response, and regression slopes meta-analyses)* and across multiple prespecified subgroup analyses.

- The observed level of heterogeneity (77%) in Miranda et al. (2021) is not unusual in small meta-analyses such as theirs; however, it is also worth noting that the authors did not attempt to investigate any sources of heterogeneity in their analysis.

6b.Q:

- Since meta-analyses can be performed with data from as few as two studies, it is more appropriate to include only studies that are clinically and methodologically similar and sound. Pooling results from all studies with significant differences and biased results is not appropriate (ref. Cochrane Handbook) and is likely to overestimate the effects of exposure. This seems to be the case for the meta-analysis of regression slopes (urine level studies, n=6) which reported medium heterogeneity among low-risk, individual-level fluoride measurement.

  Response: Disagree (no change)

  - The purpose of the meta-analysis is to combine results from multiple studies with a variety of features to examine data collectively and more precisely quantify the overall (pooled) association. The Cochrane Handbook does not say it is "not appropriate" to pool results from studies with differences in design and potential sources of bias. Rather, they recommend that differences in study design, study biases, variation in exposure characterization and outcome assessment across studies, and reporting biases be carefully considered. Moreover, excluding studies from systematic reviews or excluding studies from systematic reviews with meta-analyses is not considered a best practice in the systematic review community (Higgins et al. 2021). As a well-documented systematic review and meta-analysis, this evaluation followed a protocol where inclusion and exclusion criteria were defined a priori. Among included studies, our risk-of-bias assessment carefully considered study-specific potential for bias. Our analyses stratified results by risk-of-bias status to evaluate the potential impact on the overall effect estimates from studies that have high potential for bias versus studies that have low potential for bias. We carefully considered other differences between studies by conducting additional prespecified subgroup analyses by factors such as exposure type, outcome assessment, and country.
  - In addition, as previously mentioned, the NASEM Committee agreed with the methods used in the meta-analysis, which includes pooling results from included studies:

    "The critical information regarding comparison of study results comes from the new meta-analysis, which seeks to extract and integrate comparable findings from selected studies as discussed further below. The overall approach appears to be sound in comparing mean IQ scores for the most and least highly exposed to fluoride even though the absolute fluoride concentrations are not comparable among studies."

    "The meta-analysis applied standard, broadly accepted methods, and the data shown in Figure A5-1 and the related evaluations are especially informative."

6b.R: Three strategies were used to assess/address the heterogeneity:

Random-effects models to address heterogeneity:

<u>From the manuscript:</u>

"Data from individual studies were pooled using a random-effects model."

Comments:

- Random-effects models, as opposed to fixed-effect models, are typically used in meta-analyses when there is unexplained heterogeneity. Such models assume that the effects estimated within each study are not identical, but do follow a specific distribution (Cochrane handbook, Section 9.5). However, according to the Cochrane Handbook, random-effect models can only be used "if the heterogeneity cannot be explained clinically or methodologically. It does not remove heterogeneity, so results need to be carefully interpreted." (Cochrane handbook, Section 9.5). This is relevant because the decision to use random effect models seems to be based on statistical findings of heterogeneity, with no attempt to ensure the clinical and methodologic similarity among the studies before conducting the meta-analyses. The State of the Science document mentioned that "heterogeneity within the available evidence was evaluated to determine if a quantitative synthesis (i.e., meta-analysis) is appropriate." (p.19) but no presentation or discussion of the outcomes of this process can be found.

  Response: Disagree (no change)
  - We are unable to find the quote that [redacted] cites or any text in the Cochrane Handbook that states that random-effects models can only be used if the heterogeneity cannot be explained clinically or methodologically.
  - We followed Cochrane guidance (Section 10.10) that recommends using a random-effects model instead of a fixed-effects model when the assumption of a common (fixed) effect size is not appropriate. Figure 2 shows heterogeneity in study-specific effect estimates, clearly indicating that a fixed effect is not appropriate in our meta-analysis (see Figure 2 below). In addition, as recommended by the Cochrane Handbook, even though we used random-effects models, we still investigated potential sources of heterogeneity (Cochrane, Section 10.10; Deeks et al. 2021).

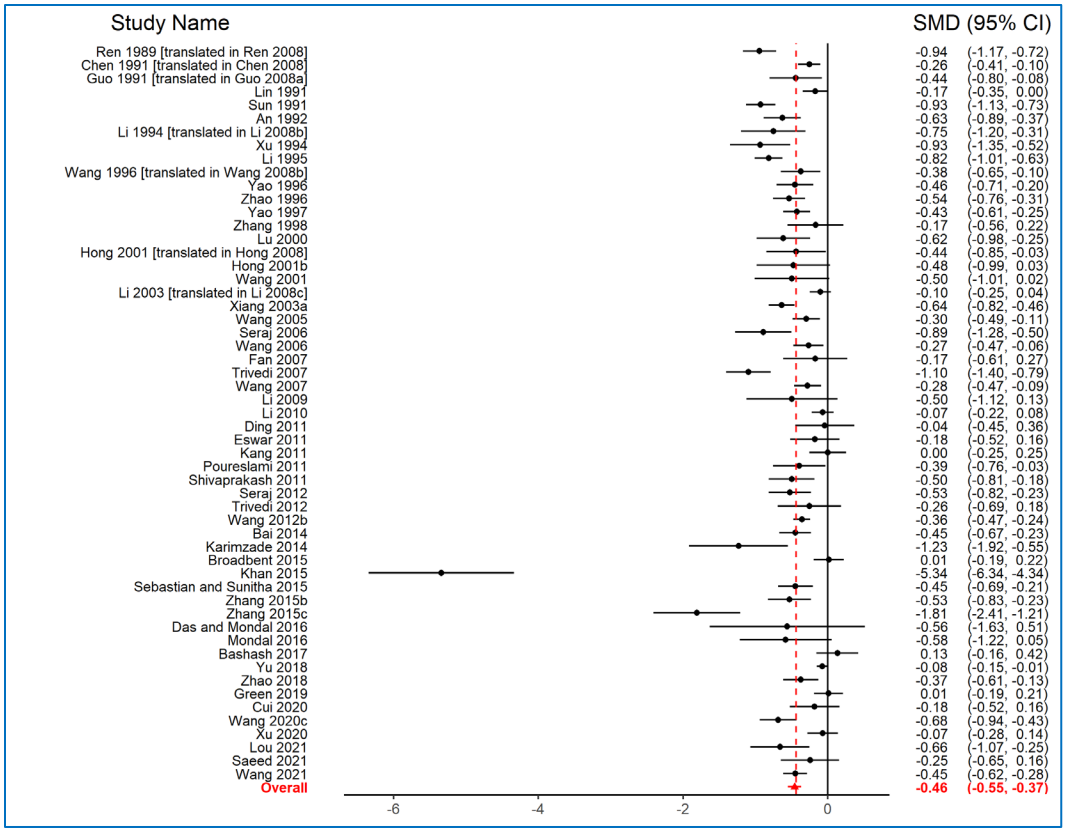


Figure 2. Association Between Fluoride Exposure and IQ Scores in Children

6b.S:

- According to the Cochrane Handbook, "*a pragmatic approach is to plan to undertake both a fixed-effect and a random-effects meta-analysis, with an intention to present the random-effects result if there is no indication of funnel plot asymmetry. If there is an indication of funnel plot asymmetry, then both methods are problematic*." (Cochrane Handbook, Section 10.10.4.1).

Response: Disagree (no change)

- This quote comes from a list of considerations for authors to contemplate when making a choice about whether to use a fixed-effects model or random-effects model. The Handbook is clear that there are a variety of factors to consider and that there is no universal recommendation on which model to use. As explained in the previous response, we have demonstrated that the random-effects model was the appropriate choice for these data.

6b.T:

A fundamental assumption of the random effects model is that the true effects in different studies represent a random sample from some population. The majority of studies are from one country, with unique environmental, economic, and sociopolitical conditions that can hardly be regarded as a random sample that allows generalization to other countries.

Response: Agree (no change requested)

- We agree, which is why we investigated potential sources of heterogeneity, including country.

6b.U: Using sensitivity analyses to address heterogeneity

From the manuscript:

"Multiple sensitivity analyses were conducted as part of the mean-effect meta-analysis (e-table 3) and meta-analysis of regression slopes (e-table 6). Four additional analyses were conducted as per NASEM's recommendation (not shown). The authors concluded that no substantial changes in the pooled SMD estimate were revealed when studies were excluded."

Comment:

Sensitivity analyses do appear necessary, as there is at least one very clear outlier (Khan, 2015) in Figure 2 and in several figures in the supplementary materials. Nonetheless, removal of one or two studies does not eliminate the heterogeneity of results; the magnitude and direction of the effect remain unknown because of lack of adequate testing for heterogeneity.

Response: Disagree (no change)

- We disagree that there was a lack of adequate testing for heterogeneity. As previously mentioned in an earlier response, the NASEM Committee agreed with our use of the prespecified subgroup analyses to investigate sources of heterogeneity, finding them informative and directly responsive to some of the Committee's previous concerns. They also recommended additional subgroup and sensitivity analyses that were subsequently added to the manuscript. Also, as previously explained, the goal of these subgroup analyses was not to eliminate heterogeneity.
- We also disagree that the direction of effect remains unknown. The NASEM Committee agreed with our conclusion that the results (i.e., the direction of the association) were consistent: "As noted in the revised monograph, 44 of the 46 studies represented in that figure had effect estimates to the left of zero-results that indicate an association between higher fluoride exposures and lower IQ. Those results highlight the marked consistency in the current epidemiological literature on fluoride and childhood IQ."
- The NASEM Committee also commented that "NTP notes that 44 of the 46 studies (96%) in its meta-analysis of childhood IQ have effect estimates to the left of zero. That finding should be emphasized more, and its meaning with respect to evaluating and quantifying heterogeneity should be mentioned. To assess heterogeneity, NTP primarily used the Cochran's Q test. However, heterogeneity can also be assessed by providing a count or percentage of the number of studies to the right or left of the null value. Some would consider that a much simpler, more intuitive, and perhaps more useful way of assessing heterogeneity, especially in light of the marked differences between the studies in design, study populations, exposure and outcome assessment methods, and statistical analyses."

6b.V: Using subgroup analyses to address heterogeneity

From the manuscript:

"Subgroup analyses were performed to investigate sources of heterogeneity"

Comment:

Sub-group analyses described in the manuscript do not appear to be pre-specified or justified at the protocol stage based on a clear theoretical, biological, or clinical basis. According to the Cochrane Handbook, "subgroup analysis should be kept to a minimum, and pre-specified and justified at the protocol stage of the review. The planned analyses should be followed at review stage (if sufficient data are available) to minimize selective reporting or over-interpretation of the results based on findings." (Cochrane Handbook, Section 9.6). Furthermore, "reliable conclusions can only be drawn from analyses that are truly pre-specified before inspecting the studies' results, and even these conclusions should be interpreted with caution."

Response: Disagree (no change)

- We disagree with the implication that the subgroup analyses were not based on a biological or other scientific basis. These analyses were based on established scientific evidence which includes the National Research Council's 2006 report (NRC 2006), two previous meta-analyses by Choi et al. (2012) and Duan et al. (2018), and the prepublication 2022 NTP Monograph.
- As previously mentioned in an earlier response, the NASEM Committee agreed with our use of the prespecified subgroup analyses, finding them informative and directly responsive to some of their previous concerns. They also recommended additional subgroup and sensitivity analyses that were subsequently added to the manuscript.
- We agree that subgroup and sensitivity analyses should be prespecified and followed to minimize selective reporting or over-interpretation of the results based on findings, and that results should be interpreted carefully. All our subgroup and sensitivity analyses were prespecified in the protocol or included at the recommendation of peer review.

6b.W:

From the manuscript:

"Sub-group analyses suggested that our conclusions were consistent across high and low risk-of-bias studies."

Comments:

- Results are clearly not numerically consistent within high- and low-risk of bias studies (effect size -0.55 versus -0.24). "Such a difference [in effect estimates between high and low risk of bias studies] is a common finding because biased studies are more likely to overestimate the effects of treatment." (Harrer et al (2021). Furthermore, there is substantial heterogeneity even in the low risk of bias studies. Five of the eight points lie outside the 95% pseudo confidence interval on the right side of eFigure 8. Moreover, several pairs of confidence intervals among the 9 low

risk of bias studies in eFigure 7 have completely non-overlapping confidence intervals, a strong indication that the true effects are different in different studies.

Response: Disagree (no change)

- Variation in the exact numerical estimate is expected in subgroup analyses, as different individual studies contribute to the different pooled effect estimates. However, the direction of the association, which we consider a more important indication of consistency in the literature, was consistent. The NASEM Committee, in its peer review report of the 2020 draft NTP Monograph, agreed with our statements on consistency:

  "As noted in the revised monograph, 44 of the 46 studies represented in that figure had effect estimates to the left of zero—results that indicate an association between higher fluoride exposures and lower IQ. Those results highlight the marked consistency in the current epidemiologic literature on fluoride and childhood IQ."

  "NTP notes that 44 of the 46 studies (96%) in its meta-analysis of childhood IQ have effect estimates to the left of zero. That finding should be emphasized more".

- We are not clear on the point of the Harrer et al. (2021) quote, as we were able to demonstrate differences in effect estimates by using subgroup and sensitivity analyses. As described above in response to another comment, studies that reported unadjusted and adjusted effect estimates did not provide evidence that higher potential for bias (as would be reflected in unadjusted effect estimates) resulted in an overestimation of the effect. This is captured in our sensitivity analysis for the *regression slopes meta-analysis* that used unadjusted effect estimates from Bashash et al. (2017), Cui et al. (2018), Green et al. (2019), and Yu et al. (2018) (see excerpt of eTable 6 below). Also, the comment fails to note that there are other stratified estimates that would be considered to underestimate, rather than overestimate, the pooled effect estimates in both the *mean-effects* and *regression slopes meta-analyses*.

- Note: We acknowledge that there is high heterogeneity in both the high and low risk-of-bias subgroups in the subgroup analysis by risk-of-bias status in the *mean-effects meta-analysis*. In a previous response above, we describe the new text added to the *Discussion* section with further interpretations of these subgroup analyses.

- [redacted] refers to eFigure 8, which is a funnel plot by risk of bias in the *mean-effects meta-analysis*. The funnel plots are not used to illustrate or evaluate heterogeneity, as implied by [redacted], but to evaluate potential for publication bias. We performed and reported on statistical tests for heterogeneity, while also transparently discussing limitations of such tests in the *Discussion* section.

- As [redacted] noticed, eFigure 7 illustrates that there is heterogeneity in the low risk-of-bias studies as well, as illustrated by the I^2 of 83% reported in Table 2.

Excerpt of eTable 1. Regression Slopes Meta-analysis

Analysis	Number of Studies	Beta (95% CI)	Heterogeneity	
			p-value	I^2
Overall Estimate				
Full-scale IQ	9	−1.81 (−2.80, −0.81)	<0.001	77%
Sensitivity Analyses				
Using unadjusted estimates from Bashash et al. (2017),[112] Cui et al. (2018),[76] Green et al. (2019)[113], Yu et al. (2018)[3]				
Full-scale IQ	9	−1.82 (−2.81, −0.83)	<0.001	76%

6b.X:

- In addition, sub-group analyses are "purely observational, so we should always keep in the mind that effect differences may also be caused by confounding variables." (Harrer et al., 2021). Of course, the studies in this meta-analysis are all observational, so confounding is a major concern even if there were no subgroup analyses. It is possible that fluoride or a combination of other factors is to blame for these differences. There is no sub-group analysis by exposure to fluoride or known neurotoxic chemicals such as lead and arsenic.

  Response: Disagree (no change)
  - As previously mentioned, potential confounding, including concurrent exposure to other neurotoxic chemicals (e.g., lead and arsenic), was assessed extensively as a key component of the risk-of-bias assessment. In addition, our analysis includes a subgroup analysis by exposure to other chemicals (such as arsenic, iodine, coal). Also, concerns about confounding among individual studies may be minimized or ruled out if consistent results are seen across different study populations, study designs, exposure settings, and studies that adjust for different sets of confounders (Arroyave et al. 2020; Steenland et al. 2020).

6b.Y:

From the manuscript:

"Heterogeneity remained low or moderate (I^2 < 48%) for all subgroup analyses except gender (I^2 > 52%)."

Comments:

- Most sub-group analyses, as shown in Table 2, had between-studies heterogeneity of 79% or higher, suggesting significant unexplained heterogeneity among sub-groups that needs to be further investigated. Subgroup analyses are intended to explain some of the heterogeneity, not to introduce more heterogeneity.

  Response: Disagree (no change)
  - In response to a previous [redacted] comment, we further investigated potential sources of heterogeneity. The results are presented in the supplemental materials.
  - We disagree that the heterogeneity within the subgroup analyses "needs to be further investigated." As stated before, the subgroup analyses were planned a priori to

investigate potential sources of heterogeneity in the overall effect estimate. It would not be informative nor is it common practice to then start investigating sources of heterogeneity within subgroup analyses.

6b.Z:

- Several of the included studies have overlapping high and low fluoride groups (i.e., what is labeled as low exposure groups in one study is considered high in another), which likely contributed to the study's high heterogeneity. Sub-group analysis based on precise cut-off points for exposure levels may help explain the considerable variability in the studies.

  Response: Disagree (no change)
  - This comment is referring to the *mean-effects meta-analysis* which was not designed to evaluate dose-response. However, we did address fluoride exposure levels separately in the *dose-response meta-analysis* using studies included in the *mean-effects meta-analysis*. The *dose-response meta-analysis* includes subgroup analyses based on precise cut-offs points for exposure levels (0 to <4 mg/L, 0 to <2 mg/L, and 0 to <1.5 mg/L). Within each of these subgroup analyses (i.e., <4 mg/L, <2 mg/L, and <1.5 mg/L), the data do not overlap.

6b.AA:

- According to Harrer et al. (2021), sub-group meta-analyses may lack power to detect small differences between groups. One solution is performing a subgroup statistical power analysis beforehand to determine the minimum detectable effect size difference with a subgroup analysis. An alternative approach would be to include a minimum of 10 studies for each level/unit (e.g. , country, gender, exposure type, etc.) analyzed in a sub-group analysis. However, many of the level/unit subgroup analyses in the manuscript included fewer than 10 studies.

  Response: Disagree (no change)
  - As we mentioned in a previous response and explained in our protocol, the purpose of the subgroup analyses was to explore potential sources of heterogeneity, not to detect small differences between subgroups. Even so, we would like to note that, except for certain countries and for studies with other sources of fluoride exposure, all the subgroup analyses of the *mean-effects meta-analysis* included at least 10 studies. We would also like to note that all the subgroup analyses across both the *mean-effects meta-analysis* and the *regression slopes meta-analysis* with 2–9 studies per group detected a statistically significant association.

6b.BB: Addressing publication bias

From the manuscript:

- Using funnel plot and Egger regression, evidence of publication bias was observed in all analyses including studies with a high risk of bias, but not those using studies with a low risk of bias.

- Adjusting for possible publication bias through trim-and-fill analysis resulted in an adjusted pooled SMD of – 0.36 (95% CI: −0.48, −0.24) under the mean effect meta-analysis, and an adjusted pooled effect estimate of −0.87 (95% CI: −1.93, 0.19; p value = 0.302) in the metanalysis of regression slopes.

Comment:

Since all but three studies show negative association to begin with, adjusting for publication bias would only center the studies around a distinctly negative pooled effect estimate, with incomplete representation of potential studies showing an association between fluoride and lower IQ. The effect size from all published studies (SMD= -0.49) is 26% larger than the adjusted effect size that imputes unpublished or excluded studies (SMD= -0.36). This imputation produces an adjusted estimate that is closer to a small effect than a medium effect based on Cohen's d. These variable results should be emphasized in the discussion/abstract, given the possibility that there is a bias against the publication of studies that have neutral outcomes.

Response: Disagree (no change)

- Accounting for publication bias is meant to account for potentially missing studies with neutral or positive effects, likely shifting the effect towards 0, as illustrated in the pooled SMD. As [redacted] points out, this was the case. In addition, the pooled SMD remained statistically significant even after the trim-and-fill analysis, which highlights the consistency of the overall association between fluoride exposure and lower IQ in children. We have reported these results in the *Results* section and provided more information in the supplemental materials; we disagree that more emphasis is needed.
- Note that, because there was no longer evidence of publication bias once the literature was updated in November 2021, the trim-and-fill analysis for the *regression slopes meta-analysis* was removed from the current draft.

6b.CC: Issues with the dose-response analyses

From the manuscript:

There is "a dose-response relationship between mean children's IQ and group-level fluoride exposure measures."……and that… “associations for drinking water appeared to be non-linear and associations for urine appeared to be linear.”

Comments:

- Associations for studies using urine fluoride exposure levels appeared to be linear only for all studies combined, and not for the low risk-of-bias studies. The relationship is non-linear for low RoB studies and in fact, the non-linear relationship appears to include a supra-linearity component when looking at low RoB studies of fluoride from both water and urine studies (See e-Table 4), indicating a dose-response curve that corresponds to greater effects at low doses than implied by linearity. For example, drinking water studies with a fluoride level of <2 mg/L (beta=-0.34) have a greater change in SMD than studies with a fluoride level of 4 mg/L. (beta=-0.22). There was no physiologic explanation provided.

Response: Disagree (no change)

- Because of small difference in AICs between the different models, and for ease of interpretability, only the linear model results were reported for the low risk-of-bias studies, so it is unclear why [redacted] says that the low risk-of-bias studies have non-linear relationships or what they mean by evidence of "supra-linearity." However, we do not find it surprising, considering the heterogeneity discussed earlier, that doubling the number of studies included in the <4-mg/L model would result in a different beta coefficient. We do not consider the cited example as convincing evidence of supra-linearity.

6b.DD:

- Regarding the mechanistic or physiologic explanation of supra-linear relationships between environmental measures and IQ, Bowers and Beck (2006) asserted that, "one must take care when interpreting statistical relationships with unexpected results that have no apparent underlying biological or other scientific basis… [and that] consistency of findings in numerous epidemiological studies is an insufficient basis for concluding that the finding is of biological significance, *as all studies share a common alternative explanation*."

Response: Disagree (no change)

- We disagree with the implication that the results were unexpected or have no apparent underlying biological or other scientific basis. Our hypothesis, that higher fluoride exposure would be associated with lower IQ in children, was based on established scientific evidence. This evidence includes the National Research Council's 2006 report (NRC 2006), two previous meta-analyses by Choi et al. (2012) and Duan et al. (2018), and the prepublication 2022 NTP Monograph.

6b.EE:

From the manuscript:

"We also examined whether there was a dose-response relationship at lower exposure levels that corresponded with the U.S. Environmental Protection Agency drinking water standards and World Health Organization drinking water guidelines."

Comment(s):

- Fluoride levels and their studies included in the dose-response assessment appear to be overlapping (eTable 4), i.e., studies with <1.5 mg/L are also included in the < 2mg/L group, and studies with <2mg/L and <1.5mg/L are included in the <4mg/L category. When exposure categories overlap, interpretation of results comparing fluoride exposures and IQ outcomes difficult or impossible.

Response: Disagree (edited for clarity)

- [redacted] may be misinterpreting what the different columns in eTable 4 (and eTable 5) represent. Each column presents results from a separate dose-response analysis restricted to a specified fluoride exposure range [i.e., all data, <4 mg/L (i.e., 0 to <4 mg/L), <2 mg/L (i.e., 0 to <2 mg/L), and <1.5 mg/L (i.e., 0 to <1.5 mg/L). Each row of the

two tables report dose-response results by statistical model used (i.e., linear, quadratic, or restricted cubic spline) for each of these different fluoride exposure ranges. It would be incorrect to interpret eTables 4 and 5 as each row showing results from one dose-response analysis where a trend across different exposure groups (i.e., different columns) could be evaluated. The following new text in the supplemental materials describes eTable 4 as providing results of different dose-response analyses based on restrictions to various fluoride exposure ranges:

"When analyses were restricted to exposed groups with <4 mg/L (i.e., 0 to <4 mg/L) fluoride in drinking water (n = 21 publications [6 low and 15 high risk-of-bias studies]), there was a statistically significant inverse association between fluoride exposure and children's IQ (SMD: −0.22; 95% CI: −0.27, −0.17; p-value < 0.001) (eTable 4). When restricted to <2 mg/L (i.e., 0 to <2 mg/L) in drinking water (n = 7 publications [3 low and 4 high risk-of-bias studies]), the magnitude of the effect estimate did not substantially change (SMD: −0.15; 95% CI: −0.41, 0.12; p-value = 0.274). However, when restricted to exposed groups with <1.5 mg/L (i.e., 0 to <1.5 mg/L) in drinking water (n = 7 publications [3 low and 4 high risk-of-bias studies]), there was no longer an association between fluoride in drinking water and children's IQ (SMD: 0.05; 95% CI: −0.36, 0.45; p-value = 0.816). When analyses were further restricted to low risk-of-bias publications at <4 mg/L, <2 mg/L, and <1.5 mg/L, the associations remained in the same direction and were larger in magnitude compared to when data from both low and high risk-of-bias studies were combined (eTable 4 and eTable 5)."

6b.FF:

- It is also valuable to establish the lowest fluoride dose that can trigger a response. According to the supplement (p.33), "when assessment was restricted to exposed groups with <1.5 mg/L in drinking water for all studies and again for low risk of bias studies there was no longer an association between fluoride in drinking water and children's IQ." However, at this low dose, there seems to be F-IQ association in urine studies. This appears contradictory since urine studies represent measurement of total fluoride intake. i.e., for individuals with fluoride level of 1.5 mg/L in urine, fluoride level in drinking water must be less than 1.5 mg/L, since intake from water is only a portion of the total intake.

  Response: Disagree (no change)

  - Establishing the "lowest fluoride dose that can trigger a response" is beyond the purpose or scope of this analysis. Also, the data used in the *dose-response meta-analyses* assessing <1.5 mg/L fluoride in urine and <1.5 mg/L fluoride in drinking water come from completely different sets of studies and have different total sample sizes (2,935 and 4,317, respectively), which could easily account for differences in the results. As [redacted] previously pointed out, the planned analyses should be followed and post-hoc explorations should be limited, which is what we did.

References [see missing and additional references below]

1. Harrer M, Cuijpers P, Furukawa TA, Ebert DD. Doing meta-analysis with R: a hands-on guide.

2. Higgins JP, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA, editors. Cochrane handbook for systematic reviews of interventions. John Wiley & Sons; 2019 Sep 23.
3. Schneeweiss S. Sensitivity analysis and external adjustment for unmeasured confounders in epidemiologic database studies of therapeutics. Pharmacoepidemiol Drug Saf. 2006 May;15(5):291-303. doi: 10.1002/pds.1200. PMID: 16447304.
4. Shi, L., & Lin, L. (2019). The trim-and-fill method for publication bias: practical guidelines and recommendations based on a large database of meta-analyses. *Medicine*, *98*(23), e15987. https://doi.org/10.1097/MD.0000000000015987
5. https://opal.latrobe.edu.au/articles/journal_contribution/Heterogeneity_and_subgroup_analysis/6818882
6. Miranda GH, Alvarenga MO, Ferreira MK, Puty B, Bittencourt LO, Fagundes NC, Pessan JP, Buzalaf MA, Lima RR. A systematic review and meta-analysis of the association between fluoride exposure and neurological disorders. Scientific reports. 2021 Nov 22;11(1):1-5.
7. Atal I, Porcher R, Boutron I, Ravaud P. The statistical significance of meta-analyses is frequently fragile: definition of a fragility index for meta-analyses. J Clin Epi 111 (2019) 32-40.

Missing References

The following references were cited in the comments without a citation.

- Cohen, 1988

  Possible citation: Cohen, 1988: Statistical Power for the Behavioral Sciences. Second Edition. Lawrence Erlbaum Associates, Publishers. 567 pages.

- Bowers and Beck, 2006

  Possible citation: Bowers TS and Beck BD 2007. What is the meaning of non-linear dose-response relationships between blood lead concentrations and IQ? Neurotoxicology 27 (4) 520-524.

- Mickael and Merja, 2021

  Unclear which reference the comments were referring to.

- Egger et al., 2008

  Unclear which reference the comments were referring to.

Additional References

Anglemyer A, Horvath HT, Bero L. 2014. Health care outcomes assessed with observational study designs compared with those assessed in randomized trials. Cochrane Database of Systematic Reviews, Issue 4; Art No.: MR000034. DOI: 10.1002/14651858.MR000034.pub2.

Arroyave WD, Mehta SS, Guha N, Schwingl P, Taylor KW, Glenn B, Radke EG, Vilahur N, Carreón T, Nachman RM, Lunn RM. 2021. Challenges and recommendations on the conduct of systematic reviews of observational epidemiologic studies in environmental and occupational health. J Expo Sci Environ Epidemiol. 31(1):21-30.

Benson K, Hartz A. 2000. A comparison of observational studies and randomized, controlled trials. New England Journal of Medicine. 342(25):1878-86.

Choi AL, Sun G, Zhang Y, Grandjean P. 2012. Developmental fluoride neurotoxicity: A systematic review and meta-analysis. Environ Health Perspect. 120:1362-1368.

Deeks, JJ, Higgins JPT, Altman, DG on behalf of the Cochrane Statistical Methods Group (2021) Cochrane Handbook Chapter 10: Analysing data and undertaking meta-analyses https://training.cochrane.org/handbook/current/chapter-10

Duan Q, Jiao J, Chen X, Wang X. 2018. Association between water fluoride and the level of children's intelligence: A dose-response meta-analysis. Public Health. 154:87-97. https://doi.org/10.1016/j.puhe.2017.08.013

Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). 2021. Cochrane Handbook for Systematic Reviews of Interventions version 6.2 (updated February 2021). Cochrane. Available from www.training.cochrane.org/handbook

Iheozor-Ejiofor Z, Worthington HV, Walsh T, O'Malley L, Clarkson JE, Macey R, Alam R, Tugwell P, Welch V, Glenny A-M. 2015. Water fluoridation for the prevention of dental caries. Cochrane Database Syst Rev (6):CD010856.

National Research Council (NRC). 2006. Committee on fluoride in drinking water, board on environmental studies and toxicology. Fluoride in drinking water: A scientific review of EPA's standards.: National Research Council. http://www.nap.edu/catalog/11571/fluoride-in-drinking-water-a-scientific-review-of-epas-standards.

Schwingshackl, L., Balduzzi, S., Beyerbach, J., Bröckelmann, N., Werner, S. S., Zähringer, J., Nagavci, B., & Meerpohl, J. J. 2021. Evaluating agreement between bodies of evidence from randomised controlled trials and cohort studies in nutrition research: meta-epidemiological study. BMJ (Clinical research ed.), 374, n1864. https://doi.org/10.1136/bmj.n1864

Steenland K, Schubauer-Berigan MK, Vermeulen R, Lunn RM, Straif K, Zahm S, Stewart P, Arroyave WD, Mehta SS, Pearce N. 2020. Risk of Bias Assessments and Evidence Syntheses for Observational Epidemiologic Studies of Environmental and Occupational Exposures: Strengths and Limitations. Environ Health Perspect. 128(9): 095002. https://doi.org/10.1289/EHP6980

Till C, Green R, Grundy JG, Hornung R, Neufeld R, Martinez-Mier EA, Ayotte P, Muckle G, Lanphear B. 2018. Community water fluoridation and urinary fluoride concentrations in a national sample of pregnant women in Canada. Environ Health Perspect. 126(10):107001. https://doi.org/10.1289/ehp3546

In June 2022, the [redacted] provided comments to NIEHS/DNTP on the prepublication 2022 *NTP Monograph on the State of the Science concerning Fluoride Exposure and Neurodevelopmental and Cognitive Health Effects: A Systematic Review* and a draft manuscript on a meta-analysis of fluoride exposure and IQ in children. This document contains a subset of the overall [redacted] comments related to the meta-analysis manuscript along with the NIEHS/DNTP responses. The meta-analysis-related comments from the [redacted] are reproduced here in black text, and the NIEHS/DNTP responses have been inserted in blue text following each of the comments beginning with the word "**Response**" in bold font. Formatting has been applied to aid in reading.

[redacted]

June 2022

Association between fluoride exposure and children's intelligence: A systematic review and meta-analysis

7.A:

1) (Questions control) In science, one of the hardest things to do is to frame and ask the question to have not necessarily the intended impact, but the optimal impact. In this case, the study examines a previously reported association between fluoride exposure and children's intelligence. The next step particular to federal research is not affirming prior association but building a model to examine the overall cost benefit of fluoride exposure and oral health, given that a decrement in intelligence might be a factor (among others). The proposed analysis seems like the next academic step rather than a federal, public health grounded one. Ideally, the next step could be tackled in a larger context of overall well-being and include dimensions of behavioral health (mental health and substance use challenges) and address the challenges and indeed meaning of measuring intelligence. Updating the evidence base without shifting to the relevant more public health question presents communications and policy risks that might actually decrease overall health.

Response: Agree (no change)

- We agree that the question of whether exposure to fluoride at any level can influence cognitive and neurobehavioral development is not new. The prepublication 2022 NTP Monograph and meta-analysis manuscript point out the evolving concern over this issue by first describing a prior 2006 review of this question by a committee convened by the National Academy of Sciences (NRC 2006). The monograph and meta-analysis manuscript go on to describe and further evaluate the rapidly expanding database of human epidemiological studies with improved quality and precision. The prepublication 2022 *NTP Monograph on the State of the Science Concerning Fluoride Exposure and Neurodevelopmental and Cognitive Health Effects: A Systematic Review* provides the most comprehensive assessment of this literature to date and explains the reasoning behind our determination of moderate confidence in the evidence base for an association between higher fluoride exposures and lower IQ in children.
- However, we do not agree that, prior to this assessment, federal research had affirmed, or in fact even formally examined, the question of whether fluoride exposure could lead to decrements in cognition or neurodevelopment.

- We do agree that a federal effort to examine the overall cost-benefit (or risk-benefit) of current fluoride exposure and oral health is an appropriate next step, and there is a precedent for this. In 2010, accumulating evidence of increasing prevalence of dental fluorosis from the CDC National Survey of Oral Health in U.S. School Children, and an earlier CDC Division of Oral Health publication estimating the attributable risk of water fluoridation to dental fluorosis among children (Griffin et al. 2002), prompted the Public Health Service to convene a panel to suggest a strategy to reverse this trend. The committee met multiple times over several months and ultimately proposed to decrease the recommended level of fluoride added to community water systems from a range of 0.7 to 1.2 mg/L (depending on ambient temperature) to a single consistent level of 0.7 mg/L (HHS 2015). Note, that the 2002 CDC publication (Griffin et al. 2002) that was used as the genesis for the Public Health Service panel, did not attempt to examine the cost-benefit relationship between reducing fluorosis and concomitantly diminishing oral health. Rather, the 2002 CDC publication pointed out and quantified the problem, as do our prepublication 2022 NTP Monograph and meta-analysis manuscript, respectively.
- The meta-analysis was a recommendation of the NASEM Committee that evaluated an earlier (2019) draft of the NTP Monograph. In a subsequent review, the Committee provided constructive criticisms of the meta-analysis that we performed. The current version of the meta-analysis manuscript has been revised in response to NASEM Committee suggestions and provides a quantitative estimate of risk. In addition to the prepublication 2022 NTP Monograph, the results of our meta-analyses would be a necessary component of a comprehensive effort to quantify risks in any larger public health risk-benefit evaluation of fluoride. Furthermore, NIEHS/DNTP has provided comprehensive responses to the Committee's comments (see Sup01_Meta-analysis for responses to the NASEM Committee's comments on the meta-analysis) and considers the meta-analysis manuscript is ready for, and appropriate for, submission for further peer review by a journal.

7.B:

2) (Methods) Others have commented on the quality of the research included in the meta-analysis, generalizability to the U.S. populous, and statistical methodology. [redacted] recognizes the value in these constructive criticisms.

Response: Agree (no change requested)

- Like [redacted], we recognize the value of constructive comments. Our responses to those comments have addressed the concerns that were raised, and the meta-analysis manuscript has been improved through responding to peer-review comments.

7.C:

3) (Casual inference) The strength of the work rests in strong grounding in toxicology science but yet the focus on the independent variable is out of balance with a parallel focus on the dependent one, intelligence. The science of measuring intelligence is vast and complex, and its meaning in living a healthier life, unclear. That perspective is absent from the question structure and thin in the concluding interpretation. Although the authors make a case with references 46-51, the larger net effect given issues of oral health are not incorporated, nor seem to be a

principal motivation of prospective question formation, critical in population-based studies. This approach risks being a one-sided toxicology story without the balance of the harm from decreased oral health.

Response: Disagree (no change)

- While we recognize that the science of measuring intelligence is complex, the field has evolved to become more standardized in many respects (see NIEHS 2022), and psychometric test results have played increasingly important roles in the regulation of environmental neurotoxins such as methylmercury (EPA 2001). We also recognize that further examination of the relationship between cognitive and oral health effects related to fluoride would be valuable, as we are unaware of any population-based study that has attempted to assess both the benefits of decreasing fluoride exposures on improved cognition and the concomitant potential risks to oral health. However, as indicated above, the results of our analyses would be necessary components of a comprehensive effort to quantify risks in any larger public health risk-benefit evaluation of fluoride. It is our view that the topic is of such high public health importance that the integration of our confidence assessment of the complete evidence base on increased fluoride exposure and neurodevelopmental and cognitive health effects with an assessment of the potential risk to oral health from decreased fluoride exposure would require a collective effort by the larger public health community that also considers the appropriate method and timing of population exposures to fluoride to benefit oral health.

7.D:

4) (Communication science) It is hard to definitively discern the applicable operative range of exposure levels and how they correspond to exposure in children in the United States. While there is mention of the issue in stating there are levels over 2 mg/L occurring through natural exposure, this not referenced, mapped, along with more local quantitative assessment of the burden in the United States in a fashion that readily enables an interpretive impact. The monograph makes clear this is a low percentage of people.

Response: Agree (change made)

- We agree that the lack of both U.S. studies of fluoride exposures in relation to children's cognition and the absence of publicly available U.S. data on total fluoride exposures for children make it difficult to directly apply our findings to fluoride exposures in the United States. The absence of U.S. studies is currently identified as a limitation in the meta-analysis manuscript. In response to this comment, we have updated the reference to CDC data and added a citation to the manuscript for a publication that maps fluoride concentrations in untreated groundwater from public supply and domestic wells (McMahon et al. 2020).

7.E:

5) (Limitations, bias) Methods do not clearly delineate the question history and approach to multiple comparisons (other measures of cognition, other age groups) which ultimately might undermine the integrity of the measured inference. This is an important limitation not described.

Response: Disagree (no change)

- We are unclear about what [redacted] is identifying as a limitation. Adjustments for multiple tests are not commonly used in meta-analyses and the Cochrane Handbook advises against their use (Section 16.7.2). We did not rely solely on p-values when describing results, and the prespecified subgroup analyses included stratification by intelligence assessment type. In addition, many of the studies in the meta-analysis specifically excluded children with obvious cognitive disabilities, and the findings with respect to IQ deficits were reported over an age range from as young as 3–4 years old to as old as 16–18 years old, suggesting that deficits persist. Measures of neurodevelopmental and cognitive effects other than IQ, such as ADHD behaviors, were evaluated in relation to fluoride exposure in the prepublication 2022 NTP Monograph. However, there was low confidence in the evidence of an association between these other effects and fluoride, suggesting that other measures of neurodevelopment and cognition were not responsible for the IQ findings.

7.F:

6) (Statistics, math) Is there an assessment and explanation of the variation in outcomes as it might affect the interpretation of the statistical measurement (drawing a line with a positive slope through a scattergram).

Response: Agree (no change)

- The authors of studies with individual-level measures of exposure and outcome frequently attempted to apply linear and non-linear regression models to determine the best fit to the scatterplots. In a meta-analysis, a funnel plot illustrating the effect estimate against the inverse of the standard error is equivalent to the scattergram suggested by [redacted]. Our *regression slopes meta-analysis* found that non-linear models did not provide a significantly better fit than linear models and, therefore, we elected to accept an assumption of linearity for the purposes of discussion. The overall pooled effect estimate was determined from studies that reported individual urine levels, which is considered a reasonable estimate of total exposure to fluoride from all sources. Presumably, this effect estimate would be more precise than the effect estimates derived from the individual studies.

7.G:

7) Each point begins with the type of concern raised listed in parentheses. The details of each point is one example of each type, where upon are at times others.

Response: Agree (no change requested)

- This comment describes the structure of the [redacted] comments in this file, which we have found helpful. We have responded to each comment above.

Note: The [redacted] comments on the prepublication 2022 NTP Monograph are not reproduced here as they are not relevant to the meta-analysis. See DocD_Monograph for the monograph-relevant comments and responses.

References

Griffin SO, Beltrán ED, Lockwood SA, Barker LK. 2002. Esthetically objectionable fluorosis attributable to water fluoridation. Community Dent Oral Epidemiol. 30(3):199-209. doi: 10.1034/j.1600-0528.2002.300306.x. PMID: 12000343.

McMahon PB, Brown CJ, Johnson TD, Belitz K, Lindsey BD. 2020. Fluoride occurrence in United States groundwater. Sci Total Environ. 732:139217. doi:https://doi.org/10.1016/j.scitotenv.2020.139217

National Research Council (NRC). 2006. Committee on fluoride in drinking water, board on environmental studies and toxicology. Fluoride in drinking water: A scientific review of EPA's standards.: National Research Council. http://www.nap.edu/catalog/11571/fluoride-in-drinking-water-a-scientific-review-of-epas-standards.

U.S. Department of Health and Human Services (HHS) Federal Panel on Community Water Fluoridation. 2015. U.S. Public Health Service Recommendation for Fluoride Concentration in Drinking Water for the Prevention of Dental Caries. Public Health Reports. July-August;130: 318-331.

U.S. Environmental Protection Agency (EPA). 2001. Methylmercury (MeHg); CASRN 22967-92-6: Chemical Assessment Summary. Retrieved from: https://iris.epa.gov/ChemicalLanding/&substance_nmbr=73

In November 2021, the [redacted] provided comments to NIEHS/DNTP on the 2021 *Draft NTP Monograph on the State of the Science concerning Fluoride Exposure and Neurodevelopmental and Cognitive Health Effects: A Systematic Review* (the NTP Monograph) and a draft manuscript on a meta-analysis of fluoride exposure and IQ in children (the meta-analysis manuscript). NIEHS/DNTP prepared responses and shared those responses back to [redacted] in April 2022.

In July 2022, the [redacted] provided two sets of comments to NIEHS/DNTP, again on the NTP Monograph and the meta-analysis manuscript.

- The first set of [redacted] comments were provided as a new layer of input on top of the original [redacted] comments (from November 2021) and NIEHS/DNTP responses. This document contains a subset of the overall [redacted] comments (from November 2021 and July 2022) related to the meta-analysis manuscript along with the NIEHS/DNTP responses. The meta-analysis-related comments from the [redacted] are reproduced below in black text and the NIEHS/DNTP responses have been inserted in blue text following each of the comments beginning with the word "**Response**" in bold font.

- The second set of the [redacted] comments were provided in track changes embedded in the draft meta-analysis manuscript in a Microsoft Word document. The full text of [redacted] comments has been reproduced below verbatim in black text along with the specific sentence referred to by [redacted] as quotes under a heading for the specific section of the document (e.g., "Abstract section"). When the [redacted] comments were inserted on a particular word or phrase, that word or phrase is highlighted in grey within the quoted sentence. Again, the NIEHS/DNTP responses have been added in blue text following each of the comments beginning with the word "**Response**" in bold font.

Formatting has been applied to aid in reading

[redacted] comments from November 2021 and July 2022

Summary of [redacted] comments on the *"Draft NTP monograph on the state of the science concerning fluoride exposure and neurodevelopmental and cognitive health effects: a systematic review"* ("SoS document") and draft Taylor *et al. Association between fluoride exposure and children's intelligence: A systematic review and meta- analysis* manuscript ("meta-analysis document")

Note: [redacted] comments on the monograph are not reproduced here as they are not relevant to the meta-analysis. See DocB1_Monograph for the monograph-relevant comments and responses.

8.A:

- [redacted] **comment on SoS document (November 2021):** The revised NTP monograph seems to address concerns from prior comments as NTP removed the hazard assessment and is now calling this a "state of the science" document. However, the meta-analysis that NTP removed from the original monograph is now being published independently. Although it will be in a scientific review publication (JAMA pediatrics), [redacted] think that this may raise questions regarding exposure levels and neurodevelopmental effects, as the publication does not seem to put the exposure levels into context.

  Response: Disagree (no change)
  - We appreciate the need to provide context concerning fluoride exposure levels and neurodevelopmental effects and presume that this comment concerns fluoride

exposures in the United States. As the comment points out, this topic is more fully addressed in the NTP state of the science monograph, which is referenced in the meta-analysis, and we have added reference to the U.S. Public Health Service recommendations for optimal water fluoridation in the meta-analysis manuscript; however, we also stress that the subject of our fluoride monograph and meta-analysis is total fluoride exposures from all sources. The November 2022 literature search update of the meta-analysis includes a number of new non-U.S. studies that further inform the relationship between IQ deficits in children and exposures to fluoride that were not available for inclusion in the 2020 draft NTP monograph. These studies have provided additional information to sharpen the dose-response mean-effects estimates and improve the *regression slopes meta-analysis*. Although the clarity of effects at lower fluoride exposures, which are presumed to be applicable to exposures in the United States, is improving, providing further context is speculative because there are no studies of the potential association between fluoride exposures and IQ in children in the United States, and nationally representative urinary fluoride levels are not available. These facts make it difficult to make more specific statements about the relevance of our meta-analysis findings to the U.S. population.

8.B:

- [redacted] **comment on SoS document (July 2022):** The systematic review finds, with moderate confidence, that higher fluoride exposure [e.g., represented by populations whose total fluoride exposure approximates or exceeds the WHO Guidelines for Drinking-water Quality of 1.5 mg/L of fluoride (WHO 2017)] is consistently associated with lower IQ in children, and that more studies are needed to fully understand the potential for lower fluoride exposure to affect children's IQ. In this regard:

  a) What is the overall confidence in the conclusions of the meta-analysis? No study was excluded from the meta-analysis based on concerns for risk of bias – how does this affect the overall confidence in the conclusions?

  Response: No change requested

  - The meta-analysis itself does not have confidence conclusions, but the finding of moderate confidence in the body of evidence reviewed in the prepublication 2022 NTP Monograph has been added to the *Discussion* section of the revised meta-analysis manuscript as follows:

    "This meta-analysis complements a larger systematic review[8] that concluded moderate confidence in the body of evidence that fluoride exposure is associated with lower IQ in children."

  - While the meta-analysis includes more studies than the prepublication 2022 NTP Monograph, which resulted from a literature search update for the meta-analysis in November 2021, our review of these additional studies has given us no reason to believe that they would either increase or decrease our confidence in that body of evidence.

8.C:

b) For the dose-response mean-effects meta-analysis and regression slopes meta-analysis, were subgroup analyses stratified by risk of bias (high or low), study location (e.g., country), outcome assessment, exposure matrix (e.g., urine or water), pre- or post-natal exposure, sex, and age group conducted? If not, is there a reason why?

Response: No change requested

- The *dose-response meta-analysis* was stratified by risk of bias and exposure level as was pre-specified in the protocol. Because the purpose of the subgroup analyses was to explore sources of heterogeneity, and the *dose-response meta-analysis* included many of the same studies included in the *mean-effects meta-analysis*, there was no reason to add further subgroup analyses post-hoc. The *regression slopes meta-analysis* was stratified by risk of bias, exposure type, country, outcome assessment type, sex, and pre- and post-natal exposure.

8.D:

- [redacted] **comment on SoS and meta-analysis documents (July 2022):** [redacted] raised concerns regarding exposure measurement in previous comments. The current Discussion sections in each document cover some exposure measurement limitations but may not sufficiently address [redacted] previous comments or other important issues potentially impacting individual and group urinary fluoride measurement, such as variation in period of urine collection, variations/transient increases in excretion, variations in clearance times, as well as total fluoride exposure by age, sex, developmental stage, and over time.

  Note: The above comment refers to previous [redacted] concerns regarding exposure measurement that were focused on the monograph and are, therefore, not reproduced here. See DocB1_Monograph for the monograph-relevant comments and responses. The meta-analysis-relevant response is provided immediately below.

  Response: Disagree (edited for clarity)

  - In responses to earlier comments from [redacted] and others, we have pointed out reasons we consider these concerns are overstated and speculative. We have addressed exposure measurements as part of the evaluations in both documents. These include our requirement for creatinine or specific gravity adjustments for measurements of urinary fluoride to be considered lower risk of bias for exposure. We also cite studies reporting reasonable agreements between 24-hour and repeated volume corrected spot urine fluorides in the monograph. We also would point out that to account for the consistent direction of effect of an inverse relationship between fluoride in urine and children's IQ would require that one or all of the cited factors would need to affect children's IQ, as well as produce the speculated spurious correlated fluoride measurements. We are happy to entertain such evidence if [redacted] wish to provide.
  - However, we acknowledge that the type and timing of urinary sample collection is important to consider and have extended the *Discussion* section of the meta-analysis to acknowledge concerns related to the issues associated with individual and group urinary fluoride measurement:

"Another limitation of the mean-effects meta-analyses is that exposure values are assumed to be the same for each child in an exposure group, either because the study used a community-level water fluoride measure or a median, mean, or midpoint in water or urine as the exposure value. Fluoride exposure may vary considerably depending on individual behaviors and is best captured by individual-level measures of total exposure, such as urinary fluoride measures. Because drinking water measures capture only some of a person's total exposure to fluoride, it is reasonable to assume that some children in the meta-analysis had higher exposure to fluoride and those children may have skewed the mean IQ deficits of the entire group. Urinary fluoride levels include all ingested fluoride and are considered a valid measure to estimate total fluoride exposure.[61, 62] When compared with 24-hour urine samples, spot urine samples are more prone to the influence of timing of exposure (e.g., when water was last consumed, when teeth were last brushed) and can also be affected by differences in dilution. However, correlations between urinary fluoride concentrations from 24-hour samples and spot samples adjusted for urinary dilution have been described,[63] and with one exception[35] all studies in the regression slopes meta-analysis, accounted for dilution."

8.E:

- [redacted] **comment on meta-analysis document (July 2022):** Given that in the Results section, heterogeneity was evaluated and found to be high, suggest that the Discussion section should address those findings with some coverage of potential sources of high heterogeneity. This would be consistent with the objectives outlined in the cited protocol [*National Toxicology Program (NTP). Protocol for systematic review of effects of fluoride exposure on neurodevelopment. U.S. Department of Health and Human Services, Public Health Service, National Institutes of Health; 2020*], where it is indicated that one of the specific aims of the meta-analysis is to "[s]ynthesize the evidence across studies that assessed learning and memory using a narrative approach or meta-analysis (if appropriate) and evaluate sources of heterogeneity."

  Response: Agree (change made)

  - In the *Discussion* section of the manuscript, we added the following new text with further interpretations of the subgroup analyses as they relate to potential sources of heterogeneity:

    "With a couple exceptions, the subgroup analyses in the mean-effects meta-analysis did not explain a large amount of the overall heterogeneity. However, the heterogeneity in the regression slopes meta-analysis was explained by subgroup analyses. This suggests that the aggregate nature of the mean-effects meta-analysis might not be sufficiently sensitive to capture potential sources of heterogeneity, as seen possible when using studies with individual-level data in the regression slopes meta-analysis. However, the large number of studies included in the mean-effects meta-analysis and the consistency in the direction of the association across the analyses make this is less of a concern."

  - We also further investigated potential sources of heterogeneity by conducting a meta-regression analysis using mean age in years and year of publication in each study. In the supplemental material we added:

"The results of the meta-regression models indicate that year of publication and mean age of study children did not explain a large degree of heterogeneity as neither were significant predictors of the relationship between fluoride and children's intelligence, and the residual I^2 remained high (85% and 87%, respectively). Year of publication (SMD = 0.01, 95% CI: -0.01, 0.02) and mean age (SMD = -0.04, 95% CI: -0.13, 0.04) explained relatively little between-study variance (adjusted R^2 of 12% and 5%, respectively). When both year of publication and mean age were included in the model, there were no notable improvements to the amount of between-study variance explained (adjusted R^2 = 13%) or percent residual variation due to heterogeneity (residual I^2 = 85%).

Excluding the outlier study34 resulted in a slightly lower heterogeneity for the overall effect estimate (I^2=84%) and for the India-specific effect estimate (I^2=69%). The meta-regression indicates that mean age is a significant predictor of the effect (SMD = -0.06, 95% CI: −0.12, −0.01, p-value =0.025), explaining 9% of the between-study variance. Year of publication (SMD = 0.01, 95% CI: 0.001, 0.02, p-value=0.028) explained a larger degree of between-study variance (R^2 = 19 %)."

[redacted] comments embedded in the Microsoft Word meta-analysis document July 2022

[redacted] *general comments*

8.F: Abstract Section: "Water and water-based beverages are the main source of systemic fluoride intake; however, an individual's total exposure to fluoride also reflects contributions from other sources such as food, dental products, industrial emissions, and some pharmaceuticals and pesticides."

[redacted] **comment:** Although this statement is true, it may relay a misleading impression that the authors also measured these other sources of exposure in the study. The authors may want to check in the pertinent EPA documentation whether this [pesticides] is a significant source of exposure worth mentioning here since sulfuryl fluoride is a fumigant and dissipates rapidly. [**Note**: the text in brackets has been added by NIEHS/DNTP to clarify [redacted] comment.]

Response: Agree (change made)

- We agree that sulfuryl fluoride is a minor contributor to a person's total fluoride exposure and have removed "pesticides" from the list of contributors to fluoride exposure in the manuscript.

8.G: *Abstract section:* (the underlined text was inserted by [redacted]) **"**To perform a systematic review and meta-analysis to investigate associations between fluoride exposure, based primarily on urinary and water fluoride levels, and children's intelligence."

[redacted] **comment:** [redacted] suggested [redacted] add "based primarily on urinary and water fluoride levels" to the sentence.

Response: Disagree (no change)

- The objective describes the intent of the systematic review, which was to include studies that captured any source of fluoride exposure, not just from urinary and water

levels. However, our results make it clear that urinary and water fluoride levels were the main measures of fluoride exposure in the relevant studies.

8.H: ***Results section:*** *"*For studies that had more than one exposed group (n = 17), a sensitivity analysis was performed to evaluate the impact of combining all exposed groups and comparing them to the reference group."

[redacted] **comment:** A decision was made to use the highest exposed groups in comparison to the reference group. Was a sensitivity analysis performed on this decision?

Response: Agree (edited for clarity)

- We have revised this sentence to clarify that the sensitivity analysis to evaluate the impact of combining all exposed groups to compare them to the reference group was applied to all studies in the *mean-effects meta-analysis* and not limited to these 17 studies. We clarified this aspect in the Results section as follows:

  "The sensitivity analysis to evaluate the impact of combining all exposed groups and comparing them to the reference group did not appreciably change the effect estimates".

- However, we are unclear if [redacted] is suggesting an additional sensitivity analysis using the second highest exposed group compared to the reference. If so, because the number of studies that have a second highest exposed group (n=17) is the same as the sensitivity analysis combining all exposed groups, it is unlikely such an analysis would provide further value.

8.I: ***Results section:*** *"*The dose-response mean-effects meta-analysis combining data from 29 studies with group-level fluoride measurements in drinking water (23 high risk-of-bias and 6 low risk-of-bias studies) and 17 studies with children's group-level mean urinary fluoride levels (9 high risk-of-bias and 8 low risk-of-bias studies) show statistically significantly lower children's IQ scores with increasing fluoride exposures."

[redacted] **comment:** Is the small number of low risk of bias studies [in the drinking water dose-response meta-analysis] of concern? [**Note**: the text in brackets has been added by NIEHS/DNTP to clarify [redacted] comment.]

Response: Disagree (no change)

- We are unclear on the exact concern to which the comment is referring. The results of the *dose-response meta-analysis* are presented in the supplemental material and are not emphasized in the main manuscript. However, the six low risk-of-bias studies in the drinking water fluoride analysis includes 4,355 children and the nine low risk of bias studies in the urinary fluoride analysis includes 5,713 children. For perspective one might consider the NHANES assessments. The annual sample size for NHANES is 5,000 people (all ages) and is considered a representative sample of the United States population. The number of participants who provide biomonitoring samples is about 1/3 of that total, so it is recommended at least 4 years of data (2 NHANES cycles) be

combined to obtain a sample size with an acceptable level of reliability for most of the sampling domains. This usually works out to ~3-5k participants.

8.J: ***Results section:*** "Adjusting for possible publication bias through trim-and-fill analysis supports the conclusion that a 1-mg/L increase in urinary fluoride was associated with lower IQ, with an adjusted pooled effect estimate of −0.87 (95% CI: −1.93, 0.19; p- value = 0.302) (**eFigure 22**). The results for fluoride intake and water fluoride levels are available in **Supplemental Materials**."

[redacted] **comment:** How significant is this change [1-mg/L increase in urinary fluoride] in relation to "normal" urinary fluoride levels? Indicating what these levels are would help contextualizing this conclusion. [**Note**: the text in brackets has been added by NIEHS/DNTP to clarify [redacted] comment.]

Response: Disagree (no change)

- What constitutes "normal" urinary fluoride levels depends entirely on the population being examined. However, using data from Green et al. (2019), the difference between water fluoride concentration in a fluoridated area v. non-fluoridated area is roughly half of 1mg/L.
- Note: After updating the *regression slopes meta-analysis* with new studies from the updated literature search, there was no longer evidence of publication bias, so the specific quoted text related to the adjusted pooled effect estimate has been removed from the manuscript.

[redacted] *comments on defining "higher":*

8.K: ***Abstract section:*** "The meta-analysis of 46 studies (N = 15,538 children) with group-level exposures found that children exposed to higher fluoride levels had lower mean IQ scores (pooled SMD: −0.49; 95% CI:−0.60, −0.38; p-value < 0.001)."

[redacted] **comment:** Please specify the meaning of "higher". For example, "greater than XX mg/mL"

Response: Agree (change made)

- The *mean-effects meta-analysis* pooled results from individual studies that compared mean IQ in "higher" vs. "lower" fluoride areas. Each individual study had its own definition of what exposure level constituted "higher" and "lower", so the data do not support defining a threshold for "higher" in the pooled SMD of the *mean-effects meta-analysis*. To clarify that "higher" exposure is simply being used relative to "lower", we have revised the quoted sentence as follows (please note the numbers have changed due to a literature search update):

  "The meta-analysis of 55 studies (N = 18,845 children) with group-level exposures found that, when compared to children exposed to lower fluoride levels, children exposed to higher fluoride levels had lower mean IQ scores (pooled SMD: −0.46; 95% CI: −0.55, −0.37; p-value < 0.001)."

- For transparency, Table 1 (excerpt below) includes the exposure levels that were compared in the *mean-effects meta-analysis*. Please note that we did explore lower levels of fluoride exposure in the *dose-response meta-analysis* (see supplemental materials).

Excerpt of Table 1. Characteristics of Studies Included in the Meta-analysis

Reference[a] *Study Design*	Study Location	Age Range (Years)	Fluoride Exposure: Assessment (Metric, Exposure Groups)	Fluoride Exposure: Levels	Intelligence Assessment	Overall RoB Rating	Confounders Considered
Ren et al. (1989)[66] [translated in Ren et al. 2008][me, o] *Cross-sectional*	China	8–14	No fluoride measurement Low iodine village/high fluoride and low iodine village	Not specified	Wechsler Intelligence Scale for Children	High	Sex; iodine
Chen et al. (1991)[68] [translated in Chen et al. 2008][me, w] *Cross-sectional*	China	7–14	Drinking water Nonendemic/endemic fluorosis village	0.89 mg/L (nonendemic) 4.55 mg/L (endemic)	Chinese Standardized Raven Test	High	Age; sex
Guo et al. (1991)[70] [translated in Guo et al. 2008a][me, o] *Cross-sectional*	China	7–13	Serum Reference area using wood/coal burning-related fluoride endemic area	0.1044 ± 0.0652 mg/L (reference) 0.1483 ± 0.0473 mg/L (endemic)	Chinese Binet Intelligence Test	High	Age; sex; SES

8.L: ***Abstract section:*** (the underlined text was inserted by [redacted] and the strikethrough text was deleted by [redacted]) "Our meta-analysis confirms results of previous meta-analyses and extends them by including newer, more precise studies with individual-level exposure measures to assess associations between high fluoride exposure (e.g., >1.5 mg/L, the World Health Organization Guideline for Drinking-water Quality) and lower IQ levels of children. Associations between lower fluoride exposure (e.g., < 1.5 mg/L) and children's IQ remain uncertain. ~~The data support a consistent inverse association between fluoride exposure and children's IQ~~. Additional prospective cohort studies with individual urinary fluoride measures, along with studies conducted in the United States, would increase the confidence in this body of evidence."

[redacted] **comment:** The term "high" is used throughout the manuscript to characterize exposure and should be defined.

Response: Disagree (no change)

- [redacted] has suggested we define "high" as greater than 1.5mg/L and "low" as less than 1.5 mg/L and references 1.5 mg/L, the WHO Guideline for fluoride in drinking water. In the prepublication 2022 NTP Monograph we used this description of "higher" because, in that qualitative assessment of the epidemiology literature, the WHO Guideline represented a useful total fluoride exposure equivalent metric. However, in the meta-analysis, we were able to explore lower exposure levels by limiting the dose-response analyses to include study groups where exposure levels were equal to or lower than the U.S. Environmental Protection Agency drinking water standards (i.e., <4mg/L and <2mg/L)[20] and World Health Organization drinking water guidelines (<1.5mg/L).

8.M: ***Results section:*** *"The meta-analysis of 46 studies (37 high risk-of-bias studies and 9 low risk-of-bias studies) that provided mean IQ scores shows that children exposed to higher fluoride levels had statistically significantly lower IQ scores (random-effects pooled SMD, −0.49; 95% CI: −0.60, −0.38; p-value < 0.001) (**Figure 2**)."*

[redacted] **comment:** Please define ["higher"] [**Note**: the text in brackets has been added by NIEHS/DNTP to clarify [redacted] comment.]

Response: Agree (change made)

- To clarify that "higher" exposure is simply being used relative to "lower", we have revised the quoted sentence as follows (please note the numbers have changed due to a literature search update):

  "The meta-analysis of 55 studies (45 high risk-of-bias studies and 10 low risk-of-bias studies) that provided mean IQ scores shows that, when compared to children exposed to lower levels of fluoride, children exposed to higher fluoride levels had statistically significantly lower IQ scores (random-effects pooled SMD,¬ −0.46; 95% CI: −0.55, −0.37; p value < 0.001) (Table 2, Figure 2)."

8.N: ***Discussion section:*** *"The results of our mean-effects meta-analysis are consistent with two previous meta-analyses that reported statistically significantly lower IQ scores in children exposed to higher fluoride levels (p < 0.001) (**Table 2**)."*

[redacted] **comment:** Please define ["higher"] [**Note**: the text in brackets has been added by NIEHS/DNTP to clarify [redacted] comment.]

Response: Agree (change made)

- To clarify that "higher" exposure is simply being used relative to "lower", we have revised the quoted sentence as follows (please note the numbers have changed due to a literature search update):

  "The results of the mean-effects meta-analysis are consistent with two previous meta-analyses that, when comparing children exposed to lower fluoride levels, reported statistically significantly lower IQ scores in children exposed to higher fluoride levels (p < 0.001) (Table 2)."

8.O: ***Discussion section:*** *"Therefore*, the data support a consistent inverse association between fluoride exposure and children's IQ."

[redacted] **comment:** At all levels or only documented at certain levels? It is important to contextualize this statement.

Response: Disagree (no change)

- Based on previous comments, our interpretation of this comment is that [redacted] is recommending we contextualize the sentence with a threshold (e.g., >1.5mg/L) which is why we disagree with the comment. To answer [redacted] question, the meta-analysis includes fluoride exposures at all levels, some of which were below 1.5mg/L. Therefore, the evidence does not support excluding lower levels from this statement.

8.P: ***Discussion section:*** *"Although the estimated decreases in IQ may seem small, research on other neurotoxicants has shown that subtle shifts in IQ at the population level can have a profound*

impact on the number of people who fall within the high and low ranges of the population's IQ distribution."

[redacted] **comment:** Does this imply that fluoride causes a shift in intelligence at all levels of exposure (e.g., including at 0.7 mg/L)? If that is not the intent, this passage could be misleading.

Response: Disagree (no change)

- We do not consider this statement to be misleading. Using [redacted] example, total fluoride exposure among individuals living in optimally fluoridated areas (0.7mg/L in drinking water) may be higher than 0.7mg/L, dependent on personal behaviors and habits. We discuss the potential for this type of variation in the manuscript.

[redacted] *comments or questions on defining a threshold:*

8.Q: ***Abstract section:*** "*The meta-analysis of the association between individual-level measures of fluoride and children's IQ found a decrease of 1.58 IQ points (95% CI: −2.63, −0.53; p-value = 0.003) per 1-mg/L increase in urinary fluoride*."

[redacted] **comment:** Was there a threshold for this effect?

Response: No change requested

- Because we used a linear model, there is no threshold for this effect.
- Note: After adding new studies the sentence has been updated:

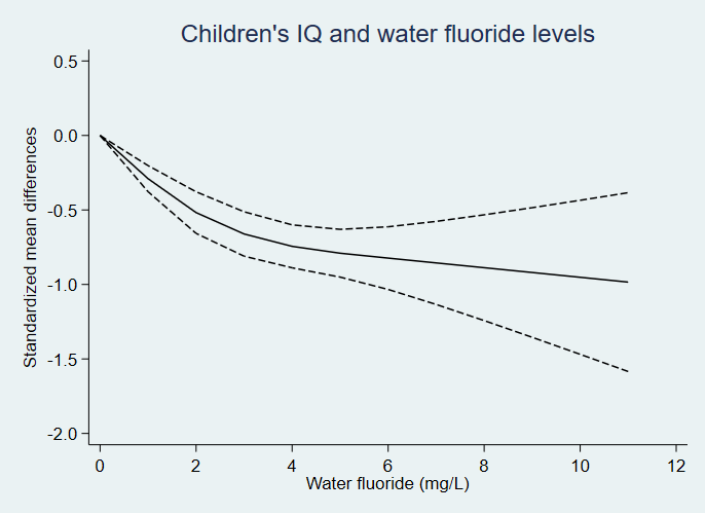
  "The meta-analysis of studies that reported individual-level measures of fluoride and children's IQ scores found a decrease of 1.81 points (95% CI: −2.80, −0.81; p-value < 0.001) per 1-mg/L increase in urinary fluoride."

8.R: ***Discussion section:*** "There is also evidence of a dose- response relationship between lower children's IQ and higher fluoride exposures."

[redacted] **comment:** Was a threshold for such relationship considered?

Response: No change requested

- As previously mentioned, results for the dose-response relationship restricted to lower fluoride exposure levels (i.e., <4mg/L and <2mg/L, <1.5mg/L) in both drinking water and urine are reported in the supplemental materials.
- The restricted cubic splines model for water fit slightly better than the linear model, however there was no obvious threshold as illustrated by the figure at either of the modelled knots.



eFigure 17. Pooled Dose-Response Association Between Fluoride in Water and Standardized Mean Differences in Children's IQ

eFigure 17 note: Water fluoride levels were modeled with quadratic restricted cubic splines terms in a random-effects model (solid line). Dashed lines represent the 95 % confidence intervals for the quadratic spline model.

8.S: ***Discussion section:*** *"Associations for drinking water appeared to be non-linear and associations for urine appeared to be linear. The Duan et al.[4] meta-analysis reported a significant non-linear dose-response relationship above 3 ppm [3 mg/L] in water."*

[redacted] **comment:** Was there a threshold for water?

Response: No change requested

- Results for the dose-response relationship restricted to lower fluoride levels (i.e., <4mg/L and <2mg/L, <1.5mg/L fluoride in drinking water) are reported in the supplemental materials.
- As described in the previous response, there was no obvious threshold for water.

8.T: ***Discussion section:*** *(underlined text inserted by* [redacted]*)* "However, among the low risk-of-bias cross-sectional studies, most provided information to indicate that exposure preceded the outcome (e.g., only including children who had lived in the area since birth, children had dental fluorosis, linked to fluoride levels greater than XX)."

[redacted] **comment:** [redacted] recommended adding "linked to fluoride levels greater than XX".

Response: Disagree (no change)

- The "e.g.," of this sentence is meant to provide examples for how cross-sectional studies provided information that establishes temporality and is not linked to any fluoride level.

8.U: ***Discussion section:*** *"*In addition, there is inconsistency in which model is the best fit at lower exposure levels (**eTable 4** and **eTable 5**) leading to uncertainty in the shape of the dose-response curve at these levels."

[redacted] **comment:** Was a threshold considered?

Response: No change requested

- See above responses concerning the various models explored for best model fit.

Supplemental Materials

NTP Monograph on the State of the Science Concerning Fluoride Exposure and Neurodevelopmental and Cognitive Health Effects: A Systematic Review

NTP Monograph 08

May 2022

National Toxicology Program
Public Health Service
U.S. Department of Health and Human Services
ISSN: 2378-5144

Research Triangle Park, North Carolina, USA

Foreword

The National Toxicology Program (NTP), established in 1978, is an interagency collaboration within the Public Health Service of the U.S. Department of Health and Human Services. Its activities are executed through a partnership of the National Institute for Occupational Safety and Health (part of the Centers for Disease Control and Prevention), the Food and Drug Administration (primarily at the National Center for Toxicological Research), and the National Institute of Environmental Health Sciences (part of the National Institutes of Health), where this virtual program is administratively located. NTP's work focuses on the testing, research, and analysis of agents of concern to identify toxic and biological effects, provide information that strengthens the science base, and inform decisions by health regulatory and research agencies to safeguard public health. NTP also works to develop and apply new and improved methods and approaches that advance toxicology and better assess health effects from environmental exposures.

Literature-based evaluations are one means by which NTP assesses whether exposure to environmental substances (e.g., chemicals, physical agents, and mixtures) may be associated with adverse health effects. These evaluations result in hazard conclusions or characterize the extent of the evidence and are published in the NTP Monograph series, which began in 2011. NTP monographs serve as an environmental health resource to provide information that can be used to make informed decisions about whether exposure to a substance may be of concern for human health.

These health effects evaluations follow prespecified protocols that apply the general methods outlined in the "Handbook for Conducting a Literature-Based Health Assessment Using the OHAT Approach for Systematic Review and Evidence Integration."[†] The protocol describes project-specific procedures tailored to each systematic review in a process that facilitates evaluation and integration of scientific evidence from published human, experimental animal, and mechanistic studies.

Systematic review procedures are not algorithms, and the methods require scientific judgments. The key feature of the systematic review approach is the application of a transparent framework to document the evaluation methods and the basis for scientific judgments. This process includes steps to comprehensively search for studies, select relevant evidence, assess individual study quality, rate confidence in bodies of evidence across studies, and then integrate evidence to develop conclusions for the specific research question. Draft monographs undergo external peer review prior to being finalized and published.

NTP monographs are available free of charge on the NTP website and cataloged in PubMed, a free resource developed and maintained by the National Library of Medicine (part of the National Institutes of Health). Data for these evaluations are included in the Health Assessment and Workspace Collaborative.

For questions about the monographs, please email NTP or call 984-287-3211.

[†]OHAT is the abbreviation for Office of Health Assessment and Translation, which has become the Health Assessment and Translation group in the Integrative Health Assessment Branch of the Division of the National Toxicology Program at the National Institute of Environmental Health Sciences.

Table of Contents

Tables

Figures

About This Review

National Toxicology Program[1]

[1]Division of the National Toxicology Program, National Institute of Environmental Health Sciences, Research Triangle Park, North Carolina, USA

Collaborators

Kyla W. Taylor, John R. Bucher, Robyn B. Blain, Christopher A. Sibrizzi, Pamela A. Hartman, Kristen Magnuson, Sorina E. Eftim, and Andrew A. Rooney

Division of the National Toxicology Program, National Institute of Environmental Health Sciences, Research Triangle Park, North Carolina, USA
Developed protocol and conducted systematic review through oversight and review of literature search, screening, data extraction, risk-of-bias assessment, and meta-analysis; wrote and edited monograph
Kyla W. Taylor, Ph.D., Project Lead
John R. Bucher, Ph.D.
Andrew A. Rooney, Ph.D.

ICF, Fairfax, Virginia, USA
Conducted literature screening, data extraction, risk-of-bias assessment, and meta-analysis; reviewed data, results, and analyses; wrote, edited, and formatted monograph; provided database and HAWC support
Robyn B. Blain, Ph.D.
Sorina E. Eftim, Ph.D.
Pamela A. Hartman, M.E.M.
Kristen Magnuson, M.E.S.M., Deputy Work Assignment Manager
Christopher A. Sibrizzi, M.P.H., Lead Work Assignment Manager

Contributors

Division of the National Toxicology Program, National Institute of Environmental Health Sciences, Research Triangle Park, North Carolina, USA
Provided oversight for external peer review
Mary S. Wolfe, Ph.D.

Provided literature search, screening, Distiller and HAWC support
Vickie R. Walker, B.S.

Reviewed meta-analysis protocol
Suril Mehta, Dr.P.H.

Conducted a technical review of the draft monograph
Freya Kamel, Ph.D. (Retired)

Conducted literature screening
Kris Thayer, Ph.D. (currently Chemical and Pollution Assessment Division, United States Environmental Protection Agency, Washington, D.C., USA)

Division of Extramural Research and Training, National Institute of Environmental Health Sciences, Research Triangle Park, North Carolina, USA
Conducted literature screening
Heather King, Ph.D.

Kelly Government Services, Research Triangle Park, North Carolina, USA
Supported external peer review
Elizabeth A. Maull, Ph.D. (retired from NIEHS, Research Triangle Park, North Carolina, USA)

ICF, Fairfax, Virginia, USA
Provided contract oversight
David F. Burch, M.E.M., Principal Investigator
Cara Henning, Ph.D.
Jessica A. Wignall, M.S.P.H.

Reviewed data, results, and analyses
William Mendez, Ph.D.

Reviewed data, results, and analyses, conducted literature screening and risk-of-bias assessment, and provided database and HAWC support
Anna Engstrom, Ph.D.

Provided database and HAWC support
Steven Black, M.P.H.
Natalie Blanton, M.P.H.
Kaitlin Geary, B.S.
Courtney Lemeris, B.A.
Rachel McGill, B.S.
Daniel Peters, B.A.
Alessandria Schumacher, B.A.
Anna Stamatogiannakis, B.S.
River Williams, B.S.

Conducted literature screening
Carlye A. Austin, Ph.D.
Ryan Cronk, Ph.D.
Lauren Fitzharris, M.P.H.
Samantha Goodman, M.S.
Angela Hensel, B.A.
Melinda Hoang, M.P.H.
Allison Killius, B.S.
Alex Kliminsky, M.S.
Shannon McGinnis, Ph.D.

Devon Morgan, M.S.
Joei Robertson, M.Sc.
Courtney Rosenthal, M.S.
Amanda Ross, M.Ed.
Karen E. Setty, Ph.D.
Kelly Shipkowski, Ph.D.
Raquel A. Silva, Ph.D.
Parnian Soleymani, M.S.

Conducted literature screening and data extraction
Katie Duke, Ph.D.
Susan Goldhaber, M.P.H.
Revathi Muralidharan, B.S.
Jennifer Seed, Ph.D.
Codi Sharp, M.P.H.

Conducted literature screening, data extraction, and risk-of-bias assessment
Johanna Rochester, Ph.D.

Conducted literature screening and risk-of-bias assessment
Pam Ross, M.S.P.H.

Conducted data extraction
Robert Shin, M.H.S.

Conducted data extraction and risk-of-bias assessment
Tao Hong, Ph.D.
Alex Lindahl, M.P.H.
Samantha J. Snow, Ph.D., DABT

Conducted risk-of-bias assessment
Louise Assem, Ph.D.
Kristin Bornstein, Ph.D.
Jennifer Hsieh, M.S.P.H.
Yi Lu, Ph.D.

Conducted risk-of-bias assessment and reviewed data to support meta-analysis
Alexandra Goldstone, M.P.H.

Reviewed data to support meta-analysis
Jonathan Cohen, Ph.D.
Hannah Eglinton, B.A.
Cynthia J. Lin, Ph.D.

Supported meta-analysis visualizations
Rebecca Gray, M.P.H.

Coordinated peer review
Canden N. Byrd, B.S.
Lindsey Green, M.P.H

Designed and executed literature searches, and managed references
Michelle Cawley, M.A., M.L.S.
Jeremy S. Frye, M.S.L.S.
Nicole Vetter, M.S.L.S.

Edited and formatted monograph
Penelope Kellar, M.S.
Whitney Mitchell, B.S.
Kevin O'Donovan, B.A.

Integrated Laboratory Systems, LLC, Research Triangle Park, North Carolina, USA
Conducted a technical review of the draft monograph
Cynthia J. Willson, D.V.M., Ph.D., DACVP

Boston University, Boston, Massachusetts, USA
Reviewed the systematic review protocol
Thomas Webster, Ph.D.

Brown University, Providence, Rhode Island, USA
Reviewed the systematic review protocol
Joseph Braun, Ph.D.

Columbia University, New York, New York, USA
Reviewed the systematic review protocol
Gail Wasserman, Ph.D.

Health Research Board, Dublin, Ireland
Reviewed the systematic review protocol
Marie Sutton, Ph.D.

University of Massachusetts, Amherst, Massachusetts, USA
Reviewed the systematic review protocol
Thomas Zoeller, Ph.D.

University of Colorado, Denver, Colorado, USA
Reviewed meta-analysis protocol
Tianjing Li, Ph.D.

Peer Review

The National Toxicology Program (NTP) conducted a peer review of the draft *NTP Monograph on the State of the Science Concerning Fluoride Exposure and Neurodevelopmental and Cognitive Health Effects: A Systematic Review* by letter in December 2021. Reviewer selection and document review followed established NTP practices. The reviewers were charged to:

(1) Comment on the technical accuracy and whether the draft *NTP Monograph on the State of the Science Concerning Fluoride Exposure and Neurodevelopmental and Cognitive Health Effects* is clearly stated and objectively presented.

(2) Determine whether the scientific evidence supports the NTP's confidence ratings for the bodies of evidence regarding neurodevelopmental and cognitive health effects associated with exposure to fluoride.

NTP carefully considered reviewer comments in finalizing this monograph.

Peer Reviewers

Ethan Balk, M.D., M.P.H.
Associate Professor (Research)
Brown University School of Public Health
Providence, Rhode Island, USA

Pam Factor-Litvak, Ph.D.
Professor of Epidemiology
Columbia University Medical Center
New York, New York, USA

Erin Haynes, Dr.P.H.
Kurt W. Deuschle Professor in Preventive Medicine and Environmental Health
University of Kentucky College of Public Health
Lexington, Kentucky, USA

Julie Obbagy, Ph.D., RD
Nutritionist, Center for Nutrition Policy and Promotion
Office of Nutrition Guidance and Analysis
U.S. Department of Agriculture
Medfield, Massachusetts, USA

Heather Volk, Ph.D.
Associate Professor
Johns Hopkins Bloomberg School of Public Health
Baltimore, Maryland, USA

Publication Details

Publisher: National Toxicology Program

Publishing Location: Research Triangle Park, NC

ISSN: 2378-5144

DOI: https://doi.org/10.22427/NTP-MGRAPH-8

Report Series: NTP Monograph Series

Report Series Number: 08

Official citation: National Toxicology Program (NTP). 2022. NTP monograph on the state of the science concerning fluoride exposure and neurodevelopmental and cognitive health effects: a systematic review. Research Triangle Park, NC: National Toxicology Program. NTP Monograph 08.

Acknowledgements

This work was supported by the Intramural Research Program (ES103316, ES103317) at the National Institute of Environmental Health Sciences, National Institutes of Health and performed for the National Toxicology Program, Public Health Service, U.S. Department of Health and Human Services under contract GS00Q14OADU417 (Order No. HHSN273201600015U).

Conflict of Interest

Individuals who reviewed the systematic review protocol or meta-analysis protocol, conducted a technical review of the draft monograph, or served on the peer review panel have certified that they have no known real or apparent conflict of interest related to fluoride exposure or neurodevelopmental and cognitive health effects.

Abstract

Background: Fluoride is a common exposure in our environment that comes from a variety of sources and is widely promoted for its dental and overall oral health benefits. A 2006 evaluation by the National Research Council (NRC) found support for an association between consumption of high levels of naturally occurring fluoride in drinking water and adverse neurological effects in humans and recommended further investigation. The evidence reviewed at that time was from dental and skeletal fluorosis-endemic regions of China. Since the NRC evaluation, the number and location of studies examining cognitive and neurobehavioral effects of fluoride in humans have grown considerably, including several recent North American prospective cohort studies evaluating prenatal fluoride exposure.

In 2016, the National Toxicology Program (NTP) published a systematic review of the evidence from experimental animal studies on the effects of fluoride on learning and memory. That systematic review found a low-to-moderate level of evidence that deficits in learning and memory occur in non-human mammals exposed to fluoride.

Objective: To conduct a systematic review of the human, experimental animal, and mechanistic literature to evaluate the extent and quality of the evidence linking fluoride exposure to neurodevelopmental and cognitive effects in humans.

Method: A systematic review protocol was developed and utilized following the standardized OHAT systematic review approach for conducting literature-based health assessments. This monograph presents the current state of evidence associating fluoride exposure with neurocognitive or neurodevelopmental health effects and incorporated predefined assessments of study quality and confidence levels. Benefits of fluoride with respect to oral health are not addressed in this monograph.

Results: The current bodies of experimental animal studies and human mechanistic evidence do not provide clarity on the association between fluoride exposure and neurocognitive or neurodevelopmental human health effects.

This systematic review identified studies that assessed the association between fluoride exposure and cognitive or neurodevelopmental effects in both adults and children, which were evaluated separately. In adults, only two high-quality cross-sectional studies examining cognitive effects were available. The literature in children was more extensive and was separated into studies assessing intelligence quotient (IQ) and studies assessing other cognitive or neurodevelopmental outcomes. Eight of nine high-quality studies examining other cognitive or neurodevelopmental outcomes reported associations with fluoride exposure. Seventy-two studies assessed the association between fluoride exposure and IQ in children. Nineteen of those studies were considered to be high quality; of these, 18 reported an association between higher fluoride exposure and lower IQ in children. The 18 studies, which include 3 prospective cohort studies and 15 cross-sectional studies, were conducted in 5 different countries. Forty-six of the 53 low-quality studies in children also found evidence of an association between higher fluoride exposure and lower IQ in children.

Discussion: Existing animal studies provide little insight into the question of whether fluoride exposure affects IQ. In addition, studies that evaluated fluoride exposure and mechanistic data in humans were too heterogenous and limited in number to make any determination on biological plausibility. The body of evidence from studies in adults is also limited and provides low

confidence that fluoride exposure is associated with adverse effects on adult cognition. There is, however, a large body of evidence on IQ effects in children. There is also some evidence that fluoride exposure is associated with other neurodevelopmental and cognitive effects in children; although, because of the heterogeneity of the outcomes, there is low confidence in the literature for these other effects. This review finds, with moderate confidence, that higher fluoride exposure (e.g., represented by populations whose total fluoride exposure approximates or exceeds the World Health Organization Guidelines for Drinking-water Quality of 1.5 mg/L of fluoride) is consistently associated with lower IQ in children. More studies are needed to fully understand the potential for lower fluoride exposure to affect children's IQ.

Preface

The National Toxicology Program (NTP) conducted a systematic review of the published scientific literature because of public concern regarding the potential association between fluoride exposure and adverse neurodevelopmental and cognitive health effects.

NTP initially published a systematic review of the experimental animal literature in 2016 that was subsequently expanded to include human epidemiological studies, mechanistic studies, and newer experimental animal literature. Because of the high public interest in fluoride's benefits and potential risks, NTP asked the National Academies of Sciences, Engineering, and Medicine (NASEM) to conduct an independent evaluation of the draft *NTP Monograph on Fluoride Exposure and Neurodevelopmental and Cognitive Health Effects* (2019 draft monograph dated September 6, 2019) and the revised draft (2020 draft monograph dated September 16, 2020), which addressed the NASEM committee's recommendations for improvement. The NASEM committee determined that, "Overall the revised monograph seems to include a wealth of evidence and a number of evaluations that support its main conclusion, but the monograph falls short of providing a clear and convincing argument that supports its assessments…." Thus, NTP has removed the hazard assessment step and retitled this systematic review of fluoride exposure and neurodevelopmental and cognitive health effects as a "state-of-the-science" document to indicate the change. This state-of-the-science document does not include the meta-analysis of epidemiological studies or hazard conclusions found in previous draft monographs; however, it provides a comprehensive and current assessment of the scientific literature on fluoride as an important resource to inform safe and appropriate use.

NTP has responded to the NASEM committee's comments on the revised draft (September 16, 2020) in a separate document (placeholder for URL) and revised relevant sections of this monograph.

Introduction

Fluoride is a common exposure in our environment from a variety of sources and is widely promoted for its dental and overall oral health benefits. Approximately 67% of the U.S. population receives fluoridated water through a community water system (CDC 2013). In other countries, fluoride supplementation has been achieved by fluoridating food products such as salt or milk. Fluoride supplementation has been recommended to prevent bone fractures (Jones et al. 2005). Fluoride also can occur naturally in drinking water. Other sources of human exposure include other foods and beverages, industrial emissions, pharmaceuticals, and pesticides (e.g., cryolite, sulfuryl fluoride). Soil ingestion is another source of fluoride exposure in young children (US EPA 2010).

The U.S. Public Health Service (PHS) first recommended that communities add fluoride to drinking water in 1962. PHS guidance is advisory, not regulatory, which means that while PHS recommends community water fluoridation as a public health intervention, the decision to fluoridate water systems is made by state and local governments. For many years, most fluoridated community water systems used fluoride concentrations ranging from 0.8 to 1.2 milligrams/liter (mg/L) (US DHHS 2015). For community water systems that add fluoride, PHS now recommends a fluoride concentration of 0.7 mg/L (equal to 0.7 parts per million [ppm]). Under the Safe Drinking Water Act, the U.S. Environmental Protection Agency (EPA) sets maximum exposure level standards for drinking water quality. The current enforceable drinking water standard for fluoride, or the maximum contaminant level (MCL), is 4.0 mg/L. This level is the maximum amount of fluoride contamination (naturally occurring, not from water fluoridation) that is allowed in water from public water systems and is set to protect against increased risk of skeletal fluorosis, a condition characterized by pain and tenderness of the major joints. EPA also has a non-enforceable secondary drinking water standard of 2.0 mg/L of fluoride, which is recommended to protect children against the tooth discoloration and/or pitting that can be caused by severe dental fluorosis during the formative period prior to eruption of teeth. Although the secondary standard is not enforceable, EPA requires that public water systems notify the public if and when average fluoride levels exceed 2.0 mg/L (NRC 2006). The World Health Organization (WHO) set a safe water guideline of 1.5 mg/L of fluoride in drinking water (first established in 1984 and reaffirmed in 1993 and 2011), which is recommended to protect against increasing risk of dental and skeletal fluorosis (WHO 2017).

As of April 2020, 1.08% of persons living in the United States (~3.5 million people) were served by community water systems (CWS) containing ≥1.1 mg/L naturally occurring fluoride. CWS supplying water with ≥1.5 mg/L naturally occurring fluoride served 0.59% of the U.S. population (~1.9 million people), and systems supplying water with ≥2 mg/L naturally occurring fluoride served 0.31% of the U.S. population (~1 million people) (CDC Division of Oral Health 2020).

Commonly cited health concerns related to fluoride are bone fractures and skeletal fluorosis, lower intelligence quotient (IQ) and other neurological effects, cancer, and endocrine disruption.

Effects on neurological function, endocrine function (e.g., thyroid,[1] parathyroid, pineal), metabolic function (e.g., glucose metabolism), and carcinogenicity were assessed in the 2006 NRC report, *Fluoride in Drinking Water: A Scientific Review of EPA's Standards* (NRC 2006). The NRC review considered adverse effects of water fluoride, focusing on a range of concentrations (2–4 mg/L) above the current 0.7-mg/L recommendation for community water fluoridation. The NRC report concluded that the Maximum Contaminant Level Goal (MCLG), 4 mg/L, should be lowered to protect against severe enamel fluorosis and reduce the risk of bone fractures associated with skeletal fluorosis (NRC 2006). Other than severe fluorosis, NRC did not find sufficient evidence of negative health effects at fluoride levels below 4 mg/L; however, it concluded that the consistency of the results of IQ deficits in children exposed to fluoride at 2.5 to 4 mg/L in drinking water from a few epidemiological studies of Chinese populations appeared significant enough to warrant additional research on the effects of fluoride on intelligence. The NRC report noted several challenges to evaluating the literature, including deficiencies in reporting quality, lack of consideration of all sources of fluoride exposure, incomplete consideration of potential confounding, selection of inappropriate control subject populations in epidemiological studies, absence of demonstrated clinical significance of reported endocrine effects, and incomplete understanding of the biological relationship between histological, biochemical, and molecular alterations with behavioral effects.

In 2016, the National Toxicology Program (NTP) 2016, NTP published a systematic review of the evidence from experimental animal studies on the potential effects of fluoride exposure on learning and memory (NTP 2016). That systematic review found a low-to-moderate level of evidence that deficits in learning and memory occur in experimental animals exposed to fluoride. Given these findings, NTP decided to conduct additional animal studies before carrying out this full systematic review and integrate human, animal, and potentially relevant mechanistic evidence in order to reach human health hazard identification conclusions for fluoride and learning and memory effects. As the NTP (2016) report on the experimental animal evidence focused on learning and memory and developed confidence ratings for bodies of evidence by life stage of exposure (i.e., exposure during development or adulthood), this monograph also evaluates two different age groups in humans (i.e., children and adults) with a focus on cognitive neurodevelopmental effects in children and cognitive effects in adults in order to address potential differences in health impacts based on time frame of exposure (i.e., during development or during adulthood). The evaluation of experimental animal studies in this monograph has been conducted separately from the 2016 experimental animal assessment; however, like the 2016 assessment, it assessed mainly learning and memory effects in experimental animal studies to determine whether the findings inform the assessment of cognitive neurodevelopmental effects in children and cognitive effects in adults.

A committee convened by the National Academies of Sciences, Engineering, and Medicine (NASEM) reviewed earlier drafts of this monograph (September 6, 2019, and September 16, 2020) (NASEM 2020; 2021). The current document incorporates changes stemming from those reviews, and responses to the 2020 review are available at (placeholder to cite NTP 2021

[1]The current review has evaluated the fluoride literature with an eye toward potential thyroid effects because a large literature base has accumulated examining the interaction of fluoride with iodine uptake by the thyroid gland and consequential effects on synthesis of thyroid hormones, which are recognized to play significant roles in neurodevelopment in utero and during early childhood. This literature, along with a detailed proposed mechanism of action, was recently reviewed by Waugh (2019).

Response to NASEM comments). See Appendix B, Table B-1 for a timeline of key activities contributing to this 2022 NTP monograph, including document review activities that have occurred since 2016.

Objective and Specific Aims

Objective

The overall objective of this evaluation was to undertake a systematic review to develop NTP human health hazard identification conclusions on the association between exposure to fluoride and neurodevelopmental and cognitive effects based on assessing levels of evidence from human and non-human animal studies with consideration of the degree of support from mechanistic data. However, the NASEM Committee's reviews (NASEM 2020; 2021) of the 2019 and 2020 drafts of the monograph indicated that, "Overall the revised monograph seems to include a wealth of evidence and a number of evaluations that support its main conclusion, but the monograph falls short of providing a clear and convincing argument that supports its assessments…." For this reason, our methods were revised to remove the hazard assessment step (i.e., the section "Integrate Evidence to Develop Hazard Identification Conclusions" and the associated section "Translate Confidence Ratings into Level of Evidence for Health Effect"). In addition, a meta-analysis of the epidemiological studies examining children's IQ in relation to fluoride exposure added to the 2020 draft in response to NASEM comments (NASEM 2020) will be published separately and is not part of this document.

Therefore, the objective of this monograph is to undertake a systematic review of the literature concerning the association between fluoride exposure and neurodevelopmental and cognitive effects and to determine the level of confidence in that evidence. The assessment was based on evidence from human and non-human animal studies with consideration of mechanistic information.

Specific Aims

- Identify literature that assessed neurodevelopmental and cognitive health effects, especially outcomes related to learning, memory, and intelligence, following exposure to fluoride in human, animal, and relevant in vitro/mechanistic studies.
- Extract data on potential neurodevelopmental and cognitive health effects from relevant studies.
- Assess the internal validity (risk of bias) of individual studies using pre-defined criteria.
- Assess effects on thyroid function to help evaluate potential mechanisms of impaired neurobehavioral[2] function.
- Summarize the extent and types of health effects evidence available.

[2]The specific aim in the protocol refers to "impaired neurological function"; however, it was changed to "impaired neurobehavior function" in this document to use more precise terminology. The overall aim from the protocol remained the same for this evaluation.

- Describe limitations of the systematic review, strengths and limitations of the evidence base, identify areas of uncertainty, as well as data gaps and research needs for neurodevelopmental and cognitive health effects of fluoride.

Depending on the extent and nature of the available evidence:

- Synthesize the evidence using a narrative approach.
- Rate confidence in the body of evidence for human and animal studies separately according to one of four statements: High, Moderate, Low, or Very Low/No Evidence Available.

Methods

Problem Formulation and Protocol Development

The research question and specific aims stated above were developed and refined through a series of problem formulation steps, including:

(1) receipt of a nomination from the public in June 2015 to conduct analyses of fluoride and developmental neurobehavioral toxicity;

(2) analysis of the extent of evidence available and the merit of pursuing systematic reviews, given factors such as the extent of new research published since previous evaluations and whether these new reports address or correct the deficiencies noted in the literature (OEHHA 2011; NRC 2006; SCHER 2011);

(3) request for information in a Federal Register notice (dated October 7, 2015);

(4) consideration of comments providing a list of studies to review through Federal Register notice and public comment period from October 7, 2015, to November 6, 2015;

(5) release of draft concept titled *Proposed NTP Evaluation on Fluoride Exposure and Potential for Developmental Neurobehavioral Effects* in November 2015;

(6) presentation of draft concept at the NTP Board of Scientific Counselors (BSC) meeting on December 1–2, 2015;

(7) consideration of comments on NTP's draft concept from the NTP BSC meeting in December 2015; and

(8) consideration of input on the draft protocol from review by technical advisors.

The protocol used to conduct this systematic review was posted in June 2017 with updates posted in May 2019 and September 2020 (https://ntp.niehs.nih.gov/go/785076).[3] The protocol served as the complete set of methods followed for the conduct of this systematic review. The OHAT Handbook for Conducting a Literature-Based Health Assessment (http://ntp.niehs.nih.gov/go/38673) is a source of general systematic review methods that were selected and tailored in developing this protocol. Options in the OHAT handbook that were not specifically referred to in the protocol were not part of the methods for the systematic review.

A brief summary of the methods is presented below. Although the methods were revised to remove the hazard assessment step and meta-analysis from this document, the protocol was not further revised.

PECO Statements

PECO (**P**opulation, **E**xposure, **C**omparators and **O**utcomes) statements were developed as an aid to identify search terms and appropriate inclusion/exclusion criteria for addressing the overall research question (effects on neurodevelopmental or cognitive function and thyroid associated

[3]NTP conducts systematic reviews following prespecified protocols that describe the review procedures selected and applied from the general methods outlined in the OHAT Handbook for Conducting a Literature-Based Health Assessment (http://ntp.niehs.nih.gov/go/38673). The protocol describes project-specific procedures tailored to each systematic review that supersede the methods in the OHAT Handbook.

with fluoride exposure) for the systematic review (Higgins and Green 2011). The PECO statements are listed below for human, animal, and in vitro/mechanistic studies (see Table 1, Table 2, and Table 3).

Using the PECO statements, the evaluation searched human studies, controlled exposure animal studies, and mechanistic/in vitro studies for evidence of neurodevelopmental or cognitive function and thyroid effects associated with fluoride exposure. Mechanistic data can come from a wide variety of studies that are not intended to identify a disease phenotype. This source of experimental data includes in vitro and in vivo laboratory studies directed at cellular, biochemical, and molecular mechanisms and attempt to explain how a substance produces particular adverse health effects. The mechanistic data were first organized by general categories (e.g., biochemical effects in the brain and neurons, neurotransmitters, oxidative stress) to evaluate the available information. Categories focused on were those with more robust data at levels of fluoride more relevant to human exposure. The intent was not to develop a mechanism for fluoride induction of effects on learning and memory but to evaluate whether a plausible series of mechanistic events exists to support effects observed in the low-dose region (below approximate drinking-water-equivalent concentrations of 20 ppm for animal studies) that may strengthen a hazard conclusion if one is derived.

Table 1. Human PECO (Population, Exposure, Comparator and Outcome) Statement

PECO Element	Evidence
Population	Humans without restriction as to age or sex, geographic location, or life stage at exposure or outcome assessment
Exposure	Exposure to fluoride based on administered dose or concentration, biomonitoring data (e.g., urine, blood, other specimens), environmental measures (e.g., air, water levels), or job title or residence. Relevant forms are those used as additives for water fluoridation: • Fluorosilicic acid (also called hydrofluorosilicate; Chemical Abstracts Service Registry Number [CASRN] 16961-83-4) • Sodium hexafluorosilicate (also called disodium hexafluorosilicate or sodium fluorosilicate; CASRN 16893-85-9) • Sodium fluoride (CASRN 7681-49-4) • Other forms of fluoride that readily dissociate into free fluoride ions (e.g., potassium fluoride, calcium fluoride, ammonium fluoride)
Comparators	Comparable populations not exposed to fluoride or exposed to lower levels of fluoride (e.g., exposure below detection levels)
Outcomes	Neurodevelopmental outcomes, including learning, memory, intelligence, other forms of cognitive behavior, other neurological/neurobehavioral[4] outcomes (e.g., anxiety, aggression, motor activity), and biochemical changes in the brain or nervous system tissue; measures of thyroid function, biochemical changes, or thyroid tissue pathology

Table 2. Animal PECO Statement

[4]The human PECO statement in the protocol refers to "neurological outcomes"; however, it was changed to "neurological/neurobehavioral outcomes" in this document to use more precise terminology for the outcomes included.

PECO Element	Evidence
Population	Non-human mammalian animal species (whole organism)
Exposure	Exposure to fluoride based on administered dose or concentration and biomonitoring data (e.g., urine, blood, other specimens). Relevant forms are those used as additives for water fluoridation: • Fluorosilicic acid (also called hydrofluorosilicate; CASRN 16961-83-4) • Sodium hexafluorosilicate (also called disodium hexafluorosilicate or sodium fluorosilicate; CASRN 16893-85-9) • Sodium fluoride (CASRN 7681-49-4) • Other forms of fluoride that readily dissociate into free fluoride ions (e.g., potassium fluoride, calcium fluoride, ammonium fluoride)
Comparators	Comparable animals that were untreated or exposed to vehicle-only treatment
Outcomes	Neurodevelopmental outcomes, including learning, memory, intelligence, other forms of cognitive behavior, other neurological/neurobehavioral[5] outcomes (e.g., anxiety, aggression, motor activity), and biochemical changes in the brain or nervous system tissue; measures of thyroid function, biochemical changes, or thyroid tissue pathology

Table 3. In Vitro/Mechanistic PECO Statement

PECO Element	Evidence
Population	Human or animal cells, tissues, or biochemical reactions (e.g., ligand binding assays)
Exposure	Exposure to fluoride based on administered dose or concentration. Relevant forms are those used as additives for water fluoridation: • Fluorosilicic acid (also called hydrofluorosilicate; CASRN 16961-83-4) • Sodium hexafluorosilicate (also called disodium hexafluorosilicate or sodium fluorosilicate; CASRN 16893-85-9) • Sodium fluoride (CASRN 7681-49-4) • Other forms of fluoride that readily dissociate into free fluoride ions (e.g., potassium fluoride, calcium fluoride, ammonium fluoride)
Comparators	Comparable cells or tissues that were untreated or exposed to vehicle-only treatment
Outcomes	Endpoints related to neurological and thyroid function, including neuronal electrophysiology; mRNA, gene, or protein expression; cell proliferation or death in brain or thyroid tissue/cells; neuronal signaling; synaptogenesis, etc.

Literature Search

Main Literature Search

Search terms were developed to identify all relevant published evidence on developmental neurobehavioral toxicity or thyroid-related health effects potentially associated with exposure to fluoride by reviewing Medical Subject Headings for relevant and appropriate neurobehavioral

[5]The animal PECO statement in the protocol refers to "neurological outcomes"; however, it was changed to "neurological/neurobehavioral outcomes" in this document to use more precise terminology for the outcomes included.

and thyroid-related terms and by extracting key neurobehavioral and thyroid-related health effects and developmental neurobehavioral terminology from reviews and a sample of relevant studies.[6] Combinations of relevant subject headings and keywords were subsequently identified. A test set of relevant studies was used to ensure the search terms retrieved 100% of the test set. Six electronic databases were searched (see Main Literature Database Search) using a search strategy tailored for each database (specific search terms used for the PubMed search are presented in Appendix B; the search strategy for other databases are available in the protocol (https://ntp.niehs.nih.gov/go/785076). A search of PubChem indicated that sodium fluoride was not found in either the Tox21 or ToxCast databases; therefore, these databases were not included in the search. No language restrictions or publication-year limits were imposed. These six databases were searched in December 2016, and the search was regularly updated during the review process through April 1, 2019.

An additional search was conducted on May 1, 2020, where human epidemiological studies with primary neurodevelopmental or cognitive outcomes (learning, memory, and intelligence) were prioritized during screening. The review of the 2020 search results focused only on the human studies because they formed the basis of the confidence ratings (see Figure 1 for framework to assess confidence) and conclusions in the September 6, 2019, draft. A supplemental literature search of Chinese-language databases (described below) was also conducted. See Appendix B, Table B-1 for a timeline of key activities contributing to this 2022 NTP monograph, including information relevant to the timing of multiple literature searches.

Publications identified in these searches are categorized as "references identified through database searches" in Figure 2. Studies identified from other sources or manual review that might impact conclusions are considered under "references identified through other sources" in Figure 2. Literature searches for this systematic review were conducted independently from the literature search conducted for NTP (2016). The current literature search strategy was based on the search terms used for NTP (2016) and refined for the current evaluation, including the addition of search terms to identify human studies. Although the review process identified experimental animal studies prior to 2015, the current assessment did not evaluate these studies and relied on the NTP (2016) assessment. The focus of the literature searches for this systematic review was to identify and evaluate relevant animal studies that were published since completion of the literature searches for the NTP (2016) assessment in addition to the human and mechanistic data that were not previously evaluated.

Supplemental Chinese Database Literature Search

In order to identify non-English-language studies that might not appear in databases for the main literature search, additional searches were developed for non-English-language databases. No definitive guidance was found on the most comprehensive, highest quality, or otherwise most appropriate non-English-language databases for health studies of fluoride. Therefore, databases were chosen that identified non-English-language studies that were not captured in searches of databases from the main literature search—those previously identified from other resources (see the Searching Other Resources section below). Multiple non-English-language databases were explored before two were identified, CNKI and Wanfang, that covered studies previously

[6]The terms "study" and "publication" are used interchangeably in this document to refer to a published work drawn from an original body of research conducted on a defined population.

identified from other sources. These two Chinese electronic databases were searched in May 2020 with no language restrictions or publication year limits. Search terms from the main literature search were refined to focus on human epidemiological studies. The CNKI and Wanfang databases have character limits in the search strings; therefore, key terms were prioritized using text analytics to identify the most prevalent terms from neurodevelopmental or cognitive human epidemiological studies previously identified as relevant. Search strings were designed to capture known relevant studies that were previously identified from searching other resources without identifying large numbers of non-relevant studies (the search strategy for both databases is available in the protocol [https://ntp.niehs.nih.gov/go/785076]). Publications retrieved were compared with publications retrieved from the main literature search, and duplicates were removed. The remaining relevant publications are categorized as "references identified through database searches" in Figure 2.

New animal and mechanistic references retrieved were scanned for evidence that might extend the information currently in the September 6, 2019, draft. Although additional studies were identified, data that would materially advance the animal and mechanistic findings were not identified; therefore, these studies were not extracted nor were they added to the draft. A primary goal of the screening of the newly retrieved human references in the supplemental search of Chinese databases was to identify studies that evaluated primary neurodevelopmental or cognitive outcomes (i.e., learning, memory, and intelligence) that may have been missed in previous searches that did not include the Chinese databases. A secondary goal was to examine whether the non-English-language studies on the Fluoride Action Network website (http://fluoridealert.org/)—a site used as another resource to identify potentially relevant studies because it is known to index fluoride publications—had been selectively presented to list only studies reporting effects of fluoride. Newly retrieved human references were reviewed to identify studies that may have been missed using previous approaches. Studies identified that evaluated primary neurodevelopmental or cognitive outcomes were included and either translated or reviewed by an epidemiologist fluent in Chinese.

Databases Searched

Main Literature Database Search

- BIOSIS (Thomson Reuters)
- EMBASE
- PsycINFO (APA PsycNet)
- PubMed (NLM)
- Scopus (Elsevier)
- Web of Science (Thomson Reuters, Web of Science indexes the journal Fluoride)

Supplemental Chinese Database Literature Search

- CNKI
- Wanfang

Searching Other Resources

The reference lists of all included studies; relevant reviews, editorials, and commentaries; and the Fluoride Action Network website (http://fluoridealert.org/) were manually searched for additional relevant publications.

Unpublished Data

Although no unpublished data were included in the review, unpublished data were eligible for inclusion, provided the owner of the data was willing to have the data made public and peer reviewed (see protocol for more details: https://ntp.niehs.nih.gov/go/785076).

Study Selection

Evidence Selection Criteria

In order to be eligible for inclusion, studies had to satisfy eligibility criteria that reflect the PECO statements in Table 1, Table 2, and Table 3.

The following additional exclusion criteria were applied (see protocol for additional details: https://ntp.niehs.nih.gov/go/785076):

(1) Case studies and case reports. Although there are various definitions of 'case study' and 'case report,' the terms are used here to refer to publications designed to share health-related events on a single subject or patient with a disease, diagnosis, or specific outcome in the presence of a specific exposure.

(2) Articles without original data (e.g., reviews, editorials, or commentaries). Reference lists from these materials, however, were reviewed to identify potentially relevant studies not identified from the database searches. New studies identified were assessed for eligibility for inclusion.

(3) Conference abstracts, theses, dissertations, and other non-peer-reviewed reports.

Screening Process

References retrieved from the literature search were independently screened by two trained screeners at the title and abstract level to determine whether a reference met the evidence selection criteria. Screening procedures following the evidence-selection criteria in the protocol were pilot tested with experienced contract staff overseen by NTP. For citations with no abstract or non-English abstracts, articles were screened based on title relevance (the title would need to indicate clear relevance); number of pages (articles ≤2 pages were assumed to be conference reports, editorials, or letters unlikely to contain original data); and/or PubMed Medical Subject Headings (MeSH). Using this approach, literature was manually screened for relevance and eligibility against the evidence selection criteria using a structured form in SWIFT-Active Screener (Sciome) (Howard et al. 2020). While the human screeners review studies, SWIFT-Active Screener aids in this process by employing a machine-learning software program to priority-rank studies for screening (Howard et al. 2020). SWIFT-Active Screener also refines a statistical model that continually ranks the remaining studies according to their likelihood for inclusion. In addition, SWIFT-Active Screener employs active learning to continually incorporate user feedback during title and abstract screening to predict the total number of

included studies, thus providing a statistical basis for a decision about when to stop screening (Miller et al. 2016). Title and abstract screening was stopped once the statistical algorithm in SWIFT-Active Screener estimated that 98% of the predicted number of relevant studies were identified.

Studies that were not excluded during the title and abstract screening were further screened for inclusion with a full-text review by two independent reviewers using DistillerSR® (Evidence Partners), a web-based, systematic-review software program with structured forms and procedures to ensure standardization of the process. Screening conflicts were resolved through discussion and consultation with technical advisor(s), if necessary. During full-text review, studies that were considered relevant were tagged to the appropriate evidence streams (i.e., human, animal, and/or in vitro). Studies tagged to human or animal evidence streams were also categorized by outcome as primary neurodevelopmental or cognitive outcomes (learning, memory, and intelligence); secondary neurobehavioral outcomes (anxiety, aggression, motor activity, or biochemical); or related to thyroid effects. In vitro data were tagged as being related to neurological effects or thyroid effects. Translation assistance was sought to assess the relevance of non-English studies. Following full-text review, the remaining studies were "included" and used for the evaluation.

Evaluation of SWIFT-Active Screener Results

During the initial title and abstract screening of 20,883 references using SWIFT-Active Screener, approximately 38%[7] of the studies were manually screened in duplicate to identify an estimated 98.6% of the predicted number of relevant studies using the software's statistical algorithm (13,023 references were not screened). SWIFT-Active Screener predicted that there were 739 relevant studies during the initial title and abstract screening, of which 729 were identified and moved to full-text review. The SWIFT-Active Screener statistical algorithm predicted that 10 relevant studies at the title and abstract level (10 represents 1.4% × 739 predicted relevant studies; or 739 predicted relevant studies minus 729 identified relevant studies during screening) were not identified by not screening the remaining 13,023 studies.

To further consider the impact of using SWIFT-Active Screener for this systematic review, the evaluation team assessed the SWIFT-Active screening results to gain a better understanding of the relevance of the last group of studies that was screened before 98% predicted recall (i.e., 98% of the predicted number of relevant studies were identified). The goal was to determine the likelihood of having missed important studies by not screening all of the literature. To do this, the evaluation team examined subsets of studies screened in SWIFT-Active Screener for trends and followed those studies through to full-text review for a final determination of relevance and potential impact (i.e., whether the studies had data on primary outcomes). Based on this evaluation, it was estimated that the use of SWIFT-Active Screener may have resulted in missing one to two relevant human studies and one to two relevant animal studies with primary neurodevelopmental or cognitive outcomes. Therefore, the use of SWIFT-Active Screener saved

[7]Howard et al. (2020) evaluated the performance of the SWIFT-Active Screener methods for estimating total number of relevant studies using 26 diverse systematic review datasets that were previously screened manually by reviewers. The authors found that on average, 95% of the relevant articles were identified after screening 40% of the total reference list when using SWIFT-Active Screener. In the document sets with 5,000 or more references, 95% of the relevant articles were identified after screening 34% of the available references, on average, using SWIFT-Active Screener.

considerable time and resources and is expected to miss very few potentially relevant publications.

Screening of the May 2020 Literature Search Update

For the May 1, 2020, literature search, only primary human epidemiological studies were identified for data extraction. The study screening and selection process was focused on the human studies with primary outcomes for the evaluation because they form the basis of the confidence ratings and conclusions. Animal in vivo, human secondary outcome-only, and human and animal mechanistic references were identified as part of the screening process. These studies were then scanned for evidence that might extend the information in the September 6, 2019, draft. All included studies from the May 2020 literature search update appear in Appendix C; however, other than the primary human epidemiological studies, data from the new studies were not extracted unless they would materially advance the findings.

Note that NTP is aware of a conference abstract by Santa-Marina et al. on a Spanish cohort study that looked at fluoride exposure and neuropsychological development in children (Santa-Marina et al. 2019). The evaluation team conducted a targeted literature search in April 2021 to see whether the data from this study had been published. When no publication was found, the evaluation team contacted the study authors to inquire about the publication of their data. The response from the study authors indicated that the study report was being finalized but had not yet been sent to a journal for review; therefore, it was not considered here.[8]

Supplemental Chinese Database Searches and Human Epidemiological Studies

Supplemental searches were conducted in non-English-language databases (CNKI and Wanfang). Of the 910 references that were identified in the supplemental Chinese database searches, 13 relevant studies published in Chinese with primary neurobehavioral or cognitive outcomes were identified during title and abstract screening (which were not identified through the main literature searches). Full texts were not found for four studies after an extensive search. The remaining nine studies for which full texts were retrieved were included and were either professionally translated or evaluated by an epidemiologist fluent in Chinese for the data extraction and quality assessment steps described below. If necessary, author inquiries were conducted in Chinese to obtain missing information relevant to the assessment of the key risk-of-bias questions described below.

[8]NTP is aware that this study was published after April 2021 (Ibarluzea et al. 2021) and, therefore, is not included in this monograph because it is beyond the dates of the literature search. Even if it had been published earlier, the study would not have contributed to the body of evidence on children's IQ because the authors assessed other neurodevelopmental or cognitive effects, specifically the association between fluoride exposure and neuropsychological development in children aged 1 year using the Mental Development Index (MDI) of the Bayley Scales of Infant Development and in children aged 4 years using the General Cognitive Index (GCI) of the McCarthy Scales of Children’s Abilities (MSCA). The study will be examined as part of the NTP meta-analysis, which is being prepared as a separate report for publication.

Data Extraction

Extraction Process

Data were collected (i.e., extracted) from included studies by one member of the evaluation team and checked by a second member for completeness and accuracy. Any discrepancies in data extraction were resolved by discussion or consultation with a third member of the evaluation team.

Data Availability

Data extraction was completed using the Health Assessment Workspace Collaborative (HAWC), an open-source and freely available web-based application.[9] Data extraction elements are listed separately for human, animal, and in vitro studies in the protocol (https://ntp.niehs.nih.gov/go/785076). Data for primary and secondary outcomes, as well as thyroid hormone level data, were extracted from human studies. Studies evaluating only goiters or thyroid size were not extracted because they do not provide specific information on thyroid hormone levels that would inform whether a thyroid-mediated mechanism was involved in fluoride-associated changes in neurodevelopment. All primary outcomes and functional neurological secondary outcomes (e.g., motor activity) were extracted from animal studies identified since the NTP (2016) report. For animal mechanistic data, studies were tiered based on exposure dose (with preference given to fluoride drinking-water-equivalent exposures, which were calculated using the method described in the NTP (2016) report, of 20 ppm or less as deemed most relevant to exposures in humans), exposure duration or relevant time window (i.e., developmental), exposure route (with preference given to oral exposures over injection exposures), and commonality of mechanism (e.g., inflammation, oxidative stress, changes in neurotransmitters, and histopathological changes) were considered pockets of mechanistic data. Thyroid data were not extracted for animal studies due to inconsistency in the available data in humans. In vitro studies were evaluated, although data were not extracted from these studies as none of the findings were considered informative with respect to biological plausibility. The data extraction results for included studies are publicly available and can be downloaded in Excel format through HAWC (https://hawcproject.org/assessment/405/). Methods for transforming and standardizing dose levels and results from behavioral tests in experimental animals are detailed in the protocol (https://ntp.niehs.nih.gov/go/785076).

In 2016, NTP published a systematic review of the evidence from experimental animal studies on the potential effects of fluoride exposure on learning and memory (NTP 2016). The literature searches for the current assessment identified and evaluated relevant animal studies published since the 2016 assessment and also included human and mechanistic data that were not previously evaluated. Although literature search activities for the current assessment identified experimental animal studies prior to 2015, the current assessment did not re-evaluate animal studies published prior to 2015 because these were reviewed in the NTP (2016) assessment.

[9]HAWC (Health Assessment Workspace Collaborative): A Modular Web-based Interface to Facilitate Development of Human Health Assessments of Chemicals (https://hawcproject.org/portal/).

Quality Assessment of Individual Studies

Risk of bias was assessed for individual studies using the OHAT risk-of-bias tool (https://ntp.niehs.nih.gov/go/riskbias) that outlines a parallel approach to evaluating risk of bias from human, animal, and mechanistic studies to facilitate consideration of risk of bias across evidence streams with common terms and categories. The risk-of-bias tool is comprised of a common set of 11 questions that are answered based on the specific details of individual studies to develop risk-of-bias ratings for each question. Study design determines the subset of questions used to assess risk of bias for an individual study (see Table 4). When evaluating the risk of bias for an individual study, the direction and magnitude of association for any specific bias is considered.

Assessors were trained with an initial pilot phase undertaken to improve clarity of rating criteria and to improve consistency among assessors. Studies were independently evaluated by two trained assessors who answered all applicable risk-of-bias questions with one of four options in Table 5 following prespecified criteria detailed in the protocol (https://ntp.niehs.nih.gov/go/785076). The criteria describe aspects of study design, conduct, and reporting required to reach risk-of-bias ratings for each question and specify factors that can distinguish among ratings (e.g., what separates “definitely low” from “probably low” risk of bias).

Key Risk-of-bias Questions

In the OHAT approach, some risk-of-bias questions or elements are considered potentially more important when assessing studies because these issues are generally considered to have a greater impact on estimates of the effect size or on the credibility of study results in environmental health studies. There are three Key Questions for observational human studies: confounding, exposure characterization, and outcome assessment. Based on the complexity of the possible responses to these questions in epidemiological studies, considerations made and methods used for evaluating the Key Questions are provided below. There are also three Key Questions for experimental animal studies: randomization, exposure characterization, and outcome assessment. In addition, for animal developmental studies, failure to consider the litter as the unit of analysis was also a key risk-of-bias concern. When there was not enough information to assess the potential bias for a risk-of-bias question and authors did not respond to an inquiry for further information, a conservative approach was followed, and the studies were rated probably high risk of bias for that question.

Risk-of-bias Considerations for Human Studies

The risk of bias of individual studies in the body of evidence was considered in developing confidence ratings. The key risk-of-bias questions (i.e., confounding, exposure characterization, and outcome assessment for human studies) are discussed in the consideration of the body of evidence. For this assessment, the key risk-of-bias questions, if not addressed appropriately, are considered to have the greatest potential impact on the results. The other risk-of-bias questions, including selection of study participants, were also considered and were used to identify any other risk-of-bias concerns that may indicate serious issues with a study that could cause it to be considered high risk of bias. No study was excluded based on concerns for risk of bias; however, the low risk-of-bias studies generally drive the ratings on confidence in the results across the

body of evidence. Human evidence was evaluated with and without high risk-of-bias studies to assess the impact of these studies on confidence in the association.

High risk-of-bias studies: Studies rated probably high risk of bias for at least two key risk-of-bias questions or definitely high for any single question are considered studies with higher potential for bias (i.e., high risk-of-bias studies) and to be of low quality. Studies could also be considered high risk of bias if rated probably high risk of bias for one key risk-of-bias question along with other concerns, including potential for selection bias and concerns with statistical methods.

Low risk-of-bias studies: The remaining studies (i.e., other than the high risk-of-bias studies) were considered to have lower potential for bias (i.e., low risk of bias) and to be of high quality. Appendix E describes strengths and limitations of the low risk-of-bias/high-quality studies identified during the assessment and clarifies why they are considered to pose low risk of bias. Details on the statistical analyses are provided in the "Other potential threats" domain in order to evaluate the adequacy of the statistical approach for individual studies.

Given the number of non-English-language studies in this assessment, the potential for the translation to introduce bias was examined as described below, and it was determined that translation of non-English-language studies did not impact evaluation of risk of bias. Thirty-two of 100 studies included in the entire human body of evidence on neurodevelopmental and cognitive effects were initially published in a foreign language (Chinese) and were either translated and published in volume 41 of the journal *Fluoride* (n = 19) or were translated by the Fluoride Action Network (n = 13) (http://fluoridealert.org/researchers/translations/complete_archive/). Most of these studies were considered to have high potential for bias due to lack of information across the key risk-of-bias questions. Therefore, in order to assess whether the lack of information relevant to key risk-of-bias concerns was the result of a loss in translation, the original Chinese publications and the translated versions of the five studies that had the most potential for being included in the low risk-of-bias group of studies were reviewed by a team member fluent in Chinese to determine whether any of the risk-of-bias concerns could be addressed (An et al. 1992; Chen et al. 1991 [translated in Chen et al. 2008]; Du et al. 1992 [translated in Du et al. 2008]; Guo et al. 1991 [translated in Guo et al. 2008a]; Li et al. 2009). For all five studies, the translations were determined to be accurate, and there was no impact of the translations on the key risk-of-bias concerns.

Confounding

Covariates were determined a priori based on factors that are associated with neurodevelopment or cognition and could be related to fluoride exposure. Covariates that were considered key for all studies, populations, and outcomes included age, sex, and socioeconomic status (e.g., maternal education, household income, marital status, crowding). Additional covariates considered important for this evaluation, depending on the study population and outcome, included race/ethnicity; maternal demographics (e.g., maternal age, body mass index [BMI]); parental behavioral and mental health disorders (e.g., attention deficit hyperactivity disorder [ADHD], depression); smoking (e.g., maternal smoking status, secondhand tobacco smoke exposure); reproductive factors (e.g., parity); nutrition (e.g., BMI, growth, anemia); iodine deficiency/excess; minerals and other chemicals in water associated with neurotoxicity (e.g., arsenic, lead); maternal and paternal IQ; and quantity and quality of caregiving environment

(e.g., Home Observation Measurement of the Environment [HOME] score). To be assigned a rating of probably low risk of bias for the key risk-of-bias question regarding the confounding domain, studies were not required to address every important covariate listed; however, studies were required to address the three key covariates for all studies, the potential for co-exposures, if applicable (e.g., arsenic and lead, both of which could affect cognitive function), and any other potential covariates considered important for the specific study population and outcome. For example, studies of populations in China, India, and Mexico, where there is concern about co-exposures to high fluoride and high arsenic, were required to address arsenic. If the authors did not directly specify that arsenic exposures were evaluated, groundwater quality maps were evaluated (https://www.gapmaps.org/Home/Public) in order to identify areas of China, India, and Mexico where arsenic is a concern (Podgorski and Berg 2020). If no arsenic measurements were available for the area, the arsenic groundwater quality predictions from the global arsenic 2020 map were used (Podgorski and Berg 2020). If an area had less than 50% probability of having arsenic levels greater than 10 µg/L (the WHO guideline concentration), the area was considered not to have an issue with arsenic that needed to be addressed by the study authors; however, it should be noted that arsenic may be associated with neurodevelopmental effects at concentrations below 10 µg/L.

Exposure

Fluoride ion is rapidly absorbed from the gastrointestinal tract and is rapidly cleared from serum by distribution into calcified tissues and urinary excretion (IPCS 2002). There is general consensus that the best measures of long-term fluoride exposure are bone and/or tooth measurements, and other than measures of dental fluorosis, these were not performed in any of the studies reviewed in this document. Prolonged residence in an area with a given fluoride content in drinking water has been considered in many studies as a proxy for long-term exposure.

Exposure was assessed using a variety of methods in the human body of evidence. Studies provided varying levels of details on the methods used and employed different exposure characterization methods to group study subjects into exposed and reference groups. Exposure metrics included spot urine (from children or mothers during at least one trimester of gestation), serum, individual drinking water, intake from infant formula, estimated total exposure dose, municipal drinking water (with residence information), evidence of dental or skeletal fluorosis, area of residence (endemic versus a non-endemic fluorosis area, with or without individual validation of exposure), burning coal (with or without fluoride), and occupation type.

Urinary fluoride levels measured during pregnancy and in children include all ingested fluoride and are considered a valid measure to estimate total fluoride exposure (Villa et al. 2010; Watanabe et al. 1995); however, the type and timing of urinary sample collection are important to consider. Urinary fluoride is thought to reflect recent exposure but can be influenced by the timing of exposure (e.g., when water was last consumed, when teeth were last brushed). When compared with 24-hour urine samples, spot urine samples are more prone to the influence of timing of exposure and can also be affected by differences in dilution; however, many studies attempted to account for dilution either by using urinary creatinine or specific gravity. Good correlations between 24-hour samples and urinary fluoride concentrations from spot samples adjusted for urinary dilution have been described (Zohouri et al. 2006). Despite potential issues with spot urine samples, if authors made appropriate efforts to reduce the concern for bias (e.g.,

accounting for dilution), studies that used this metric were generally considered to have probably low risk of bias for exposure.

Analytical methods to measure fluoride in biological or water samples also varied, some of which included atomic absorption, ion-selective electrode methods, colorimetric methods, or the hexamethyldisiloxane microdiffusion method. Individual-level measures of exposure were generally considered more accurate than group-level measures; however, using group-level measures (e.g., endemic versus non-endemic area) in an analysis was less of a concern if the study provided water or urinary fluoride levels from some individuals to verify that there were differences in the fluoride exposure between groups. Studies that provided results by area and also reported individual urinary or serum fluoride concentrations or other biochemical measures, including dental fluorosis in the children or urinary levels in mothers during pregnancy, were considered to have probably low risk of bias. Ideally, these studies would still need to consider and adjust for area-level clustering; however, these concerns are captured in evaluations of other potential threats to internal validity.

Outcome

Studies included in this evaluation used a wide variety of methods to measure IQ and other cognitive effects. Measures of IQ were generally standardized tests of IQ; however, for these standardized methods to be considered low potential for bias, they needed to be conducted in the appropriate population or modified for the study population. Because results of many of the tests to measure neurodevelopment and cognitive function can be subjective, it was important that the outcome assessors were blind to the fluoride exposure when evaluating the results of the tests. If the study reported that the assessor was blind to the exposure, this was assumed to mean that the outcome assessor did not have any knowledge of the exposure, including whether the study subjects were from high-fluoride communities. If cross-sectional studies collected biomarker measurements at the time of an IQ assessment, this was considered indirect evidence that the outcome assessor would not have knowledge of the fluoride exposure unless there was also potential for the outcome assessor to have knowledge of varying levels of fluoride by study area. In cases wherein the study did not specify that the outcome assessors were blind, the study authors were contacted and asked whether the outcome assessors were, in fact, blind to exposure. When authors responded and indicated that outcome assessors were blind to exposure or that it was not likely that they would have had knowledge of exposure, this was considered direct or indirect evidence, respectively, that blinding was not a concern for those studies.

Any discrepancies in ratings between assessors were resolved by a senior technical specialist and through discussion when necessary to reach the final recorded risk-of-bias rating for each question along with a statement of the basis for that rating. Members of the evaluation team were consulted for assistance if additional expertise was necessary to reach final risk-of-bias ratings based on specific aspects of study design or performance reported for individual studies. Study procedures that were not reported were assumed not to have been conducted, resulting in an assessment of “probably high” risk of bias. Authors were queried by email to obtain missing information, and responses received were used to update risk-of-bias ratings.

Table 4. OHAT Risk-of-bias Questions and Applicability by Study Design

Risk-of-bias Questions	Experimental Animal[a]	Human Controlled Trials[b]	Cohort	Case-control	Cross-sectional[c]	Case Series
1. Was administered dose or exposure level adequately randomized?	X	X				
2. Was allocation to study groups adequately concealed?	X	X				
3. Did selection of study participants result in the appropriate comparison groups?			X	X	X	
4. Did study design or analysis account for important confounding and modifying variables?			X	X	X	X
5. Were experimental conditions identical across study groups?	X					
6. Were research personnel blinded to the study group during the study?	X	X				
7. Were outcome data complete without attrition or exclusion from analysis?	X	X	X	X	X	
8. Can we be confident in the exposure characterization?	X	X	X	X	X	X
9. Can we be confident in the outcome assessment (including blinding of outcome assessors)?	X	X	X	X	X	X
10. Were all measured outcomes reported?	X	X	X	X	X	X
11. Were there no other potential threats to internal validity?	X	X	X	X	X	X

[a]Experimental animal studies are controlled exposure studies. Non-human animal observational studies can be evaluated using the design features of observational human studies such as cross-sectional study design.
[b]Human Controlled Trials are studies in humans with controlled exposure (e.g., randomized controlled trials, non-randomized experimental studies).
[c]Cross-sectional studies include population surveys with individual data (e.g., NHANES) and surveys with aggregate data (i.e., ecological studies).

Answers to the risk-of-bias questions result in one of the following four risk-of-bias ratings:

Table 5. The Four Risk-of-bias Rating Options

Symbol	Description
++	***Definitely Low*** *risk of bias*: There is direct evidence of low risk-of-bias practices.
+	***Probably Low*** *risk of bias*: There is indirect evidence of low risk-of-bias practices, **OR** it is deemed that deviations from low risk-of-bias practices for these criteria during the study would not appreciably bias results, including consideration of direction and magnitude of bias.
– NR	***Probably High*** *risk of bias*: There is indirect evidence of high risk-of-bias practices (indicated with "–"), **OR** there is insufficient information provided about relevant risk-of-bias practices (indicated with "NR" for not reported). Both symbols indicate probably high risk of bias.
—	***Definitely High*** *risk of bias*: There is direct evidence of high risk-of-bias practices.

Organizing and Rating Confidence in Bodies of Evidence

Health Outcome Categories for Neurodevelopmental and Cognitive Effects

After data were extracted from all studies, the health effects results within the category of neurodevelopmental or cognitive effects were grouped across studies to develop bodies of evidence or collections of studies with data on the same or related outcomes. The grouping of health effect results was not planned a priori. The vast majority of the human studies evaluated IQ in children as the single outcome; therefore, the discussion of cognitive neurodevelopmental effects in children focuses on IQ studies with supporting information from data on other endpoints. Cognitive function in adults was evaluated separately. Consistent with the NTP (2016) assessment, the primary focus within the animal study body of evidence was on animal studies with endpoints related to learning and memory.

Considerations for Pursuing a Narrative or Quantitative Evidence Synthesis

This evaluation provides only a narrative review of the data; however, heterogeneity within the available evidence was evaluated to determine whether a quantitative synthesis (i.e., meta-analysis) would be appropriate. Choi et al. (2012) and Duan et al. (2018) conducted meta-analyses and found that high fluoride exposure was associated with lower IQ scores. Choi et al. (2012) was able to determine a risk ratio for living in an endemic fluorosis area but was unable to develop a dose-response relationship. Duan et al. (2018) reported a significant non-linear dose-response relationship between fluoride dose and intelligence with the relationship stated as most evident with exposures from drinking water above 4 mg/L (or 4 ppm) fluoride. Duan et al. (2018) found similar results as Choi et al. (2012) for the standardized mean difference; however, the majority of the available studies in both analyses compare populations with high fluoride exposure to those with lower fluoride exposure (with the lower exposure levels frequently in the range of drinking water fluoridation in the United States). The meta-analysis conducted in

association with this systematic review further informs this issue and will be published separately.

Confidence Rating: Assessment of Body of Evidence

The quality of evidence for neurodevelopmental and cognitive function outcomes was evaluated using the GRADE system for rating the confidence in the body of evidence (Guyatt et al. 2011; Rooney et al. 2014). More detailed guidance on reaching confidence ratings in the body of evidence as "high," "moderate," "low," or "very low" is provided in the protocol (https://ntp.niehs.nih.gov/go/785076). In brief, available human and animal studies on a particular health outcome were initially grouped by key study design features, and each grouping of studies was given an initial confidence rating by those features. Starting at this initial rating (see column 1 of Figure 1), potential downgrading of the confidence rating was considered for factors that decrease confidence in the results (see column 2 of Figure 1). Potential upgrading of the confidence rating was considered for factors that increase confidence in the results (see column 3 of Figure 1). Short descriptions of the factors that can decrease or increase confidence in the body of evidence for human studies are provided below (see protocol [https://ntp.niehs.nih.gov/go/785076] for additional details related to the human body of evidence, as well as considerations for experimental animal studies).

Factors to Consider for Potential Downgrading

- Risk of bias: Addresses whether the body of evidence did not account for critical factors in study quality or design, including confounding bias, selection bias, exposure assessment, and outcome assessment. Consideration for downgrading the confidence rating is based on the entire body of evidence, and the evidence is downgraded when there is substantial bias across most studies that could lead to decreased confidence in the results and when the studies without substantial bias could not support the confidence rating. Individual studies are evaluated for risk of bias based on a set of criteria (as discussed above); magnitude and direction of the bias are also considered.
- Unexplained inconsistency: Addresses inconsistencies in results across studies of similar populations and design that can be determined by assessing similarity of point estimates and extent of overlap between confidence intervals or more formally through statistical tests of heterogeneity. Sensitivity analysis can be used to assess the impact of specific variables on the outcome. Inconsistencies that can be plausibly explained by characteristics of the studies (e.g., sex-associated differences) are typically not used to support a downgrade. A downgrade would only be applied when there is an inconsistency that cannot be explained and results in reduced confidence in the body of evidence.
- Indirectness: Addresses generalizability and relevance to the objective of the assessment. As outlined in the Objective and consistent with the population specified in the PECO statement, this systematic review evaluated the extent and quality of the evidence linking fluoride exposure to neurodevelopmental and cognitive effects in humans without restriction as to age, sex, geographic location, or life stage at exposure or outcome assessment. Furthermore, the review did not exclude subjects exposed in occupational settings. All exposure levels and scenarios encountered in

human studies are considered direct (i.e., applicable, generalizable, and relevant to address the objective of the assessment); therefore, a downgrade for indirectness would not be applied to bodies of evidence from human studies.

- Imprecision: Addresses confidence associated with variability in quantitative measures such as effect sizes. Typically, 95% confidence intervals are used as the primary method to assess imprecision, but considerations can also be made on whether studies were adequately powered. Meta-analyses can also be used to determine whether the data are imprecise. When a meta-analysis is not appropriate or feasible, imprecision can be based on variability around the effect estimate. A downgrade would occur if the body of evidence was considered to be imprecise based on a meta-analysis, or if serious or very serious imprecision was consistently present in the body of evidence. A downgrade is especially likely if imprecision raised questions as to whether an overall effect was significant.
- Publication bias: Addresses evidence of biased publication practices. Downgrade if one strongly detects publication bias. Publication bias is difficult to detect but may be evident if major sections of the research community are not publishing (e.g., absence of industry, academic, or government studies) on a topic or if there are multiple instances wherein data from conference abstracts are never published in peer-reviewed journals. In addition, there are methods included in conducting a meta-analysis to detect whether there is potential for publication bias, including the use of fit-and-trim models, which help identify how publication bias may affect the results of the meta-analysis. Although a meta-analysis is not included in this systematic review, there are two published meta-analyses (Choi et al. 2012; Duan et al. 2018) in addition to the one associated with this systematic review (manuscript in progress) that can be used to address publication bias.

Factors to Consider for Potential Upgrading

- Large magnitude of effect: Factors to consider include the outcome being measured and the dose or exposure range assessed. The confidence can be upgraded if the body of evidence is suggestive of a large magnitude of effect. GRADE provides guidance on what can be considered a large magnitude of effect based on relative risk (i.e., suggests one upgrade in confidence if relative risk is greater than 2 and two upgrades in confidence if greater than 5). However, not all studies provide data as a risk estimate, and smaller changes, such as increases in blood pressure, may have greater impact on health at the population level. Consideration for an upgrade is not based on a single study, and what constitutes a large magnitude of effect will depend on the outcome and the potential public health impact.
- Dose response: Patterns of dose response are evaluated within and across studies. Confidence in the body of evidence can be increased when there is sufficient evidence of a dose-response pattern across multiple studies.
- Consistency: Does not apply in this evaluation. The consideration of a potential upgrade for consistency is primarily for non-human animal evidence in which it would be applied to address increased confidence based on an observation of consistent effects across multiple non-human animal species. For human evidence, this factor would generally not be applied. Human studies are instead evaluated for

issues of consistency that could result in downgrading confidence for unexplained inconsistency (see "Factors to Consider for Potential Downgrading" above).

- Consideration of residual confounding: Applies to observational studies and refers to consideration of unmeasured determinants that are likely to be distributed unevenly across groups. Residual confounding can push results in either direction, but confidence in the results is increased when the body of evidence is biased by factors that counter the observed effect and would cause an underestimation of the effect. Confounding that would cause an overestimation of the effect is considered under the risk-of-bias considerations for decreasing confidence.

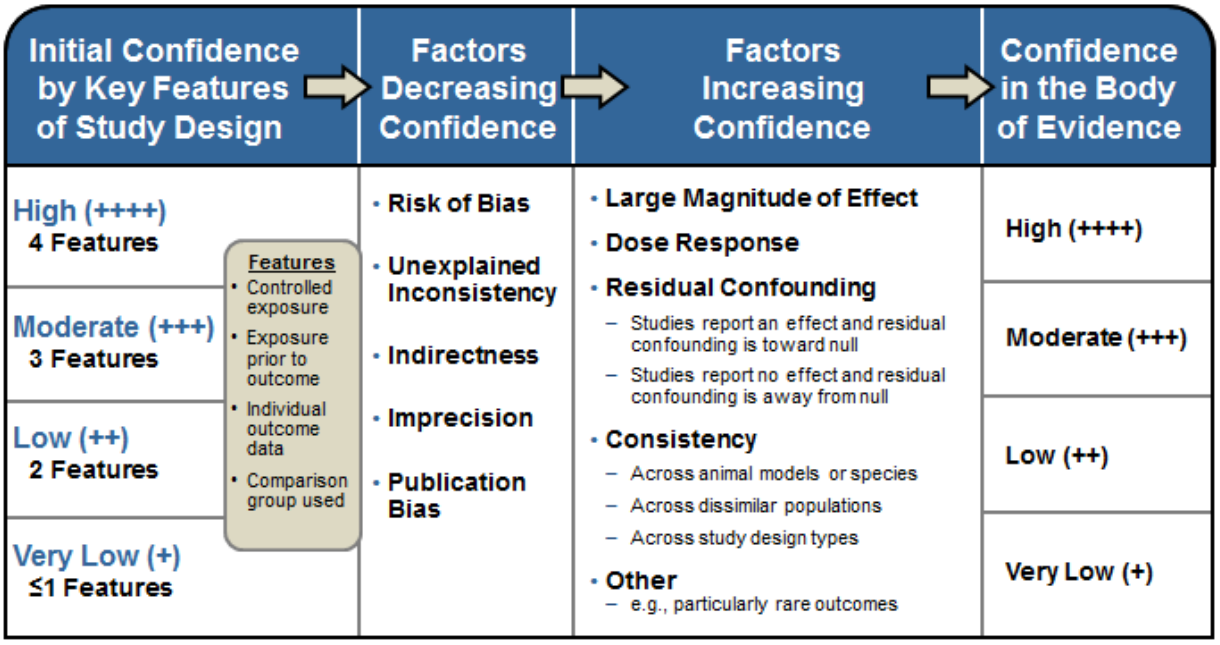


Figure 1. Assessing Confidence in the Body of Evidence

Confidence ratings were assessed by the evaluation team for accuracy and consistency, and discrepancies were resolved by consensus and consultation with technical advisors as needed. Confidence ratings for the primary outcomes are summarized in evidence profile tables for each outcome.

Results

Literature Search Results

The electronic database searches retrieved 25,450 unique references with 11 additional references[10] identified by technical advisors or obtained by manually searching the Fluoride Action Network website or reviewing reference lists of published reviews and other included studies. During title and abstract screening, 1,036 references were moved to full-text review and 24,425 were excluded (11,402 by manual screening for not satisfying the PECO criteria and 13,023 based on the SWIFT-Active Screener algorithm). Among the 1,036 references that underwent full-text review, 547 studies were considered PECO-relevant (see Appendix C for list of included studies). A few studies assessed data for more than one evidence stream (human, non-human mammal, and/or in vitro), and several studies assessed more than one type of outcome (e.g., primary and secondary outcomes). Included studies break down as follows:

- 167 human studies (84 primary only; 13 secondary only; 5 primary and secondary; 8 primary and thyroid; 2 secondary and thyroid; and 55 thyroid only);
- 339 non-human mammal studies (7 primary only; 186 secondary only; 67 primary and secondary; 6 primary, secondary, and thyroid; 4 secondary and thyroid; and 69 thyroid only); and,
- 60 in vitro/mechanistic studies (48 neurological and 12 thyroid).

Additional details on the screening results are provided in Appendix C. These screening results are outlined in a study selection diagram that reports numbers of studies excluded at each stage and documents the reason for exclusion at the full-text review stage (see Figure 2) [using reporting practices outlined in Moher et al. (2009)].

[10]These 11 studies (9 human and 2 animal studies) were not identified through the electronic database searches, as they were not indexed in any of the electronic databases searched. Note that the supplemental search of non-English-language databases was designed in part to identify non-English-language studies that are not indexed in traditional bibliographic databases such as PubMed. It was successful in this goal, as multiple studies that were initially only identified through “other sources” were subsequently captured in the supplemental Chinese database search, leaving only 11 as identified through other sources.

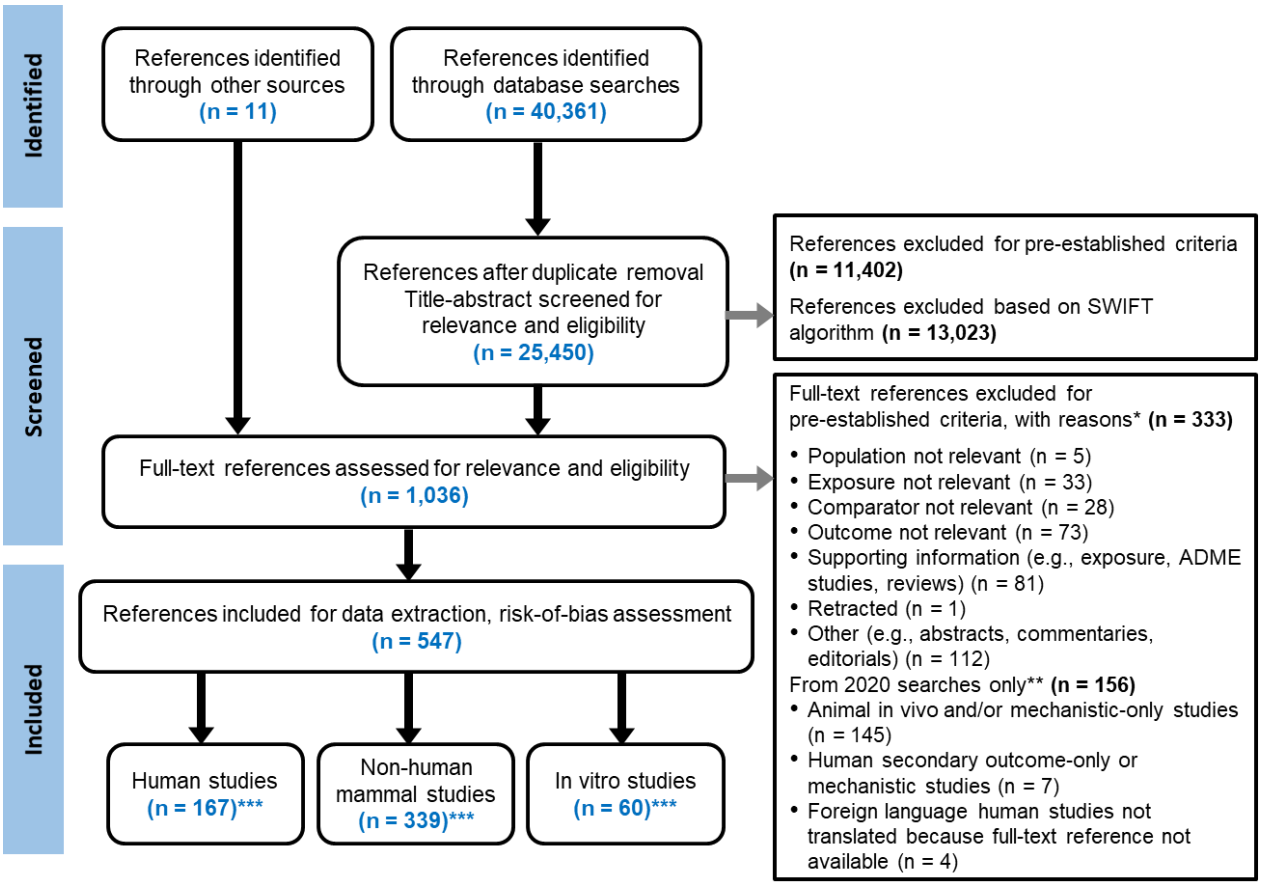


Figure 2. Study Selection Diagram[a]

[a]An interactive reference flow diagram is available here: https://hawcproject.org/summary/visual/assessment/405/Figure-2/.
*Includes studies from all literature searches conducted during the review; see the Methods section for extraction and search update information. Studies may have been excluded for more than one reason; the first reason identified was recorded.
**Includes all studies from all 2020 literature searches not otherwise excluded for pre-established criteria; see the Methods section for extraction and search update information.
***Publications may contain more than one evidence stream, so the numbers will not total the 547 included studies.

Human Neurodevelopmental and Cognitive Data

The body of literature that evaluates the association between fluoride exposure and neurodevelopmental and cognitive effects in humans is relatively robust with a large number of studies (n = 100) that cover a wide array of endpoints (see Figure 3). Seventy-two human studies investigated IQ in children. Additional studies evaluated learning and memory (n = 9 studies) or other cognitive developmental effects (e.g., total neurobehavioral scores and total mental capacity index in children, cognitive impairment in adults; n = 15 studies).[11] For this review, the evidence in children and adults was evaluated separately to address potential differences in the health impact of fluoride exposure during development versus adulthood.

[11]Some studies are included in more than one endpoint category (e.g., IQ and other cognitive developmental effects); therefore, these counts are not mutually exclusive.

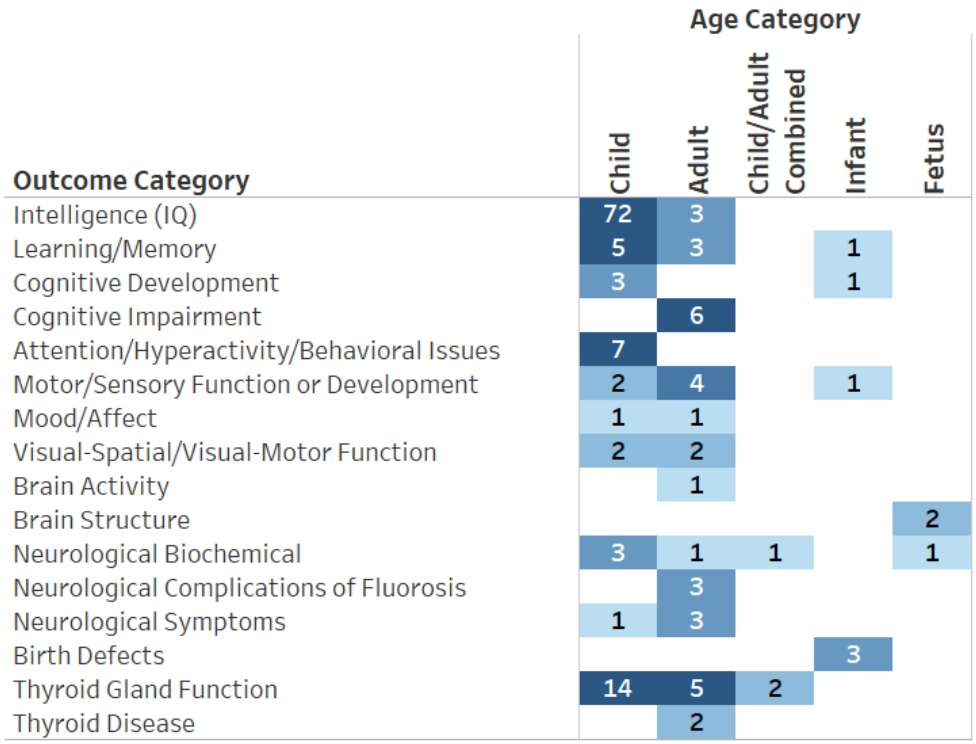


Figure 3. Number of Epidemiological Studies by Outcome and Age Categories[a]

[a]Interactive figure and additional study details in Tableau®. (https://public.tableau.com/app/profile/ntp.visuals/viz/Fluoride_Epi_2022Update/Figure3?publish=yes)
Choi et al. (2015) used subtests of the omnibus IQ test reported by the authors as Wechsler Intelligence Scale for Children-Revised (WISC-IV) to evaluate visuospatial abilities (using block design) and executive function (using digit span). These endpoints are included in the intelligence (IQ) outcome category as they are subsets of the IQ tests.
Three additional publications based on subsamples (i.e., 50–60 children) of the larger Yu et al. (2018) cohort were identified (Zhao et al. 2019; Zhao et al. 2020; Zhou et al. 2019) and are not included in the counts of this figure.

Because the majority of studies evaluated intelligence, the following section focuses on IQ effects in children followed by separate discussions on other measures of cognitive function and neurobehavioral effects in children and cognitive effects in adults. Studies that evaluated mechanistic data in humans, including effects on the thyroid, are discussed in the Mechanistic Data in Humans section. Note that a few studies were identified on congenital neurological malformations and neurological complications of fluorosis; however, they are not considered further due to the limited number of studies and the heterogeneity of outcomes evaluated in those studies.

IQ in Children

Seventy-two epidemiological studies were identified that evaluated the association between fluoride exposure and children's IQ. Nineteen of the 72 IQ studies were determined to have low potential for bias (i.e., were of high quality). Looking across the literature, there has been a progression over the years in the quality of studies conducted to assess the association between fluoride exposure and IQ in children, with more recent studies including better study designs, larger sample sizes, and more sophisticated statistical analysis. Older studies often had limitations related to study design or methods, and most of the high risk-of-bias studies (i.e.,

studies of low quality) were published prior to the 2006 NRC evaluation of fluoride in drinking water. In contrast, 18 of the low risk-of-bias studies were published after the 2006 NRC evaluation of fluoride in drinking water, and over half of those were published between 2015 and 2020 (Figure 4).

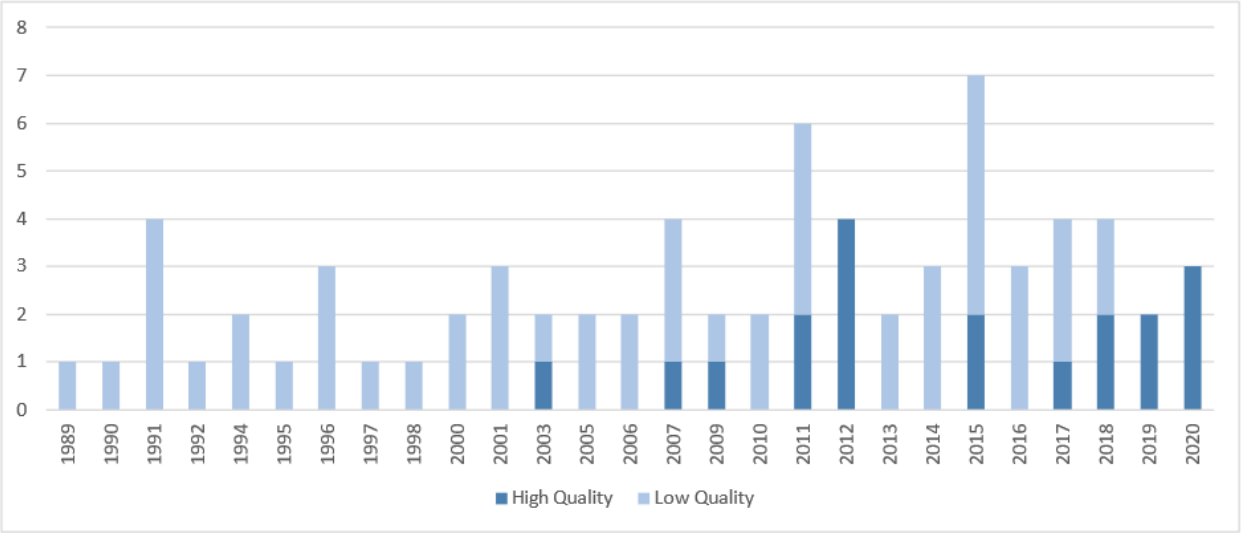


Figure 4. Number of High- and Low-quality Studies of Fluoride Exposure and IQ in Children by Year of Publication

Several characteristics of recent studies contribute to higher study quality in the overall body of literature on children's IQ and fluoride, including:

- Demonstration that exposure occurred prior to outcome assessment (an important factor when considering confidence in study results; see Figure 1) either by study design (e.g., for prospective cohort studies) or analysis (e.g., prevalence of dental fluorosis in children, limiting study populations to children who lived in the same area for long periods of time).
- Improved reporting of key study details that are necessary to evaluate study quality and allow for a more precise analysis of risk of bias.
- Increased consideration of key covariates (e.g., socioeconomic status) including potential co-exposures (e.g., arsenic or lead intake).
- Increased use of individual-level exposure measures (urine or water) as well as prenatal fluoride exposure to assess either individual-level fluoride exposure or—if still using group-level data—to confirm that regions being compared had differences in fluoride exposure.
- Utilization of more sophisticated sampling techniques for the study populations (e.g., stratified multistage random sampling).
- Application of more sophisticated regression approaches (e.g., piecewise linear regression models, multi-level regression with random effects, or generalized additive models for longitudinal measurements of fluoride).

- For studies using individual-level exposure measures, application of more sophisticated regression techniques to account for clustering at the cohort level by using cohort as a fixed or random effect and by accounting for numerous covariates that capture the cohort effect.

In addition, newer studies represent more diverse study populations across several countries (Figure 5), whereas all identified peer-reviewed studies that were published prior to 2006 took place in a single country (China). The majority of high-quality, low risk-of-bias studies exhibit these important study design and analysis characteristics, as discussed further in subsequent sections.

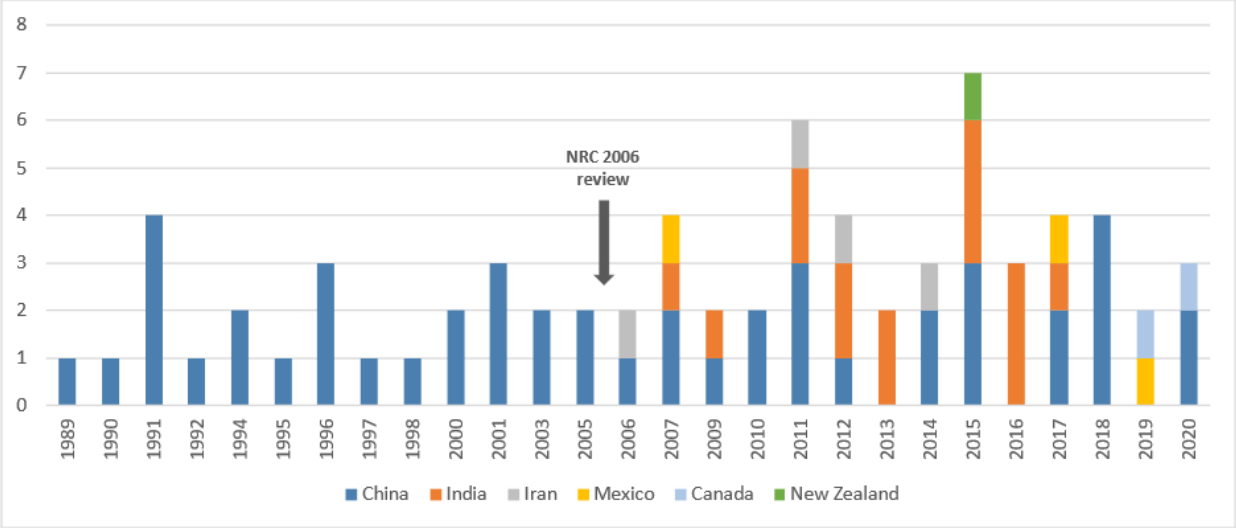


Figure 5. Number of Studies of Fluoride Exposure and IQ in Children by Country and Year of Publication

All available studies were considered in this evaluation; however, review of the body of evidence focused on the high-quality, low risk-of-bias studies for two main reasons. First, there are fewer limitations and greater confidence in the results of the high-quality studies. Second, there are a relatively large number of high-quality studies (n = 19), such that the body of evidence from these studies could be used to evaluate confidence in the association between fluoride exposure and changes in children's IQ. Therefore, the remainder of the discussion on IQ in children focuses on the 19 studies with low risk of bias. The high risk-of-bias studies are discussed briefly relative to their overall support of findings from the low risk-of-bias studies.

Low Risk-of-bias IQ Studies

Overview of Studies

Nineteen studies (3 longitudinal prospective cohort and 16 cross-sectional studies) with low potential for bias evaluated the association between fluoride exposure and IQ in children (see Quality Assessment of Individual Studies section for methods on determining which studies pose low risk of bias). These IQ studies were conducted in 15 study populations across 5 countries

and included more than 7,000 children. Specifically, of the 19 low risk-of-bias studies of IQ in children:

- ten were conducted in four areas of China on seven study populations,[12]
- three were conducted in three areas of Mexico on three study populations,
- two were conducted in Canada using the same study population,
- three were conducted in three areas of India on three study populations, and
- one was conducted in Iran.

Most studies measured fluoride in drinking water (n = 15) and/or urine (child or maternal) (n = 15). Two studies measured fluoride in serum. The IQ studies used a variety of tests to measure IQ. Because IQ tests should be culturally relevant, the tests used often differed between studies, reflecting adjustments for the range in populations studied (e.g., western vs. Asian populations). In some cases, different IQ tests were used to study similar populations. Overall, these studies used IQ tests that were population- and age-appropriate.

Table 6 provides a summary of study characteristics and key IQ and fluoride findings for the 19 low risk-of-bias studies. Several of these studies conducted multiple analyses and reported results on multiple endpoints. The purpose of the table is to summarize key findings (independent of whether an association is indicated) from each study and is not meant to be a comprehensive summary of all results from each study. For each study, results are summarized for each exposure measure assessed, but results from multiple analyses using the same exposure measure may not be presented for all studies unless multiple analyses yielded conflicting results. See Appendix E for additional information on each study in Table 6, including strengths and limitations, clarifications for why studies are considered to pose low risk of bias, and information regarding statistical analyses, important covariates, exposure assessment, and outcome assessment.

[12]In this document, "study population" refers to a defined population on which an original body of research was conducted. The published work drawn from that original body of research is often referred to as a "study." IQ studies that report on the same study populations are identified in Table 6.

Table 6. Studies on IQ in Children[a]

Study	Study Design (Location/Subjects) [n]	Exposure Measures and Summary Statistics	Assessment Timing	Outcome and Methods	Summary of IQ Results[b,c]
China					
Xiang et al. (2003a)[d]	Cross-sectional Wamiao and Xinhuai villages (Sihong County)/school children [512]	Drinking water Mean (SD): 0.36 (0.15) (control), 2.47 (0.79) (high fluoride) mg/L Children's urine Mean (SD): 1.11 (0.39) (control), 3.47 (1.95) (high fluoride) mg/L Village of residence (non-endemic vs. endemic fluorosis)	Children (ages 8–13 years)	IQ: Combined Raven's Test for Rural China	Significant dose-related association of fluoride on IQ score based on drinking water quintile levels with significantly lower IQ scores observed at water fluoride levels of 1.53 mg/L or higher; % of subjects with IQ <80 was significantly increased at water levels 2.46 mg/L or higher; significant inverse correlation between IQ and urinary fluoride (Pearson correlation coefficient of −0.164); mean IQ scores for children in non-endemic region (100.41 ± 13.21) significantly higher than endemic region (92.02 ± 13.00) No statistical adjustment for covariates
Ding et al. (2011)	Cross-sectional Inner Mongolia (Hulunbuir City)/elementary school children [331]	Children's urine Range: 0.1–3.55 mg/L Drinking water (reported but not used in analyses) Mean (SD): 1.31 (1.05) mg/L	Children (ages 7–14 years)	IQ: Combined Raven's Test for Rural China	Significant association between urinary fluoride and IQ score (each 1-mg/L increase was associated with a decrease in IQ score of 0.59 points; 95% CI: −1.09, −0.08) Adjusted for age
Xiang et al. (2011)[d]	Cross-sectional Wamiao and Xinhuai villages (Sihong County)/school children [512]	Children's serum Mean (SD): 0.041 (0.009) (control), 0.081 (0.019) (high fluoride) mg/L	Children (ages 8–13 years)	IQ: Combined Raven's Test for Rural China	Significant linear trend across quartiles of serum fluoride and children's IQ score <80 (adjusted ORs for Q1 and Q2; Q1 and Q3; and Q1 and Q4, respectively: 1; 2.22 [95% CI: 1.42, 3.47]; and 2.48 [95% CI: 1.85, 3.32]); significant associations at ≥0.05 mg/L serum fluoride Adjusted for age and sex

Study	Study Design (Location/Subjects) [n]	Exposure Measures and Summary Statistics	Assessment Timing	Outcome and Methods	Summary of IQ Results[b,c]
Wang et al. (2012)[d]	Cross-sectional Wamiao and Xinhuai villages (Sihong County)/school children [526]	Children's total fluoride intake Mean (SD): 0.78 (0.13) (control), 3.05 (0.99) (high fluoride) mg/day Village of residence (non-endemic vs. endemic fluorosis) Drinking water (reported for villages but not used in analyses) Mean (SD): 0.36 (0.11) (control), 2.45 (0.80) (high fluoride) mg/L	Children (ages 8–13 years)	IQ: Combined Raven's Test for Rural China	Significantly lower mean IQ in the endemic versus non-endemic regions, as reported in Xiang et al. (2003a); when high-exposure group was broken into four exposure groups based on fluoride intake, a dose-dependent decrease in IQ and increase in % with low IQ observed; significant correlation between total fluoride intake and IQ ($r = -0.332$); for IQ <80, adjusted OR of total fluoride intake per 1-mg/(person/day) was 1.106 (95% CI: 1.052, 1.163) Adjusted for age and sex
Choi et al. (2015)	Cross-sectional Mianning County/1st grade children [51]	Drinking water GM: 2.20 mg/L Children's urine GM: 1.64 mg/L Severity of fluorosis (Dean Index)	Children (ages 6–8 years)	IQ: WISC-IV (block design and digit span)	Compared to normal/questionable fluorosis, presence of moderate/severe fluorosis significantly associated with lower total (adjusted $\beta = -4.28$; 95% CI: −8.22, −0.33) and backward (adjusted $\beta = -2.13$; 95% CI: −4.24, −0.02) digit span scores; linear associations between total digit span and log-transformed urinary fluoride (adjusted $\beta = -1.67$; 95% CI: −5.46, 2.12) and log-transformed drinking water fluoride (adjusted $\beta = -1.39$; 95% CI: −6.76, 3.98) observed but not significant; forward digit span had similar results as backward and total but was not statistically significant; block design (square root transformed) not significantly associated with any measure of fluoride exposure Adjusted for age and sex, parity, illness before 3 years old, household income last year, and caretaker's age and education

Study	Study Design (Location/Subjects) [n]	Exposure Measures and Summary Statistics	Assessment Timing	Outcome and Methods	Summary of IQ Results[b,c]
Zhang et al. (2015b)	Cross-sectional Tianjin City (Jinnan District)/school children [180]	Drinking water Mean: 0.63 (control), 1.40 (endemic fluorosis) mg/L (SD not reported) Children's urine Mean (SD): 1.1 (0.67) (control), 2.4 (1.01) (endemic fluorosis) mg/L Children's serum Mean (SD): 0.06 (0.03) (control), 0.18 (0.11) (endemic fluorosis) mg/L	Children (ages 10–12 years)	IQ: Combined Raven's Test for Rural China	Significant correlation between IQ score and children's serum fluoride ($r = -0.47$) and urinary fluoride ($r = -0.45$); significant difference in mean IQ score for high-fluoride area (defined as >1 mg/L in drinking water; 102.33 ± 13.46) compared with control area (109.42 ± 13.30); % of subjects with IQ <90 significantly increased in high-fluoride area (28.7%) vs. low-fluoride area (8.33%); not significantly correlated with water fluoride Adjusted for age and sex, if applicable
Cui et al. (2018)	Cross-sectional Tianjin City (districts Jinghai and Dagang)/school children [323]	Children's urine Median (Q1–Q3): 1.3 (0.9–1.7) mg/L (boys), 1.2 (0.9–1.6) mg/L (girls)	Children (ages 7–12 years)	IQ: Combined Raven's Test for Rural China	Significant association between IQ score and log-transformed urinary fluoride (adjusted $\beta = -2.47$; 95% CI: −4.93, −0.01) Adjusted for age, mother's education, family member smoking, stress, and anger

Study	Study Design (Location/Subjects) [n]	Exposure Measures and Summary Statistics	Assessment Timing	Outcome and Methods	Summary of IQ Results[b,c]
Yu et al. (2018)[e,f]	Cross-sectional Tianjin City (7 towns)/children [2,886]	Drinking water Mean (SD): 0.50 (0.27) (normal), 2.00 (0.75) (high) mg/L Children's urine Mean (SD): 0.41 (0.49) (normal), 1.37 (1.08) (high) mg/L	Children (ages 7–13 years)	IQ: Combined Raven's Test for Rural China	Significant difference in mean IQ scores in high water fluoride areas (>1.0 mg/L; 106.4 ± 12.3 IQ) compared to the normal water fluoride areas (≤1.0 mg/L; 107.4 ± 13.0); distribution of the IQ scores also significantly different (p = 0.003); every 0.5-mg/L increase in water fluoride was associated with a decrease of 4.29 in IQ score (95% CI: −8.09, −0.48) when exposure was between 3.40 and 3.90 mg/L; no significant association between 0.2 and 3.40 mg/L; every 0.5-mg/L increase in urinary fluoride was associated with a decrease of 2.67 in IQ score (95% CI: −4.67, −0.68) between 1.60 and 2.50 mg/L but not at levels of 0.01–1.60 mg/L or 2.50–5.54 mg/L. Adjusted for age and sex, maternal education, paternal education, and low birth weight
Cui et al. (2020)	Cross-sectional Tianjin City (all districts)/school children (potentially some overlap with Cui et al. (2018)) [498]	Children's urine <1.6–≥2.5 mg/L	Children (ages 7–12 years)	IQ: Combined Raven's Test	Decreasing mean (± SD) IQ score with increasing urinary fluoride levels (statistical significance not reached based on a one-way ANOVA) <1.6 mg/L: 112.16 ± 11.50 1.6–2.5 mg/L: 112.05 ± 12.01 ≥2.5 mg/L: 110 ± 14.92 No statistical adjustment for covariates

Study	Study Design (Location/Subjects) [n]	Exposure Measures and Summary Statistics	Assessment Timing	Outcome and Methods	Summary of IQ Results[b,c]
Wang et al. (2020b)[e]	Cross-sectional Tianjin City (villages not specified)/school children [571]	Drinking water Mean (SD): 1.39 (1.01) mg/L Children's urine Mean (SD): 1.28 (1.30) mg/L	Children (ages 7–13 years)	IQ: Combined Raven's Test for Rural China	Significant associations between IQ and water and urinary fluoride concentrations in boys and girls combined based on both quartiles and continuous measures (water: 1.587 decrease in IQ score per 1-mg/L increase; urine: 1.214 decrease in IQ score per 1-mg/L increase); no significant effect modification of sex Adjusted for age and sex, BMI, maternal education, paternal education, household income, and low birth weight
Mexico					
Rocha-Amador et al. (2007)	Cross-sectional Moctezuma and Salitral in San Luis Potosi State and 5 de Febrero of Durango State /elementary school children [132]	Drinking water Mean (SD): 0.8 (1.4), 5.3 (0.9), 9.4 (0.9) mg/L (3 rural areas) Children's urine Mean (SD): 1.8 (1.5), 6.0 (1.6), 5.5 (3.3) mg/L (3 rural areas)	Children (ages 6–10 years)	IQ: WISC-Revised Mexican Version	Significant associations between log-transformed fluoride and IQ scores (full IQ adjusted βs of −10.2 [water] and −16.9 [urine]; CIs not reported); arsenic also present, but the association with arsenic was smaller (full-scale IQ adjusted βs of −6.15 [water] and −5.72 [urine]; CIs not reported) Adjusted for blood lead, mother's education, SES, height-for-age z-scores, and transferrin saturation

Study	Study Design (Location/Subjects) [n]	Exposure Measures and Summary Statistics	Assessment Timing	Outcome and Methods	Summary of IQ Results[b,c]
Bashash et al. (2017)	Cohort (prospective) Mexico City/Early Life Exposures in Mexico to Environmental Toxicants (ELEMENT) participants [299] IQ analysis [211]	Maternal urine during pregnancy Mean (SD): 0.90 (0.35) mg/L Children's urine Mean (SD): 0.82 (0.38) mg/L	Children (ages 6–12 years)	IQ: WASI-Spanish Version	Significantly lower child IQ score per 0.5-mg/L increase in maternal urinary fluoride (adjusted β = −2.50; 95% CI: −4.12, −0.59); no significant association with children's urine Adjusted for sex, gestational age; weight at birth; parity (being the first child); age at outcome measurement; and maternal characteristics, including smoking history (ever smoked during the pregnancy vs. nonsmoker), marital status (married vs. not married), age at delivery, education, IQ, and cohort
Soto-Barreras et al. (2019)	Cross-sectional Chihuahua/school children [161]	Children's urine Range: 0.11–2.10 mg/L Drinking water Range: 0.05–2.93 mg/L Fluoride exposure dose (summary statistics not reported) Fluorosis index (summary statistics not reported)	Children (ages 9–10 years)	IQ: Raven's Colored Progressive Matrices	No significant difference in urinary fluoride, drinking water fluoride, fluoride exposure dose, or fluorosis index in subjects across different IQ grades No statistical adjustment for covariates

Study	Study Design (Location/Subjects) [n]	Exposure Measures and Summary Statistics	Assessment Timing	Outcome and Methods	Summary of IQ Results[b,c]
Canada					
Green et al. (2019)[g]	Cohort (prospective) 10 cities/Maternal-Infant Research on Environmental Chemicals (MIREC) [512] Non-fluoridated [238] Fluoridated [162] Boys [248] Girls [264]	Maternal urine during pregnancy Mean (SD): 0.51 (0.36) mg/L (0.40 [0.27] mg/L in non-fluoridated areas and 0.69 [0.42] mg/L in fluoridated areas) Maternal fluoride intake during pregnancy Mean (SD): 0.54 (0.44) mg/day (0.30 [0.26] and 0.93 [0.43] mg/day, respectively) Drinking water Mean (SD): 0.31 (0.23) mg/L (0.13 [0.06] and 0.59 [0.08] mg/L, respectively)	Children (ages 3–4 years)	IQ: full-scale, performance, and verbal using Wechsler Preschool and Primary Scale of Intelligence, Third Edition (WPPSI-III)	Significantly lower full-scale IQ (adjusted β = −4.49; 95% CI: −8.38, −0.60) and performance IQ (adjusted β = −4.63; 95% CI: −9.01, −0.25) per 1-mg/L increase in maternal urinary fluoride in boys but not girls (adjusted β = 2.40; 95% CI: −2.53, 7.33 and adjusted β = 4.51; 95% CI: −1.02, 10.05, respectively) or boys and girls combined (adjusted β = −1.95; 95% CI: −5.19, 1.28 and adjusted β = −1.24; 95% CI: −4.88, 2.40, respectively); significantly lower full-scale IQ (adjusted β = −3.66; 95% CI: −7.16, −0.15) per 1-mg increase in maternal fluoride intake (no sex interaction); significantly lower full-scale IQ (adjusted β = −5.29; 95% CI: −10.39, −0.19) per 1-mg/L increase in water fluoride concentration (no sex interaction); no significant associations observed between measures of fluoride and verbal IQ Adjusted for sex, city, HOME score, maternal education, race, and prenatal secondhand smoke exposure

Study	Study Design (Location/Subjects) [n]	Exposure Measures and Summary Statistics	Assessment Timing	Outcome and Methods	Summary of IQ Results[b,c]
Till et al. (2020)[g]	Cohort (prospective) 10 cities/ MIREC [398] Non-fluoridated [247] Fluoridated [151] Breastfed as infants [200] Formula-fed as infants [198]	Drinking water Mean (SD) For breastfed infants: 0.13 (0.06) mg/L in non-fluoridated areas and 0.58 (0.08) mg/L in fluoridated areas For formula-fed infants: 0.13 (0.05) mg/day in non-fluoridated areas and 0.59 (0.07) mg/L in fluoridated areas Infant fluoride intake Mean (SD) For breastfed infants: 0.02 (0.02) mg/day in non-fluoridated areas and 0.12 (0.07) mg/day in fluoridated areas For formula-fed infants: 0.08 (0.04) mg/day in non-fluoridated areas and 0.34 (0.12) mg/day in fluoridated areas Maternal urine during pregnancy	Children (ages 3–4 years)	IQ: full-scale, performance, and verbal using Wechsler Preschool and Primary Scale of Intelligence, Third Edition (WPPSI-III)	Drinking water Breastfed infants: Lower (not significant) full-scale IQ (adjusted β = −1.34, 95% CI: −5.04, 2.38) per 0.5-mg/L increase in water fluoride concentration; significantly lower performance IQ (adjusted β = −6.19, 95% CI: −10.45, −1.94) Formula-fed infants: Significantly lower full-scale IQ (adjusted β = −4.40, 95% CI: −8.34, −0.46) per 0.5-mg/L increase in water fluoride concentration; significantly lower performance IQ (adjusted β = −9.26, 95% CI: −13.77, −4.76) Infant fluoride intake Breastfed: No results reported Formula-fed: Lower (not significant) full-scale IQ (adjusted β = −2.69, 95% CI: −709, 3.21) per 0.5-mg/L increase in fluoride intake from formula; significantly lower performance IQ (adjusted β = −8.76, 95% CI: −14.18, −3.34) Maternal urine during pregnancy+

Study	Study Design (Location/Subjects) [n]	Exposure Measures and Summary Statistics	Assessment Timing	Outcome and Methods	Summary of IQ Results[b,c]
		Mean (SD) Breastfed: 0.42 (0.28) mg/L in non-fluoridated areas and 0.70 (0.39) mg/L in fluoridated areas Formula-fed: 0.38 (0.27) mg/L in non-fluoridated areas and 0.64 (0.37) mg/L in fluoridated areas			Lower (not significant) full-scale IQ (adjusted β = −1.08, 95% CI: −1.54, 0.47) per 0.5-mg/L increase in maternal urinary fluoride++; lower (not significant) performance IQ (adjusted β = −1.31, 95% CI: −3.63, 1.03)++ Lower (not significant) performance IQ (adjusted β = −1.50, 95% CI: −3.41, 0.43) per 0.5-mg/L increase in maternal urinary fluoride+++; significantly lower full-scale IQ (adjusted β = −2.38, 95% CI: −4.62, −0.27)+++ No association between verbal IQ scores and any measure of fluoride exposure +Maternal urinary fluoride analyzed as covariate in the drinking water and infant fluoride intake from formula models and not in an individual model ++After additional adjustment for drinking water and breastfeeding status +++After additional adjustment for infant fluoride intake from formula All models adjusted for maternal education, maternal race, age at IQ testing, sex, HOME total score, and secondhand smoke status in the child's home (separate analysis also adjusted for mother's urinary fluoride)

Study	Study Design (Location/Subjects) [n]	Exposure Measures and Summary Statistics	Assessment Timing	Outcome and Methods	Summary of IQ Results[b,c]
India					
Sudhir et al. (2009)	Cross-sectional Nalgonda District (Andhra Pradesh)/school children [1,000]	Drinking water Level 1: <0.7 mg/L Level 2: 0.7–1.2 mg/L Level 3: 1.3–4.0 mg/L Level 4: >4.0 mg/L	Children (ages 13–15 years)	IQ: Raven's Standard Progressive Matrices	Significant increase in mean and distributions of IQ grades (i.e., increase in proportion of children with intellectual impairment) with increasing drinking water fluoride levels No statistical adjustment for covariates
Saxena et al. (2012)	Cross-sectional Madhya Pradesh/school children [170]	Drinking water ≥1.5 mg/L (high fluoride group) Children's urine Range: 1.7–8.4 mg/L	Children (age 12 years)	IQ: Raven's Standard Progressive Matrices	Significant correlations between IQ grade and water (r = 0.534) and urinary (r = 0.542) fluoride levels; in adjusted analyses, significant increase in mean IQ grade (i.e., increase in proportion of children with intellectual impairment) with increasing urinary fluoride; no significant differences in the levels of urinary lead or arsenic in children with the different water fluoride exposure levels Covariates included in the analysis were not reported
Trivedi et al. (2012)	Cross-sectional Kachchh, Gujarat/school children (6th and 7th grades) [84]	Mean (SE) Low-fluoride villages: drinking water: 0.84 (0.38) mg/L Children's urine: 0.42 (0.23) mg/L High fluoride villages: drinking water: 2.3 (0.87) mg/L Children's urine: 2.69 (0.92) mg/L	Children (ages 12–13 years)	IQ: questionnaire prepared by Professor JH Shah (97% reliability rating)	Significantly lower mean IQ score in high fluoride villages (92.53 ± 3.13) compared to the low-fluoride villages (97.17 ± 2.54); differences significant for boys and girls combined, as well as separately No statistical adjustment for covariates

Study	Study Design (Location/Subjects) [n]	Exposure Measures and Summary Statistics	Assessment Timing	Outcome and Methods	Summary of IQ Results[b,c]
Iran					
Seraj et al. (2012)	Cross-sectional Makoo/school children [293]	Drinking water Mean (SD): 0.8 (0.3) (normal), 3.1 (0.9) (medium), 5.2 (1.1) (high) mg/L	Children (ages 6–11 years)	IQ: Raven's Colored Progressive Matrices	Significant association between water fluoride and IQ score (adjusted β = −3.865 per 1-mg/L increase in water fluoride); CIs not reported); significantly higher mean IQ score in normal area (97.77 ± 18.91) compared with medium (89.03 ± 12.99) and high (88.58 ± 16.01) areas Adjusted for age, sex, child's education level, mother's education level, father's education level, and fluorosis intensity

ANOVA = analysis of variance; GM = geometric mean; HOME = Home Observation Measurement of the Environment; IQ = intelligence quotient; Q1, Q3 = first and third quartiles; SD = standard deviations; WASI = Wechsler Abbreviated Scale of Intelligence (Spanish version); WISC-IV = Wechsler Intelligence Scale for Children-Revised (as reported by Choi et al. 2015).

[a]Includes low risk-of-bias studies.

[b]Associations between IQ and fluoride levels were reported quantitatively, when possible. For studies with multiple analyses and results, the table summarizes key findings and is not a comprehensive summary of all findings. Results also indicate when a study reported no association between IQ and fluoride, provided as a qualitative statement of no association.

[c]See Figure A-1 through Figure A-8 for additional study results.

[d]Xiang et al. (2003a), Xiang et al. (2011), and Wang et al. (2012) are based on the same study population.

[e]Yu et al. (2018) and Wang et al. (2020b) are based on the same study population.

[f]Three additional publications based on a subsample (i.e., 50–60 children) of the larger Yu et al. (2018) cohort were identified (Zhao et al. 2019; Zhao et al. 2020; Zhou et al. 2019); however, these publications focused on mechanistic considerations and are not included in the study totals for IQ because the main study by Yu et al. (2018) is considered a better representation of the IQ results.

[g]Green et al. (2019) and Till et al. (2020) are based on the same study population.

Summary of Results

Overall Findings

The results from 18 of the 19 high-quality (low risk-of-bias) studies (3 longitudinal prospective cohort studies from 2 different study populations and 15 cross-sectional studies from 13 different study populations) that evaluated IQ in children provide consistent evidence that higher fluoride exposure is associated with lower IQ scores (see "Summary of IQ Results" in Table 6) (Bashash et al. 2017; Choi et al. 2015; Cui et al. 2018; Ding et al. 2011; Green et al. 2019; Rocha-Amador et al. 2007; Saxena et al. 2012; Seraj et al. 2012; Sudhir et al. 2009; Till et al. 2020; Trivedi et al. 2012; Wang et al. 2012; Wang et al. 2020b; Xiang et al. 2003a; Xiang et al. 2011; Yu et al. 2018; Zhang et al. 2015b). Only one study (Soto-Barreras et al. 2019) did not observe an association between fluoride exposure and IQ; however, results were not provided in a manner that allowed for a direct comparison with other low risk-of-bias studies (see Appendix E for details). A strength of the findings across 18 of 19 low risk-of-bias studies was the consistent association between higher fluoride exposure [e.g., represented by populations whose total fluoride exposure approximated or exceeded the WHO Guidelines for Drinking-water Quality of 1.5 mg/L of fluoride (WHO 2017)] and lower IQ scores among studies of varying study designs, exposure measures, and study populations. In studies that analyzed the sexes separately (n = 5 studies with 2 studies reporting on the same study population), consistent findings of lower IQ associated with fluoride exposure were generally reported for both sexes. There is some indication of differential susceptibility between sexes, but ultimately, due to too few high-quality studies that analyzed exposure and outcome by sex separately and a lack of consistent findings that one sex is more susceptible, it is unclear whether one sex is more susceptible to the effects of fluoride exposure than the other. The body of evidence from the 19 low risk-of-bias studies is described in further detail below. Prospective cohort studies are discussed first, as this study design can establish a temporal relationship between exposure and outcome, which would contribute to demonstrating causality and, therefore, providing the strongest evidence for an association between fluoride exposure during development and IQ in children.

Results by Study Design – Prospective Cohort Studies

As noted above, three longitudinal prospective cohort studies, conducted in Mexico and Canada, were identified and considered to reflect a low risk for bias. All three prospective cohort studies found an association between increasing maternal or child fluoride exposure and lower IQ in children (Bashash et al. 2017; Green et al. 2019; Till et al. 2020). Two of the studies (Green et al. 2019; Till et al. 2020) were based on the same Canadian study population, but one evaluated prenatal fluoride exposure and the other evaluated postnatal fluoride exposure. Green et al. (2019) included maternal urinary fluoride, maternal fluoride intake, and water fluoride concentrations, while Till et al. (2020) used fluoride intake from formula or water concentrations in formula-fed versus breastfed infants. Multiple analyses were conducted in each prospective study, and results by analysis for the three prospective studies are discussed below. In summary, although not every analysis found a statistically significant association, together the three studies provided consistent evidence that increasing maternal fluoride levels were associated with lower IQ scores in the children.

In the Early Life Exposures in Mexico to Environmental Toxicants cohort, Bashash et al. (2017) observed a statistically significant association (p-value = 0.01) between lower IQ scores in children and prenatal fluoride exposure measured by maternal urinary fluoride (measured during

all three trimesters and included if at least one measurement was available). An increase of 0.5 mg/L of maternal urinary fluoride was associated with a 2.5-point decrease in IQ score [95% CI: −4.12, −0.59] in boys and girls combined (see Figure A-8). This study also reported an inverse association between IQ level and children's urinary fluoride levels (single spot urine sample); however, this specific result did not achieve statistical significance (a 0.5-mg/L increase of child urinary fluoride was associated with a 0.89-point decrease in IQ score [95% CI: −2.63, 0.85]) (Bashash et al. 2017).

In the Maternal-Infant Research on Environmental Chemicals cohort, consisting of 10 cities in Canada, Green et al. (2019) also reported inverse associations between IQ scores in children and multiple measures of prenatal fluoride exposure, including maternal urinary fluoride, maternal fluoride intake, and water fluoride concentrations. Green et al. (2019) observed a statistically significantly lower IQ for boys associated with maternal urinary fluoride averaged across trimesters (4.49-point decrease in IQ score [95% CI: −8.38, −0.60; p-value = 0.02] per 1-mg/L increase in maternal urinary fluoride); however, results were not significant in boys and girls combined (1.95-point decrease in IQ [95% CI: −5.19, 1.28]) and were positive but not significant in girls (2.40-point increase in IQ [95% CI: −2.53, 7.33]). Other measures of prenatal exposure (maternal fluoride intake or water fluoride concentrations) were associated with lower IQ scores in boys and girls combined; the authors found no significant effect measure modification between child sex and fluoride exposure in these analyses so they did not report boys and girls separately (Green et al. 2019). Specifically, when evaluating the association between estimated maternal fluoride intake based on maternal water and beverage consumption during pregnancy and IQ in children, a 1-mg increase in daily maternal consumption of fluoride during pregnancy was associated with a significantly decrease in IQ score of 3.66 points in boys and girls combined (95% CI: −7.16, −0.15; p-value = 0.04). Similarly, water fluoride concentrations for pregnant women from fluoridated areas (mean water fluoride levels of 0.59 ± 0.08 mg/L) versus pregnant women from non-fluoridated areas (mean water fluoride levels of 0.13 ± 0.06 mg/L) were associated with a significant 5.29-point decrease in IQ score per 1-mg/L increase in fluoride in both boys and girls combined (95% CI: −10.39, −0.19; p-value <0.05) (Green et al. 2019).

In a study of the same study population as Green et al. (2019) that used fluoride intake from formula or water concentrations in formula-fed versus breastfed infants, Till et al. (2020) observed significantly lower performance IQ scores with higher fluoride regardless of the comparison used (p-values ≤0.004). They did not observe any association with verbal IQ, and full-scale IQ was only significantly lower in formula-fed infants using water fluoride concentrations as the exposure measure (p-value = 0.03). Breastfed infants and fluoride intake from formula also showed inverse associations but were not significant.

Taken together, the three prospective cohort studies (based on two North American study populations) indicate consistency in results across different types of analysis and across two study populations that higher fluoride exposure during development is associated with lower IQ scores.

Results by Study Design – Cross-sectional Studies

As with the prospective cohort studies, the cross-sectional studies reported a consistent association between fluoride exposure and lower IQ scores in children. Fifteen of the 16 low risk-of-bias cross-sectional studies [i.e., all with the exception of Soto-Barreras et al. (2019)]

consistently demonstrate that exposure to fluoride is associated with lower IQ scores. Fourteen of these 15 studies [with the exception of Cui et al. (2020)] reported significant associations.

Cross-sectional studies can have limitations, as the study design often cannot ensure that exposure preceded outcome. This uncertainty reduces confidence in study findings compared with prospective cohort studies—which, by design, establish that exposure occurred prior to outcome—and is captured in the outcome assessment. In some cases, cross-sectional studies do provide indicators of prior exposure (e.g., prevalence of dental fluorosis, limiting study populations to subjects who lived in the same area for long periods of time). Evidence that exposure occurred prior to the outcome of interest increases the confidence in results and any potential association reported in these studies. Of the 16 low risk-of-bias cross-sectional studies, 12 established that exposure preceded the outcome assessment (Choi et al. 2015; Ding et al. 2011; Rocha-Amador et al. 2007; Saxena et al. 2012; Seraj et al. 2012; Soto-Barreras et al. 2019; Sudhir et al. 2009; Wang et al. 2012; Wang et al. 2020b; Xiang et al. 2003a; Xiang et al. 2011; Yu et al. 2018). Five studies from different study populations indicated that a large portion of the exposed children had dental fluorosis (ranging from 43% to 100%) at the time of assessment (Choi et al. 2015; Ding et al. 2011; Seraj et al. 2012; Sudhir et al. 2009; Yu et al. 2018). Because dental fluorosis occurs when fluoride is consumed during enamel formation (usually during the first 6–8 years of life), the presence of dental fluorosis suggests that exposures to fluoride occurred prior to the outcome assessment. Nine studies from six study populations (including Yu et al. (2018) and Sudhir et al. (2009) listed above) excluded subjects who had not lived in the study area for a specified period of time, sometimes since birth (Rocha-Amador et al. 2007; Saxena et al. 2012; Soto-Barreras et al. 2019; Sudhir et al. 2009; Wang et al. 2012; Wang et al. 2020b; Xiang et al. 2003a; Xiang et al. 2011; Yu et al. 2018). Because these areas were generally known to be fluoride-endemic for long periods of time, it can generally be assumed that in these nine studies, exposure occurred prior to the outcome. Taken together, 12 cross-sectional studies from 9 study populations provide indicators of prior exposure.

Results by Study Design – Cross-sectional Study Variations

Overall, the cross-sectional studies consistently provide evidence that fluoride exposure is associated with lower IQ scores in children. Several cross-sectional studies conducted multiple analyses (e.g., reported results for multiple exposure metrics, endpoints, subpopulations). Although some of these variations are heterogeneous and are not comparable across studies, the consistency of the results across multiple metrics contributes to the confidence in the data. Table 6 summarizes key results for each of the low risk-of-bias cross-sectional studies, and a few examples of the within-study variations in results are provided below.

Nine cross-sectional studies (from six study populations) assessed the association between IQ and multiple exposure measures (Choi et al. 2015; Rocha-Amador et al. 2007; Saxena et al. 2012; Wang et al. 2012; Wang et al. 2020b; Xiang et al. 2003a; Xiang et al. 2011; Yu et al. 2018; Zhang et al. 2015b). Lower IQ was consistently observed across exposure measures in these studies; however, Choi et al. (2015), a small pilot study (n = 51), did not achieve statistical significance in all results by exposure measure. Specifically, the authors reported a consistent association between all fluoride exposure measures assessed (drinking water, children's urine, and severity of fluorosis) and digit span measures (subtest of the WISC-IV omnibus IQ test); however, results were only statistically significant when fluoride exposure was based on moderate or severe dental fluorosis in children (see Figure A-7). Choi et al. (2015) also observed

some variation in results by outcome assessed (i.e., square root transformed block design and digit span [forward, backward, and total]). It was the only cross-sectional study that did not provide a full IQ score but instead provided results by specific subtests. The study authors consistently observed an inverse association between fluoride exposure and results from the digit span subtest (which specifically assesses executive function); however, results from the block design (square root transformed), a subtest of the WISC-IV omnibus IQ test that specifically assesses visuospatial function, was not associated with fluoride exposure. Note that Rocha-Amador et al. (2009) also assessed visuospatial function, and the authors reported a significant association (p-value <0.001) between fluoride exposure and decreased visuospatial constructional ability using the Rey-Osterrieth Complex Figure (ROCF) Test. Ultimately, too few studies were identified that reported results by subtest of omnibus IQ tests or assessed domains other than IQ (e.g., visuospatial function) to examine or explain the variation by outcome observed in Choi et al. (2015). The only other studies that provided a breakdown of the full IQ score were the prospective cohort studies by Green et al. (2019) and Till et al. (2020), which provided results for full-scale IQ as well as results for performance and verbal IQ. In both of these studies, lower verbal IQ was not associated with fluoride exposure, but lower performance and full-scale IQ were associated with fluoride exposure. There are too few studies to evaluate whether there is a specific aspect of IQ testing that is affected by exposure to fluoride, but the studies nonetheless consistently provide evidence that fluoride exposure is associated with lower IQ.

Yu et al. (2018) reported an overall association between lower IQ and higher fluoride exposure across multiple analyses but observed some variation in IQ results by urinary exposure level. The authors reported inverse associations between IQ and children's medium- and high-range urinary fluoride levels (1.60–2.50 mg/L and 2.50–5.54 mg/L, respectively), although change in IQ score was greater in the medium-range group (2.67 points decrease [95% CI: −4.67, −0.68]) for every 0.5-mg/L increase of urinary fluoride than in the high-range group (0.84 points decrease [95% CI: −2.18, 0.50]) (see Figure A-7). No association was reported at low-range urinary fluoride levels (0.01–1.60 mg/L). Note that Yu et al. (2018) also reported an inverse association between IQ and drinking water fluoride levels at 3.40–3.90 mg/L (4.29-point decrease in IQ score [95% CI: −8.09, −0.48]) for every 0.5-mg/L increase in water fluoride; a 0.04-point decrease in IQ score [95% CI: −0.33, 0.24] was observed for 0.5-mg/L increase in water fluoride at levels of 0.20–3.40 mg/L). The variation by exposure level in urine could not be verified in the analysis of drinking water exposures because there were only two water exposure groups (low and high). In a second study (Wang et al. 2020b), authors conducted a categorical analysis using urinary fluoride quartiles with reported betas per quartile. As observed in Yu et al. (2018), there were decreasing trends in IQ within each quartile; however, unlike Yu et al. (2018), Wang et al. (2020b) observed a larger decrease in IQ with each increasing urinary quartile and observed similar results using water fluoride quartiles (Wang et al. 2020b). Note that Wang et al. (2020b) cannot be compared directly to Yu et al. (2018) for evaluation at the higher exposure levels because the two studies do not use the same categorical exposure ranges. Although additional studies may have looked at different exposure levels, none of these studies provided results in the same manner as Yu et al. (2018) and Wang et al. (2020b) (i.e., betas by exposure category). Instead, these other studies provided an overall beta or mean IQ scores by exposure level. Despite the noted variations among these studies, the overall results still consistently support an association between fluoride exposure and lower IQ.

Two studies (Cui et al. 2018; Zhang et al. 2015b) observed associations between lower IQ in children and exposure to fluoride, with variations in results in subpopulations of children with different polymorphisms (see Figure A-7). These were the only two studies that considered polymorphism as a sub-analysis. Cui et al. (2018) observed a significant association between log-transformed children's single spot urinary fluoride and lower IQ scores (2.47-point decrease in IQ scores [95% CI: −4.93, −0.01; p-value = 0.049] per ln-mg/L increase in urinary fluoride), and the association was strongest in subjects with a TT polymorphism (compared with children with a CC or CT polymorphism) in the dopamine receptor D2 (DRD2) gene (12.31-point decrease in IQ score [95% CI: −18.69, −5.94; p-value <0.001] per ln-mg/L increase in urinary fluoride), which, according to the authors, probably resulted in a reduced D2 receptor density (Cui et al. 2018). Similarly, Zhang et al. (2015b) observed a significant association between lower IQ scores and children's single spot urinary fluoride (2.42-point decrease in IQ scores [95% CI: −4.59, −0.24; p-value = 0.030] per 1-mg/L increase in urinary fluoride), and the association was strongest in subjects with a val/val polymorphism (compared with children who carried the heterozygous or homozygous variant genotypes [met/val or met/met]) in the catechol-O-methyltransferase (COMT) gene (9.67-point decrease in IQ score [95% CI: −16.80, −2.55; p-value = 0.003] per 1-mg/L increase in urinary fluoride).

Overall, the cross-sectional studies consistently support a pattern of findings that higher fluoride exposure is associated with lower IQ scores in children. Slight within-study variations occur that may be associated with study variables such as IQ domains or subsets of IQ tests in a few studies that conducted multiple analyses, but these variations are heterogenous and cannot be further explored with the available studies. Despite these few variations, the overall evidence of an association with lower IQ is apparent.

Exposure Measure and Study Population Factors

Low risk-of-bias studies provide consistent evidence that higher fluoride exposure is associated with lower IQ scores across studies using different exposure measures. In addition to water fluoride levels, studies measured fluoride exposure using single serum samples in children (Xiang et al. 2011; Zhang et al. 2015b), single spot urine samples in children (Cui et al. 2018; Ding et al. 2011; Rocha-Amador et al. 2007; Saxena et al. 2012; Wang et al. 2020b; Xiang et al. 2003a; Yu et al. 2018; Zhang et al. 2015b), and prenatal maternal urinary measures (Bashash et al. 2017; Green et al. 2019), all of which were demonstrated to be consistently associated with lower IQ scores (see Figure A-6, Figure A-7, and Figure A-8). Urine levels encompass all sources of fluoride exposure and provide a better measure of the totality of exposure. As noted previously, even though some studies measured single spot samples, which may not be representative of peak exposure, these studies generally provided evidence that fluoride exposure had been occurring for some time. The consistency in the results across studies that used different measures of fluoride exposure and different life stages at which fluoride was measured strengthens the body of evidence.

The low risk-of-bias studies consistently provide evidence that higher fluoride exposure is associated with lower IQ scores across studies of different study populations. These 19 high-quality studies represent diverse populations (n = 15 study populations) across 5 countries. Eighteen of the 19 studies conducted in Canada (n = 2), China (n = 10), India (n = 3), Iran (n = 1), and Mexico (n = 2) provide evidence that exposure to fluoride is associated with lower IQ scores; 1 study conducted in Mexico did not observe an association but reported results in a

manner that did not allow for a direct comparison with the other studies (see Appendix E for details). The overall consistency in the study results across study populations adds strength to the body of evidence.

Exposure Levels

As described in this section, the body of evidence for studies assessing the association between fluoride exposure and IQ in children consistently provides evidence of an association between higher fluoride exposure [e.g., represented by populations whose total fluoride exposure approximates or exceeds the WHO Guidelines for Drinking-water Quality of 1.5 mg/L of fluoride (WHO 2017)] and lower IQ in children; however, there is less certainty in the evidence of an association in populations with lower fluoride exposures. In the September 6, 2019, draft of this monograph, NTP conducted a qualitative analysis of children's IQ studies that 1) evaluated lower fluoride exposures (<1.5 mg/L) in drinking water and/or urine and 2) provided information to evaluate dose response (i.e., provided three or more fluoride exposure groups or a dose-response curve in their publication) in the lower fluoride exposure range. Nine low risk-of-bias studies met these criteria, which includes the three prospective cohort studies discussed in this section. Based on the qualitative review of these studies, the evidence of an association between fluoride exposure below 1.5 mg/L and lower IQ in children appeared less consistent than results of studies at higher exposure levels.

A draft quantitative dose-response meta-analysis was prepared and included in the September 16, 2020, draft monograph (NTP 2020). This meta-analysis is undergoing further refinement in preparation for separate publication and may further inform a discussion on the association between fluoride exposure levels and IQ in children.

Sex Considerations

Recent literature suggests that adverse neurodevelopmental effects of early-life exposure to fluoride may differ depending on timing of exposure and sex of the exposed subject. In a review of the human and animal literature, Green et al. (2020) concluded that, compared with females, male offspring appear to be more sensitive to prenatal but not postnatal exposure to fluoride, with several potential sex-specific mechanisms.

Sex differences were examined in five of the low risk-of-bias studies (in four study populations) (Green et al. 2019; Trivedi et al. 2012; Wang et al. 2012; Wang et al. 2020b; Xiang et al. 2003a). In general, sex differences were difficult to assess for trends within different study populations because few studies in the body of evidence analyzed exposure and stratified results by sex. Although these five studies reported IQ scores separately for boys and girls, only two of these studies analyzed fluoride exposure for boys and girls separately (Green et al. 2019; Wang et al. 2020b), which is essential for evaluating whether a differential change in IQ by sex may be related to higher susceptibility in one sex or higher exposure in that sex. The remaining three studies stratified results by sex (Trivedi et al. 2012; Wang et al. 2012; Xiang et al. 2003a), but the analyses were based on area-level exposure data (e.g., low-fluoride village compared with high fluoride village) and not drinking water or urinary fluoride concentrations. In the five studies that reported results by sex separately, consistent findings of lower IQ associated with fluoride exposure were generally reported for both sexes. There was some variation in the results between sexes across study populations and exposure measures, but there is insufficient evidence

to determine whether one sex is more susceptible to the effects of fluoride exposure than the other.

Green et al. (2019) observed a significant inverse association between maternal urinary fluoride levels and IQ scores in boys (p-values ≤0.04) but not girls in a Canadian population. Green et al. (2019) did not find any sex differences in the association between IQ and water fluoride concentrations. Wang et al. (2020b) evaluated Chinese boys and girls separately and combined and observed statistically significant decreasing trends in IQ in all groups by urinary fluoride quartiles (p-values for trend ≤0.035) (see Figure A-7). Similarly, when evaluated as a continuous variable, spot urinary fluoride levels (per 1-mg/L increase) were significantly associated with lower IQ scores in girls (−1.379 [95% CI: −2.628, −0.129; p-value = 0.031]), boys (−1.037 [95% CI: −2.040, −0.035; p-value = 0.043]), and in the sexes combined (−1.214 [95% CI: −1.987, −0.442; p-value = 0.002]). According to water fluoride quartiles, Wang et al. (2020b) found that there was a significant trend in the sexes combined, although the decreasing trend in boys and girls separately did not achieve statistical significance (p-values = 0.077 and 0.055, respectively). When water fluoride levels were evaluated as a continuous variable (per 1-mg/L increase), there were significant associations with lower IQ scores in girls (−1.649 [95% CI: −3.201, −0.097]; p-value = 0.037), boys (−1.422 [95% CI: −2.792, −0.053; p-value = 0.042]), and the sexes combined (−1.587 [95% CI: −2.607, −0.568]; p-value = 0.002).

The remaining three studies that reported results by sex-based comparisons of areas of high and low urinary or water fluoride did not report exposure levels separately for boys and girls, which decreases the utility of the data to evaluate differential susceptibility by sex. Trivedi et al. (2012) observed significantly lower IQ in children in high fluoride Indian villages compared with low-fluoride villages with decreases observed in boys and girls separately or combined (p-values ≤0.05) (see Figure A-2). Xiang et al. (2003a) and Wang et al. (2012) provide data on the same study population in China. There was a significantly lower IQ in the high fluoride area compared with the low-fluoride area in boys and girls separately and in the sexes combined (p-values <0.01), although the difference was greater in girls. Because fluoride exposure was not analyzed for boys and girls separately, it is unclear whether the greater change in IQ scores in girls could be attributed to higher susceptibility to fluoride exposure or differences in fluoride exposure by sex.

In summary, it is unclear whether one sex is more susceptible to the effects of fluoride exposure than the other due to the limited number of studies that analyzed exposure and outcome by sex and the lack of a consistent pattern of findings that one sex is more susceptible. Green et al. (2019) did not observe an association between maternal urinary fluoride levels and IQ scores in girls but did observe a significant association in boys. Although this is an indication of higher sensitivity in boys in this analysis, the authors did not detect this sex difference using other measures of prenatal exposure (maternal fluoride intake or water fluoride concentrations). Wang et al. (2020b) and Trivedi et al. (2012) reported statistically significant associations in both boys and girls without indication that one sex may be more susceptible. Although Xiang et al. (2003a) and Wang et al. (2012) reported a greater change in IQ in girls than boys, the studies used area-level exposure data, and the authors did not determine whether fluoride exposure differed in boys versus girls. Therefore, it is unclear whether this differential result by sex is an indication of higher susceptibility in girls or whether it could be explained by a difference in exposure by sex. Overall, there are too few studies that analyzed exposure and outcome by sex separately to properly evaluate whether there is differential susceptibility to fluoride exposure by sex, and

results from the five low risk-of-bias studies that do evaluate sex differences indicate that there is no consistent difference by sex across the different study populations.

Summary of Key Findings for Low Risk-of-bias Children's IQ Studies

In summary, the high-quality studies (i.e., studies with low potential for bias) consistently demonstrate lower IQ scores with higher fluoride exposure [e.g., represented by populations whose total fluoride exposure approximates or exceeds the WHO Guidelines for Drinking-water Quality of 1.5 mg/L of fluoride (WHO 2017)]. The consistency in association is observed among studies of varying study designs, exposure measures, and study populations. Although some studies that conducted multiple analyses observed within-study variations in results (e.g., differences between subsets of IQ tests), these variations were unique to individual studies and did not detract from the overall consistency in the findings that higher fluoride is associated with lower IQ scores.

High Risk-of-bias IQ Studies

The results from 53 studies with high potential for bias that evaluated IQ in children also consistently provide supporting evidence of decrements in IQ associated with exposures to fluoride. Forty-six of the 53 studies reported an association between high fluoride exposure and lower IQ scores in children.

Risk of Bias for IQ Studies in Children

The confidence in the human body of evidence was based on studies with the lowest potential for bias. A total of 19 studies on IQ in children had little or no risk-of-bias concerns, representing a relatively large body of evidence for low risk-of-bias studies (i.e., 15 study populations across 5 countries evaluating more than 7,000 children). These 19 studies are considered low risk of bias because they were rated probably low or definitely low risk of bias for at least two of the three key risk-of-bias questions and did not have any other risk-of-bias concerns that would indicate serious issues with the studies. Thirteen of the 19 studies were rated definitely low or probably low risk of bias for all risk-of-bias questions, and the remaining 6 studies were rated probably high risk of bias for a single question that was judged to have minimal impact on overall potential for bias. None of the 19 studies had a rating of definitely high risk of bias for any question. Risk-of-bias ratings for individual studies for all questions are available in Figure D-1 through Figure D-4, with risk-of-bias ratings for IQ studies in children available in Figure D-5 through Figure D-8 and Appendix E. Although the low risk-of-bias studies had minimal or no concerns, the studies with high overall potential for bias had a number of risk-of-bias concerns, including potential confounding, poor exposure characterization, poor outcome assessment, and, in many cases, potential concern with participant selection. The key risk-of-bias questions are discussed below.

Confounding for IQ Studies in Children

Low Risk-of-bias Studies

As discussed above, there are 19 studies considered to have low risk of bias when assessed across all risk-of-bias domains. Sixteen of the 19 low risk-of-bias studies [i.e., all with the exception of Cui et al. (2020), Ding et al. (2011), and Soto-Barreras et al. (2019)] were considered to have low potential for bias due to confounding because the authors addressed the three key covariates for all studies (i.e., age, sex, and socioeconomic status) through study design

or analysis. Other important covariates, including health factors, smoking, and parental characteristics, were also addressed in many of the low risk-of-bias studies (see Figure 6).

Co-exposures to arsenic and lead were not considered a concern in 18 of 19 low risk-of-bias studies [i.e., all except for Soto-Barreras et al. (2019)] because the studies addressed the potential co-exposures, the co-exposures were not considered an issue in the study population, or the impact of the potential bias on the results was not a concern. Fifteen of 19 low risk-of-bias studies either addressed potential bias related to co-exposure to arsenic through study design or analysis or co-exposure to arsenic was unlikely in the study area. All 15 studies observed an association between lower IQ and fluoride exposure. Co-exposure to arsenic was not accounted for in the remaining four low risk-of-bias studies and was the main potential concern in these studies; however, three of these studies (Wang et al. 2012; Xiang et al. 2003a; Xiang et al. 2011) were still considered low risk of bias for confounding because although arsenic was observed in the water in the low-fluoride (and not the high-fluoride) comparison areas, which would bias the association toward the null, an association was still observed. In this case, the lack of adjustment for arsenic strengthens the evidence for an association and does not represent a potential concern. The other study did not address arsenic co-exposure and, as noted above, was conducted in an area that had potential for arsenic exposure to occur (Soto-Barreras et al. 2019); it is also the only low risk-of-bias study that did not observe an association between lower IQ and fluoride exposure (see Appendix E for further discussion of the risk-of-bias concern regarding arsenic for this study). Although Soto-Barreras et al. (2019) did not discuss arsenic, there is no direct evidence that arsenic was present in the study area. Fourteen studies accounted for co-exposure to lead through study design or analysis, and all observed an association between lower IQ and fluoride exposure. Five studies did not consider co-exposure to lead; however, for all of these studies, co-exposure to lead was considered unlikely to have an impact in these study populations as there was no evidence that lead was prevalent or occurring in relation to fluoride (Cui et al. 2018; Cui et al. 2020; Soto-Barreras et al. 2019; Till et al. 2020; Trivedi et al. 2012).

There is considerable variation in the specific covariates considered across the 19 low risk-of-bias studies. The consistency of results across these studies suggests that confounding is not a concern in this body of evidence. Each of the 18 low risk-of-bias studies that observed an association between fluoride and IQ (see Summary of Results section above) considered a unique combination of covariates. The findings of these studies consistently provide evidence of an association between lower IQ in children and exposure to fluoride regardless of the inclusion or absence of consideration of any one or combination of covariates of interest. For example, maternal or family member smoking was addressed in 7 of the 19 low risk-of-bias studies, and this did not appear to affect the conclusions. All 7 studies that accounted for smoking found evidence of an association between fluoride exposure and lower IQ scores as did 11 of the 12 studies that did not account for smoking. Similarly, all 16 studies that addressed the three key covariates (age, sex, SES) (16 of 16 studies) and two of the three studies that did not fully account for them also found evidence of an association between fluoride exposure and lower IQ scores. In summary, when considering the impact of each covariate (or combinations of covariates) on the consistency of results, no trends are discernable that would suggest that bias due to confounding has impacted or would explain the consistency in findings across the body of evidence that fluoride exposure is associated with lower IQ in children.

Five of the low risk-of-bias studies confirmed the robustness of the results by conducting sensitivity analyses (Bashash et al. 2017; Green et al. 2019; Till et al. 2020; Wang et al. 2020b;

Yu et al. 2018), and none of the sensitivity analyses adjusting for additional covariates found meaningful shifts in the association between fluoride exposure and IQ or other measures of cognitive function. Bashash et al. (2017) found that adjusting for HOME score increased the association between maternal urinary fluoride and children's IQ. Green et al. (2019) reported that adjusting for lead, mercury, manganese, perfluorooctanoic acid, and arsenic concentrations did not substantially alter the associations with IQ. Sensitivity analyses by Yu et al. (2018) that adjusted for covariates (including age, sex, and socioeconomic status) did not find differences in the results compared with the primary analyses. Wang et al. (2020b) found the results of the sensitivity analysis to be the same as the results from the primary analysis. Till et al. (2020) observed that adjusting for maternal urinary fluoride levels, as a way to consider postnatal exposure, had little impact on the results.

Among the 19 low risk-of-bias studies, three were identified that have potential for bias due to confounding (Cui et al. 2020; Ding et al. 2011; Soto-Barreras et al. 2019). This was mainly due to a lack of details on covariates considered key for all studies (i.e., age, sex, and SES). See Appendix E for further discussion of the risk-of-bias concerns regarding confounding for individual studies. Although these three studies have some potential for bias due to confounding, they are considered to be low risk of bias overall because they have low potential for bias for the other two key risk-of-bias questions (exposure characterization and outcome assessment), and no other major concerns for bias were identified. Consistent with the 16 studies that adequately addressed confounding, two of these three studies also provide evidence of an association between fluoride exposure and lower IQ scores in children.

Taken together and considering the consistency in the results despite the variability across studies in which covariates were accounted for, bias due to confounding is not considered to be a concern in the body of evidence. The potential for the consistency in results to be attributable to bias due to confounding in the 19 low risk-of-bias studies is considered low.

Study (Location)[a]	Potential Covariates Considered[b]																Notes	Reported Association with Fluoride[e]
	Subject Characteristics				Other Exposures					Socioeconomic Factors		Parental Characteristics				Other[c]		
	Age	Sex	Race/Ethnicity	Health Factors[c]	Arsenic	Smoking	Iodine	Lead	Other[c]	SES[d]	Caregiving Environment (e.g., HOME score)	Demographics[c]	Reproductive Factors[c]	Health Factors[c]	IQ			
Overall RoB Rating for Confounding: Probably Low																		
Bashash 2017 (Mexico)	√	√	–	–	√	√	–	√	√	√	√	√	√	–	√	√	Other exposures: Hg, Ca Demographics: maternal age Reproductive: parity, birth order, birth weight, gestational age at delivery Other: cohort	Yes
Choi 2015 (China)	√	√	–	√	√	–	–	√	–	√	–	√	√	√	–	√	Health: subject Fe deficiency, illnesses before age 3, medical history of subject and caretakers Demographics: parental age Reproductive: parity Other: residential history	Yes
Cui 2018 (China)	√	√	√	√	√	√	√	–	–	√	–	√	√	√	–	√	Health: subject BMI, stress/anger/anxiety/depression, psychological trauma, having a cold, in relatives: thyroid diseases, cancer, mental retardation Demographics: maternal age Reproductive: abnormal birth Other: alcohol consumption, proximity to factory, physical activity, various dietary factors, environmental noise	Yes
Green 2019 (Canada)	√	√	√	–	√	√	–	√	√	√	√	√	√	–	–	√	Other exposures: Hg, Mn, PFOA Demographics: parental age, pre-pregnancy BMI Reproductive: parity, weeks of gestation, birth weight, maternal chronic condition during pregnancy Other: alcohol consumption, birth country, voiding interval in urine sampling, breastfeeding duration	Yes
Rocha-Amador 2007 (Mexico)	√	√	–	√	√	–	–	√	–	√	–	–	–	–	–	–	Health: subject height and weight by age, transferrin saturation	Yes
Saxena 2012 (India)	√	√	–	√	√	–	√	√	–	√	–	–	–	–	–	√	Health: subject height for age ratio, weight for height ratio, medical history Other: residential history	Yes
Seraj 2012 (Iran)	√	√	–	–	√	–	√	√	–	√	–	–	–	–	–	√	Other: fluorosis intensity	Yes
Sudhir 2009 (India)	√	√	–	–	√	–	–	√	–	√	–	–	–	–	–	√	Other: staple food consumed, liquids routinely consumed, aids used for oral hygiene maintenance (fluoridated or nonfluoridated)	Yes
Till 2020 (Canada)	√	√	√	–	√	√	–	–	–	√	√	–	–	–	–	√	Other: city	Yes
Trivedi 2012 (India)	√	√	–	–	√	–	√	–	–	√	–	–	–	–	–	–		Yes
Wang 2012 (China)	√	√	–	√	–	–	√	√	–	√	–	–	–	√	–	√	Health: medical conditions Other: transportation, natural environment, and lifestyle	Yes
Wang 2020b (China)	√	√	–	√	√	√	√	√	–	√	–	–	√	–	–	√	Health: subject BMI, exclusions based on diseases affecting intelligence, history of trauma or neurological disorders, positive screening test history Reproductive: low birth weight Other: alcohol consumption	Yes
Xiang 2003 (China)	√	√	–	–	–	–	√	√	–	√	–	–	–	–	–	–		Yes
Xiang 2011 (China)	√	√	–	–	–	–	√	√	–	√	–	–	–	–	–	–		Yes
Yu 2018 (China)	√	√	–	√	√	√	√	√	√	√	–	–	√	–	–	√	Health: subject BMI Reproductive: disease history during pregnancy, delivery conditions Other: dental fluorosis prevalence, consanguineous marriage	Yes
Zhang 2015b (China)	√	√	–	√	√	–	√	√	√	√	–	–	–	–	–	√	Health: subject physical and mental health status Other: thyroid hormone levels, residential history, having knowledge of fluorosis, COMT genotype	Yes
Overall RoB Rating for Confounding: Probably High																		
Cui 2020 (China)	–	√	–	√	√	√	√	–	–	√	–	√	√	√	–	√	Health: stress/anger/anxiety, having a cold, in relatives: mental retardation Demographics: maternal age Reproductive: abnormal birth, smoking and drinking during pregnancy Other: thyroid hormone levels	Yes[f]
Ding 2011 (China)	√	–	–	–	√	–	√	√	–	–	–	–	–	–	–	–		Yes
Soto-Barreras 2019 (Mexico)	√	√	–	–	–	–	–	–	–	√	–	–	–	–	–	–		No

Figure 6. Important Covariates Considered in Low Risk-of-bias IQ Studies Conducted in Children

[a]Includes all low risk-of-bias IQ studies in children. Studies are organized as those with an overall risk-of-bias rating for confounding as probably low (green) followed by those with an overall risk-of-bias rating for confounding as probably high (yellow).
[b]Covariates represented here are those considered important for this evaluation. Depending on the specific study population, individual covariates may be considered a potential confounder, effect measure modifier, and/or co-exposure. See study details provided in HAWC for information on additional covariates.
Factors outlined in blue are key covariates for all studies (subject age, subject sex, SES) and arsenic (which is of particular importance to some study populations).

A √ indicates that a covariate was considered. Examples of what it means for a covariate to be "considered": it was adjusted for in the final model, it was considered in the model but not included in the final model because it did not change the effect estimate, it was reported to have the same distribution in both the exposed and unexposed groups, it was reported to not be associated with the exposure or outcome in that specific study population. For arsenic, a √ might also be used when arsenic was not expected to be an issue because there is no evidence to indicate that the co-exposure was prevalent or occurring in relation to fluoride. See risk-of-bias explanations in Appendix E (or HAWC) for details. A hyphen (-) indicates that the factor was not considered.
[c]See the "Notes" column for additional details.
[d]Covariates considered measures of SES include SES scaled scores, household/family income, child education, caretaker/parental education, and occupation/employment.
[e]Extent of reported associations varies by study. "Yes" indicates that study authors provided evidence of an association between lower IQ scores and fluoride exposure.
[f]Study reported lower IQ scores with increasing fluoride exposure, but the results did not achieve statistical significance.

High Risk-of-bias Studies

Most high risk-of-bias studies (n = 53) considered important covariates to some degree through study design or analysis; however, when considering the full scale of potential concerns of bias due to confounding, all but three of these studies were rated probably or definitely high risk of bias. The majority of high risk-of-bias studies accounted for one or two of the three covariates considered key for all studies (age, sex, SES) but did not address all three and did not address other covariates considered important for the specific study population and outcome. Potential confounding related to important co-exposures (e.g., arsenic) was often not addressed in high risk-of-bias studies. In studies in which there was high exposure to fluoride via drinking water with high naturally occurring fluoride or from the use of coal-containing fluoride, most researchers did not account for potential exposures to arsenic, which is commonly found in coal and drinking water in fluoride-endemic areas of China and Mexico.

Despite the lack of adequate consideration of key covariates in the vast majority of high risk-of-bias studies, the results across most of these studies (46 of 53) consistently provide evidence of an association between fluoride exposure and IQ, supporting the results observed in the low risk-of-bias studies. This finding suggests that confounding is likely less of a concern for the body of evidence as a whole than for any individual study. Although the high risk-of-bias studies may have more potential for bias due to confounding compared with the low risk-of-bias studies, the consistent IQ findings across high and low risk-of-bias studies indicate that the results cannot be explained solely by potential bias due to confounding.

Exposure Characterization in IQ Studies

Low Risk-of-bias Studies

In general, there were few, if any, risk-of-bias concerns regarding exposure characterization in the low risk-of-bias studies. These studies mainly had individual exposure data based on urine or water measures with appropriate analyses. Although there are concerns related to using urine samples (see the Risk-of-bias Considerations for Human Studies section for details), the evidence suggests that urinary fluoride is a reasonable measure of exposure (Villa et al. 2010; Watanabe et al. 1995). Using three methods to account for urine dilution, Till et al. (2018) reported that adjusted risk estimates did not differ from unadjusted estimates. Analyzing the same study population as Till et al. (2018), Green et al. (2019) found that adjusting for time of urine collection or time of collection since last void during pregnancy did not substantially affect associations with IQ results in either boys or girls. In addition, adjusting maternal urinary fluoride for creatinine did not substantially alter the observed association (Green et al. 2019). To provide a more accurate and sensitive measurement of maternal urinary fluoride than a single measurement provides, Green et al. (2019) included only participants with valid fluoride

measurements at all trimesters in their analysis. Other studies also measured urinary fluoride multiple times throughout pregnancy (Bashash et al. 2017). Some studies demonstrated correlations between urinary fluoride and fluoride in drinking water, fluorosis, or estimated dose based on drinking water concentrations and consumption (Choi et al. 2015; Ding et al. 2011; Green et al. 2019; Saxena et al. 2012; Yu et al. 2018; Zhang et al. 2015b). Till et al. (2018) demonstrated that there was a linear association between urinary fluoride concentrations in pregnant women and drinking water fluoride concentrations regardless of method used to correct for urine dilution or whether adjustments were made for dilution. Bashash et al. (2017) excluded exposure outliers and found that doing so did not substantively change the results. Taken together, these studies suggest that urinary fluoride is a reasonable measure of exposure despite some potential issues.

All but one low risk-of-bias study was rated probably or definitely low risk of bias for exposure assessment. Seraj et al. (2012) had potential exposure misclassification and was rated probably high risk of bias for exposure assessment. Villages were categorized as normal (0.5–1 ppm), medium (3.1 ± 0.9 ppm), or high (5.2 ± 1.1 ppm) based on average fluoride content in drinking water in varying seasons over a 12-year period. Mild fluorosis observed in children in the normal fluoride level group indicates that there may have been higher exposure in this group at some point in the past; however, this would bias the results toward the null, and the children in the normal fluoride group had a significantly higher IQ score compared with the medium and high fluoride groups (p-value = 0.001). There were also significant associations between lower IQ scores and fluorosis intensity (p-value = 0.014) and water fluoride concentration when evaluated as a continuous variable (p-values <0.001). Although there is potential for exposure bias, the apparent exposure misclassification and inclusion of children with higher fluoride exposure in the normal group indicate that the association may be greater than what was observed in this study.

High Risk-of-bias Studies

A frequent, critical limitation among the high risk-of-bias studies was lack of information regarding exposure or poor exposure characterization. Many of the high risk-of-bias studies compared only subjects living in two regions with differing levels of fluoride exposure, and although most of them did provide some differentiation in levels of fluoride between the areas, limited or no individual exposure information was reported. Among studies that provided drinking water levels of fluoride in two areas being compared, sufficient information to determine whether the individual study subjects were exposed to these levels was often not reported. Some studies also lacked information on fluoride analysis methods and timing of the exposure measurements. In some cases (n = 3), study areas that were considered endemic for dental and/or skeletal fluorosis were compared with non-endemic areas, or high-fluoride areas were compared with low-fluoride areas, with no other information provided on fluoride levels in the areas (Li et al. 2003 [translated in Li et al. 2008c]; Ren et al. 1989 [translated in Ren et al. 2008]; Sun et al. 1991). Although living in an area endemic for fluorosis could be an indicator of exposure, these studies did not specify whether the study subjects themselves had fluorosis. Another study used only dental fluorosis as a measure of fluoride exposure in subjects who were all from an endemic area with similar drinking water fluoride levels (Li et al. 2010). In one case, multiple sources of fluoride exposure were assessed separately without properly controlling for the other sources of exposure, which could bias the results (Broadbent et al. 2015). Broadbent et al. (2015) assessed fluoride exposure in three ways: use of community water in a fluoridated area

versus a non-fluoridated area, use of fluoride toothpaste (never, sometimes, always), or use of fluoride tablets prior to age 5 (ever, never). The same children were used for each analysis without accounting for fluoride exposure through other sources. For example, there were 99 children included in the non-fluoridated area for the community water evaluation, but there is no indication that these 99 children were not some of the 139 children that had ever used supplemental fluoride tablets or the 634 children that had always used fluoride toothpaste. Therefore, comparing fluoridated areas to non-fluoridated areas without accounting for other sources of exposure that might occur in these non-fluoridated areas would bias the results toward the null.

Outcome Assessment for IQ Studies

Low Risk-of-bias Studies

The low risk-of-bias studies have few concerns regarding outcome assessment. All 19 low risk-of-bias studies used appropriate methods for measuring IQ in the study population being assessed, and blinding of outcome assessors was not a concern in 18 of the 19 studies [i.e., all low risk-of-bias studies except Sudhir et al. (2009)]. Fourteen of these 18 studies reported blinding of the outcome assessors, or correspondence with the study authors confirmed that it was not likely an issue. For the remaining 4 of the 18 studies, it was assumed that the outcome assessors were most likely blind because exposure was assessed via urine or drinking water obtained at the same time as the outcome assessment in the general population studies. One IQ study (Sudhir et al. 2009) had concerns for potential bias in the outcome assessment due to lack of information to determine whether blinding at the time of the outcome assessment was a concern (see Appendix E for details).

High Risk-of-bias Studies

Among the studies with high risk of bias, the main limitation in the outcome assessment was the lack of reporting on blinding of the outcome assessor (i.e., whether the outcome was assessed without knowledge of exposure). Although there is little concern that the children's knowledge of their own exposure would bias the way they took the IQ tests, there is potential for bias if the tests were administered by an interviewer, or if the scoring of results could be subjective (e.g., drawing tests), and the interviewer or scorer had knowledge of the children's exposure. Most of the studies did not provide sufficient information on the person scoring or administering the tests or other information on the assessment methods to alleviate concerns for potential interviewer or reviewer bias.

High risk-of-bias studies were mainly carried out in two separate populations without information provided that the tests were conducted in a central location. In many cases, the methods indicated that the tests were conducted at the schools in the study area (indicating that there was likely knowledge of exposure). In some cases, the outcomes were not considered sensitive measures (e.g., Seguin Form Board Test to test for IQ), or the test was not considered appropriate for the study population (e.g., a test validated in a western population was used on a rural Chinese population).

Confidence Assessment of Findings on IQ in Children

We conclude that there is moderate confidence in the body of evidence that higher fluoride exposure is associated with lower IQ in children. This confidence rating was reached by starting

with an initial confidence rating based on key study design features of the body of evidence and then considering factors that may increase or decrease the confidence in that body of evidence. The initial moderate confidence rating is based on 15 of the 19 low risk-of-bias studies that have 3 of the 4 key study design features shown in Figure 1 (i.e., exposure occurred prior to outcome, individual-based outcomes were evaluated, and a comparison group was used). Three of these studies were prospective cohort studies, and 12 were cross-sectional studies that provided evidence of long-term, chronic fluoride exposure prior to outcome measurement.

There are nine factors to consider for increasing or decreasing the confidence in the body of evidence (provided in Figure 1). Discussion of each of these factors in the body of evidence on fluoride exposure and IQ in children is presented below.

- **Risk of bias**: Only studies that were considered to have low risk of bias were included in the moderate confidence rating; therefore, there was no downgrade for risk-of-bias concerns.
- **Unexplained inconsistencies:** The data are consistent, and there was no downgrade for this factor. Eighteen of the 19 low risk-of-bias studies reported associations between higher fluoride levels and lower IQ scores in children. These studies were conducted in 5 different countries on more than 7,000 children from 15 different study populations. There is consistency in results across prospective and cross-sectional study designs. There is also consistency in results across studies using different fluoride exposure measures, including urinary and drinking water fluoride. The one study that did not observe an association did not provide results in a comparable manner and therefore this body of evidence is not considered to have unexplained inconsistencies.
- **Indirectness:** IQ in humans is a direct measure of the association of interest; therefore, no adjustment in confidence is warranted.
- **Imprecision:** There is no evidence of imprecision that would warrant a downgrade. Eighteen studies reported lower IQ with higher fluoride, and no issues with imprecision were identified to challenge the significance of the effect estimate.
- **Publication bias:** There is no strong evidence of publication bias; therefore, no downgrade was applied for publication bias. Two published meta-analyses (Choi et al. 2012; Duan et al. 2018) did not indicate strong evidence of publication bias. The draft meta-analysis conducted by NTP in the September 16, 2020, draft monograph found no publication bias among the low risk-of-bias studies (NTP 2020). Among high risk-of-bias studies, adjusting for publication bias using the trim-and-fill analysis estimated that, in the absence of publication bias, the inverse direction of association and statistical significance remained, thus indicating that there was no need to downgrade for publication bias.
- **Large magnitude of effect size:** Although some individual studies indicated a large magnitude of effect size, the magnitude of effect was not the same across all studies. Therefore, the overall data would not support an upgrade due to a large magnitude of effect size.
- **Dose response:** Evidence of an exposure-response relationship that could justify an upgrade to the confidence in the body of evidence is not presented in this monograph.

While the overall findings qualitatively appear less clear in the lower exposure range, many of the studies that provide data to evaluate exposure response were judged to be high risk of bias. The meta-analysis conducted in association with this systematic review further informs this issue and will be published separately.

- **Residual confounding:** Xiang et al. (2003a), Xiang et al. (2011), and Wang et al. (2012) studied the same population where arsenic occurred in the area with low fluoride but did not occur in the area with high fluoride. This would have biased the results toward the null, but there were significantly lower IQ scores in the area with high fluoride. The remaining studies do not provide enough information to consider whether residual confounding occurred for the body of evidence. Note that parental IQ has the potential to be an important factor when considering residual confounding based on likely correlations between parental IQ and children's IQ; however, there is not sufficient evidence that parental IQ is associated with water fluoride content. Taken together, the overall data would not support an upgrade due to residual confounding.
- **Consistency:** The consideration of a potential upgrade for consistency in the methods is primarily for non-human animal evidence, where it would be applied to address increased confidence for consistent effects across multiple non-human animal species. For human evidence, it is generally not applied, and the data would only be considered in deciding whether to downgrade for unexplained inconsistency. Therefore, no upgrade is applied for consistency.

As described above, there are no changes in confidence rating based on any of the possible upgrade or downgrade factors. The magnitude of effect size and the overall strength and quality of the human literature base provide moderate confidence in the body of evidence that higher exposure to fluoride is associated with lower IQ in children (see the Discussion section for strengths and limitations of the evidence base). Note that additional, well-designed prospective cohort studies with individual-level exposure data and outcome measures could provide increased confidence in the association between fluoride exposure and lower IQ in children.

Other Neurodevelopmental or Cognitive Effects in Children

Low Risk-of-bias Studies

Overview of Studies

Nine low risk-of-bias studies (three prospective cohort and six cross-sectional studies) evaluated the association between fluoride exposure and cognitive neurodevelopmental effects other than IQ in children. These nine studies were conducted in multiple study populations in three countries, specifically:

- three were conducted in three areas of China on three study populations,
- four were conducted in two areas of Mexico on three study populations, and
- two were conducted in Canada using the same study population.

There is considerable heterogeneity across studies, particularly in the different health outcomes evaluated and ages assessed. Most studies measured fluoride in the drinking water or urine (child or maternal) with one study using severity of dental fluorosis as an exposure measure in addition

to drinking water and children's urine. Two of the studies were conducted on infants, with one evaluating effects within 72 hours of birth (Li et al. 2004 [translated in Li et al. 2008a]) and the other evaluating effects at 3 to 15 months of age (Valdez Jimenez et al. 2017). The remaining studies were conducted in children of varying ages, ranging from 4 to 17 years. Other cognitive neurodevelopmental outcomes assessed include neurobehavioral effects in infants, learning and memory impairment, and learning disabilities such as attention deficit hyperactivity disorder (ADHD). Few studies measured the same health outcomes, used the same outcome assessment methods, or evaluated the same age groups.

Table 7 provides a summary of study characteristics and key findings related to other cognitive neurodevelopmental outcomes and fluoride exposure for the nine low risk-of-bias studies. The different tests conducted and the populations on which the tests were conducted are also indicated in Table 7. Several of these studies conducted multiple analyses and reported results on multiple endpoints. The purpose of the table is to summarize key findings (independent of whether an association was found) from each study and is not meant to be a comprehensive summary of all results. For each study, results are summarized for each exposure measure assessed. Results from multiple analyses using the same exposure measure may not all be presented unless conflicting results were reported. See Appendix E for additional information on studies in Table 7, including strengths and limitations, clarifications for why they are considered to pose low risk of bias, and information regarding statistical analyses, covariates, exposure assessment, and outcome assessment.

Table 7. Studies on Other Neurodevelopmental and Cognitive Function in Children[a]

Study	Study Design (Location/Subjects) [n]	Exposure Measures and Summary Statistics	Assessment Timing	Outcome and Methods	Neurobehavioral Outcome Summary[b]
China					
Li et al. (2004) [translated in Li et al. 2008a]	Cross-sectional Zhaozhou County, Heilongjiang Province/neonates [91]	Drinking water Range: 0.5–1.0 mg/L (control); 1.7–6.0 mg/L (high) Maternal urine during pregnancy Mean (SD): 1.74 (0.96) mg/L (control); 3.58 (1.47) mg/L (high)	Neonates (24–72 hours after delivery)	Neurodevelopmental: Neonatal behavioral neurological assessment (NBNA)	Significant differences in neurobehavioral assessment total scores between high-fluoride (36.48 ± 1.09) and control groups (38.28 ± 1.10) (subjects divided into high fluoride group and control group based on drinking water fluoride levels in place of residence); significant differences in total score of behavioral capability that includes measures of non-biological visual orientation reaction and biological visual and auditory orientation reaction between the two groups (11.34 ± 0.56 in controls compared to 10.05 ± 0.94 in high-fluoride group) No statistical adjustment for covariates
Choi et al. (2015)	Cross-sectional Mianning County/1st grade children [51]	Drinking water GM: 2.20 mg/L Children's urine GM: 1.64 mg/L Severity of fluorosis (Dean Index)	Children (ages 6–8 years)	Learning and memory: Neuropsychological tests including WRAML Visual motor ability: WRAVMA Motor ability: Finger tapping task Manual dexterity: Grooved pegboard test	Outcomes unrelated to the IQ test not significantly associated with any fluoride exposure measure Adjusted for age, sex, parity, illness before 3 years old, household income last year, and caretaker's age and education
Wang et al. (2020a)	Cross-sectional Tongxu County/school children [325]	Children's urine Mean (SD): 1.54 (0.89) mg/L	Children (ages 7–13 years)	ADHD and behavior measures: Conners' Parent Rating Scale-Revised (Chinese version) (CPRS-48)	Significant association between psychosomatic problems and urinary fluoride level (per 1-mg/L increase; β = 4.01; 95% CI: 2.74, 5.28; OR for T-score >70 = 1.97; 95% CI: 1.19, 3.27); no associations between urinary fluoride level and ADHD index or other behavioral measures Adjusted for age, sex, child's BMI, urinary creatinine, mother migrated, and father migrated

Study	Study Design (Location/Subjects) [n]	Exposure Measures and Summary Statistics	Assessment Timing	Outcome and Methods	Neurobehavioral Outcome Summary[b]
Mexico					
Rocha-Amador et al. (2009)	Cross-sectional Durango/elementary school children [80]	Children's urine GM (SD): 5.6 (1.7) mg/L	Children (ages 6–11 years)	Visuospatial organization and visual memory: Rey-Osterrieth Complex Figure Test, children's version	Significant correlation between urinary fluoride and visuospatial organization ($r = -0.29$) and visual memory scores ($r = -0.27$); no significant correlation with arsenic Adjusted for age
Valdez Jimenez et al. (2017)	Cohort (Prospective) Durango City and Lagos de Moreno/infants [65]	Maternal urine Range: 0.16–8.2 mg/L (all trimesters) Drinking water Range: 0.5–12.5 mg/L (all trimesters)	Infants (ages 3–15 months)	Mental development index (MDI): Bayley Scales of Infant Development II (BSDI-II) Psychomotor developmental index (PDI): Bayley Scales of Infant Development II (BSDI-II)	Significant association between log10-mg/L maternal urinary fluoride and MDI score during first trimester (adjusted $\beta = -19.05$; SE = 8.9) and second trimester (adjusted $\beta = -19.34$; SE = 7.46); no significant associations between maternal urinary fluoride and PDI score; analyses of outcomes using drinking water fluoride not performed Adjusted for age, gestational age, marginality index, and type of drinking water
Bashash et al. (2017)[c]	Cohort (prospective) Mexico City/Early Life Exposures in Mexico to Environmental Toxicants (ELEMENT) participants [299] GCI analysis [287]	Maternal urine during pregnancy Mean (SD): 0.90 (0.35) mg/L Children's urine Mean (SD): 0.82 (0.38) mg/L	Children (age 4 years)	General cognitive index (GCI): McCarthy Scales of Children's Abilities (MSCA)	Significant association between maternal urinary fluoride and offspring GCI score (per 0.5-mg/L increase adjusted $\beta = -3.15$; 95% CI: −5.42, −0.87); associations with children's urine not significant Adjusted for gestational age; weight at birth; sex; parity (being the first child); age at outcome measurement; and maternal characteristics, including smoking history (ever smoked during the pregnancy vs. nonsmoker), marital status (married vs. not married), age at delivery, IQ, education, and cohort

Study	Study Design (Location/Subjects) [n]	Exposure Measures and Summary Statistics	Assessment Timing	Outcome and Methods	Neurobehavioral Outcome Summary[b]
Bashash et al. (2018)[c]	Cohort (prospective) Mexico City/Early Life Exposures in Mexico to Environmental Toxicants (ELEMENT) participants [210]	Maternal urine during pregnancy Mean 0.85 (95% CI: 0.81, 0.90) mg/L	Children (ages 6–12 years)	ADHD: Conners' Rating Scales-Revised (CRS-R)	Significant associations between maternal urinary fluoride (per 0.5-mg/L increase) and CRS-R scores, including Cognitive Problems + Inattention Index (adjusted β = 2.54; 95% CI: 0.44, 4.63), DSM-IV Inattention Index (adjusted β = 2.84; 95% CI: 0.84, 4.84), DSM-IV ADHD Total Index (adjusted β = 2.38; 95% CI: 0.42, 4.34), and ADHD Index (adjusted β = 2.47; 95% CI: 0.43, 4.50) Adjusted for gestational age; birth weight; sex; parity; age at outcome measurement; and maternal characteristics, including smoking history (ever smoked vs. nonsmoker), marital status (married vs. not married), education, socioeconomic status, and cohort
Canada					
Barberio et al. (2017b)[d]	Cross-sectional General population/Canadian Health Measures Survey (Cycles 2 and 3) [2,221]	Children's urine Mean Cycle 2: 32.06 (95% CI: 29.65, 34.46) µmol/L Mean Cycle 3: 26.17 (95% CI: 22.57, 29.76) µmol/L	Children (ages 3–12 years)	Learning disability, ADHD (Cycle 2 only): Parent or child self-report	Significant increase in adjusted OR for learning disability (adjusted OR = 1.02; 95% CI: 1.00, 1.03) per 1-µmol/L increase in unadjusted urinary fluoride when Cycle 2 and 3 were combined; no significant associations found between urinary fluoride and ADHD (only evaluated in Cycle 2); no significant associations found when using creatinine- or specific gravity-adjusted urinary fluoride Adjusted for age and sex, household income adequacy, and highest attained education in the household

Study	Study Design (Location/Subjects) [n]	Exposure Measures and Summary Statistics	Assessment Timing	Outcome and Methods	Neurobehavioral Outcome Summary[b]
Riddell et al. (2019)[d]	Cross-sectional General population/Canadian Health Measures Survey (Cycles 2 and 3) [3,745]	Drinking water Mean (SD): 0.23 (0.24) mg/L [non-fluoridated water: 0.04 (0.06) mg/L; fluoridated water: 0.49 (0.22)] Community water fluoridation status (yes or no) Children's urine Mean (SD): 0.61 (0.39) mg/L [non-fluoridated water: 0.46 (0.32) mg/L; fluoridated water: 0.82 (0.54)]	Children (ages 6–17 years)	Hyperactivity/inattention: Strengths and Difficulties Questionnaire (SDQ) ADHD: parent or self-reported physician diagnosis	Significantly increased risk of ADHD with fluoride in tap water (adjusted OR = 6.10 per 1-mg/L increase; 95% CI: 1.60, 22.8) or community water fluoridation status (1.21; 95% CI: 1.03, 1.42) but not with urinary fluoride; similar results observed with attention symptoms based on the SDQ scores Adjusted for age and sex, child's BMI, ethnicity, parental education, household income, blood lead, and smoking in the home

ADHD = attention-deficit/hyperactivity disorder; BMI = body mass index; GCI = General Cognitive Index; GM = geometric mean; HOME = Home Observation Measurement of the Environment; IQ = intelligence quotient; MSCA = McCarthy Scales of Children's Abilities; SD = standard deviation; WASI = Wechsler Abbreviated Scale of Intelligence (Spanish version); WISC-IV = Wechsler Intelligence Scale for Children-Revised (as reported by Choi et al. 2015); WRAML = Wide Range Assessment of Memory and Learning; WRAVMA = Wide Range Assessment of Visual Motor Ability.

[a]Includes low risk-of-bias studies.

[b]Associations between other cognitive neurodevelopmental outcomes in children and fluoride levels were reported quantitatively, when possible. For studies with multiple analyses and results, the table summarizes key findings and is not a comprehensive summary of all findings. Results also indicated when a study reported no association, provided as a qualitative statement of no association.

[c]Bashash et al. (2017) and Bashash et al. (2018) are based on the same study population.

[d]Barberio et al. (2017b) and Riddell et al. (2019) are based on the same study population.

Summary of Results

Overall Findings

Although discussed together in this section, various health outcomes were assessed in the nine low risk-of-bias studies of other neurodevelopmental outcomes, including neurobehavioral scores in infants (two studies), cognitive tests in children other than IQ (three studies), and ADHD or learning disabilities (four studies) in children. Altogether, the results from eight of nine low risk-of-bias studies (three prospective cohort studies and five cross-sectional studies from seven different study populations) provide evidence of significant associations between fluoride exposure and cognitive neurodevelopmental outcomes in children other than decrements in IQ (see Figure A-9 through Figure A-11) (Barberio et al. 2017b; Bashash et al. 2017; Bashash et al. 2018; Li et al. 2004 [translated in Li et al. 2008a]; Riddell et al. 2019; Rocha-Amador et al. 2009; Valdez Jimenez et al. 2017; Wang et al. 2020a). Only one cross-sectional study did not find a significant association between fluoride exposure and a measure of cognitive neurodevelopment (Choi et al. 2015).

Although there is heterogeneity in the outcomes assessed and a limited number of directly comparable studies, the data provide additional evidence (beyond the consistent evidence of an association between fluoride exposure and IQ) of an association between higher fluoride exposure and cognitive or neurodevelopmental effects. The body of evidence from the nine low risk-of-bias studies is described in further detail below and is grouped into outcome categories of studies that are most comparable.

Results in Infants

Two studies evaluated neurobehavioral effects in infants either shortly after birth or at 3 to 15 months of age (Li et al. 2004 [translated in Li et al. 2008a]; Valdez Jimenez et al. 2017). Both studies observed a significant association between higher fluoride exposure and lower neurobehavioral scores. In neonates (1–3 days old), the high fluoride group (3.58 ± 1.47 mg/L fluoride based on spot maternal urine collected just prior to birth) had significantly lower total neurobehavioral assessment scores (36.48 ± 1.09 versus 38.28 ± 1.10 in controls; p-value <0.05) and total behavioral capacity scores (10.05 ± 0.94 versus 11.34 ± 0.56 in controls; p-value <0.05) compared to the control group (1.74 ± 0.96 mg/L fluoride) as measured by a standard neonatal behavioral neurological assessment (NBNA) method (Li et al. 2004 [translated in Li et al. 2008a]). In infants 3 to 15 months of age, the Mental Development Index (MDI)—which measures functions including hand-eye coordination, manipulation, understanding of object relations, imitation, and early language development—was significantly inversely associated with maternal urinary fluoride in both the first and second trimesters (adjusted βs per log10-mg/L increase = −19.05 with standard error of 8.9 for first trimester [p-value = 0.04] and −19.34 with standard error of 7.46 for second trimester [p-value = 0.013]) (Valdez Jimenez et al. 2017). Note that this study did not find an association between maternal fluoride during any trimester and the Psychomotor Developmental Index (PDI), which measures gross motor development (adjusted βs = 6.28 and 5.33 for first and second trimesters, respectively; no standard errors provided) (Valdez Jimenez et al. 2017).

Results for Cognitive Tests Other Than IQ in Children

Three studies conducted tests on cognitive function in children that were not part of an IQ test (Bashash et al. 2017; Choi et al. 2015; Rocha-Amador et al. 2009). None of the studies

conducted the same tests, but two of the three studies (Bashash et al. 2017; Rocha-Amador et al. 2009) observed associations between fluoride exposure and lower test scores. The General Cognitive Index (GCI) of the McCarthy Scales of Children's Abilities (MSCA) in 4-year-old children was significantly inversely associated with maternal creatinine-adjusted urinary fluoride levels during pregnancy (collected during each trimester) (adjusted β per 0.5-mg/L increase = −3.15 [95% CI: −5.42, −0.87; p-value = 0.01] in a model adjusting for main covariates including gestational age, weight at birth, sex, maternal smoking, and indicators of socioeconomic status). The association remained even after adjusting for maternal bone lead (adjusted β per 0.5-mg/L increase = −5.63 [95% CI: −8.53, −2.72; p-value <0.01]) (Bashash et al. 2017) (see Figure A-11). Choi et al. (2015), however, evaluated cognitive function endpoints in addition to IQ and found no significant associations between concurrent log-transformed water or urinary fluoride levels and Wide Range Assessment of Visual Motor Ability (WRAVMA) scores, finger tapping test scores, and grooved pegboard test scores, although there were some significant associations based on degree of fluorosis (see Figure A-11). Another study using visuoconstructional and memory scores from the Rey-Osterrieth Complex Figure Test in children 6–11 years old observed significantly lower scores with increasing concurrent child single spot urinary fluoride even after adjusting for age (partial correlation coefficients, per log-mg/L increase = −0.29 and −0.27 for copy [p-value <0.001] and immediate recall [p-value <0.001], respectively [CIs not reported]) (Rocha-Amador et al. 2009). Although these children were also exposed to arsenic, the presence of arsenic could not explain the changes because, in the area with natural contamination by fluoride and arsenic (F–As), the test scores were not significantly associated with urinary arsenic levels (partial correlation coefficients, per log-mg/L increase = −0.05 and 0.02 for copy and immediate recall, respectively [CIs not reported]). The test scores were only marginally increased from fluoride alone when both fluoride and arsenic were included simultaneously in the model (partial correlation coefficients, per log-mg/L increase = −0.32 and −0.34 for copy and immediate recall, respectively [CIs not reported]) (Rocha-Amador et al. 2009) (see Figure A-10).

Attention-related Disorders Including ADHD and Learning Disabilities in Children

Four studies evaluated attention-related disorders or learning disabilities (Barberio et al. 2017b; Bashash et al. 2018; Riddell et al. 2019; Wang et al. 2020a). All four studies found an association between increased fluoride and increased ADHD or learning disability; however, studies varied in the exposure metrics and outcomes measure. Bashash et al. (2018) evaluated behaviors associated with ADHD in children ages 6–12 years using the Conners Rating Scales-Revised (CRS-R) and observed significant associations between maternal urinary fluoride (measured during each trimester) and ADHD-like symptoms, particularly those related to inattention (an increase in 0.5 mg/L of maternal urinary fluoride was significantly associated with a 2.84-point increase [95% CI: 0.84, 4.84; p-value = 0.0054] in the DSM-IV Inattention Index and a 2.54-point increase [95% CI: 0.44, 4.63; p-value = 0.0178] in the Cognitive Problems and Inattention Index). These two scales contributed to the global ADHD Index and the DSM-IV ADHD Total Index, which were also significantly associated with higher levels of prenatal fluoride exposure (an increase of 0.5 mg/L in maternal urinary fluoride was associated with a 2.38-point increase [95% CI: 0.42, 4.34; p-value = 0.0176] in the DSM-IV ADHD Total Index and a 2.47-point increase [95% CI: 0.43, 4.50; p-value = 0.0175] in the ADHD Index) (see Figure A-11). Significant associations were not observed between maternal urinary fluoride concentrations during pregnancy and child performance on measures of hyperactivity, nor were there any significant results in children using Conners' Continuous Performance Test (CPT-II,

2nd Edition), a computerized test of sustained attention and inhibitory control (Bashash et al. 2018). Wang et al. (2020a) also used Conners' Parent Rating Scale (Chinese version) to assess behavioral outcomes in children ages 7–13 years but found only a significant association between spot urinary fluoride concentrations in children (model adjusted for creatinine) and psychosomatic problems (adjusted OR for T-score >70 per 1-mg/L increase = 1.97 [95% CI: 1.19, 3.27; p-value = 0.009] and adjusted β per 1-mg/L increase = 4.01 [95% CI: 2.74, 5.28; p-value <0.001]). No associations were found between spot urinary fluoride and the ADHD index or other behavioral measures.

Barberio et al. (2017b) evaluated learning disabilities in children 3–12 years of age, including ADHD, attention deficit disorder (ADD), and dyslexia, as part of the Canadian Health Measures Survey and found a small but significantly increased risk in self-reported (children 12 years of age) or parent- or guardian-reported (children 3–11 years of age) learning disabilities associated with higher spot urinary fluoride levels in children (adjusted OR per 1-µmol/L increase = 1.02; 95% CI: 1.00, 1.03; p-value <0.05) (see Figure A-12); however, significant associations were not observed in analyses using creatinine- or specific gravity-adjusted urinary fluoride (Barberio et al. 2017b). Barberio et al. (2017b) also reported no associations between single spot urinary fluoride and ADHD in children ages 3 to 12 years. Riddell et al. (2019) used the same Canadian Health Measured Survey but evaluated children 6–17 years old. Riddell et al. (2019) found a significantly increased risk for ADHD diagnosis with both tap water fluoride (adjusted OR per 1-mg/L increase = 6.10; 95% CI: 1.60, 22.8; p-value <0.05) and community water fluoridation status (adjusted OR per 1-mg/L increase = 1.21; 95% CI: 1.03, 1.42; p-value <0.05). A similar increase in the hyperactivity-inattention symptoms score based on the Strengths and Difficulties Questionnaire was observed with both tap water fluoride (adjusted β per 1-mg/L increase = 0.31; 95% CI: 0.04, 0.58; p-value <0.05) and community fluoridation status (adjusted β per 1-mg/L increase = 0.11; 95% CI: 0.02, 0.20; p-value <0.05). As was observed with Barberio et al. (2017b), Riddell et al. (2019) did not observe associations between specific gravity-adjusted spot urinary fluoride concentrations and either ADHD diagnosis (adjusted OR per 1-mg/L increase = 0.96; 95% CI: 0.63, 1.46) or hyperactivity-inattention symptoms (adjusted β per 1-mg/L increase = 0.31; 95% CI: −0.04, 0.66).

Summary of Key Findings for Low Risk-of-bias Studies of Other Neurodevelopmental and Cognitive Effects in Children

In summary, the high-quality studies (i.e., studies with low potential for bias) provide evidence of an association between fluoride exposure and neurodevelopmental and cognitive effects in children other than IQ; however, the body of evidence is limited by heterogeneity in the outcomes evaluated and few directly comparable studies. Across these outcomes, eight of nine studies reported a significant association between fluoride exposure and a measure of neurodevelopment or cognition other than IQ, which provides support for the consistency in evidence based on children's IQ studies of an association between fluoride exposure and adverse effects on cognitive neurodevelopment.

High Risk-of-bias Studies

High risk-of-bias studies (n = 6) also provide some evidence of associations between fluoride exposure and neurodevelopmental or cognitive effects in children other than effects on IQ, but the results are inconsistent and address different outcomes (Jin et al. 2016; Li et al. 1994

[translated in Li et al. 2008b]; Malin and Till 2015; Morgan et al. 1998; Mustafa et al. 2018; Shannon et al. 1986).

Risk of Bias for Neurodevelopmental or Cognitive Effect Studies in Children

The confidence in the human body of evidence was based on studies with the lowest potential for bias (i.e., studies that rated probably low or definitely low risk of bias for at least two of the three key risk-of-bias questions and did not have any other risk-of-bias concerns that would indicate serious issues with the studies). Each of the nine low risk-of-bias studies on other neurodevelopmental effects in children had little or no risk-of-bias concerns. Four of the nine studies were rated definitely low or probably low risk of bias for all risk-of-bias questions, and the remaining five studies were rated probably high risk of bias for a single question that was judged to have minimal impact on overall potential bias. None of the nine studies had a rating of definitely high risk of bias for any question. Although the nine low risk-of-bias studies had minimal or no concerns, the six studies with high overall potential for bias had several risk-of-bias concerns related to one or more of the three key risk-of-bias questions (confounding, exposure characterization, and outcome assessment). The key risk-of-bias questions are discussed below. Risk-of-bias ratings for other neurodevelopmental effect studies in children are available in Figure D-9 through Figure D-12 and Appendix E for the low and high risk-of-bias studies.

Confounding for Other Neurodevelopmental Studies in Children

Low Risk-of-bias Studies

As discussed above, there are nine studies considered to have low risk of bias when assessed across all risk-of-bias domains. Seven of nine low risk-of-bias studies were considered to have low potential for bias due to confounding because the authors addressed the three key covariates for all studies (age, sex, and socioeconomic status) and also addressed arsenic as a potential co-exposure of concern through study design or analysis. Other important covariates, including health factors, smoking, and parental characteristics, were also addressed in many of the low risk-of-bias studies. One of the studies (Bashash et al. 2018) examined several covariates in sensitivity analyses involving subsets of participants, including HOME scores, child contemporaneous fluoride exposure measured by child urinary fluoride adjusted for specific gravity, and maternal lead and mercury exposures. The authors reported that none of the sensitivity analyses indicated appreciable changes in the fluoride-related association with behaviors related to ADHD, nor was there evidence of effect modification between maternal urinary fluoride and sex.

Among the nine low risk-of-bias studies, two studies were identified that have potential for bias due to confounding (Rocha-Amador et al. 2009; Valdez Jimenez et al. 2017). Although both of these studies adjusted for several covariates through analysis or study design, Valdez Jimenez et al. (2017) did not address a potential concern for co-exposure to arsenic, and Rocha-Amador et al. (2009) does not appear to adjust for SES or address why it would not be a concern in the study population (see Appendix E for further details). Although these two studies have some potential for bias due to confounding, they are considered to have low potential for bias overall because they have low potential for bias for the other two key risk-of-bias questions (exposure characterization and outcome assessment), and no other major concerns for bias were identified.

Consistent with the IQ studies, bias due to confounding is not likely a concern for the low risk-of-bias studies.

High Risk-of-bias Studies

The six high risk-of-bias studies in the human body of evidence did not adequately address important covariates through study design or analysis. The same concerns due to potential confounding noted previously for the high risk-of-bias children's IQ studies were also present in the other neurodevelopmental high risk-of-bias studies, including not addressing the three key covariates for all studies (age, sex, SES) and/or not addressing potential co-exposures (e.g., arsenic) in areas of potential concern.

Exposure Characterization in Other Neurodevelopmental Studies in Children

Low Risk-of-bias Studies

There were no risk-of-bias concerns regarding exposure assessment in the low risk-of-bias studies. All of the low risk-of-bias studies had individual exposure data based on urine or water measures with appropriate analyses, and most of the urinary fluoride studies accounted for urinary dilution when appropriate. Although there are concerns related to the timing of urine samples (see the Risk-of-bias Considerations for Human Studies section for details), the studies that used maternal urine measured urinary fluoride multiple times throughout pregnancy (Bashash et al. 2017; Bashash et al. 2018; Valdez Jimenez et al. 2017). Another study demonstrated correlations between urinary fluoride and fluoride in the drinking water, fluorosis, or estimated dose based on water (Choi et al. 2015). Bashash et al. (2017) excluded exposure measurement outliers but found that doing so did not change the results in a meaningful way.

High Risk-of-bias Studies

A frequent critical limitation among the high risk-of-bias studies was lack of information regarding exposure or poor exposure characterization. In the high risk-of-bias studies that assessed the association between fluoride exposure and other neurodevelopmental and cognitive effects in children, fluoride exposure assessment was based on dental fluorosis, municipality-level water fluoridation prevalence data, number of years living in an area with fluorinated water, or group-level water samples. See the Exposure Characterization in IQ Studies section for further discussion on the limitations of exposure assessments in high risk-of-bias studies.

Outcome Assessment in Other Neurodevelopmental Studies in Children

Low Risk-of-bias Studies

The low risk-of-bias studies have few concerns regarding outcome assessment. Seven of the nine studies [i.e., all low risk-of-bias studies except Barberio et al. (2017b) and Riddell et al. (2019)] used appropriate methods for measuring other neurodevelopmental effects in the study population, and blinding of outcome assessors was either reported or not a concern in eight of the nine studies [i.e., all with the exception of Wang et al. (2020a)].

Among the nine low risk-of-bias studies, three were identified that have a potential for bias due to outcome assessment. One of the studies (Wang et al. 2020a) had potential concern for bias due to lack of information regarding the blinding of outcome assessors. Two of the studies (Barberio et al. 2017b; Riddell et al. 2019) were based on the same study population in Canada, where different questions were asked in Cycles 2 (2009–2011) and 3 (2012–2013) of the Canadian

Health Measures Survey (CHMS) to ascertain learning disabilities including ADHD. In Cycle 2, subjects were asked whether they had a learning disability diagnosed by a health professional and, if yes, were asked what kind. In Cycle 3, CHMS did not ask what kind of learning disability was diagnosed nor was a reason for the question omission provided. Because no reason was provided for the removal of the question, and because a question on learning disability without the specific diagnosis may be more prone to bias, this change in questioning from Cycles 2 to 3 is a potential concern. Blinding was not considered an issue in these two studies, but the methods for obtaining the information are considered to be less than ideal for measuring learning disabilities including ADHD. Although the questionnaire asked about a doctor's diagnosis of a learning disability, there was no confirmation with medical records. Moreover, these questionnaires were not validated like Conners' Rating Scales, which would have been a better method for assessing ADHD. Although the outcome assessment methods are less than ideal, there was no direct evidence that they were conducted incorrectly or that the methods would have biased the results in any specific direction. Because this was the only concern in these studies, they were considered to have low risk of bias overall.

High Risk-of-bias Studies

Among the studies on other neurodevelopmental effects with high potential for bias, there were several reasons for studies to be considered probably or definitely high risk of bias for outcome assessment. One study (Shannon et al. 1986) was considered to have probably high risk of bias based on lack of information regarding blinding of outcome assessors. One study was considered definitely high risk of bias because outcome was assessed based on a parent-completed questionnaire, and the study authors noted that the parents were informed of the study's intent and were requested to provide information on fluoride history. Other studies used outcome assessment methods that were not validated or utilized group-level measurements (i.e., school performance, working memory scores).

Confidence Assessment of Findings on Other Neurodevelopmental Effects in Children

The high-quality studies (i.e., studies with low potential for bias) provide evidence of an association between fluoride exposure and other cognitive neurodevelopmental effects, including lower neurobehavioral scores in infants, cognitive effects other than IQ in children, and increased attention-related disorders including ADHD in children. However, due to limitations in the data set, including the heterogeneity in the outcomes assessed, a limited number of directly comparable studies, and differences in outcome assessment methods even when studies evaluated similar outcomes, there is low confidence based on this body of evidence that fluoride exposure is associated with other cognitive neurodevelopmental effects in children. Due to these limitations, the confidence assessment is not described in the same manner as the IQ in Children section or as outlined in Figure 1. Although there are limitations in the body of evidence, the low risk-of-bias studies demonstrate a relationship between higher fluoride exposure and neurodevelopmental effects, even in very young children, which supports the consistency in evidence shown in children's IQ studies of an association between fluoride exposure and adverse effects on cognitive neurodevelopment.

Cognitive Effects in Adults

Low Risk-of-bias Studies

Overview of Studies

Two low risk-of-bias cross-sectional studies evaluated the association between fluoride exposure and cognitive effect in adults (Jacqmin et al. 1994; Li et al. 2016). These two studies used the same test for cognitive function (i.e., Mini-Mental State or MMS Examination) and used drinking water fluoride levels to assess fluoride exposure. Li et al. (2016) also measured urinary fluoride. Both studies were cross-sectional in design. One was conducted in France (Jacqmin et al. 1994) and the other in China (Li et al. 2016). Both studies were conducted in older populations (i.e., over 60 or 65 years of age).

Table 8 provides a summary of study characteristics and key findings related to fluoride exposure and cognitive effects in adults for the two low risk-of-bias studies. The purpose of the table is to summarize key findings (independent of whether an association was found) from each study and is not meant to be a comprehensive summary of all results. For each study, results are summarized for each exposure measure assessed. Results from multiple analyses using the same exposure measure may not all be presented unless conflicting results were reported.

Table 8. Studies on Cognitive Function in Adults[a]

Study	Study Design (Location/Subjects) [n]	Exposure Measures and Summary Statistics	Assessment Timing	Outcome and Methods	Neurobehavioral Outcome Summary[b]
Jacqmin et al. (1994)	Cross-sectional France (Gironde and Dordogne)/elderly adults [3,490]	Drinking water Range: 0.03–2.03 mg	Adults (ages ≥65 years)	Cognitive function: MMS Examination	No significant increase in the prevalence of cognitive impairment with increasing fluoride quartiles No statistical adjustment for covariates for prevalence rates
Li et al. (2016)	Cross-sectional China (Inner Mongolia)/adults [511]	Drinking water daily fluoride intake Mean (SD): 2.23 (2.23) (normal group), 3.62 (6.71) (cognitive impairment group) mg Urine Mean (SD): 1.46 (1.04) (normal group), 2.47 (2.88) (cognitive impairment group) mg/L Fluorosis score Mean (SD): 0.74 (0.98) (normal group), 1.29 (1.01) (cognitive impairment group)	Adults (ages ≥60 years)	Cognitive function: MMS Examination	Subjects with cognitive impairment had a significantly higher skeletal fluorosis score and urinary fluoride concentrations; odds of increasing severity of cognitive impairment increased with urinary fluoride concentrations but were not statistically significant; no significant association with total daily water fluoride intake Adjusted for sex, age, education, marital status (married vs. not married), alcohol consumption (non-drinkers, light drinkers, moderate to heavy drinkers), smoking history (never smoker, ex-smoker, light smoker, heavy smoker), and serum homocysteine levels

GM = geometric mean; MMS = Mini-Mental State.
[a]Includes low risk-of-bias studies.
[b]Associations between cognitive effects in adults and fluoride levels were reported quantitatively, when possible. For studies with multiple analyses and results, the table summarizes key findings and is not a comprehensive summary of all findings. Results also indicate when a study reported no association, provided as a qualitative statement of no association.

Summary of Results

Results from two low risk-of-bias studies in adults did not provide enough evidence to evaluate consistency when assessing evidence for a potential association between fluoride exposure and cognitive impairment (based on the MMS Examination) (Jacqmin et al. 1994; Li et al. 2016). Jacqmin et al. (1994) did not find an association between drinking water fluoride and cognitive impairment in populations in France (n = 3,490) and found prevalence rates of cognitive impairment to be the same regardless of fluoride exposure (see Figure A-13). In contrast, Li et al. (2016) did find significantly higher urinary fluoride levels and skeletal fluorosis scores in the cognitively impaired group compared with the control group in an analysis of 38 cognitively impaired cases and 38 controls matched for several covariates, including age, sex, education, alcohol consumption, and smoking (p-value <0.05). However, the authors found no significant association between cognitive impairment and total daily water fluoride intake (adjusted ORs per 1-mg/day increase = 0.94 [95% CI: 0.85, 1.04] and 0.86 [95% CI: 0.69, 1.06] in the moderate and severe cognitive impairment groups, respectively) or urinary fluoride levels (adjusted ORs per 1-mgL increase = 1.12 [95% CI: 0.89, 1.42] and 1.25 [95% CI: 0.87, 1.81] in the moderate and severe cognitive impairment groups, respectively) in subjects from fluorosis-endemic areas of China (n = 511).

High Risk-of-bias Studies

The results from five out of eight high risk-of-bias studies provide evidence of cognitive impairment in adults associated with exposure to fluoride; however, there was heterogeneity in the outcomes assessed, a limited number of directly comparable studies, and some variability in results (e.g., variation in IQ results across studies). Due to the limited number of low risk-of-bias studies identified that assess cognitive impairment in adults, the results from the high risk-of-bias studies are summarized in greater detail below than had been done in this document for bodies of evidence for IQ in children and other neurodevelopmental and cognitive effects in children.

In aluminum factory workers (exposed to gaseous and particular fluoride emissions during the production of aluminum metal), significant decreases in IQ (Duan et al. 1995), diminished performance on several neurobehavioral core battery tests (NCTBs) (Guo et al. 2001 [translated in Guo et al. 2008b]), and impaired psychomotor performance and memory were observed (Yazdi et al. 2011). One study conducted on adult subjects with fluorosis (dental and skeletal) from a fluorosis-endemic area compared with healthy subjects from a non-endemic area observed significant differences for some cognitive function tests (i.e., tests of speech fluency, recognition, and working memory) but not others and generally did not observe a significant change in IQ except in the operation scores (Shao 2003). One prospective cohort study evaluated exposure to fluoride in children at 5 years of age, based on whether the children resided in areas with community water fluoridation or used fluoride toothpaste or fluoride tablets, and found no clear differences in IQ scores of the subjects at 38 years of age (Broadbent et al. 2015). One additional study suggested that populations living in areas with higher drinking water fluoride had lower levels of dementia (Still and Kelley 1980); however, the study was not focused on effects of fluoride but on whether fluoride was able to reduce the risk associated with aluminum by competing with aluminum and reducing its bioavailability. Therefore, the study was considered inadequate to evaluate the association between fluoride and dementia (Still and Kelley 1980). A more recent study in Scotland evaluated dementia rates associated with aluminum and fluoride drinking water concentrations and observed a significant increased risk of dementia per standard deviation increase in fluoride (p-value <0.001) with the risk of dementia

more than double in the highest quartile of fluoride exposure (56.3 µg/L) compared to the lowest quartile (<44.4 µg/L). The authors also found a significantly increased risk of dementia associated with increased aluminum levels at all quartiles compared with the reference group (p-values <0.05) but found no statistical interaction between aluminum and fluoride levels in relation to dementia (Russ et al. 2019). Conversely, a study in China did not find a significant association between fluoride concentrations in the drinking water and risk for dementia (Liang et al. 2003). In addition to studies that reported on cognitive impairment and exposure to fluoride, two high risk-of-bias studies were identified that reported impaired motor and sensory function (Rotton et al. 1982) and a higher prevalence of self-reported headaches, insomnia, and lethargy (Sharma et al. 2009) associated with fluoride exposure.

Risk of Bias for Cognitive Effect Studies in Adults

Due to the small number of studies with a low potential for bias (see Figure D-13 and Figure D-14), the key risk-of-bias domains (confounding, exposure characterization, outcome assessment) are not discussed separately in respective subsections, as was done for the IQ in Children and Other Neurodevelopmental and Cognitive Effects in Children bodies of evidence. The high risk-of-bias studies had concerns across several domains (see Figure D-15 and Figure D-16), but there were still relatively few studies. Therefore, the discussion for high risk-of-bias studies is also not separated into subsections by key domain.

Low Risk-of-bias Studies

Both low risk-of-bias studies on cognitive effects in adults had little or no risk-of-bias concerns. One study was rated definitely low or probably low risk of bias for all risk-of-bias questions (Li et al. 2016), and the other study was rated probably high risk of bias for a single question that was judged to have minimal impact on overall potential bias (Jacqmin et al. 1994). Jacqmin et al. (1994) had potential concern for bias due to confounding because smoking was not addressed, which has the potential to impact risk for Alzheimer's disease and rates could vary by parish (the target population consisted of men and women from 75 civil parishes in southwestern France).

High Risk-of-bias Studies

There were several issues in the eight studies in adults considered to have high potential for bias. Four of the eight studies had potential concern for bias due to lack of information on the comparison groups, or the comparison groups were considered inappropriate. All eight studies had potential concern for bias regarding covariates not being addressed, including possible co-exposures in occupational studies (e.g., aluminum) and smoking. Five of the eight studies had potential concern for bias due to lack of information regarding exposure characterization or poor exposure characterization with the most utilized exposure measure in these studies being a comparison between exposed and unexposed areas. In one case (Broadbent et al. 2015), multiple sources of fluoride exposure were assessed separately without properly controlling for the other sources of exposure, which could bias the results (see Exposure Characterization in IQ Studies for further details). Five studies also had potential for bias based on limitations in the outcome assessment, which was mainly due to lack of blinding of outcome assessors, lack of validation of the methods, or lack of sufficient details on how the outcomes were assessed.

Confidence Assessment of Findings on Cognitive Effects in Adults

The body of evidence available to examine the association between exposure to fluoride and cognitive effects in adults is limited to two low risk-of-bias cross-sectional studies. Due to the

limited number of studies and a lack of evidence of an effect, there is low confidence based on this body of evidence that fluoride exposure is associated with cognitive effects in adults.

Mechanistic Data in Humans

Eight low risk-of-bias studies that evaluated fluoride exposure and mechanistic data in humans were considered potentially relevant to neurological effects. Effects on the thyroid were specifically evaluated because the NRC 2006 report identified this as a possible effect of fluoride (NRC 2006), and changes in thyroid hormones have been identified as a mechanism for neurodevelopmental effects (Haschek and Rousseaux 1991). These included effects on thyroid hormones in children (Kheradpisheh et al. 2018a; Kheradpisheh et al. 2018b; Malin et al. 2018), adults (Kheradpisheh et al. 2018a; Kheradpisheh et al. 2018b; Malin et al. 2018), or children and adults combined (Barberio et al. 2017a). In addition, some studies evaluated self-reported thyroid conditions in children and adults combined (Barberio et al. 2017a) and thyroid diseases in adults (Kheradpisheh et al. 2018b; Peckham et al. 2015) (see Figure D-17 and Figure D-18). Although the low risk-of-bias studies provide some evidence of mechanistic effects (primarily changes in thyroid stimulating hormone [TSH] levels in children), the studies were too heterogeneous or limited in number to make any determination on mechanism (see Figure 7).

Among the seven low risk-of-bias studies that reported on changes in thyroid hormones, three studies were conducted in children (Kumar et al. 2018; Singh et al. 2014; Zhang et al. 2015b) and reported increases in TSH levels. Zhang et al. (2015b) reported significant increases in TSH in children from a fluorosis-endemic area (median fluoride drinking water concentration = 1.40 mg/L; interquartile range = 1.23–1.57 mg/L) compared with a non-fluorosis-endemic area (median fluoride drinking water concentration = 0.63 mg/L; interquartile range = 0.58–0.68 mg/L), whereas 3,5,3'-triiodothyronine (T_3) or thyroxine (T_4) were not significantly different between the two groups. Similarly, Singh et al. (2014) observed significantly higher TSH levels in children without dental fluorosis who lived in a fluorosis-endemic area (fluoride drinking water concentrations of 1.6–5.5 mg/L) compared with children without dental fluorosis who lived in a non-fluorosis-endemic area (fluoride drinking water concentrations of 0.98–1.00 mg/L). When all children (with and without dental fluorosis) in the endemic area were compared with children from the non-endemic area, the TSH levels were higher in children from the fluorosis-endemic area, although results did not reach statistical significance ($p = 0.057$). Significant differences in T_4 or T_3 were not observed between groups (Singh et al. 2014). Kumar et al. (2018) also observed a significant increase in TSH levels in children from a fluorosis-endemic area (1.5–5.8 mg/L fluoride) compared with a control area (0.94–1.08 mg/L fluoride). There were also decreases in T_3 and T_4, but results were not statistically significant.

Barberio et al. (2017a) evaluated associations between fluoride and TSH levels in children and adults combined and found no relationship between fluoride exposure (measures in urine and tap water) and TSH levels. In the one study that evaluated thyroid hormone levels in adults but not children, Kheradpisheh et al. (2018b) found a significant increase in TSH associated with higher fluoride concentrations in drinking water in both adults with and without thyroid diseases such as hypothyroidism, hyperthyroidism, thyroid nodules, or thyroid cancer. Significant increases in T_3 were associated with higher fluoride in drinking water in adults without thyroid diseases, but increases in T_3 were not significant in adults with thyroid diseases. A significant association

between T_4 and higher fluoride in drinking water was not observed in adults with or without thyroid diseases (Kheradpisheh et al. 2018b).

Other than changes in hormone levels, there is limited evidence of fluoride-related mechanistic effects in the three low risk-of-bias studies that evaluated thyroid-related effects. Barberio et al. (2017a) found no relationship between fluoride exposure and self-reported thyroid conditions in children and adults (children were older than 12). Kheradpisheh et al. (2018b) also found no association between fluoride exposure and hypothyroidism in an adult population in Iran. One study found a significantly higher prevalence of hypothyroidism in areas with higher fluoride concentrations in drinking water (>0.7 mg/L) compared with areas with lower fluoride drinking water concentrations (≤0.7 mg/L) (Peckham et al. 2015).

Sixteen high risk-of-bias studies were available that evaluated mechanistic data in humans associated with fluoride exposure, including effects on thyroid hormones in children (n = 9 studies), thyroid hormones in adults (Michael et al. 1996; Yasmin et al. 2013), catecholamines in adults (Michael et al. 1996) or in subjects of unknown ages (Chinoy and Narayana 1992), acetylcholinesterase (AChE) or serotonin levels in children (Lu et al. 2019; Singh et al. 2013), brain histopathology or biochemistry in aborted fetuses (Du et al. 1992 [translated in Du et al. 2008]; Yu et al. 1996 [translated in Yu et al. 2008]), and mitochondrial fission/fusion molecules in children (Zhao et al. 2019). Similar to the low risk-of-bias studies, the high risk-of-bias studies provide some evidence of mechanistic effects (primarily changes in TSH levels in children); however, the data are insufficient to identify a clear mechanism by which fluoride causes neurodevelopmental or cognitive effects in humans.

Among high risk-of-bias studies (see Figure D-19 and Figure D-20), varying results were reported in 11 studies that evaluated associations between fluoride exposure and thyroid hormones, and a few of these studies (Lin et al. 1991; Wang et al. 2001; Yang et al. 1994 [translated in Yang et al. 2008]) were complicated by high or low iodine in the high fluoride area. When considering fluoride effects on each of the hormones individually, similar to results from low risk-of-bias studies, the most consistent evidence of fluoride-associated effects on a thyroid hormone was reported as changes in TSH levels in children, although there was some variation in the direction of association. Six of the nine high risk-of-bias studies that evaluated changes in TSH levels in children reported increases in TSH levels with higher fluoride (Lin et al. 1991; Susheela et al. 2005; Wang et al. 2001; Yang et al. 1994 [translated in Yang et al. 2008]; Yao et al. 1996; Yasmin et al. 2013). Two of the nine high risk-of-bias studies reported decreases in TSH levels in children with higher fluoride (Khandare et al. 2017; Khandare et al. 2018). One of the nine studies found no significant alterations in TSH levels in children from fluorosis-endemic areas (Hosur et al. 2012) (see Figure 8).

When considering associations between fluoride and TSH, T_3, and T_4 levels together, studies that evaluated changes in all three thyroid hormones reported varying combinations of increases, decreases, or no changes in levels across the three hormones, although among the eight low and high risk-of-bias studies that evaluated associations between fluoride exposure and TSH, T_3, and T_4 levels and reported increases in TSH levels in children, seven of the eight studies found no alterations in T_3 levels (one study found an increase in T_3), and six of the eight studies found no alterations in T_4 levels (two studies found an increase in T_4). Studies also displayed variation by age in the associations between fluoride and TSH, T_3, and T_4. Due to the dynamic relationship between the thyroid gland, the pituitary gland, and the production and clearance of TSH, T_3, and

T_4, the variations in results are not unexpected and do not eliminate the possibility of a mechanistic link between thyroid effects and neurodevelopmental or cognitive effects; however, the data do not support a clear indication that thyroid effects are a mechanism by which fluoride causes these effects in humans.

Endpoint	Direction of Effect ↑	Direction of Effect NS	Grand Total
serum T3	1	3	4
serum T4		4	4
serum TSH	5	2	7
self-reported thyroid condition		1	1
thyroid disease		1	1
hypothyroidism prevalence	1		1

Figure 7. Number of Low Risk-of-bias Studies that Evaluated Thyroid Hormones in Children and Adults by Endpoint and Direction of Association

Interactive figure and additional study details in Tableau® (https://public.tableau.com/app/profile/ntp.visuals/viz/Fluoride_EpiThyroid_UPDATE/Figures7and8?publish=yes). This figure displays study counts for low risk-of-bias studies in both children and adults, as these counts are most relevant to the summary of fluoride-related mechanistic effects in low risk-of-bias studies. Counts for high risk-of bias studies and studies by age (i.e., children, adults, or children/adults combined) can also be accessed in the interactive figure in Tableau®. Study counts are tabulated by significance (unless study footnotes in Tableau indicate that statistical significance was not tested)—statistically significant increase (↑), statistically significant decrease (↓), or not significant (NS). For example, the "↑" column displays numbers of unique studies with significantly increased results.

Endpoint	Direction of Effect ↑	Direction of Effect ↓	Direction of Effect NS	Grand Total
serum T3	1	1	6	8
serum T4	3		5	8
serum TSH	6	2	1	9

Figure 8. Number of High Risk-of-bias Studies that Evaluated Thyroid Hormones in Children by Endpoint and Direction of Association

Interactive figure and additional study details in Tableau® (https://public.tableau.com/app/profile/ntp.visuals/viz/Fluoride_EpiThyroid_UPDATE/Figures7and8.). This figure displays study counts for high risk-of-bias studies in children, as these counts are most relevant to the summary of associations between fluoride and thyroid hormones in high risk-of-bias studies. Counts for low risk-of bias studies, studies in adults, or all studies combined, can also be accessed in the interactive figure in Tableau®. Study counts are tabulated by significance (unless study footnotes in Tableau indicate that statistical significance was not tested)—statistically significant increase (↑), statistically significant decrease (↓), or not significant (NS). For example, the "↑" column displays numbers of unique studies with significantly increased results.

In addition to evaluating thyroid hormone levels, a few high risk-of-bias studies evaluated other mechanistic data associated with fluoride exposure; however, the data are insufficient to identify a clear mechanism by which fluoride might cause neurodevelopmental or cognitive effects in humans. Serum epinephrine and norepinephrine were significantly increased in a fluoride-endemic region (it was not reported whether subjects were children or adults) compared with a non-endemic region (Chinoy and Narayana 1992). Serum adrenaline and noradrenaline were

significantly increased in adults in a fluoride-endemic area (fluoride in the drinking water ranged from 1.0–6.53 ppm) compared with a control area (fluoride in the drinking water ranged from 0.56–0.72 ppm) (Michael et al. 1996). Serum AChE was significantly reduced in children from a high fluoride region compared with a lower fluoride region (Singh et al. 2013). Serum serotonin was significantly increased in children from Turkey who were drinking water containing 2.5 mg/L of fluoride compared with children drinking bottled water or water containing <0.5 mg/L of fluoride (Lu et al. 2019). Aborted fetuses from high fluoride areas in China were found to have histological changes in the brain and significant changes in neurotransmitter levels compared with a control area (Du et al. 1992 [translated in Du et al. 2008]; Yu et al. 1996 [translated in Yu et al. 2008]).

There are also two more recent low risk-of-bias studies that evaluated polymorphisms in dopamine-related genes; however, a determination on mechanism cannot be made at this time due to the limited number of studies. For children (10–12 years old) with a Val158Met polymorphism in the *COMT* gene (i.e., catechol-O-methyltransferase), which results in slower degradation and greater availability of dopamine within the brain, a stronger association between increasing urinary fluoride levels and decreasing IQ was reported (Zhang et al. 2015b). For children (7–12 years old) with a dopamine receptor-2 (DRD2) Taq 1A polymorphism (which is involved in reduced D2 receptor density and availability) and the TT (variant) genotype, a significant inverse association between log urinary fluoride and IQ was observed; however, this significant relationship was not observed in children with the CC (wild-type) or CT (hybrid) genotypes (Cui et al. 2018).

Animal Learning and Memory Data

NTP provided a review of the experimental animal evidence in the earlier draft monographs (NTP 2020) and agrees with the NASEM committee's comments (NASEM 2020; 2021) (placeholder to cite NTP 2021 Response to NASEM comments) that the experimental animal database is of poor quality, with many studies suffering from major reporting deficiencies. NTP acknowledges that further efforts to disentangle the potential for motor activity deficits to influence tests of learning and memory in the fluoride literature are warranted. Overall, these general issues and deficiencies with the experimental animal database led to NTP's conclusion that the animal studies are currently *inadequate* to inform the question of an association between fluoride exposure and neurodevelopmental and cognitive effects in humans. Therefore, this systematic review does not include an experimental animal section.

Mechanistic Data in Animals

There are a wide variety of studies in animals that evaluate mechanistic effects potentially related to neurological changes following oral fluoride exposure (see Appendix F); however, the mechanisms underlying fluoride-associated cognitive neurodevelopmental effects are not well characterized, and review of the data did not identify a mode of action for fluoride effects on IQ in children. Categories of mechanistic endpoints with the largest amount of available data include changes in biochemical components of the brain or neurons, neurotransmitters, oxidative stress, histopathology, and thyroid function. Limiting the data to studies with at least one exposure at or below 20 ppm fluoride drinking water equivalents (gavage and dietary exposures were backcalculated into equivalent drinking water concentrations for comparison) still provided a sufficient number of studies for evaluation of these mechanistic endpoints. This evaluation is

provided in Appendix F. Neurotransmitter and biochemical changes in the brain and neurons were considered the mechanistic areas with the greatest potential to demonstrate effects of fluoride on the brain of animals in the lower dose range and provide evidence of changes in the brain that may relate to lower IQ in children (see Appendix F). Histological data can be useful in determining whether effects are occurring in the brain at lower fluoride concentrations; however, author descriptions of these effects may be limited, thereby making it difficult to directly link histological changes in the brain to learning and memory effects. Oxidative stress is considered a general mechanistic endpoint that cannot be specifically linked to neurodevelopmental or cognitive effects in humans; however, like histopathology, it may help in identifying changes in the brain occurring at lower concentrations of fluoride. Although any effects in the brain or neurological tissue at lower concentrations of fluoride may support reduced IQ in humans, it may be difficult to distinguish the potential effects of fluoride on learning and memory functions from other neurological or general health outcomes.

In Vitro Data on Neurodevelopmental or Cognitive Effects

Although in vitro studies were identified as part of the systematic review process, NTP determined that the information on neurological effects from these studies is too general, and results cannot necessarily be attributed to effects on learning and memory or other cognitive functions at this time. The in vitro data may help support specific mechanisms identified from in vivo mechanistic data; however, as described above, no specific mechanism has been determined for fluoride effects on learning and memory or other neurodevelopmental or cognitive outcomes.

Discussion

This systematic review evaluated the available animal and human literature concerning the association between fluoride exposure and cognitive neurodevelopment. The available data on potential mechanisms to evaluate biological plausibility were also assessed. The potential health benefits of fluoride with respect to oral health are acknowledged but are not the focus of this review.

This review extended NTP's previous evaluation of the experimental animal data (NTP 2016). Although the animal data provide some evidence of effects of fluoride on neurodevelopment, they give little insight into the question of whether fluoride influences IQ. This is due to deficiencies identified in the animal body of evidence. Mechanistic studies in humans provide some evidence of adverse neurological effects of fluoride. However, these studies were too heterogenous and limited in number to make any determination on biological plausibility.

The literature on adults is also limited; therefore, it was determined that there is low confidence in the body of evidence from studies that evaluate fluoride exposure and adult cognition. Compared to the literature in adults, there is a much more extensive literature in children.

The literature in children was separated into studies assessing IQ and studies assessing other cognitive or neurodevelopmental outcomes. There is low confidence in the body of evidence from studies that evaluate fluoride exposure and other cognitive or neurodevelopmental outcomes in children. Altogether, the results from eight of nine high-quality studies (three prospective cohort and five cross-sectional studies from seven different study populations) provide some evidence that fluoride is associated with other cognitive or neurodevelopmental outcomes in children. The data also suggest that neurodevelopmental effects occur in very young children. However, the number of studies is limited, and there is too much heterogeneity in the outcomes measured and methods used to directly compare studies of any one outcome. Additional studies on outcomes such as attention-deficit hyperactivity disorder (ADHD) and other attention-related disorders, where there is some evidence of an effect of fluoride exposure, would be necessary to critically assess the data.

Most of the epidemiological studies (n = 72) assessed the association between fluoride exposure and IQ in children. Although all studies, both high- and low-quality, were considered, this evaluation focuses on the high-quality, low risk-of-bias studies in children for two reasons. First, there are fewer limitations and greater confidence in the results of the high-quality studies. Second, there is a relatively large number of high-quality studies (n = 19), such that the body of evidence from these studies could be used to evaluate confidence in the association between fluoride exposure and changes in children's IQ.

This review finds, with moderate confidence, that fluoride exposure is associated with lower IQ in children. The association between higher fluoride exposure and lower IQ in children was consistent across different study populations, study locations, study quality/risk-of-bias determinations, study designs, exposure measures, and types of exposure data (group-level and individual-level). There were 19 low risk-of-bias studies that were conducted in 15 study populations, across 5 countries, and evaluating more than 7,000 children. Of these 19 studies, 18 reported an association between higher fluoride exposure [e.g., represented by populations whose total fluoride exposure approximated or exceeded the WHO Guidelines for Drinking-water

Quality of 1.5 mg/L of fluoride (WHO 2017)] and lower IQ. These include 3 prospective cohort studies and 15 cross-sectional studies (12 of which indicated that exposure likely preceded the outcome). Forty-six of 53 low-quality studies in children also reported an association between higher fluoride exposure and lower IQ.

Many studies in this assessment relied on drinking-water fluoride levels (both group-level measures and individual-level measures), rather than measures of total fluoride exposure, to establish exposed versus "unexposed" or reference groups. Although fluoride in water is a major source of exposure [comprising 40% to 70% of total exposure (US EPA 2010)], other sources of fluoride provide variable amounts that depend on personal preferences and habits. The use of dental products containing fluoride and consuming foods and beverages prepared with fluoridated water can also result in measurable exposures (US EPA 2010). Green et al. (2019) suggested that significant exposures occur from black tea consumption. Thus, drinking water fluoride levels may, but usually do not, reflect total fluoride exposure. This could be a potential limitation in studies that rely on water fluoride data to assess fluoride exposure (in particular, earlier studies). However, because water is only part of a person's total exposure to fluoride, this limitation would likely result in an underestimate of exposure to fluoride. In addition, this limitation is less of a concern in areas where fluoride in the drinking water is high because drinking water likely contributes a large proportion of the total fluoride intake in those areas as compared with areas where fluoride in the drinking water is lower.

This review found that the quality of exposure assessment has improved over the years. More recent studies by Valdez Jimenez et al. (2017), Bashash et al. (2017), and Green et al. (2019) used individual measures of urinary fluoride, either maternal urine collected prenatally or children's urine, which confirmed the association between higher total fluoride exposure and lower children's IQ and other cognitive neurodevelopmental effects. Studies using different types of exposure measures reported similar findings of an association, which strengthens confidence in earlier studies that reported IQ deficits with increasing group-level fluoride exposure. However, there is less certainty in the quantitative estimates of the magnitude of IQ deficits from earlier studies that used group-level exposure measures than the estimates from more recent studies that used individual-level exposure measures.

It is worth noting that there are circumstances wherein typical children's water consumption considered with water fluoride levels may substantially underestimate total fluoride exposure. One example is bottle-fed infants wherein nutrition is provided by powdered formula that is rehydrated with fluoridated water (Till et al. 2020). To decrease an exclusively formula-fed infant's exposure to fluoride, for the purpose of reducing risk of dental fluorosis, the Centers for Disease Control and Prevention recommends using low-fluoride bottled water to mix with infant formula (CDC 2015). A few studies also support the possibility of heightened sensitivities to the detrimental cognitive effects of fluoride exposure in individuals with certain genetic polymorphisms in dopamine receptor D2 or catechol-O-methyltransferase (Cui et al. 2018; Zhang et al. 2015b), potentially impacting dopamine catabolism and receptor sensitivity. Differential exposures to fluoride and genetic susceptibilities of children to fluoride may represent special situations that would appear to warrant further research.

The following section briefly recaps the strength of the epidemiological evidence for an association between fluoride exposure and cognitive neurodevelopmental deficits. This is followed by a more detailed listing of limitations of the evidence base and limitations of the

systematic review, with some suggestions of areas where further research may be most beneficial.

Strengths of the Evidence Base

Strengths in the epidemiological evidence base include:

- There are 72 studies directly addressing the relationship between fluoride exposure and children's IQ.
- There are 12 high-quality cross-sectional studies with low risk of bias providing evidence that exposure occurred prior to outcome assessment in those studies.
- Studies are from diverse geographic locations that included data for more than 7,000 children.
- There are 19 high-quality studies evaluating the same outcome (i.e., IQ) and 9 evaluating other neurodevelopmental outcomes.
- Reported responses to fluoride exposure are consistent in studies of both low and high quality.
- Reported responses to fluoride exposure are consistent across different study populations, study designs, and exposure measures.
- Findings of studies with group- and individual-level information on exposure and outcomes are similar.
- A wide variety of important covariates are either addressed by study design or captured across the evidence base, with no consistent patterns that would suggest an alternative explanation.

Limitations of the Evidence Base

Limitations in the epidemiological studies with low risk of bias include:

- Few studies are available that assessed the association between fluoride exposure and cognitive function (particularly IQ) in adults and attention-related disorders including ADHD in children and adults.
- Heterogeneity in outcomes was assessed for other neurobehavioral outcomes, limiting the assessment of other possible effects in children.
- Studies rarely separated the results by sex or provided information to indicate that sex was not a modifying factor.
- Associations between lower total fluoride exposure [e.g., represented by populations whose total fluoride exposure was lower than the WHO Guidelines for Drinking-water Quality of 1.5 mg/L of fluoride (WHO 2017)] and children's IQ remain unclear. More studies at lower exposure levels are needed to fully understand potential associations in ranges typically found in the United States (i.e., <1.5 mg/L in water). However, it should be noted that, as of April 2020, CWS supplying water with ≥1.5 mg/L naturally occurring fluoride served 0.59% of the U.S. population (~1.9 million people) (CDC Division of Oral Health 2020).

- No studies investigating the association between fluoride exposure and neurodevelopmental or cognitive effects in adults or children have been conducted in the United States.
- No studies are available to evaluate fluoride exposure over a child's lifetime and neurodevelopmental or cognitive changes over time.
- The database does not allow for comparison of ages and possible changes at different developmental stages in children to assess if there is a delay in development or if associations persist.
- The database does not allow for establishing clear correlations between prenatal and postnatal exposures.

Limitations in the epidemiological studies with high risk of bias include:

- Many of the original publications were in a non-English language and provided limited details on methodology.
- Studies lacked information regarding exposure and/or had serious limitations in the exposure assessment. Exposure assessment concerns include limited individual exposure information, a lack of information on fluoride sampling methods and timing of the exposure measurements, a lack of quantitation of levels of fluoride in drinking water in a few studies, and a lack of individual-level information on fluorosis in areas reported to be endemic for fluorosis.
- The comparison groups in studies conducted in areas endemic for fluorosis still may have been exposed to high levels of fluoride or levels similar to those used in water fluoridation in the United States. This factor may have limited the ability to detect true effects.
- Studies did not provide sufficient direct information (e.g., participation rates or methods for selection) to evaluate selection bias.
- Failure to address important covariates was an issue for many studies. Some studies conducted simple statistical analyses without accounting for any covariates in the analysis, although many noted similarities between the study populations. In cases where adjustments in analyses were made, often these studies did not account for covariates considered critical for that study population and outcome including co-exposures.
- Studies conducted in areas with high, naturally occurring fluoride levels in drinking water often did not account for potential exposures to arsenic or iodine deficiencies in study subjects in areas where these substances were likely to occur.
- Studies lacked information on whether the outcome assessors were blind to the exposure group, including studies that examined children in their schools and subjects from high-fluoride communities.

Limitations in the animal and mechanistic evidence base include:

- The overall quality of the experimental animal studies is poor, and there are relatively few well-designed and well-performed studies at lower fluoride exposure levels (i.e., <20 ppm, which is roughly equivalent to human exposure of <4 ppm).

- The understanding of the specific molecular events responsible for fluoride's adverse effects on neurobehavioral function is poor.

A key data gap in the human and animal bodies of evidence includes the need for mechanistic insight into fluoride-related neurodevelopmental or cognitive changes.

Limitations of the Systematic Review

This systematic review has few limitations. The human body of evidence included a large database of observational studies. Most of the observational studies were cross-sectional; however, 12 of these were considered to provide sufficient evidence that exposure occurred prior to the outcome. In addition, the systematic review covered a wide range of study designs, populations, and measures of fluoride exposure. The systematic review was designed to cover reports on all potential mechanistic data including effects on the thyroid. After review of the studies evaluating thyroid effects, studies that only evaluated goiters and other effects on thyroid size were not considered in this review. This is not considered a limitation because these studies did not include specific information on thyroid hormones that could indicate a mechanism for thyroid involvement in neurodevelopment. In addition, review of the mechanistic data was limited to in vivo studies with at least one concentration below 20 ppm. This is not considered a limitation for the systematic review because the mechanistic body of evidence was used to evaluate biological plausibility for the effects observed in humans; therefore, data were limited to concentrations that would be more reflective of human exposures. The decision to not more closely evaluate the in vitro data is not considered a limitation because there were sufficient in vivo data, and no key events were identified where in vitro data would provide additional insight.

The supplemental literature search for non-English-language studies not indexed in traditional databases supports the comprehensive nature of the literature search strategy for this systematic review. In the absence of guidance on the most complete non-English-language databases that may contain health studies of fluoride, databases were selected that identified non-English-language studies of fluoride that we were aware of and were not captured in searches of databases from the main literature search. This informed approach influenced the selection process; however, this is not considered a limitation because it provided an objective measure by which to compare databases. Following the recommendation of the NASEM committee in its review of the September 16, 2020, draft monograph, the experimental animal section has been removed and is not included in this monograph. Although the deficiencies identified in the animal body of evidence support this removal (see Animal Learning and Memory Data for further explanation), NTP acknowledges that the absence of the experimental animal data is a limitation of this systematic review. For the purpose of this review, NTP considers the experimental animal data to be *inadequate* to inform whether fluoride exposure is associated with cognitive effects (including cognitive neurodevelopmental effects) in humans.

Summary

This systematic review evaluated the available animal and human literature concerning the association between fluoride exposure and cognitive neurodevelopment. The available data on potential mechanisms to evaluate biological plausibility were also assessed. Existing animal studies provide little insight into the question of whether fluoride exposure affects IQ. Human mechanistic studies were too heterogenous and limited in number to make any determination on biological plausibility. The body of evidence from studies on adults is also limited and provides low confidence that fluoride exposure is associated with adverse effects on adult cognition. There is, however, a large body of evidence on IQ effects in children. There is also some evidence that fluoride exposure is associated with other neurodevelopmental and cognitive effects; although, because of the heterogeneity of the outcomes, there is low confidence in the literature for these other effects. This review finds, with moderate confidence, that higher fluoride exposure [e.g., represented by populations whose total fluoride exposure approximates or exceeds the WHO Guidelines for Drinking-water Quality of 1.5 mg/L of fluoride (WHO 2017)] is consistently associated with lower IQ in children. More studies are needed to fully understand the potential for lower fluoride exposure to affect children's IQ.

References

Adedara IA, Abolaji AO, Idris UF, Olabiyi BF, Onibiyo EM, Ojuade TD, Farombi EO. 2017a. Neuroprotective influence of taurine on fluoride-induced biochemical and behavioral deficits in rats. Chem Biol Interact. 261:1-10.

Adedara IA, Olabiyi BF, Ojuade TD, Idris UF, Onibiyo EM, Farombi EO. 2017b. Taurine reverses sodium fluoride-mediated increase in inflammation, caspase-3 activity, and oxidative damage along the brain-pituitary-gonadal axis in male rats. Can J Phys Pharmacol. 95:1019-1029. https://doi.org/10.1139/cjpp-2016-0641

Akinrinade ID, Memudu AE, Ogundele OM. 2015a. Fluoride and aluminium disturb neuronal morphology, transport functions, cholinesterase, lysosomal and cell cycle activities. Pathophysiology. 22:105-115.

Akinrinade ID, Memudu AE, Ogundele OM, Ajetunmobi OI. 2015b. Interplay of glia activation and oxidative stress formation in fluoride and aluminium exposure. Pathophysiology. 22:39-48.

An J, Mei S, Liu A, Fu Y, Wang C. 1992. [Effect of high level of fluoride on children's intelligence]. Chin J Control Endem Dis. 7(2):93-94.

Arbuckle TE, Fraser WD, Fisher M, Davis K, Liang CL, Lupien N, Bastien S, Velez MP, von Dadelszen P, Hemmings DG et al. 2013. Cohort profile: The maternal-infant research on environmental chemicals research platform. Paediatr Perinat Epidemiol. 27(4):415-425. https://doi.org/10.1111/ppe.12061

Baba NA, Raina R, Verma PK, Sultana M. 2014. Alterations in plasma and tissue acetylcholinesterase activity following repeated oral exposure of chlorpyrifos alone and in conjunction with fluoride in Wistar rats. Proc Indian Natl Sci Acad B Biol Sci. 84:969-972.

Banala R, Nagapuri K, Mohd K, Reddy M, Karnati P. 2018. Carica Papaya leaf extract as a neuroprotective agent against behavioral and neurotransmitter changes in brain of the rat treated with sodium fluoride in pre- and post-natal periods. Pharmacogn Mag. 14(55):123-131. https://doi.org/10.4103/pm.pm_378_17

Barberio AM, Hosein FS, Quinonez C, McLaren L. 2017a. Fluoride exposure and indicators of thyroid functioning in the Canadian population: Implications for community water fluoridation. J Epidemiol Community Health. 71:1019-1025. https://doi.org/10.1136/jech-2017-209129

Barberio AM, Quinonez C, Hosein FS, McLaren L. 2017b. Fluoride exposure and reported learning disability diagnosis among Canadian children: Implications for community water fluoridation. Can J Public Health. 108:229-239. https://doi.org/10.17269/cjph.108.5951

Bartos M, Gumilar F, Gallegos CE, Bras C, Dominguez S, Monaco N, Esandi MDC, Bouzat C, Cancela LM, Minetti A. 2018. Alterations in the memory of rat offspring exposed to low levels of fluoride during gestation and lactation: Involvement of the alpha7 nicotinic receptor and oxidative stress. Reprod Toxicol. 81:108-114. https://doi.org/10.1016/j.reprotox.2018.07.078

Bashash M, Thomas D, Hu H, Martinez-Mier EA, Sanchez BN, Basu N, Peterson KE, Ettinger AS, Wright R, Zhang Z et al. 2017. Prenatal fluoride exposure and cognitive outcomes in

children at 4 and 6-12 years of age in Mexico. Environ Health Perspect. 125(9):1-12. https://doi.org/10.1289/ehp655

Bashash M, Marchand M, Hu H, Till C, Martinez-Mier EA, Sanchez BN, Basu N, Peterson KE, Green R, Schnaas L et al. 2018. Prenatal fluoride exposure and attention deficit hyperactivity disorder (ADHD) symptoms in children at 6-12 years of age in Mexico City. Environ Int. 121(Pt 1):658-666. https://doi.org/10.1016/j.envint.2018.09.017

Bhatnagar M, Rao P, Sushma J, Bhatnagar R. 2002. Neurotoxicity of fluoride: Neurodegeneration in hippocampus of female mice. Indian J Exp Biol. 40:546-554.

Bhatnagar M, Sukhwal P, Suhalka P, Jain A, Joshi C, Sharma D. 2011. Effects of fluoride in drinking water on NADPH-diaphorase neurons in the forebrain of mice: A possible mechanism of fluoride neurotoxicity. Fluoride. 44:195-209.

Broadbent JM, Thomson WM, Moffitt TE, Poulton R. 2015. Community water fluoridation and intelligence response. Am J Public Health. 105:3-4.

California Office of Environmental Health Hazard Assessment (OEHHA). 2011. Meeting synopsis and slide presentations: carcinogen identification committee meeting held on October 12, 2011. California Office of Environmental Health Hazard Assessment. http://oehha.ca.gov/prop65/public_meetings/cic101211synop.html. [19 August 2019]

CDC Division of Oral Health. 2020. Personal communication. September 3, 2020.

Centers for Disease Control and Prevention (CDC). 2013. Community water fluoridation: Fluoridation statistics. Atlanta, GA: Centers for Disease Control and Prevention. https://www.cdc.gov/fluoridation/statistics/2012stats.htm. [19 August 2019]

Centers for Disease Control and Prevention (CDC). 2015. Community water fluoridation FAQs: Infant formula Atlanta, GA: Centers for Disease Control and Prevention. https://www.cdc.gov/fluoridation/faqs/infant-formula.html. [22 September 2021]

Chen Y. 2012. Organophosphate-induced brain damage: Mechanisms, neuropsychiatric and neurological consequences, and potential therapeutic strategies. Neurotox. 33:391-400.

Chen YX, Han FL, Zhoua ZL, Zhang HQ, Jiao XS, Zhang SC, Huang MC, Chang TQ, Dong YF. 1991. [Research on the intellectual development of children in high fluoride areas]. Chin J Control Endem Dis. 6(Suppl):99-100.

Chen YX, Han FL, Zhoua ZL, Zhang HQ, Jiao XS, Zhang SC, Huang MC, Chang TQ, Dong YF. 2008. Research on the intellectual development of children in high fluoride areas. Fluoride. 41:120-124.

Chinoy NJ, Narayana MV. 1992. Studies on fluorosis in Mehsana District of North Gujarat. Proc Zool Soc. 45:157-161.

Choi AL, Sun G, Zhang Y, Grandjean P. 2012. Developmental fluoride neurotoxicity: A systematic review and meta-analysis. Environ Health Perspect. 120:1362-1368.

Choi AL, Zhang Y, Sun G, Bellinger DC, Wang K, Yang XJ, Li JS, Zheng Q, Fu Y, Grandjean P. 2015. Association of lifetime exposure to fluoride and cognitive functions in Chinese children: A pilot study. Neurotoxicol Teratol. 47:96-101.

Chouhan S, Flora SJS. 2008. Effects of fluoride on the tissue oxidative stress and apoptosis in rats: Biochemical assays supported by IR spectroscopy data. Toxicology. 254:61-67.

Chouhan S, Lomash V, Flora SJ. 2010. Fluoride-induced changes in haem biosynthesis pathway, neurological variables and tissue histopathology of rats. J Appl Toxicol. 30:63-73.

Cui Y, Zhang B, Ma J, Wang Y, Zhao L, Hou C, Yu J, Zhao Y, Zhang Z, Nie J et al. 2018. Dopamine receptor D2 gene polymorphism, urine fluoride, and intelligence impairment of children in China: A school-based cross-sectional study. Ecotoxicol Environ Saf. 165:270-277. https://doi.org/10.1016/j.ecoenv.2018.09.018

Cui Y, Yu J, Zhang B, Guo B, Gao T, Liu H. 2020. The relationships between thyroid-stimulating hormone and/or dopamine levels in peripheral blood and IQ in children with different urinary iodine concentrations. Neurosci Lett. 729:134981. https://doi.org/10.1016/j.neulet.2020.134981

Ding Y, Sun H, Han H, Wang W, Ji X, Liu X, Sun D. 2011. The relationships between low levels of urine fluoride on children's intelligence, dental fluorosis in endemic fluorosis areas in Hulunbuir, Inner Mongolia, China. J Hazard Mater. 186:1942-1946.

Du L, Wan C, Cao X, Liu J. 1992. [The effect of fluorine on the developing human brain]. Chin J Pathol. 21(4):218-220.

Du L, Wan C, Cao X, Liu J. 2008. The effect of fluorine on the developing human brain. Fluoride. 41:327-330.

Duan J, Zhao M, Wang L, Fang D, Wang Y, Wang W. 1995. A comparative analysis of the results of multiple tests in patients with chronic industrial fluorosis. Guizhou Med J. 18(3):179-180.

Duan Q, Jiao J, Chen X, Wang X. 2018. Association between water fluoride and the level of children's intelligence: A dose-response meta-analysis. Public Health. 154:87-97. https://doi.org/10.1016/j.puhe.2017.08.013

Gais S, Schonauer M. 2017. Untangling a cholinergic pathway from wakefulness to memory. Neuron. 94(4):696-698.

Gao Q, Liu YJ, Wu CX, Long YG, Guan ZZ. 2008a. [Level of oxidative stress in rat brains and learning and memory function of rats with chronic fluorosis]. Chin J Endemiol. 27:371-373.

Gao Q, Liu YJ, Wu CX, Long YG, Guan ZZ. 2008b. [Effects of fluoride on learning and memory and cholinesterase activity in rat brains]. Chin J Endemiol. 27:128-130.

Gao Q, Liu YJ, Guan ZZ. 2009. Decreased learning and memory ability in rats with fluorosis: Increased oxidative stress and reduced cholinesterase activity in the brain. Fluoride. 42:277-285.

Garman RH, Li AA, Kaufmann W, Auer RN, Bolon B. 2016. Recommended methods for brain processing and quantitative analysis in rodent developmental neurotoxicity studies. Toxicol Pathol. 44(1):14-42. https://doi.org/10.1177/0192623315596858

Green R, Lanphear B, Hornung R, Flora D, Martinez-Mier EA, Neufeld R, Ayotte P, Muckle G, Till C. 2019. Association between maternal fluoride exposure during pregnancy and IQ scores in offspring in Canada. JAMA Pediatr.E1-E9.

Green R, Rubenstein J, Popoli R, Capulong R, Till C. 2020. Sex-specific neurotoxic effects of early-life exposure to fluoride: A review of the epidemiologic and animal literature. Current Epidemiology Reports. 7(4):263-273. https://doi.org/10.1007/s40471-020-00246-1

Guner S, Uyar-Bozkurt S, Haznedaroglu E, Mentes A. 2016. Dental fluorosis and catalase immunoreactivity of the brain tissues in rats exposed to high fluoride pre- and postnatally. Biol Trace Elem Res. 174:150-157.

Guo XC, Wang RY, Cheng CF, Wei WS, Tang LM, Wang QS, Tang DX, Liu GW, He GD, Li SL. 1991. [A preliminary investigation of the IQs of 7-13 year-old children from an area with coal burning-related fluoride poisoning]. Chin J Epidemiol. 10(2):98-100.

Guo XC, Wang RY, Cheng CF, Wei WS, Tang LM, Wang QS, Tang DX, Liu GW, He GD, Li SL. 2008a. A preliminary investigation of the IQs of 7-13 year-old children from an area with coal burning-related fluoride poisoning. Fluoride. 41:125-128.

Guo ZY, He YH, Zhu QX. 2001. [Research on the neurobehavioral function of workers occupationally exposed to fluoride]. Ind Health & Occup Dis. 27(6):346-348.

Guo ZY, He YH, Zhu QX. 2008b. Research on the neurobehavioral function of workers occupationally exposed to fluoride. Fluoride. 41:152-155.

Guyatt GH, Oxman AD, Akl EA, Kunz R, Vist G, Brozek J, Norris S, Falck-Ytter Y, Glasziou P, Debeer H et al. 2011. GRADE guidelines: 1. Introduction-GRADE evidence profiles and summary of findings tables. J Clin Epidemiol. 64(4):383-394. https://doi.org/10.1016/j.jclinepi.2010.04.026

Haschek W, Rousseaux C. 1991. Handbook of toxicologic pathology. 1st ed.: Academic Press.

Health Canada. 2015. Third report on human biomonitoring of environmental chemicals in Canada - Results of the Canadian Health Measures Survey Cycle 3 (2012–2013). Ottawa, Ontario. https://www.canada.ca/content/dam/hc-sc/migration/hc-sc/ewh-semt/alt_formats/pdf/pubs/contaminants/chms-ecms-cycle3/chms-ecms-cycle3-eng.pdf.

Higgins JP, Green S. 2011. Cochrane handbook for systematic reviews of interventions. New York, NY: John Wiley & Sons.

Hosur MB, Puranik RS, Vanaki S, Puranik SR. 2012. Study of thyroid hormones free triiodothyronine (FT3), free thyroxine (FT4) and thyroid stimulating hormone (TSH) in subjects with dental fluorosis. Eur J Dent. 6:184-190.

Howard BE, Phillips J, Tandon A, Maharana A, Elmore R, Mav D, Sedykh A, Thayer K, Merrick BA, Walker V et al. 2020. SWIFT-Active Screener: Accelerated document screening

through active learning and integrated recall estimation. Environ Int. 138:105623. https://doi.org/10.1016/j.envint.2020.105623

Ibarluzea J, Gallastegi M, Santa-Marina L, Jiménez Zabala A, Arranz E, Molinuevo A, Lopez-Espinosa MJ, Ballester F, Villanueva CM, Riano I et al. 2021. Prenatal exposure to fluoride and neuropsychological development in early childhood: 1-to 4 years old children. Environ Res. 207:112181. https://doi.org/10.1016/j.envres.2021.112181

International Programme on Chemical Safety (IPCS). 2002. Fluorides. Geneva: World Health Organization, International Programme on Chemical Safety. Environmental Health Criteria 227. https://inchem.org/documents/ehc/ehc/ehc227.htm.

Jacqmin H, Commenges D, Letenneur L, Barberger-Gateau P, Dartigues JF. 1994. Components of drinking water and risk of cognitive impairment in the elderly. Am J Epidemiol. 139:48-57.

Jia B, Zong L, Lee JY, Lei J, Zhu Y, Xie H, Clemens JL, Feller MC, Na Q, Dong J et al. 2019. Maternal supplementation of low dose fluoride alleviates adverse perinatal outcomes following exposure to intrauterine inflammation. Scientific Reports. 9(1):2575. https://doi.org/10.1038/s41598-018-38241-8

Jiang C, Zhang S, Liu H, Guan Z, Zeng Q, Zhang C, Lei R, Xia T, Wang Z, Yang L et al. 2014. Low glucose utilization and neurodegenerative changes caused by sodium fluoride exposure in rat's developmental brain. Neuromolecular Med. 16:94-105.

Jin T, Han T, Wei Y, Wu Y, Wang Z, Zhang H. 2016. Investigation on working memory level of children aged 8-12 years in coal-burning fluorosis area. Journal of environment and health.409-411.

Jones S, Burt BA, Petersen PE, Lennon MA. 2005. The effective use of fluorides in public health. Bull World Health Organ. 83:670-676.

Khan AM, Raina R, Dubey N, Verma PK. 2017. Effect of deltamethrin and fluoride co-exposure on the brain antioxidant status and cholinesterase activity in Wistar rats. Drug Chem Toxicol. 41(2):1-5. https://doi.org/10.1080/01480545.2017.1321009

Khandare AL, Gourineni SR, Validandi V. 2017. Dental fluorosis, nutritional status, kidney damage, and thyroid function along with bone metabolic indicators in school-going children living in fluoride-affected hilly areas of Doda District, Jammu and Kashmir, India. Environ Monit Assess. 189:579. https://doi.org/10.1007/s10661-017-6288-5

Khandare AL, Validandi V, Gourineni SR, Gopalan V, Nagalla B. 2018. Dose-dependent effect of fluoride on clinical and subclinical indices of fluorosis in school going children and its mitigation by supply of safe drinking water for 5 years: An Indian study. Environ Monit Assess. 190:110. https://doi.org/10.1007/s10661-018-6501-1

Kheradpisheh Z, Mahvi AH, Mirzaei M, Mokhtari M, Azizi R, Fallahzadeh H, Ehrampoush MH. 2018a. Correlation between drinking water fluoride and TSH hormone by ANNs and ANFIS. J Environ Health Sci Eng. 16(1):11-18. https://doi.org/10.1007/s40201-018-0290-x

Kheradpisheh Z, Mirzaei M, Mahvi AH, Mokhtari M, Azizi R, Fallahzadeh H, Ehrampoush MH. 2018b. Impact of drinking water fluoride on human thyroid hormones: A case-control study. Sci Rep. 8:2674. https://doi.org/10.1038/s41598-018-20696-4

Kumar V, Chahar P, Kajjari S, Rahman F, Bansal DK, Kapadia JM. 2018. Fluoride, thyroid hormone derangements and its correlation with tooth eruption pattern among the pediatric population from endemic and non-endemic fluorosis areas. J Contemp Dent Pract. 19(12):1512-1516.

Li F, Chen X, Huang R, Xie Y. 2009. The impact of endemic fluorosis caused by the burning of coal on the development of intelligence in children. J Environ Health. 26(4):838-840.

Li J, Yao L, Q.L. S, Wu CY. 2004. [Effects of high fluoride level on neonatal neurobehavioral development]. Chin J Endemiol. 23(5):463-465.

Li J, Yao L, Shao QL, Wu CY. 2008a. Effects of high fluoride level on neonatal neurobehavioral development. Fluoride. 41:165-170.

Li M, Gao Y, Cui J, Li Y, Li B, Liu Y, Sun J, Liu X, Liu H, Zhao L et al. 2016. Cognitive impairment and risk factors in elderly people living in fluorosis areas in China. Biol Trace Elem Res. 172:53-60.

Li X, Hou G, Yu B, Yuan C, Liu Y, L Z. 2010. [Investigation and analysis of children's IQ and dental fluorosis in high fluoride area]. Chin J Pest Control. 26(3):230-231.

Li Y, Li X, Wei S. 1994. [Effects of high fluoride intake on child mental work capacity: Preliminary investigation into the mechanisms involved]. J West Chin University Med Sci. 25(2):188-191.

Li Y, Li X, Wei S. 2008b. Effects of high fluoride intake on child mental work capacity: Preliminary investigation into the mechanisms involved. Fluoride. 41:331-335.

Li YP, Jing XY, Chen D, Lin L, Wang ZJ. 2003. [Effects of endemic fluoride poisoning on the intellectual development of children in Baotou]. Chin J Public Health Manag. 19(4):337-338.

Li YP, Jing XY, Chen D, Lin L, Wang ZJ. 2008c. Effects of endemic fluoride poisoning on the intellectual development of children in Baotou. Fluoride. 41:161-164.

Liang C, Ji R, Cao J, Jiang Y, Yu B, Ma F, Wu Y, Ying B, Zhang Y, Sun S et al. 2003. Study on the relationship between drinking water trace elements and cognitive ability of the elderly. Health Res.436-440.

Lin F, Ai H, Zhao H, Lin J, Jhiang J, Maimaiti. 1991. [High fluoride and low iodine environment and subclinical cretinism in Xinjiang]. Endem Dis Bull. 6(2):62-67.

Liu YJ, Gao Q, Wu CX, Guan ZZ. 2010. Alterations of nAChRs and ERK1/2 in the brains of rats with chronic fluorosis and their connections with the decreased capacity of learning and memory. Toxicol Lett. 192:324-329.

Lou DD, Guan ZZ, Liu YJ, Liu YF, Zhang KL, Pan JG, Pei JJ. 2013. The influence of chronic fluorosis on mitochondrial dynamics morphology and distribution in cortical neurons of the rat brain. Arch Toxicol. 87:449-457.

Lu F, Zhang Y, Trivedi A, Jiang X, Chandra D, Zheng J, Nakano Y, Uyghurturk DA, Jalai R, Guner S et al. 2019. Fluoride related changes in behavioral outcomes may relate to increased serotonin. Physiol Behav. 206:76-83. https://doi.org/10.1016/j.physbeh.2019.02.017

Ma Q, Huang H, Sun L, Zhou T, Zhu J, Cheng X, Duan L, Li Z, Cui L, Ba Y. 2017. Gene-environment interaction: Does fluoride influence the reproductive hormones in male farmers modified by ERα gene polymorphisms? Chemosphere. 188:525-531. https://doi.org/10.1016/j.chemosphere.2017.08.166

Malin AJ, Till C. 2015. Exposure to fluoridated water and attention deficit hyperactivity disorder prevalence among children and adolescents in the United States: An ecological association. Environ Health. 14:17.

Malin AJ, Riddell J, McCague H, Till C. 2018. Fluoride exposure and thyroid function among adults living in Canada: Effect modification by iodine status. Environ Int. 121(Pt 1):667-674. https://doi.org/10.1016/j.envint.2018.09.026

McPherson CA, Zhang G, Gilliam R, Brar SS, Wilson R, Brix A, Picut C, Harry GJ. 2018. An evaluation of neurotoxicity following fluoride exposure from gestational through adult ages in Long-Evans hooded rats. Neurotoxicol Res.1-18. https://doi.org/10.1007/s12640-018-9870-x

Mesram N, Nagapuri K, Banala RR, Nalagoni CR, Karnati PR. 2016. Quercetin treatment against NaF induced oxidative stress related neuronal and learning changes in developing rats. J King Saud Univ Sci. 29:221-229.

Michael M, Barot VV, Chinoy NJ. 1996. Investigations of soft tissue functions in fluorotic individuals of North Gujarat. Fluoride. 29:63-71.

Miller K, Howard B, Phillips J, Shah M, Mav D, Thayer K, Shah R. 2016. SWIFT-Active screener: Reducing literature screening effort through machine learning for systematic reviews. Cochrane Colloquium Seoul; October 25 2016; Seoul, Korea.

Moher D, Liberati A, Tetzlaff J, Altman D. 2009. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. PLoS Med. 6(7):e1000097. https://doi.org/10.1371/journal.pmed.1000097

Morgan L, Allred E, Tavares M, Bellinger D, Needleman H. 1998. Investigation of the possible associations between fluorosis, fluoride exposure, and childhood behavior problems. Pediatr Dent. 20:244-252.

Mustafa DE, Younis UM, Elhag SAA. 2018. The relationship between the fluoride levels in drinking water and the schooling performance of children in rural areas of Khartoum State, Sudan. Fluoride. 51(2):102-113.

Nageshwar M, Sudhakar K, Reddy KP. 2018. Quercetin ameliorates oxidative stress, neural damage of brain and behavioral impairment of rat with fluoride exposure. Int J Pharm Sci Res. 9(8):3247-3256. https://doi.org/10.13040/IJPSR.0975-8232.9(8).3247-56

National Academies of Sciences Engineering and Medicine (NASEM). 2020. Review of the draft NTP monograph: Systematic review of fluoride exposure and neurodevelopmental and cognitive health effects. Washington, DC: National Academies of Sciences, Engineering and Medicine. https://doi.org/10.17226/25715.

National Academies of Sciences Engineering and Medicine (NASEM). 2021. Review of the revised NTP Monograph on the systematic review of fluoride exposure and neurodevelopmental and cognitive health effects: A letter report. Washington, DC: National Academies of Sciences, Engineering and Medicine. https://doi.org/10.17226/26030.

National Institute for Occupational Safety and Health (NIOSH). 1984. Fluoride in urine. Washington, DC: National Institute for Occupational Safety and Health. Method 8308.

National Research Council (NRC). 2006. Committee on fluoride in drinking water, board on environmental studies and toxicology. Fluoride in drinking water: A scientific review of EPA's standards.: National Research Council. http://www.nap.edu/catalog/11571/fluoride-in-drinking-water-a-scientific-review-of-epas-standards. [19 August 2019]

National Toxicology Program (NTP). 2016. Systematic literature review on the effects of fluoride on learning and memory in animal studies. Research Triangle Park, NC: National Toxicology Program. NTP Research Report 1. https://ntp.niehs.nih.gov/ntp/results/pubs/rr/reports/01fluoride_508.pdf. [19 August 2019]

National Toxicology Program (NTP). 2020. Revised draft NTP Monograph on the systematic review of fluoride exposure and neurodevelopmental and cognitive health effects. Research Triangle Park, NC: National Toxicology Program. https://www.nationalacademies.org/event/10-19-2020/docs/DDA97C9170D1A255D69C004CEB77B698E8D005011EFB.

Niu Q, Chen J, Xia T, Li P, Zhou G, Xu C, Zhao Q, Dong L, Zhang S, Wang A. 2018. Excessive ER stress and the resulting autophagic flux dysfunction contribute to fluoride-induced neurotoxicity. Environ Pollut. 233:889-899. https://doi.org/10.1016/j.envpol.2017.09.015

Nkpaa KW, Onyeso GI. 2018. Rutin attenuates neurobehavioral deficits, oxidative stress, neuro-inflammation and apoptosis in fluoride treated rats. Neurosci Lett. 682:92-99. https://doi.org/10.1016/j.neulet.2018.06.023

Peckham S, Lowery D, Spencer S. 2015. Are fluoride levels in drinking water associated with hypothyroidism prevalence in England? A large observational study of GP practice data and fluoride levels in drinking water. J Epidemiol Community Health. 69:619-624.

Podgorski J, Berg M. 2020. Global threat of arsenic in groundwater. Science. 368(6493):845-850. https://doi.org/10.1126/science.aba1510

Pulungan ZS, Sofro ZM, Partadiredja G. 2016. Sodium fluoride does not affect the working memory and number of pyramidal cells in rat medial prefrontal cortex. Anat Sci Int. 93(1):128-138.

Reddy YP, Tiwari SK, Shaik AP, Alsaeed A, Sultana A, Reddy PK. 2014. Effect of sodium fluoride on neuroimmunological parameters, oxidative stress and antioxidative defenses. Toxicol Mech Methods. 24:31-36.

Ren D, Li K, Liu D. 1989. [A study of the intellectual ability of 8-14 year-old children in high fluoride, low iodine areas]. Chin J Control Endem Dis. 4(4):251.

Ren D, Li K, Liu D. 2008. A study of the intellectual ability of 8-14 year-old children in high fluoride, low iodine areas. Fluoride. 41:319-320.

Riddell JK, Malin AJ, Flora D, McCague H, Till C. 2019. Association of water fluoride and urinary fluoride concentrations with attention deficit hyperactivity disorder in Canadian youth. Environ Int. 133:105190. https://doi.org/10.1016/j.envint.2019.105190

Rocha-Amador D, Navarro ME, Carrizales L, Morales R, Calderon J. 2007. Decreased intelligence in children and exposure to fluoride and arsenic in drinking water. Cad Saude Publica. 23(Suppl 4):S579-587.

Rocha-Amador D, Navarro M, Trejo-Acevedo A, Carrizales L, Perez-Maldonado I, Diaz-Barriga F, Calderon J. 2009. Use of the Rey-Osterrieth Complex Figure Test for neurotoxicity evaluation of mixtures in children. Neurotox. 30:1149-1154.

Rooney AA, Boyles AL, Wolfe MS, Bucher JR, Thayer KA. 2014. Systematic review and evidence integration for literature-based environmental health science assessments. Environ Health Perspect. 122(7):711-718.

Rotton J, Tikofsky RS, Feldman HT. 1982. Behavioral effects of chemicals in drinking water. J Appl Psychol. 67:230-238.

Russ TC, Killin LOJ, Hannah J, Batty GD, Deary IJ, Starr JM. 2019. Aluminium and fluoride in drinking water in relation to later dementia risk. Brit J Psychol. 14:1-6. https://doi.org/10.1192/bjp.2018.287

Santa-Marina L, Jimenez-Zabala A, Molinuevo A, Lopez-Espinosa M, Villanueva C, Riano I, Ballester F, Sunyer J, Tardon A, Ibarluzea J. 2019. Fluorinated water consumption in pregnancy and neuropsychological development of children at 14 months and 4 years of age. Environ Epidemiol. 3:386-387. https://dx.doi.org/10.1097/01.EE9.0000610304.33479.18

Saxena S, Sahay A, Goel P. 2012. Effect of fluoride exposure on the intelligence of school children in Madhya Pradesh, India. J Neurosci Rural Pract. 3:144-149.

Scientific Committee on Health and Environmental Risks (SCHER). 2011. Critical review of any new evidence on the hazard profile, health effects, and human exposure to fluoride and the fluoridating agents of drinking water. Scientific Committee on Health and Environmental Risks. http://ec.europa.eu/health/scientific_committees/environmental_risks/docs/scher_o_139.pdf. [19 August 2019]

Seraj B, Shahrabi M, Shadfar M, Ahmadi R, Fallahzadeh M, Eslamlu HF, Kharazifard MJ. 2012. Effect of high water fluoride concentration on the intellectual development of children in Makoo, Iran. J Dent. 9:221-229.

Shan KR, Qi XL, Long YG, Nordberg A, Guan ZZ. 2004. Decreased nicotinic receptors in PC12 cells and rat brains influenced by fluoride toxicity: A mechanism relating to a damage at the level in post-transcription of the receptor genes. Toxicology. 200:169-177.

Shannon FT, Fergusson DM, Horwood LJ. 1986. Exposure to fluoridated public water supplies and child health and behaviour. New Zeal Med J. 99:416-418.

Shao Q. 2003. [Study of cognitive function impairment caused by chronic fluorosis]. Chin J Endemiol. 22(4):336-338.

Sharma JD, Sohu D, Jain P. 2009. Prevalence of neurological manifestations in a human population exposed to fluoride in drinking water. Fluoride. 42:127-132.

Singh N, Verma KG, Verma P, Sidhu GK, Sachdeva S. 2014. A comparative study of fluoride ingestion levels, serum thyroid hormone & TSH level derangements, dental fluorosis status among school children from endemic and non-endemic fluorosis areas. Springerplus. 3:7.

Singh V, Singh C, Sandeep T, Sandeep K, Vikas G, Mukesh T, Anurag T. 2013. A correlation between serum vitamin, acetylcholinesterase activity and IQ in children with excessive endemic fluoride exposure in Rajasthan, India. Int Res J Medical Sci. 1(3):12-16.

Soto-Barreras U, Escalante-Villalobos KY, Holguin-Loya B, Perez-Aguirre B, Nevarez-Rascon A, Martinez-Martinez RE, Loyola-Rodriguez JP. 2019. Effect of fluoride in drinking water on dental caries and IQ in children. Fluoride. 52:474-482.

Still CN, Kelley P. 1980. The incidence of primary degenerative dementia vs. water fluoride content in South Carolina, USA. Neurotox. 1:125-132.

Sudhakar K, Reddy KP. 2018. Protective effects of Abelmoschus moschatus seed extract on neurotransmitter system of developing brain of Wistar rats with gestational and post-natal exposure of sodium fluoride. Int J Green Pharm. 12:S131-S139.

Sudhir KM, Chandu GN, Prashant GM, Subba Reddy VV. 2009. Effect of fluoride exposure on intelligence quotient (IQ) among 13-15 year old school children of known endemic area of fluorosis, Nalgonda District, Andhra Pradesh. J Indian Assoc Public Health Dent. 2009(13):88-94.

Sun M, Li S, Wang Y, Li F. 1991. [Measurement of intelligence by drawing test among the children in the endemic area of Al-F combined toxicosis]. J Guiyang Med Coll. 16(3):204-206.

Sun ZR, Liu FZ, Wu LN, Lu Y, Yu DK. 2000. [Effects of high fluoride drinking water on the cerebral function of mice]. Chin J Epidemiol. 19:262-263.

Sun ZR, Liu FZ, Wu LN, Lu Y, Yu DK. 2008. Effects of high fluoride drinking water on the cerebral function of mice. Fluoride. 41:148-151.

Susheela AK, Bhatnagar M, Vig K, Mondal NK. 2005. Excess fluoride ingestion and thyroid hormone derangements in children living in Delhi, India. Fluoride. 38:98-108.

Till C, Green R, Grundy JG, Hornung R, Neufeld R, Martinez-Mier EA, Ayotte P, Muckle G, Lanphear B. 2018. Community water fluoridation and urinary fluoride concentrations in a national sample of pregnant women in Canada. Environ Health Perspect. 126(10):107001. https://doi.org/10.1289/ehp3546

Till C, Green R, Flora D, Hornung R, Martinez-Mier EA, Blazer M, Farmus L, Ayotte P, Muckle G, Lanphear B. 2020. Fluoride exposure from infant formula and child IQ in a Canadian birth cohort. Environ Int. 134:105315. https://doi.org/10.1016/j.envint.2019.105315

Trivedi M, Sangai N, Patel R, Payak M, Vyas S. 2012. Assessment of groundwater quality with special reference to fluoride and its impact on IQ of schoolchildren in six villages of the Mundra Region, Kachchh, Gujurat, India. Fluoride. 45(4):377-383.

Trivedi MH, Verma RJ, Chinoy NJ, Patel RS, Sathawara NG. 2007. Effect of high fluoride water on intelligence of school children in India. Fluoride. 40:178-183.

Tsunoda M, Aizawa Y, Nakano K, Liu Y, Horiuchi T, Itai K, Tsunoda H. 2005. Changes in fluoride levels in the liver, kidney, and brain and in neurotransmitters of mice after subacute administration of fluoride. Fluoride. 38:284-292.

U.S. Department of Health and Human Services (US DHHS). 2015. U.S. Public Health Service recommendation for fluoride concentration in drinking water for the prevention of dental caries. U.S. Department of Health and Human Services. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4547570/. [19 August 2019]

U.S. Environmental Protection Agency (US EPA). 2010. Fluoride: Exposure and relative source contribution analysis. Washington, DC: U.S. Environmental Protection Agency. 820-R-10-015. http://www.epa.gov/dwstandardsregulations/fluoride-risk-assessment-and-relative-source-contribution. [19 August 2019]

Valdez Jimenez L, Lopez Guzman OD, Cervantes Flores M, Costilla-Salazar R, Calderon Hernandez J, Alcaraz Contreras Y, Rocha-Amador DO. 2017. In utero exposure to fluoride and cognitive development delay in infants. Neurotox. 59:65-70. https://doi.org/10.1016/j.neuro.2016.12.011

Villa A, Anabalon M, Zohouri V, Maguire A, Franco AM, Rugg-Gunn A. 2010. Relationships between fluoride intake, urinary fluoride excretion and fluoride retention in children and adults: An analysis of available data. Caries Res. 44(1):60-68. https://doi.org/10.1159/000279325

Wang A, Duan L, Huang H, Ma J, Zhang Y, Ma Q, Guo Y, Li Z, Cheng X, Zhu J et al. 2020a. Association between fluoride exposure and behavioural outcomes of school-age children: A pilot study in China. Int J Environ Health Res.1-10. https://doi.org/10.1080/09603123.2020.1747601

Wang G, Gao M, Zhang M, Yang M, Xiang Q. 2012. Correlation between total fluoride intake and children's IQ. J Southeast Univ Med Ed.743-746.

Wang M, Liu L, Li H, Li Y, Liu H, Hou C, Zeng Q, Li P, Zhao Q, Dong L et al. 2020b. Thyroid function, intelligence, and low-moderate fluoride exposure among Chinese school-age children. Environ Int. 134:105229. https://doi.org/10.1016/j.envint.2019.105229

Wang X, Wang L, Hu P, Guo X, Luo X. 2001. [Effects of high iodine and high fluorine on children’s intelligence and thyroid function]. Chin J Endemiol. 20(4):288-290.

Watanabe M, Kono K, Orita Y, Dote T, Usuda K, Takahashi Y, Yoshida Y. 1995. Influence of dietary fluoride intake on urinary fluoride concentration and evaluation of corrected levels in spot urine. Fluoride. 28(2):61-70.

Waugh DT. 2019. Fluoride exposure induces inhibition of sodium/iodide symporter (NIS) contributing to impaired iodine absorption and iodine deficiency: Molecular mechanisms of inhibition and implications for public health. Int J Environ Res Public Health. 16(6). https://doi.org/10.3390/ijerph16061086

World Health Organization (WHO). 2008. Guidelines for drinking-water quality [electronic resource]: Incorporating 1st and 2nd addenda. Geneva, Switzerland: World Health Organization. Third Edition. Vol. 1. https://apps.who.int/iris/bitstream/handle/10665/204411/9789241547611_eng.pdf?sequence=1&isAllowed=y.

World Health Organization (WHO). 2017. Guidelines for drinking-water quality. World Health Organization. 4th ed. + 1st add. https://apps.who.int/iris/handle/10665/254637.

Xiang Q, Liang Y, Chen L, Wang C, Chen B, Chen X, Zhou M. 2003a. Effect of fluoride in drinking water on children's intelligence. Fluoride. 36:84-94.

Xiang Q, Liang Y, Chen B, Chen L. 2011. Analysis of children's serum fluoride levels in relation to intelligence scores in a high and low fluoride water village in China. Fluoride. 44:191-194.

Xiang Q, Wang Y, Yang M, Zhang M, Xu Y. 2013. Level of fluoride and arsenic in household shallow well water in Wamiao and Xinhuai villages in Jiangsu province, China. Fluoride. 46:192-197.

Xiang QY, Liang YX, Zhou MS, Zang HB. 2003b. Blood lead of children in Wamiao-Xinhuai intelligence study. Fluoride. 36:198-199.

Yang Y, Wang X, X. G, Hu P. 1994. [The effects of high levels of fluoride and iodine on child intellectual ability and the metabolism of fluoride and iodine]. Chin J Epidemiol. 15(4):296-298.

Yang Y, Wang X, Guo X, Hu P. 2008. The effects of high levels of fluoride and iodine on child intellectual ability and the metabolism of fluoride and iodine. Fluoride. 41:336-339.

Yao L, Zhou J, Wang S, Cui K, Lin F. 1996. Analysis on TSH and intelligence level of children with dental Fluorosis in a high fluoride area. Lit Inf Prev Med. 2(1):26-27.

Yasmin S, Ranjan S, D'Souza D. 2013. Effect of excess fluoride ingestion on human thyroid function in Gaya Region, Bihar, India. Toxicol Environ Chem. 95:1235-1243.

Yazdi SM, Sharifian A, Dehghani-Beshne M, Momeni VR, Aminian O. 2011. Effects of fluoride on psychomotor performance and memory of aluminum potroom workers. Fluoride. 44:158-162.

Yu X, Chen J, Li Y, Liu H, Hou C, Zeng Q, Cui Y, Zhao L, Li P, Zhou Z et al. 2018. Threshold effects of moderately excessive fluoride exposure on children's health: A potential association between dental fluorosis and loss of excellent intelligence. Environ Int. 118:116-124. https://doi.org/10.1016/j.envint.2018.05.042

Yu YN, Yang WX, Dong Z, Wan CW, Zhang JT, Liu JL, Xiao KQ, Huang YS, Lu BF. 1996. [Neurotransmitter and receptor changes in the brains of fetuses from areas of endemic fluorosis]. Chin J Endemiol. 15(5):257-259.

Yu YN, Yang WX, Dong Z, Wan CW, Zhang JT, Liu JL, Xiao KQ, Huang YS, Lu BF. 2008. Neurotransmitter and receptor changes in the brains of fetuses from areas of endemic fluorosis. Fluoride. 41:134-138.

Zhang KL, Lou DD, Guan ZZ. 2015a. Activation of the AGE/RAGE system in the brains of rats and in SH-SY5Y cells exposed to high level of fluoride might connect to oxidative stress. Neurotoxicol Teratol. 48:49-55.

Zhang S, Zhang X, Liu H, Qu W, Guan Z, Zeng Q, Jiang C, Gao H, Zhang C, Lei R et al. 2015b. Modifying effect of COMT gene polymorphism and a predictive role for proteomics analysis in children's intelligence in endemic fluorosis area in Tianjin, China. Toxicol Sci. 144:238-245.

Zhao Q, Niu Q, Chen J, Xia T, Zhou G, Li P, Dong L, Xu C, Tian Z, Luo C et al. 2019. Roles of mitochondrial fission inhibition in developmental fluoride neurotoxicity: Mechanisms of action in vitro and associations with cognition in rats and children. Arch Toxicol. 93(3):709-726. https://doi.org/10.1007/s00204-019-02390-0

Zhao Q, Tian Z, Zhou G, Niu Q, Chen J, Li P, Dong L, Xia T, Zhang S, Wang A. 2020. SIRT1-dependent mitochondrial biogenesis supports therapeutic effects of resveratrol against neurodevelopment damage by fluoride. Theranostics. 10:4822-4838. https://doi.org/10.7150/thno.42387

Zhou G, Tang S, Yang L, Niu Q, Chen J, Xia T, Wang S, Wang M, Zhao Q, Liu L et al. 2019. Effects of long-term fluoride exposure on cognitive ability and the underlying mechanisms: Role of autophagy and its association with apoptosis. Toxicol Appl Pharmacol. 378:114608. https://doi.org/10.1016/j.taap.2019.114608

Zhou T, Duan L-J, Ding Z, Yang R-P, Li S-H, Xi Y, Cheng X-M, Hou J-X, Wen S-B, Chen J et al. 2012. Environmental fluoride exposure and reproductive hormones in male living in endemic fluorosis villages in China. Life Sci J. 9(4):1-7.

Zohouri FV, Swinbank CM, Maguire A, Moynihan PJ. 2006. Is the fluoride/creatinine ratio of a spot urine sample indicative of 24-h urinary fluoride? Community Dent Oral Epidemiol. 34(2):130-138. https://doi.org/10.1111/j.1600-0528.2006.00269.x

Appendix A. Data Figures: Neurodevelopmental or Cognitive Effects and Outcomes

Figures

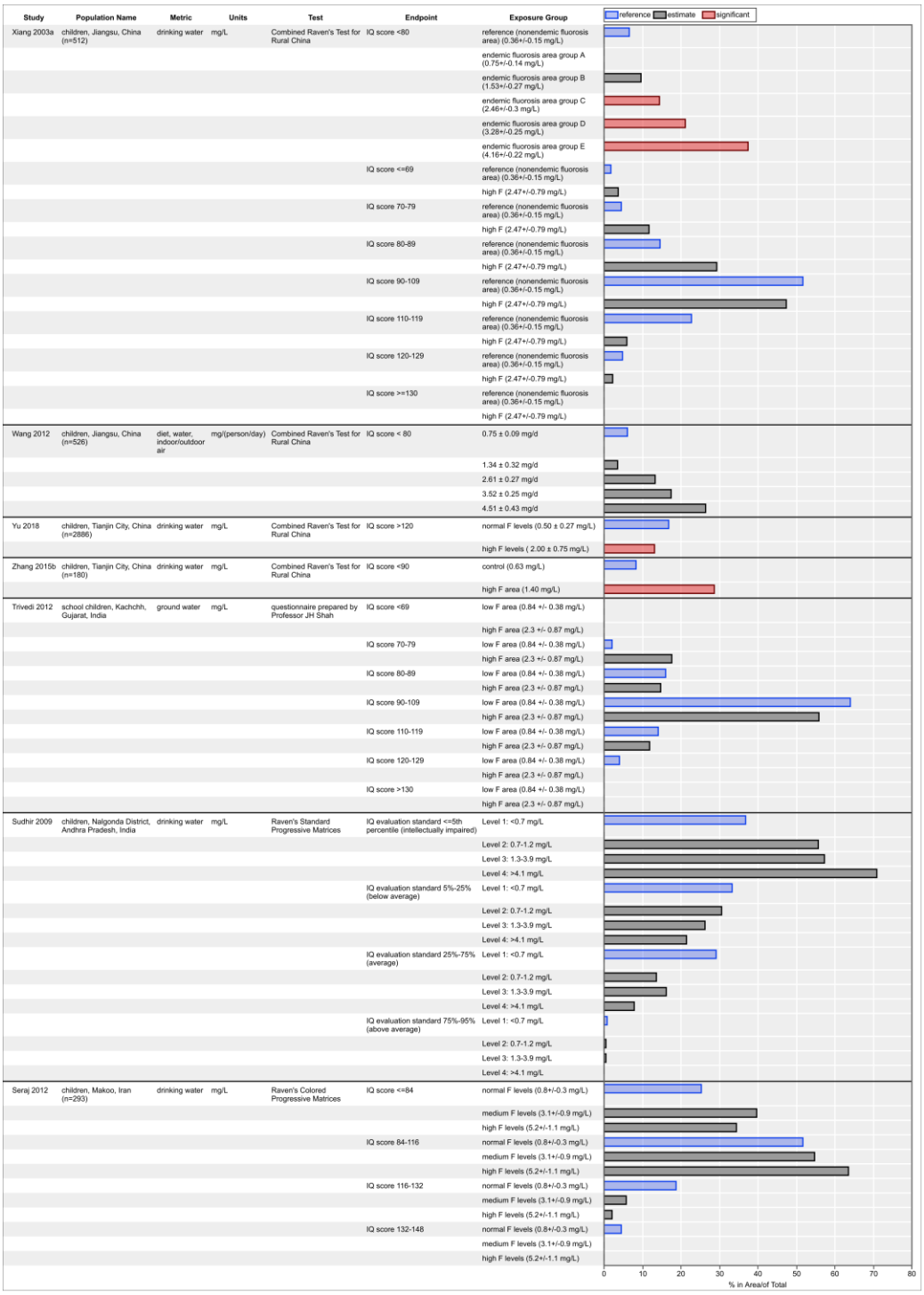


Figure A-1. Distribution of IQ in Children by Fluoride Exposure (Low Risk-of-bias Studies; Presented as % in Area or % of Total Group)

Reference group indicated by blue bars; other bars represent response estimates with red indicating statistical significance compared with the reference group.
An interactive version of Figure A-1 and additional study details in HAWC here. "F" represents fluoride. For IQ distribution results by drinking water fluoride level provided in Xiang et al. (2003a), Trivedi et al. (2012), Sudhir et al. (2009), and Seraj et al. (2012) and rate of low IQ scores by fluoride intake provided in Wang et al. (2012), statistical significance was not evaluated.

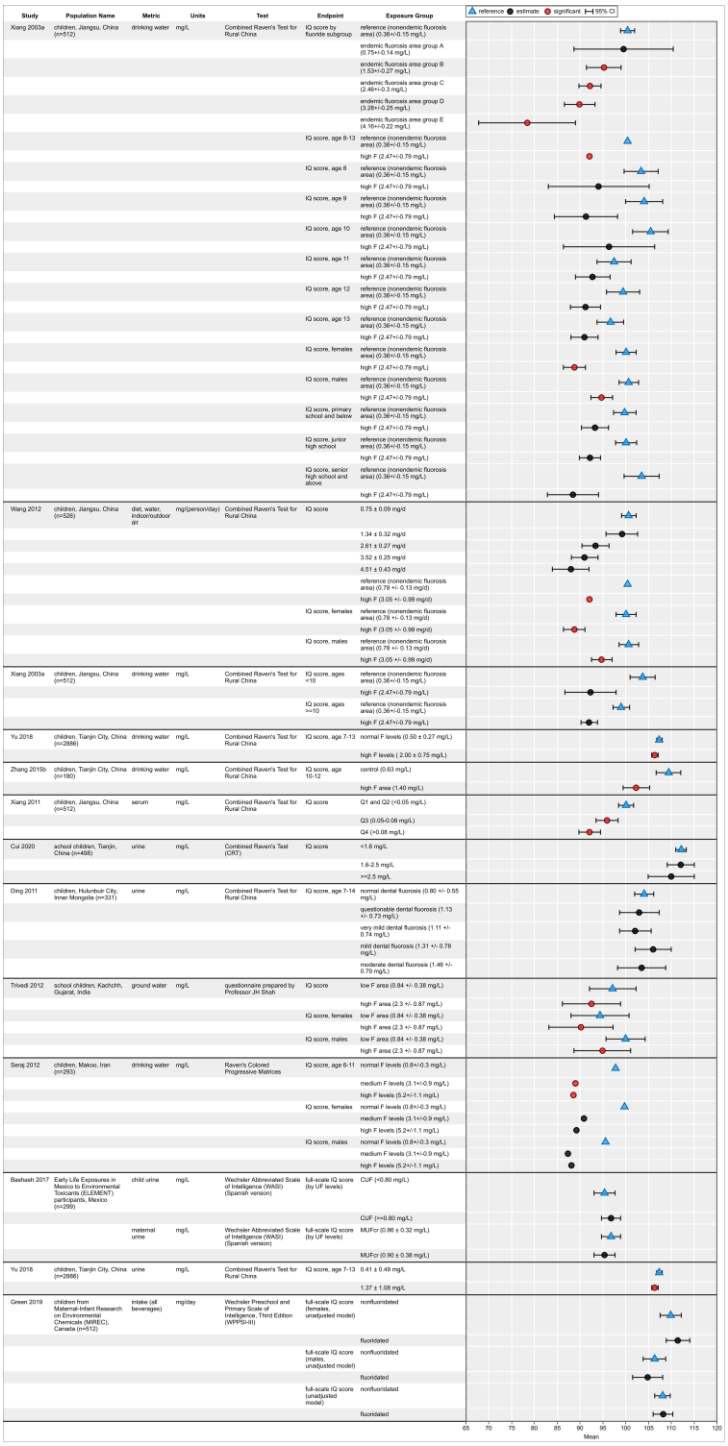


Figure A-2. Mean IQ in Children by Fluoride Exposure (Low Risk-of-bias Studies)

Reference group indicated by blue triangles; circles represent response estimates with red indicating statistical significance. An interactive version of Figure A-2 and additional study details in HAWC here. "F" represents fluoride. Three additional publications based on subsample of the larger Yu et al. (2018) cohort were identified (Zhao et al. 2019; Zhao et al. 2020; Zhou et al. 2019); however, results from these studies are not presented here. The main study by Yu et al. (2018) is considered a better representation of the IQ results. For all studies, SDs are available and can be viewed in HAWC by clicking the data points within the plot area; however, 95% CIs could not be calculated for Seraj et al. (2012) because Ns are not available for exposure groups.

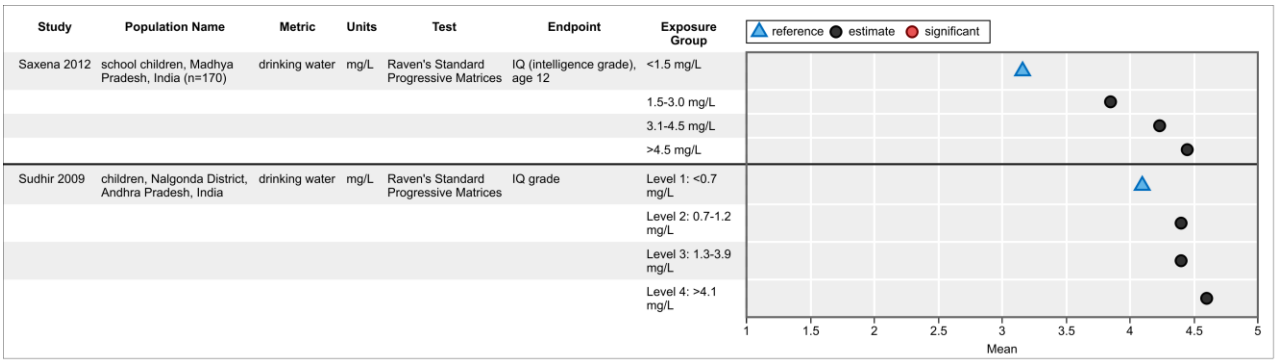


Figure A-3. Intelligence Grade in Children by Fluoride Exposure (Low Risk-of-bias Studies; Presented as Mean)

Reference group indicated by blue triangles; circles represent response estimates with red indicating statistical significance. An interactive version of Figure A-3 and additional study details in HAWC here. For Saxena et al. (2012), children's intelligence was measured using Raven's Standard Progressive Matrices. Children's scores were converted to percentile, and specific grades were allotted based on the percentiles. Grades ranged from intellectually superior (Grade I) to intellectually impaired (Grade V). Results for Soto-Barreras et al. (2019) are not presented here. Outcomes in the study were presented as levels of fluoride exposure associated with each intelligence grade. Results reported were not significant.

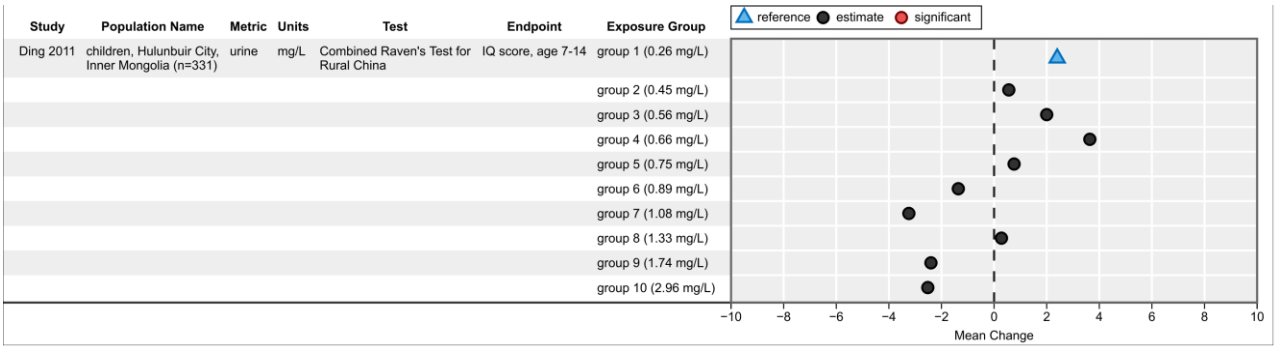
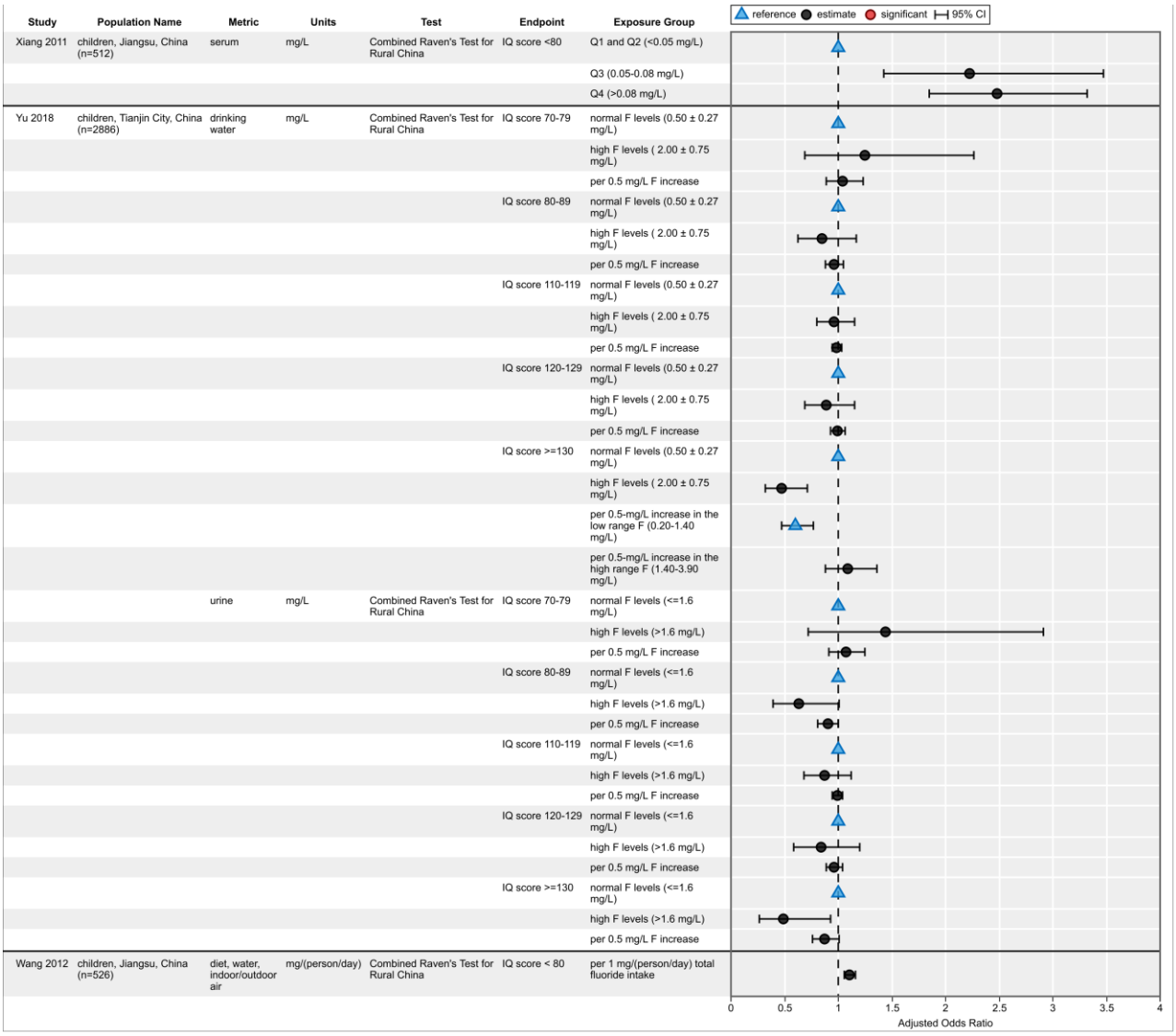


Figure A-4. Mean Change in IQ in Children by Fluoride Exposure (Low Risk-of-bias Studies)

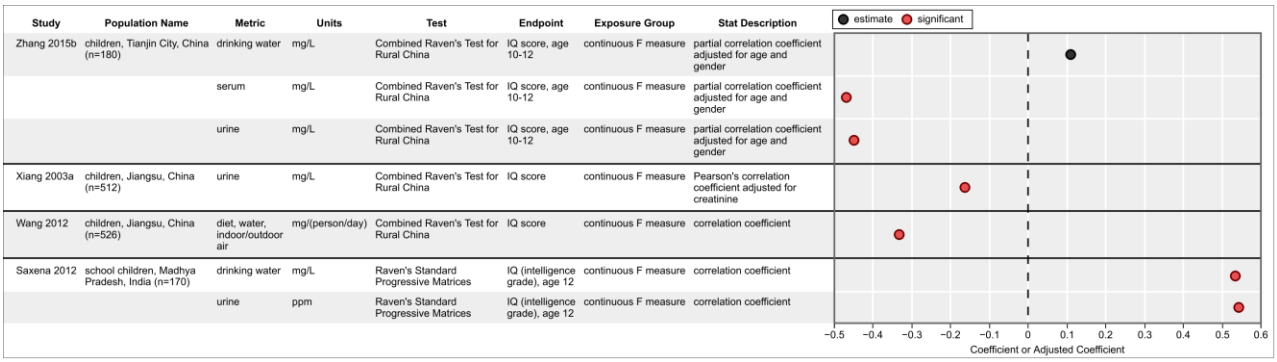
Reference group indicated by blue triangles; circles represent response estimates with red indicating statistical significance. An interactive version of Figure A-4 and additional study details in HAWC here. For Ding et al. (2011), SDs are available and can be viewed in HAWC by clicking the data points within the plot area; however, 95% CIs could not be calculated because Ns for each exposure group are not available.



Study	Population Name	Metric	Units	Test	Endpoint	Exposure Group
Xiang 2011	children, Jiangsu, China (n=512)	serum	mg/L	Combined Raven's Test for Rural China	IQ score <80	Q1 and Q2 (<0.05 mg/L)
						Q3 (0.05-0.08 mg/L)
						Q4 (>0.08 mg/L)
Yu 2018	children, Tianjin City, China (n=2886)	drinking water	mg/L	Combined Raven's Test for Rural China	IQ score 70-79	normal F levels (0.50 ± 0.27 mg/L)
						high F levels (2.00 ± 0.75 mg/L)
						per 0.5 mg/L F increase
					IQ score 80-89	normal F levels (0.50 ± 0.27 mg/L)
						high F levels (2.00 ± 0.75 mg/L)
						per 0.5 mg/L F increase
					IQ score 110-119	normal F levels (0.50 ± 0.27 mg/L)
						high F levels (2.00 ± 0.75 mg/L)
						per 0.5 mg/L F increase
					IQ score 120-129	normal F levels (0.50 ± 0.27 mg/L)
						high F levels (2.00 ± 0.75 mg/L)
						per 0.5 mg/L F increase
					IQ score >=130	normal F levels (0.50 ± 0.27 mg/L)
						high F levels (2.00 ± 0.75 mg/L)
						per 0.5-mg/L increase in the low range F (0.20-1.40 mg/L)
						per 0.5-mg/L increase in the high range F (1.40-3.90 mg/L)
		urine	mg/L	Combined Raven's Test for Rural China	IQ score 70-79	normal F levels (<=1.6 mg/L)
						high F levels (>1.6 mg/L)
						per 0.5 mg/L F increase
					IQ score 80-89	normal F levels (<=1.6 mg/L)
						high F levels (>1.6 mg/L)
						per 0.5 mg/L F increase
					IQ score 110-119	normal F levels (<=1.6 mg/L)
						high F levels (>1.6 mg/L)
						per 0.5 mg/L F increase
					IQ score 120-129	normal F levels (<=1.6 mg/L)
						high F levels (>1.6 mg/L)
						per 0.5 mg/L F increase
					IQ score >=130	normal F levels (<=1.6 mg/L)
						high F levels (>1.6 mg/L)
						per 0.5 mg/L F increase
Wang 2012	children, Jiangsu, China (n=526)	diet, water, indoor/outdoor air	mg/(person/day)	Combined Raven's Test for Rural China	IQ score < 80	per 1 mg/(person/day) total fluoride intake

Figure A-5. Associations between Fluoride Exposure and IQ Scores in Children (Low Risk-of-bias Studies; Presented as Adjusted OR)

Reference group indicated by blue triangles; circles represent response estimates with red indicating statistical significance. Cutoffs for the dichotomous outcome are listed in the Endpoint column.
An interactive version of Figure A-5 and additional study details in HAWC here. For Xiang et al. (2011), there was a significant linear trend across different levels of serum fluoride for IQ score <80 ($p < 0.001$). For Yu et al. (2018), significance levels by IQ score were not reported.



Study	Population Name	Metric	Units	Test	Endpoint	Exposure Group	Stat Description
Zhang 2015b	children, Tianjin City, China (n=180)	drinking water	mg/L	Combined Raven's Test for Rural China	IQ score, age 10-12	continuous F measure	partial correlation coefficient adjusted for age and gender
		serum	mg/L	Combined Raven's Test for Rural China	IQ score, age 10-12	continuous F measure	partial correlation coefficient adjusted for age and gender
		urine	mg/L	Combined Raven's Test for Rural China	IQ score, age 10-12	continuous F measure	partial correlation coefficient adjusted for age and gender
Xiang 2003a	children, Jiangsu, China (n=512)	urine	mg/L	Combined Raven's Test for Rural China	IQ score	continuous F measure	Pearson's correlation coefficient adjusted for creatinine
Wang 2012	children, Jiangsu, China (n=526)	diet, water, indoor/outdoor air	mg/(person/day)	Combined Raven's Test for Rural China	IQ score	continuous F measure	correlation coefficient
Saxena 2012	school children, Madhya Pradesh, India (n=170)	drinking water	mg/L	Raven's Standard Progressive Matrices	IQ (intelligence grade), age 12	continuous F measure	correlation coefficient
		urine	ppm	Raven's Standard Progressive Matrices	IQ (intelligence grade), age 12	continuous F measure	correlation coefficient

Figure A-6. Correlations between Fluoride Exposure and IQ Score in Children (Low Risk-of-bias Studies; Presented as Correlation Coefficient)

Circles represent response estimates with red indicating statistical significance.
An interactive version of Figure A-6 and additional study details in HAWC here. "F" represents fluoride. For Saxena et al. (2012), a significant relationship between water fluoride level and intelligence grade was observed. Increasing intelligence grades reflected increasing levels of impairment (reduced intelligence) in children.

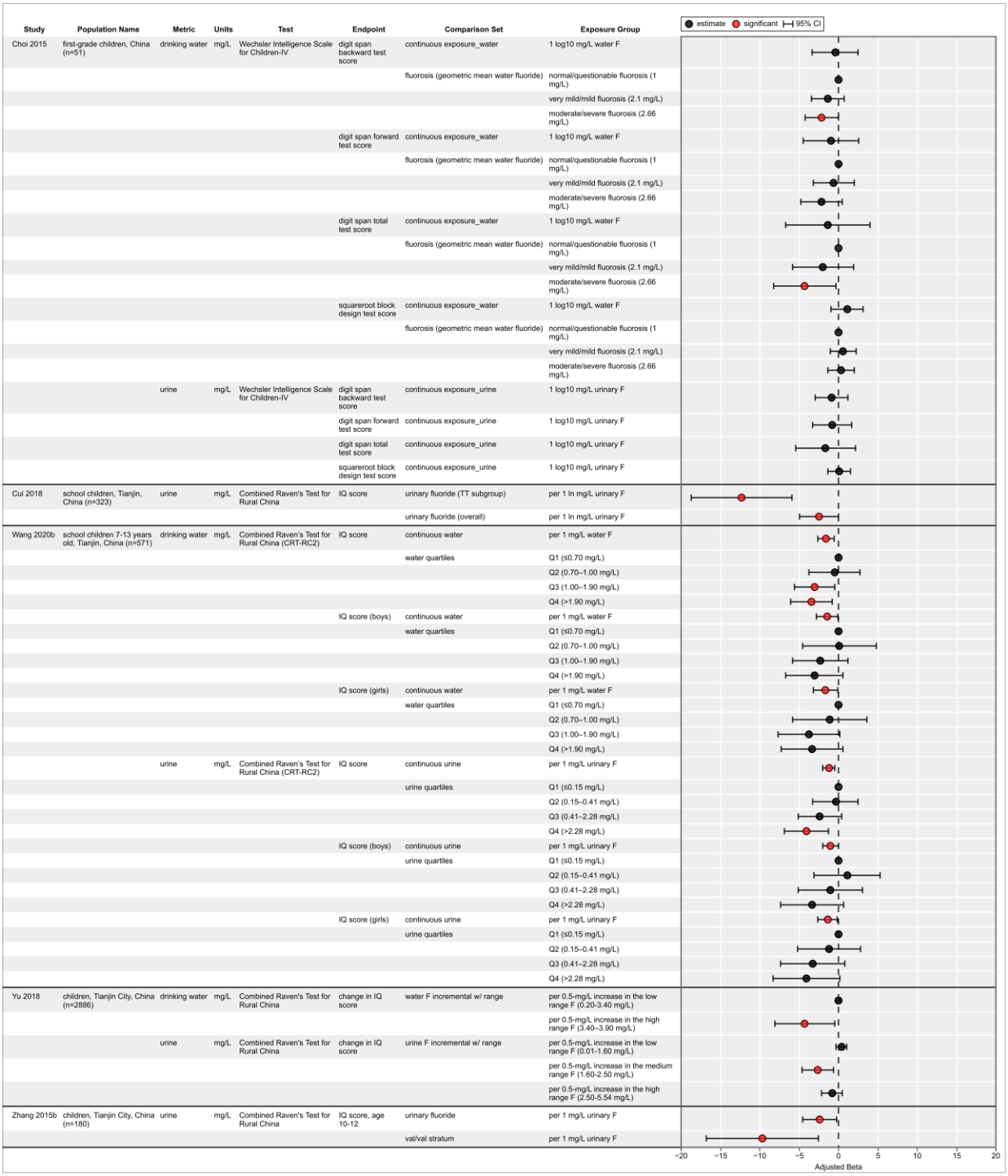


Figure A-7. Associations between Fluoride Exposure and IQ Score in Children (Low Risk-of-bias Studies; Presented as Adjusted Beta)—China

Circles represent response estimates with red indicating statistical significance.
An interactive version of Figure A-7 and additional study details in HAWC here. "F" represents fluoride. For Yu et al. (2018), authors note an obvious decrease in the IQ score at water fluoride exposure levels between 3.40 mg/L and 3.90 mg/L and a similar adverse effect on IQ scores at urinary fluoride exposure levels from 1.60 mg/L to 2.50 mg/L, and so the changes in IQ score are indicated as significant; however, significance levels for change in IQ score were not reported.

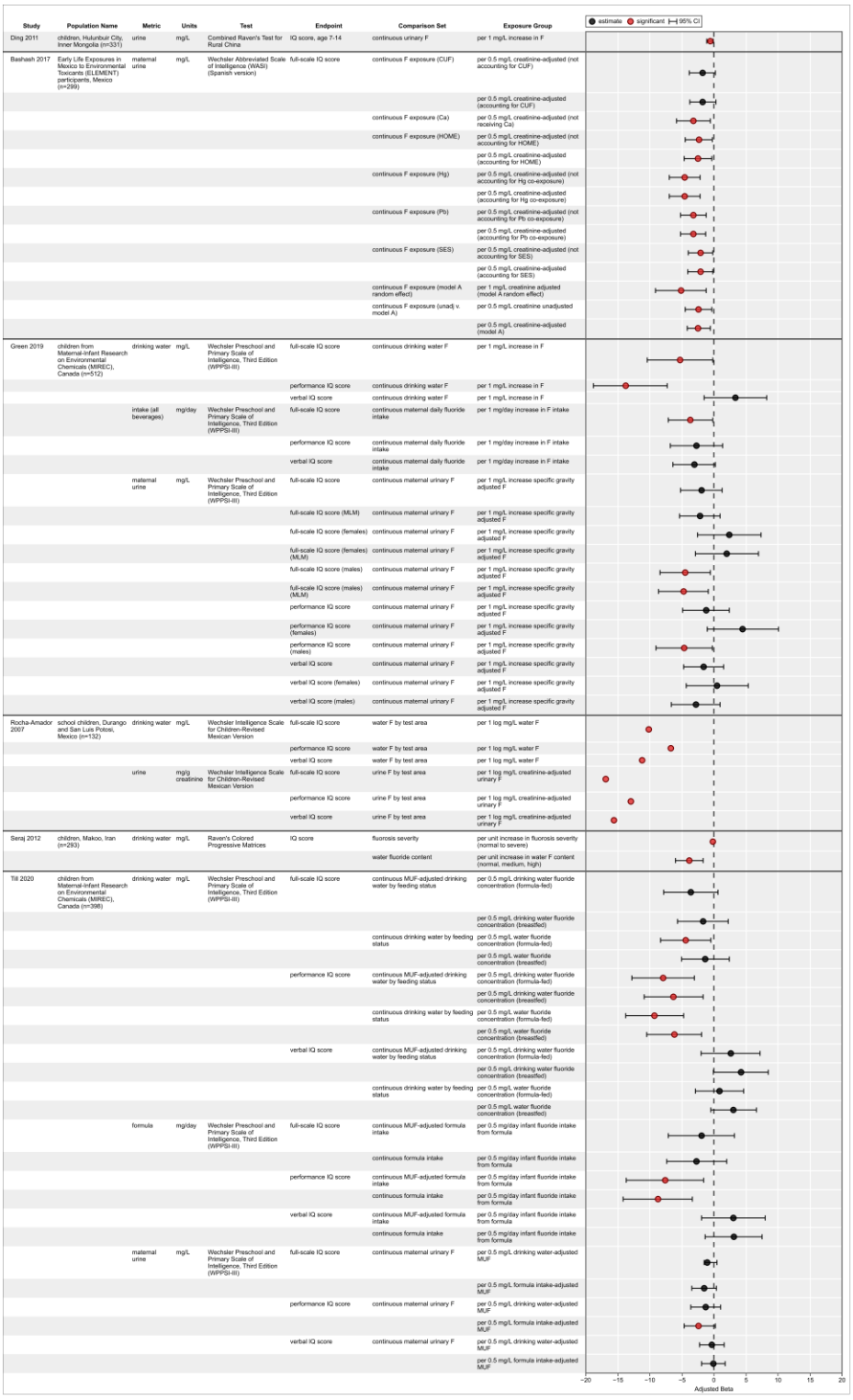


Figure A-8. Associations between Fluoride Exposure and IQ Score in Children (Low Risk-of-bias Studies; Presented as Adjusted Beta)—Areas Other Than China

Circles represent response estimates with red indicating statistical significance.
An interactive version of Figure A-8 and additional study details in HAWC here. "F" represents fluoride.

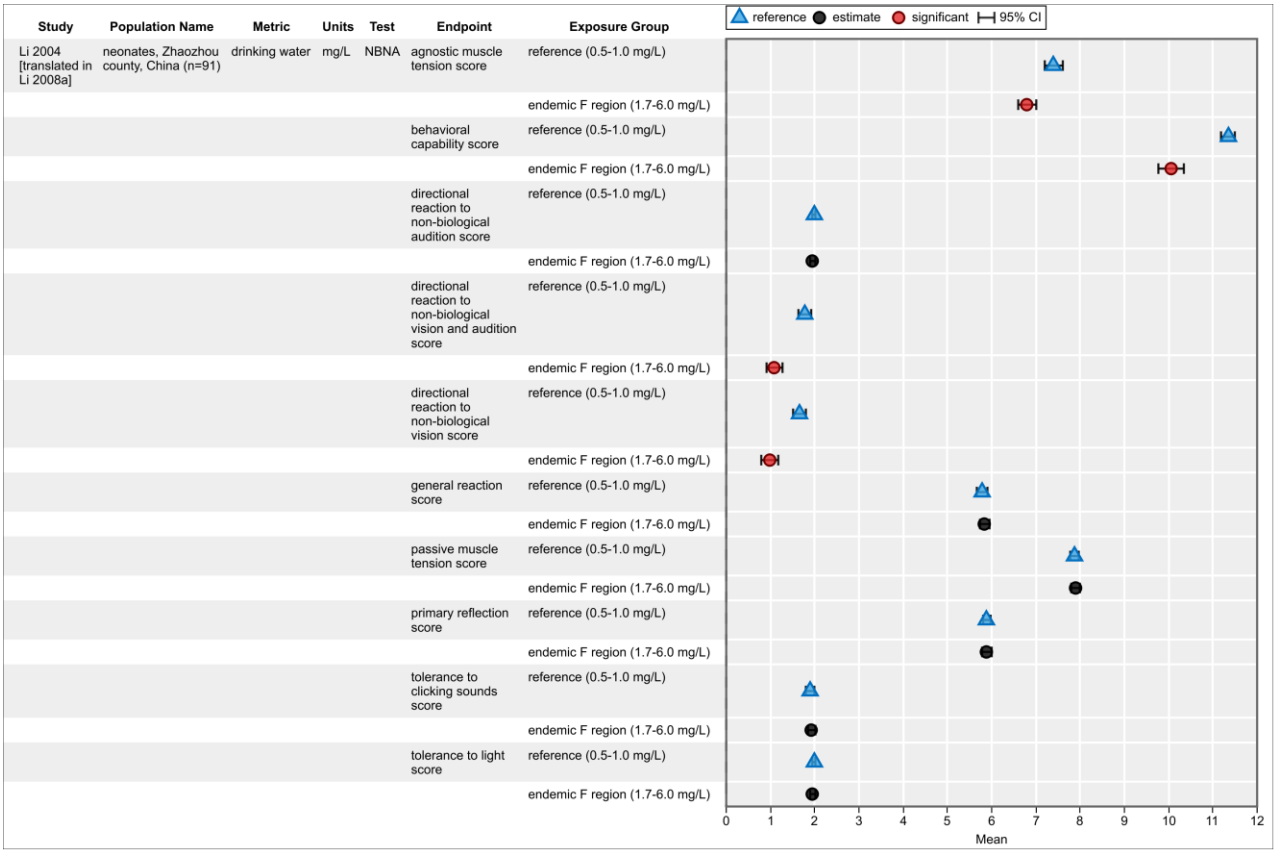


Figure A-9. Mean Motor/Sensory Scores in Children by Fluoride Exposure (Low Risk-of-bias Studies)

Reference group indicated by blue triangles; circles represent response estimates with red indicating statistical significance.
An interactive version of Figure A-9 and additional study details in HAWC here. "F" represents fluoride. 95% CIs are small and are within figure symbols and may be difficult to see. Values for SDs and 95% CIs can be viewed in HAWC by clicking the data points within the plot area. Total neonatal behavioral neurological assessment (NBNA) score was also significantly reduced in the endemic F region versus reference region (not shown).

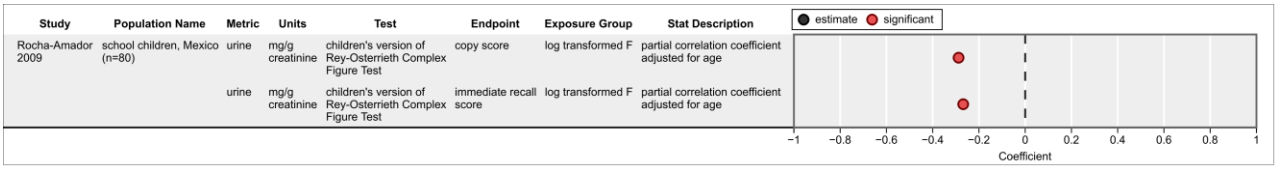


Figure A-10. Correlations between Fluoride Exposure and Other Cognitive Effects in Children (Low Risk-of-bias Studies; Presented as Correlation Coefficient)

Circles represent response estimates with red indicating statistical significance.
An interactive version of Figure A-10 and additional study details in HAWC here. "F" represents fluoride.

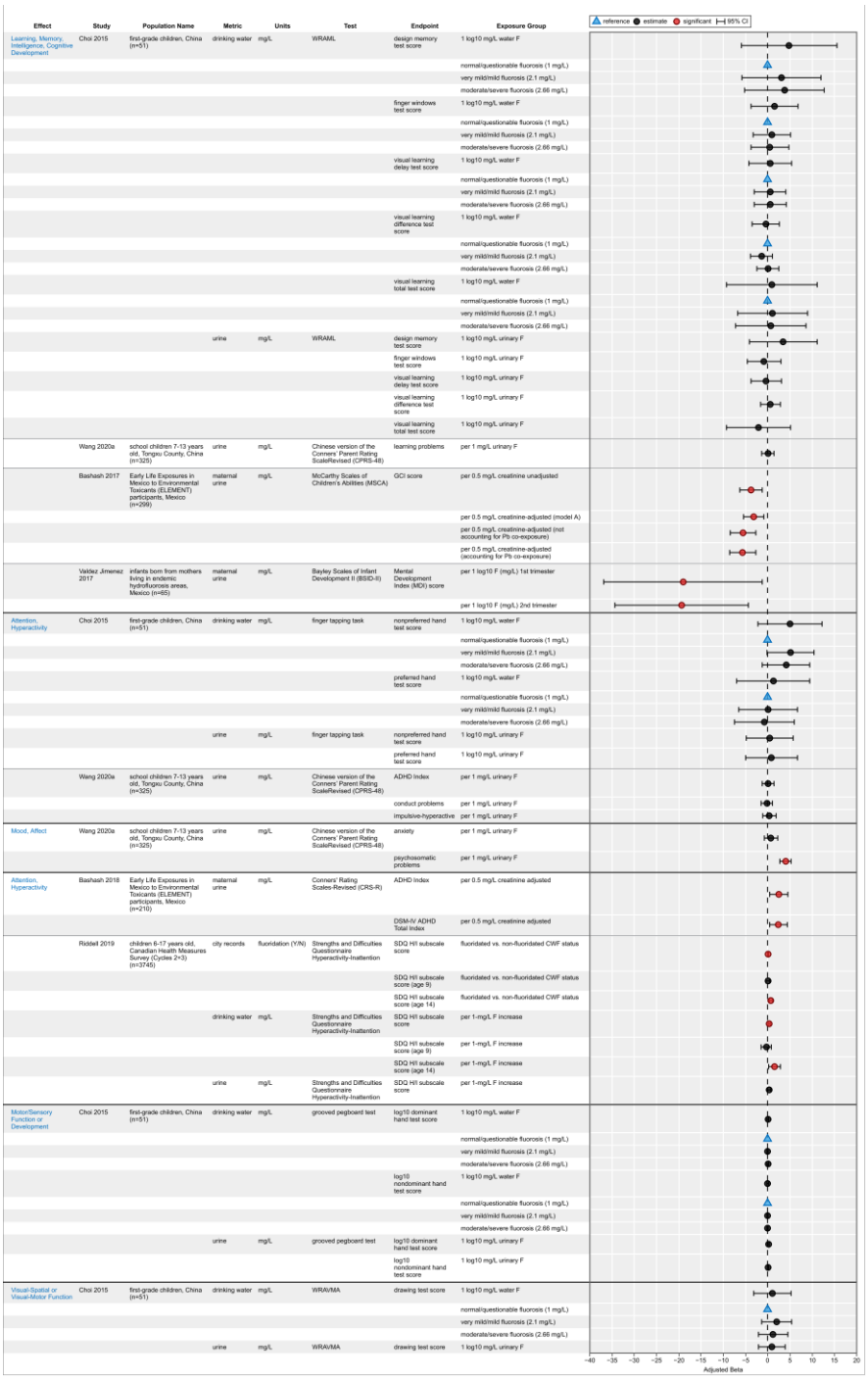


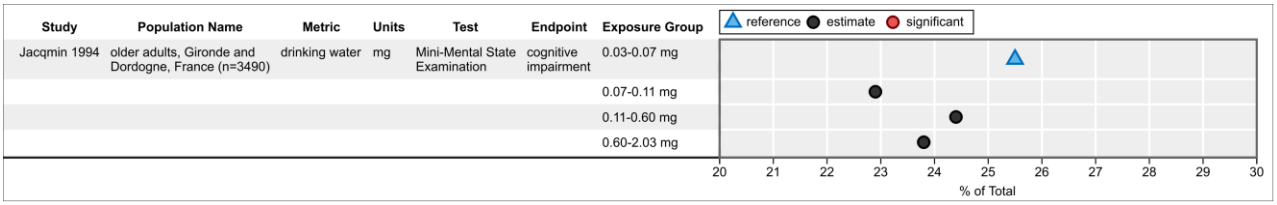
Figure A-11. Associations between Fluoride Exposure and Other Neurodevelopmental Effects in Children (Low Risk-of-bias Studies; Presented as Adjusted Beta)

Reference group indicated by blue triangles; circles represent response estimates with red indicating statistical significance. An interactive version of Figure A-11 and additional study details in HAWC here. "F" represents fluoride. Bashash et al. (2018) observed significant associations between maternal urinary fluoride and ADHD-like symptoms related to inattention (an increase in 0.5 mg/L of maternal urinary fluoride was associated with a 2.84-point increase in the DSM-IV Inattention Index and a 2.54-point increase in Cognitive Problems and Inattention Index). These two scales contributed to the global ADHD Index and the DSM-IV ADHD Total Index shown here.

Effect	Study	Population Name	Metric	Units	Test	Endpoint	Comparison Set	Exposure Group
Learning, Memory, Intelligence, Cognitive Development	Wang 2020a	school children 7-13 years old, Tongxu County, China (n=325)	urine	mg/L	Chinese version of the Conners' Parent Rating ScaleRevised (CPRS-48)	learning problems	continuous urinary F	per 1 mg/L urinary F
	Barberio 2017b	children 3-12 years old, Canadian Health Measures Survey (Cycle 2) (n=1120)	urine	umol/L		learning disability	Cycle 2 urine	per 1-umol urinary F
							Cycle 2 SG-adjusted urine	per 1-umol urinary F
				umol/mmol creatinine		learning disability	Cycle 2 creatinine-adjusted urine	per 1-umol urinary F
		children 3-12 years old, Canadian Health Measures Survey (Cycle 3) (n=1101)	drinking water	mg/L		learning disability	Cycle 3 drinking water	per 1 mg/L water F
			urine	umol/L		learning disability	Cycle 3 urine	per 1-umol urinary F
							Cycle 3 SG-adjusted urine	per 1-umol urinary F
				umol/mmol creatinine		learning disability	Cycle 3 creatinine-adjusted urine	per 1-umol urinary F
		children 3-12 years old, Canadian Health Measures Survey (Cycles 2+3) (n=2221)	urine	umol/L		learning disability	pooled urine	per 1-umol urinary F
							pooled SG-adjusted urine	per 1-umol urinary F
				umol/mmol creatinine		learning disability	pooled creatinine-adjusted urine	per 1-umol urinary F
Attention, Hyperactivity	Wang 2020a	school children 7-13 years old, Tongxu County, China (n=325)	urine	mg/L	Chinese version of the Conners' Parent Rating ScaleRevised (CPRS-48)	ADHD Index	continuous urinary F	per 1 mg/L urinary F
						conduct problems	continuous urinary F	per 1 mg/L urinary F
						impulsive-hyperactive	continuous urinary F	per 1 mg/L urinary F
	Barberio 2017b	children 3-12 years old, Canadian Health Measures Survey (Cycle 2) (n=1120)	urine	umol/L		ADHD	Cycle 2 urine	per 1-umol urinary F
							Cycle 2 SG-adjusted urine	per 1-umol urinary F
				umol/mmol creatinine		ADHD	Cycle 2 creatinine-adjusted urine	per 1-umol urinary F
	Riddell 2019	children 6-17 years old, Canadian Health Measures Survey (Cycles 2+3) (n=3745)	city records	fluoridation (Y/N)		ADHD	Cycle 2/3 CWF	fluoridated vs. non-fluoridated CWF status
						ADHD (age 9)	Cycle 2/3 CWF	fluoridated vs. non-fluoridated CWF status
						ADHD (age 14)	Cycle 2/3 CWF	fluoridated vs. non-fluoridated CWF status
			drinking water	mg/L		ADHD	Cycle 3 drinking water	per 1-mg/L F increase
			urine	mg/L		ADHD	Cycle 2/3 SG-adjusted urine	per 1-mg/L F increase
Mood, Affect	Wang 2020a	school children 7-13 years old, Tongxu County, China (n=325)	urine	mg/L	Chinese version of the Conners' Parent Rating ScaleRevised (CPRS-48)	anxiety	continuous urinary F	per 1 mg/L urinary F
						psychosomatic problems	continuous urinary F	per 1 mg/L urinary F

Figure A-12. Associations between Fluoride Exposure and Other Neurodevelopmental Effects in Children (Low Risk-of-bias Studies; Presented as Adjusted OR)

Circles represent response estimates with red indicating statistical significance.
An interactive version of Figure A-12 and additional study details in HAWC here. "F" represents fluoride. Drinking water results for Barberio et al. (2017b) have a large confidence interval and are not completely visible in the figure. 95% CIs are 0.068–11.33 and can be viewed in HAWC by clicking the OR within the plot area.



Study	Population Name	Metric	Units	Test	Endpoint	Exposure Group
Jacqmin 1994	older adults, Gironde and Dordogne, France (n=3490)	drinking water	mg	Mini-Mental State Examination	cognitive impairment	0.03-0.07 mg
						0.07-0.11 mg
						0.11-0.60 mg
						0.60-2.03 mg

Figure A-13. Cognitive Impairment in Adults by Fluoride Exposure (Low Risk-of-bias Studies; Presented as % of Total Group)

Reference group indicated by blue triangles; circles represent response estimates with red indicating statistical significance.
An interactive version of Figure A-13 and additional study details in HAWC here. Results from Li et al. (2016) suggested that fluoride exposure may be a risk factor for cognitive impairment in elderly subjects; however, results from the study were not conducive to presentation in this visualization.

Appendix B. Literature Search and Document Review Details

Table of Contents

Tables

B.1. Introduction

NTP initially published a systematic review of the experimental animal literature in 2016 that was subsequently expanded to include human epidemiological studies, mechanistic studies, and newer experimental animal literature. Table B-1 provides a timeline of key activities contributing to the 2022 NTP monograph including the multiple literature searches, draft monographs, and document review activities that have occurred since 2016.

Table B-2 is a summary of the specific search terms used for the PubMed database. In order to ensure inclusion of relevant papers, the strategy for this search was broad for the consideration of neurodevelopmental or cognitive endpoints and comprehensive for fluoride as an exposure or treatment. The specific search strategies for other databases are available in the protocol (https://ntp.niehs.nih.gov/go/785076).

Table B-1. Literature Search and Document Review Timeline

Date	Action
July 2016	Published 2016 NTP monograph of the systematic literature review on the effects of fluoride on learning and memory in animals only
June 2017	Published protocol for a new NTP monograph on systematic review on effects of fluoride on neurodevelopment and cognition from evidence in human, experimental animal, and mechanistic data
April 2019	Completed final literature search for 2019 draft NTP monograph on human, experimental animal, and mechanistic data (i.e., updated through April 2019)
May 2019	Published 2019 revised protocol for 2019 draft NTP monograph
September 2019	Sent 2019 draft NTP monograph for review by NASEM committee
February 2020	Received NASEM committee's review report of 2019 draft NTP monograph; began the following key changes in response to NASEM report: • Expanded literature search to non-English-language databases • Conducted meta-analysis on children's IQ and fluoride exposure • Revised protocol for monograph to include additional information.
May 2020	Completed final literature search for 2020 draft NTP monograph on human experimental animal and mechanistic data (i.e., updated through May 2020 and expanded to include non-English-language databases)
September 2020	Published 2020 revised protocol for 2020 draft NTP monograph
September 2020	Sent 2020 draft NTP monograph for second review by NASEM committee
February 2021	Received NASEM committee's review report of revised 2020 draft NTP monograph; made the following key changes in response to NASEM report: • Removed hazard step and hazard conclusions • Removed meta-analysis to publish separately.
December 2021	Sent 2021 draft NTP monograph on the state of the science for external peer review
April 2022	Published final 2022 NTP monograph on the state of the science

Table B-2. PubMed Search Terms

Database	Search Terms
PUBMED	((Fluorides[mh:noexp] OR fluorides, topical[mh] OR sodium fluoride[mh] OR Fluorosis, Dental[mh] OR fluorosis[tiab] OR fluorid*[tiab] OR flurid*[tiab] OR fluorin*[tiab] OR florin*[tiab]) NOT (18F[tiab] OR f-18[tiab] OR 19F[tiab] OR f-19[tiab] OR f-labeled[tiab] OR "fluorine-18"[tiab] OR "fluorine-19"[tiab] OR pet-scan[tiab] OR radioligand*[tiab])) AND ((Aryl Hydrocarbon Hydroxylases[mh] OR Aryl Hydrocarbon Receptor Nuclear Translocator[mh] OR Behavior and Behavior Mechanisms[mh] OR Gene Expression Regulation[mh] OR Glucuronosyltransferase[mh] OR Intelligence tests[mh] OR Malate Dehydrogenase[mh] OR Mediator Complex Subunit 1[mh] OR Mental disorders[mh] OR Mental processes[mh] OR Monocarboxylic Acid Transporters[mh] OR Myelin Basic Protein[mh] OR nervous system[mh] OR nervous system diseases[mh] OR nervous system physiological phenomena[mh] OR Neurogranin[mh] OR Oligodendroglia[mh] OR Peroxisome Proliferator-Activated Receptors[mh] OR Psychological Phenomena and Processes[mh] OR Receptors, thyroid hormone[mh] OR Receptors, thyrotropin[mh] OR Retinoid X Receptors[mh] OR thyroid diseases[mh] OR thyroid hormones[mh] OR Thyrotropin-releasing hormone[mh] OR Thyroxine-Binding Proteins[mh] OR Pregnane X Receptor[supplementary concept] OR thyroid-hormone-receptor interacting protein[supplementary concept] OR Constitutive androstane receptor[supplementary concept] OR Academic performance[tiab] OR auditory[tiab] OR cortical[tiab] OR delayed development[tiab] OR developmental impairment[tiab] OR developmental-delay*[tiab] OR developmental-disorder*[tiab] OR euthyroid[tiab] OR gait[tiab] OR glia*[tiab] OR gliogenesis[tiab] OR hyperactiv*[tiab] OR impulse-control[tiab] OR iodide peroxidase[tiab] OR IQ[tiab] OR ischemi*[tiab] OR locomotor[tiab] OR mental deficiency[tiab] OR mental development[tiab] OR mental illness[tiab] OR mental-deficit[tiab] OR mobility[tiab] OR mood[tiab] OR morris-maze[tiab] OR morris-water[tiab] OR motor abilit*[tiab] OR Motor activities[tiab] OR motor performance[tiab] OR nerve[tiab] OR neural[tiab] OR neurobehav*[tiab] OR Neurocognitive impairment[tiab] OR neurodegenerat*[tiab] OR Neurodevelopment*[tiab] OR neurodisease*[tiab] OR neurologic*[tiab] OR neuromuscular[tiab] OR neuron*[tiab] OR neuropath*[tiab] OR obsessive compulsive[tiab] OR OCD[tiab] OR olfaction[tiab] OR olfactory[tiab] OR open-field-test[tiab] OR passive avoidance[tiab] OR plasticity[tiab] OR senil*[tiab] OR sociab*[tiab] OR speech*[tiab] OR spelling[tiab] OR stereotypic-movement*[tiab] OR synap*[tiab] OR tauopath*[tiab] OR Thyroglobulin[tiab] OR Thyroid disease*[tiab] OR thyroid gland[tiab] OR thyroid hormone*[tiab] OR thyronine*[tiab] OR visual motor[tiab] OR Visuospatial processing[tiab] OR water maze[tiab]) OR ((active-avoidance[tiab] OR ADHD[tiab] OR alzheimer*[tiab] OR amygdala[tiab] OR antisocial[tiab] OR anxiety[tiab] OR anxious[tiab] OR asperger*[tiab] OR attention deficit[tiab] OR autism[tiab] OR autistic[tiab] OR behavioral[tiab] OR behaviors[tiab] OR behavioural[tiab] OR behaviours[tiab] OR bipolar[tiab] OR cerebellum[tiab] OR cognition[tiab] OR cognitive[tiab] OR communication-disorder*[tiab] OR comprehension[tiab] OR cranial[tiab] OR dementia[tiab] OR dendrit*[tiab] OR dentate-gyrus[tiab] OR depression[tiab] OR dextrothyroxine[tiab] OR diiodothyronine*[tiab] OR diiodotyrosine[tiab] OR down syndrome[tiab] OR dyslexia[tiab] OR entorhinal cortex[tiab] OR epilep*[tiab] OR gangli*[tiab] OR goiter[tiab] OR graves-disease[tiab] OR hearing[tiab] OR hippocamp*[tiab] OR human development[tiab] OR hyperthyroid*[tiab] OR hypothalam*[tiab] OR hypothyroid*[tiab] OR impulsiv*[tiab] OR Intellectual disability[tiab] OR intelligence[tiab] OR language[tiab] OR learning[tiab] OR lewy bod*[tiab] OR long-term potentiation[tiab] OR long-term synaptic depression[tiab] OR memory[tiab] OR mental disorder*[tiab] OR mental recall[tiab] OR monoiodotyrosine[tiab] OR Motor activity[tiab] OR motor skill*[tiab] OR multiple sclerosis[tiab] OR myxedema[tiab] OR Nervous system[tiab] OR nervous-system[tiab] OR neurit*[tiab] OR optic[tiab] OR palsy[tiab] OR panic[tiab] OR parahippocamp*[tiab] OR paranoia[tiab] OR paranoid[tiab] OR parkinson*[tiab] OR perception[tiab] OR perforant*[tiab] OR personality[tiab] OR phobia[tiab] OR problem solving[tiab] OR proprioception[tiab] OR psychomotor[tiab] OR reflex[tiab] OR risk taking[tiab] OR schizophrenia[tiab] OR seizure*[tiab] OR sensation*[tiab] OR sleep[tiab] OR smell[tiab] OR spatial behavior[tiab] OR stroke[tiab] OR substantia-nigra[tiab] OR taste[tiab] OR thyroiditis[tiab] OR thyrotoxicosis[tiab] OR Thyrotropin[tiab] OR thyroxine[tiab] OR triiodothyronine[tiab] OR vision[tiab]) NOT medline[sb]))

Appendix C. Detailed Literature Search Results and List of Included Studies

Table of Contents

C.1. Detailed Literature Search Results

C.1.1. Literature Search Results Counts and Title and Abstract Screening

The electronic database searches retrieved 25,450 unique references in total (20,883 references during the initial search conducted in December 2016, 3,657 references during the literature search updates [including the final updated search conducted for the primary epidemiological studies on May 1, 2020], and 910 references from the supplemental Chinese database searches); 11 additional references were identified by technical advisors or from reviewing reference lists in published reviews and included studies. As a result of title and abstract screening, 1,036 references were moved to full-text review, and 24,425 references were excluded (11,402 by manual screening for not satisfying the PECO criteria and 13,023 based on the SWIFT-Active Screener algorithm).

C.1.2. Full-text Review

Among the 1,036 references that underwent full-text review, 489 were excluded at that stage with reasons for exclusion documented; 333 references were excluded for not satisfying the PECO criteria; and 156 references from the May 2020 searches (main literature search update and supplemental Chinese database searches) were excluded for not including information that would materially advance the human, animal in vivo, or mechanistic findings (see the Main Literature Search section for a description of the methodology). These screening results are outlined in a study selection diagram that reports numbers of studies excluded for each reason at the full-text review stage (see Figure 2) [using reporting practices outlined in Moher et al. (2009)]. After full-text review, 547 studies were considered relevant with primary neurodevelopmental or cognitive outcomes, secondary neurobehavioral outcomes, and/or outcomes related to thyroid function. A few studies assessed data for more than one evidence stream (human, non-human mammal, and/or in vitro), and several human and animal studies assessed more than one type of outcome (e.g., primary and secondary outcomes). The number of included studies is summarized below:

- 167 human studies (84 primary only; 13 secondary only; 5 primary and secondary; 8 primary and thyroid; 2 secondary and thyroid; and 55 thyroid only);
- 339 non-human mammal studies (7 primary only; 186 secondary only; 67 primary and secondary; 6 primary, secondary, and thyroid; 4 secondary and thyroid; and 69 thyroid only); and,
- 60 in vitro/mechanistic studies (48 neurological and 12 thyroid).

One publication contained human, experimental non-human mammal, and in vitro data. Three publications contained both human and experimental non-human mammal data. Fourteen publications contained data relevant to both experimental non-human mammal studies and in vitro studies.

C.2. List of Included Studies

C.2.1. Studies in Humans

As described in Figure 2, 167 human studies were included; however, full data extraction was conducted only on studies with neurological outcomes or thyroid hormone data. Data extraction was completed using HAWC. Data were extracted from a subset of included studies in humans (n = 124) and are available in HAWC based on outcome. The following lists of references are organized as studies that are available in HAWC followed by studies that are not available in HAWC. Specifically, data for primary neurodevelopmental or cognitive outcomes (learning, memory, and intelligence) and secondary neurobehavioral outcomes (anxiety, aggression, motor activity, or biochemical changes), as well as thyroid hormone level data, were extracted from included human studies and are available in HAWC. Data for included studies identified through the 2020 literature search update were extracted only for primary neurodevelopmental or cognitive outcomes; a subset of these studies (n = 7) also included secondary neurobehavioral outcomes and/or thyroid hormone level data that were not extracted because those data would not materially advance the human or mechanistic findings. Included human studies that evaluated only other thyroid-related effects such as goiters or thyroid size (n = 43) were not extracted and are not available in HAWC. The list below presents the 167 human studies that were included in the review. An overview of the screening results is outlined in the study selection diagram (Figure 2) that reports numbers of included studies as well as numbers of studies excluded for each reason at the full-text review stage.

C.2.1.1. Studies Available in HAWC

An J, Mei S, Liu A, Fu Y, Wang C. 1992. [Effect of high level of fluoride on children's intelligence]. *Chin J Control Endem Dis* 7(2): 93-94.

Aravind A, Dhanya RS, Narayan A, Sam G, Adarsh VJ, Kiran M. 2016. Effect of fluoridated water on intelligence in 10-12-year-old school children. *J Int Soc Prev Community Dent* 6(Suppl 3): S237-S242.

Bai A, Li Y, Fan Z, Li X, Li P. 2014. [Intelligence and growth development of children in coal-burning-borne arsenism and fluorosis areas: An investigation study]. *Chin J Endemiol* 33(2): 160-163.

Barberio AM, Hosein FS, Quinonez C, McLaren L. 2017. Fluoride exposure and indicators of thyroid functioning in the Canadian population: Implications for community water fluoridation. *J Epidemiol Community Health* 71: 1019-1025.

Barberio AM, Quinonez C, Hosein FS, McLaren L. 2017. Fluoride exposure and reported learning disability diagnosis among Canadian children: Implications for community water fluoridation. *Can J Public Health* 108: 229-239.

Bashash M, Thomas D, Hu H, Martinez-Mier EA, Sanchez BN, Basu N, Peterson KE, Ettinger AS, Wright R, Zhang Z, Liu Y, Schnaas L, Mercado-Garcia A, Tellez-Rojo MM, Hernandez-Avila M. 2017. Prenatal fluoride exposure and cognitive outcomes in children at 4 and 6-12 years of age in Mexico. *Environ Health Perspect* 125(9): 1-12.

Bashash M, Marchand M, Hu H, Till C, Martinez-Mier EA, Sanchez BN, Basu N, Peterson KE, Green R, Schnaas L, Mercado-Garcia A, Hernandez-Avila M, Tellez-Rojo MM. 2018. Prenatal fluoride exposure and attention deficit hyperactivity disorder (ADHD) symptoms in children at 6-12 years of age in Mexico City. *Environ Int* 121(Pt 1): 658-666.

Broadbent JM, Thomson WM, Moffitt TE, Poulton R. 2015. Community water fluoridation and intelligence response. *Am J Public Health* 105: 3-4.

Chen YX, Han FL, Zhoua ZL, Zhang HQ, Jiao XS, Zhang SC, Huang MC, Chang TQ, Dong YF. 1991. [Research on the intellectual development of children in high fluoride areas]. *Chin J Control Endem Dis* 6(Suppl): 99-100.

Chen YX, Han FL, Zhoua ZL, Zhang HQ, Jiao XS, Zhang SC, Huang MC, Chang TQ, Dong YF. 2008. Research on the intellectual development of children in high fluoride areas. *Fluoride* 41: 120-124.

Chinoy NJ, Narayana MV. 1992. Studies on fluorosis in Mehsana District of North Gujarat. *Proc Zool Soc* 45: 157-161.

Choi AL, Zhang Y, Sun G, Bellinger DC, Wang K, Yang XJ, Li JS, Zheng Q, Fu Y, Grandjean P. 2015. Association of lifetime exposure to fluoride and cognitive functions in Chinese children: A pilot study. *Neurotoxicol Teratol* 47: 96-101.

Cui Y, Zhang B, Ma J, Wang Y, Zhao L, Hou C, Yu J, Zhao Y, Zhang Z, Nie J, Gao T, Zhou G, Liu H. 2018. Dopamine receptor D2 gene polymorphism, urine fluoride, and intelligence impairment of children in China: A school-based cross-sectional study. *Ecotoxicol Environ Saf* 165: 270-277.

Cui Y, Yu J, Zhang B, Guo B, Gao T, Liu H. 2020. The relationships between thyroid-stimulating hormone and/or dopamine levels in peripheral blood and IQ in children with different urinary iodine concentrations. *Neurosci Lett* 729: 134981.

Das K, Mondal NK. 2016. Dental fluorosis and urinary fluoride concentration as a reflection of fluoride exposure and its impact on IQ level and BMI of children of Laxmisagar, Simlapal Block of Bankura District, W.B., India. *Environ Monit Assess* 188: 218.

Ding Y, Sun H, Han H, Wang W, Ji X, Liu X, Sun D. 2011. The relationships between low levels of urine fluoride on children's intelligence, dental fluorosis in endemic fluorosis areas in Hulunbuir, Inner Mongolia, China. *J Hazard Mater* 186: 1942-1946.

Du L, Wan C, Cao X, Liu J. 1992. [The effect of fluorine on the developing human brain]. *Chin J Pathol* 21(4): 218-220.

Du L, Wan C, Cao X, Liu J. 2008. The effect of fluorine on the developing human brain. *Fluoride* 41: 327-330.

Duan J, Zhao M, Wang L, Fang D, Wang Y, Wang W. 1995. A comparative analysis of the results of multiple tests in patients with chronic industrial fluorosis. Guizhou Med J 18(3): 179-180. Erickson JD, Hay S. 1976. Water fluoridation and congenital malformations: No association. *J Am Dent Assoc* 93: 981-984.

Erickson JD, Hay S. 1976. Water fluoridation and congenital malformations: No association. *J Am Dent Assoc* 93: 981-984.

Erickson JD. 1980. Down syndrome, water fluoridation, and maternal age. *Teratology* 21: 177-180.

Eswar P, Nagesh L, Devaraj CG. 2011. Intelligent quotients of 12-14 year old school children in a high and low fluoride village in India. *Fluoride* 44: 168-172.

Fan Z, Dai H, Bai A, Li P, Li T, Li G. 2007. Effect of high fluoride exposure in children's intelligence. *J Environ Health* 24(10): 802-803.

Green R, Lanphear B, Hornung R, Flora D, Martinez-Mier EA, Neufeld R, Ayotte P, Muckle G, Till C. 2019. Association between maternal fluoride exposure during pregnancy and IQ scores in offspring in Canada. *JAMA Pediatr*: E1-E9.

Guo XC, Wang RY, Cheng CF, Wei WS, Tang LM, Wang QS, Tang DX, Liu GW, He GD, Li SL. 1991. [A preliminary investigation of the IQs of 7-13 year-old children from an area with coal burning-related fluoride poisoning]. *Chin J Epidemiol* 10(2): 98-100.

Guo XC, Wang RY, Cheng CF, Wei WS, Tang LM, Wang QS, Tang DX, Liu GW, He GD, Li SL. 2008. A preliminary investigation of the IQs of 7-13 year-old children from an area with coal burning-related fluoride poisoning. *Fluoride* 41: 125-128.

Guo ZY, He YH, Zhu QX. 2001. [Research on the neurobehavioral function of workers occupationally exposed to fluoride]. *Ind Hlth & Occup Dis* 27(6): 346-348.

Guo ZY, He YH, Zhu QX. 2008. Research on the neurobehavioral function of workers occupationally exposed to fluoride. *Fluoride* 41: 152-155.

He H, Cheng ZS, Liu WQ. 1989. [Effects of fluorine on the human fetus]. *J Control Endem Dis* 4(3): 136-138.

He H, Cheng ZS, Liu WQ. 2008. Effects of fluorine on the human fetus. *Fluoride* 41: 321-326.

He MX, Zhang CN. 2010. [Investigation of children's intelligence quotient and dental fluorosis in drinking water-type of endemic fluorosis area in Pucheng County, Shaanxi Province before and after drinking water change]. *Chin J Endemiol* 29: 547-548.

Hong F, Wang H, Yang D, Zhang Z. 2001. [Investigation on the intelligence and metabolism of iodine and fluoride in children with high iodine and fluoride]. *Chin J Control Endem Dis* 12-14.

Hong FG, Cao YX, Yang D, Wang H. 2001. [Research on the effects of fluoride on child intellectual development under different environmental conditions]. *Chin Prim Health Care* 15(3): 56-57.

Hong FG, Cao YX, Yang D, Wang H. 2008. Research on the effects of fluoride on child intellectual development under different environmental conditions. *Fluoride* 41: 156-160.

Hosur MB, Puranik RS, Vanaki S, Puranik SR. 2012. Study of thyroid hormones free triiodothyronine (FT3), free thyroxine (FT4) and thyroid stimulating hormone (TSH) in subjects with dental fluorosis. *Eur J Dent* 6: 184-190.

Jacqmin H, Commenges D, Letenneur L, Barberger-Gateau P, Dartigues JF. 1994. Components of drinking water and risk of cognitive impairment in the elderly. *Am J Epidemiol* 139: 48-57.

Jin T, Han T, Wei Y, Wu Y, Wang Z, Zhang, H. 2016. [Investigation on working memory level of children aged 8-12 years in coal-burning fluorosis area]. *J Environ Health* 33(5): 409-411.

Jin T, Wang Z, Wei Y, Wu Y, Han T, Zhang H. 2017. [Investigation on intelligence level of children aged 8-12 years old in coal-burning fluorosis area]. *J Environ Health* 34(3): 229-231.

Kang J, Cheng Y, Wu K, Lin S, He G, Jin Y. 2011. Effect of exposure to fluoride and arsenic in drinking water of Hangjinhouqi on children's intelligence. *Chinese School Health*: 679-681.

Karimzade S, Aghaei M, Mahvi AH. 2014. Investigation of intelligence quotient in 9-12 year-old children exposed to high- and low-drinking water fluoride in West Azerbaijan Province, Iran. *Fluoride* 47: 9-14.

Khan SA, Singh RK, Navit S, Chadha D, Johri N, Navit P, Sharma A, Bahuguna R. 2015. Relationship between dental fluorosis and intelligence quotient of school going children in and around Lucknow District: A cross-sectional study. *J Clin Diagn Res* 9(11): 10-15.

Khandare AL, Gourineni SR, Validandi V. 2017. Dental fluorosis, nutritional status, kidney damage, and thyroid function along with bone metabolic indicators in school-going children living in fluoride-affected hilly areas of Doda District, Jammu and Kashmir, India. *Environ Monit Assess* 189: 579.

Khandare AL, Validandi V, Gourineni SR, Gopalan V, Nagalla B. 2018. Dose-dependent effect of fluoride on clinical and subclinical indices of fluorosis in school going children and its mitigation by supply of safe drinking water for 5 years: An Indian study. *Environ Monit Assess* 190: 110.

Kheradpisheh Z, Mahvi AH, Mirzaei M, Mokhtari M, Azizi R, Fallahzadeh H, Ehrampoush MH. 2018. Correlation between drinking water fluoride and TSH hormone by ANNs and ANFIS. *J Environ Health Sci Eng* 16(1): 11-18.

Kheradpisheh Z, Mirzaei M, Mahvi AH, Mokhtari M, Azizi R, Fallahzadeh H, Ehrampoush MH. 2018. Impact of drinking water fluoride on human thyroid hormones: A case-control study. *Sci Rep* 8: 2674.

Kumar V, Chahar P, Kajjari S, Rahman F, Bansal DK, Kapadia JM. 2018. Fluoride, thyroid hormone derangements and its correlation with tooth eruption pattern among the pediatric population from endemic and non-endemic fluorosis areas. *J Contemp Dent Pract* 19(12): 1512-1516.

Kundu H, Basavaraj P, Singla A, Gupta R, Singh K, Jain S. 2015. Effect of fluoride in drinking water on children's intelligence in high and low fluoride areas of Delhi. *J Indian Assoc Public Health Dent* 13(2): 116-121.

Lamberg M, Hausen H, Vartiainen T. 1997. Symptoms experienced during periods of actual and supposed water fluoridation. *Community Dent Oral Epidemiol* 25: 291-295.

Li F, Chen X, Huang R, Xie Y. 2009. The impact of endemic fluorosis caused by the burning of coal on the development of intelligence in children. *J Environ Health* 26(4): 838-840.

Li J, Yao L, Shao QL, Wu CY. 2004. [Effects of high fluoride level on neonatal neurobehavioral development]. *Chin J Endemiol* 23(5): 463-465.

Li J, Yao L, Shao QL, Wu CY. 2008. Effects of high fluoride level on neonatal neurobehavioral development. *Fluoride* 41: 165-170.

Li M, Gao Y, Cui J, Li Y, Li B, Liu Y, Sun J, Liu X, Liu H, Zhao L, Sun D. 2016. Cognitive impairment and risk factors in elderly people living in fluorosis areas in China. *Biol Trace Elem Res* 172: 53-60.

Li X, Hou G, Yu B, Yuan C, Liu Y, L Z. 2010. Investigation and analysis of children's intelligence and dental fluorosis in high fluoride area. *J Med Pest Control* 26(3): 230-231.

Li XS, Zhi JL, Gao RO. 1995. Effect of fluoride exposure on the intelligence of children. *Fluoride* 28: 189-192.

Li Y, Li X, Wei S. 1994. [Effects of high fluoride intake on child mental work capacity: Preliminary investigation into the mechanisms involved]. *J West Chin University Med Sci* 25(2): 188-191.

Li Y, Li X, Wei S. 2008. Effects of high fluoride intake on child mental work capacity: Preliminary investigation into the mechanisms involved. *Fluoride* 41: 331-335.

Li YP, Jing XY, Chen D, Lin L, Wang ZJ. 2003. [Effects of endemic fluoride poisoning on the intellectual development of children in Baotou]. *Chin J Public Health Manag* 19(4): 337-338.

Li YP, Jing XY, Chen D, Lin L, Wang ZJ. 2008. Effects of endemic fluoride poisoning on the intellectual development of children in Baotou. *Fluoride* 41: 161-164.

Liang C, Ji R, Cao J, Jiang Y, Yu B, Ma F, Wu Y, Ying B, Zhang Y, Sun S, Li Y, Emsley CL, Gao S, Hall KS, Hendrie HC. 2003. [Study on the relationship between drinking water trace elements and cognitive ability of the elderly]. *Health Res* 436-440.

Lin F, Ai H, Zhao H, Lin J, Jhiang J, Maimaiti. 1991. High fluoride and low iodine environment and subclinical cretinism in Xinjiang. *Endem Dis Bull* 6(2): 62-67.

Liu S, Lu Y, Sun Z, Wu L, Wang X, Yan S. 2000. [Report on the intellectual ability of children living in high-fluoride water areas]. *Chin J Control Endem Dis* 15(4): 231-232.

Liu SL, Lu Y, Sun ZR, Wu L, Lu WL, Wang XW, Yan S. 2008. Report on the intellectual ability of children living in high-fluoride water areas. *Fluoride* 41: 144-147.

Lu F, Zhang Y, Trivedi A, Jiang X, Chandra D, Zheng J, Nakano Y, Uyghurturk DA, Jalai R, Guner S, Mentes A, DenBesten PK. 2019. Fluoride related changes in behavioral outcomes may relate to increased serotonin. *Physiol Behav* 206: 76-83.

Lu Y, Sun ZR, Wu LN, Wang X, Lu W, Liu SS. 2000. Effect of high-fluoride water on intelligence in children. *Fluoride* 33: 74-78.

Luo Y, Ma R, Liu Z, Guan Z, Lou D, Zheng D. 2018. [Intelligence investigation and forensic significance of children in coal-burning fluorosis area]. *Chin J Forensic Medicine* 33(6): 590-593.

Malin AJ, Till C. 2015. Exposure to fluoridated water and attention deficit hyperactivity disorder prevalence among children and adolescents in the United States: An ecological association. *Environ Health* 14: 17.

Malin AJ, Riddell J, McCague H, Till C. 2018. Fluoride exposure and thyroid function among adults living in Canada: Effect modification by iodine status. *Environ Int* 121(Pt 1): 667-674.

Michael M, Barot VV, Chinoy NJ. 1996. Investigations of soft tissue functions in fluorotic individuals of North Gujarat. *Fluoride* 29: 63-71.

Mondal D, Dutta G, Gupta S. 2016. Inferring the fluoride hydrogeochemistry and effect of consuming fluoride-contaminated drinking water on human health in some endemic areas of Birbhum District, West Bengal. *Environ Geochem Health* 38: 557-576.

Morgan L, Allred E, Tavares M, Bellinger D, Needleman H. 1998. Investigation of the possible associations between fluorosis, fluoride exposure, and childhood behavior problems. *Pediatr Dent* 20: 244-252.

Mustafa DE, Younis UM, Elhag SAA. 2018. The relationship between the fluoride levels in drinking water and the schooling performance of children in rural areas of Khartoum State, Sudan. *Fluoride* 51(2): 102-113.

Nagarajappa R, Pujara P, Sharda AJ, Asawa K, Tak M, Aapaliya P, Bhanushali N. 2013. Comparative assessment of intelligence quotient among children living in high and low fluoride areas of Kutch, India: A pilot study. *Iran J Public Health* 42: 813-818.

Peckham S, Lowery D, Spencer S. 2015. Are fluoride levels in drinking water associated with hypothyroidism prevalence in England? A large observational study of GP practice data and fluoride levels in drinking water. *J Epidemiol Community Health* 69: 619-624.

Poureslami HR, Horri A, Garrusi B. 2011. A comparative study of the IQ of children age 7-9 in a high and a low fluoride water city in Iran. *Fluoride* 44: 163-167.

Qin LS, Huo SY, Chen RL, Chang YZ, Zhao MY. 1990. [Using the Raven's Standard Progressive Matrices to determine the effects of the level of fluoride in drinking water on the intellectual ability of school-age children]. *J Control Endem Dis* 5(4): 203-204.

Qin LS, Huo SY, Chen RL, Chang YZ, Zhao MY. 2008. Using the Raven's Standard Progressive Matrices to determine the effects of the level of fluoride in drinking water on the intellectual ability of school-age children. *Fluoride* 41: 115-119.

Ratan SK, Rattan KN, Pandey RM, Singhal S, Kharab S, Bala M, Singh V, Jhanwar A. 2008. Evaluation of the levels of folate, vitamin B12, homocysteine and fluoride in the parents and the affected neonates with neural tube defect and their matched controls. *Pediatr Surg Int* 24: 803-808.

Razdan P, Patthi B, Kumar JK, Agnihotri N, Chaudhari P, Prasad M. 2017. Effect of fluoride concentration in drinking water on intelligence quotient of 12-14-year-old children in Mathura District: A cross-sectional study. *J Int Soc Prev Community Dent* 7: 252-258.

Ren D, Li K, Liu D. 1989. [A study of the intellectual ability of 8-14 year-old children in high fluoride, low iodine areas]. *Chin J Control Endem Dis* 4(4): 251.

Ren D, Li K, Liu D. 2008. A study of the intellectual ability of 8-14 year-old children in high fluoride, low iodine areas. *Fluoride* 41: 319-320.

Riddell JK, Malin AJ, Flora D, McCague H, Till C. 2019. Association of water fluoride and urinary fluoride concentrations with attention deficit hyperactivity disorder in Canadian youth. *Environ Int* 133: 105190.

Rocha-Amador D, Navarro ME, Carrizales L, Morales R, Calderon J. 2007. Decreased intelligence in children and exposure to fluoride and arsenic in drinking water. *Cad Saude Publica* 23 Suppl 4: S579-587.

Rocha-Amador D, Navarro M, Trejo-Acevedo A, Carrizales L, Perez-Maldonado I, Diaz-Barriga F, Calderon J. 2009. Use of the Rey-Osterrieth Complex Figure Test for neurotoxicity evaluation of mixtures in children. *Neurotox* 30: 1149-1154.

Rotton J, Tikofsky RS, Feldman HT. 1982. Behavioral effects of chemicals in drinking water. *J Appl Psychol* 67: 230-238.

Russ TC, Killin LOJ, Hannah J, Batty GD, Deary IJ, Starr JM. 2019. Aluminium and fluoride in drinking water in relation to later dementia risk. *Brit J Psychol* 14: 1-6.

Saxena S, Sahay A, Goel P. 2012. Effect of fluoride exposure on the intelligence of school children in Madhya Pradesh, India. *J Neurosci Rural Pract* 3: 144-149.

Sebastian ST, Sunitha S. 2015. A cross-sectional study to assess the intelligence quotient (IQ) of school going children aged 10-12 years in villages of Mysore District, India with different fluoride levels. *J Indian Soc Pedod Prev Dent* 33: 307-311.

Seraj B, Shahrabi M, Falahzade M, Falahzade F, Akhondi N. 2006. Effect of high fluoride concentration in drinking water on children's intelligence. *J Dent Med* 19(2): 80-86.

Seraj B, Shahrabi M, Shadfar M, Ahmadi R, Fallahzadeh M, Eslamlu HF, Kharazifard MJ. 2012. Effect of high water fluoride concentration on the intellectual development of children in Makoo, Iran. *J Dent* 9: 221-229.

Shannon FT, Fergusson DM, Horwood LJ. 1986. Exposure to fluoridated public water supplies and child health and behaviour. *New Zeal Med J* 99: 416-418.

Shao Q. 2003. Study of cognitive function impairment caused by chronic fluorosis. *Chin J Endemiol* 22(4): 336-338.

Sharma JD, Sohu D, Jain P. 2009. Prevalence of neurological manifestations in a human population exposed to fluoride in drinking water. *Fluoride* 42: 127-132.

Shivaprakash PK, Ohri K, Noorani H. 2011. Relation between dental fluorosis and intelligence quotient in school children of Bagalkot District. *J Indian Soc Pedod Prev Dent* 29: 117-120.

Singh A, Jolly SS, Devi P, Bansal BC, Singh SS. 1962. Endemic fluorosis: An epidemiological, biochemical and clinical study in the Bhatinda District of Panjab. *Indian J Med Res* 50: 387-398.

Singh N, Verma KG, Verma P, Sidhu GK, Sachdeva S. 2014. A comparative study of fluoride ingestion levels, serum thyroid hormone & TSH level derangements, dental fluorosis status among school children from endemic and non-endemic fluorosis areas. *Springerplus* 3: 7.

Singh V, Singh C, Sandeep T, Sandeep K, Vikas G, Mukesh T, Anurag T. 2013. A correlation between serum vitamin, acetylcholinesterase activity and IQ in children with excessive endemic fluoride exposure in Rajasthan, India. *Int Res J Medical Sci* 1(3): 12-16.

Soto-Barreras U, Escalante-Villalobos KY, Holguin-Loya B, Perez-Aguirre B, Nevarez-Rascon A, Martinez-Martinez RE, Loyola-Rodriguez JP. 2019. Effect of fluoride in drinking water on dental caries and IQ in children. *Fluoride* 52: 474-482.

Still CN, Kelley P. 1980. The incidence of primary degenerative dementia vs. water fluoride content in South Carolina, USA. *Neurotox* 1: 125-132.

Sudhir KM, Chandu GN, Prashant GM, Subba Reddy VV. 2009. Effect of fluoride exposure on intelligence quotient (IQ) among 13-15 year old school children of known endemic area of fluorosis, Nalgonda District, Andhra Pradesh. *J Indian Assoc Public Health Dent* 2009(13): 88-94.

Sun M, Li S, Wang Y, Li F. 1991. Measurement of intelligence by drawing test among the children in the endemic area of Al-F combined toxicosis. *J Guiyang Med Coll* 16(3): 204-206.

Susheela AK, Bhatnagar M, Vig K, Mondal NK. 2005. Excess fluoride ingestion and thyroid hormone derangements in children living in Delhi, India. *Fluoride* 38: 98-108.

Tamboli BL, Mathur GM, Mathur AP, Lalla SK, Goyal OP. 1980. Prevalence of fluorosis in Pratabpura and Surajpura villages, District Ajmer (Rajasthan). *Indian J Med Res* 71: 57-67.

Till C, Green R, Flora D, Hornung R, Martinez-Mier EA, Blazer M, Farmus L, Ayotte P, Muckle G, Lanphear B. 2020. Fluoride exposure from infant formula and child IQ in a Canadian birth cohort. *Environ Int* 134: 105315.

Tripathi P, Sultana N. 2007. Fluoride content of groundwater and prevalence of dental, skeletal and neurological stage of fluorosis in Tehsil Purwa of Unnao. *Indian J Environ Prot* 27: 737-739.

Trivedi MH, Verma RJ, Chinoy NJ, Patel RS, Sathawara NG. 2007. Effect of high fluoride water on intelligence of school children in India. *Fluoride* 40: 178-183.

Trivedi M, Sangai N, Patel R, Payak M, Vyas S. 2012. Assessment of groundwater quality with special reference to fluoride and its impact on IQ of schoolchildren in six villages of the Mundra Region, Kachchh, Gujurat, India. *Fluoride* 45(4): 377-383.

Valdez Jimenez L, Lopez Guzman OD, Cervantes Flores M, Costilla-Salazar R, Calderon Hernandez J, Alcaraz Contreras Y, Rocha-Amador DO. 2017. In utero exposure to fluoride and cognitive development delay in infants. *Neurotox* 59: 65-70.

Wang A, Duan L, Huang H, Ma J, Zhang Y, Ma Q, Guo Y, Li Z, Cheng X, Zhu J, Zhou G, Ba Y. 2020. Association between fluoride exposure and behavioural outcomes of school-age children: A pilot study in China. *Int J Environ Health Res*: 1-10.

Wang G, Yang D, Jia F, Wang H. 1996. [A study of the IQ levels of four- to seven-year-old children in high fluoride areas]. *Endem Dis Bull* 11(1): 60-62.

Wang G, Yang D, Jia F, Wang H. 2008. A study of the IQ levels of four- to seven-year-old children in high fluoride areas. *Fluoride* 41: 340-343.

Wang G, Gao M, Zhang M, Yang M, Xiang Q. 2012. Correlation between total fluoride intake and children's IQ. *J Southeast Univ Med Ed*: 743-746.

Wang G, Zhang M, Wang Q, Han A, Gao M, Lin P, Xiang Q. 2017. [Investigation on the relationship between serum fluoride content and IQ of children before and after reducing fluoride to water]. *Capital Public Health* 11(6): 274-277.

Wang M, Liu L, Li H, Li Y, Liu H, Hou C, Zeng Q, Li P, Zhao Q, Dong L, Zhou G, Yu X, Liu L, Guan Q, Zhang S, Wang A. 2020. Thyroid function, intelligence, and low-moderate fluoride exposure among Chinese school-age children. *Environ Int* 134: 105229.

Wang S, Wang L, Hu P, Guo S, Law S. 2001. [Effects of high iodine and high fluorine on children’s intelligence and thyroid function]. *Chin J Endemiol* 20(4): 288-290.

Wang SX, Wang ZH, Cheng XT, Li J, Sang ZP, Zhang XD, Han LL, Qiao XY, Wu ZM, Wang ZQ. 2005. [Investigation and evaluation on intelligence and growth of children in endemic fluorosis and arsenism areas]. *Chin J Endemiol* 24: 179-182.

Wang SX, Wang ZH, Cheng XT, Li J, Sang ZP, Zhang XD, Han LL, Qiao XY, Wu ZM, Wang ZQ. 2007. Arsenic and fluoride exposure in drinking water: Children's IQ and growth in Shanyin County, Shanxi Province, China. *Environ Health Perspect* 115: 643-647.

Wang SY, Zhang HX, Fan W, Fang SJ, Kang PP, Chen XG, Yu MJ. 2005. [The effects of endemic fluoride poisoning caused by coal burning on the physical development and intelligence of children]. *J Appl Clin Ped* 20(9): 897-899.

Wang SY, Zhang HX, Fan W, Fang SJ, Kang PP, Chen XG, Yu MJ. 2008. The effects of endemic fluoride poisoning caused by coal burning on the physical development and intelligence of children. *Fluoride* 41: 344-348.

Wang Z, Wang S, Zhang X, Li J, Zheng X, Hu C. 2006. Investigation of children’s growth and development under long-term fluoride exposure. *Chin J Control Endem Dis* 21(4): 239-241.

Wei N, Li Y, Deng J, Xu S, Guan Z. 2014. [The effects of comprehensive control measures on intelligence of school-age children in coal-burning-borne endemic fluorosis areas]. *Chin J Endemiol* 33(2): 320-322.

Xiang Q, Liang Y, Chen L, Wang C, Chen B, Chen X, Zhou M. 2003. Effect of fluoride in drinking water on children's intelligence. *Fluoride* 36: 84-94.

Xiang Q, Liang Y, Chen B, Chen L. 2011. Analysis of children's serum fluoride levels in relation to intelligence scores in a high and low fluoride water village in China. *Fluoride* 44: 191-194.

Xu Y, Lu C, Zhang X. 1994. Effect of fluoride on children's intelligence. *Endem Dis Bull* 2: 83-84.

Yang Y, Wang X, Guo X, Hu P. 1994. [The effects of high levels of fluoride and iodine on child intellectual ability and the metabolism of fluoride and iodine]. *Chin J Epidemiol* 15(4): 296-298.

Yang Y, Wang X, Guo X, Hu P. 2008. The effects of high levels of fluoride and iodine on child intellectual ability and the metabolism of fluoride and iodine. *Fluoride* 41: 336-339.

Yao L, Zhou J, Wang S, Cui K, Lin F. 1996. Analysis of TSH levels and intelligence of children residing in high fluorosis areas. *Lit Inf Prev Med* 2(1): 26-27.

Yao Y. 1997. Comparative assessment of the physical and mental development of children in endemic fluorosis area with water improvement and without water improvement. *Lit Inf Prev Med* 3(1): 42-43.

Yasmin S, Ranjan S, D'Souza D. 2013. Effect of excess fluoride ingestion on human thyroid function in Gaya Region, Bihar, India. *Toxicol Environ Chem* 95: 1235-1243.

Yazdi SM, Sharifian A, Dehghani-Beshne M, Momeni VR, Aminian O. 2011. Effects of fluoride on psychomotor performance and memory of aluminum potroom workers. *Fluoride* 44: 158-162.

Yu YN, Yang WX, Dong Z, Wan CW, Zhang JT, Liu JL, Xiao KQ, Huang YS, Lu BF. 1996. [Neurotransmitter and receptor changes in the brains of fetuses from areas of endemic fluorosis]. *Chin J Endemiol* 15(5): 257-259.

Yu YN, Yang WX, Dong Z, Wan CW, Zhang JT, Liu JL, Xiao KQ, Huang YS, Lu BF. 2008. Neurotransmitter and receptor changes in the brains of fetuses from areas of endemic fluorosis. *Fluoride* 41: 134-138.

Yu X, Chen J, Li Y, Liu H, Hou C, Zeng Q, Cui Y, Zhao L, Li P, Zhou Z, Pang S, Tang S, Tian K, Zhao Q, Dong L, Xu C, Zhang X, Zhang S, Liu L, Wang A. 2018. Threshold effects of moderately excessive fluoride exposure on children's health: A potential association between dental fluorosis and loss of excellent intelligence. *Environ Int* 118: 116-124.

Zhang J, Yao H, Chen Y. 1998. [Effect of high level of fluoride and arsenic on children's intelligence]. *Chin J Public Health* 17(2): 57.

Zhang P, Cheng L. 2015. [Effect of coal-burning endemic fluorosis on children's physical development and intellectual level]. *Chin J Control Endem Dis* 30(6): 458-459.

Zhang S, Zhang X, Liu H, Qu W, Guan Z, Zeng Q, Jiang C, Gao H, Zhang C, Lei R, Xia T, Wang Z, Yang L, Chen Y, Wu X, Cui Y, Yu L, Wang A. 2015. Modifying effect of COMT gene polymorphism and a predictive role for proteomics analysis in children's intelligence in endemic fluorosis area in Tianjin, China. *Toxicol Sci* 144: 238-245.

Zhao LB, Liang GH, Zhang DN, Wu XR. 1996. Effect of a high fluoride water supply on children's intelligence. *Fluoride* 29: 190-192.

Zhao Q, Niu Q, Chen J, Xia T, Zhou G, Li P, Dong L, Xu C, Tian Z, Luo C, Liu L, Zhang S, Wang A. 2019. Roles of mitochondrial fission inhibition in developmental fluoride neurotoxicity: Mechanisms of action in vitro and associations with cognition in rats and children. *Arch Toxicol* 93(3): 709-726.

Zhao Q, Tian Z, Zhou G, Niu Q, Chen J, Li P, Dong L, Xia T, Zhang S, Wang A. 2020. SIRT1-dependent mitochondrial biogenesis supports therapeutic effects of resveratrol against neurodevelopment damage by fluoride. *Theranostics* 10: 4822-4838.

Zhao Y, Cui Y, Yu J, Zhang B, Nie J, Zhao L, Zhang Z, Liu H. 2018. [Study on the relationship between water-borne high iodine and thyroid hormone and children's intelligence level]. *J Environ Health* 35(1): 6-9.

Zhou G, Tang S, Yang L, Niu Q, Chen J, Xia T, Wang S, Wang M, Zhao Q, Liu L, Li P, Dong L, Yang K, Zhang S, Wang A. 2019. Effects of long-term fluoride exposure on cognitive ability and the underlying mechanisms: Role of autophagy and its association with apoptosis. *Toxicol Appl Pharmacol* 378: 114608.

C.2.1.2. Studies Not Available in HAWC

Bachinskii PP, Gutsalenko OA, Naryzhniuk ND, Sidora VD, Shliakhta AI. 1985. [Action of the body fluorine of healthy persons and thyroidopathy patients on the function of hypophyseal-thyroid the system]. *Probl Endokrinol* 31: 25-29.

Balabolkin MI, Mikhailets ND, Lobovskaia RN, Chernousova NV. 1995. [The interrelationship of the thyroid and immune statuses of workers with long-term fluorine exposure]. *Ter Arkh* 67: 41-42.

Baum K, Boerner W, Reiners C, Moll E. 1981. [Bone density and thyroid function in adolescents in relation to fluoride content of drinking water]. *Fortschr Med* 99: 1470-1472.

Berry WTC, Whittles JH. 1963. Absence of effect of fluoride upon the incidence of thyroid enlargements in Wiltshire schoolgirls. *Mon Bull Minist Health Public Health Lab Serv* 22: 50-52.

Cherkinskii SN, Zaslavskaia RM. 1956. [Significance of fluorides in potable water in the development of endemic goiter]. *Probl Endokrinol Gormonoter* 2: 70-75.

Choubisa SL. 2001. Endemic fluorosis in southern Rajasthan, India. *Fluoride* 34: 61-70.

Chuka A, Zhukovskil V, Mirku I, Postel'Niku D. 1964. Prezhdevremennoe starenie naseleniya v zone rasprostraneniya endemicheskogo zoba. *Vestnik Akad Med Nauk Sssr* 19: 23-27.

Dai HX, Zeng P, Wang KY, Zhang XG, Ma ZJ, Zhou YG, Fan ZX, Guo SH. 2013. [Analysis of a survey results of patients with suspected high iodine goiter in Liuji Town Fuping County of Shaanxi Province]. *Chin J Endemiol* 32: 408-411.

Day T, Powell-Jackson P. 1972. Fluoride, water hardness, and endemic goitre. *Lancet* 299(7761): 1135-1138.

Desai VK, Solanki DM, Bansal RK. 1993. Epidemiological study of goitre in endemic fluorosis district of Gujarat. *Fluoride* 26: 187-190.

Díaz-Cadórniga FJ, Delgado E, Tartón T, Valdés MM, Méndez A, Fernández MT, Rojo C. 2003. Endemic goiter associated with high iodine intake in primary school children in the Saharawi Arab Democratic Republic. *Endocrinol Nutr* 50: 357-362.

Eichner R, Borner W, Henschler D, Kohler W, Moll E. 1981. [Osteoporosis therapy and thyroid function. Influence of 6 months of sodium fluoride treatment on thyroid function and bone density]. *Fortschr Med* 99: 342-348.

Fiorentini S, Galeazzi M, Visintin B. 1947. II fluoro in natura come agente morbigeno II. La fluorosi die Campagnano di Roma. III. Un focolaio di fluorosi umana a Campagnano di Roma. IV. Osservazioni radiologiche sui processi alveolari, sulle ossa mascellari, e sul paradenzio degli abitanti die Campagnano. V. Zona fluorotica intorno a Campagnano di Roma. VI. Frequenza e caratteri clinici della carie dentale in soggetti fluorotici. *Rend Ist Superiore Sanita* 10: 721-804.

Gabovich RD, Verzhikovskaia NV. 1958. [Effect of fluoride compounds on absorption of radioactive iodine by the thyroid gland in humans and in experimental conditions]. *Probl Endokrinol Gormonoter* 4: 49-54.

Galletti P, Held HR, Korrodi H, Wegmann T. 1956. [Investigations on the possible damaging effect of fluorine on the thyroid gland]. *Helv Med Acta* 23: 601-605.

Galletti PM, Joyet G, Jallut O. 1957. [Effect of sodium fluoride on thyroid function in Basedow's Disease]. *Helv Med Acta* 24: 209-215.

Galletti PM, Joyet G. 1958. Effect of fluorine on thyroidal iodine metabolism in hyperthyroidism. *J Clin Endocrinol Metab* 18: 1102-1110.

Gas'kov AI, Savchenkov MF, Iushkov NN. 2005. [The specific features of the development of iodine deficiencies in children living under environmental pollution with fluorine compounds]. *Gig Sanit*: 53-55.

Gedalia I, Brand N. 1963. The relationship of fluoride and iodine in drinking water in the occurrence of goiter. *Arch Int Pharmacodyn Ther* 142: 312-315.

Grimm H. 1973. [The physical development of schoolchildren under the influence of drinking water fluoridation in Karl Marx Stadt]. *Dtsch Gesundheitsw* 28: 2363-2369.

Hasling C, Nielsen HE, Melsen F, Mosekilde L. 1987. Safety of osteoporosis treatment with sodium fluoride, calcium phosphate and vitamin D. *Miner Electrolyte Metab* 13: 96-103.

Hidehiko T. 1958. On the relation between the distribution of endemic goiter and the fluorine content of natural water in Hidaka Province, Hokkaido. *Folia Pharmacol Jpn* 54: 225-229.

Hoffmann-Axthelm W. 1953. [Observations on the influence of fluorine on dental enamel and thyroid gland]. *Dtsch Zahnarztl Z* 8: 757-765.

Jentzer A. 1956. [Effect of fluorine on the iodine content of the human thyroid gland]. *Bull Schweiz Akad Med Wiss* 12: 539-543.

Jooste PL, Weight MJ, Kriek JA, Louw AJ. 1999. Endemic goitre in the absence of iodine deficiency in schoolchildren of the Northern Cape Province of South Africa. *Eur J Clin Nutr* 53: 8-12.

Kolomiitseva MG. 1961. [The content of fluorine in the external environment of the Upper Altai autonomous region and its role in the etiology of endemic goiter]. *Gig Sanit* 26: 101-103.

Korrodi H, Wegmann T, Galleti P, Held HR. 1955. [Caries prophylaxis and the untoward effects of fluor on the thyroid gland]. *Schweiz Med Wochenschr* 85: 1016-1019.

Kutlucan A, Kale Koroglu B, Numan Tamer M, Aydin Y, Baltaci D, Akdogan M, Ozturk M, Vural H, Ermis F. 2013. The investigation of effects of fluorosis on thyroid volume in school-age children. *Med Glas* 10: 93-98.

Latham MC, Grech P. 1967. The effects of excessive fluoride intake. *Am J Public Health* 57: 651-660.

Leone NC, Leatherwood EC, Petrie IM, Lieberman L. 1964. Effect of fluoride on thyroid gland: Clinical study. *J Am Dent Assoc* 69: 179-180.

Levi JE, Silberstein HE. 1955. Lack of effect of fluorine ingestion on uptake of iodine 131 by the thyroid gland. *J Lab Clin Med* 45: 348-351.

McGlashan N, Chelkowska E, Sasananan S. 2010. A survey of goiter morbidity in Ban Mae Toen, northwest Thailand. *Southeast Asian J Trop Med Public Health* 41: 1200-1208.

Rathore S, Meena C, Gonmei Z, Dwivedi S, Toteja GS, Bala K. 2018. Study of excess fluoride ingestion and thyroid hormone derangement in relation with different fluoride levels in drinking water among children of Jodhpur District, Rajasthan, India. *Asian J Microbiol Biotechnol Environ Sci* 20: 327-331.

Reisenauer R, Rezler D, Křemenová J, Preininger Q. 1961. [Fluorization of the waters in Czechoslovakia. IV. Endocrinological control of results of two years' fluorization of drinking-water in school children]. *Cesk Stomatol* 61: 91-97.

Romer TEZ, Kowalczyk B, Kacprzak M, Wiktorowski M. 1976. [Incidence of goiter in pubertal girls of the Piotrkow Region and iodide content in drinking water]. *Endokrynol Pol* 27: 373-380.

Savchenkov MF, Efimova NV, Manueva RS, Nikolaeva LA, Shin NS. 2016. [Thyroid gland pathology in children population exposed to the combination of iodine deficiency and fluoride pollution of environment]. *Gig Sanit* 95: 1201-1205.

Shtifanova AK. 1962. [The fluorine content in water, soil and vegetal products of the Alma-Atinsk District areas and its role in the etiology of dental caries and endemic goiter]. *Zdravookhranenie Kazakhstana*: 60-63.

Siddiqui AH. 1969. Incidence of simple goiter in areas of endemic fluorosis in Nalgonda District, Andhra Pradesh, India. *Fluoride* 2: 200-205.

Sidora VD, Shliakhta AI, Iugov VK, Kas'ianenko AS, Piatenko VG. 1983. [Indices of the pituitary-thyroid system in residents of cities with various fluorine concentrations in drinking water]. *Probl Endokrinol* 29: 32-35.

Sung FC, Chen KP, Chen CY, Tai PW, Yang CF. 1973. Studies of the effect of salt iodization on endemic goiter in Taiwan. IV. A survey of drinking water in relation to endemic goiter. *J Fomosan Med Assoc* 72: 96-103.

Tokar VI, Voroshnin VV, Sherbakov SV. 1989. [Chronic effects of fluorides on the pituitary-thyroid system in industrial workers]. *Gig Tr Prof Zabol*: 19-22.

Wespi HJ. 1954. [Iodized-fluoridized salt for the prevention of goiter and caries]. *Schweiz Med Wochenschr* 84: 885-890.

Yu YN. 1985. [Effects of chronic fluorosis on the thyroid gland]. *Chin Med J* 65: 747-7479.

C.2.2. Studies in Non-human Animals

As described in Figure 2, 339 non-human mammal studies were included; however, full data extraction was conducted only on studies with primary neurological outcomes and/or secondary functional neurological outcomes (e.g., motor activity). Data extraction was completed using HAWC. Data were extracted from a subset of included studies in animals (n = 123) and are available in HAWC based on outcome. The following lists of references are organized as studies that are available in HAWC followed by studies that are not available in HAWC. Specifically, all primary outcomes and functional neurological secondary outcomes (e.g., motor activity) were extracted from animal studies and are available in HAWC, including studies from the NTP (2016) assessment. Studies are also available in HAWC that evaluated mechanistic effects related to oral fluoride exposure at or below 20 ppm fluoride drinking water equivalents for categories of mechanistic endpoints with the largest amount of available data (i.e., biochemistry of the brain or neurons, neurotransmission, oxidative stress, and histopathology [n = 70]); however, these mechanistic data were generally not extracted. Several animal studies assessed primary neurological outcomes and/or functional neurological secondary outcomes and mechanistic effects in the four mechanistic categories listed above (n = 56). In total, 140 animal studies are available in HAWC (70 with primary neurological outcomes and/or secondary functional neurological outcomes without relevant mechanistic data; 15 with relevant mechanistic data only; and 55 with primary and/or secondary functional neurological outcomes with relevant mechanistic data). Studies that evaluated other mechanistic endpoints, as well as studies that assessed only mechanistic effects at fluoride levels above 20 ppm fluoride drinking water equivalents, are not available in HAWC (n = 199). The list below presents the 339 non-human animal studies that were included in the review. An overview of the screening results is outlined in the study selection diagram (Figure 2) that reports numbers of included studies as well as numbers of studies excluded for each reason at the full-text review stage.

C.2.2.1. Studies Available in HAWC

Adedara IA, Abolaji AO, Idris UF, Olabiyi BF, Onibiyo EM, Ojuade TD, Farombi EO. 2017. Neuroprotective influence of taurine on fluoride-induced biochemical and behavioral deficits in rats. *Chem Biol Interact* 261: 1-10.

Adedara IA, Olabiyi BF, Ojuade TD, Idris UF, Onibiyo EM, Farombi EO. 2017. Taurine reverses sodium fluoride-mediated increase in inflammation, caspase-3 activity, and oxidative damage along the brain-pituitary-gonadal axis in male rats. *Can J Phys Pharmacol* 95: 1019-1029.

Agustina F, Sofro ZM, Partadiredja G. 2018. Subchronic administration of high-dose sodium fluoride causes deficits in cerebellar purkinje cells but not motor coordination of rats. *Biol Trace Elem Res* 188(2): 424-433.

Ahmad KR, Noor S, Jabeen S, Nauroze T, Kanwal MA, Raees K, Abbas T. 2017. Amelioration by jambul fruit extract of fluoride-induced hepato-nephronal histopathologies and impaired neuromotor capacity in mice. *Fluoride* 50: 2-14.

Akinrinade ID, Memudu AE, Ogundele OM. 2015. Fluoride and aluminium disturb neuronal morphology, transport functions, cholinesterase, lysosomal and cell cycle activities. *Pathophysiology* 22: 105-115.

Akinrinade ID, Memudu AE, Ogundele OM, Ajetunmobi OI. 2015. Interplay of glia activation and oxidative stress formation in fluoride and aluminium exposure. *Pathophysiology* 22: 39-48.

Baba NA, Raina R, Verma PK, Sultana M. 2014. Alterations in plasma and tissue acetylcholinesterase activity following repeated oral exposure of chlorpyrifos alone and in conjunction with fluoride in Wistar rats. *Proc Indian Natl Sci Acad B Biol Sci* 84: 969-972.

Bagmut I, Kolisnyk I, Titkova A, Babiy L, Filipchenko S. 2018. The antioxidant system enzymes' activity in rats' brain, intoxicated with sodium fluoride in subtoxic doses. *Arch Balkan Med Union* 53(4): 506-511.

Balaji B, Kumar EP, Kumar A. 2015. Evaluation of standardized bacopa monniera extract in sodium fluoride-induced behavioural, biochemical, and histopathological alterations in mice. *Toxicol Ind Health* 31: 18-30.

Balayssac D, Richard D, Authier N, Nicolay A, Jourdan D, Eschalier A, Coudore F. 2002. Absence of painful neuropathy after chronic oral fluoride intake in Sprague-Dawley and Lou/C rats. *Neurosci Lett* 327: 169-172.

Banala RR, Karnati PR. 2015. Vitamin A deficiency: An oxidative stress marker in sodium fluoride (NaF) induced oxidative damage in developing rat brain. *Int J Dev Neurosci* 47: 298-303.

Banala R, Nagapuri K, Mohd K, Reddy M, Karnati P. 2018. Carica Papaya leaf extract as a neuroprotective agent against behavioral and neurotransmitter changes in brain of the rat treated with sodium fluoride in pre- and post-natal periods. *Pharmacogn Mag* 14(55): 123-131.

Banji D, Banji OJ, Pratusha NG, Annamalai AR. 2013. Investigation on the role of spirulina platensis in ameliorating behavioural changes, thyroid dysfunction and oxidative stress in offspring of pregnant rats exposed to fluoride. *Food Chem* 140: 321-331.

Baran-Poesina V, Negres S, Dobrescu D, Dimcevici-Poesina N, Dimcevici-Poesina A, Feghiu A, Soare T, Militaru M. 2013. Experimental pharmacological researches regarding the influence of sodium fluoride in allopathic and homeopathic doses on central nervous system's performances:

A correlation between behavioral response in classic maze test and morphological aspects of cerebral cortex. *Farmacia* 61: 781-799.

Bartos M, Gumilar F, Bras C, Gallegos CE, Giannuzzi L, Cancela LM, Minetti A. 2015. Neurobehavioural effects of exposure to fluoride in the earliest stages of rat development. *Physiol Behav* 147: 205-212.

Bartos M, Gumilar F, Gallegos CE, Bras C, Dominguez S, Monaco N, Esandi MDC, Bouzat C, Cancela LM, Minetti A. 2018. Alterations in the memory of rat offspring exposed to low levels of fluoride during gestation and lactation: Involvement of the alpha7 nicotinic receptor and oxidative stress. *Reprod Toxicol* 81: 108-114.

Basha PM, Rai P, Begum S. 2011. Fluoride toxicity and status of serum thyroid hormones, brain histopathology, and learning memory in rats: A multigenerational assessment. *Biol Trace Elem Res* 144: 1083-1094.

Basha PM, Sujitha NS. 2012. Combined impact of exercise and temperature in learning and memory performance of fluoride toxicated rats. *Biol Trace Elem Res* 150: 306-313.

Bataineh HN, Nusier MK. 2006. Impact of 12-week ingestion of sodium fluoride on aggression, sexual behavior, and fertility in adult male rats. *Fluoride* 39: 293-301.

Bera I, Sabatini R, Auteri P, Flace P, Sisto G, Montagnani M, Potenza MA, Marasciulo FL, Carratu MR, Coluccia A, Borracci P, Tarullo A, Cagiano R. 2007. Neurofunctional effects of developmental sodium fluoride exposure in rats. *Eur Rev Med Pharmacol Sci* 11: 211-224.

Bhatnagar M, Rao P, Sushma J, Bhatnagar R. 2002. Neurotoxicity of fluoride: Neurodegeneration in hippocampus of female mice. *Indian J Exp Biol* 40: 546-554.

Bhatnagar M, Sukhwal P, Suhalka P, Jain A, Joshi C, Sharma D. 2011. Effects of fluoride in drinking water on NADPH-diaphorase neurons in the forebrain of mice: A possible mechanism of fluoride neurotoxicity. *Fluoride* 44: 195-209.

Chen H, Geng D. 2011. [The change of cognition induced by chronic fluoride in rats]. *Acta Academiae Medicinae Xuzhou* 31(5): 319-322.

Chen J, Niu Q, Xia T, Zhou G, Li P, Zhao Q, Xu C, Dong L, Zhang S, Wang A. 2018. ERK1/2-mediated disruption of BDNF-TrkB signaling causes synaptic impairment contributing to fluoride-induced developmental neurotoxicity. *Toxicology* 410: 222-230.

Chinoy NJ, Shah SD. 2004. Biochemical effects of sodium fluoride and arsenic trioxide toxicity and their reversal in the brain of mice. *Fluoride* 37: 80-87.

Chioca LR, Raupp IM, Da Cunha C, Losso EM, Andreatini R. 2008. Subchronic fluoride intake induces impairment in habituation and active avoidance tasks in rats. *Eur J Pharmacol* 579: 196-201.

Chouhan S, Flora SJS. 2008. Effects of fluoride on the tissue oxidative stress and apoptosis in rats: Biochemical assays supported by IR spectroscopy data. *Toxicology* 254: 61-67.

Chouhan S, Lomash V, Flora SJ. 2010. Fluoride-induced changes in haem biosynthesis pathway, neurological variables and tissue histopathology of rats. *J Appl Toxicol* 30: 63-73.

Cui YS, Zhong Q, Li WF, Liu ZH, Wang Y, Hou CC. 2017. [Effects of fluoride exposure on thyroid hormone level and intelligence in rats]. *Chin J Ind Hyg Occup Dis* 35: 888-892.

Dabrowska E. 1997. Effect of different fluorine doses on the supraoptic nucleus of the rat. *Folia Histochem Cytobiol* 35: 115-116.

Dong Y, Wang Y, Wei N, Guan Z. 2015. [Expression levels of brain muscarinic acetylcholine receptor in offspring rats of drinking-water borne fluorosis]. *Chin J Endemiol* 34: 326-330.

Dong YT, Wang Y, Wei N, Zhang QF, Guan ZZ. 2015. Deficit in learning and memory of rats with chronic fluorosis correlates with the decreased expressions of M1 and M3 muscarinic acetylcholine receptors. *Arch Toxicol* 89: 1981-1991.

Dong YT, Wei N, Qi XL, Liu XH, Chen D, Zeng XX, Guan ZZ. 2017. Attenuating effect of vitamin E on the deficit of learning and memory of rats with chronic fluorosis: The mechanism may involve muscarinic acetylcholine receptors. *Fluoride* 50: 354-364.

Dong YW, Y. Wei, N. Guan, Z. 2015. [Expression of muscarinic acetylcholine receptors in the brain of rats with chronic fluorosis]. *Chin J Endemiol* 34(2): 84-88.

Ekambaram P, Paul V. 2001. Calcium preventing locomotor behavioral and dental toxicities of fluoride by decreasing serum fluoride level in rats. *Environ Toxicol Pharmacol* 9: 141-146.

Ekambaram P, Paul V. 2002. Modulation of fluoride toxicity in rats by calcium carbonate and by withdrawal of fluoride exposure. *Pharmacol Toxicol* 90: 53-58.

Ekambaram P, Paul V. 2003. Effect of vitamin D on chronic behavioral and dental toxicities of sodium fluoride in rats. *Fluoride* 36: 189-197.

El-lethey HS, Kamel MM, Shaheed IB. 2010. Neurobehavioral toxicity produced by sodium fluoride in drinking water of laboratory rats. *J Am Sci* 6(5): 54-63.

El-lethey HS, Kamel MM. 2011. Effects of black tea in mitigation of sodium fluoride potency to suppress motor activity and coordination in laboratory rats. *J Am Sci* 7(4): 243-254.

El-lethey HS, Shaheed IB. 2011. Potential health impact of black tea against Na-F-induced alterations in territorial aggression, sexual behaviour and fertility of male rats. *Life Sci J* 8: 828-839.

Elliott L. 1967. Lack of effect of administration of fluoride on the central nervous system of rats. *Acta Pharmacol Toxicol (Copenh)* 25: 323-328.

Flace P, Benagiano V, Vermesan D, Sabatini R, Inchingolo AM, Auteri P, Ambrosi G, Tarullo A, Cagiano R. 2010. Effects of developmental fluoride exposure on rat ultrasonic vocalization, acoustic startle reflex and pre-pulse inhibition. *Eur Rev Med Pharmacol Sci* 14: 507-512.

Gabovich RD. 1962. [On the problem of the effect of fluorine in drinking water on the functional state of the central nervous system]. *Gig Sanit* 27: 10-12.

Gao Q, Liu YJ, Wu CX, Long YG, Guan ZZ. 2008. [Level of oxidative stress in rat brains and learning and memory function of rats with chronic fluorosis]. *Chin J Endemiol* 27: 371-373.

Gao Q, Liu YJ, Wu CX, Long YG, Guan ZZ. 2008. [Effects of fluoride on learning and memory and cholinesterase activity in rat brains]. *Chin J Endemiol* 27: 128-130.

Gao Q, Liu YJ, Guan ZZ. 2009. Decreased learning and memory ability in rats with fluorosis: Increased oxidative stress and reduced cholinesterase activity in the brain. *Fluoride* 42: 277-285.

Gao Y, Liu L, Young L, Huan L, Jin H. 2009. Effects of learning and memory of fluoride and the antagonism of selenium in rats. *Studies of Trace Elements and Health* 26(2): 1-3.

Ge QD, Tan Y, Luo Y, Wang WJ, Zhang H, Xie C. 2018. MiR-132, miR-204 and BDNF-TrkB signaling pathway may be involved in spatial learning and memory impairment of the offspring rats caused by fluorine and aluminum exposure during the embryonic stage and into adulthood. *Environ Toxicol Pharmacol* 63: 60-68.

Ge Y, Chen L, Yin Z, Song X, Ruan T, Hua L, Liu J, Wang J, Ning H. 2018. Fluoride-induced alterations of synapse-related proteins in the cerebral cortex of ICR offspring mouse brain. *Chemosphere* 201: 874-883.

Gopal K, Saxena R, Gupta GSD, Rana MD, Agrawal D. 2006. Fluoride induced alterations in neurobehavioural and cardiovascular responses in rats. *J Adv Zool* 27: 1-7.

Gui CZ, Ran LY, Wu CX, Long YG, He J, Zhang H, Guan ZZ. 2009. [Changes in learning and memory ability and brain cholinesterase activity in the rats with coal burning fluorosis]. *Chin J Endemiol* 28: 497-500.

Gui CZ, Ran LY, Li JP, Guan ZZ. 2010. Changes of learning and memory ability and brain nicotinic receptors of rat offspring with coal burning fluorosis. *Neurotoxicol Teratol* 32: 536-541.

Gui CZ, Ran LY, Guan ZZ. 2011. [Expression levels of brain nicotinic acetylcholine receptor mRNA and protein in coal-burning type of fluorosis rats]. *Chin J Endemiol* 30: 239-242.

Guner S, Uyar-Bozkurt S, Haznedaroglu E, Mentes A. 2016. Dental fluorosis and catalase immunoreactivity of the brain tissues in rats exposed to high fluoride pre- and postnatally. *Biol Trace Elem Res* 174: 150-157.

Han H, Du W, Zhou B, Zhang W, Xu G, Niu R, Sun Z. 2014. Effects of chronic fluoride exposure on object recognition memory and mRNA expression of SNARE complex in hippocampus of male mice. *Biol Trace Elem Res* 158: 58-64.

Hong JH, Ge YM, Ning HM, Wang JD. 2005. [Effects of High Fluoride and Low Iodine on Learning-Memory and TchE of Brain in Offspring Rats]. *Chin Prev Med* 6: 489-491.

Inkielewicz I, Krechniak J. 2004. Fluoride effects on glutathione peroxidase and lipid peroxidation in rats. *Fluoride* 37: 7-12.

Jain A, Mehta VK, Chittora RA, Mahdi A, Bhatnagar M. 2015. Melatonin ameliorates fluoride induced neurotoxicity in young rats: An in vivo evidence. *Asian J Pharm Clin Res* 8: 164-167.

Jetti R, Raghuveer CV, Mallikarjuna RC. 2016. Protective effect of ascorbic acid and Ginkgo biloba against learning and memory deficits caused by fluoride. *Toxicol Ind Health* 32: 183-187.

Jia B, Zong L, Lee JY, Lei J, Zhu Y, Xie H, Clemens JL, Feller MC, Na Q, Dong J, McLane MW, Jones-Beatty K, Burd I. 2019. Maternal supplementation of low dose fluoride alleviates adverse perinatal outcomes following exposure to intrauterine inflammation. *Sci Rep* 9(1): 2575.

Jiang C, Zhang S, Liu H, Guan Z, Zeng Q, Zhang C, Lei R, Xia T, Wang Z, Yang L, Chen Y, Wu X, Zhang X, Cui Y, Yu L, Wang A. 2014. Low glucose utilization and neurodegenerative changes caused by sodium fluoride exposure in rat's developmental brain. *Neuromolecular Med* 16: 94-105.

Jiang S, Su J, Yao S, Zhang Y, Cao F, Wang F, Wang H, Li J, Xi S. 2014. Fluoride and arsenic exposure impairs learning and memory and decreases mGluR5 expression in the hippocampus and cortex in rats. *PLoS One* 9(4): e96041.

Khan AM, Raina R, Dubey N, Verma PK. 2017. Effect of deltamethrin and fluoride co-exposure on the brain antioxidant status and cholinesterase activity in Wistar rats. *Drug Chem Toxicol* 41(2): 1-5.

Kinawy AA, Al-Eidan AA. 2018. Impact of prenatal and postnatal treatment of sodium fluoride and aluminum chloride on some hormonal and sensorimotor aspects in rats. *Biol Trace Elem Res*: 1-8.

Kivrak Y. 2012. Effects of fluoride on anxiety and depression in mice. *Fluoride* 45: 302-306.

Li M, Cui J, Gao YH, Zhang W, Sun LY, Liu XN, Liu Y, Sun DJ. 2015. Pathological changes and effect on the learning and memory ability in rats exposed to fluoride and aluminum. *Toxicol Res* 4: 1366-1373.

Li X, Zhang J, Niu R, Manthari RK, Yang K, Wang J. 2019. Effect of fluoride exposure on anxiety- and depression-like behavior in mouse. *Chemosphere* 215: 454-460.

Liu F, Ma J, Zhang H, Liu P, Liu YP, Xing B, Dang YH. 2014. Fluoride exposure during development affects both cognition and emotion in mice. *Physiol Behav* 124: 1-7.

Liu WX. 1989. [Experimental study of behavior and cerebral morphology of rat pups generated by fluorotic female rat]. *Chin J Pathol* 18: 290-292.

Liu YJ, Gao Q, Wu CX, Long YG, Guan ZZ. 2009. [Modified expression of extracellular signal-regulated protein kinase signal transduction in rat brains and changed capacity of learning and memory of rats with chronic fluorosis]. *Chin J Endemiol* 28: 32-35.

Liu YJ, Gao Q, Wu CX, Guan ZZ. 2010. Alterations of nAChRs and ERK1/2 in the brains of rats with chronic fluorosis and their connections with the decreased capacity of learning and memory. *Toxicol Lett* 192: 324-329.

Liu YJ, Gao Q, Long YG, Yu YN, Guan ZZ. 2011. [Influence of chronic fluorosis on expression of phospho-Elk-1 in rat brains]. *Chin J Endemiol* 30: 251-255.

Lou DD, Guan ZZ, Liu YJ, Liu YF, Zhang KL, Pan JG, Pei JJ. 2013. The influence of chronic fluorosis on mitochondrial dynamics morphology and distribution in cortical neurons of the rat brain. *Arch Toxicol* 87: 449-457.

Lu F, Zhang Y, Trivedi A, Jiang X, Chandra D, Zheng J, Nakano Y, Uyghurturk DA, Jalai R, Guner S, Mentes A, DenBesten PK. 2019. Fluoride related changes in behavioral outcomes may relate to increased serotonin. *Physiol Behav* 206: 76-83.

Ma J, Liu F, Liu P, Dong YY, Chu Z, Hou TZ, Dang YH. 2015. Impact of early developmental fluoride exposure on the peripheral pain sensitivity in mice. *Int J Dev Neurosci* 47: 165-171.

Manusha S, Sudhakar K, Reddy KP. 2019. Protective effects of allium sativum extract against sodium fluoride induced neurotoxicity. *Int J Pharm Sci Res* 10(2): 625-633.

McPherson CA, Zhang G, Gilliam R, Brar SS, Wilson R, Brix A, Picut C, Harry GJ. 2018. An evaluation of neurotoxicity following fluoride exposure from gestational through adult ages in Long-Evans hooded rats. *Neurotoxicol Res*: 1-18.

Mesram N, Nagapuri K, Banala RR, Nalagoni CR, Karnati PR. 2016. Quercetin treatment against NaF induced oxidative stress related neuronal and learning changes in developing rats. *J King Saud Univ Sci* 29: 221-229.

Mullenix PJ, Denbesten PK, Schunior A, Kernan WJ. 1995. Neurotoxicity of sodium fluoride in rats. *Neurotoxicol Teratol* 17: 169-177.

Nageshwar M, Sudhakar K, Reddy NCC, Reddy KP. 2017. Neuroprotective effects of curcumin on sodium fluoride induced behavioural and enzymatic changes in brain and muscles of rat. *J Environ Biol* 38: 675-681.

Nageshwar M, Sudhakar K, Reddy KP. 2018. Quercetin ameliorates oxidative stress, neural damage of brain and behavioral impairment of rat with fluoride exposure. *Int J Pharm Sci Res* 9(8): 3247-3256.

Nian W, Wang X, Shao D, Yu Q, Ouyang W, Zhang Z, Ruan Q. 2018. Effects of subchronic exposure to fluorine on hippocampal injury in mice and its molecular mechanism. *Acta Sci Circumst* 38(11): 4512-4519.

Niu Q, Chen J, Xia T, Li P, Zhou G, Xu C, Zhao Q, Dong L, Zhang S, Wang A. 2018. Excessive ER stress and the resulting autophagic flux dysfunction contribute to fluoride-induced neurotoxicity. *Environ Pollut* 233: 889-899.

Niu R, Sun Z, Wang J, Cheng Z. 2008. Effects of fluoride and lead on locomotor behavior and expression of nissl body in brain of adult rats. *Fluoride* 41: 276-282.

Niu R, Sun Z, Cheng Z, Li Z, Wang J. 2009. Decreased learning ability and low hippocampus glutamate in offspring rats exposed to fluoride and lead. *Environ Toxicol Pharmacol* 28: 254-258.

Niu R, Liu S, Wang J, Zhang J, Sun Z. 2014. Proteomic analysis of hippocampus in offspring male mice exposed to fluoride and lead. *Biol Trace Elem Res* 162: 227-233.

Niu R, Xue X, Zhao Y, Sun Z, Yan X, Li X, Feng C, Wang J. 2015. Effects of fluoride on microtubule ultrastructure and expression of Tubalpha1a and Tubbeta2a in mouse hippocampus. *Chemosphere* 139: 422-427.

Niu R, Chen H, Manthari RK, Sun Z, Wang J, Zhang J, Wang J. 2018. Effects of fluoride on synapse morphology and myelin damage in mouse hippocampus. *Chemosphere* 194: 628-633.

Nkpaa KW, Onyeso GI. 2018. Rutin attenuates neurobehavioral deficits, oxidative stress, neuro-inflammation and apoptosis in fluoride treated rats. *Neurosci Lett* 682: 92-99.

Paul V, Ekambaram P, Jayakumar AR. 1998. Effects of sodium fluoride on locomotor behavior and a few biochemical parameters in rats. *Environ Toxicol Pharmacol* 6: 187-191.

Pereira M, Dombrowski PA, Losso EM, Chioca LR, Da Cunha C, Andreatini R. 2011. Memory impairment induced by sodium fluoride is associated with changes in brain monoamine levels. *Neurotoxicol Res* 19: 55-62.

Pulungan ZS, Sofro ZM, Partadiredja G. 2016. Sodium fluoride does not affect the working memory and number of pyramidal cells in rat medial prefrontal cortex. *Anat Sci Int* 93(1): 128-138.

Raghu J, Raghuveer VC, Rao MC, Somayaji NS, Babu PB. 2013. The ameliorative effect of ascorbic acid and Ginkgo biloba on learning and memory deficits associated with fluoride exposure. *Interdiscip Toxicol* 6: 217-221.

Raju S, Sivanesan SK, Gudemalla K. 2019. Cognitive enhancement effect of Ginkgo biloba extract on memory and learning impairments induced by fluoride neurotoxicity. *Int J Res Pharm Sci* 10(1): 129-134.

Reddy MM, Karnati PR. 2015. Protective effects of aqueous extract of fruit pulp of tamarindus indica on motor activity and metabolism of the gastrocnemius muscle of rats treated with fluoride. *Int J Toxicol Pharmacol Res* 7: 241-246.

Reddy YP, Tiwari SK, Shaik AP, Alsaeed A, Sultana A, Reddy PK. 2014. Effect of sodium fluoride on neuroimmunological parameters, oxidative stress and antioxidative defenses. *Toxicol Mech Methods* 24: 31-36.

Rumiantsev GI, Novikov SM, Mel'nikova NN, Levchenko NI, Kozeeva EE. 1988. [Experimental study of the biological effect of salts of hydrofluosilicic acid]. *Gig Sanit*: 80-82.

Sarkozi K, Horvath E, Vezer T, Papp A, Paulik E. 2015. Behavioral and general effects of subacute oral arsenic exposure in rats with and without fluoride. *Int J Environ Health Res* 25: 418-431.

Shah SD, Chinoy NJ. 2004. Adverse effects of fluoride and/or arsenic on the cerebral hemisphere of mice and recovery by some antidotes. *Fluoride* 37: 162-171.

Shalini B, Sharma JD. 2015. Beneficial effects of emblica officinalis on fluoride-induced toxicity on brain biochemical indexes and learning-memory in rats. *Toxicol Int* 22: 35-39.