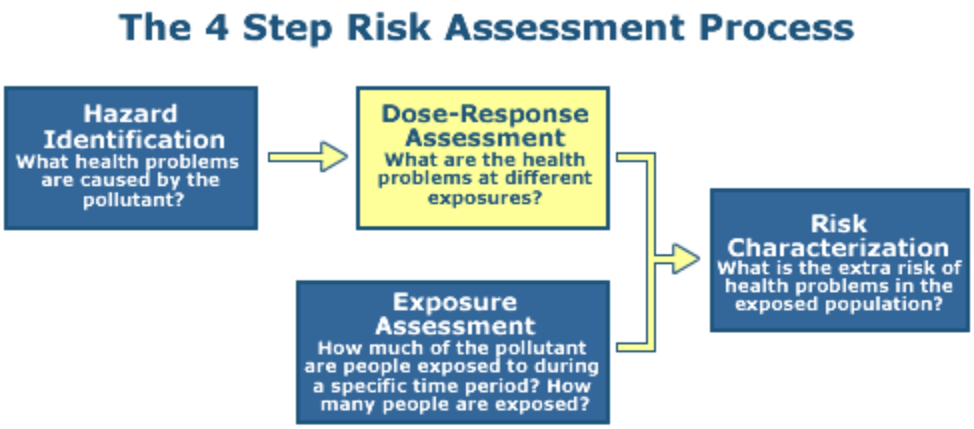
Neurath et al ISEE 2020 Conference poster:
Exposure classification protocol and individual study details for dose-response assessment

Notes on determining whether specific analyses of individual studies found adverse effects at fluoride (F) exposure levels below 0.7 mg/L or 1.5 mg/L water F or its equivalent.

The EPA has a good summary of how it conducts risk assessment and dose-response assessments:



https://www.epa.gov/risk/conducting-human-health-risk-assessment#tab-3

> "Upon considering all available studies, the response (adverse effect), or a measure of response that leads to an adverse effect (known as a 'precursor' to the effect), that occurs at the lowest dose is selected as the critical effect for risk assessment. The underlying assumption is that if the critical effect is prevented from occurring, then no other effects of concern will occur."
>
> "A No-Observed-Adverse-Effect Level (NOAEL) is the highest exposure level at which no statistically or biologically significant increases are seen in the frequency or severity of adverse effect between the exposed population and its appropriate control population. In an experiment with several NOAELs, the regulatory focus is normally on the highest one, leading to the common usage of the term NOAEL as the highest experimentally determined dose without a statistically or biologically significant adverse effect. In cases in which a NOAEL has not been demonstrated experimentally, the term "lowest-observed-adverse-effect level (LOAEL)" is used, which this is the lowest dose tested.
>
> Mathematical modeling, which can incorporate more than one effect level (i.e., evaluates more data than a single NOAEL or LOAEL), is sometimes used to develop an alternative to a NOAEL known as a Benchmark Dose (BMD) or Benchmark Dose Lower-confidence Limit (BMDL). In developing the BMDL, a predetermined change in the response rate of an adverse effect (called the benchmark response or BMR; generaly in the range of 1 to 10% depending on the power of a toxicity study) is selected, and the BMDL is a statistical lower confidence limit on the dose that produces the selected response. When the non-linear approach is applied, the LOAEL, NOAEL, or BMDL is used as the point of departure for extrapolation to lower doses."

The key principles and methods of EPA quoted above have been adopted for classifying the 29 higher quality fluoride neurotoxicity studies scored by the NTP as "lower Risk of Bias".

Protocol: Specific criteria and assumptions for classifying studies into three exposure levels

1. "Equivalence" between urine F concentration and drinking water F concentration will be assumed to mean urine F = water F.

However, Till 2018 Table 2 suggests for a given water F concentration the creatinine-adjusted urine F may be 40% greater than water F. Thus, a creatinine-adjusted urine F of 1.0 mg/L may be equivalent to a water F of 0.7 mg/L. This may be due to additional sources of F exposure besides water F, or it may be due to F metabolism factors. For unadjusted urine F, Till 2018 reported a mean value of 0.70 in fluoridated areas having an average water F of 0.61 mg/L. The ratio is thus 1.15:1 and a water F of 0.7 mg/L would be equivalent to an unadjusted urine F of 0.8 mg/L.

Some studies have reported urine F concentrations adjusted for creatinine in units of mg F per g creatinine, or mg/g_{cr}. The Till 2018 paper, Table 2, found that in fluoridated areas with a mean water F of 0.61 mg/L, the average creatinine-adjusted urine F in units of mg/g_{cr} was 1.15. Thus the ratio was 1.9:1 for this measurement unit, a urine F of 1.3 mg/g_{cr} would be equivalent to a water F of 0.7 mg/L.

Nevertheless, for this analysis the assumption will be that unadjusted urine F = water F. When urine F is in units of mg/g_{cr} the equivalence will be determined by the ratio 1.9:1.

When a study has both water F and urine F measures, the measure with lower values will be used if it has a statistically significant association with neurotoxic outcomes, and this will usually be water F. This is because we wish to focus on the contribution from water F rather than all sources of F.

2. The cutoff of 0.7 mg/L was chosen because that is currently the most common level of artificial water fluoridation in the USA. However, there are water systems in the USA that are still fluoridating at higher levels, and most state or local regulations allow levels up to 1.2 mg/L or even 1.5 mg/L. There are no federal regulations on the concentrations used in artificial water fluoridation, only the recommendation from the CDC that the level be 0.7 mg/L.

3. The cutoff of 1.5 mg/L was chosen because the NTP monograph has chosen that as a level, below which, exposures are considered relevant for most of the USA. NTP does not distinguish between artificial fluoridation and natural fluoridation. The EPA has estimated that several percent of the US population, or several million people, may be exposed to levels above 1.5 mg/L from natural fluoridation.

4. For studies with multiple subpopulations, outcomes or exposure measures, the most sensitive significant association was chosen, consistent with standard risk assessment

practice. For example, when one gender is more sensitive; one genetic polymorphism is more sensitive; one outcome is more sensitive; or one exposure measure is associated with larger effect sizes for equivalent exposure; that most sensitive association is chosen.

5. For studies with ecological (group-level) exposure data, the study must report at least 3 levels of exposure to be suitable for dose-response assessment, consistent with standard EPA risk assessment methods. Thus, when only 2 exposure levels are available, such as from the average water F concentration of two comparison villages, no dose-response assessment was done except to determine whether the higher exposure population had a mean or median below a cut-off of 0.7 or 1.5 mg/L.

6. When possible, Benchmark Dose methods were applied to estimate a BMD and BMDL level, consistent with standard EPA risk assessment methods. Benchmark Dose methods use all available data to more reliably produce an estimated dose-response than methods that use just part of the available data, such as the NOAEL method.

If a Benchmark Dose assessment found a BMD that was lower than the lowest observed dose, then the shape of the predicted dose-response curve was assessed in relationship to observed doses to determine whether a threshold might exist above 0.7 or 1.5 mg/L. If no such threshold is suggested by the curve then the study was classified as showing an effect below 0.7 or 1.5 mg/L. This criteria requires that a sufficient number of observations have values below the cutoff.

7. When suitable data was available Benchmark Dose assessments were done using EFSA's PROAST software. PROAST does not include linear dose-response relationships. PROAST only includes exponential and Hill models. PROAST website: https://proastweb.rivm.nl/

8. When insufficient data was available to conduct Benchmark Dose assessments, but published papers had evaluated the dose-response relationship in their data, the published findings were used. So, for example, if an analysis found a statistically significant linear dose-response and the observed data included a substantial number with exposure values below 0.7 mg/L, then that analysis was considered to meet the cutoff of 0.7 mg/L. Some studies conducted additional assessments to try to determine whether there might be a threshold or a non-linear dose-response. For such studies, those findings were noted and considered when deciding whether it met a cutoff of 0.7 or 1.5 mg/L. An example is Bashash 2017, which evaluated the data for non-linear dose-response relationships using GAM (Generalized Additive Models). For the outcome IQ at age 6-12 y, they found what appeared to be a possible threshold around 0.8 mg/L. For this outcome, the dose would not meet the 0.7 mg/L cutoff, but would meet the 1.5 mg/L cutoff. Note that this paper reported analyses for other outcomes that were more sensitive (IQ measured using the GCI measure at age 4 y) and those outcomes did not show evidence of a threshold, so the Bashash 2017 paper is considered to have found a statistically significant adverse effect below 0.7 mg/L, consistent with the principle explained above of using the most sensitive analysis of a paper.

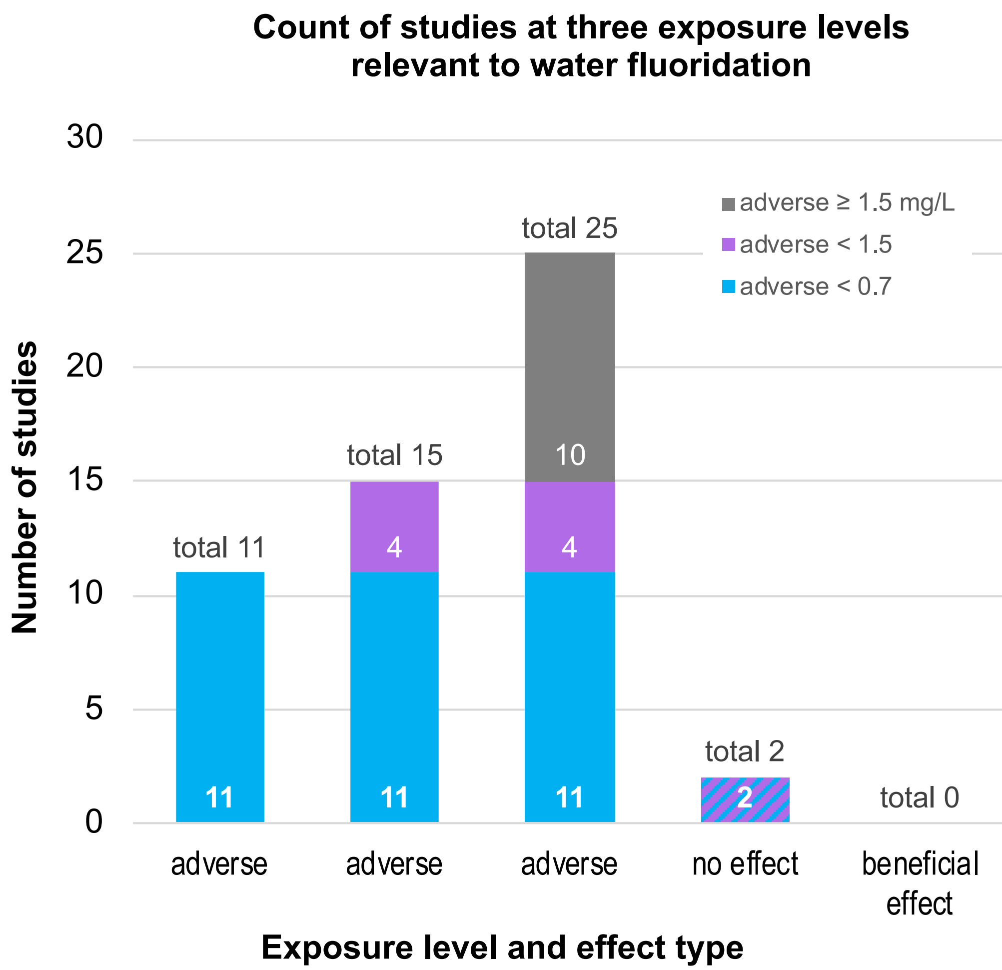
Summary of study exposure classifications

NTP rated 29 studies as lower Risk of Bias (higher quality). Two of the studies were essentially duplications of a third so have been excluded. Of the 27 different high quality studies, 25 found statistically significant adverse neurotoxic effects and 2 found no significant effect. None found a significant beneficial effect.

Of the 25 finding adverse effects, 11 found effects below 0.7 mg/L in water F or urine F, 4 found effects below 1.5 mg/L, and 10 found effects at or above 1.5 mg/L. The classifications of individual studies are shown in this table:

Study	Year	Exposure Classification	↓ – ↑	adverse ↓ no effect – beneficial ↑	Basis for Classification
Barberio	2017b	**below 0.7**	↓	adverse	mean concentration
Bashash	2017	**below 0.7**	↓	adverse	mean concentration, no threshold
Bashash	2018	**below 0.7**	↓	adverse	mean concentration, no threshold
Choi	2015	equal or above 1.5	↓	adverse	mean concentration
Cui	2018	**below 1.5**	↓	adverse	BMD analysis, individual-level data
Cui	2020	**below 1.5**	–	no effect	BMD analysis, group-level data
Ding	2011	**below 0.7**	↓	adverse	BMD analysis, group-level data
Green	2019	**below 0.7**	↓	adverse	mean concentration, no threshold
Li	2004	equal or above 1.5	↓	adverse	mean concentration
Riddell	2019	**below 0.7**	↓	adverse	mean concentration
Rocha-Amador	2007	equal or above 1.5	↓	adverse	mean concentration
Rocha-Amador	2009	equal or above 1.5	↓	adverse	mean concentration
Saxena	2012	equal or above 1.5	↓	adverse	mean concentration
Seraj	2012	**below 0.7**	↓	adverse	BMD analysis, group-level data
Soto-Barreras	2019	**below 0.7**	–	**no effect**	mean concentration
Sudhir	2009	**below 0.7**	↓	adverse	BMD analysis, group-level data
Till	2020	**below 0.7**	↓	adverse	mean concentration
Trivedi	2012	equal or above 1.5	↓	adverse	mean concentration
Valdez-Jimenez	2017	**below 1.5**	↓	adverse	BMD analysis, individual-level data
Wang	2012	**below 1.5**	↓	adverse	BMD analysis, group-level data
Wang	2020a	**below 1.5**	↓	adverse	BMD analysis, individual-level data
Wang	2020b	**below 1.5**	↓	adverse	mean concentration
Xiang	2003a	equal or above 1.5	↓	adverse	BMD analysis, individual-level data
Xiang	2011	equal or above 1.5	↓	adverse	BMD analysis, individual-level data
Yu	2018	**below 0.7**	↓	adverse	BMD analysis, individual-level data
Zhang	2019b	**below 0.7**	↓	adverse	BMD analysis, individual-level data
Zhao	2019	**below 0.7**	↓	adverse	[duplicate of Zhou 2019]
Zhao	2020	**below 0.7**	↓	adverse	[duplicate of Zhou 2019]
Zhou	2019	**below 0.7**	↓	adverse	BMD analysis, individual-level data

Summary of exposure classifications of 27 studies considered higher quality by NTP:



Individual study classification details

Barberio 2017b

This study was in a sample considered nationally representative of Canada and included people in areas with and without artificial fluoridation. The exposure distribution is expected to be similar to that found in Till 2018. Canada's artificial fluoridation level averages about 0.6 mg/L. Therefore, this study is considered to have found an adverse effect **below 0.7 mg/L** water F.

Bashash 2017

This study found a statistically significant adverse association (lowered IQ at age 4 y) that was judged to be best fit by a linear dose-response. Figure 2 shows that a considerable proportion of the sample had maternal urine F levels below 0.7 mg/L. Table 2 provides more specific information on the distribution of exposures. For the age 4 y analysis (GCI, Total) it shows that the 25th percentile had maternal urine F of 0.65 mg/L and the minimum was 0.23 mg/L.

The distribution of maternal urine F exposures in Bashash 2017 for the ELEMENT cohort is very similar to the distribution in Till 2018 and Green 2019 papers for the MIREC cohort. The mean levels are almost the same. The distribution and mean are also very similar to the study in pregnant California women [ref]. Thus, although the main source of F in this study was fluoridated salt, the exposure levels are very similar to those for the ELEMENT study in which the areas with artificial fluoridation had water F averaging 0.6 mg/L.

Therefore, this study is classified as finding adverse effects **below 0.7 mg/L**.

Supporting this finding are separate BMD analyses using the reported data that found a BMD of 0.16 mg/L and a BMDL of 0.10 mg/L [Grandjean 2020 TSCA case declaration; Grandjean 2019 review].

Additional information on exposures in this study and other ELEMENT studies can be found in Thomas 2016: http://dx.doi.org/10.1016/j.envres.2016.06.046

Bashash 2018

Bashash 2018 is based on the same ELEMENT cohort as Bashash 2017, however the outcome measures are related to ADHD rather than IQ. For similar reasons that Bashash 2017 was considered to have found a statistically significant adverse effect **below 0.7 mg/L**, Bashash 2018 was concluded to also have met this criteria. The regression line shown in Figure 2 for the outcome "DSM IV Total" in Bashash 2018 is very similar to that in Figure 2 in Bashash 2017.

Both show a statistically significant approximately linear dose-response that does not appear to have a lower threshold.

All the ELEMENT studies have useful exposure information reported in Thomas 2016: http://dx.doi.org/10.1016/j.envres.2016.06.046

Choi 2015

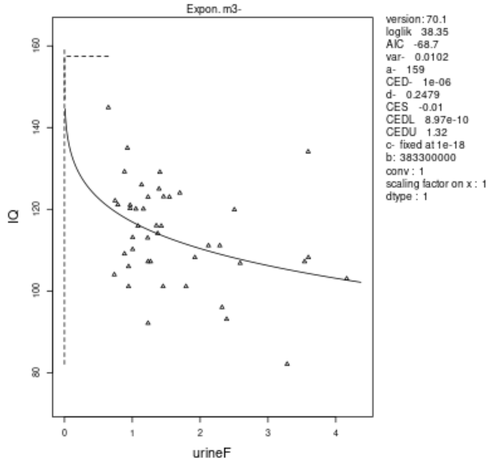
The Choi 2015 study had a mean water F level of 2.20 mg/L and a mean child urine F of 1.64 mg/L. The only statistically significant associations at relatively low doses were for differences between the group with very mild or mild compared to those with normal or questionable dental fluorosis. The mean water F and urine F for those with very mild or mild fluorosis was 1.91 and 2.10 mg/L respectively, so this study is classified as showing a significant adverse effect at levels **above 1.5 mg/L**.

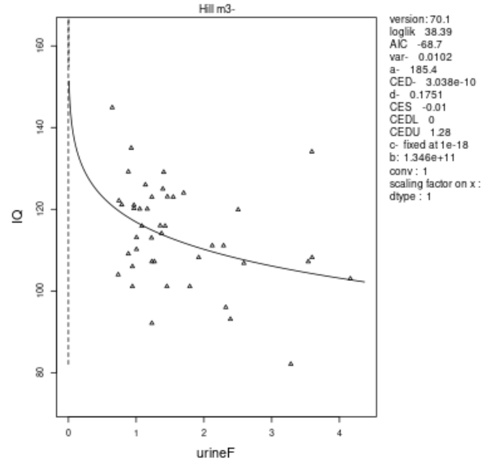
Cui 2018

Data was extracted from Figure 2C for the association between urine F (log transformed) and IQ for the subset with genetic variant TT, which was the more sensitive genotype. A larger proportion of observations had exposures below 1.5 mg/L but there were no observations with exposures below 0.7 mg/L.

BMD analysis produced a BMDL that was essentially 0.0 mg/L urine F which is for a dose-response relationship that was for the best fit Exponential m3- model. The BMD analyses predicts a larger effect magnitude at lower doses than higher doses, so no threshold is indicated. However, because there is no data below 0.7 mg/L, we classified this study as showing a significant adverse effect at levels **below 1.5 mg/L**.

Here are the graphs of the best fitting models from the PROAST Benchmark Dose analysis. PROAST uses different terminology from EPA. To translate: CED = BMD, CESL = BMDL, CES = BMR.





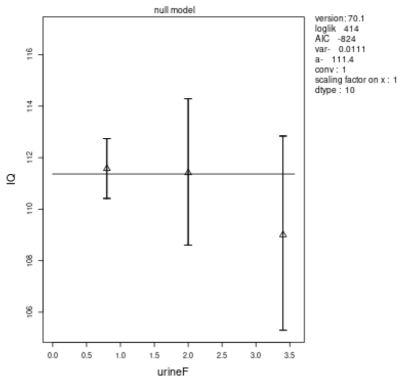
Cui 2020

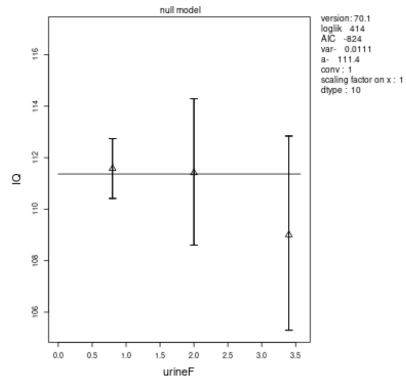
The limited data on F and IQ is all from Table 1 and is for urine F. However, this study's focus was on the association between iodine and IQ, not F and IQ, so F was a covariate in regression models with the main exposure iodine. There were no multivariable regression models with the main exposure being F. Table 1 has only unadjusted mean IQ scores for three exposure level groups. Since water iodine varied greatly and may have been associated with F, iodine may have confounded the association between F and IQ. Therefore, this study is not really suitable for assessing the association between F and IQ. It may have been improperly scored as lower Risk of Bias because it was scored for its analysis of iodine and IQ, rather than F and IQ.

Although there is a trend of slightly decreasing IQ with increasing F, the trend does not appear to be statistically significant and pair-wise comparisons between the 3 exposure group levels are not significantly different. It is not clear whether these extracted results have been adjusted for iodine, which was the main focus of the study. There was a wide range of iodine in the study and it might have been associated with F levels, since the source was drinking water.

Since no statistically significant adverse neurological effect was found, this study was considered a "no effect" study. Most of the children had urine F levels <1.6 mg/L so this study was considered to have assessed effects at levels **below 1.5 mg/L, but found none that reached statistical significance.**

A BMD analyses using the group-level data found no statistically significant trends, confirming the one-way anova between group differences assessment of not-significant. The exposure levels were estimated using the NTP method for groups reported with ranges rather than means or medians.

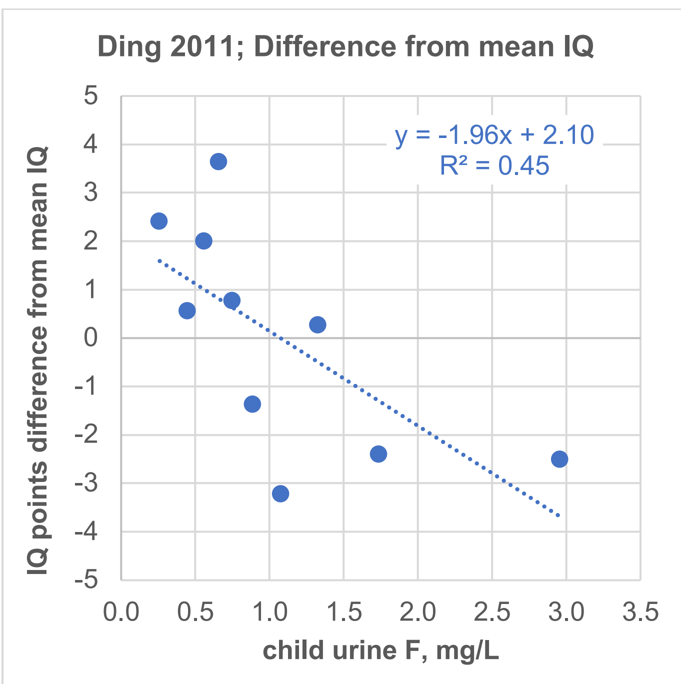




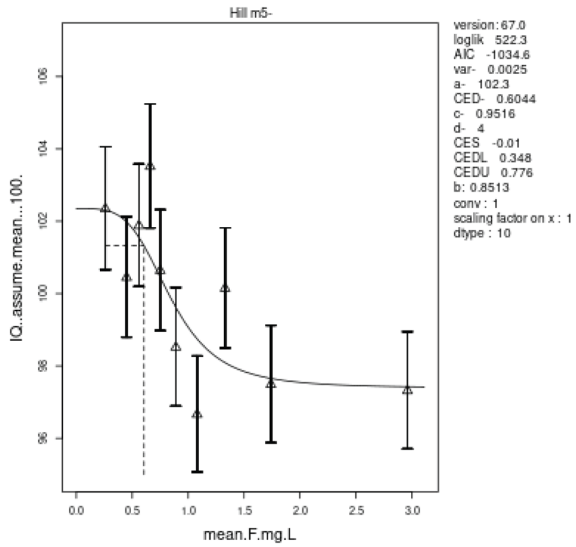
Ding 2011

BMD analysis was possible with data extracted from Figure 2. The NTP extracted the data and provides values for mean urine F and IQ (±SD) for the 10 exposure groups (n = 11 for each group):
https://hawcproject.org/epi/outcome/25328/

A scattergram of this data suggests a clear dose-response relationship:



BMD analysis of the extracted data:



BMD = 0.60 mg/L; BMDL = 0.35 mg/L; BMR = -1 IQ point

BMD is shown with vertical dashed line.

This Benchmark Dose analysis suggests there may be a threshold, but it is at about 0.3 mg/L. The estimated BMD and BMDL are both below 0.7 mg/L, so this study will be classified as finding a significant effect **below 0.7 mg/L**.

Green 2019

Additional information on exposure levels in Green 2019 and the MIREC cohort are available in Till 2018.

Mean water F concentration in fluoridated areas was 0.6 mg/L. Also, analyses to see if there was a threshold were conducted and no threshold detected. Linear dose-response fit data as well as any other model. Grandjean [2019] conducted BMD analyses and estimated BMD of 0.51 mg/L for boys and girls combined. Many of the mothers had maternal urine F below 0.51 mg/L. Therefore, this study classified as finding adverse effect **below 0.7 mg/L**. Grandjean 2019: https://doi.org/10.1186/s12940-019-0551-x

Li 2004 (translated in Li 2008a)

Ecological exposure study comparing mean water F and mean maternal urine F in a "high exposure" group to a "low exposure group". The sample comprised children from several "high F" and several reference level villages. Drinking water in the "high F" villages ranged from 1.7 – 6.0 mg/L and in the reference villages from 0.5 – 1.0 mg/L. Urine F was 3.58±1.47 mg/L in "high F" group and 1.74±0.96 mg/L in the reference group. With the "high F" group having water F and urine F above 1.5 mg/L this study is classified as finding an effect **above 1.5 mg/L**.

This study is of interest for looking at the ratio of urine F to water F. Instead of the assumed 1:1 it is closer to 2:1 in the reference group that has water F levels in the same range as most of the USA. This suggests that our assumption of a 1:1 ratio when evaluating studies that measured urine F may be non-protective.

A closer comparison to the USA can be made with the Till 2018 data on urine F distribution in the Canadian MIREC cohort. In the women in fluoridated areas with an average concentration of 0.59 mg/L the average unadjusted urine F was 0.69 mg/L. For these women, the creatine-adjusted urine F was 0.87 mg/L. This represent urine F to water F ratios of 1.2:1 and 1.5:1, respectively. These are likely underestimates because some women drank bottled water. But if we assume the ratio of unadjusted urine F to water F is 1.2 in a population with exposures very similar to the US, then it would be expected that unadjusted urine F levels would be about 0.8 mg/L when water F is 0.7 mg/L.

Riddell 2019

This study used Canadian Health Measures Survey (CHMS) data which is a nationally representative sample of Canada. The Green 2019 study found that the average water F concentration in fluoridated parts of Canada was 0.59 mg/L and in unfluoridated parts 0.13 mg/L. This study is therefore classified as finding an effect **below 0.7 mg/L**.

Rocha-Amador 2007

Water F in the three study towns averaged 0.8, 5.3, and 9.4 mg/L. Creatine-adjusted urine F averaged 1.8, 6.0, and 5.5 mg/g_{cr} in the three towns. The urine F exposure was treated as a continuous variable in the study, but no further breakdown in distribution or on the dose-response relationship was given so the possibility of a threshold cannot be assessed. Therefore, this study was classified as having found an adverse effect **above 1.5 mg/L**.

Note that for the town with water F of 0.8 mg/L, which is similar to areas with artificial fluoridation, the creatinine-adjusted urine F of 1.8 mg/g_{cr} gives a ratio of 2.25:1, which is roughly similar to the ratio 1.9:1 found in the Till 2018 study of Canadians living in fluoridated areas. It should be noted that the Rocha-Amador 2007 study was urine F in children while in Till 2018 it was in pregnant adult women.

Rocha-Amador 2009

This study included children from the highest water F town of Rocha-Amador 2007 which had a mean water F of about 9 mg/L. The creatinine-adjusted urine F averaged 5.6 mg/g_{cr}. Therefore, this was classified as having found an adverse effect **above 1.5 mg/L**.

Saxena 2012

Figures 3 and 4 show that all of the urine F (unadjusted) values were above about 1.8 mg/L, and all of the water F values were above about 0.7 mg/L. While more than ¼ of the children had water F below 1.5 mg/L, Figure 3 does not give adequate information to rule out a possible threshold below 1.5 mg/L. Therefore, this study is classified as having found an adverse effect **above 1.5 mg/L**.

Figure 3 and 4 could possibly allow extraction of individual level data on F and IQ so that a Benchmark Dose assessment could be made from individual data. However, it appears there is overlap of observed data points in the graphs. Also, the IQ scores have been categorized into just 5 grades. However, digitization of Figure 3 recognized only 132 data points rather than the 170 for the total sample, so there must be overlap of data points and insufficient data can be extracted from the figures.

Alternatively, Table 2 could be used for a Benchmark Dose assessment with the subjects grouped into 4 exposure levels, except SDs are not provided for the mean IQ scores.

Note that the effect size is quite large. The IQ grading system used in the study can be roughly converted to equivalent IQ scores on a scale with median 100 and SD 15. We assume:

Grade 1 = 130 IQ
Grade 2 = 115 IQ
Grade 3 = 100 IQ
Grade 4 = 85 IQ
Grade 5 = 70 IQ

Using these conversion values, Figure 4 shows that as urine F increased from 2 mg/L to 8 mg/L the predicted IQ went from about Grade 1 (IQ=127) to about Grade 3 (IQ=103). That is a difference of -24 IQ points for a difference of 6 mg/L or -4 IQ points per 1 mg/L increase in urine F.

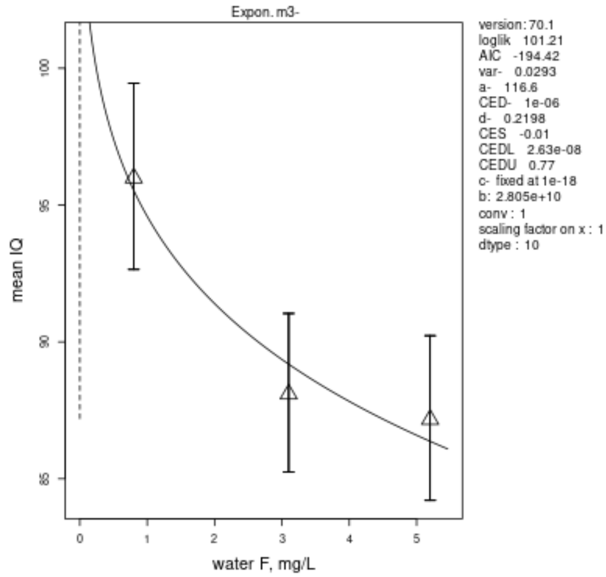
Seraj 2012

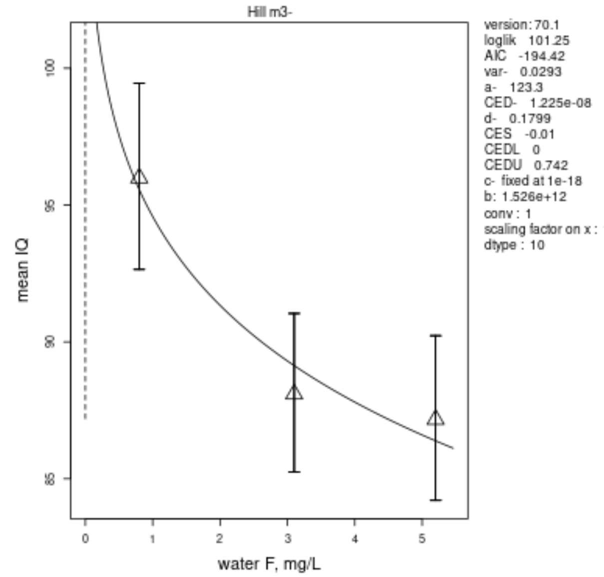
Ecological exposure measure with 3 exposure level groups:

"In this cross-sectional study, 293 children aged 6-11 years were selected from five villages in Makoo with normal fluoride (0.8±0.3 ppm), medium fluoride (3.1±0.9 ppm) and high fluoride (5.2±1.1 ppm) in their water supplies."

Therefore, this would be classified by NTP as having found an adverse effect above 1.5 mg/L. However, it is possible to do a BMD analysis since subjects were stratified into 3 exposure levels.

Results of group-level BMD analysis:





BMD=0.0; BMDL=0.0 for both Benchmark Dose models (Expon. m3- and Hill m3-).

The mean exposure level in the lowest exposure group (the "normal" group) is 0.8 mg/L with ±0.3 SD mg/L water F. There would be a large number of children with exposure below 0.7 mg/L, however, since the mean for this group is 0.8 mg/L, this will be classified as having found an adverse **effect below 1.5 mg/L**.

Soto-Barreras 2019

This is a "no effect" study.

This study probably should not have been rated as "lower RoB". It is debatable whether all the analyses described in the Methods section were actually reported in the Results section. Specifically, there were no regression analyses of the F and IQ data. No summary statistics for the whole sample were reported (e.g. mean, median, SD, or other distribution information) for exposures, outcomes, or covariates.

It is also unclear what statistical tests were used to reach the conclusion of "no statistically significant association" between F and IQ or any of the covariates. For example, Parental Education Level appears to have a strong association with child Intellectual Grade. Perhaps the statistical tests used were not appropriate or were not as sensitive as the data would justify.

Overall, it appears this study had relatively low statistical power to detect an effect. The sample size was relatively small (n = 161) and the outcome measure was crude (just 5 intelligence grades, and for the 2 extreme grades the sample size was very small with just 6 in the "intellectually superior" grade and just 4 in the "intellectually defective" grade). The exposure means and distribution information is never reported for the whole sample for any of the exposure measures (water F, urine F, dental fluorosis index).

The exposures appeared to all be relatively low with a low SD. Therefore, there may have been little contrast in exposure amongst the children. Table 2 shows that children with severe dental fluorosis had a mean water F of 1.66±0.93 mg/L while the children with no fluorosis had 0.75±0.95 mg/L, a contrast of only about 2x.

Considering the water F for 44% of the children was above 1.0 mg/L, and most children had from 0.75 to 1.66 mg/L, it is surprising that the average urine F is only about 0.6 mg/L.

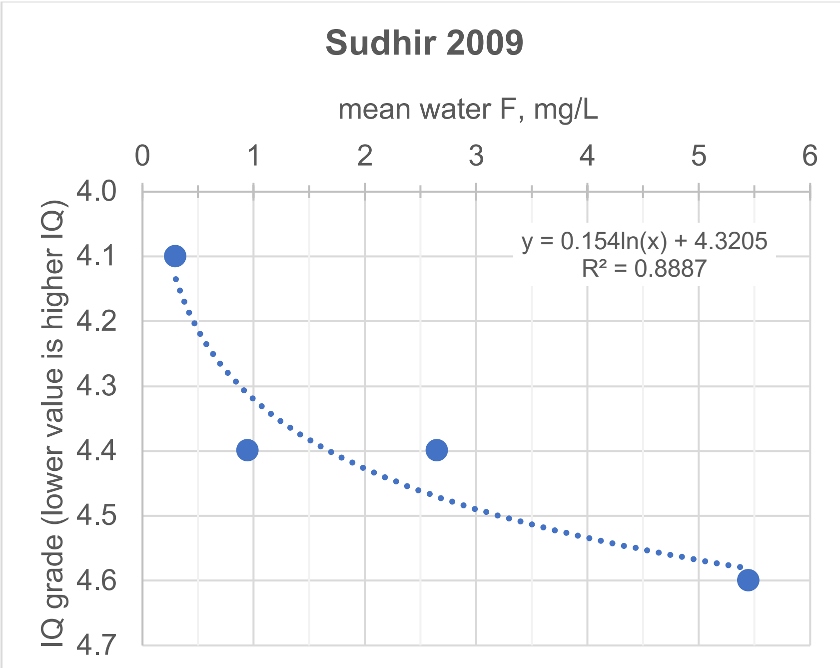
A possible explanation for the low urine F levels is because only 47% of children reported drinking public tap water. The remainder presumably drank bottled water, which may have been lower in F than tap water.

Given the low urine F levels and the narrow range of urine F levels, it is not surprising this study was not able to detect an association between F and IQ. If the EPA's study quality criteria had been used, in which sufficient contrast in dose is a factor, this study would probably have been rated poorer quality.

Because the mean urine F was below 0.7 mg/L his study is classified as having found **no effect below 0.7 mg/L, or at any higher level either**.

Sudhir 2009

This study used an ecological exposure measure with 4 levels. The lowest and highest exposure levels are open-ended range so must make assumptions for means. If we use NTP assumptions for open ended upper and midpoint of lower with zero as lowest we get this graph:



Below is the original graph, but not graphed as scatterplot:

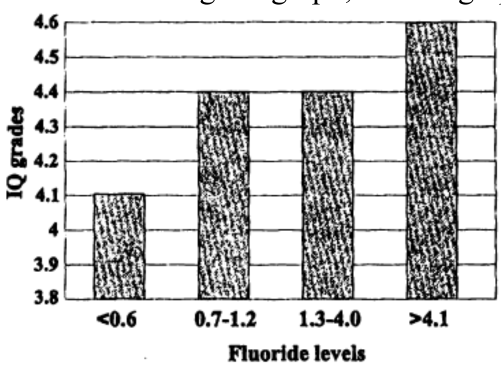


Fig. 1: IQ grades in relation to fluoride levels

Scatterplot has better logarithmic than linear fit. It appears there is loss of IQ even below 0.7 mg/L although that is a subjective observation. Certainly below 1.5 mg/L.

BMD analysis for this study was done by converting intellectual grades to mean IQ scores. The percentiles for the intellectual grades were used to make this conversion, as shown below:

IQ=130 Grade 1: "intellectually superior", if a score lies at or above 95th percentile for people of the same age-group.

IQ=116 Grade 2: "definitely above the average in intellectual capacity", if score lies at or above the 75Ih percentile.

IQ=100 Grade 3: "intellectually average", if a score lies between the 25th and 75th percentile.

IQ=84 Grade 4: "definitely below average in intellectual capacity" if a score lies at or below the 25th percentile.

IQ=70 Grade 5: "intellectually impaired", if score lies at or below the 5th percentile for that age group.

Grade 4.0 = 84.0 percentile IQ
Grade 5.0 = 70.0 percentile IQ

The four exposures groups had mean intellectual grades from 4.1 to 4.6, so conversion to percentile IQ scores over this range was done by dividing the difference between Grade 4.0 and Grade 5.0 by 10 to get an increment of 1.5 percentile IQ per increment of 0.1 Grade and applying it to each decimal Grade between 4.0 and 5.0, as shown here:

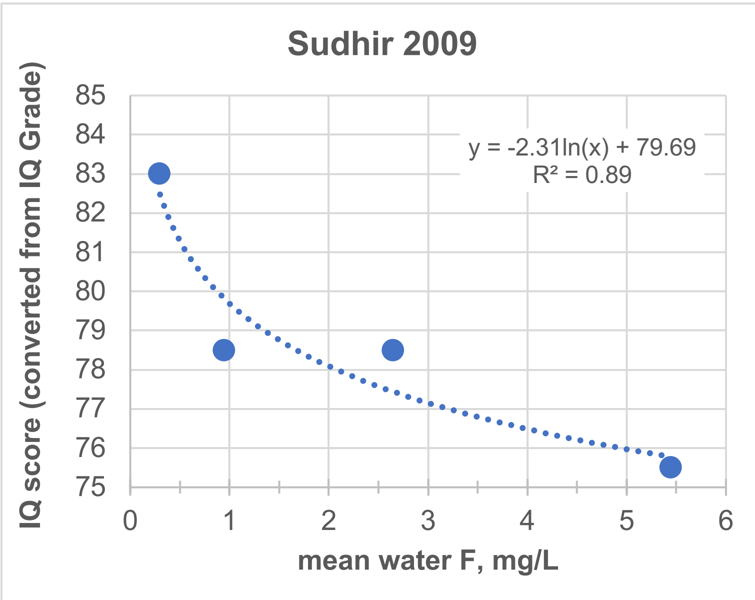
Grade 4.0 = 84.0 percentile IQ
Grade 4.1 = 83.0 IQ
Grade 4.2 = 81.5 IQ
Grade 4.3 = 80.0 IQ
Grade 4.4 = 78.5 IQ
Grade 4.5 = 77.0 IQ
Grade 4.6 = 75.5 IQ
Grade 4.7 = 74.0 IQ
Grade 4.8 = 72.5 IQ
Grade 4.9 = 71.0 IQ
Grade 5.0 = 69.5 IQ
Grade 5.0 = 70.0 percentile IQ

Then these percentile IQ were converted to IQ scores using:
https://www.edubloxtutor.com/iq-test-scores/

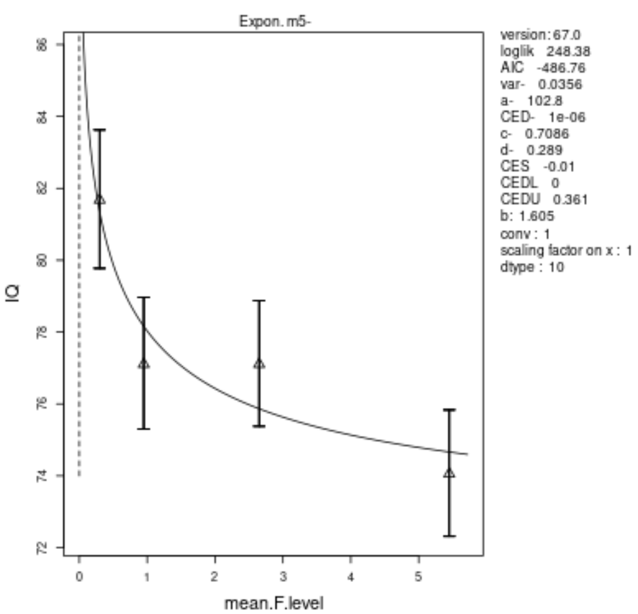
IQ	**Percentile**
65	01
70	02
75	05
80	09
85	16
90	25
95	37
100	50
105	63
110	75
115	84
120	91
125	95

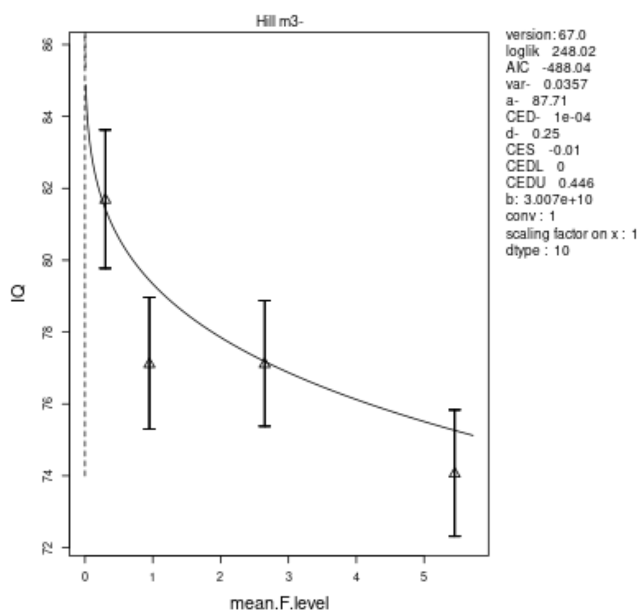
130	98
135	99

The four data points with mean IQ (converted from IQ Grade) are shown in this scatterplot:



BMD analyses of these four dose groups produced these results:





The estimated BMD is essentially zero, and so is the BMDL. There are a substantial number of observations below 0.7 mg/L and no evidence of a lower threshold. Given these Benchmark Dose results, this study was classified as having found an adverse effect **below 0.7 mg/L**.

Till 2020

This used the same MIREC cohort as Green 2019 and found statistically significant adverse effects in formula-fed infants drinking fluoridated water with average concentration of 0.6 mg/L compared to those formula-fed infants with non-fluoridated water having average water F concentration of 0.12 mg/L. There was no need to check if there was a threshold because any threshold must have been below 0.6 mg/L. Therefore, this study was classified as having found an adverse effect **below 0.7 mg/L**.

Trivedi 2012

Table 2. Drinking water and urinary F level of children living in low F villages and high F villages (Mean ± SEM)

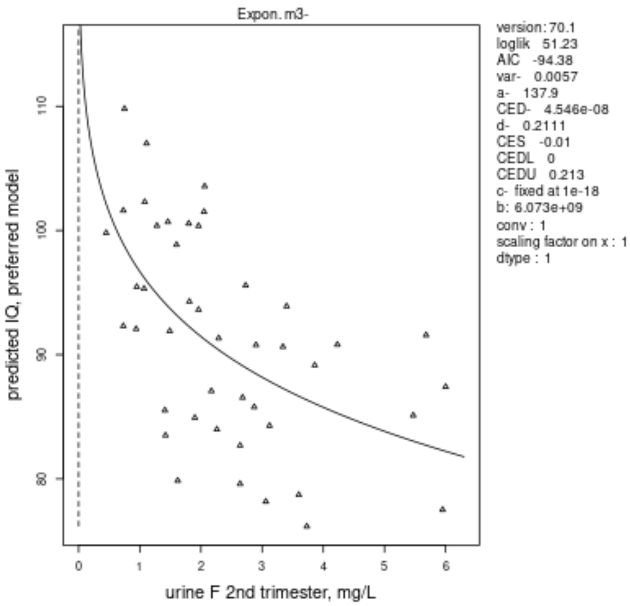
Village	Number of children examined	Level of F in GW drinking water (ppm)	Urinary F level (ppm)
Low F villages	50	0.84 ± 0.38	0.42 ± 0.23
High F villages	34	2.3 ± 0.87*	2.69 ± 0.92*

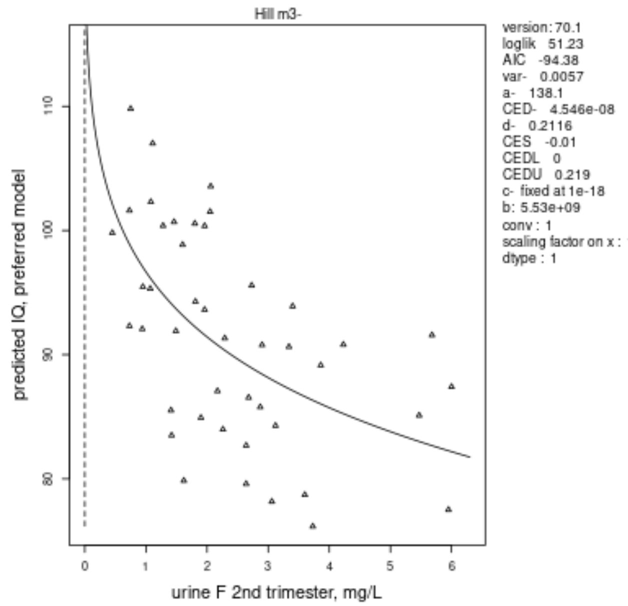
*p≤0.05 (Compared to low F contaminated villages)

This had only two exposure groups with the higher exposure group having mean water F of 2.3 mg/L and mean urine F of 2.69 mg/L. Therefore, this study was classified as having found an adverse effect **above 1.5 mg/L**.

Valdez-Jimenez 2017

Mean maternal urine F in 1st trimester was 1.9 mg/L and 96% had urine F greater than 0.65 mg/L. BMD analysis of author-provided individual-level data found BMD close to zero and there were sufficient number of observations below 1.5 mg/L to conclude there was no threshold. BMD analysis supports classification as below 0.7 mg/L, but there are not enough data points below 0.7 mg/L so this study will be classified as having found an adverse effect **below 1.5 mg/L**.





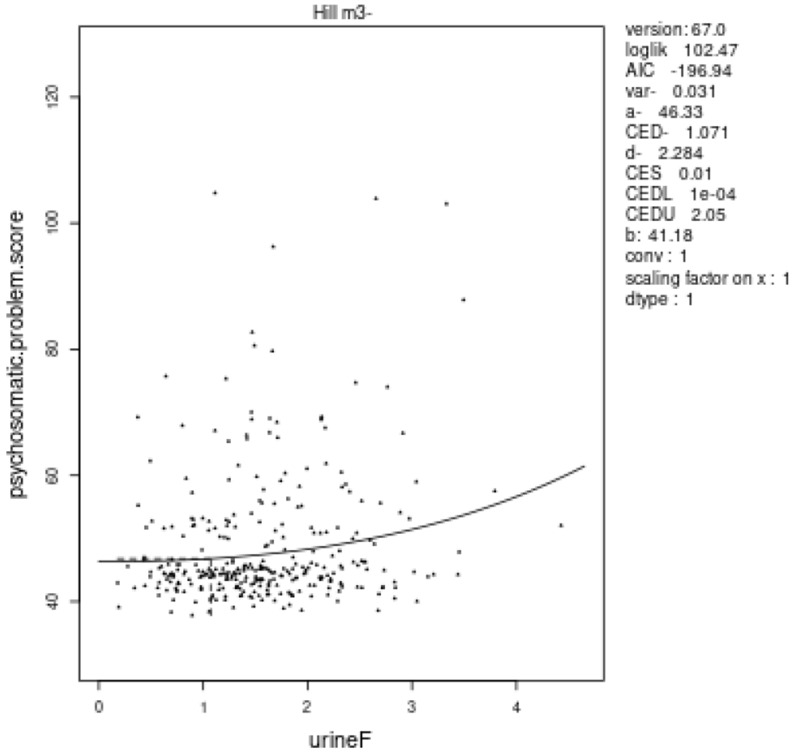
Wang 2012 (and Xiang)

This is from the Xiang group and is the same sample as used in Xiang 2003 and Xiang 2011. The exposure measure was an estimate of total F intake from drinking water, diet, and air. Exposure measure units were mg/d of F intake. Hirzy 2016 used this data and did Benchmark Dose analyses. Although Benchmark Dose analysis produces BMD of 0.3 mg/L water F for Xiang 2003 the lowest exposure group in the higher F village had mean water F of 0.75 mg/L so there were few children with less than 0.75 mg/L. About 15% of children in the high F village had water F below 1.5 mg/L, however, so this study is classified as having found an adverse effect **below 1.5 mg/L**. There was no evidence of a lower threshold.

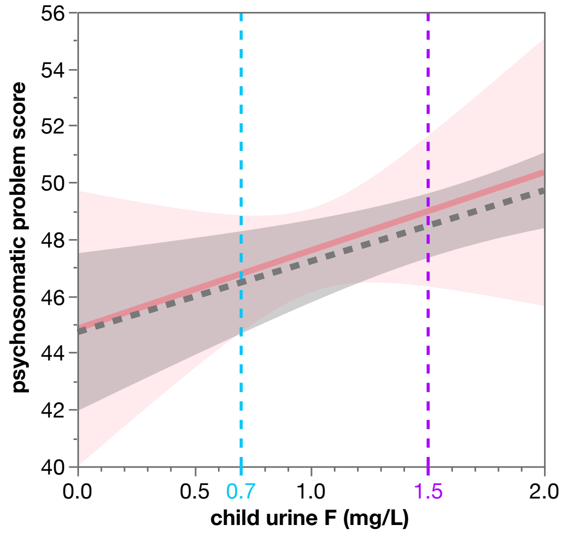
Wang, Anqi 2020a

BMD analysis of the outcome with statistically significant adverse effect (psychosomatic problems score of Conners Scales) found a BMD of 1.07 mg/L urine F and a BMDL of essentially zero. The best fitting model was the Hill m3- model although it was almost the same as the Exponential m3- model. A substantial proportion of the observations had exposures below the BMD but not below 0.7 mg/L so this study will be classified as having found an adverse effect **below 1.5 mg/L**.

Here is the graph of the best fitting model from the PROAST Benchmark Dose analysis. Different terminology is used from EPA. To translate: CED = BMD, CESL = BMDL, CES = BMR.



The next graph is of extracted data with linear dose-response model. For all data it reached statistical significance but for observations below 1.5 mg/L urine F, it did not reach significance. However, it did have the same effect size, suggesting it is reasonable to expect the dose-response relationship to apply over the full range of exposures, including below 1.5 mg/L, so any threshold would be below 1.5 mg/L.



> Wang2020a, Figure 1C extracted data
>
> Gray dashed line and shaded 95%CI zone is for all observations at all exposure levels. Red solid regression line and shaded 95%CI zone is restricted to observations with urineF ≤ 1.5 mg/L. All observations is statistically significant but ≤ 1.5 mg/L is not. However, they have virtually the same dose-response relationship (slope). The restricted sample may not reach significance because it is a smaller sample size.

Linear Fit, all observations

Linear Fit, all observations

psychosomatic problem score = 44.73 + 2.49*urineF

Summary of Fit

RSquare	0.028
RSquare Adj	0.025
Root Mean Square Error	10.32
Mean of Response	48.63
Observations (or Sum Wgts)	322

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	44.73	1.411878	31.68	<.0001*
urineF	2.49	0.822769	3.03	0.0026*

Linear Fit, observations with urine F ≤ 1.5 mg/L

Linear Fit, observations with urine F ≤ 1.5 mg/L

psychosomatic problem score = 44.85 + 2.75*urineF

Summary of Fit

RSquare	0.009
RSquare Adj	0.003
Root Mean Square Error	9.459
Mean of Response	47.65
Observations (or Sum Wgts)	160

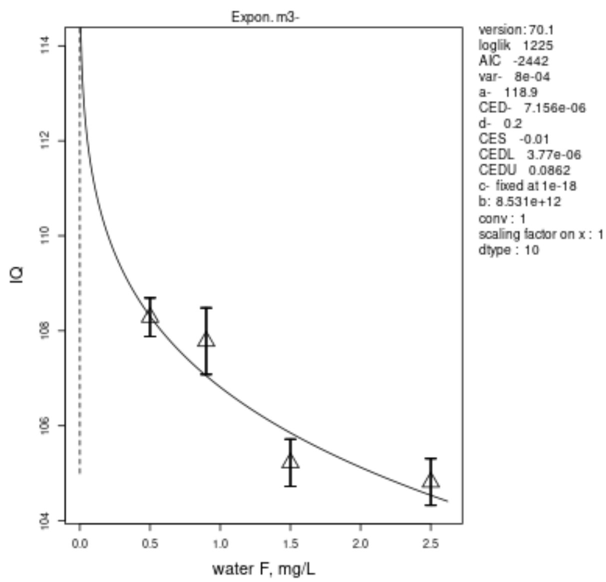
Parameter Estimates

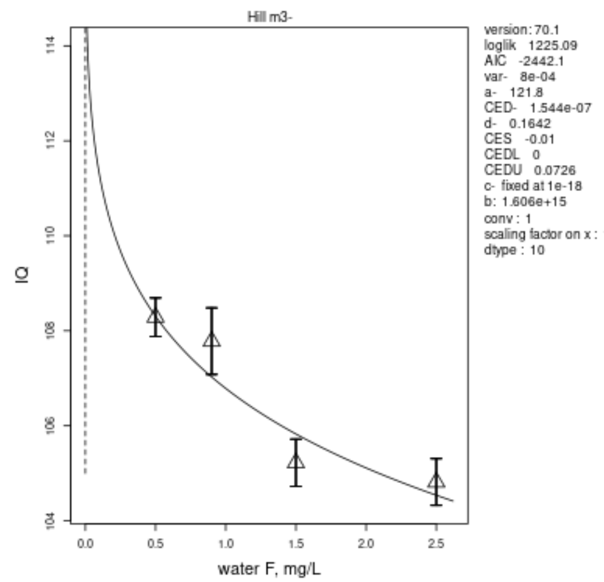
Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	44.846367	2.460294	18.23	<.0001*
urineF	2.75424	2.303674	1.20	0.2336

Wang, Mengwei 2020b

The median urine F was 0.40 mg/L and the 25th percentile was 0.15 mg/L. Although pairwise comparison between quartiles of urine F exposure were not significant, the *p*-value for trend for the association between urine F and IQ was highly significant (p=0.001), as was the regression model with exposure as a continuous exposure [Table 4]. BMD analysis using data grouped as exposure quartiles did not converge on a model due to high variance. However, Quartile 3 had a mid-point exposure level of 1.45 mg/L water F and was statistically significantly different from Quartile 1 which had water F ≤0.70 mg/L. Therefore, this study was classified as having found an adverse effect **below 1.5 mg/L**.

BMD analysis was possible but some assumptions had to be made, so it will not be the basis for classifying exposure level. Results of group-level BMD analysis, with exposure groups quartiles:





BMD and BMDL = 0.0

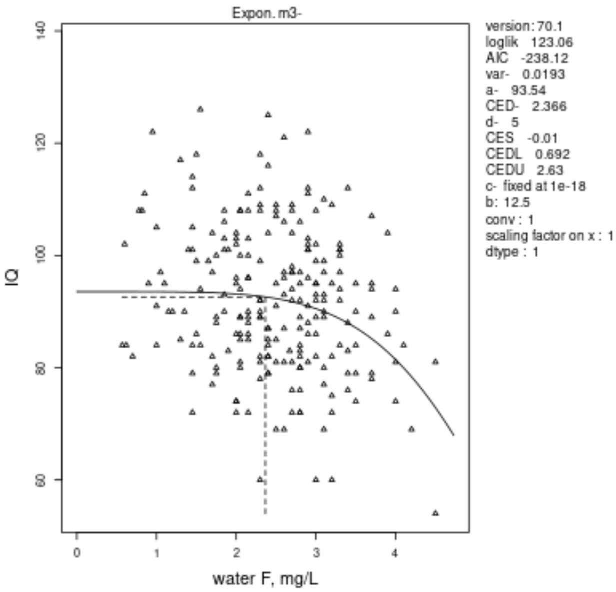
There are sufficient observations below 0.7 mg/L that this could be justified as finding an adverse effect below 0.7 mg/L.

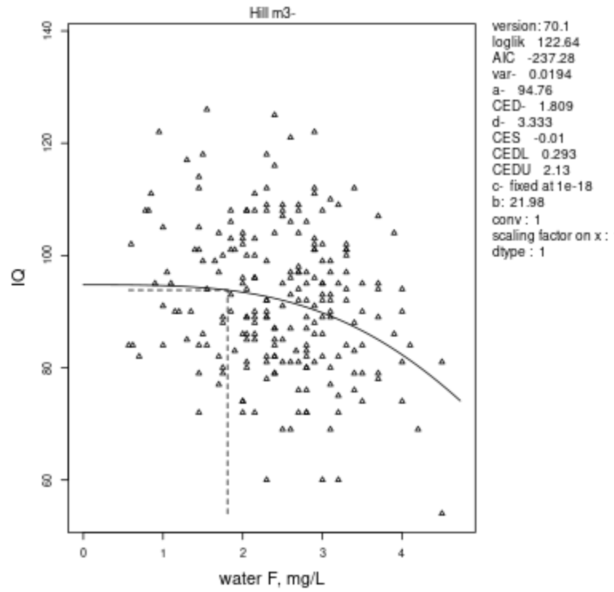
Xiang 2003

From individual-level data provided by authors we conducted BMD analyses. Our primary analysis was restricted to just the higher-F village to avoid any questions about hierarchical effects due to two villages with different F levels. The higher-F village had a wide range of F exposures so the individual-level data was amenable to BMD analysis.

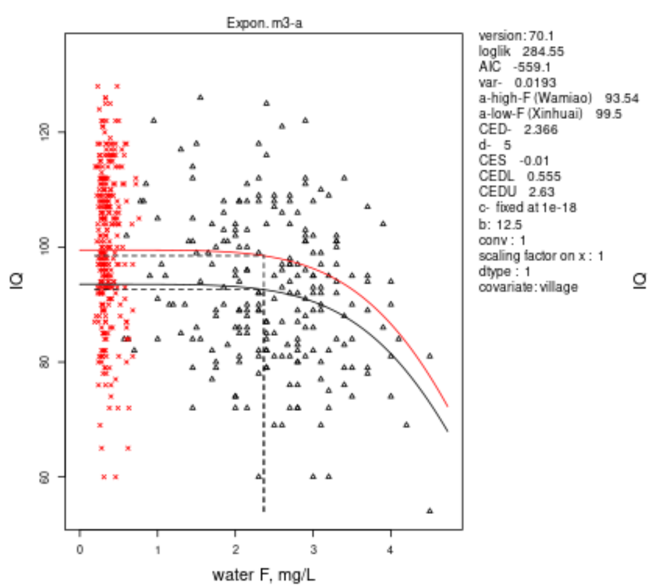
The BMD in the better fitting model (Hill m3-) is 1.81 mg/L water F, with a BMDL of 0.29 mg/L. There appears to be a threshold at roughly 1.5 mg/L so this study will be classified as having found an adverse effect **above 1.5 mg/L**.

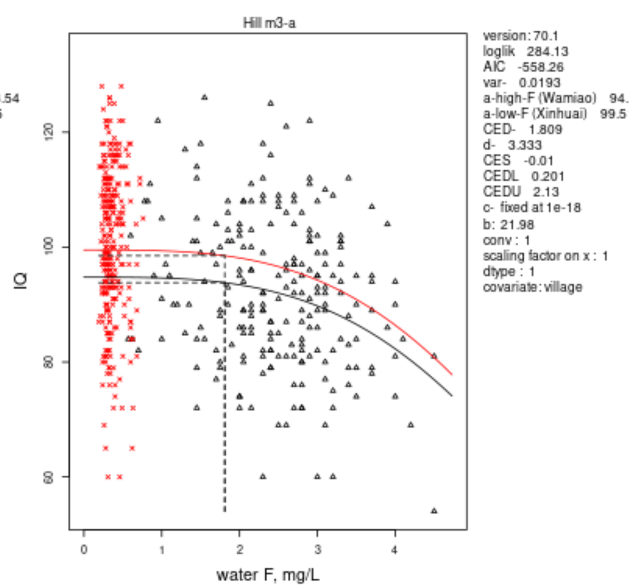
BMD for higher F village only (Wamiao):





BMD analysis for both villages (red is lower-F village, black is higher F village):





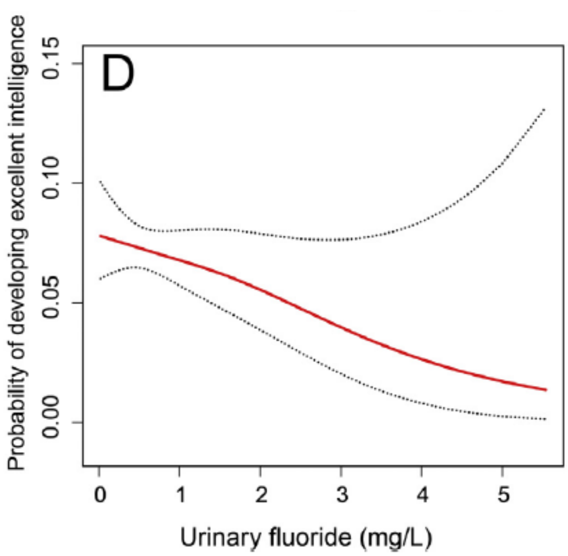
Xiang 2011

This study used the same sample as Xiang 2003 but instead of using water F as the exposure measure, it used serum F as the exposure measure. Since it is not possible to directly convert serum F to water F or urine F, this study will be assigned an exposure classification the same as Xiang 2003, as having found an adverse effect **above 1.5 mg/L**.

Yu 2018

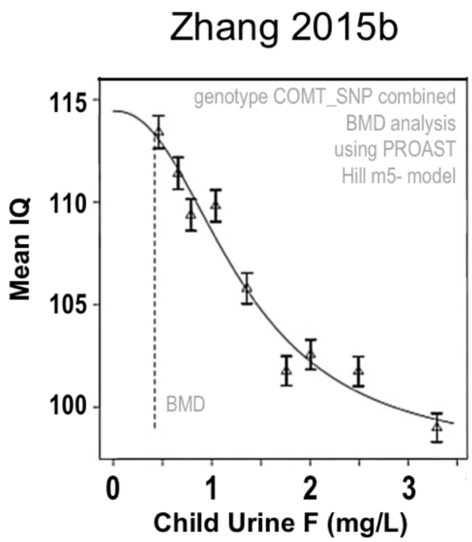
Table 1 shows that urine F was about 20% lower than water F for same exposure groups. Table 5 shows statistically significant reduction in probability of having Excellent intelligence in those with water F between 0.2 and 1.4 mg/L. This was based on piecewise linear regression for this range of exposure values. Given that urine F would have been about 0.2 to 1.2 mg/L in this group it is reasonable to expect that if a BMD analyses were done it would find a BMD below 0.7 mg/L and a large proportion of the children would have urine F below 0.7 mg/L. Therefore, this study was classified as having found an adverse effect **below 0.7 mg/L**.

LOESS regression gives further support that a BMD analysis would find BMD below 0.7 mg/L, as shown in this graph of the LOESS regression of the adjusted model:



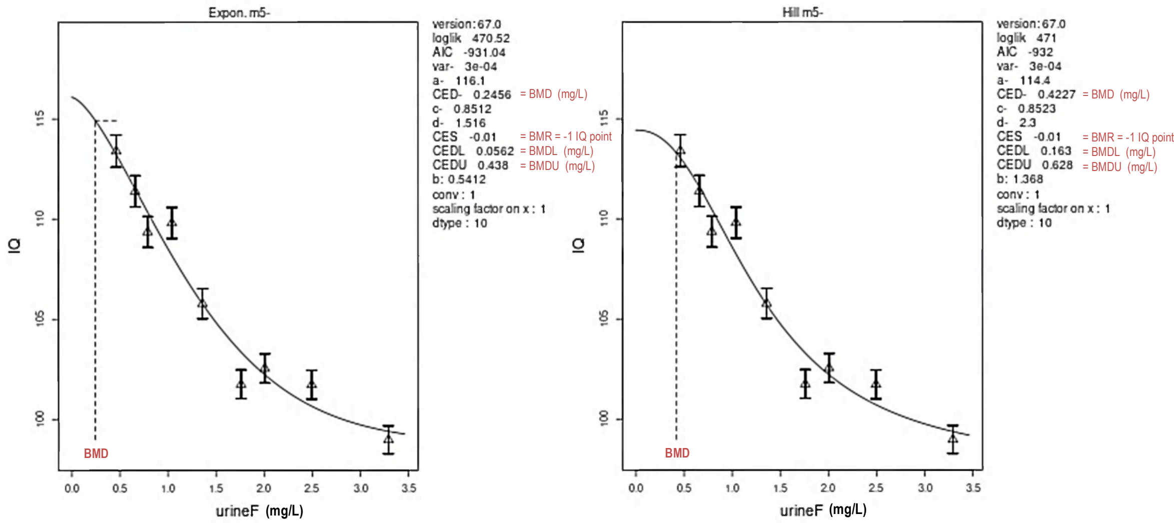
Zhang 2015b

BMD analysis shows BMD of about 0.4 mg/L child urine F, and about 20% of children with urine F below 0.7 mg/L. Therefore, this study was classified as having found an adverse effect **below 0.7 mg/L**.



Original PROAST output graphs show Hill m5- has lower AIC so preferred model, although Expon m5- is within 2 of it. For Hill m5- model the BMD = 0.42 and BMDL = 0.16. Note that this is not for just the gene variant of COMT, but for all children in study.

Benchmark Dose (BMD) Analysis of Zhang 2015, IQ vs urine F extracted from Figure 1.
PROAST online BMD method:
https://proastweb.rivm.nl

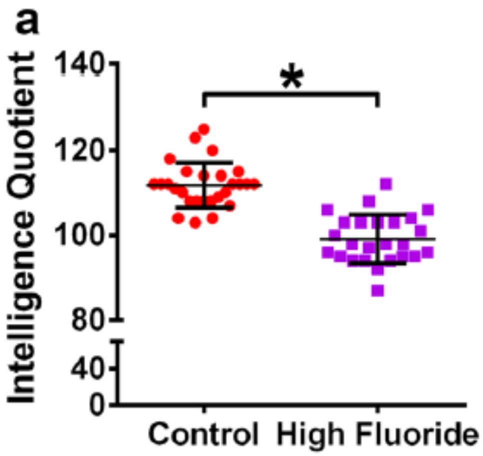


Zhao 2019, Zhao 2020, and Zhou 2019

These three studies by the same group are very similar. Zhao 2019 appears to have used a smaller sample of children than Zhao 2020 but there may be extensive overlap in the participants. Zhao 2019 reports having 27 children from the reference F area and 25 from the higher F area. Zhao 2020 reports having 30 children from both the reference and higher F areas. Zhou 2019 reports 25 children from each area. The description of the participants is otherwise identical with respect to city, year of recruitment, and age.

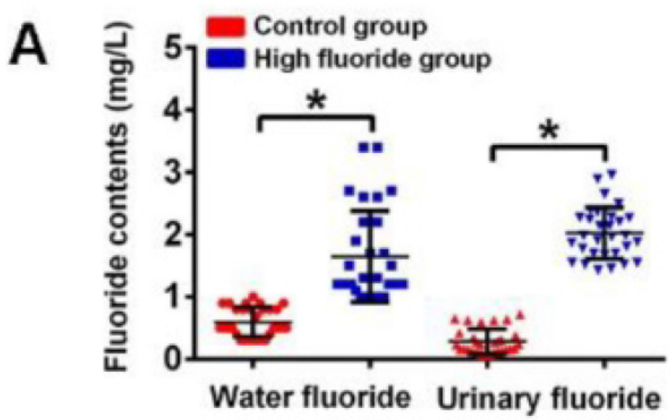
Zhao 2020 and Zhou 2019 report more details of the association between IQ and F and it will be assumed their findings apply to Zhao 2019 because they are presumed to be mostly the same participants.

Zhao 2019 results are only reported in this figure:



This is Figure 7a and it shows a statistically significant difference in mean IQ between the reference and higher F group. The paper did not report whether the exposure measure was water F or urine F or any other details about the exposure.

However, in Zhao 2020, additional details are reported in this Figure 8A:



Both water F and urine F were measured in Zhao 2020. For water F, more than half the participants in the higher F group had levels below 1.5 mg/L and all of the reference group had levels below 1.5 mg/L.

The association between IQ and F is shown in Figures 8B and 8G:

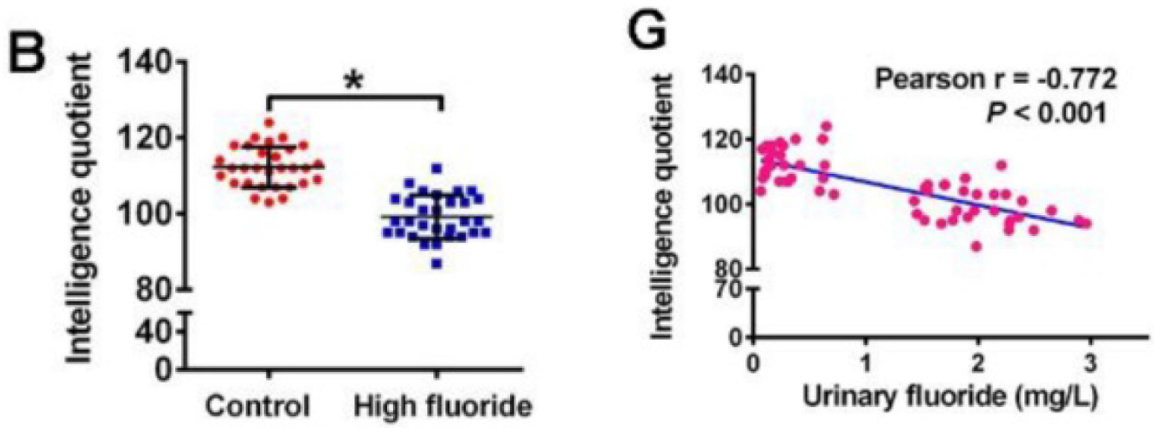
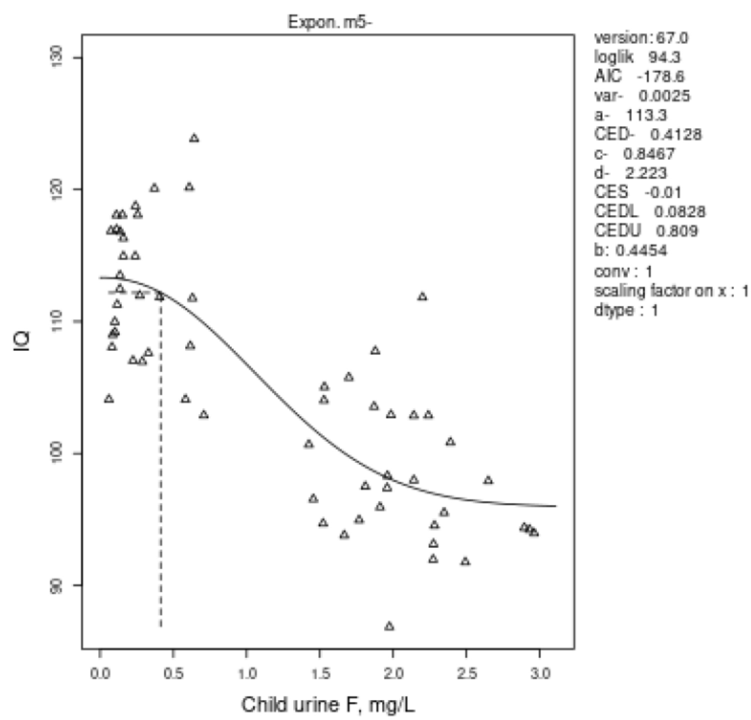


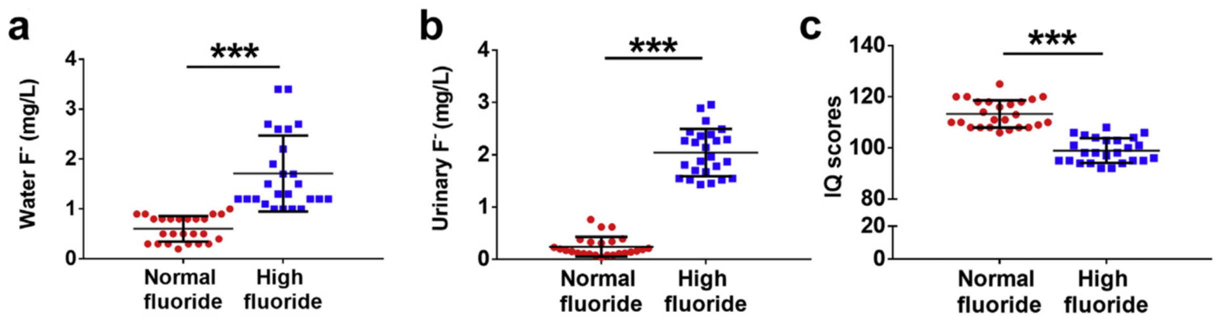
Figure 8G shows a statistically significant linear dose-response with almost half of participants having urine F below 0.7 mg/L. Since Figure 7A from Zhao 2019 shows water F also had a substantial proportion of participants below 0.7 mg/L, the Zhao 2019 results are assumed to be equivalent to Zhao 2020 with respect to dose-response below 0.7 mg/L. Therefore, Zhao 2019, Zhao 2020, and Zhou 2019 are each classified as having found an adverse effect **below 0.7 mg/L**.

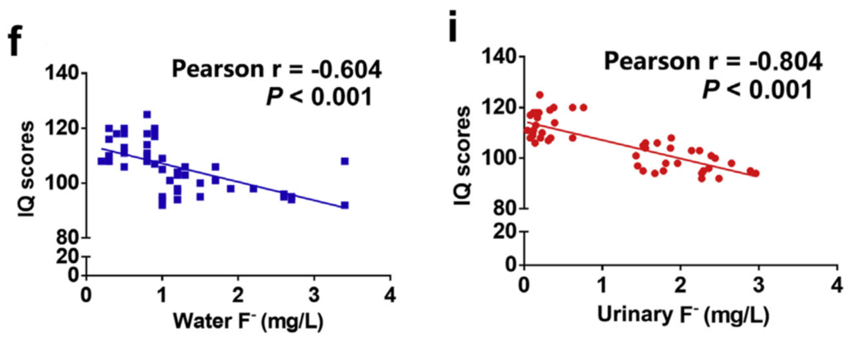
Further support for this classification comes from Benchmark Dose analysis of Figure 8G, the relationship between urine F and IQ:



The BMD is 0.41 and the BMDL is 0.08 mg/L. These are well below 0.7 mg/L and as can be seen, about half of all observations are below 0.7 mg/L.

Zhou 2019 provides more results although for 25 participants in both areas rather than 30 in each as reported by Zhao 2020. Here is Zhou 2019 Figure 6a, 6b, and 6c:

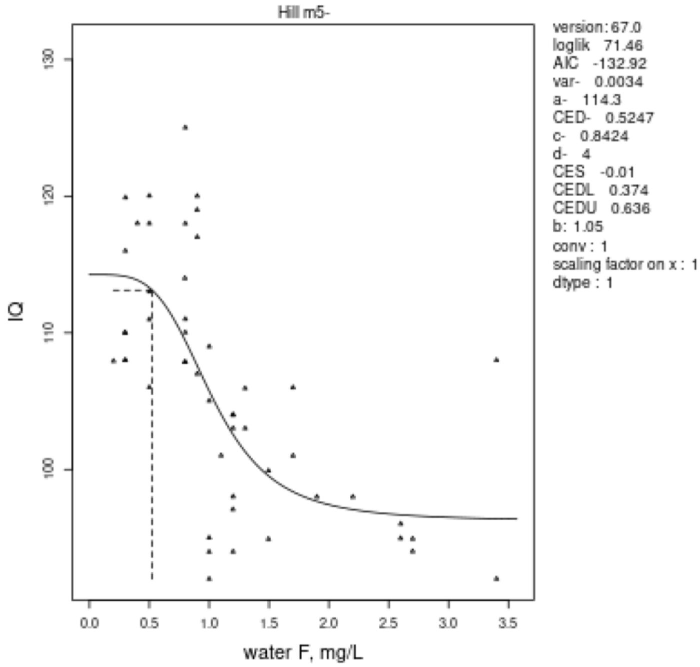




Zhou 2019

See above under **Zhao 2019, Zhao 2020, and Zhou 2019**. Zhou 2019 appears to have mostly the same participants and data as Zhao 2019 and Zhao 2020. However, it reports additional information, in particular, the association between water F and IQ. BMD analysis using digitally extracted data from Figure 6f produced a BMD of 0.52 and BMDL of 0.37 mg/L water F. Both of these are below 0.7 mg/L and almost a quarter of participants had water F of 0.7 mg/L or lower. This produces additional support for classifying Zhou 2019 as having found an adverse effect **below 0.7 mg/L**.

BMD analysis results:


Hill m5-
IQ
water F, mg/L
100
110
120
130
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
version: 67.0
loglik 71.46
AIC -132.92
var- 0.0034
a- 114.3
CED- 0.5247
c- 0.8424
d- 4
CES -0.01
CEDL 0.374
CEDU 0.636
b: 1.05
conv : 1
scaling factor on x : 1
dtype : 1