ERROR REPORT #4

Report by Matthew B. Jané, James Heathers, and David Robert Grimes

Study Title: Fluoride Exposure and Children's IQ Scores: A Systematic Review and Meta-Analysis

Journal: JAMA Pediatrics

DOI: 10.1001/jamapediatrics.2024.5542

Citations: 9

Summary: A recent meta-analysis in this journal Taylor et al (2025) produced ostensible evidence of a negative correlation between fluoride exposure and intelligence in children, which garnered major public attention. There has, however, been criticism of this undertaking and conclusions. In order to address the controversy over this work, we undertake an independent forensic meta-science review of the methodology and statistical approaches employed and inferences drawn by the authors, as well as an investigation of the underlying data integrity of the constituent studies. We find that the authors employed unjustified methodological and statistical errors which invalidate their conclusion, and demonstrate that the data cannot be analysed as the authors assert. We further find major problems with the sources employed, including reliance on studies from non-MEDLINE indexed publications with an anti-fluoridation editorial stance, and major underlying issues with the data reported in several instances, indicative of impossible or unreliable data. Taylor et al is not reliable nor are its errors remediable. It should be retracted to avoid harms to public health and scientific discourse.

Table of contents

See official report and materials on OSF . . . 3
Introduction . . . 3
Methodological Issues . . . 3
Statistical Issues . . . 4
Data Integrity from Constituent Studies . . . 7
Discussion . . . 9
Conclusion . . . 10
Supplementary materials . . . 10
References . . . 10

See official report and materials on OSF

This is essentially repost of the official (pre-print) report which can be found here: osf.io/preprints/osf/zhm54_v3. The code and analysis can be found in the associated OSF project here: osf.io/wju6r/. The pre-print has been submitted to *Meta-Psychology* and you can participate in the open peer-review process here: tinyurl.com/mp-submissions.

Introduction

Evidence to date strongly supports water fluoridation as a means of reducing dental cavities, and in the United States a level of 0.7 mg / L is recommended for optimum dental health. However, fluoride has long been the subject of conspiracy theories and a target for health disinformation. A recent meta-analysis in JAMA Pediatrics by Taylor et al (2025) purports to conduct a meta-analysis of studies published on fluoride and intelligence quota (IQ), concluding that fluoride exposure appears associated with a decrease in IQ.

This work has been criticized previously in several forums. Previous iterations have been criticised by the National Academy of Sciences for major shortcomings in inference and methodology (National Academies of Sciences 2020 & 2021), in the popular press (Oza 2025, Mole 2025), on the JAMA Pediatrics forum for a multitude of issues, and even in an editorial released to coincide with Taylor et al's work (Levy 2025). These communications highlight many deep issues with the work, including its reliance on highly biased studies and questionable selection criteria, as well as curious absence of more recent larger and more powerful studies from Spain and Australia finding no link between water fluoridation and intellectual attainment (Ibarluzea et al 2022, Do et al 2023).

These flaws, while damning, do not fully explain the extent of the problems with the text. In this special communication, we outline why the conclusion Taylor et al. reached is not justified by the data; the methodological and statistical choices made are unjustified, and the underlying data in the meta-analysis contains several results which are, at a minimum, untrustworthy.

We further demonstrate that this study is unreliable for meta-analysis, that its peer review was inadequate, and that the decision to publish the manuscript by JAMA Pediatrics was likely ill-judged and ought to be revised.

Methodological Issues

Inclusion of questionable publications

Of the 74 studies Taylor et al include at least 21 publications (28.4%) in their meta-analysis from the publication Fluoride, a publication run by International Society for Fluoride Research Inc. Its current and previous editors are public anti-fluoride activists, and it is completely unaffiliated with any professional body, scientific publisher, or academic institution. It is not MedLine indexed, and so it is clear that Taylor et al should have factored this into their risk of bias assessment (eSupp1). We show how the reported effects of fluoride differ in the journal Fluoride vs other journals in the section on statistical and data issues.

Unadjusted confounders and lurking variable problem

The bulk of the underlying studies contained within Taylor et al. (2025) are cross-sectional studies conducted on different populations in different geographic regions. Studies like this are extremely non-generalisable if no adjustment is made for confounders nor any form of matching conducted. Many of the included studies are no more detailed than the following: Area A has low fluoride water levels, Area B has high fluoride water levels, and Area A has higher mean IQ.

However, Area A might also have adequate childhood nutrition, accessible pediatric vaccination schedules, no measurable nutritional deficiencies, no environmental exposure to pediatric toxicants, etc. and Area B none of the above. Underlying studies of this nature are absolutely inappropriate for determining any causal relationship or even association between IQ and fluoride levels. Meta-analyzing many studies does not eliminate or correct the fundamental design issues, meta-analyses are only as good as the studies contained within them. These studies were not designed to answer the question: "does Fluoride ingestion cause an IQ decrease?". So even if this meta-analysis was rigorously performed (which we show in the next section was not the case) it still would not answer the question that we want it to.

Inclusion of suspect measures of fluoride levels

As other authors have covered before, urine analysis and particularly measurements of fluoride in maternal urine as an attempted proxy measure of foetal or childhood exposure is a wholly unreliable measure of fluoride ingestion, and cannot be used to make inferences due to its abysmal replicability (Guichon et al 2024, Levy 2025). The problems with spot analysis of urine for fluoride levels are manifold, and include issues with assay sensitivity and specificity, no correction for recent dietary fluoride intake (Riddell et al 2021). In this meta-analysis however, the authors ignored this issue and several of the included papers using this questionable approach as comparable with the standard metric, mass fluoride per unit mass drinking water. This is not the first time such a problem has been flagged in a JAMA pediatrics paper, with one controversial example arising in 2019 (Guichon et al 2024). In total of the 74 included studies, 18 (24.3%) used unreliable urine analysis of fluoride levels only, 9 (12.2%) had no measured fluoride level, 2 (<0.1%) used proxies like grain and wood fluoride levels, while 25 (33.8%) used standard water fluoride levels and a further 20 (27%) used both urine and water fluoride levels.

Statistical Issues

Inappropriate metric for mean-effects

When attempting an analysis for mean-effects, the authors compare low fluoride reference groups to higher exposure groups, using Standardized Mean Differences (SMDs) to normalize results for comparison. This fails on two levels. Firstly, neither 'low' (reference) nor 'high' (comparator) groups are ordinal classes and vary substantially between studies, rendering direct comparisons of SMDs fundamentally misleading. As seen in figure 1, there are stark differences in low/high across studies. In fact, some study's 'high' (exposure) groups have less fluoride than other study's 'low' (reference) group. Secondly, SMD is not readily interpretable in this application, indicating only the spread of individual study data. Since the reference and exposure groups are arbitrary, the effect size is also arbitrary. Figure 2(a) illustrates that if you have some true non-zero correlation between fluoride and IQ, then the SMDs between references can drastically change the effect size. That is, if the difference in Fluoride concentration between reference/low and exposure/high groups

In one case, the authors appear to have mistakenly inputted raw scores from a Ravens standard progressive matrices in lieu of IQ, erroneously entering a group mean IQ of 13.39 (Razdan et al 2017, cited in eTable1 of Taylor et al 2025). This clearly nonsensical data yields an SMD comparable to those reported in Taylor et al's Table 1 (-4.45), highlighting the unsuitable nature of the metric for comparison.

In studies that had more than two exposure groups, the authors only took the highest and lowest exposure group for the comparison and thus dropped groups in between. This creates multiple problems, first it causes inflation in the effect size due to direct range enhancement, that is, by taking cases at the extreme end of a variable the relationship with any other variable becomes exaggerated. Second, it decreases the precision of the effect size by unnecessarily reducing the sample size. Nor are the entries truly comparable in terms of SMD, nor even consistent with one another: for just one example, from eTable 1 again, we can see the raw data for Zhao et al (1996) with Zhang et al (1998) as shown in table 1. Notice that the “low” defined by the authors in Zhao et al is in fact 13.8% higher than the “high” Taylor et al define in Zhang et al. This is an insurmountable issue with the authors approach and, accordingly, all SMD can capture is the variability of the underlying data with no reference to the actual levels (Cummings 2011), not a base measure of any effect.

Table 1: An example of conflicting dichotomization in Taylor et al. (2025)

Study	Fluoride Exposure defined as “Low”	Fluoride Exposure defined as “High”	Fluoride Range
Zhao et al (1996)	0.91 mg / L	4.12 mg / L	3.21 mg / L
Zhang et al (1998)	0.58 mg / L	0.8 mg / L	0.22 mg / L

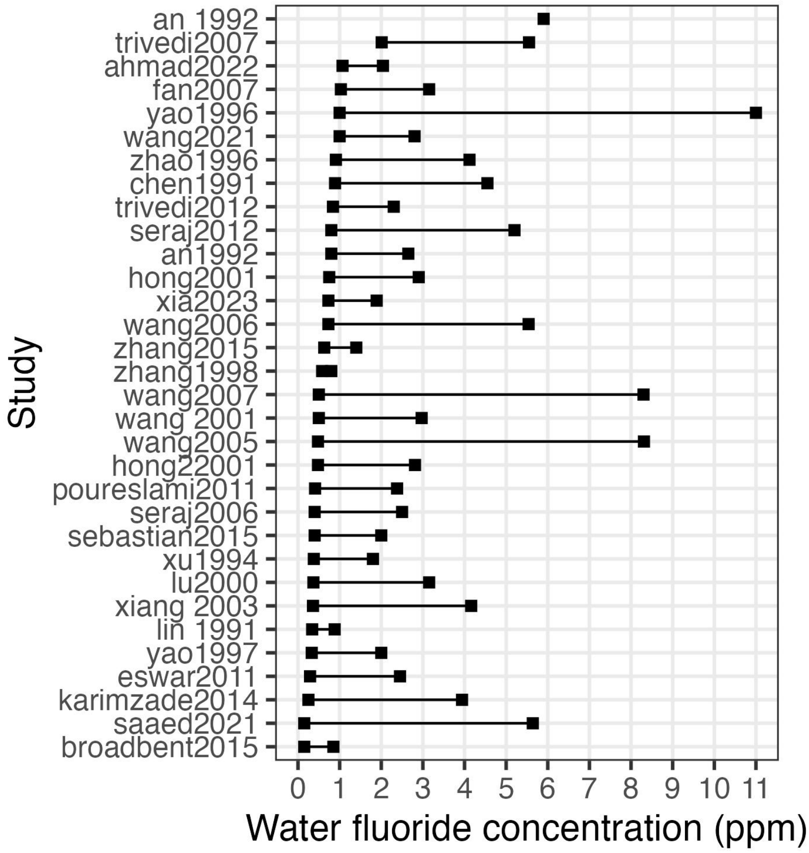

Figure 1: Reference and exposure fluoride concentration values for each study in the meta-analysis. left dot always indicates reference/’low’ group and right dot always indicates exposure/’high’ group

Inappropriate methodology for investigating dose response

The use of dose-response meta-analysis for mean differences is likewise suspect, because it pivots on comparable baselines. These studies very obviously do not have such a baseline, with reference ‘low’ levels varying markedly between studies. The arbitrary cut-off also lends itself to redaction bias (Grimes and Heathers 2021) likely to skew results. For instance, 1.5mg/L is

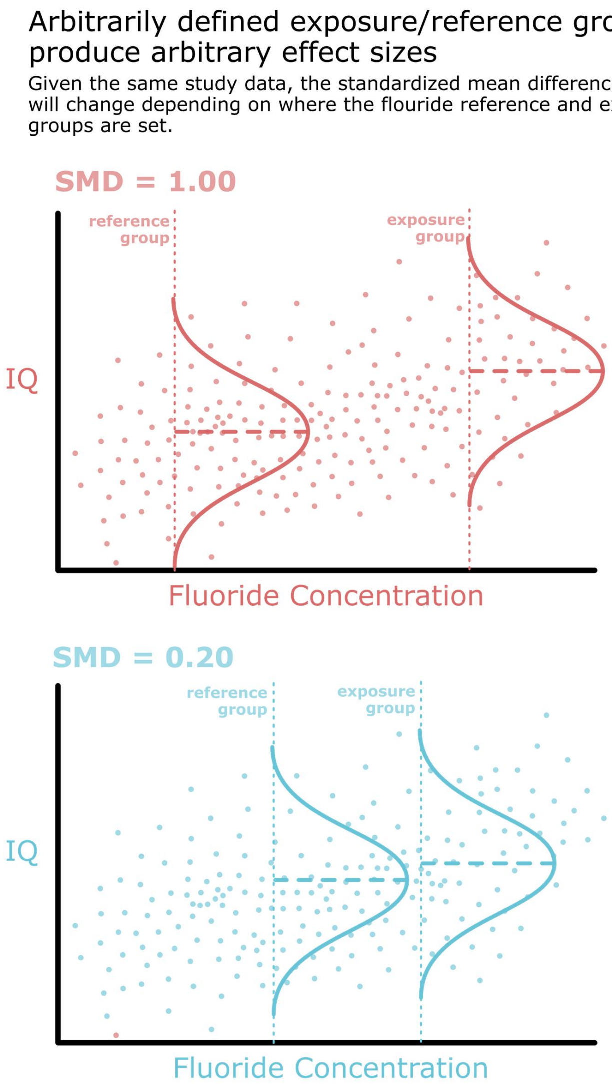

Figure 2: Identical IQ and fluoride data with two different SMDs superimposed. Top panel: When the exposure and reference group are far apart, the SMD can be arbitrarily large assuming that IQ monotonically increases with fluoride concentration. Bottom panel, when the exposure group similar to the reference group than the SMD is arbitrarily small. This can all be true even given an identical correlation between fluoride and IQ.

not a level that has any toxicological relevance, nor is any other justification given for choosing it.

Critically, both fluoride exposure level and IQ are continuous measures. Given that, a more reasonable approach would have been to employ a meta-regression to examine potential relationships between IQ means and fluoride concentration across study samples. We confined our analysis to the 45 (60.8%) studies which reported standard water fluoride levels and IQ scores, but not all of these were usable. Of these, 4 arbitrarily dichotomized the continuous measure of water fluoride level into highs or lows, a well-known error in data-handling that can lead to erroneous results (Altman 2006, Grimes and Heathers 2021b). Another 6 were excluded because they did not report at least one level or used a regression without clear raw data, with a further study (Razdan et al 2017) excluded for lacking data required to convert its metric to a standard IQ score. A final study was misreported by Taylor et al (Khan et al 2015) and did not directly link IQ and fluoride levels and was also excluded from the meta-regression.

This left 31 usable studies, consisting of 72 data pairs. As we have multiple samples within each study, we also treat study as an additional random effect to account for within-study dependence. Figure 3(a) shows the result of this meta-analysis showing a still negative and significant relationship between fluoride concentration and IQ (beta = –1.31, SE = 0.398, p = .012). However, the effect seems to be dependent on whether the study was published in the journal Fluoride (see Figure 3, right panel). The interaction effect between fluoride concentration and journal was strong (beta = –2.51, SE = 0.786, p = .010).

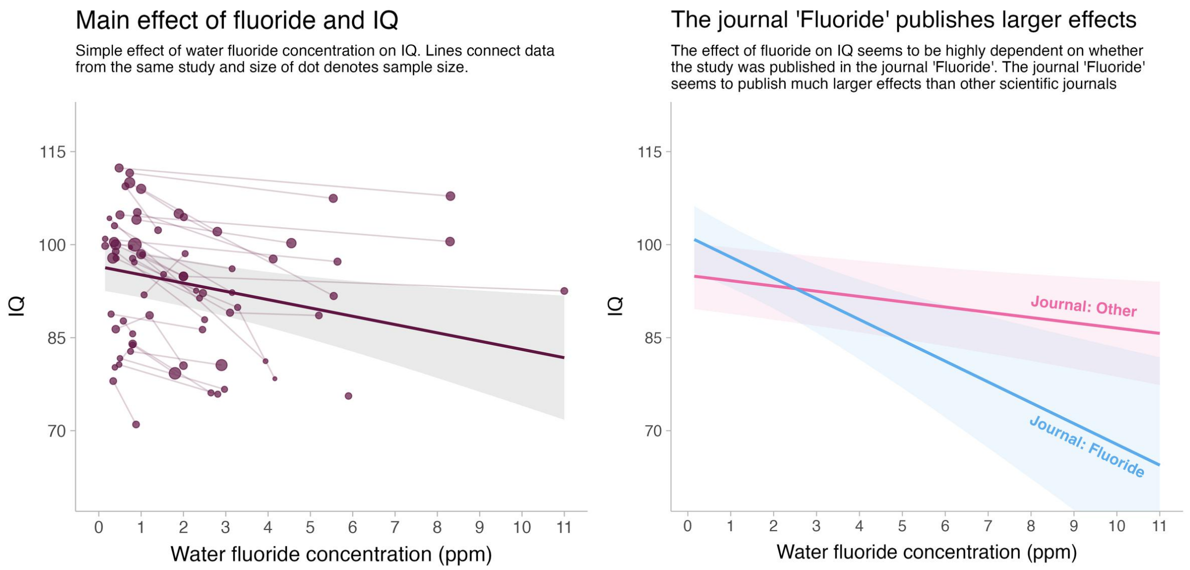

Figure 3: Left panel shows a meta-regression model estimating the relationship between fluoride concentration (horizontal axis) and IQ (vertical axis). Right panel shows another meta-regression model with a full interaction with whether the study was published in Fluoride. All shading around the regression lines denote the 95% confidence interval.

Data Integrity from Constituent Studies

Other authors have reported major issues with how Risk of Bias was calculated by Taylor et al., and that more reliable studies were left out of the presented meta-analysis. The problem is worse than envisioned: some of the included studies are extremely suspect in their own right. The following were identified simply by looking for the most extreme results on the included forest plot in Figure 1 and within the rest of the manuscript, and are not exhaustive, but encompass serious warnings of violated data integrity Taylor et al should have noted in a systematic review.

Examples of impossible statistics

Xu et al. (1994). This study has extremely unrealistic data. Chart 1 (reproduced in the supplementary material) for the “Low fluoride high iodine” group reports an IQ SD of 0.92 (N=32) in stark contrast to population SD of 15, and an extremely small SD of 2.25 in the “High fluoride” group. Another peculiar result is that the “High-fluoride low iodine” group has a mean IQ of 69.40 (intellectual disability criteria threshold is IQ=70) and a standard deviation of 20.40, suggesting there are some individuals in this group that have extraordinarily low IQs.

The next chart in this paper (see supplementary material) shows the distribution of IQs among groups, with groups not quite aligning with those in chart 1, evidenced by the discrepancy in sample sizes. Chart 2 however shows fairly typical IQ distributions where the smallest possible standard deviation for each group is given in supplementary table 1. In the text, the authors claim “there are major differences in results between a region of low iodine and fluoride, and one that only has a low level of iodine (P<0.01)”. However, based on Chart 1, an independent samples t-test between the low iodine and low fluoride group (mean=76.42, SD=7.12, N=27) and the low iodine group (mean=75.17, SD=14.16, N=62) does not show a significant difference (p-value = .667).

Zhang et al. (2015). This paper was extremely difficult to access, as it is in a journal with no DOI, no common database access, and is written in Mandarin Chinese. However, the data is unambiguous and the main results comparing IQ between fluoride exposure groups reproduced in the supplementary material. All of these results are impossible as stated, as the reported groups cannot produce the standard deviations listed. The minimum possible standard deviation for Line 1, for example (m=90.52, sd=10.37) occurs when every category member is as close as possible to the mean (for instance, if every child in the 80-89 IQ group has IQ 89). This forms an extremely unlikely dataset but one representing the lower bound of the standard deviation (Brown and Heathers 2017). The minimum sample standard deviation of the above is 13.75, and thus Line 1 is impossible, as is every other line as elucidated in table 2.

Table 2: Impossibility of data reported in Zhang et al 2015.

Line	Stated SD	Lowest possible SD
1	10.37	13.75
2	11.24	14.71
3	12.43	15.69
4	11.52	12.87

Examples of extreme and unrealistic effects

Khan et al. (2015). It has been previously noted that papers sometimes hide extremely small and questionable p-values results when they are rounded up to the stated reporting limits (i.e. p=0.001) or given as less than those limits (i.e. p<0.001) (Heathers, 2025). This obscures p-values which are unusual or unrealistic when presented in manuscripts. It is a simple distortion but one which can be easily recalculated as outlined in the supplementary material and provided code. Table 4 of Khan et al. (2015) is a 2x5 matrix (location and IQ group) for a total of n=429 children. It describes the result of a chi-squared analysis as (p<0.001). This returns a chi-squared value unstated in the paper (chi=173). The p-value is 2.7e-36, that is, p=0.0000000000000000000000000000000000027. This result is extremely unrealistic and should not be trusted without qualification. Using an analytical approach outlined in the supplementary material, these results in this table would also imply an IQ drop on average of 5.4 IQ points per additional part per million fluoride. This would render water fluoridation on par with a traumatic brain injury (Ewing-Cobb 2006), which is both patently outlandish and demonstrably false (see supplementary appendix for code).

Razdan et al. (2017). Similar to the above, Table 1 of Razdan et al. (2017), is a 3x5 matrix that tabulates (geographic location) x (IQ group) for a total of n=219 children, where geographic location is indicative of local fluoride levels in the drinking water. It describes the result of the chi-squared analysis as p=0.001. This returns a similar chi-squared value to the paper (chi=186.5587). The p-value is 4.3e-36, that is, p=0.0000000000000000000000000000000000043. This result is extremely unrealistic and should not be trusted without qualification. Equally concerning is that for this ostensible difference in IQ to be attributable to fluoride levels, simulation and analytical solutions suggest that this would imply a typical IQ drop of 5.8 IQ points every additional part per million increase, again an unbelievable figure.

Discussion

Trustworthy medical science is crucial to keeping society healthy and guiding policy. Conversely, unreliable results confound the scientific record, and undermine public health. This is especially evident when the work in question becomes the subject of intense public interest. Andrew Wakefield's discredited work in the Lancet insinuating a link between the measles-mumps-rubella vaccine and autism, for example, led to vaccine confidence crises in Western Europe after its appearance in 1998, and even now almost three decades later continues to cause quantifiable harm to public health and understanding (Godlee, Smith, and Marcovitch 2011, Motta & Stecula 2021, Grimes 2022). In the case of fluoridation, there have been decades of conspiracy theories around it, most famously parodied in the 1963 Stanley Kubrick movie Dr Strangelove, by which stage such unfounded theories were already decades old (Mausner & Mausner 1956, Glitch and Booth 2014).

These old and unfounded fears have re-emerged in public discourse in recent years, and have been incredibly politicised. As Loc Do, an epidemiologist and dentist told Statnews, Taylor et "doesn't add any new information. But the issue has become politicized. Another problem is that some people play with data differently to fit their objectives" (Oza 2025) Perhaps most egregiously, the work in question was referenced obliquely by Robert F Kennedy Junior in his senate confirmation hearings as vindication, when he stated "I was called a conspiracy theorist because I said fluoride lowered IQ. Last week, JAMA published a meta-review of 87 studies saying that there's a direct inverse correlation between IQ loss.". As we show in this work, the ostensible evidence is flawed, and hardly supports his questionable policies on this nor anything else. That it has the imprinteur of a respected journal however, gives the position a legitimacy it does not deserve, a problem seen previously in questionable reanalysis of COVID vaccines, also shown to be flawed and highly politicized (Grimes 2025).

Were this merely yet another piece of research waste, it would be merely unfortunate, but its high altmetric scores are indicative of its likely impact on public trust and understanding. Were the errors in the work salvageable with a correction, we would recommend this. But as this analysis shows, it is flawed on multiple levels from inception to execution, and cannot simply be fixed. This combination of factors motivated this special communication, and in light of these criticisms, we can only recommend blanket retraction to mitigate its potential harms on public health not only in the USA, but worldwide.

Conclusion

Taylor et al. (2025) is not reliable nor are its errors remediable. It should be retracted.

Supplementary materials

See all supplementary materials, data, and code in the OSF pre-print osf.io/preprints/osf/zhm54_v3 and the associated project osf.io/wju6r/.

References

Altman, D. G., & Royston, P. (2006). The cost of dichotomising continuous variables. Bmj, 332(7549), 1080.

Brown, N. J., & Heathers, J. A. (2017). The GRIM test: A simple technique detects numerous anomalies in the reporting of results in psychology. Social Psychological and Personality Science, 8(4), 363-369.

Cummings, Peter. “Arguments for and against standardized mean differences (effect sizes).” Archives of pediatrics & adolescent medicine 165.7 (2011): 592-596

Do, L. G., Spencer, A. J., Sawyer, A., Jones, A., Leary, S., Roberts, R., & Ha, D. H. (2023). Early childhood exposures to fluorides and child behavioral development and executive function: a population-based longitudinal study. Journal of Dental Research, 102(1), 28-36.

Ewing-Cobbs, Linda, Mary R. Prasad, Larry Kramer, Charles S. Cox, James Baumgartner, Stephen Fletcher, Donna Mendez, Marcia Barnes, Xiaoling Zhang, and Paul Swank. “Late intellectual and academic outcomes following traumatic brain injury sustained during early childhood.” Journal of Neurosurgery: Pediatrics 105, no. 4 (2006): 287-296.

Glasziou, P., & Chalmers, I. (2018). Research waste is still a scandal—an essay by Paul Glasziou and Iain Chalmers. Bmj, 363.

Glick, M., & Booth, H. A. (2014). Conspiracy ideation: a public health scourge?. The Journal of the American Dental Association, 145(8), 798-799.

Godlee, F., Smith, J., & Marcovitch, H. (2011). Wakefield’s article linking MMR vaccine and autism was fraudulent. Bmj, 342.

Grimes, D. R. (2022). Balancing benefits and potential risks of vaccination: the precautionary principle and the law of unintended consequences. BMJ Evidence-Based Medicine, 27(6), 319-323.

Grimes, D.R., 2025. Tortured confessions? Potentially erroneous statistical inferences may underpin misleading claims of harms in reanalyses of COVID-19 and HPV vaccines. Vaccine, 46, p.126657.

Grimes, D. R., & Heathers, J. (2021). Association between magnetic field exposure and miscarriage risk is not supported by the data. Scientific Reports, 11(1), 22143.

Grimes, David Robert, and James Heathers. “The new normal? Redaction bias in biomedical science.” Royal Society open science 8, no. 12 (2021): 211308.

Guichon, Juliet R., Colin Cooper, Andrew Rugg-Gunn, and James A. Dickinson. “Flawed MIREC fluoride and intelligence quotient publications: A failed attempt to undermine community water fluoridation.” Community Dentistry and Oral Epidemiology (2024).

Heathers, J. (2025). An Introduction to Forensic Metascience. www.forensicmetascience.com . https://doi.org/10.5281/zenodo.14871843

Ibarluzea, J., Gallastegi, M., Santa-Marina, L., Zabala, A. J., Arranz, E., Molinuevo, A., ... & Lertxundi, A. (2022). Prenatal exposure to fluoride and neuropsychological development in early childhood: 1-to 4 years old children. Environmental research, 207, 112181.

Khan SA, Singh RK, Navit S, Chadha D, Johri N, Navit P, Sharma A, Bahuguna R. 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 Nov;9(11):ZC10-5. doi: 10.7860/JCDR/2015/15518.6726.

Levy, S. M. (2025). Caution Needed in Interpreting the Evidence Base on Fluoride and IQ. JAMA pediatrics.

Mausner, B., & Mausner, J. (1956). Fluoridation: a Study of the Anti-Scientific Attitude. Health Education Journal, 14(3), 123-133.

Mole, B (2025), Controversial fluoride analysis published after years of failed reviews, Ars Technica, Published 6th January 2025. Available online: https://arstechnica.com/health/2025/01/controver fluoride-analysis-published-after-years-of-failed-reviews/

Motta, M., & Stecula, D. (2021). Quantifying the effect of Wakefield et al.(1998) on skepticism about MMR vaccine safety in the US. PloS one, 16(8), e0256395.

National Academies of Sciences, Division on Earth, Life Studies, Board on Environmental Studies, Committee to Review the NTP Monograph on the Systematic Review of Fluoride Exposure, Neurodevelopmental, & Cognitive Health Effects. (2020). Review of the Draft NTP Monograph: Systematic Review of Fluoride Exposure and Neurodevelopmental and Cognitive Health Effects.Available online: https://nap.nationalacademies.org/catalog/25715/review-of-the-draft-ntp-monograph-systematic-review-of-fluoride

National Academies of Sciences, Division on Earth, Life Studies, Board on Environmental Studies, Committee to Review the NTP Monograph on the Systematic Review of Fluoride Exposure, Neurodevelopmental, & Cognitive Health Effects. (2021). Review of the Revised NTP Monograph on the Systematic Review of Fluoride Exposure and Neurodevelopmental and Cognitive Health Effects: A Letter Report. Available online: https://nap.nationalacademies.org/catalog/26030/review-of-the-revised-ntp-monograph-on-the-systematic-review-of-fluoride-exposure-and-neurodevelopmental-and-cognitive-health-effects

Oza A, What to know about a controversial new study on fluoride and IQ, Stat News, Originally published January 6th 2025, Available Online: https://www.statnews.com/2025/01/06/fluoride-iq-jama-pediatrics-critiques-meta-analysis/

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 Prevent Communit Dent 2017;7:252-8.

Riddell, Julia K., Ashley J. Malin, Hugh McCague, David B. Flora, and Christine Till. “Urinary fluoride levels among Canadians with and without community water fluoridation.” International Journal of Environmental Research and Public Health 18, no. 12 (2021): 6203.

Taylor, K. W., Eftim, S. E., Sibrizzi, C. A., Blain, R. B., Magnuson, K., Hartman, P. A., ... & Bucher, J. R. (2025). Fluoride exposure and children’s IQ scores: a systematic review and meta-analysis. JAMA pediatrics.

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.

Zhang, P. H., and L. Cheng. “Effect of coal-burning endemic fluorosis on children’s physical development and intellectual level.” Chin J Control Endemic Dis 30.6 (2015): 458-60.