October 23, 2019

Rick Woychik, PhD
Acting Director, National Institute of Environmental Health Sciences
P.O. Box 12233
Mail Drop B2-01
Durham, NC 27709

Gwen W. Collman, Ph.D.
Acting Deputy Director, National Institute of Environmental Health Sciences
P.O. Box 12233
Mail Drop B2-01
Durham, NC 27709

Dear Drs. Woychik and Collman,

Research supported by the National Institute of Environmental Health Sciences (NIEHS) has tremendous potential to expand scientific knowledge about human health and positively inform health and environmental policies. We write this letter regarding the NIEHS-funded Green et al study about maternal fluoride exposure; the article about this study was recently published by *JAMA Pediatrics*.[1] We have a variety of concerns about the Green article related to methodological transparency and analytical rigor. We wish to share these concerns with you and request that NIEHS formally ask the Green authors to release the HIPAA-compliant, Research Identifiable File (RIF) data sets from their study, as well as a complete explanation of their methods and the computer program/codes used in their data management and analysis.

In recent weeks, at least two of the Green authors have declined to respond affirmatively to requests from other researchers for access to the data and analytical methods they used. Growing concerns about the replicability of scientific research makes transparency more critical than ever. Recently, the editor-in-chief of the *Journal of Neuroscience Research* and nine other research experts wrote an article whose abstract opened with this observation: "Progress in basic and clinical research is slowed when researchers fail to provide a complete and accurate report of how a study was designed, executed, and the results analyzed."[2]

By releasing their raw data and a detailed explanation of their statistical methods, the Green authors could satisfy incongruities and ensure the scientific record is accurate. Given the potential policy implications of the Green article, we believe the authors should be more transparent, as this could provide clarity amid the concerns their article has raised. In recent weeks, a number of experts in epidemiology, psychology, statistical methodology and other fields have raised numerous concerns about the Green article, including the following:

1. **Focusing on a subgroup analysis amid "noisy data":** The Green authors focused a significant portion of their narrative on the one subgroup (boys) where a lower IQ association was observed, but only in the performance part of the IQ test. Thom Baguley, professor of experimental psychology at Nottingham Trent University, is one of several experts who have raised concern about this aspect of the article. "This is an example of subgroup analysis — which is frowned upon in these kinds of studies because it is nearly always possible to identify some subgroup which shows an effect if the data are noisy," he wrote. "Here the data are very noisy."[3] Part of the reason why the data are noisy is that the
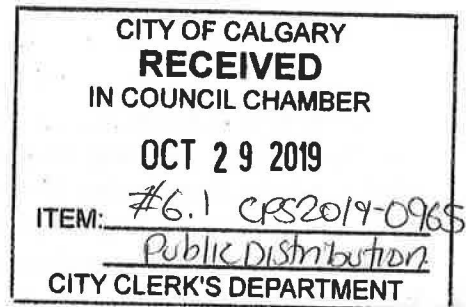
Green authors included in their analysis IQ scores that fell within 2.5 standard deviations from the mean. In other words, only a very small number of scores could have been excluded, which explains why the study includes scores in the 50s, indicating profound delays, as well as scores in the 130s, which indicate giftedness. Not excluding extreme values or outliers may have skewed the regression analysis, the impact of which could be better understood with review of the raw data used by the Green authors.

Additionally, the Green article's focus on the subgroup analysis ignores the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) recommendations, which were issued more than a decade ago. The STROBE committee cited several reasons for these recommendations, including its view that "there is too great a tendency to look for evidence of subgroup-specific associations, or effect measure modification, when overall results appear to suggest little or no effect."[4]

2. **Modeling and variable anomalies:** As you know, a p-value indicates the probability that the results observed are by chance. In its Statistical Analyses section, the Green authors reported that they "retained a covariate in the model if its p-value was less than .20 or its inclusion changed the regression coefficient of the variable associated factor by more than 10% in any of the models." This could potentially be an example of p-hacking, meaning that variables are left in the model (or excluded) to achieve some sort of statistical significance in the final p-values.[5] In this regard, the authors did not explain why they chose .20 instead of .10 or .15 so readers of the article have no basis for deciding whether this decision was warranted or not.

Typically, when a researcher is testing associations between potential confounders and an outcome of interest, they set the significance level at $p<0.20$; however, for retaining them in the final model, they typically select only those with a significance of $p<0.05$. It is unclear if the Green researchers carried out this step. The more covariates that are included in the final model, the more likely it is that researchers will find a significant association.

One approach to account for the effects of including variables in the model as the Green authors did is to use Bonferroni's correction, something they do not appear to have done. Bonferroni's correction allows for the p-values of multiple comparisons to be performed at the same time in one model. Usually, this correction decreases the acceptable p-value at which a variable is considered to be statistically significant in its contribution to the observed statistic.

Further, the Green analysis failed to account for the influence of multi-level effects on the association between the independent and dependent variables or the multi-level effects such variables would have on the outcome of interest. Moreover, selecting the covariates in a stepwise manner has the potential to bias regression coefficients. Biostatisticians have discouraged the use of stepwise selection for this reason.[6] The American Statistical Association has established six principles on the use and analysis of p-values, one of which states: "Proper inference requires full reporting and transparency."[7] By releasing the data and a detailed explanation of their analytical methods, the Green authors would enable the scientific community to better assess whether their choice of p-value was appropriate.

3. **Lacking data on relevant factors:** In recent decades, studies have revealed numerous factors that can impact children's intelligence and cognitive ability. Parental behaviors and traits are among these factors; for example, research shows children's intelligence is linked to fathers' social class[8] and mothers' IQ scores.[9] Indeed, a 2013 study supported by the Centers for Disease Control and Prevention noted that these factors are "major predictors of IQ and should be included routinely in studies of cognitive development."[9] Unfortunately, the cohort on which the Green authors relied lacks data on maternal IQs and paternal social class. Although the Green authors had access to data on maternal lead exposure, the cohort provided no data on lead exposure for children during the 3-4 years between birth and their IQ

tests. This could be another confounder. The authors themselves acknowledge that "this observational study design could not address the possibility of other unmeasured residual confounding."[1] This is more than a minor concern, and it demonstrates why the Green findings should be interpreted with significant caution.

4. **Omitting crucial findings:** In the Key Points section of their article, the Green authors summarize the question their study sought to answer: "Is maternal fluoride exposure during pregnancy associated with childhood IQ in a Canadian cohort receiving optimally fluoridated water?" Nonetheless, the authors downplayed the lack of a statistically significant difference between children's mean IQ scores in fluoridated and non-fluoridated communities of Canada. Nowhere within the narrative of their article do they share the two mean IQ scores for the Full Scale IQ (FSIQ)—108.21 among children in fluoridated communities and 108.07 among children in non-fluoridated areas. These nearly identical scores could easily be missed unless readers carefully scan the 29 rows of data within Table 1. Instead, the Green article focused on the subgroup analysis where an association was observed. Contrast how the Green article reported its findings with how the authors of a 2018 article on lead exposures[10] presented their findings. Data for both articles was sourced from the same Canadian maternal-child cohort.[11] The authors of the 2018 article included three of the Green authors, and they summarized the results in their abstract in *Environment International* (EI) by leading with their primary objective:

> "*Median blood lead concentrations for the mother at 1st trimester and 3rd trimester of pregnancy, and for cord and child blood were 0.60 μg/dL, 0.58 μg/dL, 0.79 μg/dL and 0.67 μg/dL, respectively. We found no association between cord blood lead concentrations and WPPSI-III scores in multivariable analyses. However, cord blood lead concentrations showed a negative association with Performance IQ in boys but not in girls (B = 3.44; SE = 1.62; 95% CI: 0.82, 5.98). No associations were found between WPPSI-III scores and prenatal maternal blood or concurrent child blood lead concentrations.*"

By stark contrast, the Green authors proceeded to their subgroup analyses without mentioning the full-scale IQ scores for fluoridated and non-fluoridated areas. Additionally, they did not report the main effect result for maternal urinary fluoride (MUF) and IQ.[12] Had the Green authors reported the main effect result, it would have shown the association between MUF and IQ was non-significant—both with and without covariates. STROBE guidelines indicate that all main effect results should be reported (Guideline #16), in addition to any interactions and sensitivity analyses (Guideline #17).[4]

5. **Using invalid measures to determine individual exposures:** According to an article coauthored by the director of Columbia University's Applied Statistics Center, *the* most important assumption in linear regression is that the independent and dependent variables map to one's research question and are valid.[13] In this case, relying on MUF and a twice-administered beverage recall question estimating types of fluid consumption are not reliable ways to determine fetal fluoride exposure. Moreover, relying on these measures could threaten construct validity, a limitation which is not adequately discussed by the authors. This issue is compounded by the fact that MUF was gathered by spot urine samples rather than 24-hour samples. Alastair Hay, professor emeritus of environmental toxicology at the University of Leeds, reviewed the Green article and raised this concern: "For a substance with a short half-life, such as fluoride, urine concentrations vary hugely and are really only representative of the last drink. Validation of intake is something you must do before looking at associations."[3]

Researchers have noted the limitations of extrapolation from urine samples, observing that "daily urinary fluoride excretion is suitable for predicting fluoride intake for groups of people, but not for individuals."

In their assessment: "Thus, it can be concluded that, at this time, urinary fluoride excretion has a very limited value as a biomarker of individual fluoride exposure."[14] Dr. F. Perry Wilson, a nephrologist at the Yale School of Medicine, criticized the Green article for relying on urine samples "because urine fluoride is not a perfect proxy for blood fluoride." Moreover, Dr. Wilson identified a new variable that the authors introduced:

> "... more dilute urine will have a lower fluoride concentration, and they 'correct' this problem by dividing urinary fluoride by urine specific gravity. But this introduces a new variable. Assuming that fluoride has no effect on a child's IQ, you could get results that look like this if moms with more dilute urine tend to have kids with lower IQs."

Dr. Wilson also noted that the article's results could have been skewed because women with a higher urinary pH (due to diet or other factors) tend to have higher urinary fluoride levels.[15]

6. **Defining the final study group:** The Green study is not representative of all mothers and their children in Canada, and, therefore, not generalizable. This is clear from Figure 1 in the Green article, indicating that a significant number of mother-child pairs were excluded from the final study group.[1] The cohort was drawn from six cities, and the authors did not account for clusters of sampled pregnant women or consider multilevel models. It is not entirely clear from the Figure 1 schematic or the article's narrative why all of these pairs were excluded, and it is important to learn the authors' reasons for exclusion to understand fully the actual fluoride exposure and its effects on IQ scores. What we do know is that some of the pairs were excluded because they did not drink tap water or lived outside a water treatment zone.

This kind of exclusion would make sense only if the authors were specifically seeking to compare IQ scores based on water fluoridation status, but they did not present their results in this way. The authors should have been more explicit as to the reasons for excluding the mother-child pairs that they removed from the final study group. The fact that the association was observed only in boys could be an artifact of who was left out of the study and how the Green authors modeled sex in the regression.

7. **Assessing the impact of fluoride exposure:** Several questions arise because the details of the regression model used by the Green authors are not provided. Further, the authors assessed daily fluoride intake in mothers using a non-validated questionnaire, and their estimates of fluoride intake based on water and tea consumption appear to be crude. The narrow focus on tea-drinkers could have biased the results by overlooking other sources of fluoride intake. In this regard, the EI article (cited previously) on prenatal lead concentrations and IQs in this same cohort are instructive. The EI article demonstrated that cord blood lead level was associated with IQ in boys in this cohort (but not girls). While the supplemental table eTable 2 of *JAMA Pediatrics* shows that controlling for lead does not alter the predictor or its standard error, it does raise questions about the role of environmental lead in this cohort. Additional information concerning the measure of lead exposure in the Green study's sensitivity analysis is needed, especially given research showing that blood lead levels are higher in those who drink tap water.[16] Release of the RIF data and a detailed explanation of the modeling used by the Green authors would be valuable for clarifying the relationship among these variables.

8. **Reporting anomalies:** The authors reported that a 1 mg/L increase in the adjusted MUF concentration was associated with a 4.49-point lower IQ score in boys, but there was no statistically significant association with IQ scores in girls ($B = 2.40$; 95% CI, -2.53 to 7.33). And a 1 mg/L higher daily intake of maternal fluoride was associated with a 3.66 lower IQ score in boys and girls. The Green authors did not discuss the magnitude of change in the sex differences for IQ observed in the MUF-adjusted regression.

This difference includes an actual sign change between boys and girls (from - to +) that poses a significant threat to the validity of their results. Instead, the authors stressed the results for boys while ignoring the disconnect between boys' performance IQ and verbal IQ scores. Also unmentioned in the Green article is the overlapping confidence intervals throughout for boys and girls (see Table 2), which means there is a greater than 5% probability that the IQ measurements for boys and girls are actually *not* different from each other.

In addition, while several covariates have been found in the past to be significant determinants of IQ score and are not included in the Green study, the reported 95% confidence interval (CI) in boys (−8.38 to −0.60) is too wide to be statistically acceptable. The ordinary multivariable statistical methodology (e.g. regression analysis adjustment) used in this study, which is indeed widely known and used in epidemiology, focuses on the association with the outcome.[17] However, this kind of analysis does not adequately address complicated problems, where measured and unmeasured confounding is involved.[18,19] Alternatively, novel methodological approaches (e.g. propensity scores and Inverse Probability Weighting) are being used in the medical, epidemiological and biostatistical research to infer causal effects with less potential bias and to provide narrow and more precise CIs.[20,21,22] The 95% CI in this study for boys' IQs reveals that this score can have a value of .6, which is almost 15 times smaller than the higher value (8.38) estimated in this interval. The study's omission of significant predictors/confounders (see #3) for IQ scores resulted in this wide interval, limiting the validity of the results.

9. **Internal inconsistency of outcomes:** The Green authors reported the overall effect result for fluoride intake (FI) in this way: "A 1-mg higher daily intake of fluoride among pregnant women was associated with a 3.66 lower IQ score (95%CI, −7.16 to −0.15; P = .04) in boys and girls."[1] An attentive reader would recall that the association found in the MUF regression was not all children; rather, boys showed an associated decline on one part of the IQ test, but girls did not. (In fact, girls had an observed increase in IQ.) Here, the authors attempted to demonstrate internal consistency of analysis outcomes for both FI and MUF. In summary, they observed the FI intake effect when they combined boys and girls (overall effect), but they found the effect of MUF only in boys. Additionally, Green did not find an effect with FI in her thesis when she included the same covariates.[23] Had the Green authors reported this lack of association in the *JAMA Pediatrics* article, it would have been strong evidence of internal inconsistency.

Upon closer comparison, the Green thesis states they excluded city as a covariate to achieve statistical significance for the FI analysis. As the Green thesis explained (p. 34): "Holding all covariates constant, FI significantly predicted [Full-Scale IQ] scores **without city** in the model (B = -4.03, 95% CI: -6.82 to -1.25, p = .005*) (Figure 6). **With city** in the model, FI just missed significance (B = -3.82, 95% CI: -7.65 to 0.02, p = .05). In both models, there were no significant interactions between FI and any of the covariates" (emphasis added).[23] By contrast, in the *JAMA Pediatrics* article, Table 2 reported that FI was adjusted "for city, HOME score, maternal education, race/ethnicity, child sex, *and* prenatal secondhand smoke exposure."[1] However, these two model outputs, 3.82 in the thesis versus 3.66 in the article, are different. Allowing the raw data to be reviewed and analyzed might reveal exactly how the researchers managed to find a model with significant effect with city for inclusion in the article. Accordingly, Dr. Stuart Ritchie of the Institute of Psychiatry, Psychology and Neuroscience at King's College London, reinforced this point in his recent comments about the second analysis, which was FI:

> *"For the second analysis, where there's an overall effect, the p-value is .04—that is, it's JUST below the standard threshold used for declaring something to be significant (0.05). Given that they ran lots of other hypothesis tests in the paper, and didn't correct*

for how many times they did so, I wouldn't have much confidence in this finding being robust or replicable" (capitalization in original).[3]

It could be that the authors were able to achieve a significant effect in the FI analysis with city in the FI model because they included secondhand smoke as another covariate. This possibility presents a significant problem because the number of smokers in the Green sample of 400 is only 11. In the EI article that studied the same cohort, the authors specifically stated that they excluded the secondhand smoke exposure variable from the analysis because of the lack of variance.[10] It is possible that without the secondhand smoke variable in the FI model, the two main effect models MUF-IQ and FI-IQ would show no association. This undermines the main discussion point in the Green article that there is a "converging" of the two analyses. In fact, the two models might be remarkably similar in showing no effect of fluoride on IQ. To clarify these discrepancies, the scientific community needs to have access to the Green data sets so they can be reanalyzed using a multi-level model—or, at the least, a principled accounting for the design effect arising from cluster sampling, and adjusting the p value for multiple hypothesis testing. Such an independent analysis could help us determine whether fluoride exposure at common levels has an effect on IQ in this cohort.

10. **Overlooking research that conflicts with the authors' conclusions:** Typically, when researchers identify their study's limitations in the Discussion section, they acknowledge other research that reached different conclusions and perhaps consider possible explanations for these differences. Yet the Green authors do not acknowledge or cite several studies about fluoride and cognitive development, including one that (unlike their study) tested IQs multiple times over a 30-year period.[24,25] Although most of these studies did not focus on maternal exposures, one such study, co-led by researchers at the National Toxicology Program (NTP), examined animal exposures to fluoride during the gestational period and observed no exposure-related differences in learning skills or memory.[26] Although the NTP-led study was cited in the Green article, the authors did not mention the NTP study's conclusions. Of related concern, the Green article's citations are limited in scope and include three articles from Fluoride, a publication that has been described as not applying a high degree of rigor when publishing studies.[27] Altogether, these concerns suggest the Green authors may have conducted a selective literature search that could reflect a predetermined conclusion.

Summary and Requests

We believe the Green authors should have taken additional steps to address or at least fully acknowledge potential confounders. Moreover, they should have presented their findings in a more transparent, qualified way that reflects STROBE guidelines. Given that the NIEHS funding award was an R21 exploratory grant, the authors should have exercised more caution in the interpretation of the results.

The publication of this article in a mainstream, peer-reviewed medical journal has generated a tremendous amount of media coverage. If the Green authors had merely called for more research, our focus would be directed toward ensuring that future research in this area is more methodologically robust and reflective of STROBE guidelines. However, the article's release was followed by statements to major media by the journal's editor[28] and at least one author that are creating confusion, shaping individual behaviors and influencing public policy. For instance, the corresponding author of the Green article told *Time* magazine that instructing pregnant women to reduce their fluoride intake is "a no-brainer."[29] The fallout from the Green article is currently most visible in the Canadian city of Calgary, where the article is being cited as a reason not to resume water fluoridation after eight years of cessation and significant increases in tooth decay.[30]

The aim of science is to gain a better understanding of our natural world and to build a shared knowledge base for the benefit of all. Every scientist is interested in the truth. If fluoride at common levels of maternal exposure does lead to lower IQ scores, we would certainly want to know. This is why transparency related to the Green article is crucial. Given the concerns outlined herein:

- We urge NIEHS to ask the Green authors to release their RIF data set and provide a thorough explanation of their analytical methods. Doing so could enable an independent review that would bring clarity and ensure the scientific record is accurate.

- Should the Green researchers not voluntarily release their data, please advise us on what the process would be to have the data set released so an independent analysis of the Green data can be conducted.

Without greater transparency of its data and analytical methods, the Green article could generate unjustified fear that undermines evidence-based clinical and public health practices. So much is at stake. Hundreds of millions of people around the globe—from Brazil to Australia—live in homes that receive fluoridated drinking water. Hundreds of millions of people use toothpaste or other products with fluoride. Many millions of children receive topical fluoride treatments in clinical or other settings. Tooth decay remains one of the most common chronic diseases for children and teens, and fluoride is a crucial weapon against this disease. Decay prevalence could increase if a journal article unnecessarily frightens people to avoid water, toothpaste or other products fortified with fluoride.

Please let us know if you have any questions about our request or the issues raised in this letter. Please consider Dr. Scott Tomar (stomar@dental.ufl.edu) as the individual to whom you can direct your response. We greatly appreciate your time and consideration.

Sincerely,

Raman Bedi, BDS, MSc, DDS, DSc, FDSRCS
Chair, World Federation of Public Health
Associations, Oral Health Group
Chair, Global Child Dental Fund
Former Chief Dental Officer for England

Nigel Carter, OBE, BDS, LDS(RCS)
CEO, Oral Health Foundation,
Chair, Platform for Better Oral Health in Europe
Chair of the Royal Society for Public Health

Timothy Caulfield, BSc, LLB
Canada Research Chair in Health Law
& Policy Professor
Faculty of Law & School of Public Health
University of Alberta

Denice Curtis, DDS, MPH, DHS
Assistant Professor
Master of Public Health Program
Usha Kundu, MD College of Health
University of West Florida

Gail Douglas, BMSc, BDS, PhD, MPH, FDS
RCS(Ed), FDS(DPH), RCS
President of the British Association for the
Study of Community Dentistry
University of Leeds, U.K.

Kenneth A. Eaton, PhD, MSc, FFPH, BDS
Visiting Professor
University College London, U.K.

Michael Foley, BDSc, MPH, MEpi
Director, Research and Advocacy
Metro North Oral Health Services,
Queensland Health, Australia

Ruby L. Fried, PhD
Assistant Professor
Department of Population Health Studies
University of Alaska, Anchorage
Institute for Circumpolar Health Studies

John Furness, MBBS, DCH, FRCPCH
Consultant Paediatrician
Darlington, England (U.K.)

Mairead Harding, BDS, MFGDP (UK),
MDPH, PhD, FDS
Senior Lecturer
Deputy Director, Oral Health Services
Research Centre
Cork University Dental School and Hospital
Ireland

Catherine Hayes, DMD, SM, DMSc
Clinical Professor of Health Policy & Health
Services Research
Boston University, Henry M. Goldman
School of Dental Medicine

Elizabeth A. Hodges, PhD, MPH
Associate Professor of Public Health,
College of Health
University of Alaska, Anchorage

Alice M. Horowitz, PhD, MA, RDH
Research Associate Professor, Behavioral
and Community Health
Univ. of Maryland, School of Public Health

Jonathan E. Howard, MD
Assistant Professor, Department of Neurology
Assistant Professor, Department of Psychiatry
NYU School of Medicine

Dushanka V. Kleinman, DDS, MScD
Associate Dean for Research
Univ. of Maryland, School of Public Health

Michael A. Lennon, OBE, MDS, DPD,
FDSRCS(Ed)
Professor Emeritus in Dental Public Health
University of Sheffield (U.K.)

Lauren Lessard, PhD, MPH
Assistant Professor of Health Science
University of Alaska, Anchorage
Institute for Circumpolar Health Studies

Vasileios Margaritis, PhD, MSc, DDS
Senior Lecturer/Core Faculty
College of Health Sciences
Walden University, College of Health Sciences

Jennifer Meyer, PhD, MPH, CPH, RN
Assistant Professor of Health Sciences
University of Alaska, Anchorage

A. John Morris, DDS, MCDH, BDS
Senior Lecturer in Dental Public Health and
Deputy Senior Tutor
School of Dentistry, The University of
Birmingham, United Kingdom

Mark E. Moss, DDS, PhD
Associate Professor
ECU School of Dental Medicine
East Carolina University

Joe Mullen, BDS, BA, BSc, MA, MDPH
Former Chairman,
New and Emerging Issues Committee
Expert Body on Fluorides & Health, Ireland

René F. Najera, MPH, DrPH
Associate, Department of Epidemiology
Bloomberg School of Public Health
Johns Hopkins University

H. Grant Ritchey Jr., DDS, FAGD
Practicing dentist in Kansas, contributor to
Science-Based Medicine blog

R. Gary Rozier, DDS, MPH
Emeritus Professor
Gillings School of Global Public Health
University of North Carolina

Amy Seery, MD, FAAP
Pediatric Section Chair,
Ascension Via Christi-Wichita Hospitals
Assistant Professor,
University of Kansas School of Medicine

Scott Tomar, DMD, MPH, DrPH
Professor & Director of Institutional Analysis
and Evaluation
University of Florida College of Dentistry

Georgios Tsakos, PhD, FFPH
Professor of Dental Public Health
Department of Epidemiology and Public Health
University College London (UCL)

Paula Vassallo, BChD, MSc, DDPH, RCS,
MBA, FFPH
Director, Health Promotion and Disease
Prevention
Consultant, Dental Public Health
President, European Association of Dental
Public Health

John J. Warren, DDS, MS
Professor, Department of Preventive &
Community Dentistry
University of Iowa College of Dentistry

CC: Patrick Breysse, PhD
      Director, National Center for Environmental Health (CDC)

      Karen Hacker
      Director, National Center for Chronic Disease Prevention and Health Promotion (CDC)

      Martha J. Somerman, DDS, PhD
      Director, National Institute of Dental and Craniofacial Research

References

[1] Green R, Lanphear B, Hornung R, et al. (e-published 19 Aug 2019) Association between Maternal Fluoride Exposure during Pregnancy and IQ Scores in Offspring in Canada. JAMA Pediatr. https://doi:10.1001/jamapediatrics.2019.1729.

[2] Prager EM, Chambers KE, Plotkin JL, et al. (2019) Improving transparency and scientific rigor in academic publishing. Brain Behav. 9(1): e01141. https://doi.org/10.1002/brb3.1141.

[3] Science Media Centre. 19 Aug 2019. Expert reaction to study looking at maternal exposure to fluoride and IQ in children. Accessed in Oct 2019 at https://www.sciencemediacentre.org/expert-reaction-to-study-looking-at-maternal-exposure-to-fluoride-and-iq-in-children/.

[4] Vandenbroucke JP, von Elm E, Altman DG, et al. (2007) Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): Explanation and Elaboration. PLoS Med 4(10): e297. https://doi.org/10.1371/journal.pmed.0040297.

[5] Head ML, Holman L, Lanfear R, et al. (2015) The Extent and Consequences of P-Hacking in Science. PLoS Biol 13(3): e1002106. https://doi.org/10.1371/journal.pbio.1002106.

[6] Steyerberg E.W., Eijkemans M.J.C., Habbema J.D.F. (1999) Stepwise selection in small data sets: A simulation study of bias in logistic regression analysis. J Clin Epidemiol 52(10): 935-942. https://doi.org/10.1016/S0895-4356(99)00103-1.

[7] Wasserstein RL, Lazar NA. (2016) The ASA Statement on P-Values: Context, Process, and Purpose, The Amer Statistician, 70(2): 129-133. https://doi.org/10.1080/00031305.2016.1154108.

[8] Lawlor DA, Batty GD, Morton SMB, et al. (2005) Early life predictors of childhood intelligence: evidence from the Aberdeen children of the 1950s study. J Epidemiol Community Health, 59(8): 656–663. http://dx.doi.org/10.1136/jech.2004.030205.

[9] Eriksen H-LF, Kesmodel US, Underbjerg M, et al. (2013) Predictors of Intelligence at the Age of 5: Family, Pregnancy and Birth Characteristics, Postnatal Influences, and Postnatal Growth. PLoS ONE, 8(11): e79200. https://doi.org/10.1371/journal.pone.0079200.

[10] Desrochers-Couturea M, Oulhoteb Y, Arbuckle TE, et al. (2018) Prenatal, concurrent, and sex-specific associations between blood lead concentrations and IQ in preschool Canadian children. Environ Int. Dec 2018; 121(Pt 2): 1235-1242. https://doi.org/10.1016/j.envint.2018.10.043.

[11] Government of Canada. (15 Apr 2019) Maternal-Infant Research on Environmental Chemicals (MIREC) Study. Accessed in Oct 2019 at https://www.canada.ca/en/health-canada/services/environmental-workplace-health/environmental-contaminants/human-biomonitoring-environmental-chemicals/maternal-infant-research-environmental-chemicals-mirec-study.html.

[12] For the sake of simplicity, our letter uses MUF rather than MUFsg. The latter term was used in the Green et al study to indicate that maternal urinary fluoride was adjusted for specific gravity and averaged across three trimesters.

[13] Gelman A, Hill J. (2007) Data analysis using regression and multilevel/hierarchical models. Cambridge, UK: Cambridge University Press.

[14] Rugg- Gunn AJ, Villa AE, Buzalaf MRA. (2011) Contemporary Biological Markers of Exposure to Fluoride. Buzalaf MAR (ed): Fluoride and the Oral Environment. Monogr Oral Sci. 22: 37–51. https://doi.org/10.1159/000325137.

[15] Wilson FP. (21 Aug 2019) Is Fluoridated Water Affecting Our Kids' Intelligence? Medscape: Impact Factor. Accessed in Sep 2019 at https://www.medscape.com/viewarticle/916971.

[16] Sanders AE, Slade GD. (2018) Blood Lead Levels and Dental Caries in U.S. Children Who Do Not Drink Tap Water. Am J Prev Med. 54(2): 157-163. https://doi.org/10.1016/j.amepre.2017.09.004

[17] Margaritis V, Mamai-Homata E, Koletsi-Kounari H. (2011) Novel methods of balancing covariates for the assessment of dental erosion: A contribution to validation of a synthetic scoring system for erosive wear. J Dent. 39(5), 361-367. https://doi.org/10.1016/j.jdent.2011.02.007.

[18] Stampf S, Graf E, Schmoor C, Schumacher M. (2010) Estimators and confidence intervals for the marginal odds ratio using logistic regression and propensity score stratification. Stat Med. 29(7-8):760-9. https://doi.org/10.1002/sim.3811.

[19] Kurth T, Walker AM, Glynn RJ, et al. (2006) Results of multivariable logistic regression, propensity matching, propensity adjustment, and propensity-based weighting under conditions of nonuniform effect. Am J Epidemiol. 163(3): 262–70. https://doi.org/10.1093/aje/kwj047.

[20] Rosenbaum PR, Rubin DB. (1983) The central role of the propensity score in observational studies for causal effects. Biometrika. 70(1): 41–55. https://doi.org/10.2307/2335942.

[21] Robins JM. (1998) Marginal structural models. 1997 Proceedings of the section on Bayesian statistical science. Alexandria, VA: American Statistical Association. 1–10. Accessed in Oct 2019 at https://cdn1.sph.harvard.edu/wp-content/uploads/sites/343/2013/03/msm-web.pdf.

[22] Robins JM, Hernan MA, Brumback B. (2000) Marginal structural models and causal inference in epidemiology. Epidemiology. 11(5):550–60. Accessed in Oct 2019 at https://www.ncbi.nlm.nih.gov/pubmed/10955408.

[23] Green R. (July 2018) Prenatal Fluoride Exposure and Neurodevelopmental Outcomes in a National Birth Cohort: A Thesis Submitted to the Faculty of Graduate Studies in Partial Fulfillment of the Requirements for the Degree of Master of Arts. York Univ.: Toronto, Canada. Accessed in Oct 2019 at https://yorkspace.library.yorku.ca/xmlui/bitstream/handle/10315/35502/Green_Rivka_R_2018_Masters.pdf?sequence=2&isAllowed=y.

[24] Barberio AM, Quiñonez 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(3):e229-e239. https://doi.org/10.17269/cjph.108.5951.

[25] Broadbent JM, Thomson, WM, Ramrakha S, et al. (2015) Community Water Fluoridation and Intelligence: Prospective Study in New Zealand. Am J Public Health. 105(1):72-76. https://doi.org/10.2105/AJPH.2013.301857.

[26] McPherson CA, Zhang G, Gilliam R, et al. (2018) An Evaluation of Neurotoxicity Following Fluoride Exposure from Gestational Through Adult Ages in Long-Evans Hooded Rats. Neurotox Res. 34(4):781–798. https://doi.org/10.1007/s12640-018-9870-x.

[27] Pew Center on the States. (Apr 2012) "Not all studies carry equal weight." Philadelphia, PA: Pew Charitable Trusts. Accessed in Oct 2019 at https://tapintohealthyteeth.org/wp-content/uploads/2015/07/Not-All-Studies-Carry-Equal-Weight.pdf.

[28] Guarino B. (19 Aug 2019) Study raises questions about fluoride and children's IQ. Washington Post. Accessed in Oct 2019 at https://www.washingtonpost.com/science/2019/08/19/study-raises-questions-about-fluoride-childrens-iq/.

[29] Ducharme J. (21 Aug 2019) Is Fluoride in Drinking Water Safe? A New Study Reignites a Long-Standing Debate. Time. Accessed in Oct 2019 at https://time.com/5656476/is-fluoride-in-water-safe/.

[30] McLaren L, Patterson S, Thawer S, et al. (2016) Measuring the short-term impact of fluoridation cessation on dental caries in Grade 2 children using tooth surface indices. Community Dent Oral Epidemiol. 44(3):274-282. https://doi.org/10.1111/cdoe.12215.