Reviewer's report

Title: A Cross-Sectional Study of Well Water Arsenic and Child IQ in Maine Schoolchildren

Version: 2 Date: 2 January 2014

Reviewer: David Bellinger

Reviewer's report:

Major compulsory revisions

1. The fact that the children were recruited from 3 school districts that differed in important respects, including water arsenic levels, as well as the covariates maternal IQ and HOME score, requires more consideration. It is likely that school districts also differed by other, unmeasured variables that might function as confounders. One strategy the authors might consider would be to conduct a set of analyses stratified by school district to assess whether the exposure-outcome associations are similar. Power would be reduced, but a comparison of the district-specific effect estimates might be informative. If they are comparable across districts, this would, in my mind, reduce concern about residual confounding.

2. In my opinion, a correlation of 0.53 (as between maternal education and maternal IQ) does not reflect a sufficient amount of covariance to conclude that only one of the variables needs to be included in a model. I think that model 3 should include maternal education as well as IQ.

3. In qualitative terms, the results of this study do confirm those of the other studies conducted by this group in Bangladesh in showing an inverse association between water As and neurodevelopment. To the extent that the results can be compared quantitatively, the message does not that consistent, in my reading at least. In this study, no significant differences were observed among the 3 exposure groups with water As levels above 5. Moreover, compared to the referent group, the reductions in scores, for most endpoints, were greater among children with levels between 5 and 10 than they were for children with levels of 10 to 20 or 20 and greater. So there isn’t a coherent dose-effect relationship here. In the Bangladesh studies, children’s neurodevelopmental scores continued to decline at levels well above 5. The Discussion should include something about this apparent inconsistency in the Maine and Bangladesh studies (or else make the case for why the findings are not inconsistent). The general statement in the Discussion that 5 “may represent an important threshold” seems to be based solely on the Maine findings, as the Bangladeshi data do not appear to support this conclusion.

Minor essential revisions

1. The justification for excluding children with “neurodevelopmental disorders” or “those receiving special education services” needs to be explained. One could
make the argument that this would introduce a bias toward the null. If, indeed, higher water As level is associated with children’s cognitive development, their exclusion would have introduced a selection bias. What neurodevelopmental disorders were considered? It appears that this information was obtained by parent report? Were parents asked whether the diagnosis had been confirmed by a health professional?

2. It is stated that the cut-off of 3 years of residence was selected “to maximize sample size.” I understand what is meant, but it isn’t strictly true. Having imposed no length of residence criterion would have maximized sample size.

3. The method by which HOME subscale scores were obtained for children with missing observational items should be described. Were scores extrapolated by plugging in the mean score on scale items for which data were collected?

**Discretionary revisions**

1. In the Abstract, the phrase, “performance-related child intelligence” is not clear. I assume this means scores on subtests that contribute to Performance IQ (or, in more recent studies, to Perceptual Performance or Processing Speed composites), but it should be clarified.

2. In the second paragraph of “Background,” where generalizability of Bangladeshi findings to the US is discussed, reference is made to the fact that the tests used were “modified to increase cultural appropriateness,” but the import of this is not explained. Is the need to make these modifications mentioned because it is considered to limit the generalizability of the findings to a US population, because it complicates the interpretation of children’s scores, or perhaps both?

3. GW is identified as a “psychometrician.” This is entirely up to you, of course, but I thought this term is usually used to refer to a technician who administers tests. GW is certainly more than this!

4. Why were HOME scores categorized at all? If adjusting for potential confounding by this variable was a concern, wouldn’t treating it as a continuous variable have been better (depending of course on the shape of the relationship between HOME score and the endpoints, but this doesn’t seem to have entered into the decision to categorize).

5. I am not clear on why the absence of an association between child IQ and toenail As is viewed as a limitation. It is just an observation. A reduction in power is suggested as one reason (although the sample size for these analyses was 248, compared to 272 for water As, so this does not seem a likely explanation. Maybe toenail As just isn’t a very good biomarker for As burden.

**Level of interest:** An article of outstanding merit and interest in its field

**Quality of written English:** Acceptable

**Statistical review:** Yes, and I have assessed the statistics in my report.
Declaration of competing interests:

No competing interests.