Developmental Fluoride Neurotoxicity: A Systematic Review and Meta-Analysis

Anna L. Choi, Guifan Sun, Ying Zhang, Philippe Grandjean

http://dx.doi.org/10.1289/ehp.1104912

Online 20 July 2012
Developmental Fluoride Neurotoxicity: A Systematic Review and Meta-Analysis

Anna L. Choi¹, Guifan Sun², Ying Zhang³, Philippe Grandjean¹,⁴

¹Department of Environmental Health, Harvard School of Public Health, Boston, MA, USA
²School of Public Health, China Medical University, Shenyang, China
³School of Stomatology, China Medical University, Shenyang, China
⁴Institute of Public Health, University of Southern Denmark, Odense, Denmark

Author information and address for reprints:

Anna L. Choi, Department of Environmental Health, Harvard School of Public Health, Landmark Center 3E, 401 Park Dr., Boston, MA 02215 USA. Phone 617-384-8646; fax 617-384-8994; email achoi@hsph.harvard.edu
Running Title: Fluoride neurotoxicity

Key Words: Fluoride, Intelligence, Neurotoxicity

Acknowledgments:

We thank Dr. Vasanti Malik, Harvard School of Public Health, for the helpful advice on the meta-analysis methods.

This study was supported by internal institutional funds.

The authors declare that they have no competing financial interest.

List of Abbreviations:

CI, confidence interval
CNKI, China National Knowledge Infrastructure
SE, standard error
SMD, Standardized mean difference
TOXNET, Toxicology Data Network
Abstract

Background: Although fluoride may cause neurotoxicity in animal models and acute fluoride poisoning causes neurotoxicity in adults, very little is known of its effects on children’s neurodevelopment.

Objective: We performed a systematic review and meta-analysis of published studies to investigate the effects of increased fluoride exposure and delayed neurobehavioral development.

Methods: We searched the MEDLINE, EMBASE, Water Resources Abstracts, and TOXNET databases through 2011 for eligible studies. We also searched the China National Knowledge Infrastructure (CNKI) database, as many studies on fluoride neurotoxicity have been published in Chinese journals only. In total, we identified 27 eligible epidemiological studies with high and reference exposures, endpoints of IQ scores or related cognitive function measures with means and variances for the two exposure groups. We estimated the standardized mean difference (SMD) between exposed and reference groups across all studies using random effects models. We conducted sensitivity analyses restricted to studies using the same outcome assessment and having drinking water fluoride as the only exposure. Cochran test for heterogeneity between studies, Begg’s funnel plot and Egger test to assess publication bias were performed. Meta-regressions to explore sources of variation in mean differences among the studies were conducted.

Results: The standardized weighted mean difference in IQ score between exposed and reference populations was -0.45 (95% CI -0.56 to -0.35) using a random-effects model. Thus, children in high fluoride areas had significantly lower IQ scores than those who lived in low fluoride areas. Subgroup and sensitivity analyses also indicated inverse associations, although the substantial heterogeneity did not appear to decrease.
**Conclusions:** The results support the possibility of an adverse effect of high fluoride exposure on children’s neurodevelopment. Future research should include detailed individual-level information on prenatal exposure, neurobehavioral performance, and covariates for adjustment.
Introduction

A recent report from the US National Research Council (NRC 2006) concluded that adverse effects of high fluoride concentrations in drinking-water may be of concern and that additional research is warranted. Fluoride may cause neurotoxicity in laboratory animals, including effects on learning and memory (Chioca et al. 2008; Mullenix et al. 1995). A recent experimental study where the rat hippocampal neurons were incubated with various concentrations (20 mg/L, 40 mg/L, and 80 mg/L) of sodium fluoride in vitro showed that fluoride neurotoxicity may target hippocampal neurons (Zhang et al. 2008). Although acute fluoride poisoning may be neurotoxic to adults, most of the epidemiological information available on associations with children’s neurodevelopment is from China, where fluoride generally occurs in drinking water as a natural contaminant, and the concentration depends on local geological conditions. In many rural communities in China, populations with high exposure to fluoride in local drinking water sources may reside in close proximity to populations without high exposure (NRC 2006).

Opportunities for epidemiological studies depend on the existence of comparable population groups exposed to different levels of fluoride from drinking water. Such circumstances are difficult to find in many industrialized countries, as fluoride concentrations in community water are usually no higher than 1 mg/L, even when fluoride is added to water supplies as a public health measure to reduce tooth decay. Multiple epidemiological studies of developmental fluoride neurotoxicity were conducted in China due to the high fluoride concentrations that are substantially above 1 mg/L in well-water in many rural communities, although microbiologically safe water has been accessible to many rural households as a result of the recent five-year plan (2001-2005) by the Chinese government. It is projected that all rural
residents will have access to safe public drinking water by 2020 (World Bank 2006). However, results of the published studies have not been widely disseminated. Four studies published in English (Li et al. 1995; Lu et al. 2000; Xiang et al. 2003; Zhao et al. 1996) were cited in a recent report from the National Research Council (NRC 2006), while the World Health Organization has considered only two (Li et al. 1995; Zhao et al. 1996) in its most recent monograph on fluoride (WHO 2002).

Fluoride readily crosses the placenta (ATSDR 2003). Fluoride exposure to the developing brain, which is much more susceptible to injury caused by toxicants than is the mature brain, may possibly lead to damage of a permanent nature (US EPA 2011). Based on the considerations of health risks, and in response to the recommendation of the National Research Council (NRC 2006), the U.S. Department of Health and Human Services (HHS) and the U.S. Environmental Protection Agency (EPA) recently announced that HHS is proposing to change the recommended level of fluoride in drinking water to 0.7 mg/L from the currently recommended range of 0.7 to 1.2 mg/L, and EPA is reviewing the maximum amount of fluoride allowed in drinking water, which currently is set at 4.0 mg/L (US EPA 2011).

To summarize the available literature, we performed a systematic review and meta-analysis of published studies on increased fluoride exposure in drinking water and neurodevelopmental delays. We specifically targeted studies carried out in rural China that have not been widely disseminated, thus complementing the studies that have been included in previous reviews and risk assessment reports.
Methods

Search Strategy

We searched MEDLINE (National Library of Medicine, Bethesda, MD; http://www.ncbi.nlm.nih.gov/pubmed), EMBASE (Elsevier B.V., Amsterdam, the Netherlands; http://www.embase.com), Water Resources Abstracts (Proquest, Ann Arbor, MI; http://www.csa.com/factsheets/water-resources-set-c.php), and TOXNET (National Library of Medicine, Bethesda, MD; http://toxnet.nlm.nih.gov) databases to identify studies of drinking water fluoride and neurodevelopmental outcomes in children. In addition, we searched the China National Knowledge Infrastructure (CNKI; http://www.cnki.net) database to identify studies published in Chinese journals only. Keywords included combinations of “fluoride” or “drinking water fluoride”, “children”, “neurodevelopment” or “neurologic” or “intelligence” or “IQ”. We also used references cited in articles identified. Records were searched from 1980 to 2011. Our literature search identified 39 studies, among which 36 (95%) were studies with high and reference exposure groups, and 3 (7.7%) studies were based on individual-level measure of exposures. The latter showed dose-related deficits were found but were excluded because our meta-analysis focused on studies with the high and low exposure groups only. In addition, 2 studies were published twice, and the duplicates were excluded.

Inclusion criteria and Data Extraction

The criteria for inclusion of studies included studies with high and reference fluoride exposures, endpoints of IQ scores or other related cognitive function measures, presentation of a mean outcome measure and associated measure of variance [95% confidence intervals (CI) or standard errors (SEs) and numbers of participants]. Interpretations of statistical significance are based on
an alpha level of 0.05. Information included for each study also included the first author, location of the study, year of publication, and numbers of participants in high-fluoride and low-fluoride areas. We noted and recorded the information on age and gender of children, and parental education and income if available.

**Statistical Analysis**

STATA (version 11.0; StataCorp, College Station, TX) and available commands (Stern 2009) were used for the meta-analyses. A standardized weighted mean difference (SMD) was computed using both fixed-effects and random-effects models. The fixed-effects model uses the Mantel-Haenszel method assuming homogeneity among the studies, while the random-effects model uses the DerSimonian and Laird method, incorporating both a within-study and an additive between-studies component of variance when there is between-study heterogeneity (Egger et al. 2001). The estimate of the between-study variation is incorporated into both the standard error of the estimate of the common effect and the weight of individual studies, which was calculated as the inverse sum of the within and between study variance. Heterogeneity among studies was evaluated using the $I^2$ statistic, which represents the percentage of total variation across all studies due to between-study heterogeneity (Higgins and Thompson 2002). The potential for publication bias was evaluated using Begg and Egger tests and visual inspection of a Begg funnel plot (Begg and Mazumdar 1994; Egger et al. 1997). We also conducted independent meta-regressions to estimate the contribution of study characteristics (mean age in years from the age range and year of publication in each study) to heterogeneity among the studies. The scoring standard for the Combined Raven’s Test – The Rural edition in China (CRT-RC) test classifies a score of ≤69 and 70-79 as low and marginal intelligence,
respectively (Wang et al. 1989). We also used the random effects models to estimate risk ratios for the association between fluoride exposure and a low/marginal versus normal Raven’s test score among children in studies that used the Combined Raven’s Test –Rural in China (CRT-RC) test (Wang et al. 1989). Scores indicating low and marginal intelligence ($\leq 69$ and $70-79$, respectively) were combined as a single outcome due to small numbers of children in each outcome subgroup.

**Results**

Six of the 34 studies identified were excluded due to missing information on the number of subjects or the mean and variance of the outcome (see Figure 1 for a study selection flow chart and Supplemental Material, Table S1 for additional information on studies that were excluded from the analysis). Another study (Trivedi et al. 2007) was excluded because SDs reported for the outcome parameter were questionably small (1.13 for high fluoride group, and 1.23 for low fluoride group) and the SMD (-10.8, 95% CI -11.9, -9.6) was more than 10-times lower than the second smallest SMD (-0.95, 95% CI -1.16, -0.75) and 150-times lower than the largest SMD (0.07, 95% CI -0.083, 0.22) reported for the other studies, which had relatively consistent SMD estimates. Inclusion of this study in the meta-analysis resulted with a much smaller pooled random-effects SMD estimate and a much larger $I^2$ (-0.63 (95% CI -0.83, -0.44), $I^2$ 94.1%) compared to the estimates that excluded this study (-0.45, 95% CI -0.56, -0.34), $I^2$ 80%) (see Supplemental Material, Figure S1). Characteristics of the 27 studies included are shown in Table 1 (An et al. 1992; Chen et al. 1991; Fan et al. 2007; Guo et al. 1991; Hong et al. 2001; Li et al. 2003; Li et al. 2009; Li et al. 2010; Lin et al. 1991; Lin et al. 1994; Lin et al. 1995; Lu et al. 2000; Poureslami et al. 2011; Ren et al. 1989; Seraj et al. 2006; Sun et al. 1991; Wang et al.
1996; Wang et al. 2001; Wang et al. 2006; Wang et al. 2007; Xiang et al. 2003; Xu et al. 1994; Yang et al. 1994; Yao et al. 1996; Yao et al. 1997; Zhang et al. 1998; Zhao et al. 1996). Two of the studies included in the analysis were conducted in Iran (Poureslami et al. 2011; Seraj et al. 2006), otherwise the study cohorts were populations from China. Two cohorts were exposed to fluoride from coal burning (Guo et al. 1991; Li et al. 2010), otherwise populations were exposed to fluoride through drinking water. The CRT-RC was used to measure the children’s intelligence in 16 studies. Other intelligence measures included the Weschler Intelligence tests (3 studies), Binet IQ test (2 studies), Raven’s test (2 studies), Japan IQ test (2 studies), Chinese comparative intelligence test (1 study), and the mental work capacity index (1 study). As each of the intelligence tests used are designed to measure general intelligence, we used data from all eligible studies to estimate the possible effects of fluoride exposure on general intelligence.

In addition, we conducted a sensitivity analysis restricted to studies that used similar tests to measure the outcome (specifically, the CRT-RC, Weschler Intelligence test, Binet IQ test, or Raven’s test), and an analysis restricted to studies that used the CRT-RC. We also performed an analysis that excluded studies with co-exposures including iodine and arsenic, or with non-drinking water fluoride exposure from coal burning.

**Pooled SMD estimates**

Among the 27 studies, all but one study showed random-effect SMD estimates that indicated an inverse association, ranging from -0.95 (95% CI: -1.16, -0.75) to -0.10 (95% CI: -0.25, 0.04) (Figure 2). The study with a positive association reported a SMD estimate of 0.07 (95% CI: -0.8, 0.22). Similar results were found with the fixed-effect SMD estimates. The fixed-effects pooled SMD estimate and corresponding 95% CI were -0.40 (-0.44, -0.35), with a p-value <0.001 for
the test for homogeneity. The random-effects SMD estimate and 95% CI were -0.45 (95% CI: -0.56, -0.34) with an $I^2$ of 80% and homogeneity test p-value <0.001 (Figure 2). Because of heterogeneity (excess variability) between study results, we primarily used the random-effects model for subsequent sensitivity analyses, which is generally considered to be the more conservative method (Egger et al. 2001). Among the restricted sets of intelligence tests, the SMD for the model with only CRT-RC tests and drinking-water exposure (and to a lesser extent the model with only CRT-RC tests) was lower than that for all studies combined, although the difference did not appear to be significant. Heterogeneity, however, remained at a similar magnitude when the analyses were restricted (Table 2).

Sources of heterogeneity

We performed meta-regression models to assess study characteristics as potential predictors of effect. Information on the child’s gender and parental education were not reported in more than 80% of the studies, and only 7% of the studies reported household income. These variables were therefore not included in the models. Among the two covariates, year of publication (0.02; 95% CI: 0.006, 0.03), but not mean age of the study children (-0.02; 95% CI: -0.094, 0.04), was a significant predictor in the model with all 27 studies included. $I^2$ residual 68.7%, represented the proportion of residual between-study variation due to heterogeneity. From the adjusted $R^2$, 39.8% of between-study variance was explained by the two covariates. The overall test of the covariates was significant (p=0.004).

When the model was restricted to the 16 studies that used the CRT-RC, the child’s age (but not year of publication) was a significant predictor of the SMD. The $R^2$ of 65.6% of between-study variance was explained by the two covariates, and only 47.3% of the residual variation was due
to heterogeneity. The overall test of both covariates in the model remained significant (p = 0.0053). On further restriction of the model to exclude the 7 studies with arsenic and iodine as co-exposures and fluoride originating from coal-burning, thus including only the 9 with fluoride exposure from drinking water, neither age nor year of publication was a significant predictor, and the overall test of covariates was less important (p = 0.062), in accordance with the similarity of intelligence test outcomes and the source of exposure in the studies included. Although official reports of lead concentrations in the study villages in China were not available, some studies reported high percentage (95 to 100%) of low lead exposure (less than the standard of 0.01 mg/L) in drinking water samples in villages from several study provinces (Bi et al. 2010; Peng et al. 2008; Sun 2010).

Publication bias
A Begg’s funnel plot with the SE of SMD from each study plotted against its corresponding SMD did not show clear evidence of asymmetry, though two studies with a large SE also reported relatively large effect estimates, which may be consistent with publication bias or heterogeneity (Figure 3). The plot appears symmetrical for studies with larger SE, but with substantial variation in SMD among the more precise studies, consistent with the heterogeneity observed among the studies included in the analysis. Begg (p = 0.22) and Egger (p = 0.11) tests did not indicate significant (p < 0.05) departures from symmetry.

Pooled risk ratios
The relative risk of a low/marginal score on the CRT-RC test (<80) among children with high fluoride exposure compared to those with low exposure (16 studies total) was 1.93 (95% CI:
1.46, 2.55; I^2 58.5%). When the model was restricted to 9 studies that used the CRT-RC and included only drinking water fluoride exposure (Chen et al. 1991; Fan et al. 2007; Li et al. 1995; Li et al. 2003; Li et al. 2010; Lu et al. 2000; Wang et al. 2006; Yao et al. 1996, 1997), the estimate was similar (RR 1.75; 95% CI: 1.16, 2.65; I^2 70.6%). Although fluoride exposure showed inverse associations with test scores, the available exposure information did not allow a formal dose-response analysis. However, dose-related differences in test scores occurred at a wide range of water-fluoride concentrations.

**Discussion**

Findings from our meta-analyses of 27 studies published over 22 years suggest an inverse association between high fluoride exposure and children’s intelligence. Children who lived in areas with high fluoride exposure had lower IQ scores than those who lived in low exposure or control areas. Our findings are consistent with an earlier review (Tang et al. 2008), although ours more systematically addressed study selection and exclusion information, and more comprehensive in 1) including nine additional studies, 2) performing meta-regression to estimate the contribution of study characteristics as sources of heterogeneity, and 3) estimating pooled risk ratios for the association between fluoride exposure and a low/marginal Raven’s test score.

As noted by the NRC committee (NRC 2006), assessments of fluoride safety have relied on incomplete information on potential risks. In regard to developmental neurotoxicity, much information has in fact been published, although mainly as short reports in Chinese that have not been available to most expert committees. We carried out an extensive review that includes epidemiological studies carried out in China. While most reports were fairly brief and complete information on covariates was not available, the results tended to support the potential for
fluoride-mediated developmental neurotoxicity at relatively high levels of exposure in some studies. We did not find conclusive evidence of publication bias, though there was substantial heterogeneity among studies. Drinking-water may contain other neurotoxicants, such as arsenic, but exclusion of studies including arsenic and iodine as co-exposures in a sensitivity analysis resulted in a lower estimate, although the difference was not significant. The exposed groups had access to drinking-water with fluoride concentrations up to 11.5 mg/L (Wang et al. 2007), thus in many cases concentrations were above the levels of 0.7-1.2 mg/L (HHS) and 4.0 mg/L (US EPA) considered acceptable in the US. A recent cross-sectional study based on individual-level measure of exposures suggested that low levels of water fluoride (range 0.24 to 2.84 mg/L) had significant negative associations with child’s intelligence (Ding et al. 2011). This study was not included in our meta-analysis, which focused only on studies with exposed and reference groups, thereby precluding estimation of dose-related effects.

The results suggest that fluoride may be a developmental neurotoxicant that affects brain development at exposures much below those that can cause toxicity in adults (Grandjean 1982). For neurotoxicants, such as lead and methylmercury, adverse effects are associated with blood concentrations as low as 10 nmol/L. Serum-fluoride concentrations associated with high intakes from drinking-water may exceed 1 mg/L, or 50 μmol/L, thus more than 1000-times the levels of some other neurotoxicants that cause neurodevelopmental damage. Supporting the plausibility of our findings, rats exposed to 1 ppm (50 μmol/L) of water-fluoride for one year showed morphological alterations in the brain and increased levels of aluminum in brain tissue compared with controls (Varner et al. 1998).

The estimated decrease in average IQ associated with fluoride exposure based on our analysis may seem small and may be within the measurement error of IQ testing. However, as
research on other neurotoxicants has shown, a shift to the left of IQ distributions in a population will have substantial impacts, especially among those in the high and low ranges of the IQ distribution (Bellinger 2007).

The present study cannot be used to derive an exposure limit, as the actual exposures of the individual children are not known. Misclassification of children in both high- and low-exposure groups may have occurred if the children were drinking water from other sources (e.g., at school or in the field).

The published reports clearly represent independent studies and are not the result of duplicate publication of the same studies (we removed two duplicates). Several studies (Hong et al. 2001; Lin et al. 1991; Wang et al. 2001; Wang et al. 2007; Xiang et al. 2003; Zhao et al. 1996) report other exposures, such as iodine, and arsenic, a neurotoxicant, but our sensitivity analyses showed similar associations between high fluoride exposure and the outcomes even after these studies were excluded. Large tracts of China have superficial fluoride-rich minerals with little, if any, likelihood of contamination by other neurotoxicants that would be associated with fluoride concentrations in drinking water. From the geographical distribution of the studies, it seems unlikely that fluoride-attributed neurotoxicity could be due to other water contaminants.

Still, each of the articles reviewed had deficiencies, in some cases rather serious, which limit the conclusions that can be drawn. However, most deficiencies relate to the reporting, where key information was missing. The fact that some aspects of the study were not reported limits the extent to which the available reports allow a firm conclusion. Some methodological limitations were also noted. Most studies were cross-sectional, but this study design would seem appropriate in a stable population where water supplies and fluoride concentrations have remained unchanged for many years. The current water-fluoride level likely also reflects past
developmental exposures. In regard to the outcomes, the inverse association persisted between studies using different intelligence tests, although most studies did not report age adjustment of the cognitive test scores.

Fluoride has received much attention in China, where widespread dental fluorosis indicates the prevalence of high exposures. In 2008, the Ministry of Health reported that fluorosis was found in 28 provinces with 92 million residents (China News, 2008). Although microbiologically safe, water supplies from small springs or mountain sources created pockets of increased exposures near or within areas of low exposures, thus representing exposure settings close to the ideal, as only the fluoride exposure would differ between nearby neighborhoods. Chinese researchers took advantage of this fact and published their findings, though mainly in Chinese journals, and according to the standards of science at the time. This research dates back to the 1980s, but has not been widely cited at least in part because of limited access to Chinese journals.

In its review of fluoride, the US National Research Council (NRC 2006) emphasized that both the beneficial effects of fluoride on dental health and its adverse effects were incompletely documented. Our comprehensive review substantially extends the scope of research available for evaluation and analysis. Although the studies were generally of insufficient quality, the consistency of their findings adds support to existing evidence of fluoride-associated cognitive deficits, and suggests that potential developmental neurotoxicity of fluoride should be a high research priority. While reports from WHO and national agencies have generally focused on beneficial effects (CDC 1999; Petersen and Lennon 2004), the NRC report emphasized the need to consider potential adverse effects as well as benefits of fluoride exposure (NRC 2006).
In conclusion, our results support the possibility of adverse effects of fluoride exposures on children’s neurodevelopment. Future research should formally evaluate dose-response relations based on individual-level measures of exposure over time, including more precise prenatal exposure assessment and more extensive standardized measures of neurobehavioral performance, in addition to improving assessment and control of potential confounders.
References


China National Knowledge Infrastructure (CNKI; http://www.cnki.net) [accessed 25 May 2010].


EMBASE (Elsevier B.V., Amsterdam, the Netherlands; http://www.embase.com) [accessed 10 April 2011].


### Table 1. Characteristics of epidemiological studies of fluoride exposure and child’s cognitive outcomes

<table>
<thead>
<tr>
<th>Reference</th>
<th>Study Location</th>
<th>No. in high exposure group</th>
<th>No. in reference group</th>
<th>Age range (years)</th>
<th>Fluoride exposure Assessment</th>
<th>Fluoride exposure Range</th>
<th>Outcome Measure</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ren et al. 1989</td>
<td>Shandong, China</td>
<td>160</td>
<td>169</td>
<td>8-14</td>
<td>High/low F villages</td>
<td>Not specified</td>
<td>Wechler Intelligence test</td>
<td>Children in high F region had lower IQ scores. The average IQ of children from high fluoride area were lower than that of the reference area.</td>
</tr>
<tr>
<td>Chen et al. 1991</td>
<td>Shanxi, China</td>
<td>320</td>
<td>320</td>
<td>7-14</td>
<td>Drinking water</td>
<td>4.55 mg/L (high); 0.89 mg/L (reference)</td>
<td>CRT-RC&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Average IQ in fluoride coal burning area was lower than that in the reference area.</td>
</tr>
<tr>
<td>Guo et al. 1991</td>
<td>Hunan, China</td>
<td>60</td>
<td>61</td>
<td>7-13</td>
<td>F in coal burning</td>
<td>118.1-1361.7 mg/kg (coal burning area); Control area used wood</td>
<td>Chinese Binet</td>
<td>Children in the high fluoride (low iodine) area had lower IQ scores compared with the children from the reference fluoride (low iodine) areas.</td>
</tr>
<tr>
<td>Lin et al. 1991</td>
<td>Xinjiang, China</td>
<td>33</td>
<td>86</td>
<td>7-14</td>
<td>Drinking water</td>
<td>0.88 mg/L (high); 0.34 mg/L (reference)</td>
<td>CRT-RC&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Mean IQ was lower in all age groups except ≤7 years old group in the area with high fluoride and aluminum (limited to high fluoride population only).</td>
</tr>
<tr>
<td>Sun et al. 1991</td>
<td>Guizhang, China</td>
<td>196</td>
<td>224</td>
<td>6.5-12</td>
<td>Rate of fluorosis</td>
<td>Fluorosis: 98.36% (high); not specified (reference)</td>
<td>Japan IQ test</td>
<td>IQ scores of children in high fluoride areas were significantly lower than those of children living in reference fluoride area. Early, prolonged high fluoride intake causes a decrease in the child’s mental work capacity.</td>
</tr>
<tr>
<td>An et al. 1992</td>
<td>Inner Mongolia, China</td>
<td>121</td>
<td>121</td>
<td>7-16</td>
<td>Drinking water</td>
<td>2.1-7.6 mg/L (high); 0.6-1.0 mg/L (reference)</td>
<td>Wechsler Intelligence test</td>
<td></td>
</tr>
<tr>
<td>Li et al. 1994</td>
<td>Sichuan, China</td>
<td>106</td>
<td>49</td>
<td>12-13</td>
<td>Buring of high-fluoride coal to cook grain in high fluoride area</td>
<td>4.7-31.6 mg/kg (high); 0.5 mg/kg (reference)</td>
<td>Child mental work capacity</td>
<td></td>
</tr>
<tr>
<td>Xu et al. 1994</td>
<td>Shandong, China</td>
<td>97</td>
<td>32</td>
<td>8-14</td>
<td>Drinking water</td>
<td>1.8 mg/L (high); 0.8 mg/L (reference)</td>
<td>Binet-Siman</td>
<td>Children had lower IQ scores in high fluoride area than those who lived in the reference area. The average IQ scores was lower in children from high fluoride and iodine area than those from the reference area, but the results were not significant.</td>
</tr>
<tr>
<td>Yang et al. 1994</td>
<td>Shandong, China</td>
<td>30</td>
<td>30</td>
<td>8-14</td>
<td>Well water</td>
<td>2.97 mg/L (high); 0.5 mg/L (reference)</td>
<td>Chinese comparative intelligence test</td>
<td></td>
</tr>
<tr>
<td>Li et al. 1995</td>
<td>Guizhou, China</td>
<td>681</td>
<td>226</td>
<td>8-13</td>
<td>Urine, Dental Fluorosis Index (DFI)</td>
<td>1.81-2.69 mg/L (high); 1.02 mg/L (reference); DFI 0.8-3.2 (high); DFI &lt;0.4 (reference)</td>
<td>CRT-RC&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Average IQ score was lower in children in the high fluoride group than those in the reference group.</td>
</tr>
<tr>
<td>Wang et al. 1996</td>
<td>Shaanxi, China</td>
<td>147</td>
<td>83</td>
<td>4-7</td>
<td>Drinking water</td>
<td>1.08-8.6 mg/L (high); 0.58-1.0 mg/L (reference)</td>
<td>Wechsler Intelligence Test</td>
<td></td>
</tr>
<tr>
<td>Yao et al. 1996</td>
<td>Liaoning, China</td>
<td>266</td>
<td>270</td>
<td>8-12</td>
<td>Drinking water</td>
<td>2-11 mg/L (high); 1 mg/L (reference)</td>
<td>CRT-RC&lt;sup&gt;a&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>Zhao et al. 1996</td>
<td>Shaanxi, China</td>
<td>160</td>
<td>160</td>
<td>7-14</td>
<td>Drinking water</td>
<td>4.12 mg/L (high); 0.91 mg/L (reference)</td>
<td>CRT-RC&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Children living in high fluoride and arsenic area had significantly lower IQ scores than those living in the reference fluoride (and no arsenic) area.</td>
</tr>
<tr>
<td>Reference</td>
<td>Study Location</td>
<td>No. in high exposure group</td>
<td>No. in reference group</td>
<td>Age range (years)</td>
<td>Fluoride exposure</td>
<td>Assessment</td>
<td>Range</td>
<td>Outcome Measure</td>
</tr>
<tr>
<td>-------------------</td>
<td>----------------</td>
<td>---------------------------</td>
<td>------------------------</td>
<td>-------------------</td>
<td>-------------------</td>
<td>------------</td>
<td>---------------------------</td>
<td>----------------</td>
</tr>
<tr>
<td>Yao et al. 1997</td>
<td>Liaoning, China</td>
<td>188</td>
<td>314</td>
<td>7-14</td>
<td>Drinking water</td>
<td>2 mg/L</td>
<td>0.4 mg/L (reference)</td>
<td>CRT-RC</td>
</tr>
<tr>
<td>Zhang et al. 1998</td>
<td>Xinjiang, China</td>
<td>51</td>
<td>52</td>
<td>4-10</td>
<td>Drinking water</td>
<td>Not specified</td>
<td>Japan IQ Test</td>
<td></td>
</tr>
<tr>
<td>Lu et al. 2000</td>
<td>Tianjin, China</td>
<td>60</td>
<td>58</td>
<td>10-12</td>
<td>Drinking water</td>
<td>3.15 mg/L</td>
<td>0.37 mg/L (reference)</td>
<td>CRT-RC</td>
</tr>
<tr>
<td>Hong et al. 2001</td>
<td>Shandong, China</td>
<td>85</td>
<td>32</td>
<td>8-14</td>
<td>Drinking water</td>
<td>2.90 mg/L</td>
<td>0.75 mg/L (reference)</td>
<td>CRT-RC</td>
</tr>
<tr>
<td>Wang et al. 2001</td>
<td>Shandong, China</td>
<td>30</td>
<td>30</td>
<td>8-12</td>
<td>Drinking water</td>
<td>2.97 mg/L</td>
<td>0.5 mg/L (reference)</td>
<td>CRT-RC</td>
</tr>
<tr>
<td>Li et al. 2003</td>
<td>Inner Mongolia</td>
<td>720</td>
<td>236</td>
<td>6-13</td>
<td>Fluorosis</td>
<td>Endemic versus control regions defined by the Chinese Geological Office</td>
<td>CRT-RC</td>
<td>Average IQ of children in high fluorosis area was lower than that in the reference area</td>
</tr>
<tr>
<td>Xiang et al. 2003</td>
<td>Jiangsu, China</td>
<td>222</td>
<td>290</td>
<td>8-13</td>
<td>Drinking water</td>
<td>0.57-4.5 mg/L (high)</td>
<td>CRT-RC</td>
<td>Mean IQ score was significantly lower in children who lived in the high fluoride area than that of children in the reference exposure area (both areas also had arsenic exposure)</td>
</tr>
<tr>
<td>Seraj et al. 2006</td>
<td>Tehran, Iran</td>
<td>41</td>
<td>85</td>
<td>Not specified</td>
<td>Drinking water</td>
<td>2.5 mg/L</td>
<td>0.4 mg/L (reference)</td>
<td>Raven</td>
</tr>
<tr>
<td>Wang et al. 2006</td>
<td>Shanxi, China</td>
<td>202</td>
<td>166</td>
<td>8-12</td>
<td>Drinking water</td>
<td>5.54±3.88 mg/L (high)</td>
<td>CRT-RC</td>
<td>The IQ scores of children in the high fluoride group were significantly lower than those in the reference group</td>
</tr>
<tr>
<td>Fan et al. 2007</td>
<td>Shaanxi, China</td>
<td>42</td>
<td>37</td>
<td>7-14</td>
<td>Drinking water</td>
<td>1.14-6.09 mg/L (high)</td>
<td>CRT-C2</td>
<td>The average IQ scores of children residing in the high F area were lower than those of children residing in the reference area</td>
</tr>
<tr>
<td>Wang et al. 2007</td>
<td>Shanxi, China</td>
<td>253</td>
<td>196</td>
<td>8-12</td>
<td>Drinking water and urine</td>
<td>3.8-11.5 mg/L (water, high)</td>
<td>CRT-RC</td>
<td>Mean IQ scores were significantly lower in the high fluoride group than from the reference group in the fluoride/arsenic areas</td>
</tr>
<tr>
<td>Li et al. 2009</td>
<td>Hunan, China</td>
<td>60</td>
<td>20</td>
<td>8-12</td>
<td>Coal burning</td>
<td>1.24-2.34 mg/L (high)</td>
<td>CRT-C2</td>
<td>Mean IQ was lower in children in coal-burning areas compared to those in the reference group</td>
</tr>
<tr>
<td>Reference</td>
<td>Study Location</td>
<td>No. in high exposure group</td>
<td>No. in reference group</td>
<td>Age range (years)</td>
<td>Fluoride exposure</td>
<td>Assessment</td>
<td>Range</td>
<td>Outcome Measure</td>
</tr>
<tr>
<td>-------------------</td>
<td>----------------</td>
<td>---------------------------</td>
<td>------------------------</td>
<td>-------------------</td>
<td>-------------------</td>
<td>-------------</td>
<td>------------------</td>
<td>-----------------</td>
</tr>
<tr>
<td>Li et al. 2010</td>
<td>Henan, China</td>
<td>347</td>
<td>329</td>
<td>7-10</td>
<td>Drinking water</td>
<td>2.47±0.75mg/L (high)</td>
<td>CRT-RC</td>
<td>No significant difference in IQ scores between children in the exposed and reference groups</td>
</tr>
<tr>
<td>Poureslami et al. 2011</td>
<td>Iran</td>
<td>59</td>
<td>60</td>
<td>6-9</td>
<td>Drinking Water</td>
<td>2.38 mg/L (high) 0.41 mg/L (reference)</td>
<td>Raven</td>
<td>Children in the high fluoride group scored significantly lower than those in reference group</td>
</tr>
</tbody>
</table>

*CRT-RC denotes Chinese Standardized Raven Test, rural version (Wang et al. 1989)*
Table 2. Sensitivity analyses of pooled random-effect standardized weighted mean difference (SMD) estimates of child’s intelligence score with high exposure of fluoride

<table>
<thead>
<tr>
<th>Model</th>
<th>Available studies for analysis</th>
<th>SMD (95% CI)</th>
<th>I²</th>
<th>p-value test of heterogeneity</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Exclude non-standardized tests&lt;sup&gt;a&lt;/sup&gt;</td>
<td>23</td>
<td>-0.44 (-0.54, -0.33)</td>
<td>77.6%</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>2. Exclude non CRT-RC Tests&lt;sup&gt;b&lt;/sup&gt;</td>
<td>16</td>
<td>-0.36 (-0.48, -0.25)</td>
<td>77.8%</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>3. Exclude studies with other exposures (Iodine, Arsenic)&lt;sup&gt;c&lt;/sup&gt; or non-drinking water fluoride exposure&lt;sup&gt;d&lt;/sup&gt;</td>
<td>9</td>
<td>-0.29 (-0.44, -0.14)</td>
<td>81.8%</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

<sup>a</sup>Mental work capacity (Li et al. 1994); Japan IQ (Sun et al. 1991; Zhang et al. 1998); Chinese comparative scale of intelligence test (Yang et al. 1994)

<sup>b</sup>Weschler intelligence test (An et al. 1992; Ren et al. 1989; Wang et al. 1996); Chinese Binet IQ (Guo et al. 1991); Raven (Poureslami et al. 2011; Seraj et al. 2006); Binet-Siman (Xu et al. 1994)

<sup>c</sup>Iodine (Hong et al. 2001; Lin et al. 1991; Wang et al. 2001); Arsenic (Wang et al. 2007; Xiang et al. 2003; Zhao et al. 1996; Zhang et al. 1998 - already excluded, see footnote 1)

<sup>d</sup>Fluoride from coal-burning (Li et al. 2009; Guo et al. 1991; Li et al 1994 (already excluded, see footnotes a and b)}
Figure Legend

Figure 1. Flow diagram of the meta-analysis

Figure 2. Random-effect standardized weighted mean difference (SMD) estimates and 95% CIs of child’s intelligence score associated with high exposure to fluoride. SMs for individual studies are shown as solid diamonds (♦), and the pooled SMD is shown as a non-filled diamond (◊). Horizontal lines represent 95% CIs for the study-specific SMDs.

Figure 3. Begg’s funnel plot showing individual studies included in the analysis according to random-effect standardized weighted mean difference (SMD) estimates (x-axis) and the standard error (se) of each study-specific SMD (y-axis). The solid vertical line indicates the pooled SMD estimate for all studies combined and the dashed lines indicated pseudo 95% confidence limits around the pooled SMD estimate.
Total abstracts identified from literature search (N=39)

- Duplicate records removed (N=2)
- Studies excluded because they did not meet inclusion criteria (N=3)

Studies for retrieval of detailed information (N=34)

- Studies with missing information on outcomes (N=6)
- Studies excluded due to questionably small standard deviations (N=1)

Studies included in meta-analysis (N=27)
-0.45 (-0.56, -0.34)
-0.64 (-0.82, -0.46)
-0.62 (-0.98, -0.25)
-0.55 (-0.70, -0.39)
-0.38 (-0.65, -0.10)
-0.50 (-0.80, -0.08)
-0.44 (-0.70, -0.39)
-0.38 (-0.65, -0.10)
-0.57 (-0.93, -0.21)
-0.40 (-0.74, -0.06)
-0.93 (-1.35, -0.52)
-0.50 (-1.01, 0.02)
-0.65 (-1.01, -0.33)
-0.35 (-0.70, -0.06)
-0.34 (-0.51, -0.17)
-0.54 (-0.76, -0.31)
-0.43 (-0.61, -0.25)
-0.17 (-0.55, 0.22)
-0.62 (-0.98, -0.25)
-0.44 (-0.85, -0.03)
-0.50 (-1.01, 0.02)
-0.10 (-0.25, 0.04)
-0.64 (-0.82, -0.46)
-0.89 (-1.28, -0.50)
-0.27 (-0.47, -0.06)
-0.17 (-0.61, 0.27)
-0.26 (-0.44, -0.07)
-0.43 (-0.94, 0.08)
0.07 (-0.08, 0.22)
-0.41 (-0.77, -0.04)
-0.45 (-0.56, -0.34)