

UC Irvine

UC Irvine Previously Published Works

Title

Impacts of geocoding uncertainty on reconstructed PFOA exposures and their epidemiological association with preeclampsia.

Permalink

<https://escholarship.org/uc/item/2x03x48x>

Authors

Avanasi, Raghavendhran
Shin, Hyeong-Moo
Vieira, Veronica M
et al.

Publication Date

2016-11-01

DOI

10.1016/j.envres.2016.08.019

Peer reviewed



HHS Public Access

Author manuscript

Environ Res. Author manuscript; available in PMC 2018 March 13.

Published in final edited form as:

Environ Res. 2016 November ; 151: 505–512. doi:10.1016/j.envres.2016.08.019.

Impacts of geocoding uncertainty on reconstructed PFOA exposures and their epidemiological association with preeclampsia

Raghavendhran Avanas¹, Hyeong-Moo Shin², Veronica M. Vieira^{1,3}, and Scott M. Bartell^{1,3,4}

¹Environmental Health Sciences Graduate Program, 2032, AIRB, University of California, Irvine, CA 92697-3957

²Department of Public Health Sciences, One Shields Avenue, MS1-C, Davis, CA 95616-8638

³Program in Public Health, AIRB, University of California, Irvine, CA 92697-3957

⁴Department of Statistics and Department of Epidemiology, University of California, Irvine, California, U.S.A

Abstract

Many epidemiology studies have investigated associations of perfluorooctanoate (PFOA) exposures with a variety of adverse health outcomes for participants in the C8 Health Project. The exposure concentrations (i.e., air and groundwater) used in these studies were determined primarily based on participant's residential locations. However, for residential addresses that could not be geocoded to the street level, the exposure concentrations were assigned based on population-weighted ZIP code centroid, which may result in exposure mischaracterization. The aim of this current study is to evaluate the potential impact of mischaracterized exposure concentrations due to geocoding uncertainty on the predicted serum PFOA concentrations and the epidemiological association between PFOA exposure and preeclampsia. For both workplace addresses and incompletely geocoded residential addresses, we used Monte Carlo (MC) simulation to assign alternate geographic locations within the reported ZIP code (instead of population-weighted ZIP code centroids) and the corresponding exposure concentrations. We found that mischaracterization of residential exposure due to population-weighted ZIP code centroid assignment had no significant impact on the serum PFOA concentration predictions and the epidemiological association of PFOA exposure with preeclampsia. In contrast, the uncertainty in workplace exposure moderately impacted the rank exposure among the participants. We observed a 41% increase in the average adjusted odds ratio of preeclampsia occurrence that may be due to differing proportions of cases (64.3%) and controls (54.5%) with workplace address geocodes during pregnancy. This finding suggests that differential exposure mischaracterization can be reduced by obtaining accurate exposure information such as street addresses and tap water consumption, for both workplaces and residences. The analysis we present is one approach for

Corresponding Author: Raghavendhran Avanas¹, M.S., Ph.D., 2030D, AIRB, University of California, Irvine, Irvine, CA 92697-3957, Phone: 8062246848, ravanasi@uci.edu.

We obtained approval (HS#2013-9421) from the Institutional Review Board at the University of California, Irvine, to work with the human subject data in this current study.

estimating the potential impacts of positional errors in a geocoding-based exposure assessment on exposure estimates and epidemiological study results.

Keywords

C8 Health Project; geocoding uncertainty; differential exposure mischaracterization; perfluorooctanoate; preeclampsia

1.1 Introduction

Geographic Information Systems (GIS) have been used in numerous environmental health studies for assessing the exposure of participants to contaminants of interest via proximity analysis, integration of environmental monitoring data, individual-level exposure estimation, design of exposure metrics, and reconstructing exposure through activity patterns (Ali et al., 2002; Bell et al., 2001; Bellander et al., 2001; Beyea and Hatch, 1999; Elgethun et al., 2003; Floret et al., 2003; Nuckols et al., 2004; Reynolds et al., 2003; Rull and Ritz, 2003; Shin et al., 2011a; Vieira et al., 2010; Vieira et al., 2013). The use of GIS in environmental exposure assessment can improve our understanding of the associations between environmental exposures and adverse health outcomes (Beyea and Hatch, 1999; Nuckols et al., 2004).

Geocoding, the process of matching addresses to geographic locations (latitude and longitude), is an important step in using GIS for exposure assessment (Bonner et al., 2003). One primary application of geocoding is to assign individual-level environmental exposures based on their location in an exposed geographic area (Elgethun et al., 2003; Shin et al., 2011a; Vieira et al., 2013; Ward et al., 2005). Partial matching of addresses, such as a street name without the house number or a ZIP code without a specific street, or errors in geocoding can lead to positional errors in the exposure assessment, potentially leading to exposure mischaracterization. This can impact the validity of the epidemiological studies that use the resulting exposure estimates (Bonner et al., 2003; Elgethun et al., 2003; Vieira et al., 2010; Vieira et al., 2013). Researchers and the National Institutes of Health have called for more investigation into the potential impacts of geocoding uncertainty on the results of epidemiological studies (Henry and Boscoe, 2008; US Department of Health and Human Services, 2014; Zandbergen, 2009). A recent report from a Health and Environmental Sciences Institute (HESI) workshop also recommended the characterization and evaluation of uncertainty in environmental epidemiology studies to better understand the potential sources of bias and to utilize results from epidemiological analyses for risk assessment (Burns et al., 2014).

The C8 Science Panel studies investigated associations of perfluorooctanoate (PFOA) serum concentrations predicted by a GIS-based exposure assessment (Shin et al., 2011a, b) with a variety of adverse health outcomes such as ulcerative colitis, kidney and testicular cancer, pregnancy outcomes, abnormal thyroid function, and abnormal kidney function (Barry et al., 2013; C8 Science Panel, 2011; Lopez-Espinosa et al., 2012; Savitz et al., 2012a; Savitz et al., 2012b; Steenland et al., 2013; Watkins et al., 2013). Predicted serum PFOA concentrations for 2005–2006 were well correlated ($r_s = 0.68$) with measured serum PFOA concentrations in the same year. Geocoding was used to locate participant residential

addresses geographically to assign air and water PFOA concentrations for each year, over 58 years-1951 to 2008. This was done by spatially joining the addresses with the pipe distribution networks of the six participating public water districts (PWDs) to which all the consented participants of the C8 Health Project belonged (Shin et al., 2011b; Vieira et al., 2013). About 12% of the addresses (mostly rural addresses) with ZIP codes within the six PWDs could not be geocoded and thus population weighted ZIP code centroids were used to assign PWDs and the corresponding PFOA water concentrations. The assignment of population weighted ZIP code centroids for addresses that could not be geocoded to the street level can be considered as a single geographic imputation method (analogous to a mean imputation method). Such imputation or geocoding at a coarse spatial resolution can introduce geographic bias/positional errors in the exposure classification (Henry and Boscoe, 2008; Zandbergen, 2009). Also, it has been noted that there is greater potential for positional errors when geocoding rural addresses compared to geocoding urban addresses (Vieira et al., 2010; Ward et al., 2005).

The aim of this study is to evaluate the potential impacts of geocoding uncertainty on the estimated serum PFOA concentrations of participants in the C8 Health Project. Specifically, we examine the impacts of single geographic imputation, which may have resulted in mischaracterized water PFOA concentrations for those participants geocoded to population-weighted ZIP code centroids. We also examine the corresponding impact on the association between the estimated serum PFOA concentrations and the occurrence of preeclampsia (Savitz et al., 2012), an epidemiological analysis that has been discounted for the use of modeled rather than directly measured serum PFOA concentrations (Johnson and Sutton, 2014; Koustas et al., 2014). We use Monte Carlo (MC) simulation to assign alternate geographic locations within the reported ZIP code for all residential addresses that were geocoded to a population-weighted ZIP code centroid and the reported work addresses, and recalculate the prediction of serum PFOA concentrations and the epidemiological association with preeclampsia for each set of alternate geographic locations.

2.0 Materials and Methods

2.1. PFOA exposure assessment

The PFOA exposure assessment by Shin et al. (2011a, b) had two distinct modeling components. The first part of the PFOA exposure assessment used a suite of environmental fate and transport models to predict yearly PFOA outdoor air and groundwater concentrations for 1951–2008 in the region surrounding the Washington Works facility and the six impacted PWDs. Detailed explanation of the PFOA fate and transport modeling can be found in Shin et al. (2011a). Briefly, the modeling system utilized yearly PFOA release rates from the Washington Works facility, along with PFOA physicochemical properties, local meteorology, and hydrogeology to predict the yearly air and water concentrations of PFOA for the area serviced by the six PWDs: the City of Belpre, Little Hocking Water Association, Tappers Plains Chester Water District, the Village of Pomeroy Water District, Lubeck Public Service District, and Mason County Public Service District. The model also estimated PFOA exposure for shallower private drinking water wells located in the study area.

Next, an integrated exposure and pharmacokinetic model system was used to predict the yearly serum PFOA concentrations for all consented participants in the C8 Health Project study. This model system utilized the predicted yearly PFOA air and water concentrations (Shin et al., 2011a), standard inhalation and standard/self-reported tap water ingestion rates (U.S. EPA, 2009), PWD pipe distribution networks, along with self-reported participant information collected through a questionnaire as part of the C8 Health Project (Frisbee et al., 2009). These included detailed participant residential/work histories and participant demographics such as age, gender, and body weight. Based on the self-reported information, the drinking water source at each residential history was categorized as public, private, bottled water, or mixed. GIS was then used to link participant residential addresses with modeled air and water PFOA concentrations and predict yearly combined inhalation and ingestion (total) exposure doses for all the participants. A one-compartment pharmacokinetic model was then applied to estimate the yearly serum PFOA concentrations based on a single elimination half-life. More details on the exposure reconstruction/pharmacokinetic modeling are described by Shin et al. (2011b).

2.2. GIS

GIS methods were used to assign historical outdoor air and groundwater concentrations for each participant. With respect to ingestion exposure, GIS was used first to map the pipe distribution systems of the six PWDs included in the exposure modeling system (Shin et al., 2011b). Next, the participant residential addresses were geocoded using TeleAtlas and ArcView/NAVTEQ (Vieira et al., 2010). Among the residential addresses with ZIP codes in any of the six PWDs, approximately 12% of the addresses could not be geocoded to the street level (Shin et al., 2011b; Vieira et al., 2013) and hence, a population-weighted ZIP code centroid was used to assign environmental concentrations instead of the street level geocode. Later, within the GIS, the geocoded addresses were spatially joined with the PWD pipe distribution system to assign PWD-specific annual average PFOA water concentrations to those addresses that were serviced by any specific PWD. As described in the text and Figure 1 of the Shin et al. (2011b) study, based on the participant's geocoded residential address, the PFOA water concentrations were assigned for each reported residence for each participant. Any discrepancies between the self-reported water sources and the geocoded water sources (~ 9% of the addresses) were reviewed manually to determine the most likely source. For the participant work histories, the PFOA water concentrations were assigned based on self-reported public water sources. Street level addresses were not available for work histories but ZIP codes were reported for over half (55%) the work locations. 54.3% of the pregnancies had at least one reported work location during the year of pregnancy; this statistic was 54.8%, 52.5% and 41.1 % for 1 year, 2 years and 5 years previous to the year of the pregnancy. For participants with both residential and work histories, 70% of drinking water was assumed to come from the home and 30% from the work location (Shin et al., 2011b). For inhalation exposure, the participant's geocoded address and population-weighted ZIP centroid were used to assign PFOA outdoor air concentrations based on the annual average air concentration predictions (Shin et al., 2011a). For most of the study participants, drinking water ingestion was the major exposure route during the period of epidemiological investigations described below (Shin et al., 2011b; Vieira et al., 2013; Vieira et al., 2010).

2.3. Epidemiological study

One of the C8 Science Panel epidemiological studies focused on pregnancy outcomes including preeclampsia among 10,189 pregnancies (730 preeclampsia cases) that occurred between 1990 and 2006 in this population (Savitz et al., 2012a). The analysis used generalized estimating equations to estimate the association between preeclampsia and estimated serum PFOA in the year of pregnancy, adjusting for confounding by parity, maternal age, education level and smoking status. The study reported an adjusted odds ratio (AOR) of 1.13 (95% confidence interval (CI): 1.00, 1.28) per interquartile range (IQR) of log (natural) serum PFOA concentrations (nanograms per milliliter-ng/mL). We obtained approval from the Institutional Review Board (HS#2013-9421) at the University of California, Irvine to use those study data to conduct our MC analyses. We restricted our analysis to 10,149 pregnancies with 725 preeclampsia cases, by removing 25 mothers who had previously worked at the Washington Works facility. The resulting modified AOR per IQR was similar: 1.12 (95% CI: 1.00, 1.26). When we restricted the analysis to the subset of pregnancies which had only street-level geocoded residential addresses (n= 6883), AOR (95% CI) was found to be 1.16 (1.01, 1.33).

We utilized the same PFOA exposure assessment model system, the same epidemiological model, and MC simulation to evaluate the potential impact of positional errors due to the use of population weighted ZIP code centroids (instead of the actual known address geocodes) on the estimated serum PFOA concentrations and the association with preeclampsia. MATLAB (The Mathworks Inc., Natick, MA, 2000), R (<http://www.r-project.org/>), and ArcGIS (ESRI) were used to perform these analyses.

2.4. MC simulations I and II

In order to evaluate the impact of mischaracterized exposure due to geocoding uncertainty on PFOA serum concentration predictions and epidemiological associations with preeclampsia, we conducted two types of Monte Carlo (MC) simulations: (1) simulation I using residential addresses only and (2) simulation II using residential and work addresses.

In the MC simulation I, the geocodes (latitude and longitude coordinates) of those residential addresses that were originally assigned to a population-weighted ZIP code centroid were varied, and the serum PFOA concentration predictions and the epidemiological association with preeclampsia were re-calculated using the same exposure, pharmacokinetic, and epidemiological models. In each of 200 MC iterations (n= 200 was chosen based on the Monte Carlo error being < 1%), every residential address that had used a population-weighted ZIP code centroid was reassigned a randomly selected alternate geocode within the same ZIP code (thereby reassigning the PFOA water concentrations according to the new geocoded location). In the secondary analysis (MC simulation II), in addition to handling residential addresses as described in MC simulation I, for each work address we reassigned a randomly selected alternate geocode within the reported ZIP code and the corresponding PFOA water concentrations were assigned. The exposure assessment model and the epidemiological analysis were repeated for each MC iteration to obtain plausible new serum PFOA concentration predictions and the AOR for the association of PFOA and preeclampsia.

Approximately 7.6 % (n= 2,046) of the residential addresses reported by our study participants had originally been geocoded to a population-weighted ZIP code centroid. First, the ZIP codes (n=37) that were serviced by one of the 6 PWDs and the pipe distribution networks of the 6 PWDs are projected in the North American Datum of 1983 (NAD83) projection as shown in Figure 1. Then, a grid of points was created using the ZIP code extent (each ZIP code had at least 15 grid points and up to 489 grid points). The grid points were on average 905 meters apart in the 37 ZIP codes. Then during each iteration for MC simulation I, for each residential history address that used a population-weighted ZIP code centroid, a grid point was randomly sampled from within the corresponding ZIP code and the drinking water source and PFOA water concentration were re-assigned with those corresponding values from the sampled grid point. The random grid point represents a possible location (within the ZIP code for the specific residential history) of the participant's residence. In the MC simulation II, in addition to residential geocoding uncertainty discussed above, for each participant's work address a grid point was randomly sampled from within the corresponding ZIP code and the drinking water source and PFOA water concentration were re-assigned using the corresponding values from the sampled grid point. In the MC simulations, we studied the impact of geocoding uncertainty on the PFOA exposure only through drinking water ingestion and not through inhalation of contaminated air. Therefore, the inhalation exposures for the participants were not varied in the MC simulations I and II.

To illustrate the MC methodology, consider a participant who had a residential address in ZIP code of 45769, but a population-weighted ZIP code centroid was used due to insufficient address information. In our analysis, we created a grid of 335 points evenly spaced across this ZIP code area as seen in Figure 2 (panel a). This ZIP code is serviced by two different participating PWDs (Tuppers Plains and Pomeroy) and some parts of the ZIP code are not served by any of the participating PWDs and therefore treated as private wells as shown in Figure 2 (panel b). In the MC simulation, suppose a grid point 'A' was randomly sampled in the first iteration and used as the new residential address for that participant. PFOA water concentrations for Pomeroy PWD were then used in assigning the exposure for that iteration. For iteration 2, suppose a grid point 'B' was sampled and used as the new residential address for that participant. PFOA water concentrations for Tuppers Plains PWD would then be used to assign that participant's exposure for iteration 2. Alternately, if a grid point 'C' was sampled, the water source was treated as private and PFOA water concentrations from a shallow drinking water well in that location were assigned. Hence, for any participant with a residential address in ZIP code of 45769, there are three different possible assignments of PFOA water concentrations. The PFOA air concentrations were not varied, but were assigned as discussed in the GIS methodology section.

Following the reassignment of geocodes and ingestion exposure via new water concentration assignment, participant serum PFOA concentrations for each MC iteration were computed and the epidemiological model was fit to obtain the AOR of preeclampsia occurrence (per IQR), for each of the 200 iterations of the MC simulation. Summary statistics for the serum PFOA concentrations for the 10,149 participants were calculated for each MC simulation. We then compared the serum PFOA concentrations from the MC simulation with the

originally assigned serum PFOA concentrations by plotting the rank correlation between them for the 10,149 participants between the years 1990 and 2006. We also calculated summary statistics for the epidemiological results from the MC simulations (200 iterations) and compared them with the original AOR.

We also computed a measure of the relative contribution of geocoding uncertainty (uncertainty due to potential positional errors in the use of population-weighted ZIP code centroids versus street address geocodes) to the total uncertainty in the epidemiological association of PFOA with preeclampsia (in addition to the participant sampling variability calculated as part of the confidence interval of the epidemiological association) using the law of total variance as described in our previous uncertainty analysis (Avanasi et al., 2016a). In brief, the contribution of the geocoding uncertainty is calculated by the formula $\text{var}(b) = E(\text{var}(b|X)) + \text{var}(E(b|X))$. In this formula, b corresponds to the log odds parameter estimate, X is a collection of individual exposure estimates, E is the expected value and var is the variance. The relative contribution of geocoding uncertainty to the total uncertainty was calculated by the formula $\text{var}(E(b|X)) / \text{var}(b)$.

3.0 Results

The impact of the geocoding uncertainty on the serum PFOA concentration predictions (ng/mL) was studied by calculating median, mean, and 25th and 75th percentiles of each MC iteration for both MC simulations. These statistics were calculated for the subset of pregnancies with at least one residential history with a population-weighted ZIP code centroid (centroid subset, $n = 3,266$) and for all the 10,149 study participants in MC simulations I and II. The mean and 95% probability intervals (PI) of the above mentioned summary statistics among the 200 MC iterations in comparison with the modified Savitz et al. (2012a) serum PFOA concentrations are shown in Table 1. We found minimal to no impact on the serum PFOA concentration predictions due to the presence of the geocoding uncertainty in MC simulation I, while there was a moderate impact in MC simulation II (with the mean serum PFOA concentrations among all the participants increasing from 51.1 ng/mL in the modified original analysis to 55.5 ng/mL).

For the MC simulation I, we calculated the rank correlation between the simulated and the original serum PFOA concentrations for the centroid subset between the years 1990 and 2006 and the mean (95 % probability interval) over the 200 MC iterations was obtained for each year. The lowest mean rank correlation for the centroid subset ($n=3,266$) was that of year 1999: 0.92 (0.92, 0.93). On the other hand, the lowest mean rank correlation for all participants ($n=10,149$) was for the year 2002: 0.97 (0.96, 0.97), suggesting little change in the rank exposure among centroid subset participants after accounting for geocoding uncertainty. For the MC simulation II, the addition of geocoding uncertainty in work addresses caused a reduction of rank correlation compared with the MC simulation I. The lowest mean rank correlation for all the 10,149 participants was for the year 1999: 0.93 (0.92, 0.93).

The impact of geocoding uncertainty in the residential addresses (MC simulation I) on the AOR of preeclampsia occurrence was minimal with the mean and the 95% probability

interval of AOR being 1.12 (1.00, 1.25). This 95% probability interval includes the contribution of both the sampling variability among the participants and the geocoding uncertainty propagated in the MC simulation. Comparing it to the modified original analysis, the AOR per IQR (95% confidence interval) was 1.12 (1.00, 1.26). The contribution of the geocoding uncertainty in residential addresses only to the total uncertainty was found to be 1.1%. For the MC simulation II (geocoding uncertainty in both residential and work addresses), the AOR of preeclampsia occurrence increased with the mean and 95% probability interval of AOR was 1.17 (1.04, 1.32), which is a 41% increase in the average AOR, when compared with the AOR of 1.12 (1.00, 1.26) in the original modified analysis. The contribution of the geocoding uncertainty to the total uncertainty was found to be 2.6%.

4.0 Discussion

The preeclampsia epidemiology model was fit for the street-level geocoded subset (as described in section 2.3) and the AOR (95% CI) was found to be 1.16 (1.01, 1.33). This suggests that with accurate residential addresses without the use of participants with addresses using ZIP code centroids, we might expect a stronger association with preeclampsia in this population.

Geocoding uncertainty due to the use of population-weighted ZIP code centroids for exposure assessment had little impact on the serum PFOA concentration predictions of the participants in the Savitz et al. (2012a) study as seen in Table 1. The mean rank correlation between the MC simulation I predicted serum PFOA concentrations and the original modified serum PFOA concentration predictions was high (0.97), suggesting little change in the rank exposure among the participants. Subsequently, there was negligible impact on the association with preeclampsia. The contribution of geocoding uncertainty to total uncertainty (including participant sampling variability) was minor (1.1%). These results suggest that the use of ZIP centroids versus street level residential addresses does not substantially impact the validity of the reported association between serum PFOA concentrations and the occurrence of preeclampsia in the C8 Health Project population.

Interestingly, in MC simulation II, when we accounted for geocoding uncertainty in workplace addresses, there was a moderate increase in the mean and the 75th percentile serum PFOA concentrations (as seen in Table 1), and also a moderate decrease in the rank exposure among participants compared to that in MC simulation I. This indicates an increase for the association of PFOA and preeclampsia – a mean AOR (95% probability interval) of 1.17 (1.04, 1.32) compared to the original AOR of 1.12 (1.00, 1.26). For participants with reported work histories, addition of uncertainty in the spatial location of a work history within the self-reported ZIP code resulted in a 41% increase in the AOR of preeclampsia occurrence. Because MC simulation explores the impact of adding positional uncertainty to the geocodes rather than correcting for it (Gryparis et al., 2009; Avanasi et al., 2016a; Avanasi et al., 2016b), these results suggest that if we had more accurate locations of participant work addresses the AOR of preeclampsia occurrence might have been different than previously reported. Previous literature suggests that positional error due to inaccurate geocoding or geocoding rural route addresses can potentially lead to exposure mischaracterization and bias in epidemiological study results (Vieira et al., 2010; Vieira et

al., 2013; Elgethun et al., 2003; Bonner et al., 2003; Ward et al., 2005). We further investigated this increase in the average AOR result since previous literature suggests that non-differential exposure mischaracterization causes a bias towards the null (not away from the null) in epidemiological studies (Armstrong, 1998). In this specific epidemiological analysis, we found a different proportion of work addresses among cases (64.3%) compared to controls (53.5%) in the year of pregnancy. We think that this difference could potentially be responsible for a differential mischaracterization (instead of non-differential), with respect to the uncertainty in the work history of participants, resulting in an increase in the average AOR. In addition, from Table 1, we find that the mean serum PFOA concentrations among all the participants has increased from 51.1 (ng/mL) in the modified original analysis to 55.5 (ng/mL) for the MC simulation II, thereby contributing more to the potential differential exposure mischaracterization.

The relatively mild impact of residential address geocoding uncertainty can be expected as the residential addresses were usually available at the level of street address, and because the geocoded and self-reported water source assignments were manually crosschecked using GIS. In addition, only 7.6% of the participant residential histories used a population-weighted ZIP code centroid in this study. In contrast, more participants (as discussed in the GIS section earlier) had geocoding uncertainty in work histories due to the lack of street addresses. Alternative work location geocodes appear to be able to change participant water sources enough to modify the rank order of exposure and cause an increase in the mean AOR of preeclampsia.

We had previously studied other sources of uncertainty in this PFOA exposure assessment model (Shin et al., 2011a, b) including shared uncertainty in the PFOA water concentrations (Avanasi et al., 2016a) and inter-individual variability/epistemic uncertainty in independent exposure parameters such as the standard and self-reported water ingestions rates and pharmacokinetic parameters including PFOA elimination half-life and PFOA volume of distribution (Avanasi et al., 2016b). Our previous studies found that correlated uncertainty (shared uncertainty in the PWD PFOA water concentrations due to uncertainties in source emissions and our fate and transport model) had negligible impact on the rank order of exposure among participants and the AOR of association with preeclampsia, although it had substantial impact on the serum PFOA concentrations. In contrast, independent sources of error in water ingestion rates and pharmacokinetic parameters moderately influenced the rank exposure and caused a bias towards the null in the association with preeclampsia. Together with these two studies, the geocoding uncertainty analysis yields a detailed understanding of potential impacts of various sources of uncertainty in the PFOA exposure assessment modeling system on the specific epidemiological association with preeclampsia.

As a side analysis, we evaluated the epidemiological association based on the urban/rural status. The classification was based on the U.S. Census Bureau's recommendation ([US Census Bureau, 2012](#)), in which any ZIP code with a population density less than or equal to 500 people per square mile was considered to be a rural ZIP code. Based on this, we found that 68% (n = 6928) of the Savitz et al., 2012b study participants had at least one residential/workplace history in a rural ZIP code and were classified as rural while the rest were considered urban. The epidemiological results for the original modified Savitz analysis as

well as the MC simulation II were repeated for both groups. We found that the geocoding uncertainty in the rural group has a major impact in AOR, with an increase in AOR from 1.11 (CI: 0.97, 1.27) for the modified original analysis to 1.20 (PI: 1.04, 1.39) for the MC simulation II. Whereas, in the urban group, geocoding uncertainty did not have much impact in the AOR; with the AOR for the modified original analysis being 1.09 (0.91, 1.30) and that of the MC simulation II was 1.08 (0.90, 1.28). This supports previous literature that suggests that there is greater potential for positional errors when geocoding rural addresses compared to geocoding urban addresses (Vieira et al., 2010; Ward et al., 2005).

4.1. Limitations

Epidemiological studies of other health outcomes that were part of the C8 Science Panel studies might or might not have a similar result as they include different sets of participants with different residential and work histories. In addition, it has been suggested that the impact of errors in geocoding on exposure assessment depends on spatial variation of the exposure (Wards et al., 2005). Therefore, the results presented here can inform judgments about the reliability of the Savitz et al. (2012a) preeclampsia findings but may not be generalizable to the impact of geocoding uncertainty on other C8 Science Panel epidemiological studies, or other environmental epidemiological studies that used population-weighted ZIP code centroid geo-coordinates to represent non-geocoded addresses in their exposure assessment. Also, the current analysis investigates the impact of geocoding uncertainty (residential and work addresses) only on the PFOA exposure through drinking water ingestion. We did not consider inhalation route of exposure because the contribution of inhalation exposure to overall exposure for participants in the Savitz et al., 2012a study (between the years 1990 and 2006) was minimal as discussed in the Avanasi et al. (2016a) study. However, its inclusion could result in slight increases in the total uncertainty attributed to geocoding.

Our findings are also limited by sampling alternate residential and work locations from throughout the entire identified ZIP codes. Importantly, road maps of the region suggest that not all areas are developed or inhabited. Future analyses using MC simulation could restrict the grid to areas that are highly likely to be developed or inhabited, such as areas within a fixed distance of roadways, thereby assigning more realistic alternate residential and work locations for participants.

Our results for MC simulation II are likely to be sensitive to the proportion of drinking water obtained from residential versus work addresses. Although the assumption that 30% of drinking water came from work addresses provided valid predictions of PFOA serum concentrations (Shin et al., 2011b), the actual proportion likely differs widely among participants. Future studies in this population or in other populations with contaminated drinking water might benefit from more attention to water sources at participants' workplaces, and to the extent to which each participant consumes tap water while at work.

4.2. Conclusions

In the MC simulation study presented here, we studied the potential impact of geocoding uncertainty due to the missing street level residential addresses and self-reported ZIP codes

of work addresses (for the PFOA exposure assessment participants in the Savitz et al. (2012a) study) by assigning alternate geographic locations within the reported ZIP code and recalculating the serum PFOA concentrations. We repeated the epidemiological study associating these estimated serum PFOA concentrations with the occurrence of preeclampsia (Savitz et al., 2012a) to examine if the use of alternate residential/work locations has any impact on the study results. We found that geocoding based uncertainty in residential addresses did not have any significant impact on the serum PFOA concentration predictions and the epidemiological association with preeclampsia seems to be robust, with little bias. The addition of geocoding based uncertainty in work history moderately impacts the rank exposure among the participants and causes a 41% increase in the average AOR of preeclampsia occurrence. The analysis presented here is one approach to estimating the potential impacts of positional errors in a geocoding-based exposure assessment on exposure estimates and epidemiological study results. Future exposure studies and epidemiological studies that rely on participant locations could benefit from explicit analysis of the impacts of geocoding-based uncertainties.

Acknowledgments

Funding was provided by the National Institute of Environmental Health Sciences (Award R21ES02312). The content is the sole responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. Data collection and previous analyses were funded by the C8 Class Action Settlement Agreement (Circuit Court of Wood County, WV).

We thank Dr. David Savitz and his research group at Brown University for kindly allowing us to use/modify the preeclampsia data set in the C8 Health Project study population for our analysis.

Abbreviations

| | |
|------------|--------------------------------|
| GIS | Geographic Information Systems |
| C8 | PFOA Perfluorooctanoate |
| PWD | Public Water District |
| MC | Monte Carlo |
| AOR | Adjusted Odds Ratio |
| IQR | Inter Quartile Range |
| CI | Confidence Interval |
| PI | Probability Interval |

References

- Ali M, Emch M, Donnay JP. Spatial filtering using a raster geographic information system: Methods for scaling health and environmental data. *Heal Place*. 2002; 8:85–92. DOI: 10.1016/S1353-8292(01)00029-6
- Armstrong BG. Effect of measurement error on epidemiological studies of environmental and occupational exposures. *Occup Environ Med*. 1998; 55:651–6. [PubMed: 9930084]

- Avanasi R, Shin HM, Vieira VM, Savitz DA, Bartell SM. Impact of Exposure Uncertainty on the Association between Perfluorooctanoate and Preeclampsia in the C8 Health Project Population. *Environ Health Perspect.* 2016a; 124:126–132. DOI: 10.1289/ehp.1409044 [PubMed: 26090912]
- Avanasi R, Shin HM, Vieira VM, Bartell SM. Variability and epistemic uncertainty in water ingestion rates and pharmacokinetic parameters, and impact on the association between perfluorooctanoate and preeclampsia in the C8 Health Project population. *Environ Res.* 2016b; 146:299–307. DOI: 10.1016/j.envres.2016.01.011 [PubMed: 26796985]
- Barry V, Winquist A, Steenland K. Perfluorooctanoic acid (PFOA) exposures and incident cancers among adults living near a chemical plant. *Environ Health Perspect.* 2013; 121:1313–1318. DOI: 10.1289/ehp.1306615 [PubMed: 24007715]
- Bell EM, Hertz-picciotto I, Beaumont JJ, Epidemiology S, Mar N. A Case-Control Study of Pesticides and Fetal Death Due to Congenital Anomalies Linked references are available on JSTOR for this article. 2015; 12:148–156.
- Bellander T, Berglind N, Gustavsson P, Jonson T, Nyberg F, Pershagen G, Järup L. Using geographic information systems to assess individual historical exposure to air pollution from traffic and house heating in stockholm. *Environ Health Perspect.* 2001; 109:633–639. DOI: 10.1289/ehp.01109633
- Beya J, Hatch M. Geographic exposure modeling: A valuable extension of geographic information systems for use in environmental epidemiology. *Environ Health Perspect.* 1999; 107:181–190. DOI: 10.2307/3434482 [PubMed: 10229717]
- Bonner MR, Han D, Nie J, Rogerson P, Vena JE, Freudenheim JL. Positional accuracy of geocoded addresses in epidemiologic research. *Epidemiology.* 2003; 14:408–12. DOI: 10.1097/01.EDE.0000073121.63254.c5 [PubMed: 12843763]
- Burns CJ, Wright JM, Pierson JB, Bateson TF, Burstyn I, Goldstein DA. Evaluating Uncertainty to Strengthen Epidemiologic Data for Use in Human Health Risk Assessments. 2014; 1160:1160–1165.
- C8 Science Panel. Probable link evaluation of pregnancy induced hypertension and preeclampsia. 2011:1–6.
- Elgethun K, Fenske Ra, Yost MG, Palcisko GJ. Time-location analysis for exposure assessment studies of children using a novel global positioning system instrument. *Environ Health Perspect.* 2003; 111:115–122. DOI: 10.1289/ehp.5350 [PubMed: 12515689]
- Floret N, Mauny F, Challier B, Arveux P, Cahn JY, Viel JF. Dioxin emissions from a solid waste incinerator and risk of non-Hodgkin lymphoma. *Epidemiology.* 2003; 14:392–8. DOI: 10.1097/01.ede.0000072107.90304.01 [PubMed: 12843761]
- Frisbee SJ, Brooks aP, Maher A, Flensburg P, Arnold S, Fletcher T, Steenland K, Shankar A, Knox SS, Pollard C, Halverson Ja, Vieira VM, Jin C, Leyden KM, Ducatman AM. The C8 Health Project: Design, Methods, and Participants. *Environ Health Perspect.* 2009; 117:1873–1882. DOI: 10.1289/ehp.0800379 [PubMed: 20049206]
- Gallo V, Leonardi G, Genser B, Lopez-Espinosa MJ, Frisbee SJ, Karlsson L, Ducatman AM, Fletcher T. Serum perfluorooctanoate (PFOA) and perfluorooctane sulfonate (PFOS) concentrations and liver function biomarkers in a population with elevated PFOA exposure. *Environ Health Perspect.* 2012; 120:655–660. DOI: 10.1289/ehp.1104436 [PubMed: 22289616]
- Gryparis A, Paciorek CJ, Zeka A, Schwartz J, Coull Ba. Measurement error caused by spatial misalignment in environmental epidemiology. *Biostatistics.* 2009; 10:258–274. DOI: 10.1093/biostatistics/kxn033 [PubMed: 18927119]
- Henry KA, Boscoe FP. Estimating the accuracy of geographical imputation. *Int J Health Geogr.* 2008; 7:3.doi: 10.1186/1476-072X-7-3 [PubMed: 18215308]
- Johnson P, Sutton P. The Navigation Guide—Evidence-Based Medicine Meets Environmental Health: Systematic Review of Human Evidence for PFOA Effects on Fetal Growth. *Env Heal.* 2014; 122:1040–1051. DOI: 10.1289/ehp.1307893
- Koustaas E, Lam J, Sutton P, Johnson PI, Atchley DS, Sen S, Robinson Ka, Axelrad Da, Woodruff TJ. The Navigation Guide - evidence-based medicine meets environmental health: systematic review of nonhuman evidence for PFOA effects on fetal growth. *Environ Health Perspect.* 2014; 122:1015–27. DOI: 10.1289/ehp.1307177 [PubMed: 24968374]

- Lopez-Espinosa M, Mondal D, Armstrong B, Bloom MS, Fletcher T. Thyroid Function and Perfluoroalkyl Acids in Children Living Near a Chemical Plant. *Environ Health Perspect.* 2012; 120:1036–1041. DOI: 10.1289/ehp.1104370 [PubMed: 22453676]
- Nuckols JR, Ward MH, Jarup L. Using geographic information systems for exposure assessment in environmental epidemiology studies. *Environ Health Perspect.* 2004; 112:1007–1015. DOI: 10.1289/ehp.6738 [PubMed: 15198921]
- Reynolds P, Von Behren J, Gunier RB, Goldberg DE, Hertz A, Smith DF. Childhood cancer incidence rates and hazardous air pollutants in California: An exploratory analysis. *Environ Health Perspect.* 2003; 111:663–668. DOI: 10.1289/ehp.5986 [PubMed: 12676632]
- Rull RP, Ritz B. Historical pesticide exposure in California using pesticide use reports and land-use surveys: An assessment of misclassification error and bias. *Environ Health Perspect.* 2003; 111:1582–1589. DOI: 10.1289/ehp.6118 [PubMed: 14527836]
- Savitz DA, Stein CR, Bartell SM, Elston B, Gong J, Shin HM, Wellenius GA. Perfluorooctanoic acid exposure and pregnancy outcome in a highly exposed community. *Epidemiology.* 2012a; 23:386–392. DOI: 10.1097/EDE.0b013e31824cb93b [PubMed: 22370857]
- Savitz DA, Stein CR, Elston B, Wellenius GA, Bartell SM, Shin HM, Vieira VM, Fletcher T. Children's Health Relationship of Perfluorooctanoic Acid Exposure to Pregnancy Outcome Based on Birth Records in the Mid-Ohio Valley. *Environ Health Perspect.* 2012b; 120:1201–1207. [PubMed: 22450153]
- Shin HM, Vieira VM, Ryan PB, Detwiler R, Sanders B, Steenland K, Bartell SM. Environmental fate and transport modeling for perfluorooctanoic acid emitted from the Washington Works Facility in West Virginia. *Environ Sci Technol.* 2011a; 45:1435–1442. DOI: 10.1021/es102769t [PubMed: 21226527]
- Shin HM, Vieira VM, Ryan PB, Steenland K, Bartell SM. Retrospective exposure estimation and predicted versus observed serum perfluorooctanoic acid concentrations for participants in the C8 Health Project. *Environ Health Perspect.* 2011b; 119:1760–1765. DOI: 10.1289/ehp.1103729 [PubMed: 21813367]
- Steenland K, Zhao L, Winquist A, Parks C. Ulcerative colitis and perfluorooctanoic acid (PFOA) in a highly exposed population of community residents and workers in the Mid-Ohio Valley. *Environ Health Perspect.* 2013; 121:900–905. DOI: 10.1289/ehp.1206449 [PubMed: 23735465]
- US Census Bureau. [accesses 1 July 2016] 2010. Available: <http://www.census.gov/geo/reference/urban-rural.html>
- US Department of Health and Human Services. [accessed 1 January 2016] 2014. Available: <http://grants.nih.gov/grants/guide/pa-files/PA-15-010.html>
- Vieira V, Hoffman K, Fletcher T. Assessing the Spatial Distribution of Perfluorooctanoic Acid Exposure via Public Drinking Water Pipes Using Geographic Information Systems. *Environ Health Toxicol.* 2013; 28:e2013009.doi: 10.5620/eh.2013.28.e2013009 [PubMed: 24010064]
- Vieira VM, Howard GJ, Gallagher LG, Fletcher T. Geocoding rural addresses in a community contaminated by PFOA: a comparison of methods. *Environ Health.* 2010; 9:18.doi: 10.1186/1476-069X-9-18 [PubMed: 20406495]
- Ward MH, Nuckols JR, Giglierano J, Bonner MR, Wolter C, Airola M, Mix W, Colt JS, Hartge P. Positional accuracy of two methods of geocoding. *Epidemiology.* 2005; 16:542–547. DOI: 10.1097/01.ede.0000165364.54925.f3 [PubMed: 15951673]
- Watkins DJ, Jossion J, Elston B, Bartell SM, Shin HM, Vieira VM, Savitz DA, Fletcher T, Wellenius GA. Exposure to Perfluoroalkyl Acids and Markers of Kidney Function among Children and Adolescents Living near a Chemical Plant. *Environ Health Perspect.* 2013; 121:625–630. [PubMed: 23482063]
- Zandbergen P. Geocoding Quality and Implications for Spatial Analysis. *Geogr Compass.* 2009; 3:647–680.

Highlights

- GIS-based PFOA exposure is linked to preeclampsia by a C8 Science Panel study
- MC simulations can be used to study the impact of geocoding uncertainty
- Use of ZIP code centroid assignment for home addresses had no significant impact
- There was a 41% increase in the mean AOR for the workplace geocoding uncertainty
- More accurate information on water sources at workplaces can be useful in this cohort

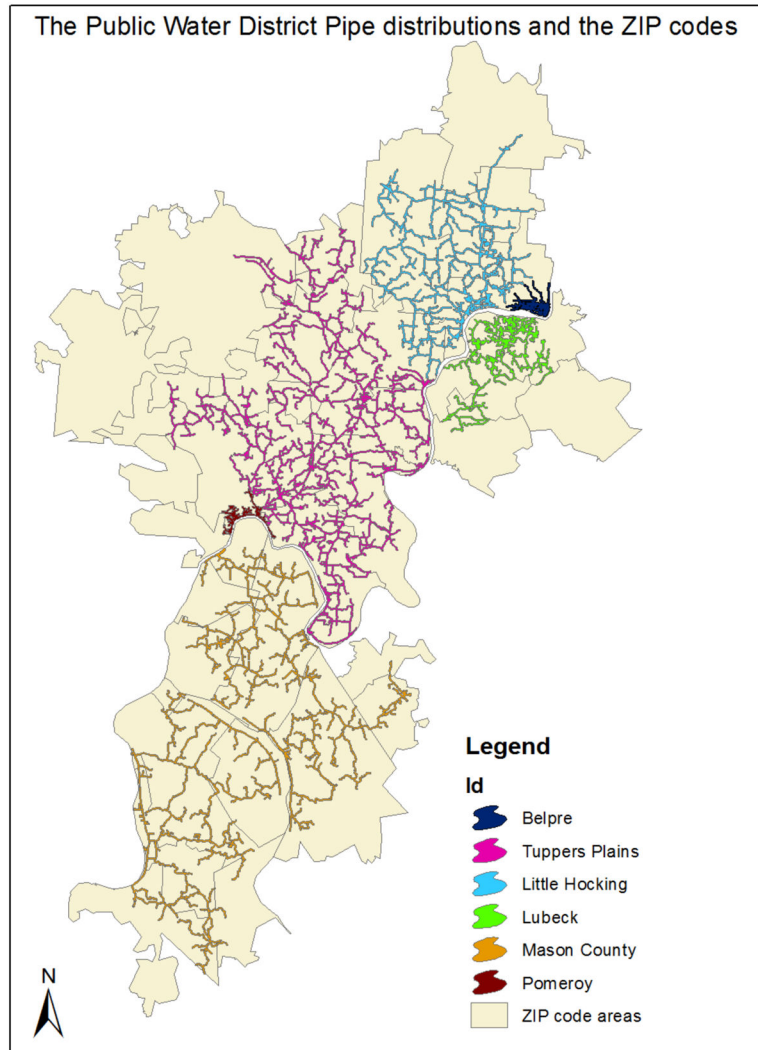


Figure 1. The image of the ZIP codes that is supplied by the pipe networks belonging to the 6 participating PWDs in the PFOA exposure assessment

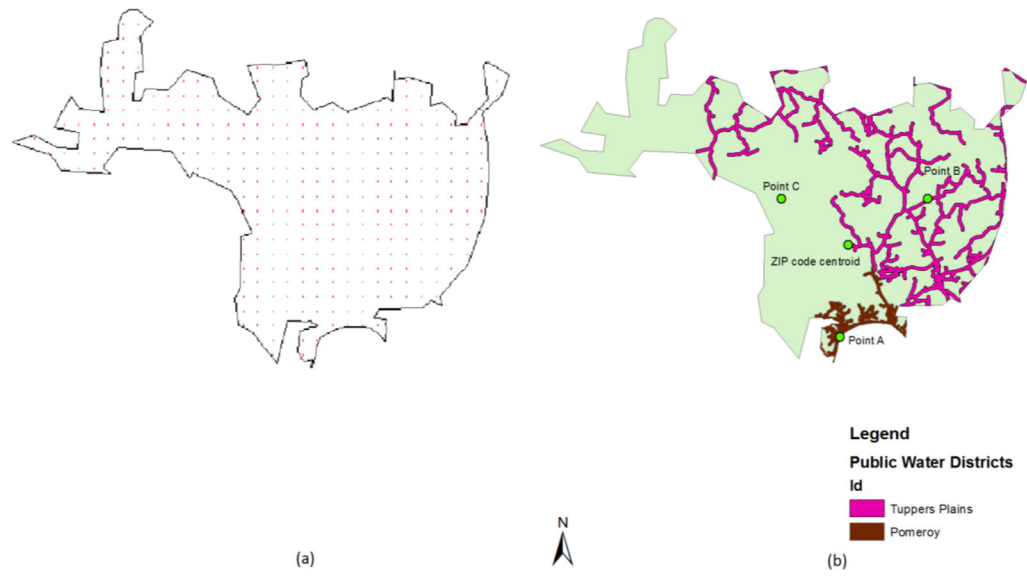


Figure 2. Outline of ZIP code 45769 with panel (a) showing the grid created for MC simulation and panel (b) illustrating the MC simulation methodology

Table 1

The mean and the 95% probability interval (PI) of the median, mean, 25th and 75th percentile serum concentrations at birth (ng/mL), for the study participants, for each of the 2 Monte Carlo simulations (200 iterations per simulation)

| Simulation | Median (95% PI) | Mean (95% PI) | 25th percentile (95% PI) | 75th percentile (95% PI) |
|---|------------------------|----------------------|--|--|
| Modified original (n=10,149) | 9.4 | 51.1 | 5.1 | 32.5 |
| Modified original Residential centroid subset (n= 3,266) | 8.3 | 50.3 | 5.0 | 27.1 |
| Street-level subset (n=6883) | 10.2 | 52.8 | 5.1 | 35.5 |
| MC simulation I Residential centroid subset (n= 3,266) | 7.4 (7.3, 7.5) | 49.9 (48.7, 51.0) | 5.1 (5.1, 5.2) | 24.1 (23.1, 25.2) |
| MC simulation I (n= 10,149) | 9.1 (9.0, 9.1) | 51.9 (51.5, 52.2) | 5.1 (5.1, 5.1) | 32.3 (31.8, 32.7) |
| MC simulation II Residential centroid subset (n= 3,266) | 8.2 (8.0, 8.4) | 53.1 (51.9, 54.3) | 5.2 (5.2, 5.2) | 31.2 (29.9, 32.6) |
| MC simulation II (n= 10,149) | 10.9 (10.8, 11.0) | 55.5 (55.1,55.9) | 5.2 (5.2, 5.2) | 40.1 (39.3, 40.8) |