

• **Brief Report**

# Assessing the Spatial Distribution of Perfluorooctanoic Acid Exposure via Public Drinking Water Pipes Using Geographic Information Systems

Verónica Vieira<sup>1,2</sup>, Kate Hoffman<sup>1,3</sup>, Tony Fletcher<sup>4</sup>

<sup>1</sup>Department of Environmental Health, Boston University School of Public Health, Boston, MA; <sup>2</sup>Program in Public Health, Chao Family Comprehensive Cancer Center, University of California, Irvine, CA; <sup>3</sup>University of North Carolina, Gillings School of Global Public Health, Chapel Hill, NC, USA; <sup>4</sup>Department of Social and Environmental Health Research, London School of Hygiene and Tropical Medicine, London, United Kingdom

**Objectives** Geographic Information Systems (GIS) is a powerful tool for assessing exposure in epidemiologic studies. We used GIS to determine the geographic extent of contamination by perfluorooctanoic acid, C8 (PFOA) that was released into the environment from the DuPont Washington Works Facility located in Parkersburg, West Virginia.

**Methods** Paper maps of pipe distribution networks were provided by six local public water districts participating in the community cross-sectional survey, the C8 Health Project. Residential histories were also collected in the survey and geocoded. We integrated the pipe networks and geocoded addresses to determine which addresses were serviced by one of the participating water districts. The GIS-based water district assignment was then compared to the participants' self-reported source of public drinking water.

**Results** There were a total of 151,871 addresses provided by the 48,800 participants of the C8 Health Project that consented to geocoding. We were able to successfully geocode 139,067 (91.6%) addresses, and of these, 118,209 (85.0%) self-reported water sources were confirmed using the GIS-based method of water district assignment. Furthermore, the GIS-based method corrected 20,858 (15.0%) self-reported public drinking water sources. Over half (54%) the participants in the lowest GIS-based exposure group self-reported being in a higher exposed water district.

**Conclusions** Not only were we able to correct erroneous self-reported water sources, we were also able to assign water districts to participants with unknown sources. Without the GIS-based method, the reliance on only self-reported data would have resulted in exposure misclassification.

**Keywords** C8, Dupont, Parkersburg, Perfluorooctanoic acid

**Correspondence:**

Verónica Vieira, DSc  
 AIRB 2042, University of California,  
 Irvine, CA 92697, USA  
 Tel: +1-949-824-7017  
 Fax: +1-949-824-0527  
 Email: [vvieira@uci.edu](mailto:vvieira@uci.edu)

Received: February 18, 2013

Accepted: June 24, 2013

Published online: August 30, 2013

This article is available from: <http://e-eh.ohg.org/>

## Introduction

The current study uses geographic information systems (GIS) to determine the distribution of perfluorooctanoic acid, C8 (PFOA) exposure via contaminated public drinking water. This work contributed to the exposure assessment used in several

studies investigating health effects of PFOA exposure among residents living near the Washington Works DuPont Teflon-manufacturing plant in Parkersburg, West Virginia (WV) [1]. PFOA is a surfactant widely used in the manufacture of stain-resistant and water-repellant consumer products, including the non-stick cookware Teflon. The DuPont facility is located near

the Ohio River (Figure 1) and released PFOA into the environment via aerial emissions and surface and groundwater discharge beginning in the 1950s, resulting in the contamination of the local drinking water in Ohio (OH) and WV [2-4]. A 2002 sampling survey revealed that the contamination of groundwater and surface water was geographically extensive, reaching as far south as the Mason County water supply (WV, Figure 1) [5]. The highest PFOA level measured in the public drinking water was 37.1  $\mu\text{g/L}$  in the Little Hocking Water Association (OH) test well located across the river from the Washington Works facility.

A class action lawsuit brought by the surrounding communities against DuPont resulted in a settlement agreement whereby Brookmar, Inc., an independent company, conducted a year-long survey (August 2005-July 2006) of over 69,000 residents called the C8 Health Project [6]. To qualify for the C8 Health Project, the drinking water of the survey participants must have been supplied from private wells in the contaminated area or at least one of the following public water supplies: the Public Service Districts of Lubeck and Mason County in WV; the Little Hocking Water Association, Tupper Plains-Chester Water District, the Village of Pomeroy Public Service District, or Belpre Public Service District in OH (Figure 1). The objectives of this study are to determine the geographic extent of public drinking

water contamination using GIS and compare the GIS-based results to the participants' self-reported water sources.

## Materials and Methods

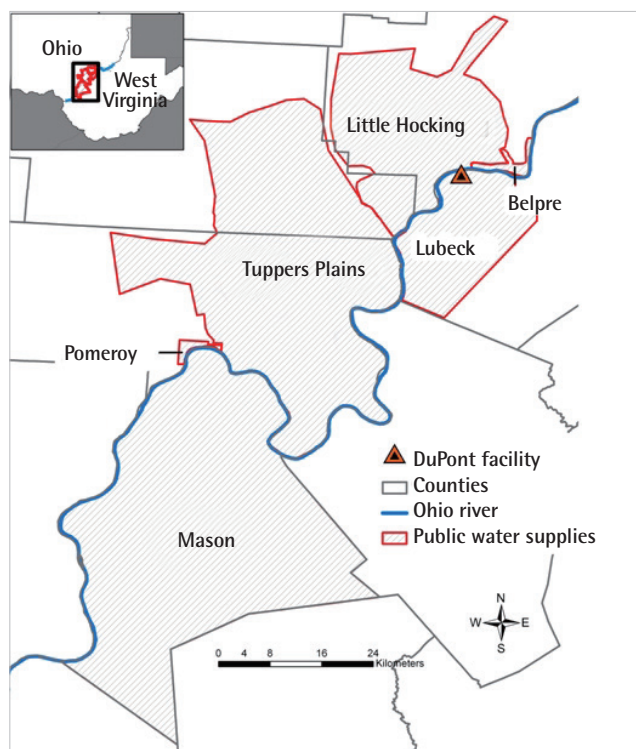
### Study Population and Data

The study area encompasses the 6 contaminated public water districts (WD) in WV and OH that surround the DuPont Washington Works facility (Figure 1) and were part of the class-action lawsuit. The C8 Health Project collected information on participants' demographics, residential history, and medical history via a self-administered questionnaire. Of the 69,030 community residents that participated in the cross-sectional survey, 48,880 provided their consent for the use of identifiable data. These data include street addresses, years of residency, and the corresponding drinking water supply [6]. Participants were asked to provide all their addresses within the study area as far back as 1951 (when the DuPont facility began operation) or their date of birth. For each address, participants reported if their drinking water was from either one of the six participating water districts, from another known source (either non-participating public water district or private water supply), or unknown. Information on bottled water use was also collected in the survey and incorporated into the final exposure assessment [7], but is not discussed here.

### GIS Methods

We first contacted the water district managers of the six participating water districts to obtain maps of their pipe distribution networks. In addition to the geographic extent of the public water supply, they also provided the years in which the pipes were installed. Paper maps were scanned and electronically added to a base map of the study area using ESRI ArcView version 9.3 (Redlands, CA, USA). The images were rectified to the base map using the Ohio River, streets, and county boundaries to determine the proper layout. Then a GIS shapefile of pipes was created using street line shapefiles as a starting point that included an attribute for pipe installation year.

Once the pipes for the water districts were digitized, the next step was to geocode the residential street addresses. There were a total of 151,871 addresses for the 48,800 participants of the C8 Health Project that consented to providing full address details to allow geocoding. The addresses were comprised primarily of street addresses (79.4%), but also included post office boxes (3.0%) and rural route boxes (17.6%). The study area included both densely populated cities and more rural areas where ru-



**Figure 1.** Study area encompassing 6 contaminated water districts surrounding the DuPont Washington Works Facility in Parkersburg, West Virginia.

ral routes were often used to refer to an address. Rural routes are problematic for geocoding because they are often not detailed on street reference shapefiles, and, without a house number, it is difficult to determine where along the route participants resided. At the same time the study was being conducted, communities were also participating in an Enhanced 911 program to assign street addresses to rural routes for improved emergency medical response. Address conversion tables for rural areas were compiled and used as an additional tool for geocoding.

We first cleaned and standardized the self-reported addresses using ZIP4 address correction software with the database and converted additional rural route boxes to street addresses using Enhanced 911 address conversion tables [8]. We created a file with the cleaned addresses, the years of residency, and the self-reported water district. The file was added to the GIS and geocoding was performed using the ESRI StreetMap Premium North America NAVTEQ 2010 enhanced street dataset as the reference address locator with a side offset of 20 meters [8]. Lastly, the geocoded shapefile was spatially joined with the pipe shapefile so that the closest pipe segment and its corresponding information were appended to the geocoded addresses. During the joining process, GIS also calculated the distance between the geocoded point and the closest pipe segment.

## Water District Assignment

After reviewing the self-reported water district data, it was apparent that a large percentage of water districts were missing or implausible. For exposure assessment purposes, it was important that participants were accurately assigned to one of the six participating water districts or else coded as being serviced by a non-participating water source. This involved an iterative process of comparing GIS-assigned water district to the self-reported water district. Once data from the nearest pipe segment was spatially joined to geocoded addresses, we examined the distance between the two to determine water district assignments. Because GIS pipe shapefiles were created along street centerlines, addresses that were within 20 meters of a pipe segment were assigned the water district of that pipe. Addresses that were outside the six participating water district service areas were assigned to the non-participating water supply category. Addresses that had a self-reported water district that differed from the one that was GIS-assigned were flagged for further review. Some of these participants may have been on a private well despite having direct access to public water. Addresses within the six water district boundaries but more than 20 meters from a pipe segment were also flagged for further review. A manual review of 17,441 addresses (11.5%) was conducted by re-contacting the

water district managers and asking them about specific streets and/or customers. Residency year and pipe installation years were also reviewed because it was also possible that a participant moved away prior to the start of water district service to that address.

To assess the extent of exposure misclassification that may result from reliance on only self-reported water district, we compared participants' exposure classification based on self-reported qualifying water districts to their GIS-based exposure measures. The methods for estimating exposure using GIS are described in detail elsewhere [7]. Briefly, participants' GIS-assigned water districts, pipe installation years, and residency durations were used as inputs for a linked fate and transport and pharmacokinetic model. The modeled C8 levels at the time of the survey were grouped by tertiles into high, medium, and low exposure groups. We compared these groupings to high, medium, and low exposure groups based on their self-reported water district and the C8 water concentrations at the time of the survey.

## Results

The number of residences per participant ranged from one to twelve with a median of 2. Table 1 shows the results of the GIS-based method of WD assignment compared to the self-report water supply by time period. Based on the residency end years, we grouped addresses by decade: <1990 (n=24,936; 16.4%), 1990-1999 (n=43,335; 28.5%), 2000-2006 (n=83,600; 55.1%). As residency within a contaminated water district was a criterion for participating in the survey, the majority of the ad-

**Table 1.** Self-reported water sources categorized by GIS-based water district (WD) assignment and time period.

GIS-assigned (geocoded)	Self-reported drinking water source			Total
	Participating WD	Other known WD	Unknown source	
Participating WD	52,997	1,615	2,504	57,116
<1990	6,277	285	585	7,147
1990-1999	12,449	518	684	13,651
>2000	34,271	812	1,235	36,318
Other known WD	6,546	65,212	10,193	81,951
<1990	853	12,452	2,228	15,533
1990-1999	1,611	21,194	3,542	26,347
>2000	4,082	31,566	4,423	40,071
Unknown source (not geocoded)	8,445	2,909	1,450	12,804
<1990	1,131	733	392	2,256
1990-1999	2,048	863	426	3,337
>2000	5,266	1,313	632	7,211
Total	67,988	69,736	14,147	151,871

GIS, geographic information systems.

dresses were current or recent residences. We were able to successfully geocode 139,067 of the 151,871 (91.6%) addresses, which we were able to subsequently determine whether they were serviced by one of the six participating WDs. Geocoding success rates were similar across time periods. Although the oldest addresses (< 1990) had the lowest rate (91.0%; 22,680 of 24,936), rates for addresses from the 1990s (92.3%; 39,998 of 43,335) and the 2000s (91.4%; 76,389 of 83,600) were only slightly better. There were 26,743 rural route addresses, and we were able to successfully assign 15,251 (57.0%) to a water district using geocoding methods.

Of the 151,871 addresses, we confirmed 118,209 (85.0%) self-reported water sources using the GIS-based method. The remaining 20,858 (15.0%) geocoded addresses had discordant WD assignments and were classified into four different categories. The majority (n = 12,697) were self-reported as unknown, but using GIS, we were able to determine that they were serviced by one of the six participating WDs. Another 7,342 self-reported service by one of the six participating WDs, but were determined to be supplied by a different drinking water source. For 819 addresses, the drinking water was self-reported as a non-participating WD when they were actually serviced by a participating WD.

Of the 151,871 addresses, participants reported 68,784 (45.3%) were serviced by one of the six participating WDs, 68,940 (45.4%) were serviced by another known water supply, and 14,147 (9.3%) had an unknown drinking water source. When we examine the data categorized by self-reported WD, we found that 52,997 (76.9%) of the addresses with a self-reported participating WD matched the GIS-assigned WD. We determined that 7,342 (10.6%) addresses were incorrectly reported being serviced by one of the six participating WDs. We also determined that 2,504 addresses with a self-reported unknown public water supply were serviced by one of the six participating WDs.

The GIS-based method was unable to provide any information on the drinking water source for the 12,804 addresses we could not geocode. In order to determine an exposure measure, we relied on the self-reported water supply (n = 12,601); for the majority (n = 12,124), we were able to verify with GIS that their ZIP codes geographically intersected the water district, and thus the reported water supply was plausibly correct. For the remainder, we concluded that the reported water district was probably incorrect as it was incompatible with the ZIP code. For the 1,450 addresses with self-reported unknown WD, 1,247 had sufficient information to assign a ZIP code-level exposure measure based on the proportion of water district pipe length in the ZIP code. Ultimately, only 203 (0.1%) addresses could not be

assigned to a WD category or a weighted average exposure measure based on ZIP code.

When we compared the participants' exposure classifications using the GIS-assigned WD to their classifications using self-reported WD, 54% of participants in the lowest GIS-based exposure group had been misclassified into higher exposure groups based on their self-reported WD. More than 40% of participants (20,850 of the 48,880) self-reported a qualifying WD in the highest exposure group when only 24% (11,811) of those were in the highest GIS-based exposure group. Conversely, only 104 participants who self-reported a qualifying WD in the lowest exposure group were assigned to the highest GIS-based exposure group.

## Discussion

By geocoding residential addresses and mapping them with water district pipe distribution networks, we determined for over 118,000 addresses whether or not the source of public drinking was one of the contaminated water districts. An important advantage to using this GIS-based method of WD assignment rather than relying only on self-reported drinking water source is that we were able to identify and correct over 20,000 WD assignments, reducing the potential for exposure misclassification. If we relied only on self-reported water districts and did not geocode the addresses, then 22,599 of the 48,880 participants would have potentially been misclassified into different exposure groups. For five of the six participating water districts, 90-95% of the addresses were correctly self-reported. However, only 76% of the self-reported Mason County addresses were correctly identified. There was some apparent confusion between the Mason County Public WD (a participating WD) and the Town of Mason WD (non-participating). We also found that water sources for older addresses were more likely to be incorrectly self-reported than more recent addresses. When we reviewed the data, it appeared that a reason for some of the reporting errors were that many of these participants correctly assigned their most recent participating water district and then simply copied that code for all of their addresses when completing the questionnaire. Although most of the addresses with missing water sources were from uncontaminated water districts, 18% were assigned to one of the participating water districts.

Despite the usefulness of geocoding and GIS-based methods, there are several limitations. Given that the study area is in WV and OH, there were many addresses with rural route boxes. We were able to recover street addresses for some of these, but in areas that were more rural, particularly Mason County in WV, our



geocoding success rates were lower. Fortunately, the Mason County area was less exposed, so it was not a serious limitation in our study, but that may not be true in other studies. Also, we had very little temporal variation in our geocoding success rates but this may not have been the case if we did not have information on changes in street names over time from the Enhanced 911 programs. For an epidemiologic study where this data is not available, geocoding of older addresses may be less successful.

There are also important exposure parameters that can only be obtained from self-reported data. Information on residency years is critical for determining when participants were first exposed and for what duration. While residency years may be subject to recall bias, it is likely non-differential with respect to exposure status. Without this self-reported data, the exposure would have been difficult to model. In a few cases, the manual review of discrepancies between participants' self-reported WDs and GIS-assigned WDs alerted us to some omissions in the water district maps we were provided. Therefore, it is important that all available information, both self-reported and modeled, is used to determine the most accurate exposure assessment possible in environmental epidemiologic studies.

In conclusion, this paper highlights the use of GIS to help assess PFOA exposure via public drinking water contaminated by an industrial facility. Exposure assessment is a critical component to epidemiologic studies conducted in this community [9]. This GIS-based method is readily adaptable and may prove useful in the exposure assessments for health studies of other contaminated sites.

## Acknowledgments

We acknowledge the staff at the Public Service Districts of Lu-beck and Mason County in West Virginia; the Little Hocking Water Association, Tappers Plains-Chester Water District, the Village of Pomeroy Public Service District, and Belpre Public Service District in Ohio; and Boston University staff Gregory Howard and Alicia Fraser for their assistance with the pipe maps. The project was supported by the C8 Class Action Settlement Agreement (Circuit Court of Wood County, WV, USA)

between DuPont and plaintiffs, which resulted from releases into drinking water of the chemical perfluorooctanoic acid (PFOA, or C8). Funds were administered by the Garden City Group (Melville, NY) that reports to the court.

## Conflict of Interest

The authors have no conflicts of interest with the material presented in this paper.

## References

1. Steenland K, Jin C, MacNeil J, Lally C, Ducatman A, Vieira V, et al. Predictors of PFOA levels in a community surrounding a chemical plant. *Environ Health Perspect* 2009;117(7):1083-1088.
2. Paustenbach DJ, Panko JM, Scott PK, Unice KM. A methodology for estimating human exposure to perfluorooctanoic acid (PFOA): a retrospective exposure assessment of a community (1951-2003). *J Toxicol Environ Health A* 2007;70(1):28-57.
3. Shin HM, Vieira VM, Ryan PB, Detwiler R, Sanders B, Steenland K, et al. Environmental fate and transport modeling for perfluorooctanoic acid emitted from the Washington Works Facility in West Virginia. *Environ Sci Technol* 2011;45(4):1435-1442.
4. Ritchey RL. Letter to Cliff D. Whyte, Assistant Director, West Virginia Department of Environmental Protection, Division of Water and Waste Management, from Robert L. Ritchey, Senior Environmental Control Consultant, DuPont Washington Works; 2006 [cited 2013 Jan 22]. Available from: <http://www.regulations.gov/#!documentDetail;D=EPA-HQ-OPPT-2003-0012-1098>.
5. Hartten AS. C-8 data summary report-consent order GWR-2001-019; 2003 [cited 2013 Jan 22]. Available from: <http://www.regulations.gov/#!documentDetail;D=EPA-HQ-OPPT-2003-0012-0039>.
6. Frisbee SJ, Brooks AP Jr, Maher A, Flensburg P, Arnold S, Fletcher T, et al. The C8 health project: design, methods, and participants. *Environ Health Perspect* 2009;117(12):1873-1882.
7. 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* 2011;119(12):1760-1765.
8. 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.
9. C8 Science Panel. Science panel news and updates; 2012 [cited 2013 Jan 22]. Available from: <http://www.c8sciencepanel.org>.