Investigating Three Classification Methods for Per/Poly-Fluoroalkyl Substance (PFAS) Exposure from Electronic Health Records And Potential for Bias

Lena M. Davidson[1], MS
Mary Regina Boland[1-4], MA, MPhil, PhD, FAMIA[1-4]
[1]Department of Biostatistics, Epidemiology and Informatics, Perelman School of Medicine, University of Pennsylvania
[2]Institute for Biomedical Informatics, University of Pennsylvania
[3]Center for Excellence in Environmental Toxicology, University of Pennsylvania
[4]Department of Biomedical and Health Informatics, Children's Hospital of Philadelphia

Abstract

Per-/poly-fluoroalkyl substances (PFAS) are a group of manmade compounds with known human toxicity and evidence of contamination in drinking water throughout the US. We augmented our electronic health record data with geospatial information to classify PFAS exposure for our patients living in New Jersey. We explored the utility of three different methods for classifying PFAS exposure that are popularly used in the literature, resulting in different boundary types: public water supplier service area boundary, municipality, and ZIP code. We also explored the intersection of the three boundaries. To study the potential for bias, we investigated known PFAS exposure-disease associations, specifically hypertension, thyroid disease and parathyroid disease. We found that both the significance of the associations and the effect size varied by the method for classifying PFAS exposure. This has important implications in knowledge discovery and also environmental justice as across cohorts, we found a larger proportion of Black/African-American patients PFAS-exposed.

1. Introduction

Per- and polyfluoroalkyl substances (PFAS) are a group of nearly 5000 of manmade organic compounds that possess high stability and mobility in the environment. Development of these synthetic compounds began in the 1970's for use in various industrial and commercial products (e.g. aqueous firefighting foam, nonstick cooking pans, grease-resistant food packaging, stain-resistant textiles). Nicknamed "forever chemicals" that are highly resistant to breakdown, PFAS molecules are composed of a chain of strong carbon-fluorine bonds. The two most recognizable PFAS compounds, perfluorooctanoic acid (PFOA) and perfluorooctanesulfonic acid (PFOS) are no longer intentionally produced in the United States (US), however they are the most widespread and bio-accumulative. Despite the widespread proliferation of PFAS in the US, many differences exist across states in terms of what exposure levels constitute risk for human health and how to measure exposure. The primary purpose of this study is to compare cohort definition differences for PFAS exposure to determine how this affects our cohorts both in terms of sample size counts and also disease risk. We also aim to explore disparities in PFAS exposure as an Environmental Justice angle of our work.

1.1 State-Level and Federal-Level PFAS Actions

As of September 2022, there are no federal regulations of PFAS, however the EPA currently has established the non-enforceable lifetime health advisory level (LHAL) of 0.07 μg/L of combined PFOS and PFOA in drinking water [1]. On October 2021, the EPA announced the PFAS strategic roadmap, outlining EPA's commitment to actions from 2021 to 2024 [2] with further revisions to the LHAL expected in future.

Statewide in New Jersey (NJ), the source of the PFAS exposure varies from military bases to manufacturers [3]. From the 1950's until about 2000, when these two types of PFAS were phased out of production, PFOA and PFOS were manufactured, sold and distributed in NJ. In 2018, New Jersey adopted a maximum contaminant level (MCL) for different PFAS species, including a MCL of 13 parts per trillion (ppt) for PFNA, followed by adoption of MCLs for PFOA (14 ppt) and PFOS (13 ppt) in 2020. The New Jersey Department of Environmental Protection (NJDEP) developed the standards, criteria, and guidance with the idea that exposures would occur throughout an individual's lifetime [4].

1.2 Challenges in Determining Contaminant Exposure in Drinking Water

A challenge in risk analysis for exposure contaminants in drinking water includes appropriate representation of the exposed cohort. Blood serum testing provides valuable, accurate information for monitoring PFAS exposure in humans. Most PFAS exposure research depends on testing biomarkers (e.g. serum, urine, breastmilk) [5–10]. Due to cost and varied access to resources, PFAS blood tests can be a challenge to implement in research.

When using electronic health record (EHR) data, patient address location can be determined at different granularities. ZIP Codes of the exposed areas can be utilized quickly and is often used in analysis of public health studies. However, ZIP Codes are assigned by the United States Postal Service for the purpose of delivering mail; the boundaries will not necessarily represent areas of the water distribution or even the municipality [11]. Using geographic information systems (GIS), patient addresses may be geocoded locally resulting in latitude and longitude coordinates for spatial analysis. Spatial analysis can be utilized to explore patterns in EHR data to analyze clinical outcomes. For example, spatial analysis has been used to investigate adverse pregnancy outcomes[12], maternal morbidity[13], pediatric surgery cancellation[14]. Researchers can then determine whether or not these patient coordinates lay within a boundary. Depending on the data available, which can vary greatly by state, the analysis may be limited to ZIP Code or municipality boundaries. Public water supplier (PWS) service area boundaries are not necessarily available in every state, limiting researchers to manually determine ZIP Codes or municipality codes to analyze exposure to contaminated drinking water. A recent related work investigated multimorbidity and PFAS exposure using EHR data and geospatial analysis using ZIP Code boundaries [15].

1.3 Objective

In this work, we aim to explore how cohort definition differences (i.e. demographics) in municipality, ZIP Code, and Public Water Supplier (PWS) Service Area boundaries in spatial analysis can alter patients' drinking water PFAS exposure status. Secondly, we refer to well-known associated diseases (i.e. thyroid disease) and analyze these associations in these cohorts, using inpatient and outpatient data. We are expanding on prior work that studied only a heavily exposed PFAS population [16] and investigating patients from the neighboring state of New Jersey (NJ) where these PFAS classification differences would likely have an effect on subsequent studies.

2. Methods

2.1 Patient Cohort Data Source

We used Electronic Health Records (EHR) data obtained from 4 different hospitals within the Penn Medicine system, the Hospital of the University of Pennsylvania (HUP), the Pennsylvania Hospital, Penn Presbyterian Hospital, and Chester County Hospital. In previous work, patient addresses were geocoded to coordinates (latitude and longitude) using ArcGIS locally [17]. The ArcGIS geocoding dataset was limited to 100 match scores. The cohort is limited to patients who had at least one address in the state of NJ, due to the limit of the NJ DEP dataset.

2.2 Identifying PFAS Exposure Boundaries

2.2.1 Spatial Boundaries Data Source

Boundaries of the public water supplier service areas are available from the New Jersey Department of Environmental Protection (NJDEP) Geographic Information System (GIS) digital data. This secondary product has not been verified by NJDEP and is not state-authorized or endorsed. According to the source information, the boundaries for New Jersey PWS service areas are approximate. Municipal boundaries of New Jersey are provided by NJ Office of Information Technology, Office of GIS (NJOGIS) in the NJ State Plane Coordinate System (NAD83) and is accessible in the New Jersey Geographic Information Network (NJGIN) Open Data portal. We used the zipcodeR() package to source all New Jersey state ZIP Codes.

2.2.2 NJ DEP Dataset

PFAS exposure was determined by a dataset of Public Water Systems with PFAS maximum contaminant level (MCL) Violations sourced from the NJ DEP Dataminer (sourced on March 16 2022). The dataset is limited to three types of PFAS: PFOA, PFOS, and PFNA. Maximum Contaminant Level of each PFAS type monitored in NJ are reported in parts per billion (ppb): 0.014 ppb of PFOA, 0.013 ppb of PFOS, and 0.013 ppb of PFNA[4]. Regulations to monitor these contaminants occurred between 2019 and 2020, while the testing requirements for all community water systems began January 1, 2020 for PFNA and January 1, 2021 for PFOS and PFOA [18].

2.3 Linking PFAS Exposure Areas and Patient Geospatial Information

Results of the NJ DEP maximum contaminant level violations were simplified in a binary matrix to indicate whether or not each PWS had at least one sample or violation above the MRL/MCL, respectively. Moreover, each corresponding municipality and ZIP code of a public water service area were coded to a binary matrix for further exposure analysis. We use binary exposure due to the complexity of contaminant exposure in drinking water.

Subsequently, we determined patient location, using patient geospatial information, within the determined boundaries: Municipality, and Public Water Supplier Service Area boundaries. For the sake of comparison to methods without the time and resources for determining geospatial information, we extracted patient address ZIP Codes for analysis. Each patient location was linked to these boundaries, resulting in a matrix of binary exposure the three PFAS types in the NJ DEP violations dataset. Patients who were found to be PFAS-exposed across all three boundary types were next classified as another cohort, further referred as intersection in our analysis. Exposure to PFAS types by address were then concatenated by patient unique medical record number (MRN).

2.3.1 Inpatient and Outpatient PFAS Exposure Cohorts

Two cohorts were determined using a combination of The International Classification of Diseases, 9th Revision, Clinical Modification (ICD-9-CM) and ICD, 10th Revision (ICD-10) billing codes. These cohorts include only patients with inpatient and outpatient diagnoses, respectively. We further divide the cohorts, indicating PFAS exposure by the following: NJ DEP Violations by PWS Service Area, NJ DEP Violations by municipality, and NJ DEP Violations by ZIP code. We collected demographics and characteristics of each cohort.

2.4 Statistical Analysis

We observe well-known PFAS exposure disease associations in literature, namely: thyroid disease [19–22] and hypertension [23–28]. We extracted ICD-9 and ICD-10 codes to capture these diseases. After code extraction, we used the 'icd_map()' function in the 'touch' package in R to crosswalk ICD-9 codes to ICD-10. Next, we included goiter diagnosis codes and categorized such as toxic and nontoxic goiter code groups. Due to differences in medical settings, resulting in differences in the EHR data, we split the cohort into inpatient and outpatient cohorts. We limited to patients who were given at least one diagnosis in the EHR record for each cohort. To improve statistical power, we limited our analysis to diagnosis codes given to at least 100 patients, and given to at least 10 PFAS-exposed patients in each cohort. In our analysis, we calculated Fisher's exact test and report *P*-value, odds ratio (OR), Bonferroni-corrected *P*-value for each diagnosis code and code group.

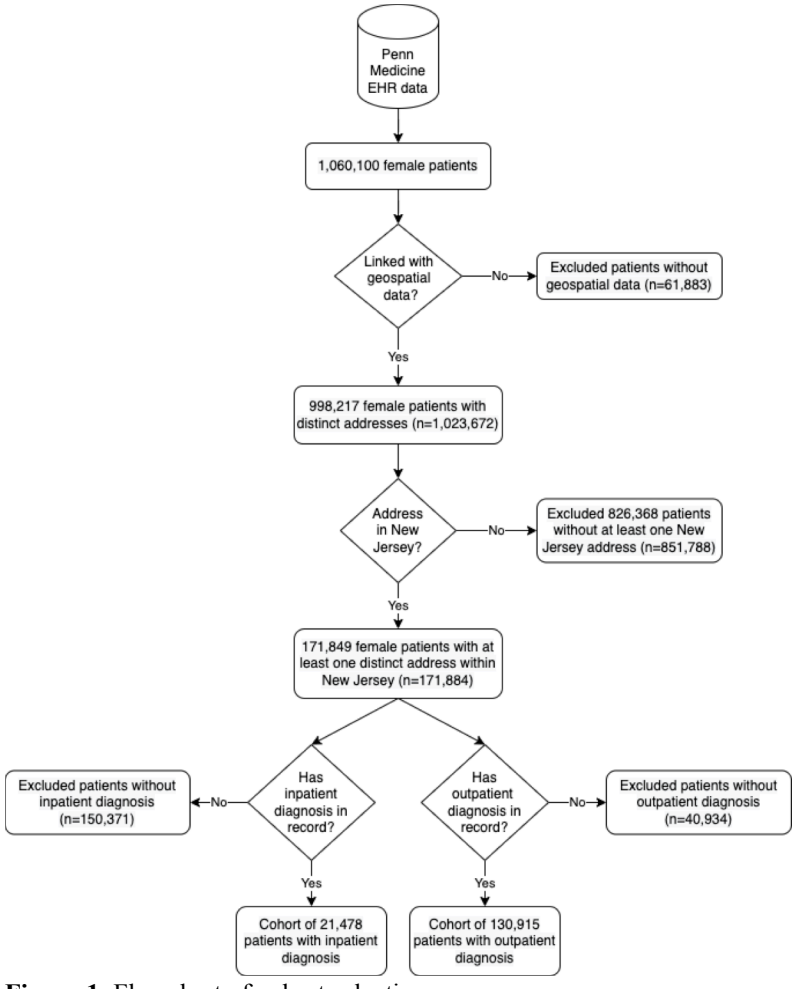

Figure 1. Flowchart of cohort selection process

3. Results

3.1 Penn Medicine Cohort Selection and Characteristics

Initially, we extracted an initial set of 1,060,100 female patients treated at Penn Medicine, with inpatient and outpatient visits between 2010-2017. We focused on female patients only for this analysis as many of the diseases correlated with PFAS exposure in the literature are more prevalent in females and our lab is focused on women's

health outcomes generally with a focus on environmental factors that affect fertility (which will be the focus of later work in this area). We geocoded all addresses using ArcGIS in prior work [17] and extracted the corresponding latitude and longitude coordinates. This resulted in a set of 998,217 female patients who had at least one address in the US that we were successfully able to geocode. We included only patients having at least one address located in New Jersey as the NJDEP data is limited to the state of New Jersey, resulting in 171,849 patients. Refer to **Figure 1** for the complete flowchart of our cohort selection process.

3.1.1. Inpatient and Outpatient Cohorts

Next, we created inpatient and outpatient cohorts and removed patients without diagnosis information, respectively. This yielded an inpatient cohort of 21,478 patients and an outpatient cohort of 130,915 patients. See **Figure 1** for illustration of the detailed process.

3.2 Identifying PFAS Exposure Boundaries

3.2.1 Spatial Boundaries Data Source

A total of 587 unique public water supplier service area boundaries were included in the NJDEP GIS PWS dataset. In our municipal boundary dataset, we found 565 unique municipality codes. Using the statistical software language R, we sourced 732 unique ZIP Codes in New Jersey state.

3.2.2 NJ DEP Dataset

The violations dataset included 97 unique PFAS MCL violations in New Jersey. We excluded noncommunity non-transient water supplier violations (n = 48) and found a total of 26 unique public water suppliers (PWSs) with at least one PFAS MCL violation. After, we manually linked these PWSs to relative ZIP Codes (n = 32) and municipalities, (n= 28). Then, we linked the PWS service area boundary shapefile dataset to the PFAS MCL violations dataset, resulting in 20 unique PWSID with at least one PFAS MCL Violation.

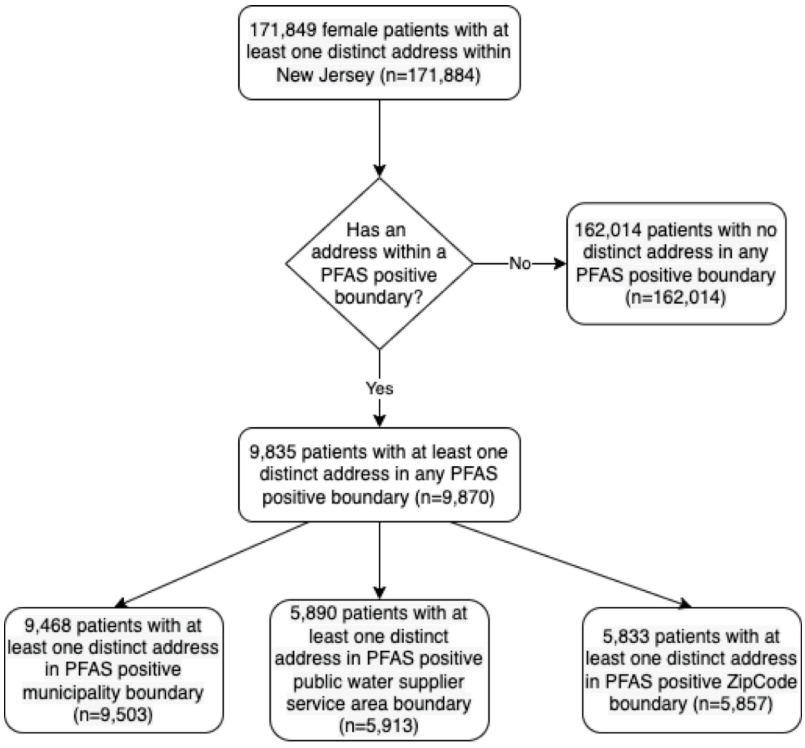

Figure 2. Flowchart of Binary classification of PFAS exposure using geospatial information

3.3 Linking PFAS Exposure Areas and Patient Geospatial Information

After linking the PFAS MCL Violations to PWSID, municipality codes, and ZIP code, we linked patient geospatial information to these respective boundaries (see **Figure 2**). This resulted in 5,890 patients exposed when using PWS service area boundary data, 9,468 patients when using municipality boundary data, and 5,833 patients when using ZIP code. We report demographics and characteristics of the complete patient cohort in **Table 1**. In comparison to other race and ethnicity categories, Black or African American patients were found to have a greater proportion exposed to PFAS in each boundary type (15.10 -19.21%).

3.3.1 Municipality Boundary Method Results in the Largest PFAS-Exposed Cohorts regardless of Inpatient vs. Outpatient Distinction

We stratified our cohorts by whether or not they were in-patient or outpatient because these different patient populations may have different reasons for traveling to PennMedicine for treatment. However, we still found differences between the cohorts in terms of the methods used to identify PFAS exposure. Some trends were consistent across cohorts, we found that the greatest number of PFAS exposed patients could be identified when using the municipality boundary method (n=934 for inpatient and n=7,684 for outpatient). For the inpatient cohort, we can increase our cohort size by 42% when using the municipality boundary method and for the outpatient data the increase is 60% of the cohort size. The actual exposure counts for PFAS exposed patients using the PWS service area boundary data (n = 684), municipality

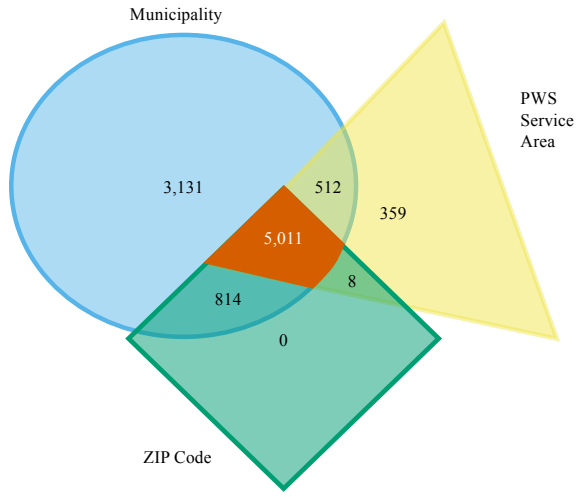

Figure 3. Intersection of PFAS exposed patients by boundary type

boundary data (n = 934), and ZIP code (n = 657) for the inpatient cohort. For the outpatient cohort, we found 4,815 PFAS exposed patients when using PWS service area boundary data, municipality boundary data (n = 7,684), and ZIP code (n = 4,792).

See **Figure 3** for the intersection of the PFAS exposure counts by boundary type. Municipality boundary data analysis resulted in the highest number of patients exposed overall (n = 9,468) and the most patients without overlap with PWS service area and ZIP code (n = 3,131). ZIP code overlapped with municipality and PWS service area, without patients exposed only due to ZIP code. PWS service area boundary, ZIP code, and municipality boundary share 5,011 patients exposed to PFAS; this resulting sample of patients is the referred to as the intersection in the subsequent analyses.

3.3.2 Environmental Justice: Black/African-American Patients Living in NJ had a Higher PFAS-Exposed Rate

We also found higher proportion of Black/African-American patients were exposed to PFAS regardless of the method selected (**Table 1**). In the inpatient cohort, while the PFAS exposed population consistently consists of a higher percent (51.16 - 58.57%) of White patients, we observe a higher proportion (13-15%) of Black or African American patients living in New Jersey exposed to PFAS than the proportion of White patients (2-3%) living in New Jersey. Similar to the inpatient cohort, for the outpatient cohort, the exposed population is mostly White patients (47.89 – 56.40%). A larger proportion of White patients represent the non-exposed across each boundary type (78.04 – 79.87%). A greater proportion of Black or African American patients were found to be exposed in all boundaries (16.61 - 21.01%) in comparison to other ethnicities.

Table 1. Overall patient cohort demographics by PFAS exposure boundary type

	PWS Boundary				**Municipality Boundary**				**ZIP Code**			
	PFAS		**No PFAS**		**PFAS**		**No PFAS**		**PFAS**		**No PFAS**	
	N	**%**	**N**	**%**	**N**	**%**	**N**	**%**	**N**	**%**	**N**	**%**
White	2,802	47.51	123,547	74.50	5,217	55.11	121,132	74.65	2,852	48.84	123,497	74.44
Black or African American	2,143	36.33	13,784	8.31	2,566	27.11	13,361	8.23	2,089	35.78	13,838	8.34
Asian	142	2.41	4,472	2.70	258	2.73	4,356	2.68	80	1.37	4,534	2.73
Other	558	9.46	16,505	9.95	985	10.41	16,078	9.91	571	9.78	16,492	9.94
Unknown	2,802	47.51	123,547	74.50	5,217	55.11	121,132	74.65	2,852	48.84	123,497	74.44
Hispanic?												
Yes	203	3.45	5,630	3.39	380	4.01	5,453	3.36	189	3.24	5,644	3.40
No	5,590	94.91	157,641	94.99	8,887	93.86	154,344	95.05	5,541	94.99	157,690	94.98
NA	97	1.65	2,688	1.62	201	2.12	2,584	1.59	103	1.77	2,682	1.62
Total Distinct Patients	5,890		165,959		9,468		162,381		5,833		166,016	

3.4 Inpatient and Outpatient Cohort and Disease Associations

After completing the ICD code crosswalk and limiting to codes diagnosed to at least 100 patients and at least 10 PFAS exposed patients, we found 6 codes given to the inpatient cohort and 31 codes given to the outpatient cohort

(including the two code groups 'toxic goiter' and 'nontoxic goiter', respectively). The direction of the relationships of the inpatient and outpatient cohorts are shown in **Figure 4A and 4B**, respectively. In these figures we report all significant associations (Bonferroni adjusted *P*-value<0.05). Overall, most diagnoses are found to be not associated with PFAS exposure (n = 4, 67%), no matter the boundary type. Consistently across boundary types, one diagnosis was found significant: essential hypertension (i.e. I10). The code hypertensive crisis, unspecified (i.e. 16.9) was found significant in three of the four classification types. In **Figure 5**, we illustrate through Manhattan plots the resulting Bonferroni adjusted *P*-values of each diagnosis code in our inpatient and outpatient cohort analyses, respectively, grouped by four disease types: goiter, hypertension, parathyroid, and thyroid.

Similar to the inpatient cohort analysis, our analysis of the outpatient cohort shows that most of the diagnosis codes have no significant association with PFAS exposure (n = 26, 74%). Of the 31 diagnosis codes, six were found to be significantly associated, albeit inconsistently across boundary types: nontoxic goiter code group, E04.2, I10, E03.9, C73, and E04.1. Of these diagnoses, only essential hypertension (i.e. I10) was found significant in more than one boundary type; those boundaries being PWS service area boundary, ZIP Code, and intersection of all boundaries. The other four diagnosis codes were found significant solely by municipality.

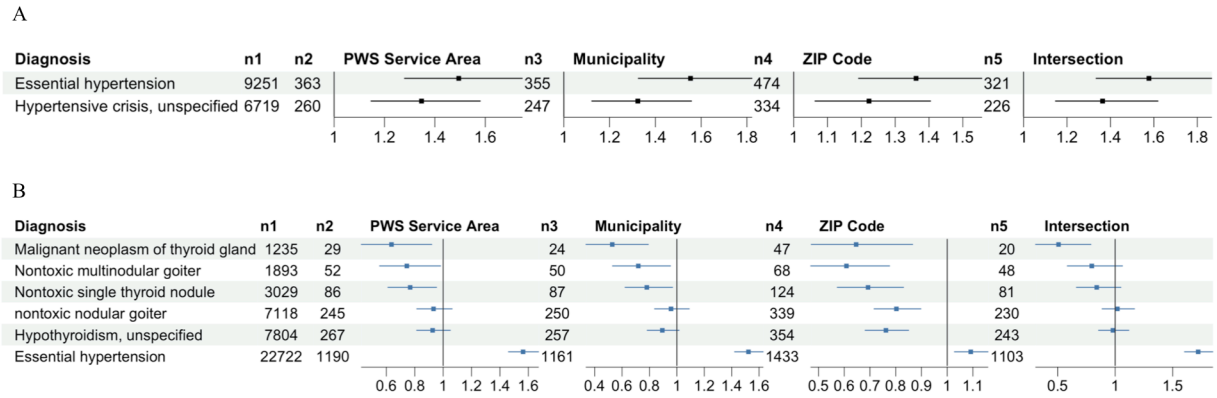

Figure 4 A. Forest Plot of Results of Inpatient Odds Ratios with 95% Confidence Intervals by Boundary Type. 4 B. Forest Plot of Results of Outpatient Odds Ratios with 95% Confidence Intervals by Boundary Type.

Columns are indicated as follows: n1(all patients with diagnosis), n2 (number of patients exposed in PWS service area with diagnosis), n3 (number of patients exposed in municipality with diagnosis), n4 (number of patients exposed in ZIP code with diagnosis), n5 (number of patients exposed in intersection of all boundaries with diagnosis).

4. Discussion

Our work demonstrates that changes in the method used to annotate PFAS exposure classification, which employs various geospatial analytic methods (i.e., different boundary data sets), will result in changes in the cohort and subsequent analyses. This can have a profound effect on PFAS exposure-disease association studies that result from such cohorts and are important for researchers to understand when utilizing address information for exposure classification. We explored the PFAS exposure-disease associations in outpatient and inpatient cohorts in order to understand if known associations in current literature can be captured in our analysis. In our case, the trends in cohort characteristics, demographics, and PFAS exposure-disease associations differ across the boundary types observed. Another profound difference is in terms of the sample sizes of our PFAS exposed populations, which also varied by PFAS exposure classification methods. Sample size can be very important in the power to detect an association.

4.1 Binary Classification of PFAS Exposure Across Different Boundary Datasets

Three types of boundaries were used in our binary classification of PFAS exposure: public water supplier (PWS) service area boundary, municipality boundary, and ZIP Code. Next, we observed the intersection of these, where patients were found to be PFAS exposed in all three boundaries. In combining the efforts of several geospatial datasets, we were able to create a unique cohort of patients who lived within the boundaries of all three boundaries. While there is no gold standard to this study, as this would require knowledge of PFAS serum levels, the intersection of these methods may prove to provide a conservative cohort of those exposed to PFAS chemicals. While it is unlikely for researchers to go to such efforts in classification, our work shows how other boundary datasets may be suitable for binary classification of exposure to contaminants through drinking water.

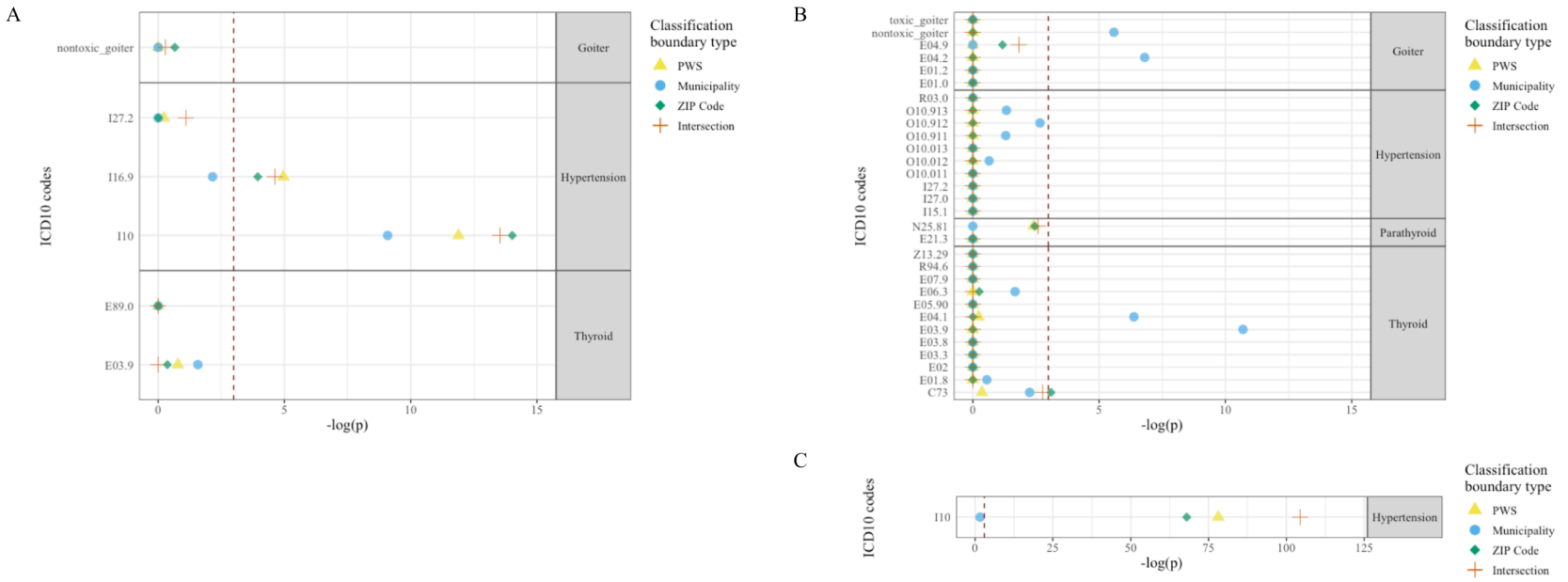

Figure 5 A. Manhattan plot of inpatient diagnosis codes by boundary type, B. Manhattan plot of outpatient diagnosis codes by boundary type, C. Manhattan plot of outpatient diagnosis of Essential hypertension (I10) by boundary type.

Classification boundary type colors coordinate with colors used in Figure 3 for clarity. The x-axis indicates -log(*P*) of Bonferroni adjusted *P*-values and the y-axis indicates the resulting ICD10 codes from the ICD code crosswalk. Codes are further organized and labeled by overarching disease types: goiter, hypertension, parathyroid and thyroid, respectively. The vertical red dashed line indicates the line of significance (-log(Bonferroni adjusted *P*-value > 0.05)). In order to fully illustrate the variance in the outpatient data, the essential hypertension diagnosis (I10) was separated from Figure 5B and illustrated in Figure 5C because the -log(*P*-value) is much more significantly associated with PFAS exposure than the other diagnoses.

ZIP Code (n = 5,833) and PWS service area boundaries (n = 5,890) demonstrate more conservative cohorts, while the classification through municipality boundaries resulted in a larger number of patients exposed (n = 9,468). It is likely the municipality cohort has misclassified several unexposed patients as PFAS exposed (i.e. false positive). In this case, it may be that municipality boundaries do not represent drinking water distribution areas well in comparison to ZIP Code and publicly available PWS service area boundaries. We delve into the limitations of using boundaries for exposure to contaminants in drinking water in our limitations and future work.

4.2 Exploring Known PFAS-Disease Associations in Our Cohorts

We observed diagnosis codes related to hypertension and thyroid disease, as these are associations with exposure to PFAS shown in previous literature. Overall, we found essential hypertension (i.e. I10) to be positively associated with PFAS exposure in our inpatient and outpatient cohorts. We further discuss results of the two cohorts in detail below. All reported OR and 95% CI below reflect those of the intersection of all boundaries' respective cohorts.

4.2.1 Inpatient Cohort

Of all the diagnoses observed, two were found positively associated with PFAS exposure, namely essential hypertension (i.e. I10) with an OR of 1.58 (85% CI: 1.33, 1.87), and unspecified hypertensive crisis (i.e. I16.9) with an OR of 1.36 (95% CI: 1.15, 1.62). This trend is observed for all four of the classification methods used, however the odds ratios and confidence intervals differ between the cohorts. We observe more consistency across boundary types in the inpatient cohort in comparison to the outpatient cohort. The odds ratios in the ZIP Code cohort tend to have wider confidence intervals in comparison to the other observed cohorts. Moreover, we found variance in the adjusted *P*-values between boundary cohorts (see **Figure 5A**).

4.2.2 Outpatient Cohort

We found more diagnosis codes in the outpatient cohort and only four of which were significant with Bonferroni adjusted *P*-values. Patients exposed to PFAS were more likely to be diagnosed with essential hypertension (i.e. I10) with an odds ratio of 1.72 (95% CI: 1.60, 1.85). The five other diagnosis codes were negatively associated with PFAS exposure. Of these, three have relatively weak, inconsistent associations, namely nontoxic goiter codes (OR: 1.02, 95% CI: 0.89, 1.17), unspecified hypothyroidism (i.e. E03.9, OR: 0.98, 95% CI:0.86, 1.12), and nontoxic multinodular goiter (i.e. E04.2, OR: 0.80, 95% CI:0.58, 1.06). A stronger negative association is found with malignant neoplasm of thyroid gland (i.e. C73), with an odds ratio of 0.51 (95% CI: 0.31, 0.79). However, it is important to note this is still a weaker association with a rather large confidence interval. In **Figure 5B and Figure 5C**, we find greater variance in the adjusted *P*-values between the boundary cohorts, especially the municipality cohort. Alone to the municipality outpatient cohort, four diseases were found significant with adjusted *P*-values that do not show significant in the other three cohorts. Overall, the outpatient cohort resulted in more variance between boundary classification methods.

4.3 Generalizability of Results beyond New Jersey

We performed our analysis using New Jersey data because we found that NJ water data from multiple sources was open access and governmental agencies within the state of NJ tested for multiple species of PFAS, including PFNA, which is often overlooked by other states health departments. Therefore, the comparison of methods using a state with multiple data sources was advantageous to us to perform the comparative analysis included in this paper. Some states do not have such rich data resources at their disposal and others rely more heavily on well water sources, which may or may not be required to test regular (depending again on state regulations). We believe our work is generalizable in that we provide information on how PFAS as an exposure label can vary depending on the method in the literature that is used to define this exposure. Others with data from other states or countries may not be able to perform these comparisons because their state or country may not provide such rich resources to the public and therefore, they can learn from our work to understand the gaps in the various methods and how the choice of method may alter the results obtained in their own studies.

4.4 Limitations and Further Work

The NJ DEP violations dataset is limited to three types of PFAS chemicals, due to monitoring rules in the state for PFOA, PFOS, and PFNA. As previously stated, PFAS are a large group of nearly 5000 human-derived compounds. There is limited information on new generation PFAS compounds, such as Hexafluoropropylene Oxide (HFPO) Dimer Acid and its Ammonium Salt (i.e. GenX chemicals). A majority of the evidence on lesser known PFAS derive from animal studies, with limited information from human exposure [29]. It is likely patients in the cohort are exposed to lesser known PFAS, as this information is outside the scope of our data. True exposure to PFAS-contaminated drinking water cannot be simply determined from spatial data; living within a PWS service area alone does not delve into the complexities of exposure to PFAS from drinking water (e.g. bottled water use, work location, distribution of well water from PWS, water filtration system use) let alone observance of other exposure pathways and sources (e.g. food consumption, indoor environment, outdoor air) [30]. We therefore determined binary exposure provides a simple first step to understanding PFAS exposure. Moreover, the NJ MCL violations dataset used in our

analysis is resulting from regulations that were implemented a few years after (PFNA: January 2020, PFOS & PFOA: January 2021) the years of the extracted EHR data included in our analysis (2010-2017). We did not adjust for residential mobility (i.e., patients moving to another residence) in this analysis because we are focused on methods to classify exposure for a given address at a given time-point. It is worth mentioning that residential mobility is an important issue when performing geospatial analyses using EHR data [31]. We also will include distance from Penn Medicine sites in our future work. This is rather complex given that there are multiple Penn Medicine sites and clinics and there are often multiple visits per patient. However, we acknowledge that this will be important in further delving into the disparities that we observed in this study. Future work will also explore the role of age in the relationship between PFAS exposure and disease risk.

5. Conclusion

We explored the utility of three common methods for linking geospatial information to PFAS exposure in drinking water and the effect that these different methods had on our cohort and findings. We investigated this using four resulting boundary types: public water supplier (PWS) service area boundary, municipality, ZIP code, and the intersection of the three boundaries (this is our own classification not used in the literature). In these boundaries, we found 3.39-5.83% of the complete cohort exposed to PFAS. The size of the cohort varied between boundary types, with municipality boundary classification resulting in 3,131 patients not captured within ZIP Code and PWS service area boundaries. Moreover, a larger proportion of Black/African-American patients were found to be exposed to PFAS across all cohorts. In effort to validate known PFAS exposure-disease associations, we investigated associations to respective diagnosis codes for hypertension, thyroid disease, and parathyroid disease. Consistent across inpatient and outpatient cohorts, we found positive associations between PFAS-exposure and essential hypertension (i.e. I10). Significant disease-exposure associations varied between boundary classification methods, especially in our outpatient cohorts. Our work demonstrates that changes in geospatial analytic methods, specifically depending on different boundary data sets, will lead to changes in the cohort and subsequent analyses.

Data Sharing: We plan on making code and other shareable resources available on our github page: https://github.com/bolandlab

Funding: This research was supported by the Institutional Clinical and Translational Science Award (CTSA) with Dr. Boland as a co-Investigator (UL1-TR-001878) with Dr. Garret Fitzgerald as PI. Generous funding also provided by the Perelman School of Medicine at the University of Pennsylvania.

References

1. US EPA. Drinking Water Health Advisories for PFOA and PFOS. [cited 2019 Oct 11]; Available from: https://www.epa.gov/ground-water-and-drinking-water/drinking-water-health-advisories-pfoa-and-pfos
2. US EPA O. PFAS Strategic Roadmap: EPA's Commitments to Action 2021-2024 [Internet]. 2021 [cited 2022 Mar 10]. Available from: https://www.epa.gov/pfas/pfas-strategic-roadmap-epas-commitments-action-2021-2024
3. Krietzman S. PFAS in New Jersey's Environment - NJDEP Evaluation and Response [Internet]. Webinar presented at; 2018 Nov 30. Available from: https://files.dep.state.pa.us/water/DrinkingWater/Perfluorinated%20Chemicals/Presentations/_New%20Jersey.pdf
4. NJDEP | PFAS | PFAS Standards and Regulations [Internet]. [cited 2022 Sep 14]. Available from: https://www.nj.gov/dep/pfas/standards.html
5. Berg V, Nøst TH, Huber S, Rylander C, Hansen S, Veyhe AS, et al. Maternal serum concentrations of per- and polyfluoroalkyl substances and their predictors in years with reduced production and use. 2014 Aug 1;69:58–66.
6. Pitter G, Da Re F, Canova C, Barbieri G, Zare Jeddi M, Daprà F, et al. Serum Levels of Perfluoroalkyl Substances (PFAS) in Adolescents and Young Adults Exposed to Contaminated Drinking Water in the Veneto Region, Italy: A Cross-Sectional Study Based on a Health Surveillance Program. Environ Health Perspect. 2020 Feb 18;128(2):027007.
7. Steenland K, Tinker S, Frisbee S, Ducatman A, Vaccarino V. Association of Perfluorooctanoic Acid and Perfluorooctane Sulfonate With Serum Lipids Among Adults Living Near a Chemical Plant. 2009 Nov 15;170(10):1268–78.
8. Barry V, Winquist A, Steenland K. Perfluorooctanoic acid (PFOA) exposures and incident cancers among adults living near a chemical plant. 2013;121(11–12):1313–8.
9. Xu Y, Fletcher T, Pineda D, Lindh CH, Nilsson C, Glynn A, et al. Serum Half-Lives for Short- and Long-Chain Perfluoroalkyl Acids after Ceasing Exposure from Drinking Water Contaminated by Firefighting Foam. Environ Health Perspect [Internet]. 2020 Jul 10 [cited 2021 May 13];128(7). Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7351026/

10. LaKind JS, Verner MA, Rogers RD, Goeden H, Naiman DQ, Marchitti SA, et al. Current Breast Milk PFAS Levels in the United States and Canada: After All This Time, Why Don't We Know More? Environ Health Perspect. 2022 Feb 23;130(2):025002.
11. Sadler RC. Misalignment Between ZIP Codes and Municipal Boundaries: A Problem for Public Health. Cityscape Wash DC. 2019;21(3):335–40.
12. Balocchi C, Bai R, Liu J, Canelón SP, George EI, Chen Y, et al. Uncovering Patterns for Adverse Pregnancy Outcomes with Spatial Analysis: Evidence from Philadelphia [Internet]. arXiv; 2022 [cited 2022 Sep 14]. Available from: http://arxiv.org/abs/2105.04981
13. Meeker JR, Canelón SP, Bai R, Levine LD, Boland MR. Individual-Level and Neighborhood-Level Risk Factors for Severe Maternal Morbidity. Obstet Gynecol. 2021 May;137(5):847–54.
14. Liu L, Ni Y, Beck AF, Brokamp C, Ramphul RC, Highfield LD, et al. Understanding Pediatric Surgery Cancellation: Geospatial Analysis. J Med Internet Res. 2021 Sep 10;23(9):e26231.
15. Ward-Caviness CK, Moyer J, Weaver A, Devlin R, Diaz-Sanchez D. Associations between PFAS occurrence and multimorbidity as observed in an electronic health record cohort. Environ Epidemiol. 2022 Jul 14;6(4):e217.
16. Boland MR, Davidson LM, Canelón SP, Meeker J, Penning T, Holmes JH, et al. Harnessing electronic health records to study emerging environmental disasters: a proof of concept with perfluoroalkyl substances (PFAS). Npj Digit Med. 2021 Aug 11;4(1):1–10.
17. Boland MR, Liu J, Balocchi C, Meeker J, Bai R, Mellis I, et al. Association of Neighborhood-Level Factors and COVID-19 Infection Patterns in Philadelphia Using Spatial Regression. AMIA Annu Symp Proc. 2021 May 17;2021:545–54.
18. NJDEP | PFAS | PFAS in Drinking Water [Internet]. [cited 2022 Sep 9]. Available from: https://www.nj.gov/dep/pfas/drinking-water.html
19. Ballesteros V, Costa O, Iñiguez C, Fletcher T, Ballester F, Lopez-Espinosa MJ. Exposure to perfluoroalkyl substances and thyroid function in pregnant women and children: A systematic review of epidemiologic studies. 2017 Feb 1;99:15–28.
20. Blake BE, Pinney SM, Hines EP, Fenton SE, Ferguson KK. Associations between longitudinal serum perfluoroalkyl substance (PFAS) levels and measures of thyroid hormone, kidney function, and body mass index in the Fernald Community Cohort. Environ Pollut Barking Essex 1987. 2018 Nov;242(Pt A):894–904.
21. C8 Science Panel. C8 Probable Link Reports: Probable Link Evaluation of Thyroid Disease. 2012 p. 1–12.
22. Coperchini F, Croce L, Ricci G, Magri F, Rotondi M, Imbriani M, et al. Thyroid Disrupting Effects of Old and New Generation PFAS. Front Endocrinol. 2021 Jan 19;11:612320.
23. C8 Science Panel. Probable Link Evaluation of Pregnancy Induced Hypertension and Preeclampsia. 2011 p. 1–6.
24. Ding N, Karvonen-Gutierrez CA, Mukherjee B, Calafat AM, Harlow SD, Park SK. Per- and Polyfluoroalkyl Substances and Incident Hypertension in Multi-Racial/Ethnic Women: The Study of Women's Health Across the Nation. Hypertens Dallas Tex 1979. 2022 Aug;79(8):1876–86.
25. Huang R, Chen Q, Zhang L, Luo K, Chen L, Zhao S, et al. Prenatal exposure to perfluoroalkyl and polyfluoroalkyl substances and the risk of hypertensive disorders of pregnancy. 2019 Dec 9;18(1):5.
26. Min JY, Lee KJ, Park JB, Min KB. Perfluorooctanoic acid exposure is associated with elevated homocysteine and hypertension in US adults. Occup Environ Med. 2012 Sep;69(9):658–62.
27. Pitter G, Zare Jeddi M, Barbieri G, Gion M, Fabricio ASC, Daprà F, et al. Perfluoroalkyl substances are associated with elevated blood pressure and hypertension in highly exposed young adults. Environ Health Glob Access Sci Source. 2020 Sep 21;19(1):102.
28. Exposure to Perfluoroalkyl Chemicals and Cardiovascular Disease: Experimental and Epidemiological Evidence - PubMed [Internet]. [cited 2022 Sep 14]. Available from: https://pubmed.ncbi.nlm.nih.gov/34305819/
29. North Carolina Department of Health and Human Services. Biological sampling for GenX and other Per- and Polyfluoroalkyl Substances (PFAS)—North Carolina, 2018 [Internet]. 2018 [cited 2022 Mar 3]. Available from: https://epi.dph.ncdhhs.gov/oee/pfas/NCDHHS_PFAS%20Biomonitoring%20Report_8Nov2018.pdf
30. De Silva AO, Armitage JM, Bruton TA, Dassuncao C, Heiger-Bernays W, Hu XC, et al. PFAS Exposure Pathways for Humans and Wildlife: A Synthesis of Current Knowledge and Key Gaps in Understanding. Environ Toxicol Chem. 2021;40(3):631–57.
31. Meeker JR, Burris H, Boland MR. An algorithm to identify residential mobility from electronic health-record data. Int J Epidemiol. 2021 Dec 1;50(6):2048–57.