

Original Paper

Geospatial Imprecision With Constraints for Precision Public Health: Algorithm Development and Validation

Daniel Harris¹, PhD; Chris Delcher², PhD

¹Institute for Biomedical Informatics, University of Kentucky, Lexington, KY, United States

²Department of Pharmacy Practice and Science, College of Pharmacy, University of Kentucky, Lexington, KY, United States

Corresponding Author:

Daniel Harris, PhD

Institute for Biomedical Informatics

University of Kentucky

Healthy Kentucky Research Building, Room 361

760 Press Avenue

Lexington, KY, 40508

United States

Phone: 1 8593237100

Email: daniel.harris@uky.edu

Abstract

Background: Location and environmental social determinants of health are increasingly important factors in both an individual's health and the monitoring of community-level public health issues.

Objective: We aimed to measure the extent to which location obfuscation techniques, designed to protect an individual's privacy, can unintentionally shift geographical coordinates into neighborhoods with significantly different socioeconomic demographics, which limits the precision of findings for public health stakeholders.

Methods: Point obfuscation techniques intentionally blur geographic coordinates to conceal the original location. The pinwheel obfuscation method is an existing technique in which a point is moved along a pinwheel-like path given a randomly chosen angle and a maximum radius; we evaluate the impact of this technique using 2 data sets by comparing the demographics of the original point and the resulting shifted point by cross-referencing data from the United States Census Bureau.

Results: Using poverty measures showed that points from regions of low poverty may be shifted to regions of high poverty; similarly, points in regions with high poverty may be shifted into regions of low poverty. We varied the maximum allowable obfuscation radius; the mean difference in poverty rate before and after obfuscation ranged from 6.5% to 11.7%. Additionally, obfuscation inadvertently caused false hot spots for deaths by suicide in Cook County, Illinois.

Conclusions: Privacy concerns require patient locations to be imprecise to protect against risk of identification; precision public health requires accuracy. We propose a modified obfuscation technique that is constrained to generate a new point within a specified census-designated region to preserve both privacy and analytical accuracy by avoiding demographic shifts.

(*Online J Public Health Inform* 2024;16:e54958) doi: [10.2196/54958](https://doi.org/10.2196/54958)

KEYWORDS

social determinants of health; geocoding; privacy; poverty; obfuscation; security; confidentiality; low income; geography; geographic; location; locations; spatial; geospatial; precision

Introduction

Geographic information systems (GISs) are increasingly important for public health research and policy makers and are instrumental in measuring socioeconomic equity in health care [1,2]. Social determinants of health are the conditions in which individuals are born, live, work, and age; mesolevel determinants are from the physical environment and encompass items such

as geographic location and access to resources [3]. Location-based exposures tied to geographic location are a pivotal element to one's health [4-6]; ongoing research suggests that zip code is on par with genetic code in influencing individual health [7-10]. Even greater research utility lies in more precisely geolocating patients beyond the zip code level, yet privacy regulations often prevent high-resolution patient residential address data from being shared for research purposes

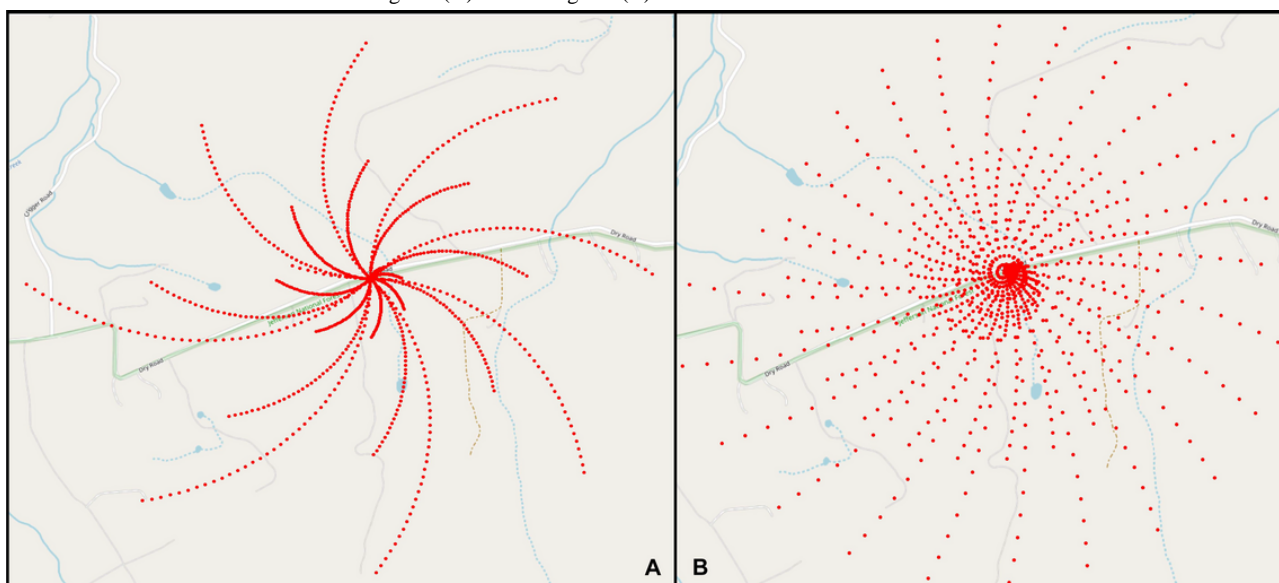
[11]. Privacy is paramount when working with health care data and access is regulated at different levels through both institutional policies and government-mandated legal protections [12]. The Health Insurance Portability and Accountability Act (HIPAA) mandates privacy protections of personal health information in the United States; it outlines which data elements are considered private, including patient addresses needed for geospatial analysis for place-based epidemiology.

Although institutional review boards may grant researchers and other agencies access to identified data that pose minimal risk to the patient, there may exist institutional hesitancy to disclose this data due to the inherent privacy and sensitivity of residential addresses. As a current example, this tension is apparent between state and federal public health and safety agencies using the Overdose Detection Mapping Application Program (ODMAP), which maps in real time the exact locations of suspected drug overdoses, often occurring in residential locations [13,14]. Geocoding a patient's address (ie, converting to geographic coordinates) is often an intermediate step in secondary data analyses; it is either used to link the patient to external geographic units (eg, census tracts to obtain neighborhood socioeconomic status from the United States Census Bureau) or to calculate distance from other entities, such as health care providers and facilities. For example, accessibility of buprenorphine, a medication for opioid use disorder, may be determined using addresses of health care providers that are authorized to prescribe the medication [15]. In these examples, imprecise locations may be sufficient for confident linkage to administrative units or approximate distance measures and preferred for research to preserve privacy. In these scenarios, thoughtful and controlled techniques designed to generate inexact data are needed to reduce precision [12,16].

To this end, geospatial or location-based privacy methods seek to maintain an appropriate level of confidentiality for a given task, service, or application while balancing the utility that these offer [17-19]. For example, users of location-based services on a cellular phone expect some level of privacy when sending personal data, and different strategies exist that anonymize pools of people by anonymizing data at point of collection. Location-based k -anonymity provides a method where one's data and location are indistinguishable from $k-1$ other people [20]. Other methods, such as geographic masking, alter coordinates systematically to limit the risk of reidentification when releasing data [21]; no universally accepted method exists for protecting geospatial privacy [16].

Point obfuscation refers to the deliberate degrading of the resolution of coordinate information with the goal of protecting the privacy of the individual represented by the point [22]; this may be referred to as geographic masking, geomasking, jittering, or dithering and relies upon transformations or perturbations using randomness or artificial noise [22]. The N -RAND algorithm generates N candidate points in a given area and selects the furthest point [23]; θ -RAND limits candidate points to a specific area defined by a chosen angle [24]. We introduce modifications to the pinwheel obfuscation method which shifts points along pinwheel-like paths for a randomly chosen angle [25]; examples are shown in Figure 1. The noise added by this method is asymmetrical and highly variable, making it less open to privacy attacks designed to eliminate uniform and predictable noise [25]. However, the limitation of any point obfuscation technique is that coordinate shifts may change real-world locations and distort the linked health-related metrics. For example, a study defining a participant's rurality based on administrative units may be impacted if obfuscated coordinates move the participant across boundaries into an urban area.

Figure 1. Pinwheel obfuscation at theta 45 degrees (A) and 15 degrees (B).



Address *correction* positively impacts the accuracy of assigning a patient to a geographic area [26]. On the other hand, the goal of point obfuscation is to *intentionally generate an incorrect* address to preserve patient privacy without compromising

analytical conclusions. Our paper demonstrates that indiscriminate point obfuscation impacts studies linking points to neighborhood-level socioeconomic demographics and subsequently provides new methods needed to constrain

pinwheel obfuscation to yield results confined in specific census-designated regions, such as blocks, block groups, or tracts. The constraint reduces concerns that neighborhood-level measures are inappropriately assigned at the patient level, leading to misclassification bias. We use poverty status as an example of data recorded by the United States Census Bureau to explore the potential impact of unintentional administrative boundary shift. Also, we demonstrate how indiscriminate point obfuscation impacts hot spot analysis at the census tract level. The role of this work is to provide evidence that point obfuscation techniques may substantially alter neighborhood-level socioeconomic demographics and that the intentional imprecision in these techniques must be constrained to support precision public health.

Methods

Overview

We implemented our methods using PostGIS, an open-source project that adds geospatial objects and procedures to the PostgreSQL database. We previously demonstrated PostGIS as a capable environment for geospatial privacy research [27,28]. We make our custom PostGIS functions available as open-source software [29]. The pinwheel technique was originally designed because other point obfuscation methods could be reversed by methods designed to filter uniform noise; the randomness of the pinwheel has been shown to maintain high variability, making it less susceptible to privacy attack [25]. Our geographically constrained pinwheel algorithm leverages the same concept as the original pinwheel algorithm and improves its research utility by adding constraint checking that controls how the new obfuscated point is selected. There is a function, *PINWHEEL*, that obfuscates a single point given a specific *theta*, maximum radius *r*, and a calculated random degree *a*; these are used to calculate a projection distance that can leverage PostGIS's projection function, *ST_PROJECT*, to calculate the resulting obfuscated point.

The left-hand side of Figure 1 shows 1000 candidate points for a given seed point using a 45-degree angle; we sequentially varied the maximum radius while keeping *theta* constant at 45 degrees. Similarly, the right-hand side of Figure 1 shows the same simulation but with 15-degree angles; this demonstrates that *theta* controls the width of the pinwheel layers and that small angles naturally yield closer layers. In practice, a random degree may be used to further obfuscate the results.

Census Bureau geometries are input as reference data. The smallest census-based geographic boundary, the census block, is contained in the block group; a census block group typically represents between 600 and 3000 people. Block groups are organized into census tracts, which typically contain between 1200 and 8000 people and have an optimum size of 4000 people [30]. The United States Census Bureau publishes and updates files containing the geometries for these regions; these geometries are used in point intersection calculations to assign a region to a given point.

To obfuscate, points from the original list, *P1*, are fed into the *PINWHEEL* function and saved into *P2*. To constrain

obfuscation, our method recalculates the pinwheel candidate for any generated candidates falling outside a specific region associated with the original point; we can constrain to standard administrative units: state, county, tract, block group, and block geometry. Furthermore, we can constrain obfuscation to custom geometries, such as buffer zones or other areas that may be relevant for research projects. The region of the points can be calculated with PostGIS's *ST_CONTAINS* function, which tests the intersection of the points with the Census Bureau geometries. The regions are compared for every matching pair of existing and new points in *P1* and *P2*. If the regions are dissimilar, *PINWHEEL* is rerun on the existing point. This continues until all points in *P1* have a matching obfuscated point in *P2* where *P1* regions align with *P2* regions.

We tested our methods with 2 different data sets, with geographic coverage ranging from multiple states to a single large urban area. Our first data set contained 1,000,000 records formatted in the Observational Medical Outcomes Partnership (OMOP) common data model [31]; we previously leveraged this data for geospatial research on open data and privacy [27]. Our second data set contains 58,102 case records from the Medical Examiner Case Archive from Cook County, Illinois, which contains the city of Chicago; we previously used this open data for research on geographic clustering of fatal overdoses [32] and to create an open data pipeline for spatial analyses on substance use disorders [33]. This open data set was released by the Cook County Medical Examiner's Office (CCMEO) and offers details on all deaths recorded by the CCMEO from August 2014 to April 2022, including the address where the incident occurred and the address of the death. Deaths by suicide recorded in the CCMEO data set were also used for our hot spot analysis example.

Ethical Considerations

This study was exempt from ethical review since no private health data were used and no human subjects were involved.

Results

As expected, the pinwheel obfuscation method initially resulted in points shifting into different geographic regions; the frequency of these region shifts is summarized in Table 1 for our OMOP and CCMEO data. We categorized these shifts by census-designated regions by increasing population size (block, block group, tract, county, and state). Shifts were proportional to the maximum radius allowed; a radius of 1000 meters was the largest distance tested and naturally generated the most region shifts. Blocks are the smallest geographic unit and experienced the most change: 747,934 of 1,000,000 (74.8%) of the OMOP points and 53,415 of 58,102 (91.9%) of the CCMEO points were moved to different census blocks after obfuscation up to 1000 meters away. This empirically indicates that smaller geographical regions are more likely to shift when using any significant distance in the obfuscation method. Blocks are the smallest of the census-designated administrative boundaries. There is no maximum size for a census block; minimum block size is 30,000 square feet (2787.1 m²) for polygons bounded by roads or 40,000 square feet (3716.1 m²) otherwise, which is

smaller than the largest obfuscation distance selected for our testing [34].

Additionally, the results demonstrate that even with very small distances unintentional consequences may occur; even moving the point using a radius of 1 meter resulted in misclassification. Although relatively rare, point obfuscation 1 meter away could change a point’s county in 0.0007% of the OMOP data (7/1,000,000) or a point’s census tract in 0.09% of the CCMEO data (58/58,102). There are no universally accepted best practices for point obfuscation, and the most effective allowed distance may vary with study area and application [16]. Due to the inclusion of the city of Chicago in the Cook County data, the census-designated regions in the CCMEO data are geographically smaller than those in our OMOP data, which cover multiple states; census tracts generally contain between 1200 and 8000 residents, meaning urban census tracts are geographically smaller than rural census tracts. This can be seen in our results, where a radius of 100 meters or larger yielded a higher percentage of shifts in our CCMEO data than our OMOP data.

Our results demonstrate that indiscriminate point obfuscation can shift a point into different census-designated geographical regions; this is a natural and expected consequence of moving a point. However, we now discuss and quantify the potential impact of shifting to linked administrative units (ie, neighborhoods) by comparing census demographics before and

after obfuscation. The United States Census Bureau conducts large-scale surveys, such as the decennial census and the American Community Survey (ACS). The yearly ACS samples approximately 250,000 household units monthly. From the ACS, we picked the estimated number of “individuals with income in the past 12 months below poverty level” as an example demographic; these data are publicly available at the census tract level. We chose poverty status due to its saliency in health outcomes research [35,36].

We give a high-level overview of the obfuscation impact on poverty status measurement in Table 2 to justify the need for a geographically constrained obfuscation technique when assigning points to a population-based rate of individuals living under the poverty line (as a percentage of the total population). In our OMOP data, a pinwheel distance of 1000 meters resulted in 22.4% (n=224,065) of records with a different poverty rate after obfuscation where those changes were, on average, a mean 7.3% (SD 7.4%) away from the original rate (median 5%, range 91.9% to -78.6%). We also show the magnitude of the difference between the original and obfuscated address by showing minimum and maximum differences of rates. Negative differences imply the obfuscated record had a lower assigned poverty rate while positive differences imply the obfuscated record had a higher assigned poverty rate. For completeness and to complement Table 1, we include small distances of 10 and 1 meters in Table 2, although we did not anticipate such small distances would impact poverty rate assignments.

Table 1. The number of records shifted into a different region varied per obfuscation radius (in meters) and size of region.

Data set and radius (meters)	Records, n×1000 (%)				
	Block	Block group	Tract	County	State
Observational Medical Outcomes Partnership (n=1,000,000)					
1000	747 (75)	367 (37)	226 (22)	16 (1.6)	1.1 (0.11)
500	616 (62)	210 (21)	114 (12)	8 (0.8)	0.479 (0.049)
100	236 (24)	35 (3.5)	16 (1.6)	1.3 (0.13)	0.071 (0.007)
1	3.2 (0.34)	0.359 (0.03)	0.194 (0.02)	0.08 (0.008)	0.004 (0.0004)
1	0.304 (0.03)	0.044 (0.004)	0.018 (0.001)	0.007 (0.0007)	0 (0)
Cook County Medical Examiner’s Office (n=58,102)					
1000	53 (91.9)	40 (70.4)	31 (53.3)	0.161 (0.27)	0.034 (0.05)
500	49 (85.4)	29 (50.7)	19 (33)	0.082 (0.14)	0.01 (0.02)
100	27 (47.7)	7.8 (13.4)	4.5 (7.8)	0.012 (0.02)	0.001 (0.002)
10	1.2 (2.22)	0.495 (0.85)	0.343 (0.59)	0.001 (0.002)	0 (0)
1	0.187 (.032)	0.082 (0.14)	0.058 (0.09)	0 (0)	0 (0)

Table 2. Poverty rates are substantially different before and after obfuscation.

Data set and distance (meters)	Records with changed poverty rate, n (%)	Difference in rate (%), mean (SD)	Difference in rate (%), median (maximum to minimum)
Observational Medical Outcomes Partnership (n=1,000,000)			
1000	224,065 (22.4)	7.3 (7.4)	5 (91.9 to -78.6)
500	113,682 (11.3)	7.3 (7.5)	5 (91.9 to -78.6)
100	16,121 (1.6)	7.4 (7.7)	5.1 (78.6 to -78.6)
10	193 (0.01)	6.5 (6.4)	4.2 (35.9 to -32.6)
1	18 (0)	11.7 (8.1)	10.5 (38.8 to -12)
Cook County Medical Examiner's Office (n=58,102)			
1000	30,843 (53.1)	8.3 (7.7)	6 (61.7 to -69)
500	19,139 (32.9)	8.2 (7.6)	5.9 (58.4 to -61.7)
100	4506 (7.8)	8.2 (7.5)	6 (58.4 to -58.4)
10	343 (0.6)	8.1 (7.3)	6.2 (33.2 to -55.4)
1	58 (0.1)	6.8 (6.5)	4.4 (31 to -18.2)

A larger percentage of records in the CCMEO data experienced rate changes in comparison to our OMOP data. With a distance of 1000 meters, 53.1% (n=30,843) of records were assigned into a region with a different poverty rate where those changes were a mean 8.3% (SD 7.7%) away from the original rate (median 6%, range 61.7% to -69%). The magnitude of changes was not substantially different from our OMOP data (averages of 7.3% vs 8.3% and medians of 5% vs 6%, respectively, for 1000 meters); yet the frequency of these changes was notably higher (22.4% vs 53.1%, respectively, for 1000 meters).

Figure 2 shows an example census tract (17031031100) in Cook County, Illinois; 33 deaths were recorded in this area. Figure 2A shows an example simulation using pinwheel obfuscation with a 1000-meter radius; Figure 2B shows the results of our geographically constrained pinwheel obfuscation. The original point is orange, and the obfuscated point is blue; the census tracts are colored according to quintile of our poverty measure, where lightly colored areas have the lowest poverty rates. For this example, pinwheel obfuscation resulted in 22 of 33 (66%) of the points shifting census tracts; 12 of 33 (36%) of these were shifted into areas of higher poverty, while 10 of 33 (33%) were shifted into areas of lower poverty. Of those positive shifts, 4 of 12 were pushed to the highest category (33%) while the other 8 were moved into the second-highest poverty quintile. This example shows how obfuscation may move points from one extreme to another.

Hot spot analysis is known to be an effective tool for understanding how health outcomes and social determinants of

health concentrate and cluster together [37-39]. We explored the impact of indiscriminate point obfuscation on a hot spot analysis of deaths by suicide from our CCMEO data; suicides were identified by the manner of death field in the CCMEO data and span the time period August 2014 to April 2022. Figure 3 shows the results of hot spot analyses using ArcGIS Pro [40] on the original data (Figure 3A) and data obfuscated with the pinwheel method using an unconstrained 1000-meter radius (Figure 3B). The obfuscation naturally blurred the correct hot spots, but (unexpectedly) new hot spots emerged, as identified by the pink boxes in Figure 3B. Most notably, a hot spot (95% confidence) spills into the neighboring and uninhabited Lake Michigan (census tract 17031990000). Highlighted in green are regions of interest with substantial change; the upper green box demonstrates the disappearance of a hot spot (99% confidence), and the lower green box demonstrates how indiscriminate obfuscation can bridge 2 hot spots together and weaken the signal that distinct clusters exist. By definition, the hot spots corresponding to the geographically constrained pinwheel method are identical to the true clusters in Figure 3A because of the confinement to the point's original census tract. When linked to an administrative boundary such as census tract, results are consistent before and after obfuscation when the pinwheel method is constrained; only distance-based results would be impacted by moving the original point. By constraining the pinwheel process to a specific geographic region, the results of any method depending upon aggregation within those regions will not be impacted by our method.

Figure 2. Poverty and point obfuscation of original points (orange) to obfuscated points (blue) using a pinwheel (A) and a geographically constrained pinwheel (B).

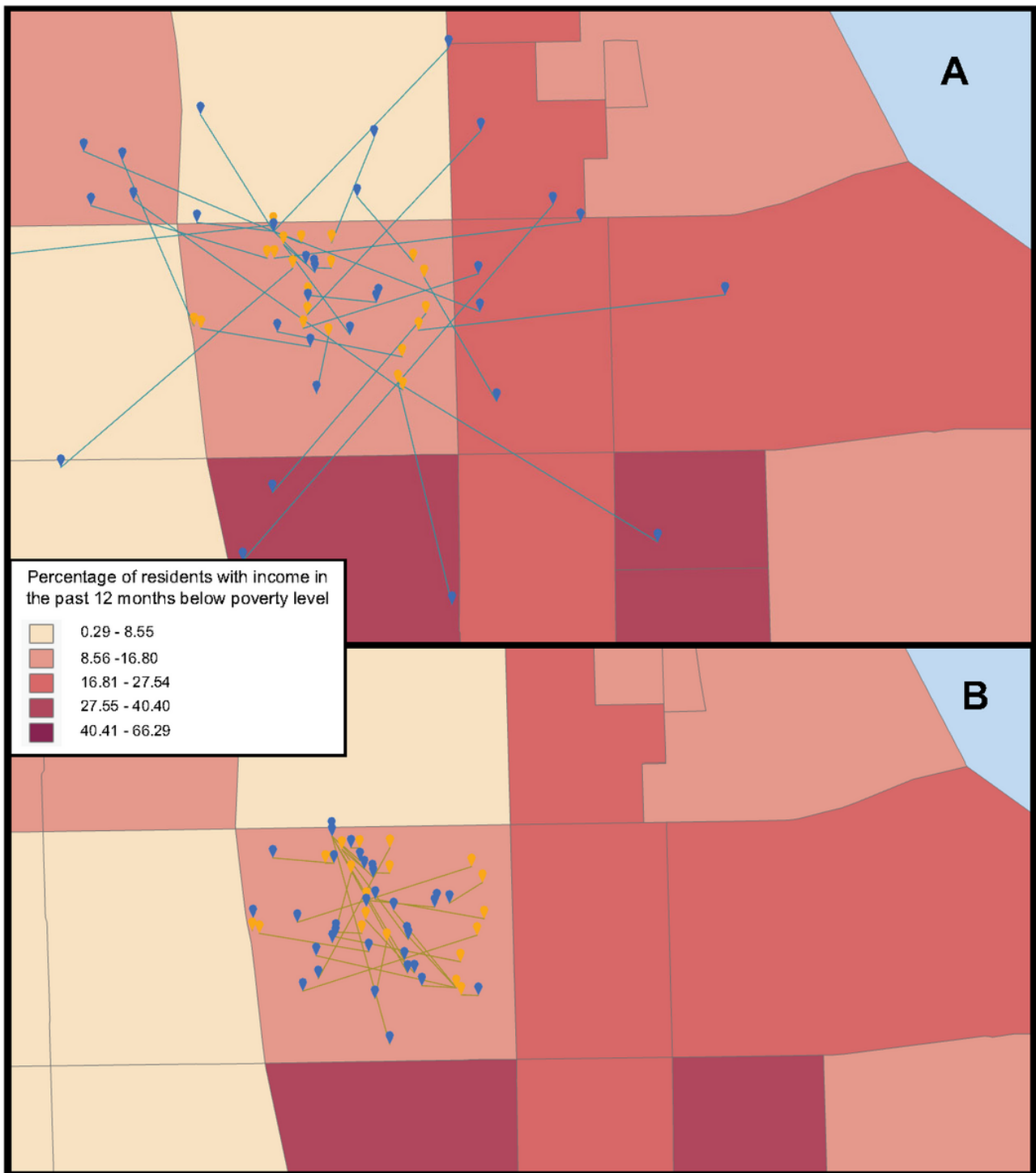
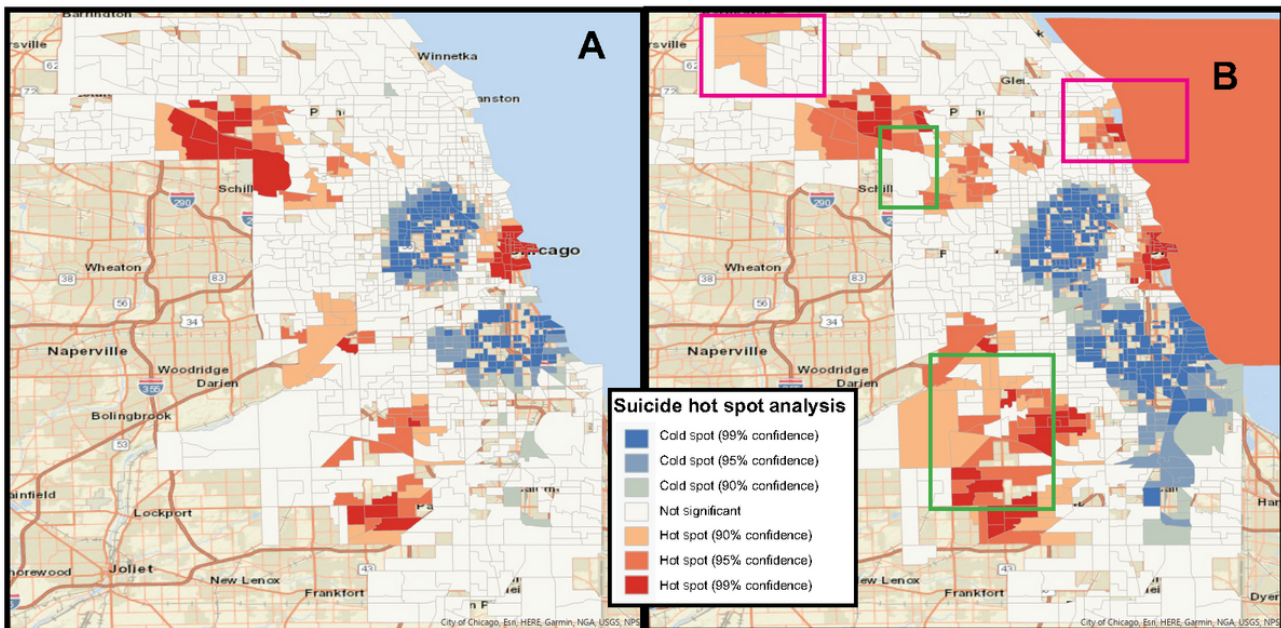


Figure 3. Hot spot analysis of suicides using original data (A) and obfuscated data (B).

Discussion

Principal Findings

We demonstrated that imprecise point obfuscation results in shifts across geographic regions and showed that these shifts do result in points geolocating in regions with vastly different socioeconomic contexts. This justifies the need for more precise point obfuscation techniques; our method constrains the candidate points into a specific region, which guarantees identical regional demographics after privacy protection is applied. The official poverty rate for the United States was 11.5% in 2022, and in the state of Kentucky, it was 16.5%, which places it 46th in a ranking of poverty rates in the United States [41]. For comparison, New Hampshire was ranked first and has the lowest poverty rate of 7.2% [41]. These example rates for the United States indicate areas experiencing differences of 5% to 7% in poverty rate, such as those reported in Table 2, represent vastly different socioeconomic dynamics. The extreme of this is illustrated in Table 2, where the largest difference between rates before and after obfuscation was 91.9% when records having relatively low poverty rates were assigned into areas having extreme poverty rates of 100% after obfuscation. The frequency of rate differences was substantially higher in the CCMEQ data, which represents only Cook County, Illinois, and is home to Chicago, the third-largest urban area in the United States. In the 1990s, there were notable declines in the concentration of poverty in Chicago [42]. This decline, mixed with concerns regarding how gentrification has impacted the socioeconomic dynamics of Chicago, may explain why changes in poverty rate would occur at a higher frequency than in our larger OMOP data [43]. The impact of these shifts is important to understand when working with sensitive, protected health information and social determinants of health to correctly identify associations between place and health. As an example of poverty and health, women in high-poverty places are at greatest risk of being diagnosed with late-stage breast cancer [44]. Furthermore, different research studies may require

different definitions of “neighborhood” when accessing socioeconomic statuses; for example, a person with multiple economic disadvantages may have a much narrower spatial range and limited social mobility.

We included small distances in our analysis to show that shifts occur even at very small distances. In obfuscation practice, small distances, such as 1 meter, would likely not be used due to the shifted point being too close to the original point and therefore not providing privacy; the balance between protecting privacy and protecting utility is context sensitive [21].

Limitations

Different analytical applications may be variably sensitive to shifts in demographics; our method eliminates analytical concerns by avoiding shifts altogether. The caveat to our method is that constraining inherently limits the maximum distance a point can travel, which may not be suitable for all applications in terms of privacy requirements. For example, in Figure 2, pinwheel obfuscation moved points on average 482 meters away, while our geographically constrained pinwheel algorithm moved points on average 217 meters away. Some applications may not be suitable for point obfuscation; for example, research studies requiring distance to be preserved between subjects and waypoints such as hospitals or clinics may not tolerate any shifting of geographic coordinates.

Public Health Impact

A previous evaluation of the pinwheel obfuscation method indicated that it had no major impact on geospatial analyses such as heat maps and hot spots [45]. However, we demonstrated that erroneous hot spots may be generated when analyzing deaths by suicide in Cook County, Illinois, including hot spots in uninhabitable areas. The issue presented in this paper is not a deficiency of the pinwheel technique, but a deficiency of data linkage using obfuscated points generated from any technique; if the variable that data linkage depends upon changes during obfuscation, then utility is harmed. We contend that any point

obfuscation technique may be constrained to specific geographies.

An alternative solution could be that socioeconomic demographics are calculated using the real address data before data release, but data owners, especially state and local health agencies, have varying degrees of technical sophistication and may not be able to compute demographics. Our research group geocodes electronic health records on behalf of our local health care enterprise on campus, and we make the data available to any university researcher using our local data warehouse. Mobile applications, networking, and research with Internet of Things technology have explored privacy at different levels; our work is a step closer to context-aware point obfuscation within the epidemiology domain.

Conclusions

A growing number of publicly available data sets are including precision geographic data for analysis. Our own work has

explored decedent data published from open data portals for use in precision public health [14,33]. Point obfuscation can naturally shift a point into a different census-designated region; the regional differences before and after shifting highlight significantly different socioeconomic demographics. This is a natural consequence of moving a point and is not a weakness of the techniques themselves. We chose poverty as an example demographic due to its popularity in public health research; we also wish to explore the results of linking other census-level demographics. As future work, we will evaluate other techniques of point obfuscation and explore how these techniques may differ from those presented here. We show that it is possible to enhance point obfuscation by constraining where the new point may be placed; this ensures that the original point is obfuscated in a way that will not impact analyses depending upon the linkage to external region-based data.

Acknowledgments

This project is fully supported by the US Centers for Disease Control and Prevention of the US Department of Health and Human Services as part of grant 1R01CE003360-01-00. The contents are those of the author(s) and do not necessarily represent the official views of, nor an endorsement by, the US Centers for Disease Control and Prevention, the US Department of Health and Human Services, or the US government.

Conflicts of Interest

None declared.

References

1. Wang F. Why public health needs GIS: a methodological overview. *Ann GIS*. 2020;26(1):1-12. [FREE Full text] [doi: [10.1080/19475683.2019.1702099](https://doi.org/10.1080/19475683.2019.1702099)] [Medline: [32547679](https://pubmed.ncbi.nlm.nih.gov/32547679/)]
2. Pearson J, Jacobson C, Ugochukwu N, Asare E, Kan K, Pace N, et al. Geospatial analysis of patients' social determinants of health for health systems science and disparity research. *Int Anesthesiol Clin*. Jan 01, 2023;61(1):49-62. [FREE Full text] [doi: [10.1097/AIA.0000000000000389](https://doi.org/10.1097/AIA.0000000000000389)] [Medline: [36480649](https://pubmed.ncbi.nlm.nih.gov/36480649/)]
3. Dally D, Amith M, Mauldin RL, Thomas L, Dang Y, Tao C. A semantic approach to describe social and economic characteristics that impact health outcomes (social determinants of health): ontology development study. *Online J Public Health Inform*. Mar 13, 2024;16:e52845. [FREE Full text] [doi: [10.2196/52845](https://doi.org/10.2196/52845)] [Medline: [38477963](https://pubmed.ncbi.nlm.nih.gov/38477963/)]
4. Hussein M, Diez Roux AV, Field RI. Neighborhood socioeconomic status and primary health care: usual points of access and temporal trends in a major US urban area. *J Urban Health*. Dec 2016;93(6):1027-1045. [FREE Full text] [doi: [10.1007/s11524-016-0085-2](https://doi.org/10.1007/s11524-016-0085-2)] [Medline: [27718048](https://pubmed.ncbi.nlm.nih.gov/27718048/)]
5. Mohnen SM, Schneider S, Droomers M. Neighborhood characteristics as determinants of healthcare utilization - a theoretical model. *Health Econ Rev*. Mar 06, 2019;9(1):7. [FREE Full text] [doi: [10.1186/s13561-019-0226-x](https://doi.org/10.1186/s13561-019-0226-x)] [Medline: [30840211](https://pubmed.ncbi.nlm.nih.gov/30840211/)]
6. Awuor L, Melles S. The influence of environmental and health indicators on premature mortality: An empirical analysis of the city of Toronto's 140 neighborhoods. *Health Place*. Jul 2019;58:102155. [doi: [10.1016/j.healthplace.2019.102155](https://doi.org/10.1016/j.healthplace.2019.102155)] [Medline: [31252289](https://pubmed.ncbi.nlm.nih.gov/31252289/)]
7. Lakhani CM, Tierney BT, Manrai AK, Yang J, Visscher PM, Patel CJ. Repurposing large health insurance claims data to estimate genetic and environmental contributions in 560 phenotypes. *Nat Genet*. Feb 2019;51(2):327-334. [FREE Full text] [doi: [10.1038/s41588-018-0313-7](https://doi.org/10.1038/s41588-018-0313-7)] [Medline: [30643253](https://pubmed.ncbi.nlm.nih.gov/30643253/)]
8. Elkbuli A, Sanchez C, Boneva D, Hai S, McKenney M. Zip code-targeted injury prevention community outreach initiatives: zip code is as important as genetic code. *Am Surg*. Nov 01, 2019;85(11):e546-e548. [Medline: [31775988](https://pubmed.ncbi.nlm.nih.gov/31775988/)]
9. Graham GN. Why your zip code matters more than your genetic code: promoting healthy outcomes from mother to child. *Breastfeed Med*. Oct 2016;11:396-397. [doi: [10.1089/bfm.2016.0113](https://doi.org/10.1089/bfm.2016.0113)] [Medline: [27513279](https://pubmed.ncbi.nlm.nih.gov/27513279/)]
10. Place and health. Agency for Toxic Substances and Disease Registry. URL: <https://www.atsdr.cdc.gov/placeandhealth/index.html> [accessed 2023-11-06]
11. Harris D. Geographic information systems as data sharing infrastructure for clinical data warehouses. *J Soc Clin Data Manag*. Nov 8, 2023;3(4):1-8. [FREE Full text] [doi: [10.47912/jscdm.240](https://doi.org/10.47912/jscdm.240)]

12. Blatt A. Geospatial data mining and knowledge discovery. In: Health, Science, and Place. Cham, Switzerland. Springer; 2015:77-87.
13. Delcher C, Horne N, McDonnell C, Bae J, Surratt H. Overdose Detection Mapping Application Program expansion evaluation—A qualitative study. *Crim Public Policy*. May 10, 2023;22(3):491-516. [doi: [10.1111/1745-9133.12628](https://doi.org/10.1111/1745-9133.12628)]
14. Delcher C, Harris DR, Anthony N, Mir M. Opioid overdoses increase at home during the COVID-19 stay-at-home order period in Cook County, Illinois. *AJPM Focus*. Sep 2022;1(1):100007. [FREE Full text] [doi: [10.1016/j.focus.2022.100007](https://doi.org/10.1016/j.focus.2022.100007)] [Medline: [36942018](https://pubmed.ncbi.nlm.nih.gov/36942018/)]
15. Shrestha S, Lindstrom MR, Harris D, Rock P, Srinivasan S, Pustz JC, et al. Spatial access to buprenorphine-waivered prescribers in the HEALing communities study: enhanced 2-step floating catchment area analyses in Massachusetts, Ohio, and Kentucky. *J Subst Use Addict Treat*. Jul 2023;150:209077. [FREE Full text] [doi: [10.1016/j.josat.2023.209077](https://doi.org/10.1016/j.josat.2023.209077)] [Medline: [37211155](https://pubmed.ncbi.nlm.nih.gov/37211155/)]
16. Zandbergen PA. Ensuring confidentiality of geocoded health data: assessing geographic masking strategies for individual-level data. *Adv Med*. 2014;2014:567049. [FREE Full text] [doi: [10.1155/2014/567049](https://doi.org/10.1155/2014/567049)] [Medline: [26556417](https://pubmed.ncbi.nlm.nih.gov/26556417/)]
17. McKenzie G, Keßler C, Andris C. Geospatial privacy and security. *J Spat Inf Sci*. Dec 26, 2019;2019(19):53-55. [doi: [10.5311/josis.2019.19.608](https://doi.org/10.5311/josis.2019.19.608)]
18. Ghinita G. Privacy for Location-based Services. Cham, Switzerland. Springer; Apr 30, 2013:1-85.
19. Rodgers SE, Demmler JC, Dsilva R, Lyons RA. Protecting health data privacy while using residence-based environment and demographic data. *Health Place*. Mar 2012;18(2):209-217. [doi: [10.1016/j.healthplace.2011.09.006](https://doi.org/10.1016/j.healthplace.2011.09.006)] [Medline: [21996431](https://pubmed.ncbi.nlm.nih.gov/21996431/)]
20. Gedik B, Liu L. Location privacy in mobile systems: A personalized anonymization model. In: 25th IEEE International Conference on Distributed Computing Systems (ICDCS'05). New York, NY. IEEE; 2005.
21. Hampton KH, Fitch MK, Allshouse WB, Doherty IA, Gesink DC, Leone PA, et al. Mapping health data: improved privacy protection with donut method geomasking. *Am J Epidemiol*. Nov 01, 2010;172(9):1062-1069. [FREE Full text] [doi: [10.1093/aje/kwq248](https://doi.org/10.1093/aje/kwq248)] [Medline: [20817785](https://pubmed.ncbi.nlm.nih.gov/20817785/)]
22. Duckham M, Kulik L. A formal model of obfuscation and negotiation for location privacy. In: Gellersen HW, Want R, Schmidt A, editors. *Pervasive Computing*. Berlin, Germany. Springer; 2005:152-170.
23. Wightman P, Coronell W, Jabba D, Jimeno M, Labrador M. Evaluation of location obfuscation techniques for privacy in location based information systems. In: 2011 IEEE Third Latin-American Conference on Communications. New York, NY. IEEE; 2011:1-6. [doi: [10.1109/LatinCOM.2011.6107399](https://doi.org/10.1109/LatinCOM.2011.6107399)]
24. Wightman P, Zurbarán M, Zurek E, Salazar A, Jabba D, Jimeno M. Theta-rand: random noise-based location obfuscation based on circle sectors. In: 2013 IEEE Symposium on Industrial Electronics & Applications. New York, NY. IEEE; 2013:100-104. [doi: [10.1109/ISIEA.2013.6738976](https://doi.org/10.1109/ISIEA.2013.6738976)]
25. Wightman P, Zurbarán M, Santander A. High variability geographical obfuscation for location privacy. In: 2013 47th International Carnahan Conference on Security Technology (ICCST). New York, NY. IEEE; 2013:1-6. [doi: [10.1109/CCST.2013.6922079](https://doi.org/10.1109/CCST.2013.6922079)]
26. Bissette JM, Stover JA, Newman LM, Delcher PC, Bernstein KT, Matthews L. Assessment of geographic information systems and data confidentiality guidelines in STD programs. *Public Health Rep*. 2009;124 Suppl 2(Suppl 2):58-64. [FREE Full text] [doi: [10.1177/00333549091240S209](https://doi.org/10.1177/00333549091240S209)] [Medline: [27382655](https://pubmed.ncbi.nlm.nih.gov/27382655/)]
27. Harris DR. Leveraging differential privacy in geospatial analyses of standardized healthcare data. *Proc IEEE Int Conf Big Data*. Dec 2020;2020:3119-3122. [FREE Full text] [doi: [10.1109/bigdata50022.2020.9378390](https://doi.org/10.1109/bigdata50022.2020.9378390)] [Medline: [35253022](https://pubmed.ncbi.nlm.nih.gov/35253022/)]
28. Harris DR, Delcher C. bench4gis: benchmarking privacy-aware geocoding with open big data. *Proc IEEE Int Conf Big Data*. Dec 2019;2019:4067-4070. [FREE Full text] [doi: [10.1109/bigdata47090.2019.9006234](https://doi.org/10.1109/bigdata47090.2019.9006234)] [Medline: [32185372](https://pubmed.ncbi.nlm.nih.gov/32185372/)]
29. Harris D. Harris/pinwheelmod. BitBucket. URL: <https://bitbucket.org/harris/pinwheelmod> [accessed 2024-05-06]
30. Glossary. United States Census Bureau. URL: <https://www.census.gov/programs-surveys/geography/about/glossary.html> [accessed 2023-11-06]
31. Rosenbloom ST, Carroll RJ, Warner JL, Matheny ME, Denny JC. Representing knowledge consistently across health systems. *Yearb Med Inform*. Aug 2017;26(1):139-147. [FREE Full text] [doi: [10.15265/IY-2017-018](https://doi.org/10.15265/IY-2017-018)] [Medline: [29063555](https://pubmed.ncbi.nlm.nih.gov/29063555/)]
32. Delcher C, Anthony N, Mir M. Xylazine-involved fatal overdoses and localized geographic clustering: Cook County, IL, 2019-2022. *Drug Alcohol Depend*. Aug 01, 2023;249:110833. [doi: [10.1016/j.drugalcdep.2023.110833](https://doi.org/10.1016/j.drugalcdep.2023.110833)] [Medline: [37352735](https://pubmed.ncbi.nlm.nih.gov/37352735/)]
33. Harris DR, Anthony N, Mir M, Delcher C. geoPIPE: geospatial pipeline for enhancing open data for substance use disorders research. *AMIA Annu Symp Proc*. 2022;2022:522-531. [FREE Full text] [Medline: [37128463](https://pubmed.ncbi.nlm.nih.gov/37128463/)]
34. Geographic Areas Reference Manual. United States Census Bureau. 1994. URL: <https://www2.census.gov/geo/pdfs/reference/GARM/GARMcont.pdf> [accessed 2024-05-15]
35. Pickett KE, Pearl M. Multilevel analyses of neighbourhood socioeconomic context and health outcomes: a critical review. *J Epidemiol Community Health*. Feb 2001;55(2):111-122. [FREE Full text] [doi: [10.1136/jech.55.2.111](https://doi.org/10.1136/jech.55.2.111)] [Medline: [11154250](https://pubmed.ncbi.nlm.nih.gov/11154250/)]
36. Diez Roux AV, Mair C. Neighborhoods and health. *Ann N Y Acad Sci*. Feb 2010;1186(1):125-145. [FREE Full text] [doi: [10.1111/j.1749-6632.2009.05333.x](https://doi.org/10.1111/j.1749-6632.2009.05333.x)] [Medline: [20201871](https://pubmed.ncbi.nlm.nih.gov/20201871/)]
37. Maroko AR, Nash D, Pavilonis BT. COVID-19 and inequity: a comparative spatial analysis of New York City and Chicago hot spots. *J Urban Health*. Aug 2020;97(4):461-470. [FREE Full text] [doi: [10.1007/s11524-020-00468-0](https://doi.org/10.1007/s11524-020-00468-0)] [Medline: [32691212](https://pubmed.ncbi.nlm.nih.gov/32691212/)]

38. Noble D, Smith D, Mathur R, Robson J, Greenhalgh T. Feasibility study of geospatial mapping of chronic disease risk to inform public health commissioning. *BMJ Open*. 2012;2(1):e000711. [FREE Full text] [doi: [10.1136/bmjopen-2011-000711](https://doi.org/10.1136/bmjopen-2011-000711)] [Medline: [22337817](https://pubmed.ncbi.nlm.nih.gov/22337817/)]
39. Kurani SS, McCoy RG, Lampman MA, Doubeni CA, Finney Rutten LJ, Inselman JW, et al. Association of neighborhood measures of social determinants of health with breast, cervical, and colorectal cancer screening rates in the US Midwest. *JAMA Netw Open*. Mar 02, 2020;3(3):e200618. [FREE Full text] [doi: [10.1001/jamanetworkopen.2020.0618](https://doi.org/10.1001/jamanetworkopen.2020.0618)] [Medline: [32150271](https://pubmed.ncbi.nlm.nih.gov/32150271/)]
40. ArcGIS Pro Resources | Tutorials, Documentation, Videos and More. Esri. URL: <https://www.esri.com/en-us/arcgis/products/arcgis-pro/resources> [accessed 2023-11-06]
41. Poverty in the United States: explore the map. Center for American Progress. URL: <https://www.americanprogress.org/data-view/poverty-data/poverty-data-map-tool/> [accessed 2023-09-01]
42. McDonald JF. The deconcentration of poverty in Chicago: 1990-2000. *Urban Stud*. Jul 02, 2016;41(11):2119-2137. [doi: [10.1080/0042098042000268375](https://doi.org/10.1080/0042098042000268375)]
43. Hwang J, Sampson RJ. Divergent pathways of gentrification: racial inequality and the social order of renewal in Chicago neighborhoods. *Am Sociol Rev*. Jun 12, 2014;79(4):726-751. [doi: [10.1177/0003122414535774](https://doi.org/10.1177/0003122414535774)]
44. Henry KA, Sherman R, Farber S, Cockburn M, Goldberg DW, Stroup AM. The joint effects of census tract poverty and geographic access on late-stage breast cancer diagnosis in 10 US States. *Health Place*. May 2013;21:110-121. [doi: [10.1016/j.healthplace.2013.01.007](https://doi.org/10.1016/j.healthplace.2013.01.007)] [Medline: [23454732](https://pubmed.ncbi.nlm.nih.gov/23454732/)]
45. Wightman P, Zurbarán M. An initial evaluation of the impact of location obfuscation mechanisms on geospatial analysis. In: Ukkusuri S, Yang C, editors. *Transportation Analytics in the Era of Big Data*. Cham, Switzerland. Springer; 2019:153-180.

Abbreviations

ACS: American Community Survey

CCMEO: Cook County Medical Examiner's Office

GIS: geographic information system

HIPAA: Health Insurance Portability and Accountability Act

ODMAP: Overdose Detection Mapping Application Program

OMOP: Observational Medical Outcomes Partnership

Edited by E Mensah; submitted 30.11.23; peer-reviewed by L Zhang, F Wang; comments to author 31.01.24; revised version received 28.03.24; accepted 03.04.24; published 21.05.24

Please cite as:

Harris D, Delcher C

Geospatial Imprecision With Constraints for Precision Public Health: Algorithm Development and Validation

Online J Public Health Inform 2024;16:e54958

URL: <https://ojphi.jmir.org/2024/1/e54958>

doi: [10.2196/54958](https://doi.org/10.2196/54958)

PMID:

©Daniel Harris, Chris Delcher. Originally published in the Online Journal of Public Health Informatics (<https://ojphi.jmir.org/>), 21.05.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Online Journal of Public Health Informatics, is properly cited. The complete bibliographic information, a link to the original publication on <https://ojphi.jmir.org/>, as well as this copyright and license information must be included.