A Generalizable Evaluated Approach, Applying Advanced Geospatial Statistical Methods, to Identify High Lead Exposure Locations at Census Tract Scale: Michigan Case Study

Jianping Xue,1 Valerie Zartarian,1 Rogelio Tornero-Velez,1,© Lindsay W. Stanek,1© Antonios Poulakos,1,© Alan Walts,1© Kathy Triantafilou,1© Maryann Suero,1© and Nicholas Grokhowsky1©

1U.S. Environmental Protection Agency (EPA), Office of Research and Development, Research Triangle Park, North Carolina, USA
2LinTech Global, Inc., Boston, Massachusetts, USA
3U.S. EPA, Region 5, Chicago, Illinois, USA
4Oak Ridge Institute for Science and Education, Research Triangle Park, North Carolina, USA

**BACKGROUND:** Despite great progress in reducing environmental lead (Pb) levels, many children in the United States are still being exposed.

**OBJECTIVE:** Our aim was to develop a generalizable approach for systematically identifying, verifying, and analyzing locations with high prevalence of children’s elevated blood Pb levels (EBLLs) and to assess available Pb models/indices as surrogates, using a Michigan case study.

**METHODS:** We obtained ~1.9 million BLL test results of children <6 years of age in Michigan from 2006–2016; we then evaluated them for data representativeness by comparing two percentage EBLL (%EBLL) rates (number of children tested with EBLL divided by both number of children tested and total population). We analyzed %EBLLs across census tracts over three time periods and between two EBLL reference values (≥5 vs. ≥10 μg/dL) to evaluate consistency. Locations with high %EBLLs were identified by a top 20 percentile method and a Getis-Ord Gi*

**RESULTS:** Analyses of 2014–2016 %EBLL data identified 11 Michigan locations via cluster analysis and 80 additional locations via the top 20 percentile method and their associated census tracts. Data representativeness and consistency were supported by a 0.93 correlation coefficient between the two EBLL rates over 11 y, and a Kappa score of ~0.8 of %EBLL hotspots across the time periods (2014–2016) and reference values. Many EBLL hotspot locations converge with current Pb exposure models/indices; others diverge, suggesting additional Pb sources for targeted interventions.

**DISCUSSION:** This analysis confirmed known Pb hotspot locations and revealed new ones at a finer geographic resolution than previously available, using advanced geospatial statistical methods and mapping/visualization. It also assessed the utility of surrogates in the absence of blood Pb data. This approach could be applied to other states to inform Pb mitigation and prevention efforts. [https://doi.org/10.1289/EHP9705](https://doi.org/10.1289/EHP9705)

**Introduction**

Despite great progress over the last several decades, childhood lead (Pb) exposure remains a public health challenge, and “substantial individual- and community-level disparities persist.”1,2 As summarized in the “Federal Action Plan to Reduce Childhood Lead Exposures and Associated Health Impacts”1:

> Lead exposure to children can result from multiple sources and can cause irreversible and lifelong health effects. No safe blood Pb level in children has been identified. Even low levels of Pb in blood have been shown to affect intelligence quotient (IQ), ability to pay attention and academic achievement.

The Action Plan calls for data, maps, and mapping tools to identify high exposure locations and disparities for prioritization efforts to reduce children’s blood Pb levels (BLLs).

Local BLL and environmental data are often difficult to obtain, incomplete, or unavailable and state-to-state testing, data collection, and reporting vary.3–7 This is not a new issue.8 Thus, different groups have developed surrogate predictive models (e.g., regression models) and indices to identify disproportionately impacted places for action.3,6,9–16 Predictive modeling remains challenging at both national and local scales. Generalizable approaches with state-specific models and robust data at the census tract or finer-scale level are needed.

By focusing on a state with extensive, robust BLL data, we can improve mapping resolution and analysis of available measured BLLs in this state while also providing a generalizable approach that could be applied in other states and geographic regions to identify and address high Pb exposure locations, including using available Pb indices and predictive models. Once locations with a high prevalence of elevated BLLs (EBLLs; defined below) are identified, determining the key Pb exposure factors and indicators at those locations (e.g., old housing, which may have Pb paint or Pb service lines for drinking water, and other sources of Pb exposures; sociodemographics, such as poverty and minority variables; or environmental sources other than from old housing) based on exposure pathway/source apportionment modeling can inform effective mitigation and prevention targeting efforts.7,10

This case study focuses on Michigan, a state with an extensive, robust BLL data set and with the federal–state partnerships in place to facilitate collaboration and ground truthing during this study. We also evaluated and verified our analyses using published data from state reports17 and published BLL data maps for Michigan12 and the city of Flint where a 2014 drinking water-related public health crisis occurred.18,19 Our analysis of BLL data and convergence of available Pb indices in a state with robust BLL data can also inform Pb “hotspot” identification in states with less robust BLL data.

**Goals and Objectives**

This paper presents a generalizable systematic approach, using new applications of geospatial and statistical methods and models with available BLL and other data at the census tract level, to inform Pb exposure reduction efforts—with a focus on Michigan.
as an initial case study. We present an innovative data analysis to identify locations in Michigan with high prevalence of EBLL and initial analyses of key select Pb exposure indicators, namely, old housing and sociodemographic factors. Environmental sources of exposure are of interest for planning, but analysis related to these sources is outside the scope of this article.

Specific research objectives in this study were to:

- Statistically evaluate and analyze Michigan BLL measurements over census tracts and time.
- Identify high percentage EBLL (%EBLL) census tracts, and locations near or containing them, using two different statistical methods (top 20 percentiles and an advanced geospatial statistical cluster analysis).
- Analyze how well the highest %EBLL locations are explained by Pb exposure models/indices.
- Verify identified Michigan hotspots based on previous studies.

Methods

We processed and evaluated available Michigan BLL data, used those data to identify high %EBLL locations and statistical clusters, and analyzed changes over time. We then compared three available Pb exposure indices/models (based on age of housing and sociodemographic variables) as potential surrogates when such robust BLL data are absent. We also developed and applied an initial approach to distinguish locations with a high prevalence of EBLLs ("high %EBLL locations") associated with old housing and sociodemographics from those associated with other exposure factors and indicators.

BLL Data Summary, Processing, and Evaluation

The institutional review boards at the Michigan Department of Health and Human Services (MDHHS) and the University of North Carolina at Chapel Hill, North Carolina, reviewed and approved this study. Through a data use agreement with the MDHHS, we obtained ~1.9 million BLL data points (~50% venous and ~50% capillary samples; with 31,600 un-identified) from the MDHHS Childhood Lead Poisoning Prevention Program (CLPPP) for the years 2006–2016. We geocoded test results data for children <6 years of age using ArcGIS (version 10.6.1; ESRI Inc.) for all 1,076,964 address data points provided and matched 1,016,564 addresses for a 94.39% using ArcGIS (version 10.6.1; ESRI Inc.) for all 1,076,964 address samples; with 31,600 un-identifiable factors and indicators.

The <6-years-of-age group was selected to be consistent with the U.S. EPA’s age groupings guidance. We applied quality assurance/quality control to all aspects of the data analysis and mapping by having multiple coauthors cross-check work. We applied the following exclusion criteria to 2006–2016 data for each census tract: total population of <50 children 0 to <6 years of age with BLL tests; <10 children tested in a given year; and not containing data for greater than half of the total years assessed.

To identify hotspots, we randomly selected one measure regardless of capillary or venous sampling in Michigan if there were multiple measurements within 1 y for a given person. We conducted sensitivity analyses on the selection of one sample from multiple samples per person per year using three methods: a) random selection (the main method for this paper), b) selection of the highest value, and c) selection of the first sample of the given year.

There are 2,813 census tracts in Michigan (according to the 2010 Census). We removed 39 unpopulated census tracts, for a total of 2,774 in our geospatial analyses. After applying geocoding and test exclusion criteria, we analyzed 2,411 census tracts for 2006–2016 and 2,401 census tracts for 2014–2016 BLL data (n = 363 and n = 373 excluded, respectively). We combined venous and capillary BLL data to increase the total sample size within census tracts. For 2006–2016, the total sample size of BLL test results for 0- to <6-y-olds was 1,930,943 (N = 958,215 for venous and N = 972,28 for capillary samples, after deleting duplicates). For these data, there is 0.58 correlation between venous and capillary samples for the same person and same year.

For the analyses, we defined an EBLL as one above the 2012 Centers for Disease Control and Prevention (CDC) blood Pb reference value of 5 μg/dL, which is the 97.5th percentile of BLL of 1- to 5-y-old children in the United States based on the 2007–2008 and 2009–2010 National Health and Nutrition Examination Surveys (NHANES). We analyzed the Michigan BLL data set to evaluate its robustness based on sample quantity and the degree to which samples were statistically representative of the general population of 0- to <6-y-olds in Michigan. Concretely, "representativeness" was evaluated by comparing two defined % EBLL metrics or rates, shown in Equations 1 and 2, with the same numerators and different denominators.

\[
\text{Representativeness, by our definition, does not necessarily capture specific sociodemographic differences within a census tract because we did not have those data. Comparing these rates is a check that the BLL data used for the hotspot analyses statistically reflect sampling proportional to the population, providing confidence in the analyses.}
\]

\[
\text{Michigan uses a targeted testing approach for blood Pb surveillance that focuses on children at greater risk for Pb exposure. We note that Michigan CLPPP EBLL rate calculations are in accordance with Equation 1. To assess BLL data representativeness for our analysis, we calculated the correlation coefficient between the two EBLL rates over different time periods.}
\]

\[
\text{We calculated correlation coefficients between EBLL measures by blood sample type to determine whether it was statistically appropriate (based on the 0.58 correlation coefficient) to combine venous and capillary results for analysis. Note that the threshold for statistical significance used in this paper is } p < 0.01.
\]

\[
\text{To assess the consistency of the %EBLL statistical hotspots, we calculated Cohen’s Kappa statistics of hotspot locations over different time periods (2006–2013 and 2011–2013 vs. 2014–2016); between EBLL reference values (≥5 and ≥10 μg/dL); and between different age groups (5-, 6-, and 7-y-olds for a}
\]
Identifying Locations with High Prevalence of EBLLs

We used a top 20 percentile method and an advanced geospatial cluster analysis method to identify high %EBLL locations for census tracts and “reference locations” using 2014–2016 BLL data. We define a reference location in our analysis as the highest population city near or containing a top 20 percentile census tract or more than one geospatial cluster analysis statistical hotspot census tract.

The top 20 percentile method identified 80th–100th percentile census tracts (i.e., the census tracts with BLL data, considering exclusion criteria) for %EBLLs in Michigan (2014–2016; ≥5 μg/dL). The 80th–100th percentile metric was based on the U.S. EPA’s Office of Environmental Justice EJSCREEN 2017 technical documentation.15

The geospatial cluster analysis method, Getis-Ord Gi*23 involved applying the ArcGIS Hot Spot Analysis tool [ESRI ArcGIS (version 10.6.1) and ArcGIS Pro (version 2.8)]24 to census tract exceedance rates in Michigan from 2014 to 2016 (using an EBLL reference value of ≥5 μg/dL). We conducted additional analyses to compare exceedance rates for this time period using the latest CDC EBLL reference value of ≥3.5 μg/dL.25 Getis-Ord Gi* provided a range of statistical confidence levels when evaluating geospatial patterns of EBLL prevalence; hotspots are defined as locations where %EBLLs are clustered statistically. We queried data to only include census tracts that the tool determined to have confidence intervals (CIs) ≥95% for the Getis-Ord Gi* statistical testing significance level and then identified reference locations for the EBLL clusters.

The ArcGIS Hot Spot Analysis tool is a spatial cluster analysis function that calculates the Getis-Ord Gi* statistic for each feature (i.e., census tract) in a data set. This tool looks at each feature within the context of neighboring features. In this analysis, to be identified as a statistically significant high %EBLL area, a given census tract will have a high %EBLL value and will also be adjacent to other census tracts with high %EBLL values.24 Basically, the standardized Gi* statistic compares the weighted average of values of the outcome of interest in the neighborhood around the focus area to the overall average of values of the outcome of interest. The Gi* follows the Z distribution, and areas with high Gi* with low p-values indicate hotspots. More detail can be found in Ord and Getis.25

After locations were detected by the top 20 percentile and geospatial cluster analysis methods, we assigned each location or cluster a reference location by comparing and spatially joining the spatial cluster analysis methods, we assigned each location or cluster a reference location by comparing and spatially joining the resulting data to the U.S. Census Bureau 2010 “Places” file.26 using the “Select by Location” function in ESRI ArcGIS (version 10.6.1). We used the U.S. Census 2010 decennial file year;27 given that it fell in the middle of the original time frame of the Michigan BLL data set (2006–2016) and because Michigan city boundaries for the “Places” data set did not change between 2010 and 2016. In the rare case of reference locations that were completely contained/embedded within a larger city (e.g., the cities of Hamtramck and Highland Park, both embedded in the city of Detroit), the larger city was selected as the “reference location.” Associating census tracts with reference locations was not a straightforward spatial assessment, especially in rural areas/townships, because not all census tracts are associated with a unique city or township.

To ground truth and verify locations identified in our Michigan analyses, we compared results with the Michigan MDHHS CLPPP annual reports of blood Pb testing from 2010–2017, which included “targeted communities.”17 To evaluate the 2014–2016 high %EBLL (exceedance rate) locations, we mapped the nine targeted communities in the MDHHS 2016 report,28 added a 1-mi (1.6 kilometers) buffer, then overlaid the resulting nine buffers/circles with results of Getis-Ord Gi* of 2014–2016 BLLs and the EJSCREEN Pb Paint EJ Index. In addition, we conducted a visual comparison with Reuters ZIP code mapping and, specifically for Flint, a visual comparison with the ward mapping shown in Figure 2 of Hanna-Attisha et al.18

For both the exceedance rate (Equation 1) and population-adjusted rate (Equation 2), we conducted a time trend regression analysis of %EBLLs from 2006–2016, regressing the log-transformation of the mean rate on years elapsed since 2006. We also analyzed and visualized %EBLL changes over time by census tracts.

Initial Exploration of Exposure-Related Pb Indicators

Old housing (defined as pre-1960) is prevalent throughout Michigan in both rural and urban areas. After identifying high % EBLL (exceedance rate) census tracts, we used three available Pb exposure models/indices for analyses of old housing data and sociodemographic variables with both the top 20 percentile method and the geospatial cluster analysis method. The three Pb exposure models/indices were the U.S. EPA’s EJSCREEN 2017 Pb Paint EJ Index,15 modeled BLLs of Schultz et al.,13 and U.S. Department of Housing and Urban Development’s (HUD) Deteriorated Paint Index.9

The U.S. EPA’s EJSCREEN 2017 Pb Paint EJ Index, based on pre-1960 housing and sociodemographic data, was originally developed at the census block group level. We aggregated the data by averaging census block group index values to the census tract. This aggregation was done for data compatibility across index/model data sets and to analyze with BLL data at the census tract scale. EJSCREEN Pb Paint EJ Index values are based on Equation 3.

EJSCREEN Pb Paint EJ Index = ((% low income + % minority)/2) per block group
− ((% low income + % minority)/2) per national population
× (population count per block group)).

This index uses the 2011–2015 American Community Survey (ACS) data (low income and minority are defined on page 20 of the EJSCREEN 2017 Technical Document15). Children’s BLLs are not considered in this index.

Schultz et al. presented a multiple regression model approach to predict 1- and 2-year-old children’s geometric mean (GM) BLLs at the census tract level across the United States; they evaluated their model against measured BLL data from three states (Massachusetts, Michigan, Texas).13 We generated modeled 2015 GM venous BLL predictions for a 2-year-old child for all Michigan census tracts by implementing the Schultz et al. model13 in the R programming language. Schultz et al. originally used the 2005–2009 ACS information on three key variables: percentage pre-1960 homes, percentage below the poverty line, and percentage non-Hispanic African Americans13; we applied the model using 2012–2016 ACS data to better overlap with EJSCREEN 2017 and to generate modeled BLL values for a year (i.e., 2015) that falls in the middle of our analytical time frame of 2014–2016.

HUD’s Deteriorated Paint Index was developed to help HUD grantees better target home remediation and abatement efforts.
The Deteriorated Paint Index uses microdata from the 2011 American Housing Survey and the 2009–2013 ACS to develop a household-level predicted risk metric that identifies housing units at risk of containing large areas of deteriorated paint. This predicted risk is defined as the mean predicted percentage of occupied pre-1980 housing units at risk of containing deteriorated paint within a given jurisdiction. Results were summarized by state, county, and census tract.9

We analyzed and mapped these three Pb exposure models/indices and %EBLL (exceedance rate) data at census tract resolution using geospatial and traditional statistical functions within ArcGIS (version 10.6.1; ESRI) and SAS (version 9.4; SAS Institute, Inc.). Analyzing the model approaches among each other (using Kappa scores and visual inspection) provided a model-to-model comparison, analyzing them against %EBLL data (using Kappa scores, sensitivity and specificity, and visual inspection) helped assess their usefulness as surrogates in the absence of BLL data and to initially identify key indicators. Visual inspection refers to visual observation of mapped hotspot locations identified in our analyses compared with hotspot maps in the other papers,12,18 which were used to draw conclusions about the consistency of the areas identified.

Census 2010 data from Summary File 127 H2 (rural and urban housing) and P14 (population by age) were joined to the Getis-Ord Gi* hotspot and top 20 percentile data that were produced in this study. Any tract with >50% urban housing was designated as urban, and any tract with >50% rural housing was designated as rural. Using the resulting joined values, ratios for urban-to-rural census tracts and urban-to-rural children 0 to <6 years of age were calculated and categorized for the urban vs. rural analysis presented in the “Results” section.

We merged a selection of environmental justice-related variables from EJSCREEN 2017 (less than high school education, total minority population, low income to less than two times the federal poverty level),15 and table DP05 (non-Hispanic African American population) of the 2011–2015 ACS29 data to the clusters identified by applying the Getis-Ord Gi* based hotspot analysis on Michigan 2014–2016 EBLLs. This merging was done in ArcGIS via geospatial join by census tract number. The 2011–2015 ACS was chosen to exactly match with the underlying data presented in EJSCREEN 2017. Using the resulting joined values, population totals, population percentages, and ratios for hotspot vs. non-hotspot were calculated for the environmental justice variables and Getis-Ord Gi* hotspot cluster analysis presented in the “Results” section.

To demonstrate and quantify robustness of the BLL and exposure data sets, and to corroborate analytical and visual findings, we ranked the various measures (predictions or observations) from their lowest to highest values, and used a threshold (i.e., 80th percentile; 90th percentile for a sensitivity analysis) to identify the top 20 or 10 for sensitivity analysis) percentage highest values. This corresponded to a set of census tracts indicated as the most severely affected areas, by a particular measure. To compare two measures, we generated a 2 x 2 table tally of the inclusions (1) and exclusions (0), and summarized findings using sensitivity, specificity (with measured BLL used as the gold standard/true value), and Kappa statistics. This procedure, which enabled comparison of measures with disparate units, is referred to in this paper as “data convergence analysis.”

Results

Identifying Statistical Hotspot %EBLL Locations, and Evaluation

There were 36 census tracts in 2014–2016 with %EBLL ≥20% for children 0 to <6 years of age (Table 1). This table reveals the decreasing average percentage of children with EBLLs over time (14.8% for 2006–2007 to 4.1% in 2014–2016). As a point of reference, 2.5% is the national exceedance rate for the United States, based on the CDC 2012 reference value of ≥5 µg/dL. We also used the new CDC blood Pb reference value of ≥3.5 µg/dL and found that the pattern of hotspot changes over time is similar, but overall exceedance rates are higher, as expected with the lower reference value [Figure S1(b)].

Cluster analysis with Getis-Ord Gi*.

The Getis-Ord Gi* geospatial cluster analysis identified 306 census tracts across 11 “reference locations.” Figure 1 and Table 2 illustrate %EBLL clusters identified with the Getis-Ord Gi* geospatial cluster analysis using 2014–2016 data. Locations with the highest %EBLLs (from 7.1% to 12.4%, “random” selection methodology—i.e., the main method of our analysis) had varying numbers of census tracts (from 4 in Flint and Kalamazoo to 190 in Detroit) and numbers of children from 643 to 39,323. More than 74,000 children lived in the identified clusters (~11% of the population of 0 to <6 years of age). Table 2 also shows that the results using two additional sampling selection methods (i.e., highest and first) are consistent with the random selection results.

High %EBLL locations were consistent across time periods (2006–2013 and 2014–2016; 2011–2013 and 2014–2016) (Figure 2). There was strong statistical agreement between periods, with Kappa = 0.83 between 2006–2013 and 2014–2016 and Kappa = 0.84 between 2011–2013 and 2014–2016. The 11 geospatial cluster reference locations identified in Figure 1 were also identified over time (≥95% confidence), but with some minor differences in specific associated census tracts making up those clusters.

High %EBLL locations were also consistent between ≥5 µg/dL and ≥10 µg/dL BLL reference values across the 2014–2016 time period (Kappa = 0.78) (Figure 3). Locations identified were consistent between the two EBLL reference values, but with some census tract differences within the clusters and one additional cluster or reference location (Laurium) identified with the ≥10 µg/dL reference value. The Kappa value of the %EBLL statistical hotspots for 2014–2016 in Michigan between 0- to 6- and 0- to 7-y-olds was 0.95, and between 0- to 5- and 0- to 6-y-olds it was 0.99. We compared Pb hotspots with EBLL reference values of ≥3.5 µg/dL vs. ≥5 µg/dL, resulting in high agreement (Kappa = 0.94) [Figure S1(a)].

| Time period | Census tracts (%) |
|-------------|------------------|
| 2006–2007   | 2010–2013        | 2014–2016        |
| Percentage category for exceedance rate | Percentage category for exceedance rate | Percentage category for exceedance rate |
| ≥5 µg/dL (%) | ≥5 µg/dL (%)     | ≥5 µg/dL (%)     |
| 0–5         | 765               | 1,394            | 1,848             | 1,901             |
| >5–10       | 665               | 564              | 355               | 315               |
| >10–20      | 503               | 274              | 152               | 149               |
| >20–40      | 235               | 144              | 48                | 36                |
| ≥40–70      | 76                | 7                | 2                 | 0                 |
| Total       | 2,244             | 2,383            | 2,405             | 2,401             |

Note: Exceedance rate is the exceedance rate of EBLLs (≥5 µg/dL) in children 0 to <6 years of age. EBLL, elevated blood lead level.
Comparison of locations against other published analyses. Figure S4 demonstrates the high %EBLL locations identified in Flint, Michigan, using both Getis-Ord Gi* and top 20 percentiles. Eight of the 9 targeted communities in the 2016 MDHHS report intersected with the BLL Getis-Ord Gi*–predicted hotspots (Lansing is the one location missed), and 8 of 9 intersected with the EJSCREEN Pb Paint EJ Index Getis-Ord Gi*–predicted hotspots (Adrian is the one location missed). Of the 11 reference locations selected using the Getis-Ord Gi* method, when comparing against the MDHHS CLPPP annual reports, 6 (Adrian, Detroit, Flint, Grand Rapids, Jackson, and Muskegon) were targeted communities in 2016, 3 (Battle Creek, Kalamazoo, and Saginaw) were targeted communities in previous years, and 2 (Bay City and Ludington) were not identified as targeted communities. Three MDHHS CLPPP–targeted communities were not identified as reference locations by the Getis-Ord Gi* method (Hamtramck, Highland Park, and Lansing). Of the 91 reference locations selected using the top 20 percentile method, 7 of 9 MDHHS CLPPP–targeted communities were identified (Adrian, Detroit, Flint, Grand Rapids, Jackson, Lansing, and Muskegon).

Representativeness of EBLL Data

Our geospatial and statistical analyses show that Michigan BLLs from 2006–2016 were statistically robust and consistent across census tracts, over time, and between two reference values (Figures 1–3). The correlation coefficient was \( r = 0.93 \) between the two EBLL rates over 11 y; Kappa scores of \( \geq 0.8 \) for analysis of EBLLs over different time periods and a Kappa score of 0.78 for analysis of EBLLs \( \geq 5 \mu g/dL \) vs. \( \geq 10 \mu g/dL \) indicate substantial agreement between time periods and EBLL definition levels (Table 3). Dots inside the two reference lines indicate census tracts with representative data, and dots outside the two reference lines will be helpful to identify census tracts to improve sampling for the representativeness [as illustrated in Figure S5(b)].

The correlation coefficient calculated between %EBLL and population-adjusted %EBLL with both venous and capillary BLLs (0.93; \( p < 0.001 \)) was higher than with venous only (0.70; \( p < 0.001 \)) or capillary only (0.81; \( p < 0.001 \)), which supports combining the data. For 2014–2016, \( N = 1,951 \) for venous and \( N = 2,297 \) for capillary samples. There was also good agreement in the %EBLL clusters using venous and capillary BLL data (Figure S6).

EBLL data over time

Our analysis of the BLLs over time indicates data representativeness. Figure S5 illustrates the %EBLL and population-adjusted %EBLL rate (\( \geq 5 \mu g/dL \)) over time for the state of Michigan for children 0 to <6 years of age. Figure S5(a) shows that the two metrics (defined in Equations 1 and 2) track well with each other over time and have a strong correlation of 0.93. Figure S5(b) shows the relationship of the two metrics: Each circle represents one census tract. The overwhelming majority of observations fell within the 95% CI, which demonstrates minor sampling bias in terms of the proportionality to the total population. Figure S5(c) shows that from 2006 to 2016 there was an average 12.5% decrease per year in the exceedance rate and an average 16.4% decrease per year in the population-adjusted rate. The regression analysis revealed a statistically significant time trend (respectively, \( p = 1.73 \times 10^{-5} \) and \( p = 6.52 \times 10^{-6} \)).

Table S1 data show that the %EBLL in Michigan declined greatly from 2006 to 2011 (median 7.7% to 2.4%; 95th percentile 35.9% to 15.8%) and were similar after 2011. As Table 1 reveals, from 2011 to 2016 about 8% of census tracts in Michigan had %
EBLLs >10% (based on ≥5 μg/dL). Figure 4 shows changes in children’s %EBLLs over time by census tracts in Michigan, and Figures S7 and S8 show the percentage of increases and decreases of %EBLLs by census tract.

Analyses of Pb Exposure Models/Indices with %EBLLs

Comparison of Pb exposure models/indices. The census tracts identified by the three exposure models/indices were highly consistent. Kappa values among the three Pb exposure models/indices ranged from 0.83 to 0.88 using Getis-Ord Gi* (Table 3) and from 0.72 to 0.87 using the top 20 percentile method (Table 4), indicating strong statistical convergence between models.

Getis-Ord Gi* geospatial hotspot cluster analysis. High Pb exposure locations identified with the three Pb exposure models/indices, based on old housing and sociodemographics, have moderate statistical agreement (convergence) with the 2014–2016 %EBLLs using the Getis-Ord Gi* geospatial cluster analysis.
Comparison between 2014 and 2016 %EBLLs and the three Pb exposure models/indices showed 75–80%, 88–90%, and 0.54–0.55 for sensitivity, specificity, and Kappa values, respectively (Table 3). Table S2 provides descriptive statistics of EJSCREEN variables, summarized by census tracts in the cluster analysis of exceedance rate of 2014–2016 EBLLs (≥5 μg/dL; children 0 to <6 years of age). Table S3 provides additional environmental justice-related variables geospatially joined to the cluster analysis of exceedance rate of 2014–2016 EBLLs (≥5 μg/dL; children 0 to <6 years of age).

**Top 20 percentile method.** High Pb exposure locations identified with the three Pb exposure models/indices, based on old housing and sociodemographics, also have moderate statistical agreement (convergence) with the 2014–2016 %EBLLs using the top 20 percentile method [i.e., 80th–100th percentiles of census tracts with available Pb exposure models/indices data (Table 4; Figure S9)]. Comparison between 2014–2016 %EBLLs and the three Pb exposure models/indices shows 57–67%, 88–90%, and 0.44–0.55 for sensitivity, specificity, and Kappa values, respectively (Table 4). Furthermore, there were 140 census tracts that had high %EBLLs but did not converge with the three Pb exposure models/indices (Figure S10).

**Rural vs. Urban.** An analysis using urban and rural housing data from the 2010 U.S. Census shows that the three indices/models overwhelmingly indicate urban hotspots with either method, but especially with the geospatial cluster analysis method (Table 5). The HUD index identifies a few more hotspots in rural areas in comparison with the EJSCREEN Pb Paint EJ Index and the Schultz et al. model. If indices are used for identifying and addressing hotspots, more housing and sociodemographic data and other information (e.g., environmental sources) are needed in rural areas.

**Environmental Justice-Related variables and Getis-Ord Gi* geospatial hotspot cluster analysis.** As seen in Table S3, an analysis combining environmental justice-related data, extracted from EJSCREEN 2017 and table DP05 of the 2011–2015 ACS with Getis-Ord Gi* geospatial statistical cluster analysis hotspots identified for the exceedance rate of Michigan 2014–2016 EBLLs, was conducted to demonstrate the relationship between environmental justice variables (non-Hispanic African American, less than high school education, low income, total minority population) and these hotspot data. The total number of census tracts evaluated was 2,401. There were 306 hotspot census tracts and 2,095 non-hotspot census tracts. The ratios were 4.09, 1.97, 1.86, and 2.93 for hotspot...
Table 3. Statistical agreement analyses for Getis-Ord G* geospatial cluster analysis of EJSCREEN 2017 Pb Paint EJ Index, Schultz et al. regression model13 approach BLL data, HUD Deteriorated Paint Index, and exceedance rate of EBLLs (≥5 μg/dL) (census tracts; children 0 to <6 years of age).

| Data comparison | Census tracts (n) | Sensitivity | Specificity | Kappa |
|-----------------|------------------|-------------|-------------|-------|
| Exceedance rate of EBLLs | 2006–2013 vs. 2014–2016 | 2.378 | — | — | 0.83 |
| | 2006–2007 vs. 2014–2016 | 2.236 | — | — | 0.81 |
| | 2008–2010 vs. 2014–2016 | 2.362 | — | — | 0.81 |
| | 2011–2013 vs. 2014–2016 | 2.383 | — | — | 0.84 |
| | ≥5 vs. ≥10 μg/dL (2014–2016) | 2.401 | — | — | 0.78 |
| | 2014–2016 vs. EJSCREEN* | 2.401 | 75 | 90 | 0.55 |
| | 2014–2016 vs. Schultz12 | 2.401 | 80 | 88 | 0.54 |
| | 2014–2016 vs. HUD10 | 2.401 | 76 | 89 | 0.54 |
| | EJSCREEN vs. Schultz | 2.752 | — | — | 0.88 |
| | EJSCREEN vs. HUD | 2.740 | — | — | 0.83 |
| | Schultz vs. HUD | 2.741 | — | — | 0.84 |

Note: Exceedance rate is the exceedance rate of EBLLs (≥5 μg/dL) in children 0 to <6 years of age; —, not applicable; BLL, blood lead level; EBLL, elevated blood lead level; EJ, environmental justice; EPA, Environmental Protection Agency; GM, geometric mean; HUD, U.S. Department of Housing and Urban Development; Pb, lead.
*Cohen’s Kappa agreement statistic: 0.41–0.6, moderate; 0.61–0.8, substantial; 0.81–0.99, near perfect agreement.
13U.S. EPA EJSCREEN 2017 Pb Paint EJ Index.15
12U.S. EPA regression model, 2015 Census BLL GM, 2017 publication.13
10HUD Deteriorated Paint Index.2

Discussion

Advancing the Science

This paper presents a new science-based approach, developed through a federal–state collaboration, for better identifying locations with high prevalence of children’s EBLLs. Identifying locations with the highest %EBLLs, and the contributing sources and exposures of those EBLLs, can assist with targeting and prioritization for Pb exposure risk reduction, prevention, and mitigation efforts as called for in the Federal Lead Action Plan.1

Ideally, robust, consistently measured and reported BLL data at the individual level would be available across all U.S. states to use in identifying locations with high prevalence of EBLLs. Given the challenges in implementing such a program and the significant limitations in BLL data availability and quality, such as privacy concerns and lack of spatial resolution and representativeness,3–5 surrogate Pb exposure models/indices have been developed.6,9,11,13,16,21,30 Differences in variables and data used for housing age, sociodemographics, environmental risk factors, and BLL data (e.g., state reference values for EBLL, age categories used for analysis) can produce different results. State-to-state differences in collection, analysis, and reporting of BLL data can Pb to uncertainty in national-scale analyses and maps to identify high Pb exposure locations.6 Our analysis provided finer (i.e., census tract level) spatial resolution than county or ZIP code. Aggregating individual data to the census tract level allows for more spatially detailed analyses while avoiding data privacy issues.

To advance the science through a specific state case study analysis, we geocoded an extensive Michigan BLL data set from the MDHHS, statistically analyzed the data for representativeness, and identified high %EBLL locations by applying several geospatial and statistical methods. For locations identified, we used three available Pb exposure models/indices that combine old housing and sociodemographic variables for two purposes:

a) to assess their use as surrogates for EBLL locations, and
b) to help identify key indicators of EBLLs to inform Pb reduction efforts at the local scale. The resulting maps and analyses identify areas of potential high Pb exposures.

Key Issues

Use of BLL data. Comparing %EBLL vs. population-adjusted %EBLL [Figure S5(b)] can help identify census tracts where the exceedance rate and population-adjusted rates differ. These census tracts can be investigated to determine why rates differed and how BLL sampling improvements could be considered in the future (in Michigan and other states). In addition, one study using cumulative case ascertainment ratios of reported EBLL counts over expected counts based on NHANES survey data from 39 states participating in CLPPP calculated an ascertainment ratio of 0.77 for Michigan, which was better than the ascertainment ratios of 31 other states and the overall CDC estimate.4 This further suggests robustness of the Michigan data.

Venous test results are more accurate than capillary test results, but they are usually not the first test a child receives; and follow-up testing after capillary screening is not always done. Of a total 1,992,311 test results used, 49.62% samples were venous and 50.34% were capillary. The advantage of an effectively doubled sample size far outweighs the uncertainty added by combining venous and capillary data for our analysis. Our correlation analysis (comparing the two %EBLL rates, with correlation coefficients of 0.93 for combined venous and capillary, 0.7 for venous only, and 0.81 for capillary only), supported determining that it was statistically appropriate to combine venous and capillary BLL data (Figure S6). Although venous blood Pb samples are considered the gold standard,31 one study showed that “capillary blood draws are suitable alternatives to venous blood draws when screening children <6 years of age to determine Pb exposure and provide reasonable estimates at the population level.”32,33 Combining venous and capillary data is also supported by the 2018 report, “Screening for Elevated Blood Lead Levels in Children: A Systematic Review for the U.S. Preventive Services Task Force”34.

Four studies of capillary blood Pb testing demonstrated sensitivity of 87% to 91% and specificity >90% (range 92% to 99%) compared with venous measurement.

Other studies have also used both venous and capillary data.5 Furthermore, because we used large sample sizes and assess BLLs >5 μg/dL, we believe the influence of detection limits and accuracy in venous vs. capillary sampling to be negligible in our analysis.

We used EBLL prevalence rather than actual BLL concentrations to avoid statistical challenges in using laboratory-derived concentrations that are subject to detection limit issues at the low end of the distribution. The Michigan blood Pb surveillance database did not have the capacity to store indicators for tests below analyzer limits of detection until 2018. Consequently, there was no way of determining whether a reported BLL from a given laboratory was a true low value or whether it was a test result below the limit of detection. Treating test results below limit of detection as a true test value will create biased estimates. Once the limit of detection is known, there are statistical methods for adjusting BLLs to account for results below limits of detection.35,36 Still, there is more analytical error in quantifying BLLs at 1 μg/dL than at 5 μg/dL. In a review of 5 of results for target blood Pb values <11 μg/dL for U.S. clinical laboratories participating in CDC’s voluntary Lead and Multi-Element Proficiency (LAMP) quality assurance program, Caldwell et al. observed that 40% of the...
participating laboratories reported a nondetectable result at a target blood Pb value of 1.48 µg/dL compared with 5.5% at a target blood Pb of 4.60 µg/dL. On the other hand, using the EBLL does not require consideration of blood Pb limit of detection, greatly simplifying analyses and enabling greater confidence in results. Analyzing and visualizing EBLL is also consistent with how Michigan and other states routinely report blood Pb surveillance results. We note that the date of geocoded address may not match the date during blood draw, which may result in a small misclassification, but this should not impact the results because of the large sample size and because we used year as the time period.

**Figure 4.** Changes over time for exceedance rate of %EBLLs by census tracts—lower peninsula of Michigan (number of census tracts evaluated: (A) 2006–2007, 2,244; (B) 2008–2010, 2,383; (C) 2011–2013, 2,405; (D) 2014–2016, 2,401). See also Table 1 and Table S1. Note: %EBLLs, percentage elevated blood lead levels.

**Table 4.** Statistical agreement analyses for the top 20 percentile method (80th–100th percentiles of census tracts with available data) of EJSCREEN 2017 Pb Paint EJ Index, Schultz et al. 2017 regression model approach, BLL data, HUD Deteriorated Paint Index, and exceedance rate of EBLLs (≥5 µg/dL) (census tracts; children 0 to <6 years of age).

| Data comparison | Census tracts (n) | Sensitivity | Specificity | Kappa
|-----------------|------------------|-------------|-------------|-----|
| Exceedance rate of EBLLs | | | | |
| 2014–2016 vs. EJSCREEN | 2,401 | 65 | 89 | 0.53 |
| 2014–2016 vs. Schultz | 2,401 | 67 | 90 | 0.55 |
| 2014–2016 vs. HUD | 2,401 | 57 | 88 | 0.44 |
| EJSCREEN vs. Schultz | 2,752 | — | — | — |
| EJSCREEN vs. HUD | 2,740 | — | — | 0.75 |
| Schultz vs. HUD | 2,741 | — | — | 0.72 |

Note: Exceedance rate is the exceedance rate of EBLLs (≥5 µg/dL) in children 0 to <6 years of age. —, not applicable; BLL, blood lead level; EBLL, elevated blood lead level; EJ, environmental justice; EPA, Environmental Protection Agency; GM, geometric mean; HUD, U.S. Department of Housing and Urban Development; Pb, lead.

*Cohen’s Kappa agreement statistic: 0.41–0.6, moderate; 0.61–0.8, substantial; 0.81–0.99, near perfect agreement.

**EJSCREEN** 2017 Pb Paint EJ Index.15

**U.S. EPA regression model, 2015 Census BLL GM, 2017 publication.13**

**HUD Deteriorated Paint Index.9**

%EBLL trends over time. Tracking changes in %EBLL over time may serve as “lagging indicators” of progress. Analyzing and visualizing %EBLL changes over time by census tracts showed an overall dramatic decline in %EBLLs (Figure 4). Locations where data show continuing high %EBLL (Figures S7 and S8) should be investigated to determine why rates have not declined and to guide future Pb reduction efforts. Before 2010, most census tracts in Michigan had decreasing rates of %EBLL; after 2010, there were still reductions, although the trend was much slower. Further investigation drawing on the expertise of state and local partners is needed to confirm these trends and interpret the results.
Figure 5. Getis-Ord Gi* geospatial cluster analysis for (A) EJSCREEN 2017 Pb Paint EJ Index (2,752 census tracts evaluated), (B) Schultz et al. regression model13 approach BLL Data (2,774 census tracts evaluated), and (C) HUD Deteriorated Paint Index (2,741 census tracts evaluated) displayed side by side with (D) the Getis-Ord Gi* cluster analysis for exceedance rate of EBLLs (2,401 census tracts evaluated; ≥ 5 μg/dL; 2014–2016; children 0 to <6 years of age). See also Table 3. Note: BLL, blood lead level; EBLL, elevated blood lead level; EJ, environmental justice; EPA, Environmental Protection Agency; HUD, U.S. Department of Housing and Urban Development; Kappa, Cohen’s Kappa agreement statistic (0.41–0.6, moderate; 0.61–0.8, substantial; 0.81–0.99, near perfect agreement); Pb, lead.

Table 5. Urban vs. rural census tract summary statistics for Getis-Ord Gi* geospatial cluster and top 20 percentile analyses of EJSCREEN 2017 Pb Paint EJ Index (2,752 census tracts evaluated), Schultz et al. regression model13 approach BLL data (2,774 census tracts evaluated), HUD Deteriorated Paint Index (2,741 census tracts evaluated), and exceedance rate of 2014–2016 EBLLs (≥ 5 μg/dL; children 0 to <6 years of age; 2,401 census tracts evaluated).

| Analysis | Number of census tracts (n) | Total 0- to <6 years of age population |
|----------|-----------------------------|---------------------------------------|
|          | Urban | Rural | Ratio | Urban | Rural | Ratio |
| Getis-Ord Gi* of 2014–2016 exceedance rate of EBLLs | No | 1,572 | 523 | 3 | 447,815 | 132,485 | 3 |
| | Yes | 295 | 11 | 27 | 71,430 | 2,669 | 27 |
| Top 20% of 2014–2016 exceedance rate of EBLLs | No | 1,439 | 482 | 3 | 410,752 | 123,428 | 3 |
| | Yes | 428 | 52 | 8 | 108,493 | 11,726 | 9 |
| Getis-Ord Gi* of EJSCREEN | No | 1,638 | 670 | 2 | 456,344 | 158,310 | 3 |
| | Yes | 443 | 1 | 443 | 105,651 | 2 | 52,826 |
| Top 20% of EJSCREEN | No | 1,532 | 669 | 2 | 426,110 | 158,039 | 3 |
| | Yes | 549 | 2 | 275 | 135,885 | 273 | 498 |
| Getis-Ord Gi* of Schultz | No | 1,581 | 682 | 2 | 439,846 | 158,310 | 3 |
| | Yes | 507 | 3 | 169 | 122,156 | 2 | 61,078 |
| Top 20% of Schultz | No | 1,536 | 682 | 2 | 426,141 | 158,308 | 3 |
| | Yes | 552 | 3 | 184 | 135,861 | 4 | 33,965 |
| Getis-Ord Gi* of HUD | No | 1,583 | 662 | 2 | 447,249 | 158,306 | 3 |
| | Yes | 494 | 2 | 247 | 114,747 | 4 | 28,687 |
| Top 20% of HUD | No | 1,552 | 660 | 2 | 429,012 | 157,969 | 3 |
| | Yes | 545 | 4 | 136 | 132,984 | 341 | 390 |

Note: Exceedance rate is the exceedance rate of EBLLs (≥ 5 μg/dL) in children 0 to <6 years of age. BLL, blood lead level; EBLL, elevated blood lead level; EJ, environmental justice; EPA, Environmental Protection Agency; GM, geometric mean; HUD, U.S. Department of Housing and Urban Development; Pb, lead.

*aRatio of urban to rural census tracts.
*bRatio of urban to rural population for children 0 to <6 years of age.
1 U.S. EPA EJSCREEN 2017 Pb Paint EJ Index.15
2 U.S. EPA regression model, 2015 Census BLL GM, 2017 publication.13
3 HUD Deteriorated Paint Index.9
**Locations identified with 2014–2016 data.** This study identified clusters of high %EBLL census tracts using the top 20 percentile analysis and geospatial cluster analysis. Ninety-one reference locations containing high %EBLL census tracts were identified: 11 (mainly urban locations) were identified with both analyses and an additional 80 with the top 20 percentile analysis. The Michigan BLL data we analyzed were predominantly from highly populated areas, although some data were from less populated areas. Blood Pb surveillance efforts in Michigan “have been concentrated on the geographic areas and populations where the exposure problem is greatest. . . However, it is possible to use the testing data to... characterize the population currently being tested.”

The top 20 percentile analysis identified rural or sparsely populated census tracts not identified by the geospatial cluster analysis because it does not take into consideration the neighboring census tracts around each selected census tract. In rural areas, where census tracts are large and encompass more than one municipality, cluster detection methods will overlook a single census tract with a high %EBLL because it has no neighbors with comparable %EBLL.

The focus of the MDHHS CLPPP is to ensure that all children with EBLLs in the state of Michigan are identified and receive appropriate attention. Consequently, CLPPP preferentially selects the highest and most reliable blood Pb test result to assign to an individual for each year. This approach creates a data set that is intentionally biased toward higher blood Pb test results. The purpose of the analysis in this paper, which involved randomly selecting blood tests results for each individual in the study, was to identify EBLL hotspots across the state of Michigan and to accurately characterize the population. Random selection results in a data set with less bias toward higher BLLs and estimates based on this data set provide results that are more representative of all blood Pb test results in the state. We conducted sensitivity analyses with the 2014–2016 BLLs using the MDHHS CLPPP sampling methodology to compare respective results. Table 2 demonstrates some minor statistical differences between the “highest” (i.e., MDHHS CLPPP) and “random” (i.e., our analysis) sampling selection methodologies. Figure S11 shows almost identical results between both approaches regarding reference locations identified, with one exception (Holland).

Our methodology is an important advancement for identifying and protecting vulnerable populations of children with high Pb exposures; however, it may miss some locations, for example, in rural or disadvantaged communities. It is important to check the BLL data representativeness with the two rates in Equations 1 and 2. Even though the correlation coefficient is 0.93 (venous and capillary data combined), there are some hotspots that our analysis could miss owing to the sampling strategy. However, our analysis of BLL data and convergence of available Pb indices in a state with robust BLL data can inform hotspot identification in Michigan and other states with less robust BLL data. As noted above, Roberts et al. distinguished BLL data quality across states based on their estimated “ascertainment ratios.”

The Getis-Ord Gi* analysis was able to detect clusters in areas of lower population density by using only venous blood tests, in comparison with cluster detection using all blood tests or only capillary blood tests (Figure S6). This included one cluster around an MDHHS CLPPP–targeted community that was not identified using all blood tests (city of Lansing), and clusters in counties in southwestern Michigan (Cass, St. Joseph, Van Buren) and the Upper Peninsula (Houghton, Marquette). Most algorithms used to identify communities for assistance favor locations where there is a large underlying population and large numbers of children tested and may overlook areas with smaller populations and low testing numbers. This may reflect a focus on taking action in areas where more children are potentially impacted, as well as difficulty in producing measures that are statistically valid in areas where less testing data are available (e.g., the small-number problems when computing rates based on small denominators). Cluster detection using venous blood Pb tests has the potential to overcome these limitations. Use of the venous-only test results appears to be able to detect clusters in low population areas, but further investigation is needed to ensure that the bias associated with using venous tests has been accounted for.

**Comparison to other studies.** Several reference locations identified in this study are consistent with high EBLL locations identified in previous studies.4,12,17,18,38 This study improves upon those earlier studies by identifying locations at a finer geographic resolution (Pell and Schneyer,12 used ZIP codes in Michigan); identifying locations throughout the state of Michigan (Hanna-Attisha et al.18 and Gómez et al.38 focused on Flint); and applying geospatial statistical methods for identifying potential locations for Pb targeting. Our analysis provides better temporal and geospatial resolution (e.g., analyzed BLL trends over 11 y of data, including more recent years; census tract vs. ZIP code scale) than Pell and Schneyer12 (with different time periods). We note that our reference location boundaries did not always align with how cities with high levels of EBLLs are identified by the MDHHS CLPPP, given that CLPPP uses the mailing address city of the subject provided by the testing laboratory and mailing address cities do not conform to census tract boundaries.

Many of the identified locations are in cities targeted by the MDHHS CLPPP for actions that are underway by state and local agencies to reduce Pb exposure in children. As reported in the “Results” section, targeted communities in the 2016 MDHHS report are mostly consistent with locations identified in our analyses. Three MDHHS CLPPP–targeted communities were not identified as reference locations by the Getis-Ord Gi* method. Of these three, two targeted communities (Hamtramck and Highland Park) were not selected as reference locations, even though they contained Getis-Ord Gi* hotspot census tracts, because they are municipalities embedded/located within the boundaries of the city of Detroit (Detroit was selected as the reference location instead). Using the top 20 percentile method, two MDHHS CLPPP–targeted communities (Hamtramck and Highland Park) were also not selected as “reference locations,” for the same reasons mentioned above.

A number of locations our analysis identified in Michigan have higher %EBLL rates than Flint (Figure 1), which is consistent with findings by other researchers.4,12,38 Gómez et al.38 observed that other areas in Michigan (specifically, the city of Detroit, as well as Kent, Jackson, Lenawee, Mason, and Menominee counties) experienced higher percentages of BLL $\geq 5 \mu g/dL$ than Flint during the 2015 Flint drinking water crisis. With the exception of Menominee County, these counties contain reference locations—namely, the cities of Grand Rapids, Jackson, Adrian, and Ludington—indicated by our geospatial cluster analysis. Our analysis of results for the state of Michigan are consistent with previous reports at the state level4,12 and for the city of Flint (Figure S4 compared with Figure 2 of Hanna-Attisha et al.,18 which used wards rather than census tracts for geographic aggregation). Although areas outside Flint had higher EBLLs during the period of interest, Flint continues to be of great concern because changes in the drinking water supply led to a public health crisis there.18 Flint has benefited from the attention through support from a variety of governmental and non-governmental organizations to improve conditions in the city and by increasing awareness about childhood Pb exposure among residents of Flint and the rest of Michigan. Other higher EBLL communities deserve attention and action also, but government resources for...
identifying and assisting these communities are not always readily available. Tools, such as cluster detection used in this paper, can be used to help identify other communities needing focused attention and resources.

We note that the CDC’s recent lowering of the blood Pb reference value to $\geq 3.5$ μg/dL does not impact our cluster analysis hotspot results for EBLL reference value $\geq 5$ μg/dL, based on the high Kappa value ($\sim 0.94$). This can also be seen in Figure S1.

Pb models/indices as surrogates for identifying high %EBLL locations. Comparison among the three Pb exposure models/indices showed close statistical convergence. This is to be expected given their use of similar variables and data despite being developed for different purposes. Importantly, differences in the exposure variables used—such as percentage minority versus percentage non-Hispanic African Americans, use of variables specific to visible peeling Pb paint, and differences in blood Pb data used (age categories, years covered)—can impact the convergence statistics and index-specific maps.

Although many higher %EBLL locations identified in this study were associated with old housing and sociodemographic variables by the three Pb exposure models/indices, 140 census tracts with high %EBLLs were not (when using the top 20% percentile method). This suggests that other sources of Pb exposure may be more important in those places than the housing and sociodemographic factors considered in this study as potential surrogates. The 140 census tracts with high %EBLLs that do not con verge with the old housing and sociodemographics-based models/indices and the moderate ($\sim 0.5$) Kappa scores (Tables 3 and 4) suggest other potential Pb exposure sources (Figure S10). As a potential next phase, these results can be used to determine potential causes of EBLs (e.g., old housing or other environmental Pb exposures through ambient air, outdoor soils, drinking water, and consumer food products) along with population characteristics in those locations.

Some Michigan EBLL locations identified here have been confirmed by previous efforts; however, we advanced the science beyond traditional mapping/visualization approaches, by a) identifying additional locations state-wide at a finer geographic resolution, and b) applying geospatial statistical methods for identifying potential locations for Pb targeting. Locations identified were further ground truthed (with the MDHHS CLPPP annual childhood Pb reports) where actions are currently being taken by state and local agencies.

Implications and future needs. The key findings in this paper—identifying high %EBLL locations at the census tract level using two different approaches, and the initial exploration of the causes of Pb exposure in those locations—have important implications for federal, state, and local policies and actions. Our analysis showed that available Pb exposure models/indices can be used as reasonable surrogates to identify high BLL locations in the absence of robust, representative state BLL data but should be verified with available blood Pb data and local knowledge of Pb sources and exposures. Prior literature reviews and published modeling efforts have found and produced results that suggest sociodemographic (e.g., poverty, race, housing, owner/renter status, education) factors, independent of environmental factors (e.g., ambient air levels), are useful as predictors for childhood BLLs.

The approach presented in this paper relies on completeness and accuracy of census tract-level BLL data. The analysis using available Pb indices is not dependent on the BLL data; however, it is dependent on the completeness and accuracy of U.S. Census/ACS and American Housing Survey data.

It is challenging to identify the key factors driving EBLLs in specific communities when available data (e.g., Pb concentrations in soil, dust, drinking water, air, food, blood; exposure factors, such as housing variables, sociodemographics data, bioavailability, soil/dust ingestion rate, premise plumbing) are assembled from disparate studies for use in multimedia source apportionment modeling. Our data convergence analysis highlights the need for more complete and coordinated environmental data collection and modeling to better pinpoint environmental sources for targeting Pb reduction actions.

Future research can also focus on further unraveling of the housing and demographic components, such as old housing contributions to BLLs from Pb paint, Pb service lines, or other sources not directly related to sociodemographics. The planned next phase of this research includes analyzing available local environmental data together with the census tract-level BLL, housing, and sociodemographic data and developing and applying new models to help understand the exposure sources contributing to higher BLL locations. Another future research need is analyzing the reason for BLL changes over time, such as the impacts of various environmental actions.

As a result of 55.6% of the Michigan MDHHS BLL data missing race variables and no data received for income and education level, we cannot directly estimate impacts of environmental justice-related variables on the Michigan %EBLL hotspots. As a surrogate for this missing data, we used U.S. Census ACS 2011–2015 data and merged it with the Michigan 2014–2016 exceedance rate of BLLs Getis-Ord Gi$^2$ geospatial cluster hotspots produced in this analysis. This data merging and the resulting statistics presented in Table S3 provide a general understanding of the environmental justice-related variable breakdown in hotspots vs. non-hotspots by population and respective percentages. These results, more specifically the high ratios presented in Table S3 (e.g., 4.09 means that there are four times more non-Hispanic African American people in hotspots than in non-hotspots), also demonstrate potential correlations between high percentages of environmental justice/vulnerable communities and high Pb exposure locations. The data in Table S2 also corroborates what is seen in Table S3, as the mean percentiles of environmental justice variables in %EBLL hotspot census tracts are consistently higher than in non-hotspot census tracts.

Landrigan et al. observed that Pb poisoning is disproportionately concentrated in poor minority communities in the United States because older housing units that are in disrepair are disproportionately concentrated in these neighborhoods. In an investigation of socioeconomic factors impacting childhood BLLs in Milwaukee County, Lynch and Meier stressed that community Pb prevention efforts should be informed by the intersection of multiple indicators of sociodemographic disadvantage. Additional techniques may be required to project the impact of Pb point sources on neighboring communities. Moody and Grady modeled airborne Pb in the Detroit Metropolitan Area using the U.S. EPA’s AERMOD model and found depositional and airborne Pb contributed directly and significantly to children’s BLLs, even after controlling for housing age and social structures, with black children at a continued disadvantage. This was consistent with the authors’ prior finding that Pb-emitting facilities were primarily located in, and moving to, highly Black segregated neighborhoods regardless of poverty levels. Egan et al., in their review of national survey data on BLLs collected over four decades, concluded that non-Hispanic Black race/ethnicity and poverty are consistently associated with higher BLLs.

This study suggests a systematic approach to help identify high Pb exposure locations. Where direct BLL surveillance data are available, this paper provides two statistical methods to identify hotspots, or areas of high potential Pb exposure. In the absence of sufficient BLL data, it provides a data convergence methodology using available surrogates in the absence of...
sufficient BLL data. These indices each have uncertainties given the limitations of the underlying survey data, but collectively they are useful for identifying sources of Pb hotspots related to old housing (e.g., Pb paint, Pb service lines). Where the indices did not show a strong statistical agreement with EBLL hotspots, collaborative engagement with state and local agencies is important to identify and address contributing sources.

This paper focused on a generalizable state-specific approach, using Michigan as a case study, with the goal of enhancing targeting efforts for Pb mitigation and prevention. This approach, using available BLL data and Pb exposure models/indices, can be used by other states in identifying Pb focus areas even where robust BLL data are lacking. This case study methodology for one state could inform potential how-to guides and training to identify high Pb exposure locations as called for by the U.S. EPA Lead Strategy and Federal Lead Action Plan. Effective risk communication will be critical when presenting the results of these analyses to stakeholders to inform the development of strategies to mitigate Pb exposures in children and when communicating these findings to the general public. Further interagency collaborations will be vital in obtaining, interpreting, and communicating available data, knowledge, and results to identify locations of high EBLL prevalence and key exposure sources. Such collaborations, supported by the scientific approach presented in this paper, can advance the public health protection goals described in the Federal Lead Action Plan.

Acknowledgments

We thank the U.S. Environmental Protection Agency (EPA) Office of Research and Development (ORD) managers, Sustainable and Healthy Communities research program, and communications team; U.S. EPA Region 5 managers; Michigan Department of Health and Human Services (MDHHS) Childhood Lead Poisoning Prevention Program (CLPPP) managers, including T. Largo and contractors R. Miller and D. Albright; C. Lee of the U.S. EPA Office of Environmental Justice; V. Garrison and P. Ashley of the U.S. Department of Housing and Urban Development; A. Ettinger (formerly of the Centers for Disease Control and Prevention, currently with Rutgers University); and J. Brown and S. Liu of the U.S. EPA/ORD for technical review of the manuscript.

This paper has been reviewed by the U.S. EPA and approved for publication. The views expressed in this manuscript are those of the authors and do not necessarily reflect the views or policies of the U.S. EPA. The research described in this article has been funded in part by the U.S. EPA under the ECHO–Multi Region Information Technology (IT) Services Support Contract and the U.S. EPA Region 1 IT Services Support Task Order; specifically, LitTech Global, Inc.’s ECHO Contract No. GS-35F-0334W/68HE0319FO0020 with the U.S. Environmental Protection Agency, Regions 1, 2 & 3. In addition, Contractor’s role did not include establishing Agency policy. This project was also supported in part by an appointment to the Research Participation Program at the Office of Research and Development, Center for Public Health and Environmental Assessment, U.S. Environmental Protection Agency, administered by the Oak Ridge Institute for Science and Education through an interagency agreement between the U.S. Department of Energy and EPA. We also acknowledge the Michigan Department of Health and Human Services for providing the statewide blood lead data under data use agreement 955-17.

The data reported in this paper are available at: https://edg.epa.gov/metadata/catalog/search/resource/details.page?uuid=https://doi.org/10.23719/1520754.

References

1. President’s Task Force on Environmental Health Risks and Safety Risks to Children. 2018. Federal Action Plan to Reduce Childhood Lead Exposures and Associated Health Impacts. https://www.epa.gov/sites/production/files/2018-12/documents/fedactionplan_lead_final.pdf [accessed 7 October 2021].
2. Hauptman M, Niles JK, Kaufman HW. 2021. Individual- and community-level factors associated with detectable and elevated blood lead levels in US children: results from a national clinical laboratory. JAMA Pediatr 175(12):1252, PMID: 34570188, https://doi.org/10.1001/jamapediatrics.2021.3518.
3. Roberts EM, English PB. 2018. Analysis of multiple-variable missing-not-at-random survey data for child lead surveillance using NHANES. Stat Med 35(29):5417–5429, PMID: 31723866, https://doi.org/10.1002/sim.7687.
4. Roberts EM, Madrigal D, Valle J, King G, Kite L. 2017. Assessing child lead poisoning case ascertainment in the US, 1999–2010. Pediatrics 139(5):e20164266, PMID: 28557759, https://doi.org/10.1542/peds.2016-4266.
5. McClure LF, Niles JK, Kaufman HW. 2016. Blood lead levels in young children: US, 2009–2015. J Pediatr 175:173–181, PMID: 27297207, https://doi.org/10.1016/j.jspeds.2016.05.005.
6. Egan K, Courtney J, the CDC/ATSDR LERI Development Teams. 2021. CDC-ATSDR Lead Exposure Risk Index (LERI). In: Lead Exposure and Prevention Advisory Committee Meeting, Centers for Disease Control and Prevention, December 2021. https://stacks.cdc.gov/view/cdc/114400.
7. Zartarian V, Xue J, Torner-Velez R, Brown J. 2017. Children’s lead exposure: a multimedia modeling analysis to guide public health decision-making. Environ Health Perspect 125(8):097009, PMID: 28694066, https://doi.org/10.1289/EHP1605.
8. Payne-Sturges DC, Braugelmann-JG. 2001. Local lead data are needed for local decision making. Am J Public Health 91(9):1396–1397, PMID: 11527786, https://doi.org/10.2105/ajph.91.9.1396.
9. Garrison VEH, Ashley PJ. 2021. Identifying jurisdictions at risk of containing housing units with deteriorated paint: results and targeting implications for the US Department of Housing and Urban Development. J Public Health Manag Pract 27(6):546–557, PMID: 32659085, https://doi.org/10.1097/PHH.000000000001191.
10. Zartarian V, Torner-Velez R, Poukoulos A, Xue J, Stanek J, Lee C, et al. 2019. Collaborative Evidence-Based Approach for Identifying High Pb Exposure Locations with Draft Examples from Several States. In: Proceedings of the 83rd Annual Educational Conference and Exhibition of the National Environmental Health Association. 9–12 July 2018. Nashville, TN: National Environmental Health Association. https://cfpub.epa.gov/si/si_public_record_report.cfm?Lab=NERL&dirEntryId=346407.
11. Frostenson S, Kiff S. 2016. Where is the lead exposure risk in your community? [Map.] In: The risk of lead poisoning isn’t just in Flint. So we mapped the risk in every neighborhood in America. Vox, 6 April 2016. https://www.vox.com/a/lead-exposure-risk-map [accessed 7 October 2021].
12. Pell MB, Schneyer J. 2016. Unsafe at any level; off the charts—the thousands of U.S. locales where lead poisoning is worse than in Flint. Reuters Investigates, 19 December 2016. https://www.reuters.com/investigates/special-report/usa-lead-testing/interactive-lead [accessed 7 October 2021].
13. Schultz BD, Morara M, Buxton BE, Weintraub M. 2017. Predicting blood-lead levels among U.S. children at the census tract level. Environ Justice 10(5):129–138, https://doi.org/10.1089/env.2017.0055.
14. U.S. EPA (U.S. Environmental Protection Agency). 2007. Final Report For The Pilot Study Of Targeting Elevated Blood-Lead Levels In Children. EPA 260-R-08-003. 3 October 2007. https://www.epa.gov/lead/final-report-pilot-study-targeting-elevated-lead-levels-children-epa-260-003-003/october [accessed 7 October 2021].
15. U.S. EPA. 2017. EJSCREEN 2017 technical documentation and general website. https://www.epa.gov/sites/production/files/2017-09/documents/2017_ejscreen_technical_document.pdf; https://www.epa.gov/ejscreen [accessed 7 October 2021].
16. Wheeler DC, Jones RM, Schootman M, Nelson EJ. 2019. Explaining variation in elevated blood lead levels among children in Minnesota using neighborhood socioeconomic variables. Sci Total Environ 650(1):970–977, PMID: 30368871, https://doi.org/10.1016/j.scitotenv.2018.09.088.
17. MDHHS CLPP (Michigan Department of Health and Human Services Childhood Lead Poisoning Prevention Program). 2021. 2018 Annual Data Report on Childhood Blood Lead Testing in Michigan, Issued 15 June 2021. https://www.michigan.gov/documents/mi/2018_CLPP_Data_Report_727774_7.pdf [accessed 7 October 2021].
18. Hanna-Attisha M, LaChance J, Sadney M, Champaign Schnepf A. 2016. Elevated blood lead levels in children associated with the Flint drinking water crisis: a spatial analysis of risk and public health response. Am J Public Health 106(2):283–290, PMID: 26951115, https://doi.org/10.2105/AJPH.2015.303003.
19. Laidlaw MAS, Filippelli GM, Sadler RC, Gonzales CR, Ball AS, Mielke HW. 2016. Children’s blood lead seasonality in Flint, Michigan (USA), and soil-sourced lead hazard risks. Int J Environ Res Public Health 13(4):358, PMID: 27022578, https://doi.org/10.3390/ijerph13040358.
20. U.S. EPA. 2005. Guidance on Selecting Age Groups for Monitoring and Assessing Childhood Exposures to Environmental Contaminants. EPA/630/P-03/003F. November 2005. Washington, DC: U.S. EPA, Risk Assessment Forum.
21. Centers for Disease Control and Prevention. 2021. Childhood Lead Poisoning Prevention Program. Last reviewed 6 June 2022. https://www.cdc.gov/ncceh/lead/prevention/blood-lead-levels.htm [accessed 7 October 2021].
22. Landis JR, Koch GG. 1977. The measurement of observer agreement for categorical data. Biometrics 33(1):159–174, PMID: 843571, https://doi.org/10.2307/252910.
23. Ord JK, Getis A. 1995. Local spatial autocorrelation statistics: distributional issues and an application. Geogr Anal 27(4):286–306, https://doi.org/10.1111/j.1539-4632.1995.tb00912.x.
24. ESRI ArcGIS Pro. 2021. How Hot Spot Analysis (Getis-Ord Gi*) works. https://pro.arcgis.com/en/pro-app/tool-reference/spatial-statistics/h-how-hot-spot-analysis-getis-ord-gi-spatial-stat.htm [accessed 1 October 2021].
25. Ruckart PZ, Jones RL, Courtney JG, LeBlanc TT, Jackson W, Karwowski MP, et al. 2021. Update of the blood lead reference value—United States, 2021. MMWR Morb Mortal Wkly Rep 70(43):1509–1512, PMID: 34710078, https://doi.org/10.15585/mmwr.mm7043a4.
26. U.S. Census Bureau. 2010. Places. https://www.census.gov/cgi-bin/geo/shapereflayer?index.php?year=2010&layergroup=Places [accessed 16 June 2022].
27. U.S. Census Bureau. 2010. 2010 Decennial Census. https://data.census.gov/cedsci/all?g=0100000US%241400000&y=2010&d=DEC%20Summary%20File%201 [accessed 16 June 2022].
28. MDHHS CLPPP (Michigan Department of Health and Human Services Childhood Lead Poisoning Prevention Program). 2021. Update of the blood lead reference value—United States, 2021. MMWR Morb Mortal Wkly Rep 70(43):1509–1512, PMID: 34710078, https://doi.org/10.15585/mmwr.mm7043a4.
29. U.S. Census Bureau. 2010. Detailed Tables, 2011–2015 American Community Survey 5-year estimates. https://data.census.gov/cedsci/all?g=0100000US%241400000&y=2015&d=ACS%205-Year%20Estimates%20Detailed%20Tables [accessed 16 May 2022].
30. Frostenson S. 2017. 1.2 million children in the US have lead poisoning. We’re only treating half of them. Vox, 27 April 2017. https://www.vox.com/science-and-health/2017/4/27/14524050/us-underreports-lead-poisoning-cases-map-community [accessed 7 October 2021].
31. Wang A, Rezania Z, Haugen KMB, Baertlein L, Yandell SJ. 2019. Screening for elevated blood lead levels: rate-positive-frequencies of tests on capillary samples, Minneapolis, 2011–2017. J Public Health Manag Pract 25(suppl 1):S44–S50, PMID: 30507769, https://doi.org/10.1097/PHH.0000000000000879.
32. Raymond J, Brown MJ. 2017. Childhood blood lead levels in children aged < 5 years—United States, 2009–2014. MMWR Surveill Summ 66(3):1–10, PMID: 28103215, https://doi.org/10.15585/mmwr.ss6603a1.
33. Schlenker TL, Fritz CJ, Mark D, Layde M, Linke B, Murphy A, et al. 1994. Screening for pediatric lead poisoning: comparability of simultaneously drawn capillary and venous blood samples. JAMA 271(17):1346–1348, PMID: 8158820, https://doi.org/10.1001/jama.1994.035104105833.
34. Cantor AG, Hendrickson R, Blazina I, Griffin J, Grusing S, McDonagh MS. 2019. Screening for Elevated Blood Lead Levels in Children: A Systematic Review for the U.S. Preventive Services Task Force. Evidence Synthesis, No. 174. Report No.: 18-05245-EF-1. Rockville, MD: Agency for Healthcare Research and Quality. https://www.ncbi.nlm.nih.gov/books/NBK540622/.
35. Crogan CW, Eggehy PP. 2003. Methods of dealing with values below the limit of detection using SAS. Presented at Southern SAS User Group, St. Petersburg, FL, 22–24 September 2003. https://analytics.ncsu.edu/sesug/2003/ S008-Croghan.pdf [accessed 7 October 2021].
36. LaFleur B, Lee W, Billiheimer D, Lockhart C, Liu J, Merchant N. 2011. Statistical methods for assays with limits of detection: serum bile acid as a differentiator between patients with normal colons, adenos, and colorectal cancer. J Carcinog 10(1):12, PMID: 21712958, https://doi.org/10.4103/1477-3163.79681.
37. Caldwell KL, Cheng PY, Jarrett JM, Mahmudov A, Vance K, Ward CD, et al. 2017. Measurement challenges at low blood lead levels. Pediatrics 140(2): e20170272, PMID: 28771411, https://doi.org/10.1542/peds.2017-0272.
38. Gómez HF, Borgiai DA, Sharman M, Shah KK, Scolpino AJ, Oleske JM, et al. 2018. Blood lead levels of children in Flint, Michigan: 2008–2016. J Pediatr 197:158–164, PMID: 29599069, https://doi.org/10.1016/j.jpeds.2017.12.063.
39. Frank JJ, Poulokas AG, Tornero-Velez R, Xue J. 2019. Systematic review and meta-analyses of lead (Pb) concentrations in environmental media (soil, dust, water, food, and air) reported in the United States from 1996 to 2016. Sci Total Environ 694:133489, PMID: 31756826, https://doi.org/10.1016/j.scitotenv.2019.07.295.
40. U.S. EPA. 2013. Integrated Science Assessment (ISA) for Lead (Final Report, Jul 2013). EPA/600-R-10/075F. Washington, DC: U.S. EPA.
41. Akkus C, Ozdenerol E. 2014. Exploring childhood lead exposure through GIS: a review of the recent literature. Int J Environ Res Public Health 11(6):6334–6334, PMID: 24945189, https://doi.org/10.3390/ijerph11060634.
42. Landrigan PJ, Rauh VA, Galvez MP. 2010. Environmental justice and the health of children. Mt Sinai J Med 77(1):178–187, PMID: 20309928, https://doi.org/10.1002/mjs.20173.
43. Lynch EE, Meier HCS. 2020. The intersectional effect of poverty, home ownership, and racial/ethnic composition on mean childhood blood lead levels in Milwaukee County neighborhoods. PLoS One 15(6):e0234995, PMID:32559243, https://doi.org/10.1371/journal.pone.0234995.
44. Moody HA, Grady SC. 2021. Lead emissions and population vulnerability in the Detroit Metropolitan Area, 2006–2013: impact of pollution, housing age and neighborhood racial isolation and poverty on blood lead in children. Int J Environ Res Public Health 18(5):2747, PMID: 33800525, https://doi.org/10.3390/ijerph18052747.
45. Moody H, Grady SC. 2017. Lead emissions and population vulnerability in the Detroit (Michigan, USA) Metropolitan Area, 2006–2013: a spatial and temporal analysis. Int J Environ Res Public Health14(12):1445, PMID: 29168799, https://doi.org/10.3390/ijerph14121445.
46. Egan KB, Cornwell CR, Courtney JG, Ettenger AS. 2021. Blood lead levels in US children ages 1–11 years, 1976–2016. Environ Health Perspect 129(3):37003, PMID: 33730866, https://doi.org/10.1289/EHP7932.