Tapping into Lead Service Line Information: Two City Case Study

Rogelio Tornero-Velez[1], Valerie Zartarian[1], Antonios Poulakos[1,2] ,
Nicole Shao[1], Lindsay Stanek[1], Timothy Buckley[1], Michael Goldberg[3]
[1] US EPA Office of Research and Development
[2] LinTech Global,Inc.,Boston,Massachusetts,USA
[3] US EPA Office of Ground Water and Drinking Water

Lead Data Mapping: Methods and Tools for Lead Prioritization, Prevention, and Mitigation

2023 NEHA Annual Educational Conference & Exhibition
July 31-August 3, 2023

New Orleans, Louisiana

Background

- Bipartisan Infrastructure Law * provides $15 billion for “lead service line (LSL) replacement projects and associated activities directly connected to the identification, planning, design, and replacement of LSL.”

- This law creates an opportunity to evaluate the impact of LSL prevalence on lead exposure

- Goal of current study is to use existing data to assess association between LSL and children’s elevated blood lead levels (EBLL) in two midwestern cities.

*https://www.whitehouse.gov/briefing-room/statements-releases/2021/11/06/fact-sheet-the-bipartisan-infrastructure-deal/

Data & Approach

- LSL prevalence
  - Ohio Utility #3 and Michigan Utility #5 provided Lead Service Lines (LSL) data to EPA
  - EPA aggregated LSL data to 2010 census tracts and calculated %LSL per tract
  - Weighting by the population size of children aged 0-5 year in census block groups

- $\text{EBLL prevalence} = \frac{\#\text{children tested in the census tract with EBLL}}{\#\text{children tested in the census tract}} \times 100,$

  where an elevated blood lead level (**EBLL**) is when child's blood lead level ≥ 5µg/dL, per Xue et al. 2022 (https://ehp.niehs.nih.gov/doi/10.1289/EHP9705)

- Approach
  - Compare the predictive value of LSL prevalence to other Pb exposure indices and models (EJSCREEN Index, HUD Index, Random Forest Regression EBLL Prediction Model)
  - Using linear regression or weighted quantile sum (WQS) regression
  - We regressed logit, $\ln(\theta/(1-\theta))$, on standardized predictors to compare them
  - where θ is the EBLL prevalence

Methods- Compare LSL prevalence with Pb indices

- EPA EJSCREEN 2017 Pb Paint EJ Index (www.epa.gov/ejscreen) – "EJSCREEN"
  - Originally developed at census block group level by EPA OEJ
  - Uses American Community Survey 2011-2015 5-year summary file
  - Based on pre-1960 homes,% low income, % minority, population
  - EPA ORD aggregated the data by averaging index values per census tract

- HUD Deteriorated Paint Index (Garrison & Ashley, 2020) – "HUD DPI"
  - Provided by the US Department of Housing and Urban Development (https://hudgis-hud.opendata.arcgis.com/datasets/deteriorated-paint-index-by-tract)
  - Uses 2011 American Housing Survey and 2009-2013 American Community Survey Data
    - 2011 American Housing Survey: occupied pre-1980 households that reported a large area of peeling paint
    - 2009-2013 American Community Survey: presence of children in household, housing tenure status (owned, rented, or other), household income, race (white, black, other), ethnicity (Hispanic or non-Hispanic), and education level

- EPA/ORD Work-in-Progress Random Forest Regression EBLL Prediction Model – "RF Model"
  - Based on Ohio 2007-2011 BLL data and 2013 Ohio Dept Health report model, and currently includes the following 5 predictors: % homes built prior to 1940, % homes built prior to 1950, % families whose income-to-poverty ratio was > than 2, % population with either high-school or higher education, % non-Hispanic African Americans
  - Demographic data originate from the American Community Survey 2013-2017 5-year summary file

Methods- RF Regression model of EBLL prevalence

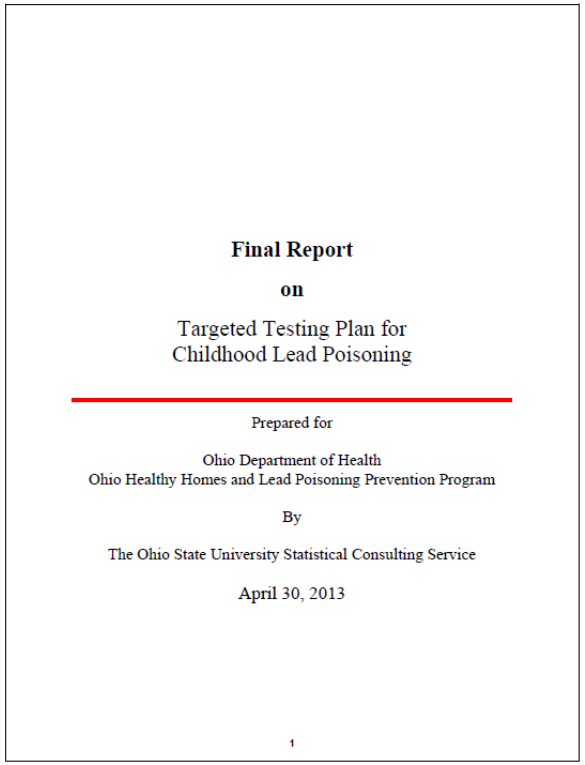
EPA/ORD Work-in-Progress Random Forest Regression EBLL Prediction Model for Children < 6 years

In prior work, we assessed data published in a 2013 report prepared for the Ohio Department of Health

A set of 29 housing-demographic variables for census tracts were identified through RF regression model developed from Ohio, 2007-2011

these were most important:

- percent of houses built before 1940 (DP04),
- percent of houses built before 1950 (DP04),
- percent of population that is African American (non-Hispanic) (DP1),
- percent of households with income to poverty ratio greater than 2 (B17026),
- percent of population with a high school degree or higher (DP02)



Final Report

on

Targeted Testing Plan for
Childhood Lead Poisoning

Prepared for

Ohio Department of Health
Ohio Healthy Homes and Lead Poisoning Prevention Program

By

The Ohio State University Statistical Consulting Service

April 30, 2013

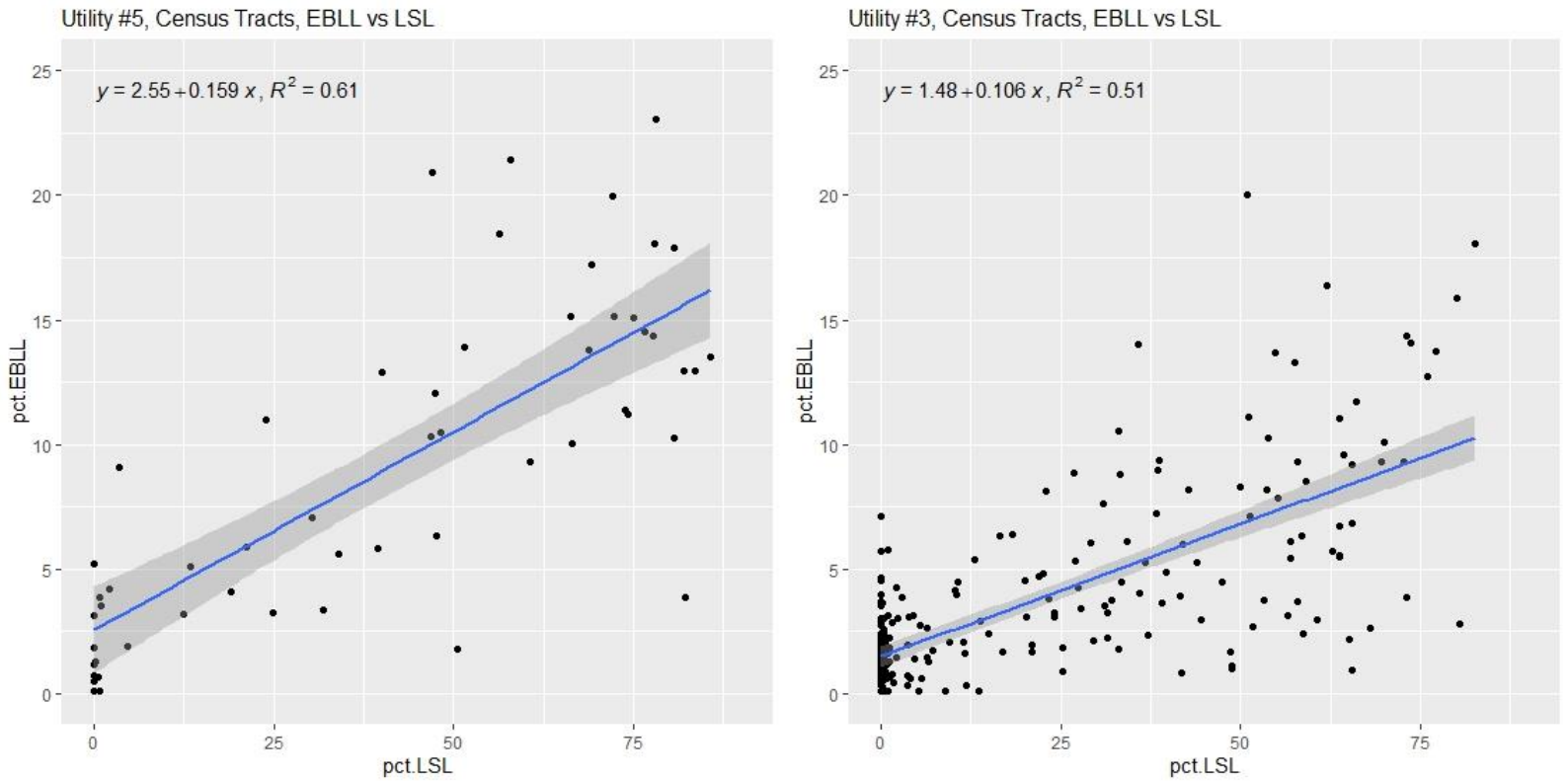
1

Moderate to strong correlation between LSL prevalence and EBLL prevalence

Utility #5 (N=57)

Utility #3 (N=232)

Percent EBLL



Percent LSL

Strong correlation between LSL prevalence and EBLL prevalence

Strong correlation between LSL prevalence and EBLL prevalence, at census tract, where EBLL is children's BLL ≥ **5µg/dL**

For Michigan **Utility #5:**

Pearson Correlation (**LSL percent**, **EBLL percent**)

= 0.78, (p = 6.9e-13); 95%CI: 0.65, 0.87 (df= 55)

For Ohio **Utility #3:**

Pearson Correlation (**LSL percent**, **EBLL percent**)

= 0.71, (p < 2.2e-16); 95%CI: 0.64, 0.77 (df= 230)

Correlation between LSL prevalence and recognized Pb Covariates

Utility #5 (N=57)

	Pct_LSL_w
Pct_LSL_w	1.00
z_Pct_LSL_w	1.00
pct.home_pre1950	0.90
pct.home_pre1940	0.89
RF.OH0711	0.84
z_RF.OH0711	0.84
Observed	0.78
logit	0.74
EJS_PbPI	0.55
z_EJS_PbPI	0.55
HUD_DPI	0.47
z_HUD_DPI	0.47
pct.black	0.26
pct.HS_higher	-0.51
pct.inc_pov_ratio_g2	-0.61

Utility #3 (N=232)

	Pct_LSL_w
Pct_LSL_w	1.00
z_Pct_LSL_w	1.00
pct.home_pre1940	0.77
pct.home_pre1950	0.75
Observed	0.71
RF.OH0711	0.70
z_RF.OH0711	0.70
logit	0.63
HUD_DPI	0.52
z_HUD_DPI	0.52
EJS_PbPI	0.43
z_EJS_PbPI	0.43
pct.black	0.33
pct.HS_higher	-0.42
pct.inc_pov_ratio_g2	-0.47

Utilities #3 and #5 (N=289)

	Pct_LSL_w
Pct_LSL_w	1.00
z_Pct_LSL_w	1.00
pct.home_pre1940	0.80
pct.home_pre1950	0.79
Observed	0.75
RF.OH0711	0.73
z_RF.OH0711	0.73
logit	0.69
HUD_DPI	0.48
EJS_PbPI	0.48
z_HUD_DPI	0.48
z_EJS_PbPI	0.48
pct.black	0.23
pct.HS_higher	-0.47
pct.inc_pov_ratio_g2	-0.50

Observed = EBLL prevalence

LSL prevalence vs. HUD DPI, EJSCREEN, or RF model, Utilities #5 (n=57)

	WQS Regression		Linear Regression					
	Component	Mean Weight	Component	Coefficient	SE	t value	P value	Signif.
			(Intercept)	-2.821	0.123	-23.004	< 2e-16	***
EJ Screen	Pct_LSL_w	0.880	z_Pct_LSL_w	0.911	0.148	6.161	0.000	***
	EJS_PbPI	0.120	z_EJS_PbPI	0.163	0.148	1.099	0.276	
			(Intercept)	-2.821	0.113	-24.930	< 2e-16	***
HUD DPI	Pct_LSL_w	0.750	z_Pct_LSL_w	0.801	0.129	6.206	0.000	***
	HUD_DPI	0.250	z_HUD_DPI	0.425	0.129	3.293	0.002	**
			(Intercept)	-2.821	0.123	-22.935	< 2e-16	***
RF Model	Pct_LSL_w	0.640	z_Pct_LSL_w	0.820	0.228	3.598	0.001	***
	RF.OH0711	0.360	z_RF.OH0711	0.213	0.228	0.936	0.353	

Signif. Codes: * 0.05, ** 0.01, *** 0.001

Result: As 'predictor of' or 'contributor to' Pb exceedance, LSL prevalence outperformed HUD DPI, EJSCREEN, or RF model.

LSL prevalence vs. HUD DPI, EJSCREEN, or RF model, Utilities #3 (n=232)

	WQS Regression		Linear Regression					
	Component	Mean Weight	Component	Coefficient	SE	t value	P value	Signif.
			(Intercept)	-3.849	0.062	-61.845	< 2e-16	***
EJ Screen	Pct_LSL_w	0.783	z_Pct_LSL_w	0.692	0.069	9.993	< 2e-16	***
	EJS_PbPI	0.217	z_EJS_PbPI	0.203	0.069	2.931	0.004	**
			(Intercept)	-3.849	0.062	-61.893	< 2e-16	***
HUD DPI	Pct_LSL_w	0.866	z_Pct_LSL_w	0.665	0.073	9.097	< 2e-16	***
	HUD_DPI	0.134	z_HUD_DPI	0.219	0.073	2.993	0.003	**
			(Intercept)	-3.849	0.059	-64.713	< 2e-16	***
RF Model	Pct_LSL_w	0.520	z_Pct_LSL_w	0.452	0.084	5.391	0.000	***
	RF.OH0711	0.480	z_RF.OH0711	0.467	0.084	5.581	0.000	***

Signif. Codes: * 0.05, ** 0.01, *** 0.001

Result: As 'predictor of' or 'contributor to' Pb exceedance, LSL prevalence outperformed HUD DPI, EJSCREEN, or RF model.

LSL prevalence vs. HUD DPI, EJSCREEN, or RF model, Utilities #3 and #5 (n=289)

	WQS Regression		Linear Regression					
	Component	Mean Weight	Component	Coefficient	SE	t value	P value	Signif.
			(Intercept)	-3.646	0.056	-64.991	< 2e-16	***
EJ Screen	Pct_LSL_w	0.867	z_Pct_LSL_w	0.819	0.064	12.769	< 2e-16	***
	EJS_PbPI	0.133	z_EJS_PbPI	0.187	0.064	2.912	0.004	**
			(Intercept)	-3.646	0.056	-65.251	< 2e-16	***
HUD DPI	Pct_LSL_w	0.769	z_Pct_LSL_w	0.808	0.064	12.667	< 2e-16	***
	HUD_DPI	0.231	z_HUD_DPI	0.210	0.064	3.291	0.001	**
			(Intercept)	-3.646	0.055	-66.618	< 2e-16	***
RF Model	Pct_LSL_w	0.772	z_Pct_LSL_w	0.627	0.080	7.835	0.000	***
	RF.OH0711	0.228	z_RF.OH0711	0.387	0.080	4.838	0.000	***

Signif. Codes: * 0.05, ** 0.01, *** 0.001

Result: As 'predictor of' or 'contributor to' Pb exceedance, LSL prevalence outperformed HUD DPI, EJSCREEN, or RF model.

Findings

In two cities with available lead service line (LSL) and blood lead level data:

- Moderate to strong correlation between LSL prevalence and: 1) prevalence of blood lead exceedance and 2) housing and sociodemographic variables known to be relevant to lead exposure
- LSL prevalence was a stronger indicator of blood-Pb exceedance than EJSCREEN Pb Paint Index, HUD Deteriorated Paint Index, or a random forest predictor of blood-Pb exceedance.
- This work suggests LSL prevalence is an important predictor of EBLL and should be considered in hotspot analyses
- Study findings relevant only to the two Midwest utilities studied

Acknowledgements/ Disclaimers

Acknowledgements

- Michigan Department of Health and Human Services (MDHHS) provided blood lead data used in this presentation, pursuant to a Data Use Agreement. EPA assumes full responsibility for the analysis and interpretation of the data.
- This presentation includes analyses with blood lead data provided by the Ohio Department of Health (ODH) , through the Ohio Public Health Information Warehouse. The Department specifically disclaims responsibility for any analyses, interpretations, or conclusions from these data.
- Research included in this analysis was approved under Institutional Review Boards (IRBs) through University of North Carolina at Chapel Hill (UNC; 16-2302), MDHHS (201703-12-EA) and ODH (2019-41).
- 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, LinTech Global, Inc.'s ECHO Contract No. GS-35F-0343W/ 68HE0319F0020 with the U.S. Environmental Protection Agency, Regions 1, 2 & 3.

Disclaimer

- The views expressed in this presentation are those of the authors and do not necessarily reflect the views or policies of the U.S. EPA.

EPA United States Environmental Protection Agency

Questions?

email: tornero-velez.rogelio@epa.gov