Prepared for the Hawaiʻi State Department of Health Family Health Services Division

Hawaiʻi Childhood Lead Poisoning Prevention Program (HI-CLPPP) Risk System Evaluation

Technical Report

2022

University of Hawaiʻi Center on the Family
Hua Zan, Javzandulam Azuma, Ji-Yeon Kim, Barbara DeBaryshe

Table of Contents

Overview

The primary goals of this evaluation are to (1) examine advantages and disadvantages of the existing Hawai‘i Childhood Lead Poisoning Prevention Program (HI-CLPPP) risk surveillance system; (2) validate the existing method to identify high-risk geographic areas to screen children; (3) propose the desired qualities of a risk surveillance system; and (4) suggest potential alternative zip code methods. This technical report provides details about data sources used in this evaluation, data cleaning and preparation procedures, and data analysis.

Data Sources

The analysis was based on three data sources: the HI-CLPPP lead surveillance database; the 2011-2015, 2012-2016, 2013-2017, 2014-2018, and 2015-2019 American Community Survey (ACS) five-year estimates from U.S. Census Bureau (census data); and the 2017 and 2021 SmartParcels® real property database. The HI-CLPPP lead surveillance database was matched with census data by census tract. The property data were matched with the geocodes of the residential addresses of children in the HI-CLPPP lead surveillance database.

HI-CLPPP Lead Surveillance Database

The HI-CLPPP lead surveillance database included information such as patient date of birth, sex, street address, blood lead level, and blood sample collection date. Table 1 details the variables we used in the lead surveillance database and how they were preprocessed for extraction.

Census Data

We downloaded Census tables of selected population, social, economic, and housing characteristics from the American Community Survey. To increase reliability and avoid suppressed data for tracts with small populations, we used five-year estimates instead of single-year estimates. The specific variable names with associated database information (i.e., year, data table number) are listed in Table 2.

Property Data

The 2021 SmartParcels® property data included information about the year the primary structure on the property was built. They were linked to the children's residential addresses using geocodes, and provided individualized housing information that was a useful complement to ACS estimates of tract level housing stock age. The 2017 property data were used to capture housing units in overlapping areas between census tracts and zip codes (see more details in the "Census Tract and Zip Code Conversion" section).

Table 1. HI-CLPPP lead surveillance database

Section	Description	Note
Patient Information	De-identified patient ID	De-identified unique identifier for each patient
	Age	Difference between specimen collection date and patient's date of birth
	Street address	
	City	
	County	
	State	
	Zip Code	
Blood Test Information	Blood lead value	Reported in mcg/dL
	Date blood sample was analyzed	in MM/DD/YYYY
	Specimen collection date	in MM/DD/YYYY
	Date blood sample was resulted	in MM/DD/YYYY
	Lab facility where sample was processed	

Table 2. ACS data five-year estimates

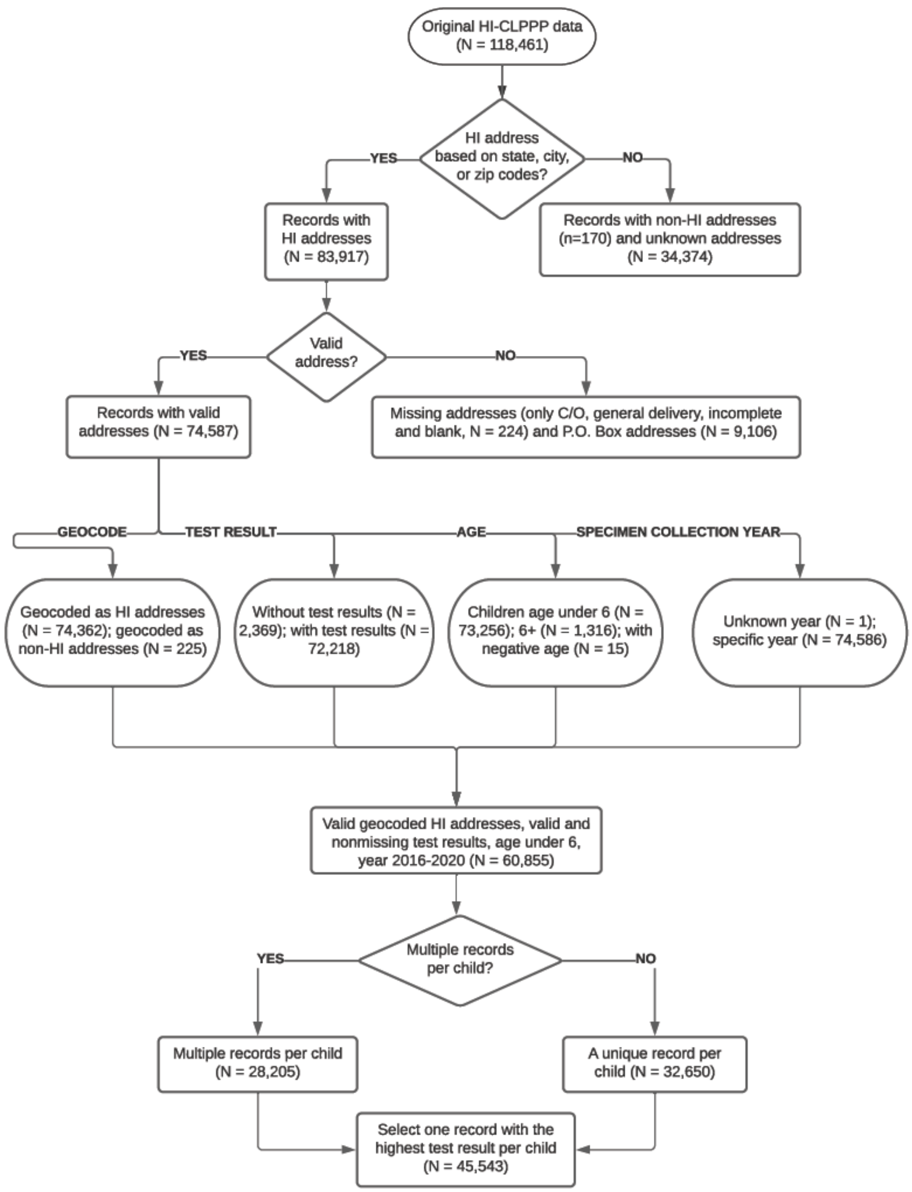
Variable	Description	Database
DP02_0059 to DP02_0068	Educational attainment	2015-2019 ACS 5-year, Table DP02
DP02_0111 to DP02_0122	Language spoken at home	2015-2019 ACS 5-year, Table DP02
DP03_0119 to DP03_0137	% of families and people with income below poverty level in the past 12 months	2015-2019 ACS 5-year, Table DP03
DP04_0016 to DP04_0026	# of housing units by year structure built	2015-2019 ACS 5-year, Table DP04
B05007_001 to B05007_106	Place of birth by year of entry and citizenship status for foreign-born population	2015-2019 ACS 5-year, Table B05007
B17024_001 to B17024_040	Age by ratio of income to poverty level	2011-2015, 2012-2016, 2013-2017, 2014-2018, and 2015-2019 ACS 5-year, Table B17024

Data Processing and Preparation

Data Processing Steps

The original data in the HI-CLPPP lead surveillance database provided by DOH included 118,461 records from 2012 to 2021. When a child was tested more than once, there were multiple records for the child. The following flowchart illustrates the data processing steps:

Figure 1. HI-CLPPP lead surveillance data processing flowchart



The first step of data preparation began with excluding records with non-Hawai‘i addresses using state, city, and zip code information, resulting in 83,917 records. Then, we excluded invalid addresses and P.O. Box addresses, resulting in 74,587 records. Among these 74,587 records, 225 were found to be out-of-state addresses after they were geocoded; 2,369 records did not have valid blood lead test results; 1,316 records were from children age six or older; 15 records had negative ages calculated from birth dates and blood collection dates; and one record was missing the specimen collection year. Further, 13.49% of the 74,587 records were collected before 2016 or after 2020 (Table 3). When we limited our sample to those cases with valid geocoded Hawai‘i addresses, valid test results, and age under six from 2016 to 2020, the sample consisted of 60,855 records.

There were multiple lead tests reported for some children as follow-up tests were common especially for children with elevated lead levels. In those cases, only one blood test result for each child was used by selecting the result with the highest blood lead level. When there were ties on the highest blood lead level for a child, the record with the most recent specimen collection date was selected. This process resulted in a final sample of 45,543 unique records.

Table 3. Records by specimen collection year

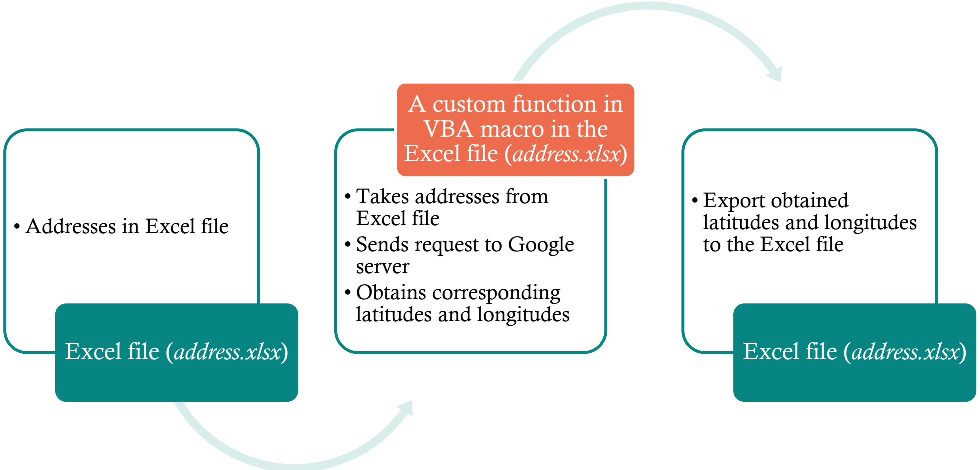
Year	# of records	%
2012	1	<0.01
2013	2,973	3.99
2014	3,245	4.35
2015	3,673	4.92
2016	11,107	14.89
2017	11,426	15.32
2018	13,962	18.72
2019	14,831	19.88
2020	13,201	17.70
2021	167	0.22
.	1	<0.01
Total	74,587	100

Source: HI-CLPPP lead surveillance database

Geocoding

Patient addresses were geocoded using Google Geocoding API. The geocoding was done using a custom function in a VBA Excel macro which imported each patient address from an Excel file, converted it into latitude and/or longitude coordinates using Google’s geocoding service, and then exported the results to the Excel file. The following figure illustrates the steps:

Figure 2. The geocoding process



Lead Test Results

For the main analysis using 5 mcg/dL as the reference value, eight records (e.g., "<65", "<10", "6.3; 3.6", and "8.5; <1.9") were excluded from the analysis as it was unclear if the actual test results were higher or lower than the reference value. For sensitivity analysis using 3.5 mcg/dL as the reference value, two additional records (i.e., "<2; 4", and "<5") were excluded as it was unclear if the actual test results were higher or lower than the reference value. For other records with more than one test result in one cell, the results were either clearly higher or lower than the reference value (e.g., "2.1; 2.1; 2.1", "<1.9; 2.1"), and we chose the highest value for the record.

Data Analysis

Testing Rate

The blood lead testing rate for each census tract was calculated using the number of children tested divided by the estimated number of children under age six living in that tract. The numerator was the unduplicated head count of children in the HI-CLPPP database with at least one valid blood test taken before age six between 2016 and 2020. The denominator was the number of children living in each tract who were under age six at any point during the years 2015-2019. The reason the time frame for the population estimates is not an exact match to the time frame of the HI-CLIPP dataset is that the 2016-2020 ACS five-year estimates had not been released at the time this report was prepared. However, we do not expect a significant impact on the results, as there are only minor fluctuations in these estimates from year to year.

The denominator was somewhat difficult to capture using the available population statistics. First, the ACS does not provide population estimates by single year of age, so we were limited to using the population under age six. Second, each year some children entered the study window of eligibility while others exited. For example, a newborn in 2015 would continue to be eligible for inclusion in the analysis through 2019. A child who was five years old in 2015 would only contribute towards the analysis head count in 2015, while a child born in 2019 would only contribute towards the analysis head count in 2019. The 2015-2019 ACS five-year population estimate of children under age six is less than the total study head count, because children who aged out during the study period are not included. To estimate the total unique head count during the study window, we took the 2015-2019 five-year estimate of children under age six and added one-sixth of the ACS estimates for the preceding years (see Table 4). This required the assumption that there was an equal number of children at each single year of age (under 12 months, one year, two years, three years, four years, five years).

Table 4. Derivation of the estimated number of children under age six between 2015 and 2019

Child's age in 2019	Child's age in 2018	Child's age in 2017	Child's age in 2016	Child's age in 2015	Data source (for cells in orange)	Estimated *N*
<1	Not yet born	Not yet born	Not yet born	Not yet born	2015-2019 ACS 5-year estimate of children < age 6	All children *N* = 104,037
1	<1	Not yet born	Not yet born	Not yet born		
2	1	<1	Not yet born	Not yet born		
3	2	1	<1	Not yet born		
4	3	2	1	<1		
5	4	3	2	1		
6	**5**	4	3	2	2014-2018 ACS 5-year estimate of children < age 6	1/6 of children *N* = 17,520
7	6	**5**	4	3	2013-2017 ACS 5-year estimate of children < age 6	1/6 of children *N* = 17,800
8	7	6	**5**	4	2012-2016 ACS 5-year estimate of children < age 6	1/6 of children *N* = 17,937
9	8	7	6	**5**	2011-2015 ACS 5-year estimate of children < age 6	1/6 of children *N* = 17,835

Notes: Each year's data came from the ACS Table B17024 for the estimated number of children under age six. The counts for cells in orange were summed to create the total estimated number of children under age six at any point between 2015 and 2019.

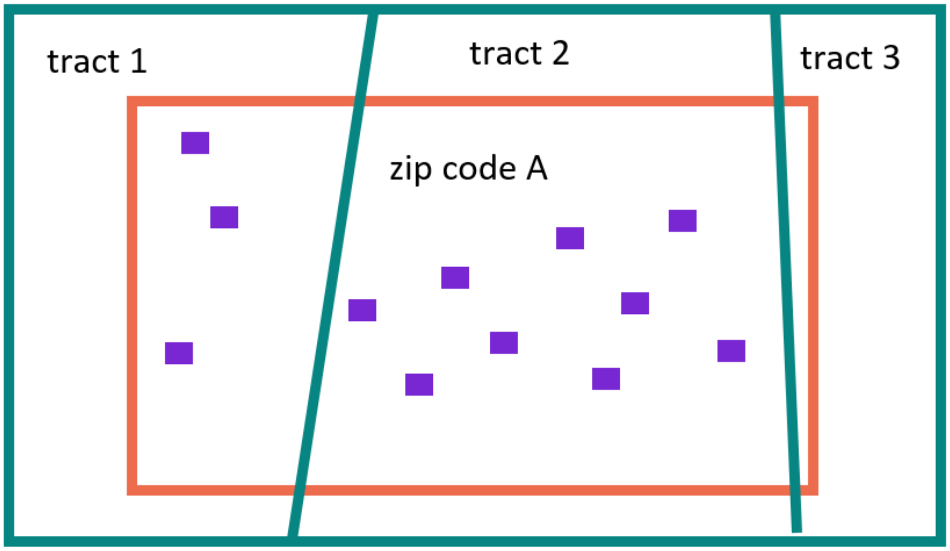
Census Tract and Zip Code Conversion

Zip code boundaries did not always align with census tract boundaries (see Figure 3). Because of this, it was necessary to use a decision rule for converting high-risk census tracts into associated zip codes. Of particular concern were cases where a zip code overlaps with part of a census tract and the resulting area of overlap contains no houses. When we converted census tracts into zip codes, we excluded these non-residential overlaps since they were not relevant to our analysis. Specifically, we counted the number of housing units in each zip code and in its overlapping census tract(s). We then calculated the proportion of total housing units in the zip code that were in the area of overlap. If the proportion was zero, we considered the overlap as non-residential and excluded it from the conversion. This novel solution greatly enhanced the precision of tract-to-zip conversions.

Figure 3. 2019 Zip code and census tract boundaries

Figure 4 shows a hypothetical situation where zip code A (orange) overlaps with three census tracts (dark teal). There are 12 housing units (purple rectangles) in zip code A. Three housing units are in both tract 1 and zip code A, nine housing units are in both tract 2 and zip code A, and no housing units are in the overlap of tract 3 and zip code A. Therefore, the corresponding proportion for each overlap is 0.25 (=3/12), 0.75 (=9/12), and 0 (=0/12).

Figure 4. Example of overlapping zip code and census tracts



Weighting

According to the U.S. Environmental Protection Agency, about 24% of homes built in 1960-1977 contain lead-based paint, whereas 69% of homes built in 1940-1959 and 87% of homes built before 1940 contain lead-based paint.[1] To incorporate the prevalence of lead hazards by housing age in weights, we applied the weights listed in Table 5 below when we conducted a sensitivity analysis (see results in Table 11, S3 in the full report).

Table 5. Weights for different old housing ages

Housing age	Weight
1960-1979	0.24/(0.24+0.69+0.87)
1940-1959	0.69/(0.24+0.69+0.87)
Before 1940	0.87/(0.24+0.69+0.87)

Note: We assumed that homes built in 1978-1979 were similar to those built in 1960-1977 and applied the same weight considering that the census data only provided relevant information in 10-year intervals.

For the composite risk score method using old housing and poverty with weights, the state of Washington applied weights of 0.58 and 0.42 for old housing and poverty, respectively.[2] We built upon the Washington approach by incorporating the prevalence of lead hazards by housing age in weights (See results in Table 11, S4 in the full report). Specifically, we applied weights for each risk factor as follows in Table 6.

Table 6. Weights for poverty and different old housing ages

Risk factor	Weight
Housing age: 1960-1979	0.58*0.24/(0.24+0.69+0.87)
Housing age: 1940-1959	0.58*0.69/(0.24+0.69+0.87)
Housing age: before 1940	0.58*0.87/(0.24+0.69+0.87)
Poverty	0.42

Note: We assumed that homes built in 1978-1979 were similar to those built in 1960-1977 and applied the same weight considering that the census data only provided relevant information in 10-year intervals.

Citations and Endnotes

[1] U.S. Environmental Protection Agency. (2021). *Protect your family from sources of lead.* Retrieved from https://www.epa.gov/lead/protect-your-family-sources-lead.
[2] In Washington's report, the weights were calculated using data from the National Health and Nutrition Examination Survey reported in the CDC's 2013 MMWR "*Blood Lead Levels in Children Aged 1-5 Years – United States 1990-2010.*"
https://www.doh.wa.gov/Portals/1/Documents/Pubs/334-383.pdf

Suggested citation:

Zan, H., Azuma, J., Kim, J., DeBaryshe, B. (2022). *Hawai'i Childhood Lead Poisoning Prevention Program risk system evaluation*: *Technical report*. Honolulu, HI: University of Hawai'i, Center on the Family.

CENTER ON THE FAMILY

UNIVERSITY of HAWAI'I MĀNOA