LEADRiskINDEX

LEAD RISK INDEX METHODOLOGY

Lead is highly toxic, especially to young children under 6. It can harm a child's brain, kidneys, bone marrow, and other body systems. The most common source of lead exposure for children today is lead paint in older housing and the contaminated dust and soil it generates. [1] To eliminate lead poisoning problems for children, it is critical to identify communities with high lead poisoning risk, so that prevention and treatment resources can be allocated to where they are most needed. To accomplish this, we modeled lead risk metrics by the demographic risk factors of gender, race/ethnicity, age, poverty status, old housing stock, as well as county level blood lead test data using Bayesian hierarchical regression models. The model data includes the 2005–2010 National Health and Nutrition Examination Survey (NHANES) [2], and CDC's state surveillance data on Blood lead tests. [3] We then applied the models to zip code demographics to estimate lead risk within zip codes. We aggregated the zip code estimates of lead risk to higher geographical levels—including federal and state legislative districts, metropolitan statistical areas (MSAs), counties and states—through crosswalk tables.

PART 1: Blood Lead Level

Outcome Variables

We used the 2005-2010 NHANES's Blood Cadmium and Lead Laboratory file to obtain blood lead measures (variable name **LBXBPB**) on the survey subjects, and derived the following lead risk metrics as outcome variables. [2]

1. **Percent Population with Blood Lead ≥ 10 µg/dL.** 10 µg/dL is frequently used as the cut point for elevated blood lead level (EBLL) in many studies, _e.g._, CDC's "Surveillance for Elevated Blood Lead Levels Among Children - United States, 1997—2001". [4]

2. **Percent Population with Blood Lead ≥ 5 µg/dL.** Recent studies have suggested many adverse effects in children with blood lead level (BLL) under 10 µg/dL.[5] We use 5 µg/dL as a "new" EBLL standard to capture a more complete population affected by lead poisoning.

3. **Average Blood Lead (µg/dL)**

Predictor Variables

Predictor variables were the individual level risk factors of gender, race/ethnicity, age, and poverty status from NHANES. An initial analysis was performed to impute missing

values for poverty status (poverty-income ratio, or PIR), because the surveys were missing these values for many survey subjects (around 16% of males and 7% of females).

Bayesian Model

We adopted multilevel Bayesian hierarchical regression models to incorporate individual-level risk differentials by gender, race/ethnicity, age, poverty status from the NHANES and county level data. The models also included a random effect term to account for extra variability in the data. Conceptually, the lead risk was modeled as a function of the individual-risk effects and random effect (Equation 1).

$$P = F(x_1, x_2, x_3, x_4, s_{random})$$ (1)

P : Probability of a person with high lead risk, e.g., BLL ≥ 10 µg/dL

x_1 : Age x_2 : Gender x_3 : Race/ethnicity x_4 : Poverty

s_{random}: random effect

On top of the individual risk model, we fitted a county level model using county level blood lead surveillance data from CDC [3]. The data year varies from state to state, with the most common years being 2006 and 2007. The final zip code estimates depend on outputs from both the individual level risk model based on NHANES and the county model based on CDC's blood lead surveillance results.

The Bayesian models were fitted in WinBUGS 1.4. [6]

Making estimates for zip codes and other geographies

After we acquired the mathematical equations from the fitted Bayesian models, we applied them to a zip code level dataset extracted from U.S. Census 2010 [7] that included all predictor variables (poverty status data for zip codes are not available from Census 2010 yet so we used Census 2000 [8] numbers instead) to calculate lead risk metrics within individual zip codes. The lead risk metrics were calculated for children under 6 by the following demographic groups (Overall, Male, Female, White, Black, Hispanic, and Other races) for every zip code in the United States.

We also developed crosswalk tables between zip codes and higher level geographical areas, such as MSAs and states (Table 1). We then calculated lead risk estimates for the higher level geographical areas from the zip code estimates. Equation 2 shows how "Percent Population with Blood Lead ≥ 10 µg/dL" was calculated.

$$PV = \frac{\sum_i (pv_i \times w_i)}{\sum_i w_i}$$ (2)

PV : Percent of population with Blood Lead ≥ 10 μg/dL in a higher level geographical area, e.g., state

pv_i : Percent of population with Blood Lead ≥ 10 μg/dL in zip code *i* within the higher level geographical area

w_i : Proper weight measure (e.g., population) for zip code *i* relative to other zip codes within the higher level geographical area

We also estimated the total number of people with elevated blood lead (≥ 10 μg/dL and ≥ 5 μg/dL) within each zip code by multiplying zip code level percentages by the zip code's population according to Census 2010. We calculated population with elevated blood lead for higher level geographical areas, such as MSAs and states, by aggregating zip code estimates through crosswalk tables between zip codes and the higher level geographical areas (Table 1). Equation 3 illustrates the aggregation:

$$N = \sum_i n_i \qquad (3)$$

N : Number of people with elevated blood lead within a higher level geographical area, e.g., state

n_i : Number of people with elevated blood lead in zip code *i* within the higher level geographical area

Table 1: Illustration of a Geographic Crosswalk Table (Zip Code to State)

Zip Code	State
19003	PA
19004	PA
19005	PA
…	…

Part 2: Old Housing Stock

Concentration of old houses

Because lead-based paint from old houses is the most common source of lead hazard, it makes sense to map the concentration of old housing stocks. Here, we mapped "*Percent of houses built before 1980*" for zip codes, which was calculated from US Census 2000 housing data [8]. (Census 2010 housing data is not yet released). The reason to use housing before 1980 is that lead-based paint was commonly used in houses until it was banned in 1978 [1], and 1980 was the closest cut off year that the Census provided housing data for zip codes.

Children living in old houses

In addition to mapping percent of old houses built before 1980, we took one step further to estimate "*number of children under age 6 living in pre-1980 housing*" for each zip code. Zip codes with high values of this metric should be where public health

interventions are focused. The formula we used to estimate this number is straightforward:

$$P_i = C_i \cdot H_i$$

P_i : Number of children under age 6 living in pre-1980 housing in zip code i

C_i : Population of children under 6 in zip code i, based on Census 2010 **[7]**

H_i : Percent of houses built before 1980, based on Census 2000

Mapping

The national map of lead risk metrics was created with ArcMap 9.2. Values were classified by the natural-breaks (Jenks) method, which identifies break points by creating the breaks that best group similar values and maximize the differences between levels.[9]

Notes

1. President's Task Force on Environmental Health Risks and Safety Risks to Children (2000), *Eliminating Childhood Lead Poisoning: A Federal Strategy Targeting Lead Paint Hazards,* http://www.cdc.gov/nceh/lead/about/fedstrategy2000.pdf.
2. National Health and Nutrition Examination Survey, *Questionnaires, Datasets, and Related Documentation* (National Center for Health Statistics, US Centers for Disease Control and Prevention), http://www.cdc.gov/nchs/nhanes/nhanes_questionnaires.htm (accessed Dec 2011).
3. State lead surveillance data, CDC http://www.cdc.gov/nceh/lead/data/state.htm (accessed Jan 2012).
4. US Centers for Disease Control and Prevention, 2003, *Surveillance for Elevated Blood Lead Levels Among Children --- United States, 1997—2001,* http://www.cdc.gov/mmwr/preview/mmwrhtml/ss5210a1.htm.
5. Tina Adler, *Questioning Lead Standards: Even Low Levels Shave Points off IQ* , Environ Health Perspective, 2005, http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1257668/.
6. Lunn, D.J., Thomas, A., Best, N., and Spiegelhalter, D. (2000), *WinBUGS—a Bayesian modeling framework: concepts, structure, and extensibility,* Statistics and Computing, 10:325–337.
7. US Census Bureau, Census 2010, Summary File 1, http://www2.census.gov/census_2010/04-Summary_File_1/ (accessed December 2011).

8. US Census Bureau, Census 2000, Summary File 3, http://www2.census.gov/census_2000/datasets/Summary_File_3/ (accessed December 2011).
9. ESRI Support Center, "Natural Breaks (Jenks)," in ArcGIS 9.2 Desktop Help, http://webhelp.esri.com/arcgisdesktop/9.2/index.cfm?topicname=natural_breaks_%28jenks%29 (accessed December 2011).