**Data Dictionary**

**Zartarian et al., 2024 – A U.S. Lead Exposure Hotspots Analysis**

The following content is a data dictionary for the key fields of the underlying GIS data layers used to generate Figures 1 and 2 of Zartarian et al., 2024. This is also a companion document to the GIS file geodatabases available for download through the data link in Zartarian et al., 2024 and found in the “Figures\_1\_2\_02082024” zipped folder.

Please note: (1) the data layer labeled “Getis\_Ord\_Gi\_and\_Top\_20\_FULL\_UNDERLYING\_DATASET” in the “Figure\_1.gdb” file also contains the data necessary to recreate the summary statistics shown in Tables 1 and 2, and (2) the data for EPA ORD Random Forest model version 2 are located in the “Supplement B” zipped folder.

**Figure 1:**

Note: The fields listed here are in order of their appearance in the GIS layer attribute tables (left to right) found in the Figure\_1.gdb file. All other fields not defined in this document for Figure 1 can be ignored.

**AveDLDPN** – US EPA EJScreen 2017 Lead (Pb) Paint Environmental Justice (EJ) Index values (census tract averages)

**TRACT** – census tract identification number (text format)

**Geography** – full census tract, county, and state names

**BLL\_ug\_dL\_** – US EPA Office of Research and Development (ORD) modeled blood lead level (BLL) data based on the Schultz et al., 2017 model

**STATE** – state code number

**COUNTY** – county code number

**Total\_05** – the total number of children ages 0 to <6 years old per 2010 Census

**GEOID** – census tract identification number (number format)

**Vox** – Vox Pb exposure risk score index values

**RF\_pred** – US EPA ORD Random Forest regression prediction model version 1 values

**Pct\_DPI** – US Department of Housing and Urban Development (HUD) Deteriorated Paint Index (DPI) values

**getis\_index5\_03242023** – the total number of indices/models (i.e., EJScreen 2017 Pb Paint EJ Index; EPA ORD – Schultz et al., 2017 model; Vox Pb exposure risk score; EPA ORD Random Forest regression prediction model version 1; and HUD DPI) that selected the respective census tract using the Getis-Ord Gi\* geospatial cluster method (values range from 0 to 5, where 0 equals no indices/models and 5 equals all indices/models)

**top20\_index5\_03242023** – the total number of indices/models (i.e., EJScreen 2017 Pb Paint EJ Index; EPA ORD – Schultz et al., 2017 model; Vox Pb exposure risk score; EPA ORD Random Forest regression prediction model version 1; and HUD DPI) that selected the respective census tract using the top 20 percentile (i.e., 80th-100th percentile) method (values range from 0 to 5, where 0 equals no indices/models and 5 equals all indices/models)

**Figure 2:**

Note: The fields listed here are in order of their appearance in the GIS layer attribute tables (left to right) found in the Figure\_2.gdb file. All other fields not defined in this document for Figure 2 can be ignored.

**TRACT** – census tract identification number (text format)

**Geography** – full census tract, county, and state names

**STATE** – state code number

**COUNTY** – county code number

**Total\_05** – the total number of children ages 0 to <6 years old per 2010 Census

**GEOID** – census tract identification number (number format)

**getis\_index5\_03242023** – the total number of indices/models (i.e., EJScreen 2017 Pb Paint EJ Index; EPA ORD – Schultz et al., 2017 model; Vox Pb exposure risk score; EPA ORD Random Forest regression prediction model version 1; and HUD DPI) that selected the respective census tract using the Getis-Ord Gi\* geospatial cluster method (values range from 0 to 5, where 0 equals no indices/models and 5 equals all indices/models)

**top20\_index5\_03242023** – the total number of indices/models (i.e., EJScreen 2017 Pb Paint EJ Index; EPA ORD – Schultz et al., 2017 model; Vox Pb exposure risk score; EPA ORD Random Forest regression prediction model version 1; and HUD DPI) that selected the respective census tract using the top 20 percentile (i.e., 80th-100th percentile) method (values range from 0 to 5, where 0 equals no indices/models and 5 equals all indices/models)