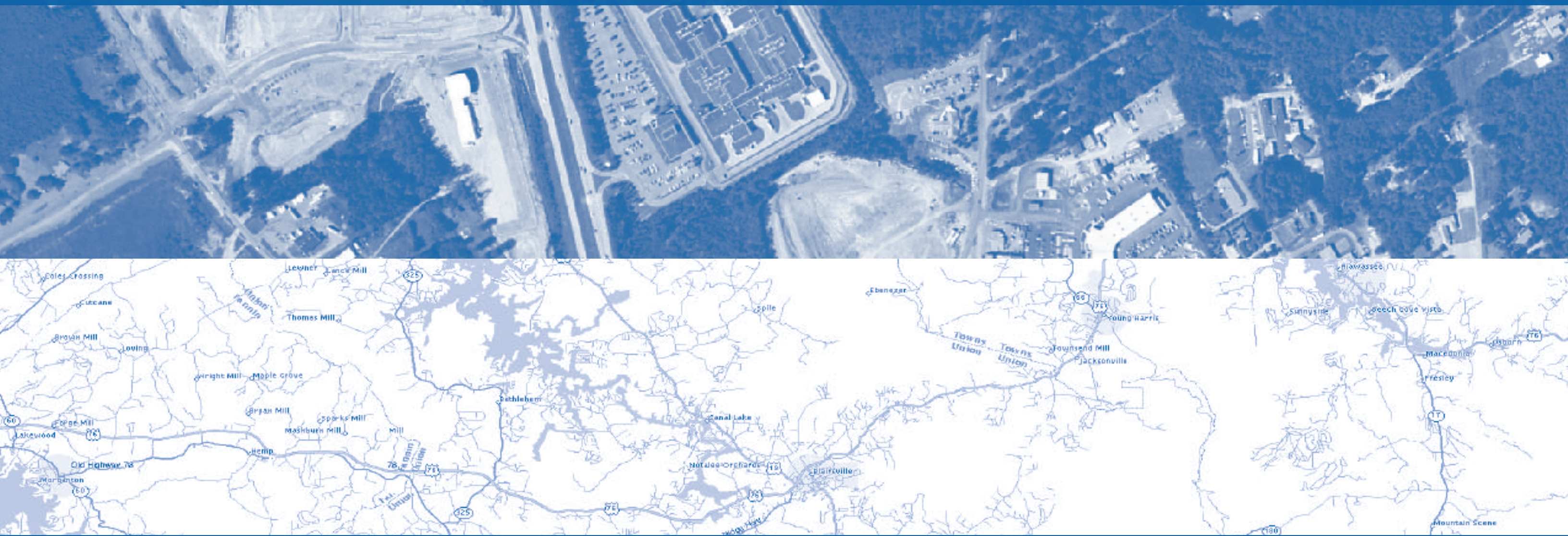
Using GIS to Assess and Direct Childhood Lead Poisoning Prevention

Guidance for State and Local Childhood Lead Poisoning Prevention Programs



DEPARTMENT OF HEALTH & HUMAN SERVICES · USA

CDC
SAFER · HEALTHIER · PEOPLE™

Developed by the Childhood Lead Poisoning Prevention Program
Geographic Information System Workgroup

December, 2004

Using GIS to Assess and Direct Childhood Lead Poisoning Prevention

Guidance for State and Local Childhood Lead Poisoning Prevention Programs

Developed by the Childhood Lead Poisoning Prevention Program Geographic Information System Workgroup

December, 2004

The Childhood Lead Poisoning Prevention Program Geographic Information System Workgroup

Gerald B. Curtis, BA, GISP, Centers for Disease Control and Prevention

John T. Braggio, PhD, MPH, Maryland Department of Health and Mental Hygiene

Frida Fokum, PhD, Illinois Department of Public Health

James R. Roberts, MD, MPH, Medical University of South Carolina

Robert Scott, PhD, Michigan Department of Community Health

Forrest Staley, MUP, Georgia Department of Human Resources

Joseph Sweatlock, PhD, DABT, New Jersey Department of Health and Senior Services

Richard Tobin, MS, MPA, Philadelphia Department of Public Health

CDC

Centers for Disease Control and Prevention™

ACKNOWLEDGEMENTS

We wish to thank the following individuals for their valuable input:

Centers for Disease Control and Prevention

Mary Jean Brown, ScD, RN

James B. Holt, PhD, MBA

C. Virginia Lee, MD, MPH, MA

Pamela Meyer, PhD, MSPH

Other

Jerome E. Dobson, PhD – University of Kansas Department of Geography

Contents

Purpose of These Guidelines

These guidelines were prepared to help new lead epidemiologists quickly learn how to use geographic information systems (GIS) mapping technology to assess and direct childhood lead poisoning elimination efforts.

Eliminating elevated blood lead levels (BLLs) $\geq$10 micrograms per deciliter (μg/dL) or higher among young children is a Healthy People 2010 goal. The Centers for Disease Control and Prevention (CDC) Lead Poisoning Prevention Branch (LPPB) is committed to attaining that goal. The adverse health effects of lead and the sources of lead are well documented. Lead-based house paint and the dust and soil it contaminates are the most common high-dose sources of lead exposure for young children in the United States today.

The challenge for public health practitioners and policy makers is to prevent childhood lead poisoning, not just react to it (1). GIS technology is a powerful tool that can be used to effectively target lead poisoning preventive interventions. The addresses of old housing units can be geocoded (geographically located) to identify areas where children at risk for lead poisoning live. Interventions can then be directed to those areas and specific properties to address lead hazards.

These guidelines will focus on mapping applications, although GIS also can be used for statistical modeling to predict risk for lead exposure (2). Examples are provided of how GIS mapping technology can use blood lead screening, tax assessor (property), and U.S. census data to develop and improve preventive interventions, especially primary prevention (before children are poisoned).

Who is at Risk for Lead Exposure?

Lead poisoning is a preventable environmental disease in children (3). Children under the age of 6 years commonly put things in their mouths that they find around them. This hand-to-mouth behavior increases the young child's risk for ingesting lead-contaminated dust and soil. CDC estimates that 434,000 children have BLLs >10 μg/dL, CDC's level of concern (4). Children at greatest risk for lead poisoning are those whose families are poor and live in substandard housing built before 1950. These children tend to be African American or of Hispanic ethnicity.

Since the 1970s, policies have been implemented to limit the use of lead in products such as gasoline, food and drink cans, solder in pipes, and residential lead paint. Those polices have resulted in dramatic reductions in BLLs for children and adults (5, 6, 7). They have also reduced lead in our environment. Today, the most common high-dose source of lead exposure for young children in the United States is leaded house paint. That includes the dust and soil that becomes contaminated as the paint deteriorates (8, 9). Although lead-based paint was banned for residential use in 1978, millions of properties built before that time are still lead hazards. House paint used before 1950 contained up to 50% lead by weight (10).

In the years between 1950 and 1977, manufacturers voluntarily reduced the concentration of lead in paint. Consequently, even though there is lead-based paint in nearly all houses built before 1977, houses built before 1950 place children at considerably higher risk (11).

At the end of the 20th century, an estimated 38 million housing units had lead-based paint, and 24 million of these had significant lead-based paint hazards. Low-income families (<$30,000/year) with children younger than 6 years of age occupied 1.2 million of those hazardous units (12).

A 1991–1994 national survey showed the prevalence of children with BLLs >10μg/dL varied by age of housing: 8.6% for children living in houses built before 1946, 4.6% for those living in houses built from 1946–1973, and 1.6% for those living in houses built after 1973 (13). Children who live in old housing units and are poor are at higher risk for having elevated BLLs than are children from higher income families. For example, the prevalence of elevated BLLs among children living in homes built before 1946 was 16.4% for those from low-income families compared with 4.1% and 0.95% among those from middle- and high-income families, respectively (14). Studies show that property valuation or the assessed values of houses (tax assessor data) can be used as a proxy measure of the structures' condition. Children living in lesser-valued houses were found to be at greater risk of having elevated BLLs, even after controlling for the age of the house (15). Rental units also have been linked with higher prevalence of lead poisoning. That may reflect a greater likelihood that paint in old rental housing is deteriorated and becomes accessible to children (16).

How Can GIS Help?

With GIS, maps can be created that show the location and age of every housing unit in an area. These maps can include information on other risk factors for lead poisoning, including population distributions, housing conditions, and BLLs of resident children during a given period. This information can be used to show the relationship between housing units and risk factors. GIS visually presents these geospatial and temporal relationships in data that have many uses:

- To identify where high-risk children live;
- To assess screening penetration among high risk groups;
- To obtain a better understanding of changes over time in areas where children at high risk live;
- To evaluate the impact of our targeted screening and other intervention efforts and improve them; and
- To identify those housing units responsible, over a period of years, for multiple cases of childhood lead poisoning (17).

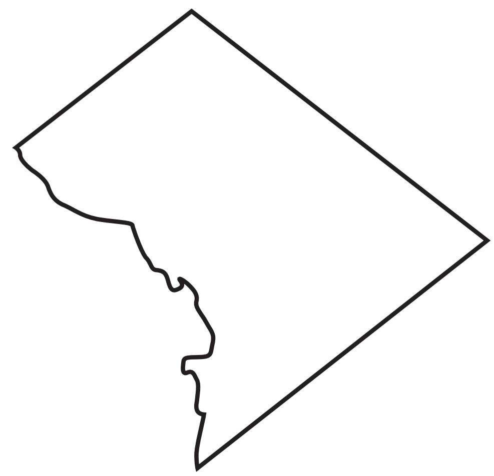
What is GIS?

GIS is a computer-assisted system for the acquisition, storage, analysis, and display of geographic data (18). GIS software allows the user to create maps that display spatially related or geographically based data. Data representing geographic features (landscape elements) can be visually displayed as points, lines, and polygons (19). Each type of feature element in a GIS is contained in its own feature layer. A layer can only contain one type of feature. For example, a map displaying a city with streets and locations of individual residences, is composed of three feature layers:

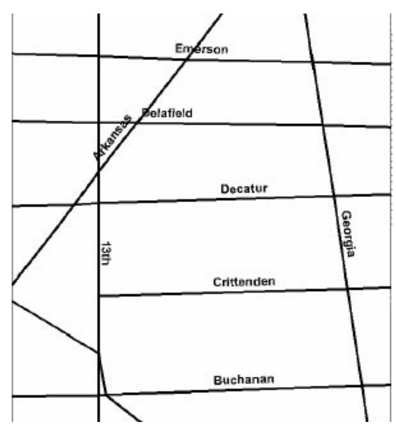
- One polygon layer denoting the area boundaries of a city,
- One line layer to denote streets,
- One point layer denoting the location of individual residences.

In a GIS, every layer's geometric relationship to all other layers is prescribed through a process called topology (not to be confused with topography). Topology represents angular relationships (order, adjacency, etc.) that remain constant regardless of map distortion.

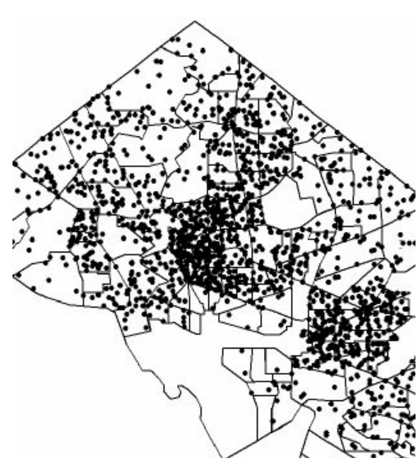
Polygon



Lines



Points



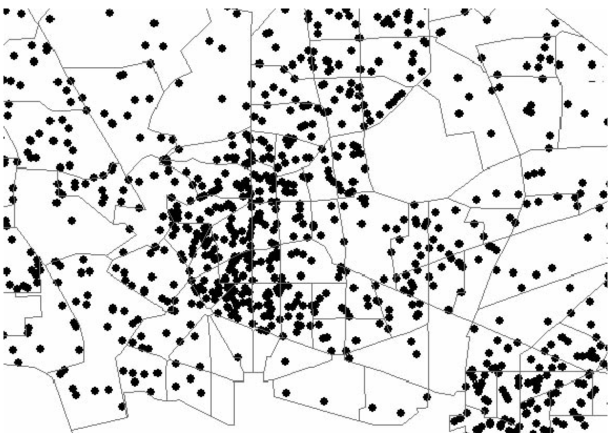
Any record containing geographic information can be mapped. The geographic data may represent any size geographic area (e.g., a single housing unit, a building, a census block or tract, municipality, ZIP code, county, state, etc.). Address Geocoding is the process whereby specialized software matches an address against a database of standardized addresses and assigns unique map coordinates for location (i.e., latitude and longitude). Some software contain “tables of alternate street names” that allow correction for streets having different names over time. Once an address is geocoded, it can be added to a GIS for spatial analysis (19). Geocoding allows us to establish relationships to other geographic identifiers (ZIP code, census area, municipality, block group) and query the data (20).

The class of maps used in epidemiology are thematic maps. Thematic maps are used to illustrate the distribution of a single attribute or the relationship between attributes. Thematic maps may be qualitative or quantitative. Qualitative thematic maps show the spatial distribution or location of a particular type of feature. Quantitative thematic maps display the spatial aspects of numeric data usually of a single variable. Two major techniques used in thematic map design are dot maps and choroplethic maps.

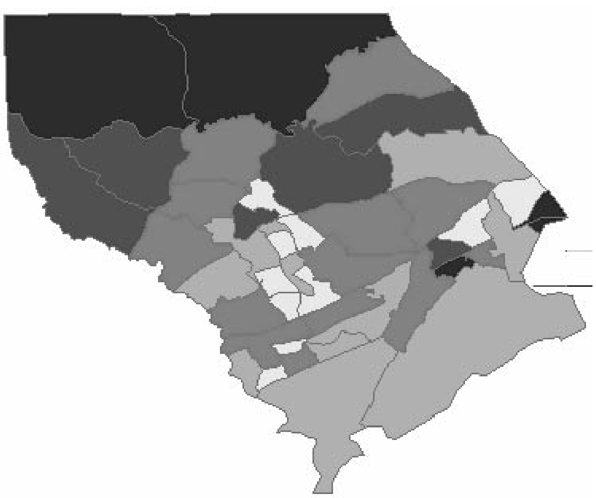
Dot mapping is used to map discrete geographical phenomena. The purpose of these maps are to show the spatial density of the feature being mapped. A key point to remember in this type of map is that the symbol form does not change. The number of symbols from place to place changes in proportion to the number of objects that are represented. There are two classes of dot maps: the dot equals one object or the dot equals more than one object. For example, to represent the distribution of housing density within census tracts one could use one dot to represent one household or one dot to represent 100 households, allowing the observer to compare the density between census tracts.

Choropleth mapping is used to show enumeration data. Choropleth maps use shading or colors for statistical or administrative areas to show the values of a selected variable within the area selected. Use choropleth maps when the data is discrete and has been collected so that it can be assigned to definite enumeration units such as census tracts. Choropleth maps may be classed or classless depending on their purpose. Classless choropleth maps tend to be complex and use a continuum of tones to display the unique values in the dataset. A classed choropleth map groups data into different classes. Each grouping is then represented by a different color. The number of classes determines how detailed the map distribution will be. For example, if one wanted to map areas of concern, a progressive color scheme can represent values from high to low (dark blue to light blue) (20).

Dot-Density



Choropleth



The symbology used to picture the data displayed reflects the classification of the map (layer) features. For example, one can represent the residences where lead poisoned children live by manipulating the shape, size, and color of a symbol. Black dots could represent housing and red triangles could represent houses of children with elevated blood lead levels.

A major advantage of GIS technology is the dynamic linkage between attribute and spatial data. Attribute data is the descriptive data within the database that is associated with a particular element or record. For example, a typical demographic data table would have attribute information such as the person's first and last name, date of birth, race, ethnicity, and assigned personal identifier number for each person (record) in the dataset. In contrast, a laboratory test result table would have columns for assigned personal identifier number (person from whom the sample was taken), test type, sample type, sample collection date, analysis date, and result. To map attribute data such as blood lead levels, the data table must have some type of spatial reference variable such as address or county in the database. Most databases also contain a data table with address information that is linked to each person. Thus, all the attribute information in the database can be linked to an address, which can be geocoded and placed on a map. When information is updated in the database, the dynamic linkage automatically updates the information on the map. Selection of specific attribute records will be highlighted on the map and conversely, selection of specific map features will be highlighted in the attribute data (21).

Spatial data contains information about the location and shape of geographic features. For example, one could map the location of all children who had a blood lead test during the 2003 calendar year. Different symbols could be used to represent those children who had a test result greater than or equal to 10 μg/dL. This picture can be further enhanced by using spatial overlay, a feature of GIS that superimposes a layer from another GIS or attribute database to create a composite map image. One could, for example, superimpose a layer of pre-1950 housing. This composite map would show any relationship between elevated blood lead level and age of housing. The composite map also may identify high-risk neighborhoods where little or no screening has occurred.

Data Sources

Data for GIS mapping are available from many sources, ranging from the CDC, U.S. Postal Service, and U.S. Census Bureau, to state and municipal offices. Keep in mind, however, that quality can be a problem, no matter what the source. (Appendix D discusses cleaning and preparing data for GIS use.)

Child Blood Lead Surveillance

Each CDC-supported childhood lead poisoning prevention program (CLPPP) collects surveillance data as part of its comprehensive program. These data contain information on all children tested and reported, and appropriate case management activities. CDC requires CLPPPs to collect and submit a core set of data fields (Appendix B). Included in the information reported to CDC are several geographically based fields, such as city, ZIP code (ZIP + 4), county Federal Information Processing Standards (FIPS) code, and census tract. An address identifier number links these geographic fields to other attribute data (e.g., blood lead test results, environmental activities, etc.). The state or local program may collect more than the "core" data required by CDC.

Census

Census data are widely used in health research and public health practice. For Census 2000, the U.S. Census Bureau distributed two forms—a short form and a long form. The short form was sent to every household. It included seven questions for each household: name, sex, age, relationship, Hispanic origin, race of each household's member, and whether the housing unit was owned or rented. About 17% of the households received a much longer questionnaire. It included the seven questions, plus questions about ancestry, income, mortgage, and age of the housing unit (Appendix C). Long form data is available from the Census as Summary File 3, which is the only file to contain "age of housing" information. The Census 2000 information is available in the following geographic levels: nation, state, county, place (cities and towns), census tract, block group, and block. Although the smallest geographic area is the block, for confidentiality purposes, detailed data is only available to the block group level. Block groups are collections of blocks, and census tracts are collections of block groups. Populations for census tracts are usually between 2, 500 and 8,000 and block groups between 600 and 3,000 (detailed descriptions of census levels can be found at http://www.census.gov). Census data related to childhood lead poisoning is available from CDC's Lead Poisoning Prevention Branch (see the How CDC Can Help CLPPPs section below for further details).

Tax Assessor

Tax assessor data (real estate property data) is a source of detailed housing information (22). County and municipal government offices, known for our purpose as the "tax assessor office," collect detailed information on each parcel of land (property) as a record of real estate transactions and to valuate property for taxes. Each parcel of property in a county has its own unique number—the "assessor's parcel number" (APN)—and every county has a unique numbering method. A tax assessor database can be linked to a geographical information system to reveal the actual street location of housing. Tax assessor data is public information, meaning every member of the public has a right to view these data.

There can be more than 400 variables in the real estate database, including whether the property is residential or commercial. Residential property type is further divided into categories such as single-family, apartment, or condominium. Each record can contain hundreds of variables, including sales price, date of sale, property address, and owner address (street number, street name, street type, ZIP code, and county). Additional information usually includes subdivision name, number of housing units (apartments), number of rooms, number of stories, whether owner or renter occupied, and most important, the year the structure was built. The database may even include information on renovation history. Address information is listed in two categories: owner information and property information. Owner information is the mailing address provided by the owner to the assessor for tax billing purposes. Property information is the actual property location. If the owner lives at the property, the mailing and property address should be the same.

Real estate data is available commercially but is costly. Tax assessor data is available at a little cost and is relatively accurate and updated frequently. As a result, it is recommended that local tax assessor data be used, when it is available.

Getting Started

GIS Office

Many counties and municipalities have a GIS office (usually in the planning department). CLPPPs are encouraged to seek out their help, especially through a personal visit. Even if the tax assessor does not have a GIS property map nor is aware of the existence of a map, a visit to the county or local GIS office may reveal the existence of one. The GIS staff also know the “lingo” of the GIS world and will probably be happy to cooperate. GIS offices can provide additional information that the tax assessor may not be knowledgeable about (resources, persons to contact, current GIS files, etc.).

Obtaining Tax Assessor Data

Tax Assessor

Plan to Meet With the Tax Assessor

In the United States, almost all property data is maintained at the county level, except for a few large cities. There are 3,153 counties in the United States. Historically tax assessors have maintained property data in a paper format. Today, more and more tax assessors are converting their records into an electronic format. If the property data is available in an electronic format, a trip to the tax assessor office will be necessary. This is important so that the tax assessor can be provided with a clear understanding of what is required. In return, it helps to learn what the tax assessor can and cannot provide. While in the area, visit the local county and city GIS offices, the library (main branch), and the local health officer. It is a good idea to set up appointments with a planned agenda for each stop. In all, expect at least a 1-day visit.

Preparing to Visit the Tax Assessor Office

To prepare for the visit, learn about the differences between census data and tax assessor data. For example, "owner address" is where the tax bill is sent, "situs address" is the property location, and "assessor parcel number (APN)" is the unique number assigned to each property.

If the tax assessor's office can supply GIS data, tell them which map projection is desired (usually provided as ESRI ArcGIS shapefiles). A key issue to note in requesting map data is that most GIS data are kept in certain projections. To correctly align the data of interest requires that all the layers are projected the same. A map projection is the mathematical transformation process to systematically arrange the earth's spherical geographic coordinate system onto the planar (flat) map surface. Keep in mind that the 3-dimensional Earth cannot be projected onto a flat surface without distortion of either area, shape, distance, or direction. Different projections minimize different types of distortions. The distortions found in different projection systems are more extreme for small scale maps, such as those of the world or of a continent rather than a large-scale map of a county. Be sure to ask the tax assessor what projection the data will be in. If one receives a map layer in a projection that is different from the layer one is working in, the tax assessor's data will not be aligned properly to the other layers in one's map (i.e. census tract boundaries). However, the newest versions of GIS software will allow users to re-project layers to allow for alignment.

If users need to measure areas in a flat map, select an equal-area map projection such as Albers Equal Area Conic and

Equal-Area Cylindrical on which the area relationships are maintained. These projections will minimize distortion of area although the shapes of the objects shown appear quite differently than they do on a globe. Conformal projections such as Lambert Conformal Conic and Mercator preserve the angles around points to preserve the shapes of small areas even though shapes for larger regions such as continents may be distorted. Conformal projections do not allow for accurate measurement of areas. In the United States, states use State Plane Coordinate System as the basis for their mapping projects. The SPCS is a rectangular grid reference system that was developed in the 1930s for simplifying work of surveyors. States are of small enough area so that the distortion is minimal. There are 3 projections that are used for the SPCS to map the states: Lambert conformal conic for states with long east-west dimensions, transverse Mercator for states with long north-south dimensions, and oblique Mercator in parts of Alaska.

All geodata files should be accompanied with corresponding spatial metadata files. The metadata file includes relevant information about the spatial attributes of the data (datum, projection, resolution, etc.) needed to make it compatible with other spatial data files. Be sure you obtain the appropriate metadata with any spatial files received from an outside source.

If county level census data of pre-1950 housing is available, prepare it in a format that can be easily understood. A map of the county with its towns and cities, showing the distribution of old housing by block group (where each dot represents a certain number of housing units), will immediately reveal all the data in a picture; and a picture is worth a thousands words. Prepare the data at the lowest level possible (block group). If one has commercial data, print out appropriate data, and map the data (if possible).
Use these materials to demonstrate and explain more detailed information is needed. The tax assessor can also assess the completeness of the commercial data set. Bring copies of the lists containing the data one has, and the data one wishes to acquire.

Preparing for the Tax Assessor Meeting

Tax assessors are usually very receptive when their data can be used for positive purposes. Explain the reason for the visit, describe the project, and show what information has been collected so far. If commercial data has been acquired, share that with the tax assessor and show where there is missing information (i.e., "year built" may be missing on some early properties).

Present a list of the data one would like to have and be prepared to leave a copy of this request with the tax assessor. A list of fields one might find useful may be found in Appendix D. Most tax assessors will have the bulk of this information. At a minimum, request the following:

- owner information (street number, name, and type; city; state; ZIP code);
- situs information (street number, name, and type; city; state; ZIP code);
- number of units;
- number of stories;
- type of housing (single family residence, apartment, condominium, duplex, triplex);
- year built (the year construction was begun); and
- effective year built (the year construction was completed).

Stress that it is important to obtain information on properties with "year built" data, as well as information on properties that are missing "year built" data entries. Housing with missing "year built" data can be checked later against the 1950 City Directory (library resource) to see if they existed at that time. If the "year built" category is not in the database, ask the tax assessor if there is another way this information could be ascertained.

Ask the tax assessor if the data are in a GIS format known as a "parcel map." Parcel maps show property boundaries. If not, ask if there is anyone in the county working on a parcel map. Finally, state the format in which one wishes to receive the data (e.g. spreadsheet, database).

Tips

The tax assessor will explain what data can be provided and what is unique about the county's record system. The following are examples of county-specific information discovered during various visits to tax assessors:

- Many properties contain more than one house, but the properties are only assigned one APN.
- APNs might not be assigned consecutively and therefore cannot be used as a method to ascertain the "year built" when "year built" was missing from a record. (In some counties, however, APNs might be consecutive.)
- There are more municipalities (census places) in the county than listed in the U.S. census, which excludes census places that do not meet criteria, such as minimum population size.

- Most municipalities are represented by a single "tax district" and each property record includes the tax district. Properties in the same tax district will probably share the same "ZIP code." This can be useful when filling in missing information, such as ZIP codes.
- City boundaries in the assessor database differ from those in the U.S. Census Bureau's database (larger taxable areas).
- The "official" county street name file can be used to verify spellings, directions, and street number ranges.
- ZIP code information may be incomplete on property data, but almost never on owner data (because the tax assessor needs the "ZIP code" to send the owner of the property the tax bill!).
- Residential properties do not include apartments; apartments are listed as commercial property.
- The county may not have collected "year built" information until a few years ago. However, whether a unit was built before 1950 could be determined through the tax category "percent good," which is a method used to fairly tax older housing. Each time property was reassessed, the assessment is reduced by a few percentage points to compensate for inflation (increased value). Therefore, on the basis of accumulating percentage reductions, a property assessed at "percent good" below a certain level (75% in this case) was determined to be a pre-1950 building.

Analysis

Questions to Consider

Housing

- How can pre-1950 housing be located?
- How to know which housing has been remediated?
- Will remediation of contaminated housing before habitation prevent a child from being lead poisoned?
- How can the age and condition of housing in which children reside be determined?
- How can housing responsible for multiple poisonings be identified?

Screening

- Have all of the children at risk been screened?
- Have some groups of high-risk children been missed?
- Are all children living in poverty at risk?
- By screening only the Medicaid population, are some children at risk not tested?
- Which providers have poor screening rates for children at highest risk?
- Where are the residential locations of our cases.
- What is the sensitivity, specificity and positive predictive value of blood lead screening for children in the Women, Infants, and Children (WIC) nutrition program vs. Medicaid vs. poverty?

Geographic Units

Geographic units used in GIS include state, county, place, census tract, ZIP code, block group, block, and parcel. To target efficiently, it is helpful to use the smallest geographic level: the parcel. Parcel-level data can only be obtained from the tax assessor. If it is not feasible to obtain parcel-level information, other geographic units can be used but it is helpful to understand their limitations.

Limitations of ZIP codes

ZIP codes are widely used in planning because they are easily recognized and nearly always a part of any standard address; virtually every adult knows his or her own ZIP code. However, ZIP codes exist solely to help the U.S. Postal Service deliver mail more efficiently. Their main limitations are:

- They are unreliable for GIS mapping, because they cross state, county and municipal boundaries, and their boundaries change (23).
- New ZIP codes are added periodically.
- They were not created with any intention of indicating homogenous populations, and the number of ZIP codes vary

greatly by state, or other geographical boundaries.

- Data geocoded in one year cannot be added to data from future years because of changes that may occur in ZIP code boundaries from year to year. All data must be re-geocoded to prevent analysis errors. (24, 25). As long as the address, city and state are correct, one can geocode without reference to the old ZIP code, and the geocoding software will provide the updated ZIP code.

Number vs. Proportion of Pre-1950 Housing

When trying to identify the highest risk areas, it may help to consider both the proportion and the number of pre-1950 housing. Relying only on the proportion may miss high-risk areas that have a large number of old housing and many newer housing units. In the table below, 2000 census data reveal that the number of pre-1950 housing units decreased since the 1990 census from 27,508,653 to 25,815,821 (Table 1). As new housing units have been built, and the number of pre-1950 housing units has decreased, the percentage of pre-1950 housing units also decreased from 26.9% to 22.3%. This change was even more pronounced in fast-growing areas.

Table 1: Changes in Number of Children and Housing Units Between 1990 and 2000.

Children & Housing	1990 Census	2000 Census	2000–1990 Census Totals	2000–1990 Census Percent
0–5 Year Olds	21,951,100	23,140,901	1,189,801	5.4%
All Housing	102,263,678	115,904,641	+13,640,963	13.3%
<1950 Housing	27,508,653	25,815,821	–1,692,832	–6.2%
<1950%	26.9%	22.3%		

Thus, screening plans based on the number of high-risk housing units may be preferable to the proportion of old housing units because proportions will change over time, depending on housing growth in an area. Although the percentage of pre-1950 housing units drops over time, in many areas the condition of housing can become worse.

Some Potential Uses of Tax Assessor Data

Tax assessor data can be linked with other databases, such as the following:

- 1990 and 2000 censuses
- State/local CLPPP screening/ surveillance
- Patient and case management
- Birth data (vital records)
- Medicaid
- WIC, and other health department service databases (immunizations)
- Provider/HMO lists of patients younger than 6 years of age

Linking these data can provide the means to answer specific questions that can be used to improve and evaluate interventions. Below are specific examples of how databases can be linked with tax assessor data to improve state and local programs.

Multiple Case Source:

Identify housing units that are associated with multiple cases over a period of time. CLPPPs have targeted these high-risk residences as priorities for remediation. Jefferson County, Kentucky, successfully applied for HUD funding to remediate housing responsible for poisoning numerous children over a period of years (17). State/local CLPPP screening and surveillance data can provide the necessary attribute and spatial data to perform this type of spatial analysis.

Target Screening:

Using maps of pre-1950 housing, a CLPPP can target blood lead testing at the right population, e.g., children under the age of 6 years who live in pre-1950 housing. A map can give a clear and instant picture of where high-risk neighborhoods are located. Providers can be instructed to screen children who live in these high-risk neighborhoods. This type of analysis uses census and vital records (birth data) information.

Measure Program Effectiveness:

Programs can evaluate current activities. By mapping the locations of children who have been screened and the locations of pre-1950 housing, programs can consider the following issues: Where is current screening taking place? Are these areas the appropriate places to find lead poisoned children? And where are most cases found? For example, a map might reveal that the area where children are currently screened only partly overlaps the area where children are at greatest risk, on the basis of housing age. State/local CLPPP screening and surveillance data, combined with census data, can provide the information needed to answer the above questions. Including information from other sources such as Medicaid, WIC, and immunization registries, which also is associated with high risk for lead poisoning, will enhance this analysis.

Direct Mailings:

Data available from the tax assessor allow us to target direct mailings of literature about lead-based paint to residents living in housing built before 1950 and to the owners of rental property. One can further target families at risk by matching housing data (from the tax assessor or the census), with birth certificates listing home addresses. That would allow the CLPPP to limit mailings to only the parents of young children. The housing database also can be matched with WIC or Aid to Families with Dependent Children (AFDC) rolls, because poverty is considered a risk factor for lead poisoning.

Housing History:

Combining data from state/local CLPPP screening and surveillance data and case management can result in maps that can serve multiple functions in tracking houses:

- Indicate housing where lead-poisoned children have been identified and where homes await environmental remediation. Jefferson County, Kentucky, has successfully applied for HUD funding by identifying many properties responsible for multiple cases occurring over a period of years (17). This is a useful tracking tool for health departments.
- Indicate which houses have been remediated through renovation and are presumably no longer hazardous.
- Monitor other children who later live in the homes to confirm that remediation was done properly and test whether this is an effective method of primary prevention.
- Map addresses where owners have applied for renovation licenses. Improperly performed renovation and remodeling can create lead hazards when old paint is disturbed. The health department may wish to target special educational programs to residents of neighborhoods where these activities are common.
- Identify homes that are associated with more than one case of lead poisoning. These data can help a health department focus on neighborhoods of very high risk. It also justifies the case to Medicaid for reimbursement of environmental investigation services to prevent future lead poisonings in the same housing unit.

Limitations

Census

There are limitations to using census data:

- Although certain basic demographic and housing questions were asked for every person in the United States, detailed information on housing (including the year a structure was built) was collected on a subset sampled at a 1-in-6 rate (long form).
- Data on the year the structure was built are susceptible to errors of response and nonreporting because respondents must rely on their memories or on estimates by persons who have lived in the neighborhood a long time (26). Available evidence indicates there

is underreporting in the older year-structure-built categories, especially "built in 1939 or earlier" (27).

- Data are only available every 10 years. In the time from one census to the next, some units are lost through attrition, and the demographics of neighborhoods changes.
- Age of housing below the block group level is not available.

Because the U.S. Census Bureau maintains confidentiality of individual responses for 72 years, 1990 individual records will not be released until the year 2062. Thus, the exact address or age of individual housing units are not available. These data are reported as housing units, as it is based upon family units and does not differentiate between a single family home or an apartment. It can only identify a quantity of older housing per census area (block group, census tract). It is a snapshot in time; variables such as the number of housing units will change from one census to the next.

Tax Assessor

There are some limitations to using tax assessor data, too. Not all counties have tax assessor information in an electronic form. Post office (P.O.) boxes and rural route addresses can only be geocoded to the ZIP code level. The tax assessor usually has up-to-date ZIP codes. However, ZIP code areas are an administrative tool of the U.S. Postal Service, which periodically revises the boundaries as demographics change.

Problems are encountered when other data are used that contain data entered many years ago (e.g., out-dated ZIP codes). Care must be taken to insure that these data are updated or it will not be properly geocoded.

The U.S. Census Bureau and the tax assessor count housing differently. The U.S. Census Bureau counts each house or apartment (residential) as an equal unit. The tax assessor counts each building (residential) on a property as one taxable unit. Therefore, when tax assessor information is mapped, only one building at each address will be mapped. The number of additional units, such as apartments, is referenced in the property record.

How CDC Can Help CLPPPs

The Lead Poisoning Prevention Program at CDC can assist CLPPPs in developing a GIS approach to the primary prevention of lead poisoning in children. This assistance can be provided in Atlanta or in the grantee's state in the following areas:

- Offering GIS expertise—knowledgeable staff can provide GIS software training in ArcView 3.x, ArcGIS 8.x and 9.x, and Maptitude
- Training on use of the LPPB GIS website
- Sharing expertise on using grantee data
- Geocoding data, including cleaning and preparing data
- Facilitating networking with other GIS users in the LPPB GIS Workgroup
- Responding to special requests for data and shapefiles

In addition, the CDC staff can provide Census 2000 data containing 78 childhood lead poisoning related variables (age of housing, children age < 6 years, poverty levels, etc.) in the following formats (see Appendix E):

- Data files for each state at county, place, census tract, block group, and 5-digit ZIP code tabulation area (ZCTA5) levels
- GIS files for each state at county, place, census tract, block group, and 5-digit ZCTA5 levels
- Web-based downloads at www.cdc.gov/nceh/lead/lead.htm

Summary

The use of geographic information systems (GIS) has revolutionized the ability of programs to target interventions to children at greatest risk and evaluate past performance through visualization of their data. GIS can help us prioritize prevention activities. Renovation of these high-risk housing units could then be carried out through the collaboration of CDC's Childhood Lead Poisoning Prevention Programs (CLPPP) and programs funded by the U. S. Department of Housing and Urban Development (HUD) or community groups. A housing history list can enable us to keep records on which housing units have been properly renovated, and when. If the structure was renovated properly, other children who eventually live in these renovated housing units will not be lead poisoned (28). Children who are lead poisoned and move into these renovated units should initially demonstrate decreases in their BLLs, followed by lead values below 10 μg/dL with time.

The capacity to achieve the 2010 elimination goal is directly related to the ability to target strategies to geographic areas (29). Geocoding (street address matching or assignment of map coordinates) will be the basis for data linkage and analysis in the 21st century.

The versatility of GIS supports the exploration of spatial relationships, patterns, and trends that may otherwise go unnoticed (30). This technology also allows for the linking of nongeographic data, such as blood lead levels, to geographic locations. GIS allows for the analysis of all data related to geographic location data. Traditional biostatistical and spatially based data analytic methods can be used to estimate risk for lead exposure (31).

GIS is a powerful tool that can precisely locate the home of a child at risk from exposure to lead. This level of information is necessary for public health professionals to accurately assess the extent of childhood lead poisoning, to identify new cases, and to evaluate the effectiveness of prevention activities. However, public presentation or release of maps at this level is discouraged. Public access to data below the county level is prohibited or severely restricted because of confidentiality and privacy issues. A major challenge in the coming decade will be to increase public access to GIS information without compromising confidentiality (32).

New methods must be developed to identify these high-risk children, whose homes may be dispersed over a large geographic area. These new methods will heavily rely upon the use of GIS. The difficulty in implementing GIS methodologies is in identifying and obtaining data that contains specific geographically related information. GIS is an important tool that supports federal, state, and local activities focused upon achieving the national goal of eliminating elevated BLLs among young children by 2010.

Case Studies

Study #1

Reissman et al. Use of geographic information system technology to aid Health Department decision making about childhood lead poisoning prevention activities (17).

This study used Jefferson County, Kentucky, tax assessor, CLPPP BLLs, and birth data to demonstrate the usefulness of a GIS in identifying children at risk for lead exposure using BLLs and residential location of at-risk children screened for lead exposure. "At-risk children" were defined as those children living in housing built before 1950 or in an area with a high proportion of older housing. Participants were the cohort of children from Jefferson County who were born in 1995 and screened from 1996 through 1997, and children <7 years who were screened from 1994 through 1998. Results revealed that only 50% of the at-risk children in the county living in pre-1950 housing were screened. In addition, between 1994 and 1998, 79 homes housed 184 (35%) of the 524 children with lead poisoning. This information was used to successfully win a HUD grant to target these houses.

Study #2

Kim et al. Relation between housing age, housing value, and childhood BLLs in children in Jefferson County, Kentucky (15).

This study used Jefferson County, Kentucky, tax assessor and CLPPP data in a GIS to demonstrate the relation between housing age, housing value, and childhood BLLs in children. Socioeconomic data were linked to tax assessor data to determine whether living in older or less-expensive housing is a risk factor for having an elevated blood lead level among children. The results were as follows:

1) more children living in older housing had elevated BLLs then those living in newer housing;
2) the older the house (1900 to 1980), the higher the mean blood lead level of resident children and the greater the proportion of resident children with elevated BLLs;
3) a dose-response trend with higher BLLs and older housing by decade built; and
4) children living in lesser-valued houses were at greater risk of having elevated BLLs, suggesting that lesser-valued homes are more likely to have deteriorated paint.

However, housing age does not indicate the condition of the paint or reflect recent remodeling that could increase or decrease a child's risk. These techniques of examining housing were applied to Metro Louisville Housing's application for another HUD Lead Hazard Control Grant, which was funded in October 2002 for $1,944,513.

Study #3

Roberts et al. Using geographic information systems to assess risk for elevated BLLs in children (11).

This study used a GIS to determine high-risk regions in Charleston County, South Carolina, to assist public health officials in developing targeted lead-screening. Using local tax assessor and CLPPP data, properties built before 1978 were geocoded from tax assessor data. Addresses of Charleston County children who had been screened for lead poisoning were also geocoded. Locations of all housing, lead poisoning cases, and negative screens were created as separate map layers. With the help of GIS mapping, the authors concluded that children living in pre-1950 housing are 4 times more likely to be lead poisoned than children living in newer homes. Even with abundant screening, they found little evidence of lead poisoning in neighborhoods with little or no pre-1950 housing. In contrast, it took very little screening effort to find cases of lead poisoning in areas with abundant pre-1950 housing. The authors also noted unexpected clustering from a potential point source, which they planned to investigate in a follow-up study.

Study #4

Braggio et al. Polygon risk score and geographic information system used to display children's residential geographic polygons that differ on lead risk (33).

In a study of Oklahoma ZIP codes, these researchers observed that housing value, along with housing age, can be a predictor of childhood lead poisoning in Oklahoma, both in large urban cities and in rural communities. Oklahoma will now use the birth certificate, county tax assessor, and lead surveillance data to establish the association between pre-1950 residences of venous-confirmed lead poisoned children in the eight communities that had the highest predictive scores. They also plan to study smaller areas where there are high concentrations of pre-1950 housing units, children <6 years old who live in poverty, and low screening rates. By increasing the screening rates, they hope to improve their ability to locate previously unidentified cases of lead-poisoned children.

Healthy People 2010 Objectives related to lead poisoning, and GIS

Healthy People 2010: Objectives for Improving Health lists two objectives that pertain to childhood lead poisoning and one that pertains to geographic information systems (34):

- Eliminate blood leads levels in children >10 µg/dL (section 8–11).
- Increase the proportion of persons living in pre-1950s housing that has been tested for the presence of lead-based paint (section 8–22).
- Increase the proportion of all major national, State, and local health data systems that use geocoding to promote nationwide use of geographic information systems (GIS) at all levels. Public health rests on information. Increased geocoding in health data systems will provide the basis for more cost effective disease surveillance and intervention (section 23–3).

Internet Resources

Listed below are some websites that provide information about using geographical information systems (GIS) for readers who want to learn more about GIS.

GIS Software, Training, and Information

- http://www.esri.com—ArcView and ArcGIS
- http://www.mapinfo.com—MapInfo
- http://www.caliper.com—Maptitude

Data

- http://www.cdc.gov/nceh/lead/lead.htm —Access to information from CDC's Lead Poisoning Prevention Branch.
- http://www.census.gov—U. S. Census 2000 data.
- http://www.census.gov—Connection to GIS information portal (created by ESRI).
- http://www.geodata.gov—Access to geospatial information from federal agencies and a growing number of state, local, tribal, and private agencies through one comprehensive and comprehensible portal.

Other

- http://www.epa.gov—Environmental Protection Agency. Provides access to datasets (emissions and effects), as well as the Geospatial Data Clearinghouse and spatial data and applications.
- http://geonames.usgs.gov/fips55.html —Searchable database for FIPS codes for populated areas, maintained by the Geographic Names Information System of the U.S. Geological Survey.
- http://www.gis.cancer.gov—National Cancer Institute. Good source of information, including Health Insurance Portability and Accountability Act of 1996 (HIPAA) regulations, confidentiality, and tools and resources.
- http://www.gisci.org—GIS Certification Institute (see www.urisa.org).
- http://www.urisa.org—Urban and Regional Information Systems Association: National non-profit association of professionals using geographic information systems (GIS). Good source of information. Offers professional GIS certification (www.gisci.org). Check site for meeting schedules in your area.
- http://www.usps.gov—U.S. Postal Service ZIP code look-up site for verification of individual addresses.
- http://www.colorbrewer.org—Web tool for selecting color schemes for thematic maps, most usually for choropleth maps. Contains color schemes that can be distinguished by colorblind readers.

Journals and Newsletters

- http://www.cdc.gov/nchs/gis.htm—National Center for Health Statistic's site for public health and GIS. All issues of Public Health GIS News and Information, a monthly publication, dedicated to CDC scientific excellence and advancement in disease, injury, and disability control and prevention, and occupational safety, using GIS.
- http://www.geoplace.com—GEOWORLD is a publication for spatial information.

Glossary of Terms

Attribute—Data about a map feature. Attributes of a property include address, year-built, value, and number of apartments

Geographic information system (GIS)—a computer-assisted system for the acquisition, storage, analysis, and display of geographic data

Geocoding—matching an address against a database of standardized addresses and assigning unique map coordinates (i.e., latitude and longitude)

Layer—a geographic set (state, county, or block group) containing the same type of information

Spatial analysis—manipulation of spatial data for analysis

Spatial data—data that occupies space that can be mapped and can be defined by a specific location

Spatial overlay—process by which map layers can be placed directly on top of one another and common features can be identified

Thematic Maps—Illustrate the distribution of attributes

Topology—every layer's relationship to other layers

References

1. Centers for Disease Control and Prevention. Preventing Lead Exposure in Young Children: A Housing-Based Approach to Primary Prevention of Lead Poisoning. Atlanta; CDC: 2004
2. Miranda ML, Dolinoy DC, Overstreet MA. Mapping for prevention: GIS models for directing childhood lead poisoning prevention programs. Environ Health Perspect 2002;110:947–53.
3. Centers for Disease Control and Prevention. Screening young children for lead poisoning: guidance for state and local public health officials. Atlanta: US Department of Health and Human Services, 1997. Available online at URL: http://www.cdc.gov/nceh/lead/guide/guide97.htm.
4. Centers for Disease Control and Prevention. Childhood Lead Poisoning Prevention Program web site at URL: http://www.cdc.gov/nceh/lead/lead.htm
5. Sargent JD, Dalton M, Demidenko E, Simon P, Klein RZ. The association between state housing policy and lead poisoning in children. Am J Public Health 1999;89:1690–5.
6. Brown MJ, Gardner J, Sargent JD, Swartz K, Hu H, Timperi R. The effectiveness of housing policies in reducing children's lead exposure. Am J Public Health 2001;91: 621–4.
7. Brown MJ. Cost benefits of enforcing housing policies to prevent childhood lead poisoning. Med Decis Making 2002;22:482–92.
8. Lanphear BP, Roghmann KJ. Pathways of lead exposure in urban children. Environ Res 1997;74:67–73.
9. Lanphear BP, Hornung R, Ho M, Howard CR, Eberle S, Knauf K. Environmental lead exposure during early childhood. J Pediatr 2002;140: 40–7.
10. US Department of Housing and Urban Development. Guidelines for evaluation and control of lead-based paint hazards in housing. Washington DC: US Department of Housing and Urban Development; 1995. pp. 1–6. Available online at URL: http://www.hud.gov/offices/lead/guidelines/hudguidelines/index.cfm.
11. Roberts JR, Hulsey TC, Curtis GB, Reigart JR. Using geographic information systems to assess risk for elevated BLLs in children. Public Health Rep 2003;118:221–9.
12. Jacobs DE, Clickner RP, Zhou JY, Viet SM, Marker DA, Rogers JW, et al. The prevalence of lead-based paint hazards in US housing. Environ Health Perspect 2002;110:A599-A606.

13. US Department of Housing and Urban Development. Current Housing Reports, American Housing Survey for the United States. Washington DC: US Department of Commerce, Bureau of the Census, and the US Department of Housing and Urban Development, 1995.

14. Pirkle JL, Kaufmann RB, Brody DJ, Hickman T, Gunter EW, Paschal DC. Exposure of the US population to lead, 1991–1994. Environ Health Perspect 1998;106:745–50.

15. Kim DY, Staley F, Curtis G, Buchanan S. Relation between housing age, housing value, and childhood blood lead levels in children in Jefferson County, Ky. Am J Public Health 2002;92:769–70.

16. Sargent JD, Stukel TA, Dalton MA, Freeman JL, Brown MJ. Iron deficiency in Massachusetts communities: Socioeconomic and demographic risk factors among children. Am J Public Health 1996;86:544–50.

17. Reissman DB, Staley F, Curtis GB, Kaufmann RB. Use of geographic information system technology to aid Health Department decision making about childhood lead poisoning prevention activities. Environ Health Perspect 2001;109:89–94.

18. Longley PA, Goodchild MF, Maguire DJ, Rhind DW. Geographic information systems and science. New York: John Wiley & Sons; 2001.

19. Clarke KC. Getting started with geographic information systems. Upper Saddle River (NJ): Prentice Hall; 1999. p. 4.

20. Cromley E, McLafferty SL. GIS and public health. 2002. New York: The Guilford Press; 2002. p. 18–37.

21. Wiggins L (editor). Using geographic information systems technology in the collection, analysis, and presentation of cancer registry data: a handbook of basic practices. Springfield (IL): North American Association of Central Cancer Registries; 2002.

22. Curtis G, Kaufmann R, Alan Bloch. Tax assessor data and GIS in targeting lead poisoning prevention interventions. Presented at the 3rd Annual Conference, GIS in Public Health, San Diego CA, 1999.

23. Krieger N, Waterman P, Chen JT, Soobader M, Subramanian SV, Carson R. ZIP code caveat: bias due to spatiotemporal mismatches between ZIP codes and US census-defined geographic areas—the Public Health Disparities Geocoding Project. Am J Public Health 2002;92:110–2.

24. Raper J. Multidimensional geographic information science. New York: Taylor & Francis; 2000.

25. Löytönen M. GIS, time geography and health. In: AC Gatrell AC, Löytönen M, editors. GIS and health. Philadelphia: Taylor & Francis, Inc; 1998. p. 97–110.

26. Schwab LT, Roberts JR, Reigert JR. Inaccuracy of parental reporting of the age of their home for lead-screening purposes. Arch Pediatr Adolesc Med 2003;157: 584–6.

27. US Census Bureau. 2000 census of population and housing; summary file 3: technical documentation; appendix B: definitions of subject characteristics. Washington, DC: US Department of Commerce, 2002: B-68. Available at URL: http://www.census.gov/prod/cen2000/doc/sf3.pdf.

28. Mielke HW, Powell ET, Shah A, Gonzales CR, Mielke PW. Multiple metal contamination from house paints: consequences of power sanding and paint scraping in New Orleans. Environ Health Perspect 2001;109:973–8.

29. Yasnoff WA, and Sondik EJ. Geographical information systems (GIS) in public health practice in the new millennium. J Public Health Manag Pract 1999;5:ix-xii.

30. Croner CM, Sperling J, Broome FR. Geographical information systems (GIS): new perspectives in understanding human health and environmental relationships. Stat Med 1996;15:1961–77.

31. Hendrix E. Geographic information systems. In: Dent BD. Cartography: thematic map design. 5th ed. New York:McGraw Hill; 1999. p. 110–36.

32. Richards TB, Croner CM, Rushton G, Brown CK, Fowler L. Geographic information systems and public health: mapping the future. Public Health Rep 1999;114:359–60.

33. Braggio JT, Cadaret K, Fletcher A, Allen J. Polygon risk score and geographic information system used to display children's residential geographic polygons that differ on lead risk. Oklahoma State Department of Health, Oklahoma, 2003: Presented at the CDC Lead Poisoning Prevention Branch Annual Conference, New Orleans LA, 2001.

34. US Department of Health and Human Services. Healthy People 2010. Washington, DC: US Department of Health and Human Services; 2000. Available online at URL: http://www.healthypeople.gov/document/.

APPENDIX A

Preparing Data for GIS Use—Problems to Avoid

Geocoding is the process whereby specialized software assigns unique map coordinates (i.e., latitude and longitude) to each address in a database. Once an address is geocoded, it can be added to a geographical information system (GIS) for spatial analysis. Geocoding software can also correct errors in address, but addresses must be set up in a proper format for this to occur.

Many problems are encountered in preparing Childhood Lead Poisoning Prevention Program (CLPPP) databases for GIS use. Table I summarizes the types of street number and name errors found within 65,000 records of a CLPPP database, of which 11,260 records did not contain any address information. The remaining 48,633 records contained more than 8,000 errors of various types. Table II summarizes the variety of potential city and ZIP code errors.

Proper Format

Street addresses are comprised of six components: number, number suffix, street direction, street name, street type, and street suffix. Many addresses are comprised of fewer components. An address such as 12½ W McBride Avenue NW would have the following components:

Number	Number Suffix	Street Direction	Street Name	Street Type	Street Suffix
12	½	W	McBride	Avenue	NW

Geocoding software requires addresses to be prepared in four components: address (number, direction, name, and type), city, state (two letters) and ZIP code (five digits). The same address prepared for geocoding would appear as follows:

Address	City	State	Zip Code
12 ½ W McBride Avenue NW	Paterson	NJ	02034

However, the same address could appear in the records as:

Address	City	State	Zip Code
W12.5#McBrideAv	Aterson		20034

Standardization of Addresses

CLPPP databases contain numerous errors that must be corrected before they can be processed by geocoding software. If each record in a database is formatted properly, the software can correct for errors such as misspellings of street names and cities. A local street map is an excellent resource for verifying street names and number ranges.

Common Problems With Data

Once the data have been received, there may be missing information and errors that may require cleaning. Typical errors include misspellings of street names, misidentified street types, invalid street numbers and cities, and the incorrect zip codes. All of these problems might even be found in the listing for the same address! A spreadsheet or database program such as Excel or Access can be used to correct the information. The good news is that this task has to be done only once.

Problems can be encountered in seven distinct categories:

- Extraneous
- Street numbering
- Street name
- Street type
- Street direction
- City name
- ZIP code

Extraneous

The first step in the data cleaning process is to remove extraneous data, representing entries that make no sense, and probably caused by hitting the wrong key or holding the shift key down (see Table A.1):

- Multiple spaces—caused by pressing the space bar more than necessary.
- Punctuation - ,, ;,., /, \, [,], ‘, =, -.
- Shift key errors - !, @, #, $, %, ^, &, *, (,), _, +, “, :, <, >, ?. Most entries of this type probably represent attempts to enter the numbers 1 through 10.

Street Numbering Errors

- Street numbers present numerous obstacles (see Table A.1 for summary):
- Address appears as single field—lack of space between number, name, type, and direction.
- Direction is placed in front, the middle, or at the end of a number, such as: W1224 or 1224W or 1W224.
- Punctuation inserted improperly, as noted above.

Street Name Errors

- Street names can present the most difficulties (see Table A.1 for summary):
- Misspellings are frequently encountered where names have been merged or separated, transposed, or have letters added or dropped. These errors are

especially difficult to find and fix when the first letter is wrong (e.g. Butenberg, rather than Gutenberg).

- Inconsistent Abbreviations—Two common errors involve the words "Mount" and "Saint" in street names. "Mount" can be spelled as Mount, Mt., or Mt (i.e. Mt Blanc, Mount Washington). "Saint" can be found as St, St., Ste, Ste., Sainte, or Saint. Geocoding software can recognize only one version. Sometimes the software can correct these errors, but usually there are other errors in the address that will prevent such a correction.

Street Type Errors

Compounding the issue are misleading or wrong street types, as in the following examples:

- The name is associated with more than one street type. For example, the "street name" is Court, but there are four street types with this name: Court Avenue, Court Drive, Court Road, and Court Street. A list of number ranges usually resolves this problem because each would have its own number range (the county and Post Office do not need this headache). A helpful resource is the U.S. Postal Service website at www.usps.com. Users can select the "Find a ZIP Code" option, enter the address, city, and state, and receive verification of the address and the current ZIP code.
- Street type abbreviations can be very troublesome. Always use the full name, not initials (i.e. Boulevard, not Blvd). A street could be listed as St, Str, St. or Street. Don't assume anything! Some abbreviations have several possible meanings: TR could be Trace, Trail, or Terrace; CR could be Court or Circle.
- Highways present difficulties; not only are the abbreviations numerous, such as Hwy 35, Hgwy 35, Hgy 35, US Hwy 35, or State Hwy 35 , but the actual spelling (and name) could be any of the following: Highway 35, Us Highway 35, State Highway 35, Old State Highway 35, Kentucky State Highway 35, Weatherton Road, and Walter Matheson Hwy.

Street Direction Errors

Another problem encountered with the address field is direction: East, North, South, and West. Number ranges could be helpful where N Rockford St is actually S Rockford St. Difficulties arise when directions all begin with the same numbering system—the ranges for N Rockford St and S Rockford St begin with the number 1.

City Name Errors

Misspellings of city names are common. See Table A.2 for examples (N=23) of various ways the city of Aurora could be misspelled. Errors also occur if city is placed in the state field or city and state are in the same field.

ZIP Code Errors

ZIP codes comprise an administrative delivery system for the U.S. Postal Service. By their very nature, they are subject to change. Many people are familiar with residential ZIP code changes that have occurred since 1990. Geocoding software can correct for these changes if the address (certainly street name and type) is correct. Table A.3 illustrates some common ZIP code errors:

- Inappropriate numbers that are transposed or those containing no digits, all zeros, punctuation and letters of the alphabet.
- Inappropriate ZIP code for the state
- 5-digit codes lacking 1, 2, 3, or 4 digits.

Table A.1. Examples of errors that can occur in street address data entries. (Extra spaces are delineated by a "^" symbol).

Sample Addresses	Errors
#3 ALCOVE AVE	Character
**1*OA ^ AVENUE	Characters and one extra space
43	Number only
*30 ^ ^ ^19VVENUE	Character, three extra spaces
0S644 EAST ST	Numbers mixed with street direction
1^0 ^W 1^TH ST	Three extra spaces, name
1^22 50T^ ^ V	Three extra spaces, questionable street name and type
1000071 WENDY LANE	Inappropriate number
10080_PARIS CTR	Number is attached to street name
10909_S KONRAD	Number is attached to street direction
116 15700 ST	Inappropriate name
1218 ^ 2^4TH AVE	Multiple spaces, name confusion
138OO REVISION	Number ends with two letters
Z65 N SPOILED ST	Letter in front of number
*U*36L**YMD***THY	Asterisks (seven) prevent deciphering of address
102 NCRYST^AL	Direction attached to name, space
1020WILSONSST	Lack of spaces between number, name, and type
103^ HAROLD *RIVE	Extra space plus asterisk attached to name
1125 N ROADW*Y	Asterisk
113 *^ DEMSTER	Asterisk and space
1228 N QUEENS STADIUM	Incomplete name
1235 E ^*LSON	Incomplete name (extra space and asterisk)
1253WOR^ ^D	Run on name, number with last letter separated by two spaces
13 ^ ^ ^ ^SHAM^MPBE^L	Spaces (six) and incomplete name
132183S GR*ENLAND	Number and name problem
153W ILSON MOUND	Correct name = "Wilson" Mound
15431S HAVLI*VE	Direction and name problem
1824257 *HAVECICER0	Correct address = 1824 257th Ave, Cicero
230QUEVADIS RD	Run on address with misspelled name
2960 DILLINGER @2	Character
334 SPEN^CE ST	Extra space

TABLE A.2. Examples of various ways the city of Aurora could be misspelled.

AURORA	AURORAAAAAAAAAA
A A	AURORN
A UROR	AURORSA
AUAORA	AURPRA
AUORA	AURQRA
AUREO	AVROR A
AURO	AVRORA
AURO *	CURORA
AUROAR	SURORA
AUROR	UROR
AUROR A	URORA
AURORA *	
AURORA M	

TABLE A.3. Examples of possible ZIP code data entry errors.

Sample City and ZIP Code		
City	**ZIP code**	**Errors**
CHICAGO	00000	Inappropriate numbers
CHICAGO	4	1 of 5 digits
CHICAGO	25	2 of 5 digits
AURORA	00504	Inappropriate numbers
NILES	606	3 of 5 digits
CHICAGO	0608	4 of 5 digits
RIV ^ RDALE	06082	City error with inappropriate numbers
CHICAGO	6032M	Inappropriate numbers
CHICAGO	606??	Inappropriate numbers with punctuation marks

Appendix B
Child Blood Lead Surveillance Data Fields

Child Data (Required)
- Child Last Name *
- Child First Name*
- Child Middle Initial*
- Child ID
- DOB (actual)
- Age (reported from laboratory or provider)*
- Sex
- Special Ethnicity
- Race (see table below)
- Pregnant at time of test (if applicable)*
- Child's previous country of residence*
- Travel outside of US*

Blood Lead Test Data (Required)
- Specimen Source for lead test (sample type: venous blood, capillary blood, etc.)
- Date sample collected (Sample Date)
- Date sample analyzed (Sample Analyze Date)
- Laboratory result report date (Result Report Date)
- Numeric result comparator (less than, greater than) *
- Numeric result value
- Numeric result units*
- Explanation for missing numeric result (e.g., clotting, quantity not sufficient, etc.)*

Child Risk Assessment Data (Required)
- Non-paint lead source - other
- Non-paint lead source – traditional medicines
- Non-paint lead source – occupation of household member
- Non-paint lead source – hobby of household member
- Non-paint lead source –pottery, imported or improperly fired
- Non-paint lead source – child occupation

Laboratory Data (Required)
- Name of Laboratory that reported test result *
- CLIA number *
- Limit of Detection for blood lead testing*

Address Data (Required)
- Street Address*
- Address ID
- City
- County FIPS
- State
- Zip Code
- Census Tract

Case Data (Required)
- Date case closed*
- Closure reason*

Investigation Data (Required)
- Referral date for investigation
- Date address investigation inspection completed
- Investigation findings of sources
- Investigation closure reason*
- Date remediation due
- Date address hazard remediation or abatement completed
- Date clearance testing completed
- Clearance testing results

Provider Data (Required)
- Provider/medical group State*
- Provider/medical group City*
- Provider/medical group County*

*new required fields

APPENDIX C

Census 2000 Content

100% characteristics (short form):

A limited number of questions were asked of every person and housing unit in the United States. Information is available on:

- Name
- Household relationship
- Sex
- Age
- Hispanic or Latino origin
- Race
- Tenure (whether the home is owned or rented)

Sample characteristics (long form):

Additional questions were asked on a sample of persons and housing units to obtain data on the following:

Population

Social Characteristics

- Marital status
- Place of birth, citizenship, and year of entry
- School enrollment and educational attainment
- Ancestry
- Residence 5 years ago (migration)
- Language spoken at home and ability to speak English
- Veteran status
- Disability
- Grandparents as caregivers

Economic Characteristics

- Labor force status
- Place of work and journey to work
- Occupation, industry, and class of worker
- Work status in 1999
- Income in 1999

Housing

Physical Characteristics

- Units in structure
- Year structure built
- Number of rooms and number of bedrooms
- Year moved into residence
- Plumbing and kitchen facilities

APPENDIX D

Desired Tax Assessor Data

Requested information	Description
Absentee or Owner Occupied	Yes/No
Assessor Parcel Number (APN)	Tax assessor's unique identifier
Book Page	Location of property map
Building Class or Type	Single family, apartment, condominium, duplex, etc.
Building Number	If more than one building
Census Tract	
Condo Name	
Effective Year Built	Year structure was completed
Neighborhood	
Number of Buildings	Number of structures on a property
Number of Stories	
Number of Units	Number of units in the structure
Owner Apartment Number	
Owner City	
Owner County	
Owner Mailing Address (full)	
Owner Name	
Owner State	
Owner Street Name	
Owner Street Number	
Owner ZIP Code	
Remodeled	Yes/No, Date
Sale Date	
Sale Price	
Situs Address (full)	Property location
Situs City	Property location
Situs County	Property location
Situs House Alpha	Property location
Situs House Number	Property location
Situs Mailing Address (full)	Property location
Situs Street Name	Property location
Situs Suffix	Property location
Situs ZIP Code	Property location
Situs ZIP + 4	Property location
Subdivision Name	
Tax District	
Year Built	Year structure was started
Zoning Code	

APPENDIX E

Available Census 2000 Variables In LPPB Shapefiles

Census 2000 Variables In LPPB Shapefiles

Field Name	Description		Field Name	Description
STUSAB	STATE ABBREVIATION	40	BPOV_0_5	BELOW_POVERTY_AGE_0_5_YRS
NAME	NAME	41	TOT_HU	TOTAL_HOUSING_UNITS
STFID	Federal Information Processing Standard ID code	42	HU_99_M2	STRUCTURE_BUILT_1999_MAR2000
TOT_POP	TOTAL_POPULATION	43	HU_95_98	STRUCTURE_BUILT_1995_98
WHITE	WHITE_ALONE	44	HU_90_94	STRUCTURE_BUILT_1990_94
BLACK	BLAC_AA_ALONE	45	HU_80_89	STRUCTURE_BUILT_1980_89
NAT_AMER	AMER_IN_NAT_ALASKAN_ALONE	46	HU_70_79	STRUCTURE_BUILT_1970_79
ASIAN	ASIAN_ALONE	47	HU_60_69	STRUCTURE_BUILT_1960_69
PAC_ISLA	NAT_HAWAIIAN_PAC_ISLANDER_ALONE	48	HU_50_59	STRUCTURE_BUILT_1950_59
OTHERRACE	OTHER_RACE_ALONE	49	HU_40_49	STRUCTURE_BUILT_1940_49
MULTIRACE	TWO_OR_MORE_RACES	50	HU_PRE_40	STRUCTURE_BUILT_1940_EARLIER
HISPANIC	HISPANIC_TOTAL	51	HU_PRE_50	STRUCTURE_BUILT_PRE_1950
HISPWHITE	HISPANIC_WHITE_ALONE	52	MEDYRBLT	MEDIAN_YEAR_STRUCTURE_BUILT
HISPBLACK	HISPANIC_BLACK_AA_ALONE	53	OCC_HU	TOTAL_OCCUPIED_HOUSING_UNITS
HISPNATAM	HISPANIC_AI_AN_ALONE	54	OWN_OCC_HU	TOTAL_OWNER_OCCUPIED_UNITS
HISPASIAN	HISPANIC_ASIAN_ALONE	55	OWN_99_M2	OWNER_OCCUP_BUILT_1999_MAR2000
HISPPACIS	HISPANIC_NH_PI_ALONE	56	OWN_95_98	OWNER_OCCUP_BUILT_1995_98
HISPOTHER	HISPANIC_OTHER_ALONE	57	OWN_90_94	OWNER_OCCUP_BUILT_1990_94
HISPMULTI	HISPANIC_TWO_OR_MORE_RACES	58	OWN_80_89	OWNER_OCCUP_BUILT_1980_89
AGE_0_5	AGE_0_5_YRS	59	OWN_70_79	OWNER_OCCUP_BUILT_1970_79
TOTAL_HH	TOTAL_HOUSEHOLDS	60	OWN_60_69	OWNER_OCCUP_BUILT_1960_69
HH10000	HOUSEHOLD_INCOME_LT_10000	61	OWN_50_59	OWNER_OCCUP_BUILT_1950_59
HH14999	HOUSEHOLD_INC_10000_14999	62	OWN_40_49	OWNER_OCCUP_BUILT_1940_49
HH19999	HOUSEHOLD_INC_15000_19999	63	OWN_PRE40	OWNER_OCCUP_BUILT_1940_EARLIER
HH24999	HOUSEHOLD_INC_20000_24999	64	OWN_PRE50	OWNER_OCCUP_BUILT_PRE_1950
HH29999	HOUSEHOLD_INC_25000_29999	65	REN_OCC_HU	TOTAL_RENTER_OCCUPIED_UNITS
HH34999	HOUSEHOLD_INC_30000_34999	66	REN_99_M2	RENTER_OCCUP_BUILT_1999_200003
HH39999	HOUSEHOLD_INC_35000_39999	67	REN_95_98	RENTER_OCCUP_BUILT_1995_98
HH44999	HOUSEHOLD_INC_40000_44999	68	REN_90_94	RENTER_OCCUP_BUILT_1990_94
HH49999	HOUSEHOLD_INC_45000_49999	69	REN_80_89	RENTER_OCCUP_BUILT_1980_89
HH59999	HOUSEHOLD_INC_50000_59999	70	REN_70_79	RENTER_OCCUP_BUILT_1970_79
HH74999	HOUSEHOLD_INC_60000_74999	71	REN_60_69	RENTER_OCCUP_BUILT_1960_69
HH99999	HOUSEHOLD_INC_75000_99999	72	REN_50_59	RENTER_OCCUP_BUILT_1950_59
HH124999	HOUSEHOLD_INC_100000_124999	73	REN_40_49	RENTER_OCCUP_BUILT_1940_49
HH149999	HOUSEHOLD_INC_125000_149999	74	REN_PRE40	RENTER_OCCUP_BUILT_1940_EARLIER
HH199999	HOUSEHOLD_INC_150000_199999	75	REN_PRE_50	RENTER_OCCUP_BUILT_PRE_1950
HH200000	HOUSEHOLD_INC_GE_200000	76	MEDYRB_OCC	MEDIAN_YEAR_BUILT_TOTAL_OCCUP
HH_MEDINC	HOUSEHOLD_MEDIAN_INCOME	77	MEDYRB_OWN	MEDIAN_YEAR_BUILT_OWNER_OCCUP
PC_INCOME	PER_CAPITA_INCOME	78	MEDYRB_REN	MEDIAN_YEAR_BUILT_RENTER_OCCUP

DEPARTMENT OF HEALTH & HUMAN SERVICES · USA

CDC
SAFER • HEALTHIER • PEOPLE™

Birth Defects Prevention • Training • Chronic Disease Prevention • Injury Prevention • Laboratories
Epidemiology • Environmental Health • Disabilities • Genetics and Public Health • Global Health • Public
Workplace Health • Health Information • HIV Prevention and Control • Health Statistics • Enironm
al • Chronic Disease Prevention • Infectious Disease Protection • Immunization • Injury Prevention • Pr
Child Health • Global Partnerships • Minority Outreach • Monitoring Health • Community Partnerships
SAFER • HEALTHIER • PEOPLE • SAFER • HEALTHIER • PEOPLE • SAFER • HEALTHIER • PEOPLE • SAFER • HEA
Health Education • Epidemiology • Workplace Health • Immunization • Workplace Safety • Training • C
lth • Prevention Research • Private Sector Partnerships • Public Health Workforce • Women's Health
Infectious Disease Protection • Public Partnerships • Healthy Aging • HIV Prevention and Control • H
• STD Control • Women's Health • Training • Health Statistics • Minority Outreach • Youth Programs • G
tion • Community Partnerships • Healthy Aging • Global Partnerships • Outbreak Investigations • Healt
Birth Defects Prevention • Training • Chronic Disease Prevention • Injury Prevention • Laboratories