

FINAL REPORT

Assessment of Populations in Proximity to Superfund National Priorities List Sites

Submitted to
Dr. Claudia Thompson
Acting Director, Superfund Basic Research Program
Acting Director, Center for Risk and Integrated Sciences
Division of Extramural Research and Training
National Institute of Environmental Health Sciences
PO Box 12233; EC-27
Research Triangle Park, NC 27709
Thomps14@niehs.nih.gov

by
Center for International Earth Science Information Network (CIESIN)
Columbia University
Palisades, New York

Meredith L. Golden, Gregory Yetman, and Tricia Chai-Onn

5 December 2008

The Center for International Earth Systems Information Network (CIESIN) at Columbia University, in response to a request by NIEHS, has conducted an assessment of populations living in proximity to Superfund National Priorities List (NPL) sites for the year 2000. CIESIN improves on earlier assessments in several key ways. First, CIESIN uses the U.S. Census Grids database to determine population totals and demographic breakdowns in proximity to NPL sites. Second, we develop a methodology to eliminate the double-counting of populations in proximity of more than one NPL site. Third, our study uses the 'majority rule' to determine when population data should be included in the summation. Fourth, the total population in proximity to two or more Superfund sites is estimated. Finally, these analyses are conducted twice: once for populations living within a 1 mile buffer, and then again for those within 4 miles of the site.

Population Data

For our analyses, we utilize the 2000 U.S. Census Grids population database (<http://sedac.ciesin.columbia.edu/usgrid/>). The standard census block and block group (vector) data set, typically used for demographic analyses, makes it difficult to estimate the number of residents living within circular buffers of NPL sites. In contrast, the gridded (raster) data set assigns Census variables (SF1 and SF3) to a regular surface consisting of grids with 30-arc seconds (about 1 kilometer) resolution. This format facilitates faster, easier, and more accurate analyses within Geographic Information Systems (GIS).

In consultation with the NIEHS staff, a set of 21 variables were selected from the U.S. Census Grids data for a closer examination of the demographic, economic, and housing characteristics of populations living in proximity to Superfund sites.

Selected Variables

<input type="radio"/> Population (SF1)	<input type="radio"/> Population age 80 and older (SF1)
<input type="radio"/> Population (SF3)	<input type="radio"/> White alone (SF1)
<input type="radio"/> Population for whom the poverty level is determined (SF3)	<input type="radio"/> Black or African American alone (SF1)
<input type="radio"/> Population under age 1 (SF1)	<input type="radio"/> American Indian or Alaska Native alone (SF1)
<input type="radio"/> Population ages 1 to 4 (SF1)	<input type="radio"/> Hispanic (SF1)
<input type="radio"/> Population age 5 to 17 (SF1)	<input type="radio"/> Asian alone (SF1)
<input type="radio"/> Population ages 65 to 79 (SF1)	<input type="radio"/> Population living below 200% of the poverty level (SF3)

○ Population living below 50% of the poverty level (severe poverty) (SF3)	○ Female headed households with related children under 18 (SF1)
○ Population living below poverty level (SF3)	○ Population in households (SF1)
○ Foreign born (SF3)	○ Owner occupied housing units (SF1)
○ Linguistically isolated households (SF3)	

The U.S. Census Bureau provides a comparison of the Summary Files (SF1, SF2, and SF3). See Appendix I http://www.factfinder.census.gov/home/en/epss/sf3_compare.html. According to the U.S. Census Bureau Glossary:

This file [SF1 or short form] presents 100-percent population and housing figures for the total population, for 63 race categories, and for many other race and Hispanic or Latino categories. This includes age, sex, households, household relationship, housing units, and tenure (whether the residence is owned or rented). Also included are selected characteristics for a limited number of race and Hispanic or Latino categories. The data are available for the U.S., regions, divisions, states, counties, county subdivisions, places, census tracts, block groups, blocks, metropolitan areas, American Indian and Alaska Native areas, tribal subdivisions, Hawaiian home lands, congressional districts, and ZIP Code Tabulation Areas. Data are available down to the block level for many tabulations, but only to the census-tract level for others.

(http://www.factfinder.census.gov/home/en/epss/glossary_s.html).

By comparison with SF1, SF3 (the long form) includes more detailed data on social, economic, and housing characteristics for a sample of the population (generally 1-in-6 of persons and housing units). The SF1 offers more limited geographic coverage with tabulations only down to the block group, not the block level. It is important to note the following, "The differences between the long form estimates in SF 3 and values in SF 1 or SF 2 are particularly noticeable for the smallest places, tracts, and block groups. The long form estimates of total population and total housing units in SF 3 will, however, match the SF 1 and SF 2 counts for larger geographic areas such as counties and states, and will be essentially the same for medium and large cities." More information on SF1 and SF3 is located on the U.S. Census Bureau website: http://www.factfinder.census.gov/servlet/MetadataBrowserServlet?type=dataset&id=DEC_2000_SF1_U&lang=en http://www.factfinder.census.gov/servlet/MetadataBrowserServlet?type=dataset&id=DEC_2000_SF3_U&lang=en.

National Priorities List Site Location Data

The U.S. EPA's Comprehensive Environment Response, Compensation, and Liability Information System (CERCLIS) online database includes latitudinal and longitudinal coordinates for the centroid of each National Priorities List (NPL) site. However, the publicly available data are not in a format that can be easily downloaded. Instead, Brette Beasley from EPA provided a data file for this study that consists of all 1,641 NPL sites listed as of June 7, 2008 by name, their identification numbers, and latitude/longitude coordinates. Within this dataset, it was necessary to correct six of the original coordinates by changing the sign before the latitude or longitude coordinates and providing the missing coordinates for the Flash Cleaners in Florida. In addition, we removed 7 sites in 4 U.S. Territorial areas (Guam, Virgin Islands, American Samoa, and Northern Mariana Islands) plus the now independent Palau, Federated States of Micronesia, since demographic data at high resolution are not available for analysis. The final EPA database used here consists of point locations for a total of 1,634 NPL sites in the U.S. and Puerto Rico. These include sites listed: Currently on the Final NPL (1,251), Deleted from the Final NPL (323), and Proposed for the NPL (60).

In examining the proximity of nearby populations, NPL point data are useful; however, polygon boundaries for the individual NPL sites provide better information for determining populations at risk of potential exposures. The Agency for Toxic Substances and Disease Registry (ATSDR) created such a database with 2,042 polygons for hazardous waste sites listed as of January 1, 1996. In the ATSDR database, 1,562 polygons have the same identification numbers as NPL sites listed in the EPA CERCLIS database. Together, the ATSDR's polygon data for 1,562 NPL sites and EPA's point data for the remaining 72 NPL sites serve as the locations for the Superfund sites in this study.

Methodology

The first step in determining the populations in close proximity to Superfund sites was to find the most up-to-date georeferenced data for the location of National Priorities List (NPL) sites. Following a request by NIEHS, EPA provided an Excel file extracted from the Comprehensive Environmental Response, Compensation and Liability Information System (CERCLIS). It lists all 1,641 NPL sites in the U.S. and its territories as of June 7, 2008 with names, identification numbers, and coordinates by latitude and longitude. After cleaning the data for minor errors (six incorrect latitude and longitude signs) and one missing value, it was then determined that gridded population data were not available for 7 sites in U.S. territorial areas. These sites were dropped leaving 1,634 NPL sites in the point database. Further investigation found that ATSDR's Geospatial Research, Analysis, and Services Program (GRASP) had a polygon database for selected hazardous waste sites created in 1996. Using EPA identification numbers,

1,562 ATSDR polygons were identified as NPL sites. For the remaining 72 NPL sites without polygon data, we used the EPA point data.

To identify populations in close proximity to the National Priorities List sites, our study created one-mile and four-mile buffer polygon datasets using ArcGIS™ with the ATSDR polygon data for the 1,562 NPL sites and the EPA point data for the remaining NPL sites. An earlier analysis¹ by Heitgerd et al. found it too difficult to avoid the problem of double-counting populations who are in close proximity to more than one Superfund site. To eliminate double-counting for our calculations, we dissolved the buffer areas that overlap (i.e., those that are within 1 or 4 miles of multiple sites), thus creating a seamless layer for summation. Heitgerd also used the “total count method”: if the buffer crossed any part of census block, no matter how small, then the entire value for that block was included in the calculation for the site total. This approach leads to overestimates of population counts within buffers. Instead, we implement the “majority rule” which reduces overestimation. If fifty percent or more of the gridded cell area fall within the buffer, the entire value is assigned to that cell for each of the selected variables is included in the calculations. Our analyses then uses the zonal statistics function to sum the 21 variables from the U.S. Census Grids.

To estimate the number of persons within two or more NPL sites, non-dissolved buffers were generated for all sites. Our GIS specialist then created geoprocessing scripts to remove areas from the buffers that are only within 1 or 4 miles of a single site. The remaining areas were then dissolved and used with the zonal statistics function to calculate the population within 1 and 4 miles of 2 or more NPL sites.

¹ Heitgerd JL, Burg JR, Strickland, HG. [A geographic information systems approach to estimating and assessing National Priorities List site demographics: racial and hispanic origin composition](#). Int J Occup Med Toxicol 1995;4(3):343-63. Also available on ATSDR website: <http://www.atsdr.cdc.gov/NER/JOURNAL/jad11030302.html>

Deliverables

Total Population Estimates

- 1) The estimates of total residential population in 2000 living within a 1-mile buffer of National Priorities List (NPL) Superfund sites (1634) in the United States and Puerto Rico include 13,217,736 (SF1) and 13,214,375 (SF3).
- 2) The estimates of total residential population in 2000 living within a 4-mile buffer of National Priorities List (NPL) Superfund sites (1634) in the United States and Puerto Rico include 74,727,924 (SF1) and 74,726,099 (SF3).
- 3) The number of children under the age of 5 in 2000 living within a 1-mile buffer of National Priorities List (NPL) Superfund sites (1634) in the United States and Puerto Rico is 972,459 (SF1).
- 4) The number of children under the age of 5 in 2000 living within a 4-mile buffer of National Priorities List (NPL) Superfund sites (1634) in the United States and Puerto Rico is 5,226,990 (SF1).
- 5) The number of people over the age of 65 in 2000 living within a 1-mile buffer of National Priorities List (NPL) Superfund sites (1634) in the United States and Puerto Rico is 1,510,549 (SF1).
- 6) The number of people over the age of 65 in 2000 living within a 4-mile buffer of National Priorities List (NPL) Superfund sites (1634) in the United States and Puerto Rico is 8,959,837 (SF1).
- 7) The estimate of the residential total population living within 1 mile of two or more NPL sites is 1,465,121 (SF1).
- 8) The estimate of the total residential population living within 4 miles of two or more NPL sites is 28,976,895 (SF1).

Census Tables

- 1) Estimates of total residential population and other selected 2000 Census variables living within a 1-mile buffer of National Priorities List (NPL) Superfund sites (1634) in the United States and Puerto Rico, Appendix II (also available on DVD).
- 2) Estimates of total residential population and other selected 2000 Census variables living within a 4-mile buffer of National Priorities List (NPL) Superfund sites (1634) in the United States and Puerto Rico, Appendix III (also available on DVD).

Site Overlaps (with 1 and 4 mile buffers)

- 1) 588 NPL sites 1-mile buffers overlap with at least one other NPL site 1-mile buffer.
- 2) Maximum number of NPL sites with overlapping 1-mile buffers is 8 (Santa Clara, CA).
- 3) 1,102 NPL sites 4-mile buffers overlap with at least one other NPL site 4-mile buffer.
- 4) Maximum number of NPL sites with overlapping 4-mile buffers is 20 (Santa Clara, CA).

Databases

- 1) NPL site polygons with 1-mile buffer (Appendix IV available on DVD)
- 2) NPL site polygons with 4-mile buffer (Appendix V available on DVD)
- 3) U. S. Environmental Protection Agency (EPA) National Priorities List (NPL) Sites Point Data with CIESIN Modifications, 2008 (Appendix VI available on DVD)
- 4) Agency for Toxic Substances and Disease Registry (ATSDR) Hazardous Waste Site Polygon Data with CIESIN Modifications, 1996 (Appendix VII available on DVD)

Appendix I

US Census Bureau

Comparing SF 3 Estimates with Corresponding Values in SF 1 and SF 2

(http://www.factfinder.census.gov/home/en/epss/sf3_compare.html)

As in earlier censuses, the responses from the sample of households reporting on long forms must be weighted to reflect the entire population. Specifically, each responding household represents, on average, six or seven other households who reported using short forms.

One consequence of the weighting procedures is that each estimate based on the long form responses has an associated confidence interval. These confidence intervals are wider (as a percentage of the estimate) for geographic areas with smaller populations and for characteristics that occur less frequently in the area being examined (such as the proportion of people in poverty in a middle-income neighborhood).

In order to release as much useful information as possible, statisticians must balance a number of factors. In particular, for Census 2000, the Bureau of the Census created weighting areas --geographic areas from which about two hundred or more long forms were completed-- which are large enough to produce good quality estimates. If smaller weighting areas had been used, the confidence intervals around the estimates would have been significantly wider, rendering many estimates less useful due to their lower reliability.

The disadvantage of using weighting areas this large is that, for smaller geographic areas within them, the estimates of characteristics that are also reported on the short form will not match the counts reported in SF 1 or SF 2. Examples of these characteristics are the total number of people, the number of people reporting specific racial categories, and the number of housing units. The official values for items reported on the short form come from SF 1 and SF 2.

The differences between the long form estimates in SF 3 and values in SF 1 or SF 2 are particularly noticeable for the smallest places, tracts, and block groups. The long form estimates of total population and total housing units in SF 3 will, however, match the SF 1 and SF 2 counts for larger geographic areas such as counties and states, and will be essentially the same for medium and large cities.

This phenomenon also occurred for the 1990 Census, although in that case, the weighting areas included relatively small places. As a result, the long form estimates matched the short form counts for those places, but the confidence intervals around the estimates of characteristics collected only on the long form were often significantly wider (as a percentage of the estimate).

SF 1 gives exact numbers even for very small groups and areas, whereas SF 3 gives estimates for small groups and areas such as tracts and small places that are less exact. The goal of SF 3 is to identify large differences among areas or large changes over time. Estimates for small areas and small population groups often do exhibit large changes from one census to the next, so having the capability to measure them is worthwhile.

Appendix II: Year 2000 Census Characteristic Totals within 1-Mile of Superfund NPL Sites



Site locations based on ATSDR polygon data for 1562 sites and EPA point data for 72 sites

Buffer Distance from NPL site - 1 Mile	United States (Continental)	Hawaii	Alaska	Puerto Rico	Total U.S.
Population (SF1)	12,870,400	108,710	78,363	160,263	13,217,736
Population (SF3)	12,866,200	109,244	78,378	160,553	13,214,375
Population for whom the poverty level is determined (SF3)	12,459,700	100,055	73,944	158,588	12,792,287
Population under age 1 (SF1)	191,193	1,880	1,719	2,620	197,413
Population ages 1 to 4 (SF1)	750,988	7,271	6,437	10,350	775,046
Population ages 5 to 17 (SF1)	2,403,950	20,271	17,211	33,943	2,475,375
Population ages 65 to 79 (SF1)	1,082,870	9,581	2,671	11,944	1,107,066
Population age 80 and older (SF1)	396,245	2,833	602	3,803	403,483
White alone (SF1)	8,855,930	26,295	49,775	129,317	9,061,317
Black or African American alone (SF1)	1,513,350	6,651	9,170	13,965	1,543,136
American Indian or Alaska Native alone (SF1)	104,609	462	6,873	533	112,477
Hispanic (SF1)	2,581,900	10,937	5,893	158,557	2,757,287
Asian alone (SF1)	749,120	42,481	3,581	307	795,488
Population living below 200% of the poverty level (SF3)	4,189,590	31,626	24,716	118,452	4,364,384
Population living below 50% of the poverty level (severe poverty) (SF3)	784,766	4,882	2,737	37,564	829,949
Population living below poverty level (SF3)	1,762,740	11,449	7,170	73,200	1,854,559
Foreign born (SF3)	2,278,340	19,137	5,620	2,309	2,305,406
Linguistically isolated households (SF3)	319,991	1,573	782	31,465	353,811
Female headed households with related children under 18 (SF1)	381,240	1,874	2,685	5,984	391,782
Population in households (SF1)	12,448,100	98,959	73,352	157,369	12,777,780
Owner occupied housing (SF1)	2,669,670	13,262	9,795	39,137	2,731,864
Total SF1 2000 U.S. Census Grid population	279,583,437	1,211,537	626,885	3,808,610	285,230,469
Total SF3 2000 U.S. Census Grid population	279,583,437	1,211,537	626,885	3,808,610	285,230,469

Appendix III: Year 2000 Census Characteristic Totals within 4-Miles of Superfund NPL Sites



Site locations based on ATSDR polygon data for 1562 sites and EPA point data for 72 sites

Buffer Distance from NPL site - 4 Miles	United States (Continental)	Hawaii	Alaska	Puerto Rico	Total U.S.
Population (SF1)	73,217,504	378,861	222,659	908,900	74,727,924
Population (SF3)	73,215,696	378,738	222,686	908,979	74,726,099
Population for whom the poverty level is determined (SF3)	71,191,200	364,606	215,025	898,633	72,669,464
Population under age 1 (SF1)	1,022,210	5,968	3,783	14,496	1,046,457
Population ages 1 to 4 (SF1)	4,084,100	23,709	14,505	58,220	4,180,533
Population ages 5 to 17 (SF1)	13,568,900	74,602	46,690	193,269	13,883,461
Population ages 65 to 79 (SF1)	6,447,110	30,782	10,622	72,482	6,560,996
Population age 80 and older (SF1)	2,366,980	7,924	2,635	21,302	2,398,841
White alone (SF1)	50,737,900	71,525	155,302	732,469	51,697,196
Black or African American alone (SF1)	10,249,400	14,899	16,652	68,618	10,349,569
American Indian or Alaska Native alone (SF1)	471,619	1,118	18,564	3,334	494,636
Hispanic (SF1)	12,462,000	31,651	13,441	900,423	13,407,515
Asian alone (SF1)	3,959,330	170,401	10,756	1,599	4,142,087
Population living below 200% of the poverty level (SF3)	21,790,600	92,387	55,831	667,656	22,606,474
Population living below 50% of the poverty level (severe poverty) (SF3)	4,383,010	14,670	8,199	218,120	4,623,998
Population living below poverty level (SF3)	9,456,840	33,343	18,432	416,111	9,924,725
Foreign born (SF3)	11,640,500	68,359	17,092	14,874	11,740,824
Linguistically isolated households (SF3)	1,674,740	5,661	2,295	178,737	1,861,433
Female headed households with related children under 18 (SF1)	2,123,050	6,734	6,962	31,400	2,168,145
Population in households (SF1)	71,057,504	364,522	213,722	896,539	72,532,287
Owner occupied housing (SF1)	16,286,700	61,716	41,388	224,561	16,614,365
Total SF1 2000 U.S. Census Grid population	279,583,437	1,211,537	626,885	3,808,610	285,230,469
Total SF3 2000 U.S. Census Grid population	279,583,437	1,211,537	626,885	3,808,610	285,230,469

Additional Appendices for Databases

Available on Final Report DVD

Appendix IV: NPL site polygons with 1-mile buffer

Appendix V: NPL site polygons with 4-mile buffer

Appendix VI: EPA NPL Sites (1,641) Point Data, June 7, 2008-- EPA Excel file with CIESIN
corrected lat/long signs & added one missing value

Appendix VII: ATSDR Hazardous Waste (1,562 NPL sites) Polygon Data, January 1, 1996 with
CIESIN modifications, August 2008