The Public Health Disparities Geocoding Project Monograph

Geocoding and Monitoring US Socioeconomic Inequalities in Health: An introduction to using area-based socioeconomic measures

CASE EXAMPLE:

Analysis of all cause mortality rates in Suffolk County, Massachusetts, 1989-1991, by CT poverty strata.

Step 1 | Step 2 | Step 3 | Step 4 | Step 5 | Step 6 | Step 7 | Step 8 | Completed Tables

We've created this case example as an opportunity for you to try out our methods. The example draws on all cause mortality data from Suffolk County, Massachusetts, between 1989 and 1991. You'll have a chance to analyze these data by census tract poverty to see the socioeconomic gradient in mortality in this county. We've divided the case example into clearly defined tasks to highlight the process of moving from raw data to summary measures of the socioeconomic disparity. Clicking on "Methods & SAS" will take you to a step by step comparison of each task, the relevant analytic methods, and sample SAS code.

• Step 1: Aggregate the numerator data.

The file rawcase.csv (click to download) is a comma delimited file containing all deaths occurring in Suffolk County, Massachusetts, between 1989 and 1991. Each person who died is represented by one line in the data file. The variable “AGE” gives the age at death. The variable “AREAKEY” is the geocode to the census tract level. Read these data into a SAS dataset, and then aggregate deaths within each census tract into the following age categories:

Age Category	AGE (years)
1	0-14
2	15-24
3	25-44
4	45-64
5	65+

View Methods & SAS Programming for Step 1

• Step 2: Aggregate the denominator data.

The file rawdenom.csv (click to download) is a comma-delimited file containing the estimated population count in 31 age categories [see Analytic Methods *Aggregating Denominator Data* section] for the 189 census tracts in Suffolk County, from the 1990 U.S. Census. Each census tract is represented by one line in the data file, with the 31 age categories arrayed horizontally.

a. Aggregate the population counts into the five broad age categories listed above.

b. Transpose the structure of the data, so that there is one record for each age stratum within a census tract, with a corresponding categorical age variable and population count. You should end up with 5 records for each census tract, with each record represented by one line of your output dataset.

c. Multiply the population count by 3, to yield a person-time denominator for three years worth of death data.

View Methods & SAS Programming for Step 2

back to top

• Step 3: Merge the numerators and denominators by AGECAT and AREAKEY.

For age cells in census tracts where no cases were reported, set the numerator to zero.

View Methods & SAS Programming for Step 3

• Step 4: Now merge the combined numerator and denominator data from Step 3 with the ABSM data, by AREAKEY.

The file rawabsm.csv (click to download) is a comma-delimited file containing the 189 census tracts in Suffolk County, and the % of persons living below poverty for each tract, categorized into 4 categories (1=0-4.9%, 2=5-9.9%, 3=10-19.9%, 4=20-100%).

View Methods & SAS Programming for Step 4

• Step 5: For each category of CT poverty, calculate the age-standardized incidence rate, using the year 2000 standard million.

In order to do this:

a. Aggregate the numerator and denominator within each age X CT poverty stratum, across all census tracts.

b. Exclude cases and denominator where CT poverty is missing.

c. Merge with the year 2000 standard million in five age categories (see table below):

Age in 11 categories	Year 2000 standard million	Age in 5 categories	Year 2000 standard million
<1	13,818	<15	214,700
1-4	55,317		
5-14	145,565		
15-24	138,646	15-24	138,646
25-34	135,573	25-44	298,186
35-44	162,613		
45-54	134,834	45-64	222,081
55-64	87,247		
65-74	66,037	65+	126,387
75-84	44,842		
85+	15,508		

d. Calculate the age-standardized incidence rate [see Analytic Methods section 1], standardized to the year 2000 standard million, and the corresponding “gamma” confidence intervals [see Analytic Methods section 2] for the direct standardized rates, to fill in the following table:

CT poverty	IR*st* (age standardized rate per 100,000)	95% confidence intervals ("gamma" intervals)	
0-4.9%			
5-9.9%			
10-19.9%			
20-100%			

View Methods & SAS Programming for Step 5

• Step 6: Estimate the age-standardized incidence rate difference and the age-standardized incidence rate ratio [see Analytic Methods section 4]
comparing the age standardized rates in each poverty stratum to the rate in the least impoverished poverty stratum (0-4.9%).

Calculate the 95% confidence limits on the incidence rate difference and incidence rate ratio. Fill out the table below:

CT poverty	IRD*rst* (age standardized incidence rate difference)	95% confidence intervals		IR*rst* (age standardized incidence rate ratio)	95% confidence intervals	
0-4.9%	0	(reference)		1	(reference)	
5-9.9%						
10-19.9%						
20-100%						

View Methods & SAS Programming for Step 6

• Step 7: Estimate the relative index of inequality (RII) [see Analytic Methods section 5]
for CT level poverty in relation to all cause mortality.

a. Estimate the approximate cumulative distribution function for CT poverty, based on the population denominator for each poverty stratum (summed up over age).

b. Calculate the expected cases in each CT poverty stratum, based on the age-standardized incidence rate.

c. Fit a Poisson log linear model, modeling the expected number of cases as a function of the approximate cumulative distribution of CT poverty, using the population denominator as an offset.

d. Exponentiate the beta term from this model to get the relative index of inequality.

		RII (relative index of inequality)	95% confidence intervals		
	Estimate				

View Methods & SAS Programming for Step 7

- **Step 8: Calculate the population attributable fraction [see Analytic Methods section 6] of all cause mortality due to CT poverty.**

a. Starting with the data from Step 4, sum up over AREAKEY into strata defined by AGECAT and CT poverty.

b. Calculate (i) the total cases in each age stratum, over poverty; and (ii) the rate in the reference group of CT poverty.

c. Calculate stratum specific rates, rate ratios, and case fractions.

d. Calculate the age-stratum-specific population attributable fractions.

e. Calculate the grand total of all cases to use in calculating weights for all age strata.

f. Finally, calculate the aggregated population attributable fraction, using the age specific weights based on proportion of cases in each age stratum.

		Aggregated population attributable fraction	
	Estimate		

View Methods & SAS Programming for Step 8

To see the completed tables click here:
Answers

This work was funded by the National Institutes of Health (1RO1HD36865-01) via the National Institute of Child Health & Human Development (NICHD) and the Office of Behavioral & Social Science Research (OBSSR).