Social and spatial variations in mortality in Pennsylvania

Project lead: Ana Diez Roux
Contact information: avd37@drexel.edu

Collaborators: Amy Auchincloss, Steven Melly, Kari Moore, Jeffrey Yaeger

Background/rationale

It has become increasingly apparent that health varies substantially across neighborhoods and may be affected by neighborhood level attributes. To our knowledge no study has investigated small area variations in mortality across the state of Pennsylvania. Characterizing and understanding the drivers of these variations are important (1) to characterize the presence and magnitude of spatial health inequalities and changes over time in these inequalities; (2) to identify are indicators associated with space and time variation; (3) to identify possible interventions; and (4) to target public health and social efforts. The availability of estimate for small areas that can be tracked over time will also provide a platform to monitor the impact of various policies and interventions through natural experiments.

The proposed study will examine spatial and temporal variations in infant mortality, in mortality for different causes, and in life expectancy in the state of Pennsylvania over a period of more than 50 years. Specifically we will examine: (1) small area variations in mortality (2) associations of mortality with various small area (neighborhood) indicators (3) spatial variations in trends over time and (4) associations of area indicators with trends over time.

Research Aims/Hypotheses

Study Aim: The overall scientific goal is to examine social and spatial variations in mortality across Pennsylvania and over time.

Study hypotheses:

(1) Infant and adult mortality will be strongly spatially patterned revealing large spatial inequities after controlling for the age, sex, and race/ethnic composition of areas.

(2) Infant and adult mortality will be associated with neighborhood socioeconomic and social environment characteristics and race/ethnic composition.

(3) Spatial and social patterning will vary over time.

Data

Sample:

The sample will consist of all birth and death records for the state of Pennsylvania for years 1960-present. No exclusions will be made. The number of annual deaths ranges from approximately 123,000 to approximately 129,000 across the years of study. The number of births ranges from 140,000 to 172,000 a year across years. Data will be obtained from the state of Pennsylvania Department of Health vital statistics records.

Exposure variables:

The study will create neighborhood and environmental variables that can be linked to birth and death records obtained from the state of Pennsylvania Department of Health using a variety of publicly and commercially available data sources.

Key exposure variables will be neighborhood (census tract) level of race/ethnicity, socioeconomic status, and racial segregation. Race/ethnicity and socioeconomic status will be obtained from the U.S. Census Bureau. Measures of socioeconomic status will include median household income, poverty, education, and employment. Racial segregation will be calculated using the Getis-Ord Gi statistic[1]. These measures will be available for the entire state of Pennsylvania.

Additional exposures will be used in locations where more detailed data is available. This includes crime, safety, social capital, built environment, and other environmental factors.

Crime data is available from the Philadelphia Police Department Mapping and Analysis Unit via OpenDataPhilly[2]. Crimes will be operationalized as the yearly number of crimes within a census tract per 1,000 population. Crime will be assessed as an overall count of total crime as well as by categories of crime (e.g. homicide, burglary, ect). Violent and non-violent crime will be assessed separately.

Safety and social capital will be assessed using the Southeastern Pennsylvania Household (SEPAHH) Survey, conducted by the Public Health Management Corporation, which is a biennial telephone survey administered to approximately 10,000 randomly selected adults in five counties in southeastern Pennsylvania[3]. Census tract level measures of safety and social cohesion will be created using empirical Bayes estimation similar to methods used elsewhere[4]. Safety is assessed by answering yes or no to the question "In the past month, did you not go someplace during the day because you felt you would not be safe?" An empirical Bayes estimate will be created using two-level logistic regression, with over dispersion for "0" responses to model the predicted probability of a "yes" response (unsafe) with level-1 being individuals nested within census tracts and level-2 being census tracts. Social capital is assessed via 3 questions: "How likely are people in your neighborhood willing to help their neighbors" – 5 point Likert scale, "Most people in my neighborhood can be trusted" and "I feel that I belong and am a part of my neighborhood" - both 4 point Likert scale. Answers will be reverse coded where appropriate such that a higher score is positive, then the scales will be standardized to range from 0-3. Empirical Bayes estimates will be calculating using three-level linear regression. Level-1 is questions nested within individual nested within census tracts, level-2 is individuals nested within census tracts, and level-3 is census tracts.

Built environment will be assessed by land use, intersection density, and public transportation. Land use data for the city of Philadelphia is available from the Pennsylvania Spatial Data Clearinghouse (PASDA)[5]. This will be used to create percentage of land use categories (retail, residential, commercial) as well as an entropy measure, which summarizes the mixture of land uses[6]. Intersection density, defined as the number of intersections per hectare, will be assessed using TomTom Dynamap/Transportation streets data to using standard methods[7, 8]. Public transportation stops and routes are available from the Delaware Valley Regional Planning Commission (DVRPC)[9] and will be as distances to nearest public transit or densities of transit stops.

Other environmental measures include green space and air pollution. Green space will be assessed via impervious surface and tree canopy. Impervious surface is the sum of roads, parking lots, sidewalks, rooftops, and other impermeable surfaces and is available from the National Land Cover Database from the Multi-Resolution Land Characteristics Consortium (MRLC)[10]. Tree canopy is the layer of leaves, branches, and stems of trees when viewed from above and this data is also available from MRLC. Both impervious surface and tree canopy will be summarized as percent impervious and percent tree canopy within the census tract. Air pollution will be obtained from the Philadelphia Department of Public Health Air Management Services which operates a network of 10 air pollution monitors across the city[11, 12].

Outcome variables:

Primary endpoints are adult mortality rates and infant mortality rates (overall and by demographic subgroups), and life expectancy estimates. Records from the birth and death records will be geocoded using ArcGIS10.3 address locator and the census tract will be assigned based on the geocoded latitude/longitude. Infant mortality per year for each census tract will be determined as the number of deaths (from death certificates) where age is under 1 year within the census tract divided by the total number of births (from birth certificates). Adult mortality per year for each census tract will be determined by the number death (from death certificate) where age is 18 years or older within the census tract divided by the total adult population from the U.S. Census. For life expectancy, all deaths (infant, child, and adult) will be used and the total population will be from the U.S. Census (see Methods below for details on calculation).

Covariates:

Potential covariates include individual sociodemographics (gender, race/ethnicity, age, education) as recorded on the birth and death certificates.

Data File Format:

Preferred data format is a SAS (.sas7bdat) dataset.

Analytical plan

To test the association of SES with infant and adult mortality rates, a multi-level poisson regression approach using cross-classified cells will be used[13]. For infant mortality, cells within census tracts will be divided by race/ethnicity and gender. For adult mortality, cells within census tracts will be divided by age groups, race/ethnicity and gender. The total number of deaths and births (for infant mortality) and total population (for adult mortality) will be calculated for each of the cross-classified cells, nested within census tracts. Two-level poisson regression models will then be fit to the cross-classified cells as:

Level 1 (cells nested within census tracts):

$$\text{Log } Y_{ij} = \beta_{0j} + \beta_{1j}*\text{age}_{ij} + \beta_{2j}*\text{gender}_{ij} + \beta_{3j}*\text{race/ethnicity}_{ij} + \log n_{ij}$$

Where: Y_{ij} = number of death for the i^{th} cross-classified cell in census tract j

Age_{ij}, gender_{ij}, and $\text{race/ethnicity}_{ij}$ are cell-level categorical variables for the classified levels of each covariate.

Log n_{ij} = the offset for each cell as the total population (or total births) for the i^{th} cross-classified cell in census tract j

Level 2 (census tracts): Each of the coefficients from Level 1 are modelled as

$$\beta_{0j} = \gamma_{00} + \gamma_{01}*\text{SES}_j + U_{0j}$$
$$\beta_{1j} = \gamma_{10} + U_{1j}$$
$$\beta_{2j} = \gamma_{20} + U_{2j}$$
$$\beta_{3j} = \gamma_{30} + U_{3j}$$

Where: γ_{00} = common intercept across census tracts

γ_{01} = Fixed effect of SES

γ_{10} - γ_{30} = common slopes associated with cell-level variables
U_{0j} - U_{3j} = random effects associated with census tracts

To test for trends over time, the above model can be expanded to a three-level hierarchical model with Level 1 being time nested in cells nested in census tracts; Level 2 being cells nested in census tracts; and Level 3 being census tracts. The inclusion of a main effect of time and interactions with SES and each of the covariates will be included to test for time trends.

To calculated life expectancy, death records and total population for each census tract will be aggregated into age groups. Age-specific mortality rates will be calculated using the number of deaths divided by the total population in each age group. These rates will then be used to calculate abridged life tables by census tract and census tract level life expectancies by standard life expectancies formulas[14]. Mapping techniques using ArcGIS 10.3 will be used to visualize spatial patterning in life expectancies. Hot spot analysis, such as Getis-Ord Gi statistic, will be used to determine spatial clustering of life expectancies.

Once life expectancies are calculated, this can be used as a continuous outcome in standard regression analysis. We will test the association of SES with life expectancy by weighted linear regression models:

$Y_i = \beta_0 + \beta_1 * x_1 + B_j * \mathbf{X}_j + \varepsilon_i$; where $Var(\varepsilon_i) \sim N(0, \sigma^2/w_i)$

Where: x_1 = Census tract level SES
$\mathbf{X}_j$ = Matrix of j covariates for adjustment
β_0 = common intercept
β_1 = slope term for SES measure
B_j = slope terms for covariates
ε_i = Error term
w_i = standard deviation of life expectancy for i^{th} census tract

To test for trends over time, the above model can be expanded to a two-level hierarchical model with Level 1 being time nested in census tracts; Level 2 being census tracts. The inclusion of a main effect of time and interactions with SES and each of the covariates will be included to test for time trends.

Timeline

Jan 2016 – Obtain data
Feb-Mar 2016 – Data cleaning, creation of variables
Apr-May 2016 – Data Analysis
Jun-Jul 2016 – Draft Manuscript
Aug 2016 – Manuscript submission to journal

UHC Staff Assistance Required

UHC staff assistance required:
GIS support – Create maps of life expectancy and death rates; Creation of neighborhood level measures.
Data Analysis support – Statistical analysis of life expectancy and death rates

References

1. Getis, A. and J.K. Ord, *The Analysis of Spatial Association by Use of Distance Statistics.* Geographical Analysis, 1992. **24**(3): p. 189-206.

2. OpenDataPhilly. *About OpenDataPhilly*. 2015 [cited 2015 2/25/2015]; Available from: https://www.opendataphilly.org/about.
3. Public Health Management Corporation. *Community Health Data Base: Southeastern Pennsylvania Household Survey 2002-2012.* 2012 [cited 2015 2/25/2015]; Available from: http://www.chdbdata.org/.
4. Mujahid, M.S., et al., *Assessing the measurement properties of neighborhood scales: from psychometrics to ecometrics.* American journal of epidemiology, 2007. **165**(8): p. 858-67.
5. City of Philadelphia and UVM Spatial Analysis Laboratory, *High Resolution Landcover, Philadelphia PA, 2008*, City of Philadelphia and University of Vermont Spatial Analysis Laboratory, Editors.
6. Rodriguez, D.A., et al., *Land use, residential density, and walking:: The Multi-Ethnic Study of Atherosclerosis.* American Journal of Preventive Medicine, 2009. **37**(5): p. 397-404.
7. Troped, P.J., et al., *Relationships between the built environment and walking and weight status among older women in three U.S. States.* J Aging Phys Act, 2014. **22**(1): p. 114-125.
8. Forsyth, A., et al., *LEAN-GIS Protocols (Local Environment for Activity and Nutrition-Geogarphic Information Systems) Version 2.1*. 2012.
9. DVRPC. *Geospatial Data*. 2014 [cited 2015 2/25/2015]; Available from: http://www.dvrpc.org/Mapping/data.htm.
10. MRLC. *National Land Cover Database (NLCD) 2011 Product Data Downloads*. 2015 [cited 2015 2/25/2015]; Available from: http://www.mrlc.gov/nlcd11_data.php.
11. Philadelphia Dept of Public Health, *2014-2015 Air Monitoring Network Plan*. 2014.
12. Philadelphia Dept of Public Health, *Philadelphia's Air Quality Report 2010*. 2010.
13. Subramanian, S.V., C. Duncan, and K. Jones, *Multilevel perspectives on modeling census data.* Environment and Planning, 2001. **33**: p. 399-417.
14. Chiang, C.L., *The Life Table and Its Applications*. 1984, Malabar, Florida: Robert E Krieger Publishing Company.