The Importance of Region of Birth and Early Life vs. Current Residence to Health and Mortality

Scott M. Lynch
Department of Sociology and Office of Population Research
Princeton University

J. Scott Brown
Department of Sociology and Gerontology and Scripps Gerontology Center
Miami University

Miles G. Taylor
Department of Sociology
Florida State University

Abstract

Survey respondents' region of residence is commonly used at least as a control in demographic research on health. An implicit rationale for including region is that regional cultures vary, and cultural practices may affect health. For example, dietary differences between Southerners and others may produce rates of obesity and related health complications that differ compared to those of other regions. Under this argument, an individual's region of birth and childhood should have greater later-life health implications than his/her current region, because early socialization is strong and enduring. Using Bayesian multistate life table methods applied to HRS data from 1998-2008, we investigate regional differences in healthy life expectancy at age 50, where region of residence is measured three ways: (1) current region, (2) birth region, and (3) region during adolescence. We find that the latter measure of region is most important in predicting and differentiating health across regions. Implications are discussed.

The Importance of Region of Birth and Early Life vs. Current Residence to Health and Mortality

[NOTE: This is a very rough extended abstract but gives an idea of our basic question]

Regional Variation in Health in Extant Studies

The use of region of residence as a control variable is pervasive in demographic research on health. Most, if not all, studies that incorporate region into models use the respondents' current region of residence at the time of interview, both in cross-sectional and in longitudinal studies. In cross-sectional studies, region is usually measured in only one way. In longitudinal studies, many studies use only baseline region of residence and ignore movement between regions across time. A few studies actually measure region of residence across time, but few if any have considered that region early in the life course may play a role in influencing health in later life.

What Does Region Mean?

From a sociological perspective social structure and culture shape individual behavior and provide a context within which individuals live out their lives. That is, individuals make choices within the constraints presented by the sociocultural context within which they live. From that view, the region in which an individual currently lives may be an important determinant of a variety of behaviors, under the assumption that culture varies across regions. Studies suggest that regional cultures do in fact vary, as evidenced by the extent to which measures of region almost always have statistically significant effects on health. In research on physical health, for example, it is known that southerners tend to have lower life expectancy and worse health than those from other regions. As a contributing factor to excess mortality, Southerners tend to have higher rates of heart disease and stroke than persons in other regions, and it is commonly assumed that these differences arise from dietary differences between regions. Southerners also tend to have higher rates of obesity, and this may be due to regional human geographic differences, like lower population density and the reduced likelihood of living within, and working within, a major city where walking is a common form of exercise: In the South, most persons drive to and from almost all destinations. Other regional differences exist in mental health. Studies have found, for example, that anxiety levels are higher in the Northeast than in other regions, while personality disorders are more prevalent in the West.

More broadly, other regional variation exists that may be relevant to health. Educational attainment is lowest in the south, as is median income, both of which have been shown to predict a variety of both mental and physical health outcomes. Physician density, as well as hospital density, are lowest in the south and greatest in the Northeast, potentially influencing access to health care.

Part of the explanation for such regional density differences is attributable to the extent to which regions vary in their rural/urban composition. The South, for example, has a much higher level of rurality, and rural residents tend to have worse health outcomes than urban and suburban residents. Occupational outcomes are also stratified by rurality and by region. In rural areas,

agricultural and related occupations (e.g., textile manufacturing) are more prevalent. These occupations are marked by lower levels of employer-based health insurance. In contrast, in the upper midwest (defined as the "east north central" region by the Census Bureau), steel and automobile manufacturing have been traditional occupations, and these industries are marked by high levels of employer-based health insurance provision due to union influence.

A life course perspective adds an important dimension to understanding the influence of region on health. This perspective suggests that, while current region may be important to health for primarily external reasons, like physician density, etc; that, is, physical and demographic—social structural characeristics of the region—the region the respondent was born into and raised may also be critically important. In particular, primary socialization occurs early in life and is enduring. Thus, individuals internalize the culture of the region they lived in during their "formative" years and likely carry this culture with them regardless of where they may later live. Thus, in this research, we ask whether the region in which the respondent currently resides, or the region in which the respondent resided in early life matters more for predicting later life health.

Data

The data for this paper are from the 1998-2008 waves of the Health and Retirement Study (HRS). In brief, the HRS is a panel study with numerous waves of measurement that began in the early 1990s as two separate studies that were later merged. In 2004, the study was augmented with the addition of new respondents. In our analyses, we limit the data to persons who were in the study in 1998 and were between the ages of 51 and 60. These restrictions yield an initial sample size of 6,566 persons who were potentially measured in 1998, 2000, 2002, 2004, 2006, and 2008.

Our key variable in the analysis—region—was measured at each wave of the study. Additionally, the region in which the respondent was born was measured, as was the region in which the respondent spent most of his/her time growing up.

We employ two strategies for the measurement of region across time. Under one approach, we measure region using the four category measure used by the Census; under the other approach we measure region using the Census' expanded nine category measure. The four category measure captures four US regions: Northeast, Midwest, South, and West. The nine category measure breaks these four regions into subregions. The Northeast is composed of New England (ME, VT, NH, MA, RI, and CT) and mid-Atlantic states (NY, NJ, PA). The midwest is composed of east north central (OH, IL, IN, MI, WI) and west north central (MN, IA, MO, KS, NE, ND, SD) states. The south is composed of west south central (LA, AR, TX, OK), east south central (KY, TN, AL, MS), and south-Atlantic (DE, MD, DC, WV, VA, WV, VA, NC, SC, GA, FL) states. Finally, the West is composed of Mountain (NV, NM, AZ, ID, UT, CO, MT, WY) and Pacific (WA, OR, CA, HI, AK) states.

Our primary health outcome measure is a self-rated health measure, which was measured with five categories, including "excellent," "very good," "good," "fair," and "poor" response categories. Self-rated health (SRH) is a commonly-used measure for assessing health status, and

it is viewed as both a reliable and valid measure. It is strongly related to objective physical and mental health and provides a good summary measure of overall health. For our life table analyses (see next section), we collapsed this health measure into a dichotomous meaure indicating whether the respondent had fair or poor (=1) versus better health (=0).

In addition to region and the health outcome measures, we also include age (measured in two-year age groups), sex (male=1; female=0), race (nonwhite=1; white=0), and years of schooling at baseline.

Methods

In addition to basic descriptive analyses documenting regional transitions from birth through adolescence through the 6 waves of the study, we use two methods of analyses: basic regression models of self-rated health as a preliminary step, and multistate life tables. In the regression models, we estimated four preliminary models. In model 1, health was regressed on current region of residence. In model 2, health was regressed on birth region. In model 3, health was regressed on region during adolescence. Finally, in model 4, health was regressed on all three measures.

In our multistate life table analyses, we estimate healthy life expectancy with all three sets of region measures (birth, adolescence, and current) as predictors. Standard multistate life table methods cannot be used in these analyses for three reasons. First, typical multistate methods require the disaggregation of data, and the sample size of the HRS is insufficient for providing stable estimates of transition probabilities—the key input for multistate life table calculations—across age by region of birth and current residence. Second, total and healthy life expectancy are dependent on sex, race, and socioeconomic status, and each of these statuses is related to region of birth and region of current residence. Therefore, any analyses of state expectancies by region must be able to control for regional differences in these characteristics. Yet, again, the sample size of the HRS limits our ability to control on these characteristics via the usual procedure of disaggregation. Third, standard multistate life table techniques assume that the data on transitions are measured at the population level. In these analyses, we use sample data rather than population data, and so our modeling strategy must be able to compensate for the uncertainty inherent in using sample data. Standard errors—i.e., estimates of sampling (and other) uncertainty—cannot be obtained via traditional methods.

Because of these limitations of traditional methods, we therefore use the method described in Lynch and Brown (2005 *Sociological Methodology*). In brief, the method involves structuring the data into spells and estimating a bivariate probit model with mortality and health at T2 predicted by birth region, region during adolescence, and T1 region, health status at T1, and the other covariates as predictors. The parameters of the bivariate probit model are simulated from their posterior distribution using Gibbs sampling (this strategy is akin to producing a sample from the sampling distribution of the parameters; see Lynch 2007 for general discussion of Gibbs sampling). This approach yields a multivariate distribution of the model parameters. The distribution of parameters is then applied to desired covariate profiles (one covariate being age, which is incremented during this process) to produce distributions of age-specific predicted z scores, which in turn are transformed into age-specific transition probabilities. These age-

specific transition probabilities are transformed into continuous time age-specific hazard matrices, and traditional multistate life tables are constructed from these distributions of smoothed hazard matrices (see Schoen 1988 for a discussion of multistate life table estimation).

The net result of this process is that we obtain distributions of state expectancies, which can then be used to construct interval estimates of healthy life expectancy, facilitating statistical comparison (i.e., hypothesis testing) of subpopulation differences. Our key comparison of interest is the magnitude of regional differences in healthy life expectancy under the different measurements for region.

Very Preliminary Results

At this point, we have not conducted the multistate life table analyses, but we have estimated some preliminary regression models. In the first model, when only current region is included as a predictor of health, regional differences tend to be statistically significant. In the second model, when only birth region is included as a predictor of health, regional differences also tend to be statistically significant, but larger in magnitude than in Model 1. In the third model, when only adolescent region is included as a predictor of health, regional differences are significant and of even greater magnitude than in Model 2. Finally, in Model 4, with all three measures of region included, only adolescent region differences reach statistical significance.

These results are very preliminary, but they support the cultural argument presented above. Differences in region of residence can be viewed as an indicator of cultural differences. Given that the most enduring effects of socialization occur at early ages, we expect that region of residence at birth and adolescence to be more important than region of residence in adulthood. Thus, when measures of region are compared, the measures of region in earlier life fully account for the effect of later life region measures. The extent to which later life region measures matter therefore may simply be a reflection of the fact that many people do not change regions across the life course. Thus, current region is merely a proxy measure for early life residence, but it has substantial measurement error, because many people do move. These and additional issues will be discussed.