Dr. Megan Andrew
University of Notre Dame

Dr. Maggie Hicken
University of Michigan

SOCIOECONOMIC SEGREGATION AND INFANT HEALTH IN THE AMERICAN METROPOLITAN, 1980-2000

INTRODUCTION AND BACKGROUND

The sociology of place has a special, long-standing status in the discipline. American sociologists in the last century theorized extensively about the importance of place and its effects on individual and group outcomes (Gans 1968, Gieryn 2000 for a review), and important streams of contemporary research have followed that tradition (e.g., Sampson 2013). This emphasis on the sociology of place has particular relevance for the study of health disparities. Scholarly interest in the relationship between place and health disparities has increased markedly over the past decade, and health scholars increasingly recognize not only the importance of place for health but the policy value of altering group contexts and infrastructure (Diez-Roux 1998). Following larger sociological traditions emphasizing the importance of place as well as recent scholarship on the relationship between place and health disparities, we study the relationship between different dimensions of residential segregation and infant health.

Residential segregation is a macro-level measure of social influence and resources, and race-ethnic segregation in particular has a robust demonstrated association with various population and individual health outcomes (e.g., Morenoff 2003, Bell et al. 2006, Kramer et al. 2008, Osypuk and Acevedo-Garcia 2008). Though a large literature demonstrates important relationships between residential segregation and health, further study of these relationships is warranted given two recent trends. First, race-ethnic residential segregation that heavily marked the last century has since declined. Second, residential segregation by income and particularly education has increased over the same period. Massey, Rothwell, and Domina (2009) note this rather marked shift in residential segregation from race-ethnicity to socioeconomic status over the last 40 years, christening the upward trend in socioeconomic segregation as "post-modern segregation."[1] Despite its likely importance, the implications of a shifting segregation regime are not well understood. Little research directly addresses newer forms of residential segregation—particularly segregation by education—and most research on residential segregation evaluates its different forms in isolation of each other (e.g., Mayer and Sarin 2005).

This gap in the literature on residential segregation and health limits our understanding of the importance of residential segregation across a variety of health outcomes. For example, income and education segregation both increased after 1970. Because education is an important progenitor of income, analyses focusing exclusively on increasing income segregation may mistakenly interpret observed relationships between income segregation and outcomes of interest. It may be that education segregation is the real driver of an outcome of interest such as health disparities. Moreover, income and education determine different aspects of health. While

[1] See recent work by Reardon and Bischoff (2011b) on trends in income segregation.

education is related to the onset of poor health, income is related to the progression of morbidity at the individual level (Herd et al. 2005, Cutler et al. 2008). Similar dynamics may play out at the population level. A further complication is the extent to which race-ethnic segregation alters relationships between income and education segregation and outcomes of interest. This possibility is highlighted by the literature on ethnic and immigrant enclaves. That body of work suggests that while individuals may live in communities marked by high levels of poverty and/or low levels of education, they may thrive regardless because of protective social relationships afforded by a community of peers similar to themselves (Cutler and Glaeser 1997, Cutler, Glaeser, and Vigdor 2008).

Given important changes in the bases of residential segregation in recent decades and remaining ambiguity about the relationships between the different dimensions of residential segregation and health, we estimate models that describe the relationships between race-ethnic, income, and education segregation and infant health at the metropolitan statistical area (MSA) and individual-levels from 1980 to 2000. We focus on infant health because of its particular sensitivity to social and economic conditions in even developed countries (Pampel & Pillai, 1986; Wise & Pursley, 1992) and given the relatively high rates of infant mortality in the U.S. compared to counterpart countries (OECD 2006). Using standardized census and Vital Statistics data, we ask and answer 3 research questions in this paper:

1) How does race-ethnic, income, and education segregation across metropolitan statistical areas in the U.S. affect infant health?
2) How do these effects change over time as socioeconomic segregation increases and race-ethnic segregation decreases?
3) How do these different dimensions of residential segregation moderate the effects of each other on health over time?

Our research makes five contributions to the literature. First, we estimate models that parse the relationships between multiple dimensions of residential segregation and infant health. Second, we employ rank-order measures of residential socioeconomic segregation that better measure differences in the concentration of persons with a given income or level of education separate from their distribution in that population (Reardon, Firebaugh, O'Sullivan, and Matthews 2006, Neckerman and Torche 2007, Reardon and Bischoff 2011a). Third, we determine the extent to which different forms, or dimensions, of residential segregation have increased in importance over time as the bases of residential segregation have changed. Fourth, we determine the extent to which the effects of different forms of residential segregation moderate and perhaps even compound one another over time. It may be that both, say, income and education segregation have moderate independent effects on infant health. But together these two forms of socioeconomic segregation may have severe (negative) effects on infant health, interaction effects that may increase over time as education and income segregation become more prevalent in a MSA. Fifth, we evaluate the relationship between changing residential segregation and both MSA and individual-level infant health outcomes. The relationship between residential segregation and MSA-level infant health follows a long tradition of ecological models and the role of metropolitan-level factors in determining population well-being (Taeuber and Taebuer 1965), while individual-level models link important ecological factors to individual well-being. Together, our ecological and individual-level models of residential segregation provide a robust analysis of the changing residential segregation regime in the U.S. and its effects on infant population health.

DATA AND METHODS

Data

We use data from the US Vital Statistics (USVS) and the Neighborhood Change Database (NCDB) to estimate the effects of different forms of residential segregation on population and individual infant health in our analysis. We estimate these models of MSA-level infant health rates and individual-level models of birth outcomes separately. The USVS provide annual data for mothers' age, education, race, ethnicity, country of birth, and marital status, for fathers' education, and for the infant's sex, birth weight, and gestational age from certificates of all live births in a given year. These birth microdata can be aggregated to the MSA-level using basic geographic identifiers in the data. We use the micro and aggregated data for the period 1980 through 2002 in our preliminary analyses given certain requisite measures for our analysis are only available from 1985 forward and because of limited data access after 2002. In planned extensions, we will add additional years of the USVS data.

We also use standardized census data from the Neighborhood Change Database in our analysis. The NCDB provides tract-level measures standardized across census years 1970-2000, allowing for easy comparisons across tracts and other, larger geographic units over time. These census tract-level data are standardized to the 2000 census tract boundaries for the 1970, 1980, and 1990 censuses. Tract standardization is especially relevant for calculating metropolitan area segregation indexes, both the standard dissimilarity index favored in residential segregation analyses as well as rank-order socioeconomic segregation measures we use here. In planned extensions, we will add 2010 standardized data as it becomes available.

Following extant research, we limit our analyses to metropolitan statistical areas with a population of at least 100,000 and with black, Latino, and white populations of at least 5000.[2] Descriptive statistics for these metropolitan areas can be found in Table 1.

Dependent Variables

We use individual-level USVS birth certificate data to construct our MSA-level dependent variables: MSA low birth weight rate and MSA pre-term birth rate. We elect to use these two measures of infant health in order to better understand how different dimensions of residential segregation may differently affect different measures of poor infant health. We calculate the MSA-level low birth weight rate as the number of singleton infants born weighing less than 2500 grams to women ages 25-54 in a given metropolitan statistical area and year divided by the total number of singleton infants born in that area and year. We calculate the infant pre-term birth weight as the number of pre-term infants born to women ages 25-54 in a given MSA and year divided by the total number of singleton infants born in that area and year. We model infant health at the individual-level using dummy indicators of singleton birth outcomes for mothers ages 25-54. Analogous to our MSA-level measures of infant health, we

[2] Some research suggests that segregation indexes are unreliable when minority populations are sufficiently small (e.g., Walton 2009). However, other research suggests indexes are insensitive to population size (Zoloth 1976, Frey and Farley 1996). We evaluate the robustness of our estimates under both conditions.

consider whether an infant was low birth weight at birth and whether an infant was born pre-term. Both metropolitan and individual-level dependent measures are taken near the time of birth in any given year between 1985 through 2002.

Independent Variables

Our key independent variables measure race-ethnic, education, and income segregation of a metropolitan area in each census year 1980-2000. We measure race-ethnic segregation using Theil's entropy index, defined as:

$$E_i = \sum_{r=1}^{R} (\pi_{ri}) \ln(\frac{1}{\pi_{ri}}) \quad , \qquad (1)$$

and

$$H_{ij} = \sum_{i=1}^{I} [\frac{t_i(E_j - E_i)}{E_j T_j}] \, , \qquad (2).$$

The first equation defines the entropy score which is a measure of diversity (Reardon and Firebaugh 2002). The second equation defines the entropy index. In equation 1, *i* indexes a given geographic subunit, in this case a census tract, and π refers to a race-ethnic group *r*'s proportion of the whole population. Note that $r \in 1 \ldots R$, where *R* represents the total number of race-ethnic groups in subunit *i*. Equation 2 defines the entropy index as the weighted average deviation of each subunit *i*'s entropyE_i from the unit *j*'s entropy E_j. In this analysis, unit *j* refers to the MSA. The t_i refers to the total population of subunit *i* and the T_j refers to the total population of unit *j* in equation 2. We prefer the entropy index over other, more standard measures of race-ethnic segregation, such as the dissimilarity index, because it similarly measures the evenness of race-ethnic groups in a given geographic unit but can include multiple race-ethnic groups. Though the residential segregation literature often focuses on blacks and whites, it makes sense to evaluate segregation among multiple groups given the marked population increase across different race-ethnic groups in the U.S., particularly Latinos.

We measure socioeconomic segregation using rank-order measures of segregation proposed by Reardon, Firebaugh, O'Sullivan, and Matthews (2006) and implemented by Reardon and Bischoff (2011). These measures build on the entropy index described in equations 1 and 2 and indicate the socioeconomic variation within a given geographic subunit relative to socioeconomic variation in the larger geographic unit. We calculate indexes separately for education and income for census tracts in MSAs. The rank-order measure is defined as:

$$E(p) = p log_2 \frac{1}{p} + (1-p) log_2 \, \frac{1}{(1-p)} \quad , \qquad (3)$$

$$H(p) = 1 - \sum_{j}^{J} \frac{t_j E_j(p)}{\mathrm{T} E(p)'} \quad , \qquad (4)$$

where *p* is the percentile rank in a given income or education distribution, *T* is the population of a MSA and t_j is the population or census tract *j* in that MSA. Following Reardon and colleagues (2006), the rank-order index of education or income segregation can be defined as:

$$H^R = 2\ln(2)\int_0^1 E(p)H(p)dp\,, \qquad (5)$$

which is the weighted average of education or income segregation across the distribution or each. This index ranges from 0 to 1 (zero indicating no segregation and one indicating complete segregation) and weights households with education or income above and below the median most heavily. Because education and income are aggregated to a limited number of categories in the census, we calculate H^R for each education or income category then estimate $\widetilde{H}^R$ using polynomial regressions and adjusting for population size and the median population income or education. Adjusting for population size ensures comparable $\widetilde{H}^R$ across geographic units since they are upwardly biased in small populations; adjusting for median population income or education ensures that ranks across vastly different population distributions of education or income refer to individuals with similar absolute levels of income or education. Reardon et al. (2006) correctly suggest their rank-order measure of segregation allows for comparisons across time and geographic units because it is independent of income (or education) distributions. However, this means that individuals with a similar rank but with vastly different incomes are treated the same. In supplementary analyses, we also estimate $\widetilde{H}^R$ for income using population size, median population income, *and* cost of living given vast geographic differences in the purchasing power of a given income level across different MSAs. To correct for bias in our estimates, we bootstrap standard errors for $\widetilde{H}^R$.

Controls

We control for a number of metropolitan and individual-level controls in our models. First, we use census information from the NCDB and the National Historical Geographic Information System (NHGIS) to construct control variables measuring the size and the age, foreign-born, gender, race-ethnicity, household structure, education, and income composition of a geographic unit's population over time. We include a single term describing the overall size of the population in the hundred thousands. We measure the age composition of the population using terms describing the percent of the population ages 17 years or younger, ages 18-30 years, ages 31-64 years, and ages 65 or older. We measure percent foreign born, female, foreign-born, black, Latino, asian, and female-headed household with a single, continuous term for each. We measure the education and income levels of the population as the percent of the population falling within mutually exclusive education and income categories. For education, we use the categories less than high school, some college, and college degree or higher. For income, we adjust census income categories for inflation (2010 dollars) and then create quartile categories for a given census year. Finally, we measure the extent of residential mobility as the percent of the population residing in the same house as five years before for individuals five years of age and older at a given census year and the percent of the population that resides in a different house in the same MSA for individuals five years of age and older.

We will consider characteristics of a MSA's economy and infrastructure that may similarly bias estimates of the effects of differences in residential segregation on differences in infant health as population composition and mobility. To construct measures of health and education infrastructure, we use information from the Area Resource File (ARF) and the Integrated Post-Secondary Data Set (IPEDS). The ARF provides data on hospitals and health care professionals in a given geographic unit. Using these data in the years it is available, we will construct MSA-level measures of the number of hospitals and the number of physicians in a MSA unit. We will use data from the IPEDS to measure the number of four-year colleges and universities at the state and MSA-level. We also use census data from the NCDB and the NHGIS to measure MSA-level economic characteristics. Along these lines, we measure the percent unemployed and the percent employed in manufacturing, construction, finance/insurance/real estate (FIRE) for civilians 16 years old or older.

In individual-level models, we control for a number of mother and child measures constructed from the USVS data. We specifically control for mothers' age in years at the birth, race-ethnicity, country of birth, and marital status. We control for mothers' and fathers' education using the categories less than high school, some college, and college degree or higher. We also control for whether the mother smoked during pregnancy. The education and smoking measures in particular are helpful for addressing individual selection into a given metropolitan area. However, even with controls for mothers' and fathers' characteristics, it remains possible that mothers select into metropolitan areas based on other relevant unobserved characteristics. We discuss modeling considerations below that help address concerns about selection on unobservables into metropolitan areas that might bias our estimates of the effects of metropolitan race-ethnic and socioeconomic segregation on individual infant health in particular.

Metropolitan Infant Health Model

In the first part of our analysis, we aim to estimate the causal effects of changes in different forms of residential segregation on changes in metropolitan infant health. We begin with a pooled model of metropolitan infant health with robust standard errors. This model is written as:

$$y_{it} = \alpha + \delta_t + \gamma_s + \tau_r + \boldsymbol{x_{it}}\boldsymbol{\beta} + \varepsilon_{it} \qquad (5),$$

where y_{it} is the infant pre-term birth rate or the low birth weight rate for the MSA *i* at time *t*. In our analysis, we evaluate time across three census years: 1980, 1990, and 2000. The term α describes a time-invariant constant across units *i* and time *t*, the term δ_t describes the time trend in infant pre-term rate across units *i*, the term γ_s describes state fixed-effects, and the term τ_r describes census region fixed-effects. The vector $\boldsymbol{x_{it}}$ contains terms describing unit and time-varying characteristics, including our primary independent variables of interest—different forms of residential segregation—as well as controls for observed differences in MSAs. The vector $\boldsymbol{\beta}$ describes the relationships between the terms in the vector $\boldsymbol{x_{it}}$ and infant pre-term rate or low birth weight rate y_{it}, averaged across units *i* and time *t*. Robust standard errors account for dependence across observations of the same MSA unit *i* over time *t*. This simple model obviously makes a number of rather strict assumptions, namely that regressors are exogenous and that the ε_{it} are serially uncorrelated within and between units *i* over time *t*, both conditional on the δ_t , γ_s , τ_r , and $\boldsymbol{x}$. Given our explicit interest in the causal effects of different forms of residential segregation on metropolitan infant health, we next estimate a model of infant health that differences out time-invariant metropolitan-specific unobserved heterogeneity. This differencing makes assumptions about exogeneity and serially uncorrelated errors more plausible. The basic first-difference model can be written as:

$$(y_{it} - y_{i,t-1}) = (\delta_{t>1} - \delta_{t=1}) + (\boldsymbol{x_{it}} - \boldsymbol{x_{i,t-1}})\boldsymbol{\beta} + (\varepsilon_{it} - \varepsilon_{i,t-1}) \qquad (6).$$

The notation is defined as before in equation 5, but key differences are evident between the two models. The model in equation 6 explicitly differences out unobserved, time-invariant unit-specific characteristics that likely induce correlations between the ε_{it} and the $\boldsymbol{x_{it}}$ as well as correlations between the ε_{it} . The case for exogeneity and no serial correlations in ε_{it} is obviously stronger in the first-difference model than in the pooled model in equation 5. However, we are unable to estimate state and census region differences as in equation (5) since these do not vary across time (i.e., $(\gamma_{s,t} - \gamma_{s,t-1}) = (\tau_{r,t} - \tau_{r,t-1}) = 0$). We further extend the first-difference model in equation 6 by making different assumptions about the structure of the correlation matrix for the ε_{it} . Specification checks will include autoregressive correlations with lag-1 as well as stationary and non-stationary correlation structures.

Based on the equation 5 defined above, we estimate a number of specific models. We first estimate the effects of different forms of residential segregation on rates of low infant birth weight and infant mortality across metropolitan areas using six basic models. We estimate these models separately each of our two dependent variables. The first model we refer to as the total residential segregation model. We simply estimate the effects of education, income, and race-ethnic segregation on each measure of infant health. We then introduce control measures described above in what we refer to as the net residential segregation model. We next estimate the net trend residential segregation model. This model allows the effects of residential segregation in a geographic unit to vary over time $t = 1, 2, \ldots T$ by interactions between the three types of residential segregation we consider and $T - 1$ dummies. (Because we have three panels spanning 1980 to 2000, we include dummy terms for two time points.) We then estimate a net trend and geographic fixed-effects segregation model that builds on the previous model to include fixed-effects for state and census region. We then determine the extent to which different forms of residential segregation moderate one another over time in a fifth model. We refer to this model as the interacted residential segregation model. As the name implies, we introduce three terms that multiply the different forms of residential segregation by one another: education segregation by income segregation, income segregation by race-ethnic segregation, and education segregation by race-ethnic segregation. Following estimation of the interaction residential segregation model, we next estimate the trend interaction residential segregation model. This sixth model replicates the interaction residential segregation model but introduces three-way interactions between different forms of residential segregation and time t. This model tests the possibility that certain forms of residential segregation increase or decrease the effects of one another over time as residential segregation shifted from race-ethnic to socioeconomic segregation. These six models constitute our pooled models of the effect of different dimensions of residential segregation on population infant health. We replicate these models to the extent possible using the first-difference specification in equation 6.[3] Having estimated the basic pooled and first-difference segregation models, we then explore the error structure of the first-difference model. To explore the error structure and basic assumptions about serially uncorrelated errors, we first estimate a lag-1 autoregressive correlation model. In separate specifications, we then estimate a stationary and non-stationary specification, respectively.

[3] Please note we are unable to include state and census region fixed-effects in our first-difference models.

Individual Infant Health Model

In the second part of our analysis, we evaluate the effects of metropolitan segregation on individual birth outcomes. These models follow from equation 5. In the individual infant health models, however, we include MSA fixed-effects in addition to state and census region fixed-effects. The geographic fixed-effects allow us to net out unobserved metropolitan, state, and census region characteristics related to infant health outcomes. Noted elsewhere, selection into neighborhoods and metropolitan areas based on unobserved individual characteristics is a major concern with research linking individuals to a given social context such as metropolitan areas (e.g., Fauth and Brooks-Gunn 2008). In that vein, we will estimate models of mothers' smoking behavior during pregnancy and mothers' prenatal care. Fauth and Brooks-Gunn (2008) note models of intermediary behaviors linking residential segregation and infant health like mothers' smoking and prenatal care are useful for tightening the link between residential segregation and infant health. We are also considering instrumental variable models using exclusion restrictions on metropolitan governance structure, public finance, and topography similar to Cutler and Glaeser (1997) to further address the issue of individual selection into metropolitan areas.