# Income Mobility, Mortality and Health

Sebastian Daza
*Center for Demography and Ecology*
*University of Wisconsin-Madison*
sdaza@ssc.wisc.edu

January 26, 2017

10,025 words

# 1 Introduction

Recent research on U.S. income mobility and health using community and individual data shows that higher mobility is associated with lower mortality risks (Palloni et al. 2016; Venkataramani et al. 2015, 2016). That relationship seems to be stronger and more consistent than the relationship between income inequality and health (a topic widely studied), although considerably smaller than the impact of income on mortality (Palloni et al. 2016). This preliminary evidence suggests that income mobility might be a relevant determinant of health and mortality. Surprisingly, this potential pathway has received little attention in the literature.

This dissertation builds on that small literature and examines the robustness of the relationship between income mobility and health at the aggregate and individual level. By using different data sources and modeling approaches, I describe the magnitude and variability of this association in the U.S., and explore the plausibility and consistency of explanations offered in the literature. The central argument is that the effect of income mobility on health is stronger and larger than the impact of income inequality and that the mechanisms behind it, although related to income inequality, are theoretically distinct and independent of those of income and inequality, and can have powerful and lasting consequences.

To ground this argument, I use two strategies. First, I analyze aggregate and individual data to assess the magnitude, robustness, and variability of the *effects* of income mobility[1], and empirically examine whether some of the potential pathways and mechanisms proposed in the literature are *supported* by the data. Secondly, building on this evidence, I use a theoretical model to assess the conditions and plausibility of the potential mechanisms involved in the association between income mobility and health. By formulating a *generative model*, where I precisely define (represent) a set of mechanisms (causal relationships) likely to bring about the observed patterns, I am able to assess the internal consistency of the theory and evaluate its generative sufficiency (Billari and Prskawetz 2003; Epstein 2006; Miller and Page 2007; Manzo 2014; Page 2015). These *virtual* representations help me explore the implications of the theory and ask *what-if* questions. The goal is to go beyond statistical models (i.e., the detection of average differences that may reflect the aggregate statistical signature of *unspecified* underlying mechanisms), and to approximate plausible mechanism behind the observed patterns.[2]

In the following sections I outline the problem, formulate a theoretical framework and propose two analytic strategies to examine it. Each of these sections will correspond to a chapter in the dissertation. I will add an introduction and conclusion. Thus, the dissertation will have an introduction, four chapters, and a conclusion.

[1] I use the term *effects* in a loose way and mean the size and sign of (path) coefficients measuring the strength of the relation between two variables. I do not presume the existence of a proven causal relation.

[2] Models are just learning tools that, even failing to create a model that generates the observed patterns through the computational representation of the mechanisms suggested by theory, help understand the system and phenomenon of interest.

## 2 The Problem: Health Disparities and Stratification

Socioeconomic inequality in health and mortality is one of the persistent regularities observed across societies since at least the turn of the 20th century. All countries, even those with high economic prosperity and human development, show systematic inequalities in mortality and morbidity. These inequalities are not small. They ranges from 5 to 10 years of life expectancy at birth, and from 10 to 20 years of disability-free life expectancy (Commission on Social Determinants of Health 2008). In a recent paper, Chetty et al. (2016) used administrative data to show that the gap in life expectancy between the 1% richest and poorest 1% of individuals in the U.S. was about 14 years. These disparities persist across multiple measures of individual social and economic status (e.g., education, income, wealth, occupation) and there is evidence they have recently widened in high-income countries (Mackenbach 2012).

These social disparities in health are usually explained by the characteristics of the social stratification system. The key argument behind the theory of *fundamental causes* of health disparities (Link and Phelan 1995; Phelan et al. 2010) is that inequality in health is caused ultimately by social stratification processes, not the exposure to proximal risk factors (e.g., smoking). Individual socioeconomic status would provide *flexible resources* (e.g., power, prestige, money, social connections) that can be used to avoid or minimize the consequences of diseases. These resources are contingent, changing and continuing to exist, reproducing the association between socioeconomic status and health over time. To explain health disparities, thus, is necessary to identify features of the stratification system that influence health status. According to Grusky (2014), for instance, three main elements define any stratification system: (1) mobility mechanisms that *sort* individuals into social strata, (2) allocation rules that distribute resources to social strata, and (3) social processes that provide more value to some resources than other. Correspondingly, health inequalities might be understood as a function of (1) social mobility and the differences of people across social strata, (2) resource distribution or differences in access to material and non-material resources across social strata, and (3) the value of those resources to avoid health problems (Mackenbach 2012). All these processes and mechanisms might vary over time and between geographic areas, giving rise to differences in the magnitude of health inequalities across regions, nations, and periods. The first mechanism of stratification (i.e., sorting) involves processes such as residential segregation or health selection. Individuals in poor health early in life, for example, might be less likely to experience upward mobility and more likely to experience both downward mobility and have poor health.[3] The remaining two mechanisms of stratification are related to the effect of allocation of resources (e.g., income) on health and its interaction with characteristics of the prevailing epidemiological regime.

Mechanisms that produce health disparities, however, are not limited to sorting, differential resource abundance, and heterogeneity of resource productivity. It has also been argued that factors such as the distribution of income in the population (i.e., income inequality), the relative position of individuals according to income, and income mobility might have a *contextual* impact on individual health and mortality. In the next section, I briefly discuss the literature on income inequality and health. Although this project is about the link between social mobility and health, it is worth reviewing the mechanisms behind the relationship between income inequality and health to gain conceptual clarity and establish common ground with that rich research tradition. I focus mainly on theoretical and conceptual issues rather than methodological problems. Then, I link that discussion to the main topic of this dissertation, income mobility and health.

[3]The only requirement for health selection is that at some point in the life course of individuals' resources depends on previous health status (Palloni et al. 2009). There are good reasons to expect selection processes. Leigh et al. (2009) note that poor health can have serious consequences on educational outcomes, employment, and marriage opportunities. It should not be surprising that reproduction of income inequality occurs, at least in part, through an attainment process: poor health of low-income children translates into lower levels of educational attainment and, as a result, lower earnings during adulthood.

## 2.1 Income Inequality and Health

Income inequality is, by definition, an ecological variable (i.e., aggregate) that describes the scale of income differences across a population. The *income inequality hypothesis* states that a *contextual* factor such as income inequality affects individual's health. The link between income inequality and health, however, could be generated by different mechanisms or processes. Evans et al. (2004) provide a useful typology that summarizes why one might observe a correlation between income inequality and health:

1. *Mechanical* effects. If individual economic status is associated with health, then an increase in economic inequality will lead to increase in inequality in health. This effect, however, is not due to material factors only. *Scarcity* may impose a cognitive *bandwidth tax* that interferes with decision making and long-term planning, increasing the adoption of risky behaviors (Mullainathan and Shafir 2013).

2. *Relational* or slope effects alter the magnitude of the relationship between economic status and health, and change the relationship between income inequality and health even when the distribution of income remains constant. These effects can be entirely exogenous (e.g., improving access to health care) and attenuate the link between income and diseases.

3. *Functional form* effects are related to the shape of the relationship between economic status and health. For instance, if the relationship is concave, the average health in society will improve as the average income increases and inequality of income decreases.

4. Finally, Evans et al. (2004) suggest the importance of *externality* effects. These effects occur when a contextual condition triggers individual processes that are detrimental to health. For instance, living in a social context of high inequality might intensify the feeling of relative deprivation among low-income persons, leading to higher levels of violent crime, even if individual economic status has no relationship to the outcome of interest. Thus, changes in other people's income can modify customs, ideals, politics, and the behavior of those whose own income remains unchanged.[4]

In order to formalize the relations conjectured by the *income inequality hypothesis*, I use the notation proposed by Wagstaff and van Doorslaer (2000). At the individual level, we can express the *income inequality hypothesis* as:

$$h_i = f_i(y_i, I_c) \tag{1}$$

where $h_i$ is the health of individual $i$, $y_i$ is individual's income, $I_c$ is a measure of income inequality at the community level[5], and $f_i$ is a function through which inputs are transformed into health. This formula represents independent effects of the terms included in the (unspecified) functional form $f_i$. It is possible to define a case where it is not the community's inequality that matters but the national level of inequality (let's say $p$), in which case we replace $I_c$ by $I_p$ (see, for instance, Zheng 2012).

[4]Differences in health between two communities might be due to characteristics of the individuals living in those communities (*composition effect* and sorting), exposure to local factors that affects individuals' health (e.g., smoke from factory, community resources), or social influence ( individuals who live in proximity are more likely to assimilate the same type of behaviors, Christakis and Fowler 2007). It is likely that these three processes are simultaneously affecting health outcomes.

[5]Community can refer to a county, state, a region, or any area smaller than a county (e.g., Census tract).

At the community level, the expression is:

$$h_c = f_c(y_c, I_c) \tag{2}$$

where $c$ is an index for the level of aggregation (e.g., census tract, county, state). Again, $I_c$ can be replaced with $I_p$ (income inequality at the national level or a larger geographic unit). Using this aggregate model, however, it would be difficult to distinguish between some of the effects discussed by Evans et al. (2004). Aggregate analyses, nevertheless, should not be discarded by invoking the ecological fallacy or other inference difficulties.[6] These analyses can provide clues on how contextual effects observed at the individual level aggregate at the community and population level. Given the complexity of the individual and contextual effects of stratification processes on health, aggregation might not be linear and straightforward.[7] Pickett and Wilkinson (2015), for instance, note that effects of income inequality are stronger in large areas because in that context income inequality serves as a measure of the scale of social stratification. Income inequality in small areas is affected by the degree of residential segregation. Higher segregation would increase the inequality between areas, but also increase homogeneity within them (Chen and Gotway Crawford 2012).

It has been argued that income inequality affects the social fabric, reducing both social capital and mutual trust, and that the resulting stresses and lack of public investments harm the health of both wealthy and poor (although one group may be more affected than the other, Wilkinson 1992; Kawachi et al. 1997). Income inequality increases social distances, accentuating social class or status differences, and strengthening causal processes whereby social class marks individuals throughout life. Differences produce negative emotions, such as shame and distrust, that could translate into poorer health by psycho-neuro-endocrine mechanisms and induce behaviors such as smoking (Wilkinson and Pickett 2009). In addition, social comparison to higher-ranking friends and neighbors might cause stress and ill-health among people who are poorer than their reference group (Wilkinson 1992; Marmot 2004). The so-called *relative deprivation hypothesis* assumes that negative effects outweigh the benefits of having richer neighbors who may pay for better neighborhood amenities. Another explanation suggests that the diffusion of innovations (knowledge or technology) among the most disadvantaged groups is slower when income inequality is high, increasing health disparities by improving the health of the wealthy when these innovations begin and benefiting the poorer only after long lags (Phelan et al. 2010).

Another approach to explain the link between income inequality and health is a neo-material reinterpretation proposed by Lynch et al. (2004). According to these authors, structural causes of inequalities are the root of health disparities. Income inequality would be only a manifestation of a cluster of *neo-material* conditions that affect population health (Lynch et al. 2004). Thus, the association between income inequality and health would be contingent and reflect people's lack of resources, both at the individual and community level. This interpretation is an explicit recognition that the social, political and economic processes that generate income inequality influence personal resources and also have an impact on public resources such as schooling, health care, social welfare, and working conditions (Lynch et al. 2004).[8] Thus, according to this perspective, one would have to account for material factors at the community level to estimate the

[6] Trends at multiple levels of analysis may or may not be similar and their causes may or may not be the same (Snijders et al. 2011).

[7] Health often reflects cumulative exposures over a long period of time. Thus, health at adult ages can be partly determined by exposure to biological and social factors early in life so that the roots of health inequalities may well lie in inequalities experienced back then.

[8] Gilens (2012) provides a political interpretation of these processes and notes that income inequality increases the political influence of the wealthy. If their policy preferences restrict public goods that provide health benefits (e.g., education, sanitation, policing), the health of the poor may be harmed.

independent effect of inequality on health.

## 2.2 Income Mobility and Health

*Can these hypotheses be extended to income mobility? If so, what consequences and patterns should be expected?* Unlike income inequality, mobility can be measured both at the individual level (inter and intra-generational socioeconomic changes over the life course) and aggregate level (earnings elasticity, rank-rank slope).[9] Thus, unlike income inequality, income mobility is not always an ecological or contextual variable. There is a large body of research on individual mobility on health and mortality. The work by Sorokin (1959) and the *mobility effects* literature, for example, suggests at least three individual pathways through which social mobility might affect health at the individual level. These mechanisms are associated with the *dissociative*, *falling from grace*, and *acculturation* hypotheses.

The *dissociative hypothesis* holds that mobility will have a negative causal effect on well-being, psychological distress, and health. Any changes in social class (up or down) could create chronic strain and generate a permanent negative impact on mental health and well-being. The *falling from grace* hypothesis predicts that *only* downward intragenerational class mobility is detrimental to health, specially in social and historical contexts that foster high social expectations for economic success (Newman 1999). Downward mobility often entails an involuntary loss of an achieved status, and signals a failure to live up to social expectations of individual success. Because downward mobility is more likely to be negative and involuntary, it should have negative consequences for health, independently of the reduction of material conditions. Finally, the *acculturation* hypothesis (Blau 1956) argues that mobility is neither a cause nor consequence of health, but that socially mobile individuals come to experience similar levels of health as those who share their social class of destination. Under this hypothesis, social mobility would be unrelated to distress and health (after controlling for income or social class).

The focus of this dissertation, however, is *not* on the health consequences of individual experiences of socioeconomic mobility but on the *contextual* impact of a place's income mobility on the health of individuals who experience such mobility regime. As in the case of income inequality, one can think on income mobility as a contextual factor that causally affects health and health disparities. Recent evidence suggests that social mobility has a positive relationship with health and mortality eve n after adjusting for income at the community and individual level, and income inequality (Palloni et al. 2016; Venkataramani et al. 2015, 2016). That is, at a given level of inequality, people living in communities with a more flexible stratification regime have lower mortality and better health. Theoretically, it is possible to think on societies with different levels of inequality and social mobility provided that both characteristics of the stratification regime, although related, are produced by different mechanisms. To empirically disentangle the (independent) effect of income inequality and mobility, however, the relationship between both aspects of the stratification system should be moderate (or not too strong), so that we can observe communities being simultaneously characterized by unequal income distribution and flexible mobility regimes, and vice-versa. Evidence on the *Great Gatsby Curve*, according to which countries with higher income inequality score low in social mobility, is mixed and sensitive to the mobility measures used (Winship 2015; Chetty et al. 2014a). However, the relationship between income inequality and opportunity, even when considering the largest estimates, is far from perfect, providing empirical grounds to estimate the independent effect of income mobility and inequality.

[9] See Torche (2015) for a review of mobility measures.

*What mechanisms would explain the link between health and mortality and income mobility at the community level?* Communities with high-income mobility may share properties that help individuals and families manage resources available to them, improve resilience to confront adverse conditions and, ultimately, reduce individual exposure to health risks irrespective of a community's income levels and aggregate income inequality.[10] For instance, credible prospects for upward income mobility might encourage people to invest more in their health (i.e., higher expected returns to health investments, Venkataramani et al. 2015). Also, optimism about the quality of one's future may raise the desire to achieve good health, independently of the possibility of obtaining greater income by doing so (Venkataramani et al. 2015). A pessimistic interpretation of this mechanism holds that a rigid stratification system fosters individual hopelessness, weakens aspirations and, more generally, diminishes the value of adoption of attitudes and behaviors that promote good health (Palloni et al. 2016). According to this explanation, not only actual scarcity might impose a cognitive *bandwidth tax* as Mullainathan and Shafir (2013) note, but expectations and outlooks of economic success could trigger the adoption of unhealthy behaviors through changes in decision making and long-term planning (e.g., time discounting).[11] Both positive and negative interpretations are not necessarily symmetric and generate different predictions. The increase of expected returns of health investments would suggest that high mobility regimes generate benefits for everybody (albeit with decreasing returns, see Grossman 1972), whereas hopelessness implies that the extent of mobility should only produce benefits for individuals at the lower end of the income hierarchy.

These lines of argument borrow from Grossman's model on health capital (Grossman 1972, 2000). In his seminal paper, *On the Concept of Health Capital and the Demand for Health*, Grossman defines health as a capital and investment good that increases the number of healthy days to work and earn income. Thus, how people allocate resources between health capital and other goods/services would depend on the prices of health care, wages and their productivity in the production of health. For instance, increasing wage rates (shifting the marginal efficiency of investment) would rise the returns from healthy days and increase the optimal health stock. The final result of an increase of income, however, might be ambiguous because of the trade-off between higher returns of health (healthy days are more valuable) and a higher opportunity costs associated with the production of health (time spent exercising). The level of education, on the other hand, is assumed to improve the efficiency to produce health investments (better knowledge of harmful effects of smoking, ability to follow medical instructions). Thus, having low expectations about socioeconomic prospects would discourage health investments (lower return of healthy days), or decrease investments in education with subsequent reductions of health production efficiency.

These conditions also imply intergenerational effects. The perception of opportunities in a community may influence how parents socialize children and favor (discourage) the adoption of positive outlooks and the value of skill acquisition that eventually affect health status (Palloni et al. 2016). Thus, early upbringing and socialization, the formation of skills associated with significant returns to extra years of healthy living, and the adoption of attitudes and behaviors that lower exposure to health risks and strengthen resiliency, would impact health throughout the life course. If early cognitive and non-cognitive skills matter for labor market success, income and wages, they may also matter for adult health not only because socioeconomic success breeds good health but because the traits responsible for success also generate adherence to low-risk behaviors.[12]

[10] Income mobility could also be just a proxy for a host of community traits which could impact health.

[11] Time discounting characterizes how individuals' preference for a reward decreases with the delay to its receipt. In general, most people tend to prefer smaller, immediate rewards to larger ones available after a delay. It is also important to differentiate *time discounting* from *time preference*. The latter describes whether and how people consider events in the past, present and future when making decisions, while time discounting captures the degree to which people devalue rewards with every additional unit of delay (Barlow et al. 2016).

[12] This also opens the possibility that health affects income mobility (feedback): the deterioration of health in a community

To formalize these ideas, one can write an expression using the notation proposed by Wagstaff and van Doorslaer (2000):

$$h_i = f_i(y_i, m_i, M_c, I_c) \tag{3}$$

where $m_i$ correspondents to a measure of individual (intergenerational) mobility, $M_c$ a measure of income mobility at the community level, $y_i$ the individual income, and $I_c$ inequality at the community level. As in the case of income inequality one can distinguish processes affecting *directly* individual behavior and delayed effects due to early socialization and intergenerational transmission.

The relationship between $m_i$ and $M_c$ deserves a comment. The hypothesized individual *mobility effects* tend to go in the opposite direction than those expected of contextual effects: while individual mobility could increase instability and stress, contextual mobility would improve expectations and outlooks. It is not hard to imagine that contextual mobility might also increase insecurity and stress, particularly among those at the top of the income distribution. However, those better off are able to protect themselves against the anticipated threat of falling in the economic ladder, and the stress to which they might be exposed can be reduced or offset by the benefits of being in a relatively high status. On the other hand, a contextual effect of social mobility through changes in *perceptions* of the mobility regime, time discounting, and eventually the adoption of behaviors, might be influenced by experiences of individual mobility.

It is also plausible to expect that *intragenerational mobility* has a similar impact on decision making and long-term planning as *intergenerational mobility*. Living in areas with more intergenerational mobility implies a higher chance of experiencing intragenerational mobility. This brings up the question of how persons define their perceptions about the mobility regime and how resistant these perceptions are over time. Hence, when the key factor being investigated is the *perception of the mobility regime*, the distinction between contextual and individual effects, intra and inter mobility, becomes fuzzy. In addition, we should expect lagged effects because it can take time to align individuals' perceptions with actual changes in the stratification regime.

The link between $M_c$ and $I_c$ also deserves a comment. First, the contextual conditions generated by income inequality might be a direct cause of income rigidity. Nepotistic relations, for instance, are stronger in communities with high-income inequality and can inhibit social mobility (Palloni et al. 2016). In this case, social mobility would be a mediator of the link between inequality and heath. Under this assumption, the estimation of the independent contextual effect of mobility requires adjusting for income inequality to avoid confounding. The reverse is also possible: changes in income mobility affect individuals' income, and by extension, the income distribution.

Let us imagine a counterfactual scenario where there are several communities with a given income inequality (measured by the Gini coefficient or any other statistic to characterize the income distribution). Those communities would also have a steady state income mobility. Suppose there is an intervention that improves access quality quality among those with fewer resources. We would expect that, in the long run, those changes would affect communities' income and its distribution. Changes in income would improve health, changes in the income distribution should also improve health (income inequality hypothesis). Perceptions of the mobility regime, in turn, will take some time to update. Once they are in line with actual socioeconomic opportunities both due to individual experiences of mobility or those of significant others (i.e., family, friends, neighbors), decision making and long-term planning will reduce the adoption of un-

might stiffen the stratification regime due to health selection processes discussed before. I will examine this question in Section 5.

healthy behaviors. Contagion and social influence might also multiply the consequences of these processes (Christakis and Fowler 2007). Suppose we compare these communities with others that are identical but where the intervention did not take place. In this case, we can estimate the total *causal* effect of social mobility on health. By adjusting for income, we estimate the part of the effect that is not due to income (functional specification issues aside). In this *ideal* scenario, adjusting additionally for income inequality would imply *controlling* for an intermediate factor (i.e., post-treatment bias). The residual difference would represent the remaining effect of income mobility that was not due to changes in income levels and income inequality, but to *individual and/or contextual* effects of income mobility, attributable, at least in part, to the behavioral mechanisms described above. This simple example allows comparing an ideal setup with the actual models and data I would use.

To finish this section, I outline some conjectures that can be extracted from the previous discussion on income mobility and health:

1. A positive relationship between community income mobility and health/mortality at the individual and aggregate is expected, after adjusting for individual income, community income, income inequality and other individual and community confounders.[13]

2. Income mobility should be related to risk behaviors such as smoking and diet habits at the individual and aggregate level, after adjusting for individual income, community income, income inequality and other individual and community confounders.

3. Benign consequences of the flexibility of a stratification regime on health should be higher among those at the bottom of the income distribution, and lower for those at the top. Those at the bottom have more to gain from a highly mobile environment than those at the top (i.e., expectations of economic success should have a higher impact on the investment in skills necessary to improve socioeconomic status). Conversely, the risk of downward mobility might have potentially harmful effects (e.g., stress), especially for those better off, although those negative consequences can be reduced or offset by the benefits of being in a relatively high status.

4. Income mobility should buffer the negative consequences of income inequality and relative deprivation. Thus, we can expect that differences in income mobility would be more consequential under highly unequal income distributions than in more equitable communities (i.e., income mobility as moderator of income inequality). This would depend, however, on the nature of income inequality and social mobility regimes. In some cases, inequality arises from the growth of the share of income to the very top of the income distribution (as in the case of the U.S.), but in other cases, inequality is due to increases in the population size at the lower income ranks. Social mobility might also occur at different levels of the income distribution.[14] In this context, exploring the variability of the relationship between income mobility and health outcomes across geographical areas becomes relevant.

5. At the aggregate level, we should expect a negative relationship between income social mobility and health disparities. As in the case of inequality and health, there is no theoretical reason to focus only on average health changes by social mobility and disregard changes in health disparities or variance (see, for instance, Neckerman and Torche 2007 and Truesdale and Jencks 2016).

[13] It is also important to avoid post-treatment bias and overfitting.

[14] Unless one is able to formally define and empirically identify *some* of the theoretically relevant configurations, we cannot aspire to formulate precise predictions, models or estimation. This is the motivation behind the idea of formulating a generative theoretical model of the impact of income mobility on health.

6. Given the theoretical mechanisms proposed for the contextual effect of social mobility, one might expect a consistent relationship between income mobility and characteristics such as hopelessness or positivism (expectations and outlooks), after adjusting for family background variables, income, income inequality, and other confounders.

# 3 Aggregate Effects of Mobility on Health and Mortality

## 3.1 Goal and Questions

In this chapter I extend the work done by Palloni et al. (2016) and Venkataramani et al. (2015), examining the association between mortality/health and income mobility at the community level (e.g., counties and commuting zones). Although I do not offer a rigorous test of causality, I provide a deep description of the association between these two phenomena. This is a necessary step to understand the link between health and income mobility.

The key questions I examine in this chapter are:

1. What is the magnitude of the association between income mobility (relative and absolute) and mortality/health in comparison with income and income inequality?
2. How does the association vary across commuting zones and states? What aggregate level characteristics do account for that variability? Are there differences in the patterns observed within states and between states?
3. How robust is the association between income mobility relative to income inequality at the community level?
    (a) How do effects differ by health outcomes: mortality and risk behaviors such as smoking and physical activity?
    (b) How do effects vary by gender, education, income, age, race/ethnicity, and cause of death?
    (c) How does income mobility impact health disparities (differences and ratios of health outcomes by income, education, and race/ethnicity)?
    (d) How do these effects vary when using different measures of income mobility (relative and absolute) and adjusting for different measures of inequality (Gini coefficient, fraction of middle class, top 1% income share)?

## 3.2 Model

In addition to an extensive descriptive and graphical inspection of the association between income mobility, inequality, and health, I will use *multilevel models* to estimate income mobility effects. I make this decision for two reasons. First, hierarchical models enable me to easily accommodate the nested structure of the data (e.g., states, commuting zones, counties), and explore where most of the action is: at the state level, commuting zones or counties. Secondly, there is no reason to think that the health's consequences of the mobility regime are constant across states and commuting zones, and making that assumption is overly restrictive and, in some cases, misleading. Exploring variability can offer hints about omitted variables that explain why some units respond more or less to changes in income mobility than others. Multilevel modeling provides great flexibility to incorporate those features.

Using death counts and exposure (person-years) by county, age, gender, race/ethnicity, and cause of

death, I can predict $\lambda$, the expected number of deaths $\mu$ per unit of time $\tau$ with a Poisson model.[15] An initial model would be:

$$
\begin{aligned}
D_i &\sim \text{Poisson}(\mu_i) \\
\log(\mu_i) &= \log(\tau_i) + \alpha + \alpha_{\text{state[i]}} + \beta \text{Mobility}_i \\
\alpha_{\text{state}} &\sim \text{Normal}(0, \sigma_{\text{state}}) \\
\sigma_{\text{state}} &\sim \text{HalfCauchy}(0, 1)
\end{aligned}
$$

where $D_i$ is the number of deaths by county, $\text{Mobility}_i$ is the mobility measure for county $i$, and $\log(\tau_i)$ is the logarithm of the exposure for county $i$ (log of mid-population). Computing the exponential of $(\alpha + \alpha_{\text{state[j]}} + \beta \text{Mobility}_i)$ provides the estimate of the mortality rate per county because $\log(\lambda_i) = \log(\mu_i) - \log(\tau_i)$. This is just an (underfitted) varying intercept model. Additional types of varying intercepts (e.g., age, gender, race/ethnicity, cause of death), predictors (e.g., income at the county level, income inequality, other confounders), and over-dispersion adjustments would be included.

For instance, if deaths and population counts are tabulated by age group (0, 1-4, 5-9, 10-14, ..., 75-79, and $\geq$80 years), gender, race/ethnicity, and county, a model using only income mobility as contextual predictor might be specified like this:[16]

$$
\begin{aligned}
D_i &\sim \text{Poisson}(\mu_i) \\
\log(\mu_i) &= \log(\tau_i) + \alpha_i + \beta \text{mobility}_i + \varepsilon_i \\
\alpha &= \alpha + \alpha_{\text{age[i]}} + \alpha_{\text{race[i]}} + \alpha_{\text{county[i]}} + \alpha_{\text{state[i]}} \\
\beta \text{mobility}_i &= \beta + \beta_{\text{age[i]}} + \beta_{\text{race[i]}} + \beta_{\text{county[i]}} + \beta_{\text{state[i]}} \\
\alpha &\sim \text{Normal}(0, 10) \\
\beta &\sim \text{Normal}(0, 1) \\
\varepsilon_i &\sim \text{Normal}(0, \sigma_\varepsilon) \\
\sigma_\varepsilon, \sigma_\alpha, \sigma_\beta &\sim \text{HalfCauchy}(0, 2)
\end{aligned}
$$

Further extensions of these models might be attempted provided they increase prediction power, goodness-of-fit, and do not show signs of overfitting.[17] The same approach would be used when exploring causes of death, and other outcomes such as smoking, obesity, and physical activity. The analysis of aggregate data (states, commuting zones, and counties) is a first step to understand the association between income mobility and health/mortality. The scope of this chapter is mainly descriptive and represents an effort of examining in depth the patterns and magnitude of that association.

[15] See Dwyer-Lindgren et al. (2016) for a recent implementation of similar models to estimate trends in mortality rates for major causes of death in the U.S.

[16] The model adjusts for overdispersion ($\varepsilon_i$), and omits multivariate priors (to avoid cluttering).

[17] I will use a Bayesian approach that avoids overfitting by using priors and (adaptive) regularization. For a discussion on how multilevel modeling helps with multiple comparisons see Gelman and Hill 2007.

### 3.3 Data and Limitations

I will use the following data sources:

1. Income mobility measures, ethnicity-adjusted life expectancy at 40, health behaviors, and a set of county-level covariates come from the *Health Inequality Project* dataset (Chetty et al. 2016).
2. Death records from the National Vital Statistics System provided by the National Center for Health Statistics (NCHS).[18] Measures of health behavior and conditions (e.g., smoking, exercise, obesity) come from the CDC county indicators, while county-level population counts by age, sex, and race/ ethnicity from 2000 to 2014 is provided by the NCHS.
3. Additional covariates might be added at the county or commuting zone level from the American Community Survey (ACS) (Firebaugh and Farrell 2016; ICPSR 2008).

Given its importance for this dissertation, I briefly discuss how income mobility measures and life expectancy at age 40 were constructed by the *Health Inequality Project*.[19] This is useful to provide a better idea of the strengths and weaknesses of the data and analyses I can do.

The *Health Inequality Project* uses data from federal income tax records from 1996 to 2012[20] to create a dataset of children born between 1980 and 1991 who were U.S. citizens by 2013 and have a valid Social Security or Individual Taxpayer Identification number. For each child, parents were defined as the first person(s) who claims the child as a dependent on a 1040 tax form. If parents were married but filled tax forms separately, the child was assigned to both parents provided the mother (or father in case of single fathers) was aged 15-40 at the birth of the child. If no such eligible match occurs in 1996, subsequent years (through 2012) were used until a valid match was found (matching was successful for 95% of the children in the 1980 and 1991 birth cohorts). Parent income was measured at the household level.[21]

The publicly available data correspond to the core sample using baseline income definitions. The core sample includes nearly 10 million children born between 1980 and 1982, for whom parents were identified, and whose mean parent income between 1996-2000 was positive (only 1.2% children were excluded). To measure intergenerational mobility, a baseline income definition was used, including labor earning and capital income as well as unemployment insurance, social security, and disability benefits. Income was always measured before the deduction of individual income taxes and employee-level payroll taxes. Family income reflects inflation-adjusted (2012) averages over 2010 to 2012 tax returns for children (mean age of 30 years), and averages over 1996 to 2000 tax returns for parents (mean age of 43 years). Children were then assigned to ZIP codes of residence based on their parents' ZIP code included in the form 1040 which served to match parents and children. Those ZIP codes were mapped to counties based on the 1999 Census crosswalk and then aggregated into counties and *commuting zones*.[22] The minimum number of children allowed by county and commuting zone to compute mobility measures was 250.

[18] Mortality Multiple Cause-of-Death Restricted Record 2000-2014.

[19] A complete description of these data is available in Chetty et al. (2014b)

[20] Including both tax returns (1040 forms) and third-party information returns (e.g., W-2 forms).

[21] It is known that there is a lower correlation between women's individual income and their parents' income than the comparable correlation for men. However, the correlation between women's family income (that is, women and their male partners) and their parents' income should be roughly similar to the one for men. This is particularly important when comparing effects by gender and can be directly checked because gender-specific income mobility measures are available at the commuting zone level.

[22] Commuting zones are aggregations of counties, similar to metro areas but including rural areas and covering the entire U.S. For more details see Tolbert and Sizer 1996.

The *Health Inequality Project* dataset includes two main measures of intergenerational mobility: relative and absolute. Relative mobility or the rank-rank slope corresponds to the correlation between a child's income rank in her birth cohort income distribution $R_i$ and parents' income rank in parents' income distribution $P_i$:[23]

$$\rho_{PR} = Corr(P_i, R_i) \tag{4}$$

Rank-rank slopes have proved to be quite robust across specifications and highly suitable for comparisons across areas (Chetty et al. 2014b). *Canonical* measures of relative mobility, such as intergenerational income elasticity,[24] tend to be more sensitive to changes in inequality across generations. The second indicator used in the data set is the absolute upward mobility score or "the mean rank (in the national income distribution) of children whose parents are at the 25$^{th}$ percentile of the national parent income distribution" (Chetty et al. 2014b, p. 7). Although at the national level both the relative and absolute measure of mobility provide similar information, when studying small areas a child's rank in the national income distribution would be an absolute outcome because income in a given areas have little impact on the national distribution.[25]

Figure 1 shows income mobility and inequality measures by the logarithm with base 10 of counties' population. Extreme values of income mobility are observed mostly in small counties. This pattern suggests that part of the variability observed in the mobility measures is just due to differences in sample sizes (the number of children born between 1980-1982 is strongly correlated with the population of counties, $\rho$ = 0.988). That is not the case of income inequality where extreme values are observed all over the population (probably, because those estimates use all the cohorts at a given period).[26]

The *Health Inequality Project* also provides estimates of race and ethnicity adjusted period life expectancy at 40 years of age and health behaviors by income quartile. Mortality was measured using the Social Security Administration death records (that match closely the data from the NCHS, Chetty et al. 2016). Life expectancy was constructed by (1) estimating age-specific mortality rates from ages 40 to 76, (2) extrapolating mortality rates beyond the age of 76 (Gompertz function) and calculating period life expectancy, and (3) adjusting for differences in the proportion of ethnic and racial across income quartile for counties, commuting zones, and states (for that the data from 2001 to 2014 were pooled). Thus, life expectancy estimates summarize the mortality experience of those who were between 40 to 63 year old over the period 2001-2014, that is, cohorts born between 1923 and 1959. Finally, indicators of health behaviors and conditions by income quartile come from the Behavioral Risk Factor Surveillance Survey 1996-2008, and include rates of smoking, obesity (BMI $\geq$ 30), and self-reported exercise during the past month.

Estimating the relationship between income mobility, mortality and health behavior using the *Health Inequality Project* data must face both a mismatch of the cohorts being studied and missing data (about 50% for counties and 20% for commuting zones). As discussed by Palloni et al. (2016), all the measures of mobility characterize cohorts that are younger (1980-1982) than the cohorts used to estimate mortality experiences (around 1923-1959) and health behaviors. This will be partially solved in this dissertation by

[23] In this case, the regression coefficients equals the correlation because both child and parents ranks follow a uniform distribution by construction.

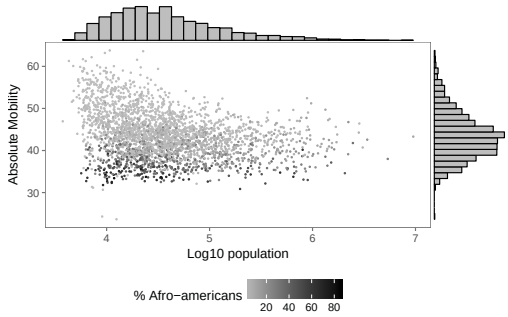
[24] The elasticity of child income with respect to parents' income.

[25] There are different versions of mobility measures. One of them uses *permanent residents*, that is, parents who stay in the same commuting zones and counties between 1996-2012. Note that children who grow up in a commuting zone or county may have moved out as adults.

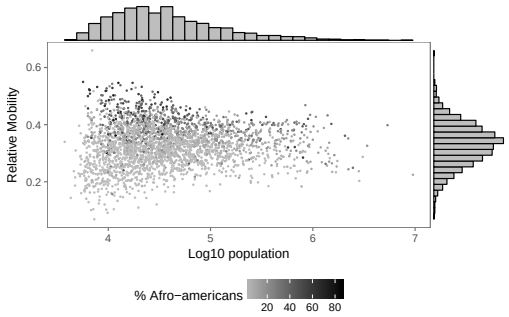
[26] This potential variability in mobility measures must be accounted for in the models we estimate.

Figure 1: Measures of Income Mobility and Inequality by County Population (N = 2768)

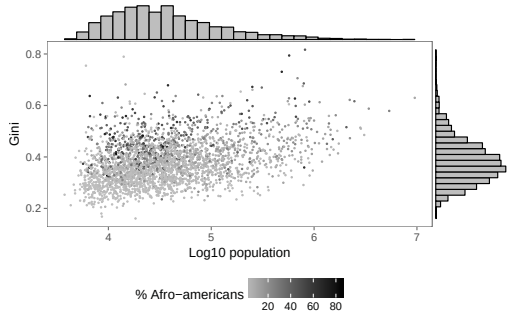
(a) Absolute Mobility




(b) Relative Mobility




(c) Gini



using mortality data from cohorts closer to those used by Chetty et al. (2014a). The *Health Inequality Project* data, however, has the advantage of providing life expectancy *by income quartile*, making possible to the estimation of mobility effects by income. An alternative approach is to use estimates of educational attainment as a proxy for socioeconomic status (see, for instance, Bound et al. 2015). Another possibility is to assume that income mobility in small areas is not a transitory feature but rather reflects entrenched local conditions that make the mobility experience of one generation an enduring feature shared by subsequent generations. According to Palloni et al. (2016), there are two pieces of evidence to support that assumption. First, there is evidence that national and local U.S. mobility trends have been steady for longer than a generation. This suggests that there must be a more than a modest correlation between the income mobility regimes that parents and offspring experience during important life cycle stages. Second, analysis of the mobility data by commuting zones and counties reveals the existence of high correlations between income mobility regimes and indicators of small area characteristics that are not transient: there are strong correlations with race composition, levels of segregation, income levels and inequalities, school quality, family structure and, finally, a number of indicators of social capital (Chetty et al. 2016). These empirical results suggest permanence, not fleetingness, of mobility regimes. Despite this evidence, we should be interpreted results with caution.[27]

Finally, it is not advisable to pre-average data (e.g., proportions or rates at the county level) when estimating effects. The reason is that averaging removes variation (Gelman et al. 1999; McElreath 2016). Thus, instead of modeling age-specific mortality rates and life expectancy directly, modeling the number of deaths after adjusting for exposure would preserve variability due to the size of counties. To assess the implications of a model, one can compute summary statistics (e.g., life expectancy) and examine different scenarios. Unfortunately, the raw mortality data used by the *Health Inequality Project* to compute life expectancy by income is not publicly available, so more detailed analyses would be only possible using the NCHS mortality data. There are also theoretical reasons to avoid focusing on aggregate indexes such as life expectancy. For instance, given the mechanisms discussed above, causes of death related to behavioral factors should be more sensitive to changes in the mobility regime given. Similarly, the consequences of hopelessness and socialization might be more relevant than other causes during middle ages and later in life.

The strategy and analyses proposed in this chapter are subject to the following shortcomings:

1. The *Health Inequality Project* data set is limited. It provides mobility information of a relatively small cohort, so it is not clear how the results of this chapter would hold when using information from other generations. One of the advantages of this dataset, however, is it was created using administrative information, reducing measurement errors. In addition, the data set only considers one dimension of social mobility, namely, income. Although variables such occupational status, class, earnings, and income mobility seem to be part of the same latent concept (socioeconomic standing and well-being), empirical evidence shows different mobility patterns when using class/occupational status and income/earnings (Beller and Hout 2006; Torche 2015). These variables appear to capture different dimensions of socioeconomic disadvantage.[28] Furthermore, from a theoretical perspective there are substantive reasons why focusing exclusively on income mobility might be limiting. Although economic capital (e.g., income, occupation) is central to understand stratification, it has been

[27] This leads us to another aspect of the consequences of income mobility on health: the timing of the effect. It is reasonable to expect lags, especially if these effects involve socialization processes and changes in perceptions. In the case of socialization, what seems relevant is the income mobility of parents during the childhood of theirs offspring. Perceptions on the mobility regime, in addition to intragenerational mobility, instead, would probably depend on previous cohort experiences.

[28] Differences should be expected when the distributions of these variables are not perfectly correlated and deviations across distributions are strongly correlated across generations (Björklund et al. 1999).

suggested that socioeconomic factors are not enough to define class (Savage 2015). According to Bourdieu (2011), for instance, *cultural capital* is a central dimension of the process of accumulation and stratification in modern societies. Thus, cultural tastes and preferences (considered as superior or more legitimate than others) could help when interact with institutions to obtain better services (Lareau 2002). These tastes and preferences have been also related to the adoption of healthy behaviors (Phelan et al. 2010). Thus, when interpreting results, we should be cautious since the *effects* observed may not necessarily be the same as those associated with other measures of socioeconomic and class mobility.

2. The analyses are mainly correlational and cross-sectional, and no strong causal inferences are possible. For instance, mortality statistics do not necessarily measure lifespans of people from places, but life spans of people who die in places. If there is migration and selectivity among those who migrate, the effects of income mobility or inequality may be contaminated by the influence of characteristics that distinguish migrants from non-migrants.

3. There is a risk to attribute individual effects to income mobility when the causal mechanisms at the aggregate level might be different from the ones observed at the individual level (i.e., ecological fallacy).

4. There exists the possibility of reverse causality, especially when there is a mismatch between cohorts.

5. It is not evident that the more salient effects of income mobility on health occur at the county or commuting zone level. It is possible that the presence of more local levels such as census tracks or neighborhoods provide a better level of aggregation to infer some of the behavioral mechanisms proposed by the theory.

# 4 Individual Effects of Income Mobility on Health Behaviors, Health Status and Mortality

## 4.1 Goal and Questions

The goal of this chapter is to empirically assess the plausibility of the contextual effect of income mobility on health behaviors and mortality at the individual level, and to examine some of the mechanisms proposed by Palloni et al. (2016) and Venkataramani et al. (2015, 2016). Thus, I extend the analyses from my previous chapter by using individual data and exploring mechanisms.

The key questions I examine in this chapter are:

1. What is the direction and size of the association of income mobility at the community level (i.e., county and commuting zone) with individual mortality, subjective and mental health, after adjusting for income, income inequality and other confounders at the individual and community level? Are the effects of income mobility sizeable and in the expected direction? How big/small is the magnitude of the mobility effect relative to individual income and income inequality?
2. What is the relationship between income mobility at the community level and individual health behaviors and conditions such as smoking, physical activity, obesity, after adjusting for income, income inequality and other confounders at the individual and community level?
3. How do these effects vary at the state and commuting zone level? How do these effects vary by income, income inequality, age, race, and gender?
4. Is there any link between income mobility and some of the mediators proposed by the theory, namely, cognitive and non-cognitive skills, outlooks and expectations?
5. After adjusting for early upbringing (type of family, parental background), early cognitive and non-cognitive traits, health behaviors and status, what is the magnitude and direction of the remaining income mobility effect on current risky behaviors, subjective health, and mortality?

## 4.2 Model

Again, I propose to use *hierarchical models* to estimate effects of income mobility at the individual level and explore the variability of the estimates across states and commuting zones. The basic model is:

$$y_i \sim (\beta_0 + b_{1,s[i]}) + (\beta_1 \text{ Mobility}_c + \beta_2 \text{ Inequality}_c + \beta_3 \text{ Income}_i + \beta_n X_i + \beta_m X_c) + \varepsilon_i \quad (5)$$

$$b_1 \sim N(0, \sigma_1^2) \quad (6)$$

$$\varepsilon_i \sim N(0, \sigma_r^2) \quad (7)$$

Where $y_i$ represent an individual outcome (e.g., mortality, health or behavior), $X_i$ and $X_c$ are a set confounders at the individual and community level. An extension of this model explores varying *mobility* effects by state/commuting zones:

$$y_i \sim (\beta_0 + b_{1,s[i]}) + (\beta_1 + b_{2,s[i]}) * \text{ Mobility}_c + \beta_2 \text{ Inequality}_c + \beta_3 \text{ Income}_i + \beta_n X_i + \beta_m X_c + \varepsilon_i \quad (8)$$

$$\{b_1, b_2\} \sim \text{MVN}(0, \Sigma_1) \quad (9)$$

$$\varepsilon_i \sim N(0, \sigma_r^2) \quad (10)$$

When modeling mortality, I will use Cox or time-discrete survival models with random effects.

### 4.3 Data and Limitations

In this chapter I will use the following data sources:

1. National Longitudinal Survey of Youth, 1997 (NLSY97).[29] The NLSY97 is a nationally representative sample of about 9,000 American youth born between 1980 and 1984, the same cohort of the children included in the *Health Inequality Project*. Surveys were conducted annually, beginning in 1997 when the youth were between 12 and 18 years of age. In the first round, both the eligible youth and one of that youth's parents were administrated personal interviews. The restricted NLSY97 geocode datafile contains information on the geographic residence of each respondent over time, and allows me to merge them to county income mobility measures. Table 1 displays some of the key variables available in the NLSY97 dataset.

2. Panel Study of Income Dynamics (PSID) is a nationally representative sample of U.S. men, women, children, and their families that has been followed for more than 40 years. The PSID began interviewing a sample of approximately 5,000 families in 1968 and were re-interviewed each year through 1997, when interviewing became biennial. Restricted geographic data allow merging individual records with county income mobility measures. Unlike the NLSY97 data, the PSID includes multiple cohorts. This involves mismatches between the cohorts used to estimate mobility measures in the *Health Inequality Project* and the ones being interviewed in the PSID. One can restrict the analytical sample to cohorts closer to the ones used by Chetty et al. (2014a), provided the sample sizes do not become too small. The PSID 1980-1984 cohort sample, for instance, consists of 2975 respondents, the cohort 1980-1986 of 3800 respondents.[30] Furthermore, PSID has data on the parents of these cohorts and mortality data from the National Death Index (NDI). Between 1968 and 2013, 6456 respondents have died (8.6% of the total sample). In addition, both intergenerational and intragenerational mobility can be estimated at the individual level, so that we can explore the link between individual and community income mobility. Table 2 shows some of the variables available.

Both datasets have strengths and weaknesses. The NLSY sample matches some of the cohorts used by Chetty et al. (2014a) (1980-1984) to compute county and commuting zones mobility measures. However, because those cohorts are relatively young, it is difficult to estimate the effect of income mobility on health/mortality. These data are more useful to examine the relationship between income mobility and

[29] Some analyses might also use the National Longitudinal Survey of Youth, 1979 (NLSY79) that originally sampled 12,686 respondents born between 1957-64 (ages 14-22). For the purposes of this proposal I only focus on NLSY97.

[30] These are PSID *gene* respondents. All 1968 sample members have the PSID *gene*, and they are followed in all subsequent waves across their entire lives, regardless of where they live. All individuals born to or adopted by somebody with the PSID *gene* acquires the gene themselves, and therefore are followed.

Table 1: Key Variables, NLSY97

| Baseline 1997 | 2010-2013 |
|---|---|
| *Covariates* | *Outcomes* |
| Family Background | Subjective Health Status |
| Achenbach's Youth Self Report | Smoking |
| Expectation Index | Substance Use |
| Optimism Scale | Exercise |
| Big Five Items (2002) | Obesity |
| | Mental Health (Depression) |
| Subjective Health Status | Mortality (Non-response record) |
| Smoking | |
| Substance use | *Current Covariates* |
| Obesity | Income |
| | Education |

Table 2: Key Variables, PSID

| |
|---|
| Mortality (NDI) |
| Subjective Health Status |
| |
| Smoking |
| Exercise |
| Obesity |
| Life Satisfaction |
| |
| Income |
| Education |
| |
| Achievement motivation |
| Word-to-picture score (cognitive) |
| Aspiration / Ambition |
| Risk Avoidance |

risky behaviors such smoking, alcohol use, and cognitive and non-cognitive skills. The PSID, although has smaller samples of the cohorts that match the ones used by the *Health Inequality Project*, also has good mortality information (National Death Index) and health status measures. The use of both data sources may offer a broader picture to examine the research questions proposed in this chapter. Both datasets also track the county of residence over a relatively long period of time, being possible to estimate models that take into account the temporal impact of county exposure on health and migration selection (Wodtke et al. 2011; Hernán and Robins 2006).

When using NSLY97, I will focus on behaviors such as smoking, substance use, exercise as well as on individual expectation indexes, optimism scores, psychological assessment (big five), and health outcomes such obesity, mental health, and depression. In the case of the PSID, I will focus on mortality, subjective health status, smoking, exercise, obesity, and life satisfaction.

This chapter also has limitations. First, models proposed will not offer reliable causal inferences on the effects of income mobility on health. However, they provide a solid base for exploring the robustness of the association at the individual level and some potential mechanisms discussed in the literature.[31] Second, depending on the cohorts used there exists the risk of reverse causality (health affecting income mobility, instead of mobility determining health). Because I am using cross-sectional measures of income mobility, I am not sure of the order of the events and I cannot estimate how changes in mobility affect health status. However, using longitudinal datasets that cover a relatively long period of observation allows for adjustment of non-random migration across counties.

[31] An alternative strategy would be to find exogenous variation from policies that raise or restrict income mobility. Unfortunately, that is not enough since policies *should not* affect population's health directly but only through income mobility. Even when those assumptions are met, we would only be able to estimate the total effect of *income mobility*. Additional assumptions are needed to estimate the causal effect of mediators (see Pearl 2014).

# 5 A Generative Model of Health and Socioeconomic Mobility

The relationship between inequality, income mobility, and health is not simple. As I discussed in the introduction, it involves individual, contextual and spatial effects, feedback loops, and cumulative processes. These processes are usually difficult to grasp and understand using just statistical models because they do not take into account the dynamic, reciprocal, discontinuous or changing relations between exposures and outcomes. For instance, individual socioeconomic position contributes to the type of neighborhood a person can afford to live in and his level of substance use. At the same time, individual socioeconomic position is a product of the types of income-generating opportunities afforded by the neighborhood socioeconomic achievement. Early exposure to poor socioeconomic environments may also impact health in later life.

Simulation models have the *potential* to be useful tools to understand and learn from these systems, by offering simplified representations of the mechanisms that generate and maintain health inequalities. Following that tradition, I propose to use a computer simulation model to specify the mechanisms discussed in the literature and assess the consequences of the socioeconomic mobility regime for health. The goal is to examine behaviors and consequences of the mechanisms proposed, and explore *what-if* questions and scenarios.

The key questions I examine in this chapter are:

1. What are the consequences of feedback mechanisms whereby the distribution of individuals' by health status influences income mobility on health disparities over several generations?[32]
2. How would health behave under different income inequality and mobility scenarios (e.g., inequality observed mainly between the top 1% and the rest of population, income mobility at different positions of the income distribution)?
3. How the system behaves when allowing selective migration (i.e., segregation) of agents across neighborhoods/counties?

## 5.1 Model Construction

I will design a simulation model that replicates as closely as possible the dynamic of a stratification system where agents are allocated to different positions according to a set of well-defined human capital traits. In principle, this would be an abstract or *low-dimensional realism* model where micro-level behaviors are assumed or known, and simulation is used to explore its aggregate consequences. This strategy is what Wilensky and Rand (2015) refer to as *exploratory modeling*. Thus, the scope of the model is mainly theoretical, and represents an attempt to verify the logic of hypotheses and refine predictions.

The model construction will proceed as follows. First, I will review/replicate previous simulation models addressing health disparities (for instance, Wolfson et al. 2017; Speybroeck et al. 2013; Zhang et al. 2014) to familiarize myself with alternative implementation strategies and identify their strengths and weaknesses. I have already replicated the Theoretical Health Inequality Model (THIM) proposed by Wolfson

[32] It is well-known that an important determinant of individuals' income/status attainment depends on labor market performance (Palloni et al. 2009). To the extent to which the latter is influenced by individual health, there is the potential of feedback effects whereby a regime of income mobility may be maintained steady, made more flexibly or rigid as a function of the distribution of children and adults by health status.

et al. (2017) (see Figure 2). That model is motivated by the strong differences across cities and countries in health/mortality and income inequality. The mechanisms specified in that model are inspired by the *neo-material* theory by Lynch et al. (2004), and the idea that an aggregate relation between income inequality and health is not necessary but contingent. The model also includes relative-income mechanisms, that is, the notion that individual's relative - rather than absolute - income affects health. However, it excludes any (direct) effect of income inequality on health and mortality. All the effects work through income. The mechanisms specified in this model include segregation, intergenerational transmission of social advantage/disadvantage (based on relative income), the effect of relative neighborhood's income on individual's education and income, and the effect of relative individual income on health and mortality (see Figure 3). Although Wolfson et al. (2017)'s model reproduces some of the pathways between inequality and health at the aggregate level, it has no explicit feedback between income and health, and intergenerational transmission only operates through the direct effect of parent's income and education on child's education and income (there are no contextual effects). I could extend the model by explicitly incorporating adoption of a risk behaviors (e.g., smoking), including feedbacks and well-defined links between income mobility and time preferences.

Second, I will define the general components of the model and assess them separately to understand how they work. Below I outline some elements that could be incorporated.

1. Education and income attainment
    (a) Income over life course (age-profile)
    (b) Health influence on income/education
    (c) Individual intergenerational transmission of advantage/disadvantage
    (d) Contextual effect of income and education (relative/absolute)
2. Preferences to adopt risky behaviors (e.g., smoking, National Academy of Sciences 2015)
    (a) Influence of early exposure to social mobility context
    (b) Effect of previous behavior
    (c) Effect of income and education (individual and contextual)
    (d) Consequences of social influence (Christakis and Fowler 2007)
3. Health over life course
    (a) Mortality
    (b) Adoption of risky behaviors over life
    (c) Influence of income and education (individual and contextual)
4. Selective migration (e.g., segregation)
    (a) Segregation based on income / education
5. Fertility (stationary population)
6. Assortative mating

Finally, the general principle in this kind of models is to simplify as much as possible. The important point is to understand the components of the model and learn which processes are the most important while designing and extending the model: *start simple and build toward the question you want to answer* (Wilensky and Rand 2015). By starting small and adding elements one by one, I can avoid the inclusion of useless and non-relevant components. A simple model is also more understandable and easier to verify (to ensure that the computational model faithfully implements its target conceptual model). The development of my model will follow the ODD protocol, "Overview, Design Concepts, and Details" (Railsback and Grimm 2011; Müller et al. 2013). This protocol was designed to create complete model descriptions, and to present information in a consistent way. The final goal of the protocol is to facilitate design and replication. The model will be build using the software Anylogic (Borshchev 2013).

## 5.2 Analysis plan

The analysis of simulation models is an iterative process where the design of the model and exploration of the results go through several small loops. I can identify in advance, however, some procedures I plan to implement to explore model's results and assess sensitivity to changes in parameters or initial conditions (Thiele and Grimm 2015).

First, I will use simple descriptive statistics, time series, and multivariate methods. Because being able to reproduce experiments is critical, I will document all the parameter values, initial conditions, analyses and results. Some of the *heuristics* I will use to explore the models are: use extreme values of parameters, search striking or strange patterns in the model output, vary some parameters while keeping the controlling parameter constant, explore several outcomes for evaluating simulation experiments (mortality rates, life expectancy, disparities), examine unrealistic patterns. I will also use standard measures of statistical distributions of results to assess the stochastic components of the model. The estimates obtained as part of the empirical chapters could be used as parameters of the simulation.

Second, I will explore the characteristics of the time series (trends, autocorrelation, time until the system become static, stability properties). In addition to these general analyses, I plan to run sensitivity and robustness analyses. Sensitivity analysis is used to explore the influence of varying inputs on the outputs of a simulation model (Saltelli et al. 2004; Thiele et al. 2014). This helps to identify parameters that have a strong influence on outputs and indicates which processes are most important. Sensitivity analysis also helps to assess the importance of the uncertainty of inputs (e.g., agents' heterogeneity).[33]

Finally and following Thiele et al. (2014), I will start with a simple procedure to explore sensitivity (e.g., graphical methods) to obtain a rough idea of the relationships of inputs and outputs and their linearity or non-linearity. I will use then more informative but computationally demanding methods to identify the most important input factors. For instance, one can use Morris's elementary effects screening to identify the most important factors and then apply the Sobol's method (variance decomposition) or the extended Fourier amplitude sensitivity test (Pujol et al. 2015). I will also use different sampling methods (e.g., Latin hypercube sampling) and meta-models (e.g., Kriging Models). Finally, robustness analysis will focus on parts of the model that seem more likely to be criticized as unrealistic, too simple or too complex.

[33] Sensitivity analysis is closely related to uncertainty analysis: if the model output does not vary strongly when the parameter values change, the uncertainties are of small importance.

Figure 2: Snapshot of Own Replication of the THIM Model using Anylogic

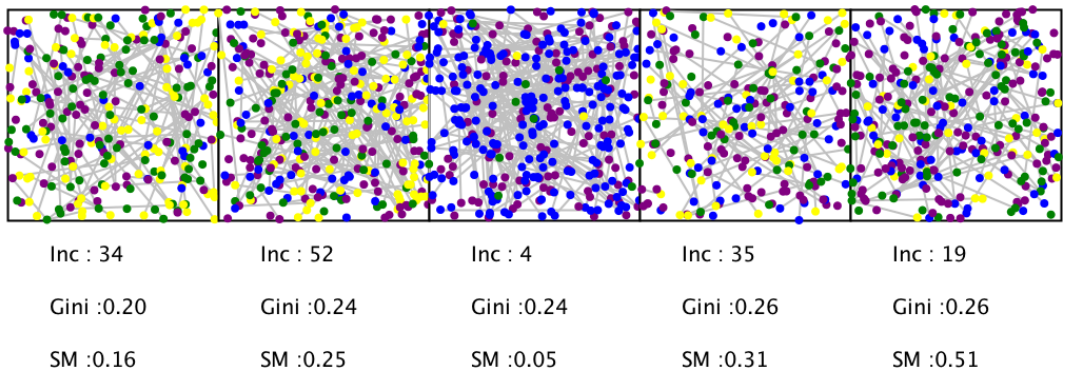
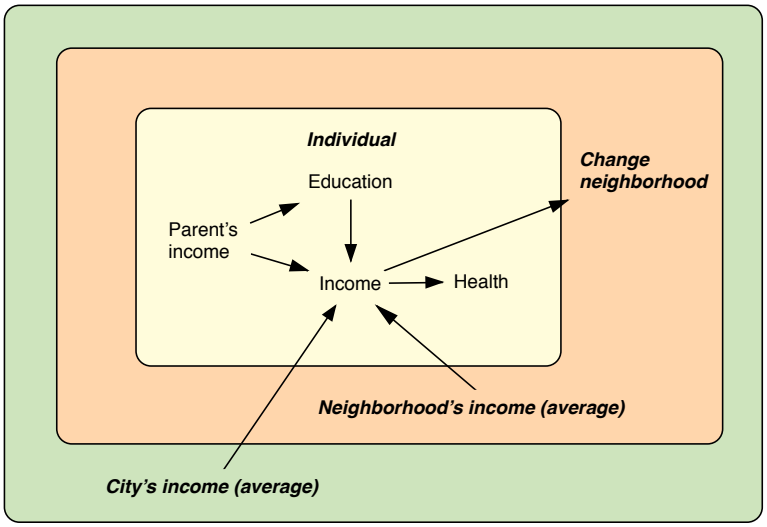



Figure 3: Representation of the THIM Model



# 6 Timeline

| Activity / Milestone | Description | Status | Feb | Mar | Apr | May | Jun | Jul | Aug | Sep |
|---|---|---|---|---|---|---|---|---|---|---|
| Proposal Defense | | In progress | 1 | | | | | | | |
| Draft Chapter 1 | Theory | In progress | | | 1 | | | | | |
| Editing and Corrections Chapter 1 | Theory | | | | 3 | | | | | |
| Draft Chapter 2 | Aggregate results | In progress | | | 1 | | | | | |
| Editing and Corrections Chapter 2 | Aggregate results | | | | 3 | | | | | |
| Draft Chapter 3 | Individual results | | | | | | 2 | | | |
| Editing and Corrections Chapter 3 | Individual results | | | | | | 4 | | | |
| Draft Chapter 4 | Theoretical model | In progress | | | | | | | 3 | |
| Editing and Corrections Chapter 4 | Theoretical model | | | | | | | | 4 | |
| Writing Introduction and Conclusion | | | | | | | | | | 1 |
| Final Defense | | | | | | | | | | 4 |

Note: Numbers represent the week of the month.

## References


Barlow, P., McKee, M., Reeves, A., Galea, G., and Stuckler, D. (2016). Time-discounting and tobacco smoking: A systematic review and network analysis. *International Journal of Epidemiology*, pages 1–10.

Beller, E. and Hout, M. (2006). Intergenerational social mobility: The United States in comparative perspective. *The Future of Children*, pages 19–36.

Billari, F. C. and Prskawetz, A., editors (2003). *Agent-Based Computational Demography: Using Simulation to Improve Our Understanding of Demographic Behaviour*. Physica, Heidelberg; New York, 2003 edition edition.

Björklund, A., Jäntti, M., and others (1999). *Intergenerational Mobility of Socio-Economic Status in Comparative Perspective*. Social Policy Research Centre, University of New South Wales.

Blau, P. M. (1956). Social mobility and interpersonal relations. *American sociological review*, 21(3):290–295.

Borshchev, A. (2013). *The Big Book of Simulation Modeling: Multimethod Modeling with AnyLogic 6*. AnyLogic North America.

Bound, J., Geronimus, A. T., Rodriguez, J. M., and Waidmann, T. A. (2015). Measuring Recent Apparent Declines In Longevity: The Role Of Increasing Educational Attainment. *Health Affairs*, 34(12):2167–2173.

Bourdieu, P. (2011). The forms of capital. In Szeman, I. and Kaposy, T., editors, *Cultural Theory: An Anthology*, pages 81–93.

Chen, Z. and Gotway Crawford, C. A. (2012). The role of geographic scale in testing the income inequality hypothesis as an explanation of health disparities. *Social Science & Medicine*, 75(6):1022–1031.

Chetty, R., Hendren, N., Kline, P., and Saez, E. (2014a). Where is the land of Opportunity? The Geography of Intergenerational Mobility in the United States. *The Quarterly Journal of Economics*, page qju022.

Chetty, R., Hendren, N., Kline, P., and Saez, E. (2014b). Where is the Land of Opportunity? The Geography of Intergenerational Mobility in the United States. Working Paper 19843, National Bureau of Economic Research.

Chetty, R., Stepner, M., Abraham, S., Lin, S., Scuderi, B., Turner, N., Bergeron, A., and Cutler, D. (2016). The Association Between Income and Life Expectancy in the United States, 2001-2014. *JAMA*, 315(16):1750.

Christakis, N. and Fowler, J. H. (2007). The spread of obesity in a large social network over 32 years. *The New England journal of medicine*, 357(4):370–9.

Commission on Social Determinants of Health (2008). Closing the gap in a generation: Health equity through action on the social determinants of health: Final report of the commission on social determinants of health.

Dwyer-Lindgren, L., Bertozzi-Villa, A., Stubbs, R. W., Morozoff, C., Kutz, M. J., Huynh, C., Barber, R. M., Shackelford, K. A., Mackenbach, J. P., van Lenthe, F. J., Flaxman, A. D., Naghavi, M., Mokdad, A. H., and Murray, C. J. L. (2016). US County-Level Trends in Mortality Rates for Major Causes of Death, 1980-2014. *JAMA*, 316(22):2385–2401.

Epstein, J. M. (2006). *Generative Social Science: Studies in Agent-Based Computational Modeling*. Princeton University Press, Princeton.

Evans, W. N., Hout, M., and Mayer, S. E. (2004). Assessing the effect of economic inequality. *Social inequality*, pages 933–68.

Firebaugh, G. and Farrell, C. (2016). Racial Neighborhood Inequality in the United States, 1980-2010. Technical report, Inter-university Consortium for Political and Social Research (ICPSR).

Gelman, A. and Hill, J. (2007). *Data Analysis Using Regression and Multivel/Hierarchical Models*. Cambridge University Press, Cambridge, MA.

Gelman, A., Price, P. N., and others (1999). All maps of parameter estimates are misleading. *Statistics in medicine*, 18(23):3221–3234.

Gilens, M. (2012). *Affluence and Influence: Economic Inequality and Political Power in America*. Princeton University Press.

Grossman, M. (1972). On the Concept of Health Capital and the Demand for Health. *Journal of Political Economy*, 80(2):223–255.

Grossman, M. (2000). Chapter 7 - The Human Capital Model*. In Newhouse, A. J. C. and P., J., editors, *Handbook of Health Economics*, volume 1, Part A of *Handbook of Health Economics*, pages 347–408. Elsevier.

Grusky, D. B. (2014). *Social Stratification: Class, Race, and Gender in Sociological Perspective*. Westview Press, Boulder, CO, 4 edition edition.

Hernán, M. A. and Robins, J. M. (2006). Estimating causal effects from epidemiological data. *Journal of Epidemiology and Community Health (1979-)*, 60(7):578–586.

ICPSR (2008). County Characteristics, 2000-2007, United States. Technical report, Inter-University Consortium for Political and Social Research (ICPSR).

Kawachi, I., Kennedy, B. P., Lochner, K., and Prothrow-Stith, D. (1997). Social capital, income inequality, and mortality. *American journal of public health*, 87(9):1491–1498.

Lareau, A. (2002). Invisible Inequality: Social Class and Childrearing in Black Families and White Families. *American Sociological Review*, 67(5):747–776.

Leigh, A., Jencks, C., and Smeeding, T. M. (2009). Health and economic inequality. *The Oxford Handbook of Economic Inequality, Oxford University Press, Oxford*.

Link, B. G. and Phelan, J. (1995). Social Conditions As Fundamental Causes of Disease. *Journal of Health and Social Behavior*, pages 80–94.

Lynch, J., Smith, G. D., Harper, S. A., Hillemeier, M., Ross, N., Kaplan, G. A., and Wolfson, M. (2004). Is income inequality a determinant of population health? Part 1. A systematic review. *Milbank Quarterly*, 82(1):5–99.

Mackenbach, J. P. (2012). The persistence of health inequalities in modern welfare states: The explanation of a paradox. *Social Science and Medicine*, 75(4):761–769.

Manzo, G., editor (2014). *Analytical Sociology: Actions and Networks*. Wiley series in computational and quantitative social science. Wiley, Hoboken.

Marmot, M. (2004). The status syndrome: How social standing affects our health and longevity. 2004. *London: Bloomsbury*.

McElreath, R. (2016). *Statistical Rethinking: A Bayesian Course with Examples in R and Stan*, volume 122. CRC Press.

Miller, J. H. and Page, S. E. (2007). *Complex Adaptive Systems: An Introduction to Computational Models of Social Life*. Princeton University Press, Princeton, N.J.

Mullainathan, S. and Shafir, E. (2013). *Scarcity: Why Having Too Little Means so Much*. Macmillan.

Müller, B., Bohn, F., Dreßler, G., Groeneveld, J., Klassert, C., Martin, R., Schlüter, M., Schulze, J., Weise, H., and Schwarz, N. (2013). Describing human decisions in agent-based models – ODD + D, an extension of the ODD protocol. *Environmental Modelling & Software*, 48:37–48.

National Academy of Sciences (2015). Assessing the Use of Agent-Based Models for Tobacco Regulation. Technical report.

Neckerman, K. M. and Torche, F. (2007). Inequality: Causes and Consequences. *Annual Review of Sociology*, 33(1):335–357.

Newman, K. S. (1999). *Falling from Grace: Downward Mobility in the Age of Affluence*. University of California Press, Berkeley, 1st california pbk. print edition edition.

Page, S. E. (2015). What Sociologists Should Know About Complexity. *Annual Review of Sociology*, 41(1):21–41.

Palloni, A., Daza, S., and Emanuel, E. (2016). Income Inequality, Social Mobility and Mortality in the U.S. *CDE Working Paper, University of Wisconsin-Madison*, (No. 2016-02).

Palloni, A., Milesi, C., White, R. G., and Turner, A. (2009). Early childhood health, reproduction of economic inequalities and the persistence of health and mortality differentials. *Social Science & Medicine*, 68(9):1574–1582.

Pearl, J. (2014). Interpretation and identification of causal mediation. *Psychological Methods*, 19(4):459–481.

Phelan, J. C., Link, B. G., and Tehranifar, P. (2010). Social conditions as fundamental causes of health inequalities theory, evidence, and policy implications. *Journal of health and social behavior*, 51(1 suppl):S28–S40.

Pickett, K. E. and Wilkinson, R. G. (2015). Income inequality and health: A causal review. *Social Science & Medicine*, 128:316–326.

Pujol, G., Iooss, B., Veiga, A. J. w. c. f. S. D., Fruth, J., Gilquin, L., Guillaume, J., Gratiet, L. L., Lemaitre, P., Ramos, B., and Touati, T. (2015). *Sensitivity: Sensitivity Analysis*. R package version 1.11.

Railsback, S. F. and Grimm, V. (2011). *Agent-Based and Individual-Based Modeling: A Practical Introduction*. Princeton University Press.

Saltelli, A., Campolongo, F., and Ratto, M. (2004). *Sensitivity Analysis in Practice: A Guide to Assessing Scientific Models*. Wiley, Hoboken, NJ.

Savage, M. (2015). *Social Class in the 21st Century*. Penguin UK.

Snijders, P. T. A. B., Bosker, P. R., Snijders, T., and Bosker, R. (2011). *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling*. Sage Publications Ltd, second edition.

Sorokin, P. A. (1959). Social and cultural mobility. *New York*, 4:99–145.

Speybroeck, N., Van Malderen, C., Harper, S., Müller, B., and Devleesschauwer, B. (2013). Simulation Models for Socioeconomic Inequalities in Health: A Systematic Review. *International Journal of Environmental Research and Public Health*, 10(11):5750–5780.

Thiele, J. C. and Grimm, V. (2015). Replicating and breaking models: Good for you and good for ecology. *Oikos*, 124(6):691–696.

Thiele, J. C., Kurth, W., and Grimm, V. (2014). Facilitating Parameter Estimation and Sensitivity Analysis of Agent-Based Models: A Cookbook Using NetLogo and R. *Journal of Artificial Societies and Social Simulation*, 17(3):11.

Tolbert, C. M. and Sizer, M. (1996). U.S. Commuting Zones and Labor Market Areas: A 1990 Update.

Torche, F. (2015). Analyses of Intergenerational Mobility An Interdisciplinary Review. *The ANNALS of the American Academy of Political and Social Science*, 657(1):37–62.

Truesdale, B. C. and Jencks, C. (2016). The Health Effects of Income Inequality: Averages and Disparities. *Annual Review of Public Health*, 37(1):413–430.

Venkataramani, A. S., Brigell, R., O'Brien, R., Chatterjee, P., Kawachi, I., and Tsai, A. C. (2016). Economic Opportunity, Health Behaviors, and Health Outcomes in the United States: Population-Based, Multilevel Study.

Venkataramani, A. S., Chatterjee, P., Kawachi, I., and Tsai, A. C. (2015). Economic Opportunity, Health Behaviors, and Mortality in the United States. *American Journal of Public Health*, 106(3):478–484.

Wagstaff, A. and van Doorslaer, E. (2000). Income Inequality and Health: What Does the Literature Tell Us? *Annual Review of Public Health*, 21(1):543–567.

Wilensky, U. and Rand, W. (2015). *An Introduction to Agent-Based Modeling: Modeling Natural, Social, and Engineered Complex Systems with NetLogo*. The MIT Press, Cambridge, Massachusetts.

Wilkinson, R. (1992). Income distribution and life expectancy. *British MEdical Journal*, 304(6820):165–6.

Wilkinson, R. and Pickett, K. E. (2009). Income inequality and social dysfunction. *Annual Review of Sociology*, 35:493–512.

Winship, S. (2015). The great Gatsby curve: All heat, no light. *Brookings post: http://brookings.edu/blogs/social-mobility-memos/posts/2015/05/20/20-opposition-great-gatsby-curve-winship*.

Wodtke, G. T., Harding, D. J., and Elwert, F. (2011). Neighborhood Effects in Temporal Perspective The Impact of Long-Term Exposure to Concentrated Disadvantage on High School Graduation. *American Sociological Review*, 76(5):713–736.

Wolfson, M., Gribble, S., and Beall, R. (2017). Exploring Contingent Inequalities: Building the Theoretical Health Inequality Model. In Grow, A. and Van Bavel, J., editors, *Agent-Based Modelling in Population Studies*, volume 41, pages 487–513. Springer International Publishing, Cham.

Zhang, D., Giabbanelli, P. J., Arah, O. A., and Zimmerman, F. J. (2014). Impact of Different Policies on Unhealthy Dietary Behaviors in an Urban Adult Population: An Agent-Based Simulation Model. *American Journal of Public Health*, 104(7):1217–1222.

Zheng, H. (2012). Do people die from income inequality of a decade ago? *Social Science & Medicine*, 75(1):36–45.