THREE ESSAYS ON MORTALITY, HEALTH, AND MIGRATION

David Frankenfield

A DISSERTATION

in

Demography

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2016

Supervisor of Dissertation

____

Michel Guillot, Professor of Sociology

Graduate Group Chairperson

____

Hans-Peter Kohler, Frederick J. Warren Professor of Demography

Dissertation Committee

Michel Guillot, Professor of Sociology
Hans-Peter Kohler, Frederick J. Warren Professor of Demography
Douglas Ewbank, Research Professor of Sociology

THREE ESSAYS ON MORTALITY, HEALTH, AND MIGRATION

ACKNOWLEDGMENT

I have learned a great deal writing this dissertation, and have many people to thank. I must first recognize my committee, Professors Michel Guillot, Hans-Peter Kohler, and Douglas Ewbank. Without their expertise, time, and effort, I would not have been able to finish this dissertation. I am deeply appreciative of my cohort, whose companionship, knowledge, and support were critical in my time at the University of Pennsylvania. Furthermore, I am grateful to the faculty, staff, and students, both past and present, in the Graduate Group in Demography, especially Tanya Yang, Dawn Ryan, and Nykia Perez. Finally, I acknowledge the generous financial support of the National Institute of Child Health and Human Development that allowed me to complete the dissertation.

ABSTRACT

THREE ESSAYS ON MORTALITY, HEALTH, AND MIGRATION

David Frankenfield

Michel Guillot

This dissertation contains three chapters covering relationships between mortality, health, migration. Using a discrete time failure model via pooled logistic regression, chapter one shows that self-rated health is a significant predictor of mortality in rural Malawi, a context that differs greatly from those in most previous studies. This indicates that the well-established relationship between self-rated health and mortality extends to even the most resource poor settings. In chapter two, life tables are created for each state in the United States that allow for the measurement of migration over the full life course. The results show that migrants are generally positively selected on their health, and more importantly that migration reduces inequality in mortality between states. This is a contrast to other research on geographical inequality in mortality, which typically does not point to migration as a driver of other observed mortality trends. Finally, using a marginal model through generalized estimating equations, analysis in chapter three shows the varying degree to which internal migrants in the United States are selected on their health. Individuals were selected most significantly on measures of disability, and analyzing only married couples gave the strongest results by showing how individuals can be selected on a spouse's health. Since couples often move together, marriage is an important dimension of health selective migration on the individual level in the United States.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: Health Perceptions and Mortality in Malawi

Introduction

Self-rated health—defined as the answer to the single item question, "In general, how would you rate your health?"—is a key measure of health status in populations. Its utility is derived in large part from its ease of collection and simplicity of interpretation. Over time it has increasingly been used to shape policy, and is influential in the distribution of health resources and the study of health inequality (O'Reilly and Rosato 2010). This is a result of a longstanding and growing body of research that has determined self-rated health to be a significant predictor of mortality. Recent scholarship indicates that this is an association which has only strengthened over time (Schnittker and Bacak 2014). However, the determinants of self-rated health are both numerous and wide in scope, so reporting is subjective in nature. Though research pertaining to this relationship is extensive, comparatively little of it examines developing countries. This is unfortunate, considering the large number of health surveys conducted in these areas, as well as the relative difficulty in assessing population health in such places. Given the subjectivity of self-rated health, the context under consideration may greatly influence the generally accepted finding.

This study uses data from the Malawi Longitudinal Study of Families and Health to determine the significance of self-rated health as a predictor of mortality in rural Malawi, a previously unstudied sub-Saharan African context. The setting of this analysis is important in that it serves to test existing knowledge concerning the relationship between self-rated health and mortality in an environment representative of other contexts in which this relationship has never been explored. The MLSFH collected self-rated health at multiple survey waves, which allows for the primary measure of interest to vary over time. In addition, at certain waves respondents were directly asked their personal mortality perceptions, which allows for a direct comparison with self-rated health.

Background

While self-rated health has not proven to be universally significant across each study and setting, better self-rated health is generally accepted to be positively associated with a decreased risk of mortality (Idler and Benyamini 1997; Mossey and Shapiro 1982). Specifically, though populations are examined at different times, using various methods, and in an assortment of places, fair or poor ratings of self-rated health are good predictors of subsequent mortality for both sexes in most contexts (Benyamini and Idler 1999; DeSalvo et al. 2006; Idler and Benyamini 1997). Previous research has investigated the relationship between self-rated health and a multitude of factors, determining that it can be related to socioeconomic status, gender, race/ethnicity, marital status, the presence of chronic illness, and limitations in daily activities or physical functioning (Burstrom and Fredlund 2001; Dowd and Todd 2011; Prus 2011; Reile and Leinsalu 2013; Xu et al. 2010). Notably, women usually report themselves to be in worse health than men (Ginneken and Groenewold 2012). These findings arise from a variety of backgrounds, and it is undeniably true that the relative importance of these determinants is not uniform in all countries (Prus 2011).

A solid conceptual understanding of self-rated health is an important component of both justifying and conducting research of this nature. In a general sense, self-rated health can be conceptualized as that which contains information not only about the health status of respondents, but also about characteristics of the respondents themselves, including education, standard of living, and beliefs about what "good health" actually means (Duncan and Frankenberg 2002). This is a good starting point, but the best understanding of self-rated health should formalize the evaluation and reporting of self-rated health. To date, this was best accomplished by Jylha (2009). In her model, an individual must first decide what actually forms ‘health’, specifically in terms of its constituent parts. Cultural and historical understandings of health are integral to this process, during which an individual might consider medical diagnoses, functional status, bodily sensations and symptoms, and risks to future health. Next, one must evaluate health in a general sense, considering reference groups, knowledge of previous health, and health expectations. Disposition and age are important factors affecting this comparative evaluation. Finally, a decision must be reached about how to express health, given the constrained nature of the answer choices. Cultural conventions of expression, both positive and negative, as well as the conceptualization of the scale, will be important here. A critical but perhaps subconscious aspect of this determination is figuring out which choice appears to be the ‘normal’ option, and then comparing one’s personal situation to that.

If the previous model governs our thinking about self-rated health, there are several aspects of Malawi that differentiate it from other settings. Culturally, this study population is quite different than almost every other previously examined, and the potential effects of this difference are not directly measureable. Previous research has shown that social networks in Malawi are structured and gendered, and that they played a significant role in the formation of AIDS prevention strategies. Such strong social ties will be important when individuals assess their health in relation to reference groups. Perhaps rural Malawians have a unique understanding of their peers in terms of health, and the gendered nature of the social networks is of note due to differences in reporting of self-rated health by gender. In addition, there is a comparative lack of formal medical knowledge in this population, as access to healthcare is low in rural Malawi. The exception would be knowledge about HIV/AIDS, as testing was offered during data collection waves. This means that study participants are more knowledgeable about HIV status than the Malawi population in general. However, before testing, research shows that individuals in fact overestimated their likelihood of infection, which negatively affected their subjective well-being. Testing in some ways then helps to normalize perceptions of HIV risk, and plausibly had divergent effects on inputs for reporting self-rated health, depending on HIV status. Finally, the study population has a high level of functional limitations (Kohler et al. 2014). These types of health problems are immediately perceptible, and thus can be evaluated in the process of reporting self-rated health. Disability may therefore play an outsize role in determining self-rated health in this population, given the relative lack of information about other health risks. Indeed, more generally it has been shown that self-rated health in the African context relates more to that of physical health (including chronic disease and functional limitations) as opposed to mental health (Onadja et al. 2013). This result mimics the inclusion of these as controls in studies taking place in developed countries (Benyamini and Idler 1999). In addition to disease and functional limitations, education and social capital are thought to be especially important determinants of self-rated health in the African context (Cramm and Nieboer 2011).

The existence of factors that may modify the reporting of self-rated health, and thus alter its relationship with mortality, require the exploration of the relationship in many contexts. Though the evidence is well documented in developed, Western settings, it is much scarcer in developing countries. However, some corroboration does exist (Frankenberg and Jones 2004; Hirve et al. 2012; Ng et al. 2012). In fact, there are a handful of studies that have examined the relationship between self-rated health and mortality in Africa (Ardington and Gasealahwe 2014; Dzekedzeke, Siziya and Fylkesnes 2008), and the results align with the typical finding. However, Malawi is unlike the settings previously studied. Malawi, located in southeastern Africa, is one of the poorest nations in the world, in which the population is predominantly rural, and the majority of individuals are employed in subsistence agriculture. Age patterns of mortality and cause of death profiles in Malawi contrast those found in currently developing nations. The disease burden is high compared to many of the countries where the relationship between self-rated health and mortality has been previously studied. Partially due to the high prevalence of HIV, resulting in a life expectancy at birth that is still under 60 years (Kohler et al. 2014). The two previous studies, which were carried out in South Africa and Zambia, occurred in countries classified as having medium human development by the UN. However, Africa is not a monolith. Malawi is classified as low human development, with only 13 other countries having a lower score (Malik 2014).

This analysis makes a significant contribution by adding to our knowledge on the ability of self-rated health to predict mortality in Africa, as the rural Malawian context has never before been studied. Malawi is also more similar to the many settings in which this relationship has not yet been examined than much of the previous work in this area, which focused on highly developed nations. In addition, it is essential to note that many of the highly cited articles that explore the relationship between mortality and self-rated health strictly use mortality follow-ups, and reference back to one measure of self-rated health that was collected at baseline (Idler and Angel 1990; Idler, Russell and Davis 2000; Miilunpalo et al. 1997; Mossey and Shapiro 1982). The advantage of this particular study is that self-rated health was collected in two year intervals (at each wave), which adds additional information to the model and allows the primary independent variable to change over time.

Data and Methods

The Malawi Longitudinal Study of Families and Health (MLSFH), formerly the Malawi Diffusion and Ideational Change Project, is a longitudinal data collection project that has been conducted in rural Malawi since 1998. The MLSFH is implemented in three regions of Malawi (Rumphi, Mchinji, and Balaka) that are similar in economic context, but are heterogeneous in marital patterns, religion, and education. The focus of the study is "studying the mechanisms that individuals, families, households, and communities develop and use in a poor rural setting to cope with the impacts of high morbidity and mortality in their immediate living environment" (Kohler et al. 2012). Entry into the sample after the initial wave in 1998 is dependent in part upon the year in which the respondent enters. Notably, in 2004 an adolescent sample was added, while, in 2008, the data was bolstered by a parent sample. More specific information regarding the data, setting, and sampling frame of the study can be found in the cohort profile (Kohler et al. 2014).

This paper focuses on self-rated health, which was not included in the survey until 2004, and the last full survey was completed in 2010. Thus, four waves of data (2004, 2006, 2008, and 2010) are included in the survival analysis, resulting in a maximum follow-up period of six years. In addition, the timing of death can only be assessed when a survey is conducted, since all that is known is whether an individual died between surveys. Without an exact date of death, time must be treated discretely in the analysis. This means that the data is constructed in person-period format, where covariates measured in one survey predict survival as recorded in the following survey. The values for all covariates except region and sex are allowed to change over time, but this only occurs at each survey time. As a result, the maximum possible number time intervals is three, which is also the maximum possible records contributed to the analysis by any individual. These records correspond to the 2004, 2006, and 2008 covariates predicting outcomes in the subsequent survey. Further, the relatively few number of time intervals results in a large number of tied event times in the outcome variable, making the use of techniques like Cox regression quite difficult. However, this survival data can be analyzed using a discrete time failure model via pooled logistic regression, where individuals are followed while they are at risk of having an event, and then not afterwards (Singer and Willett 2003). This discrete time failure model appropriately handles the large amount of ties in the dependent variable, allows for predictors to vary over time, and is statistically similar to the time dependent covariate Cox regression approach (D'Agostino et al. 1990). The model uses a logit transformation of hazard, which entails several assumptions: the model is a proportional odds model, and the shape of baseline hazards are similar, even if at different values of the covariates the relative level may change. The baseline hazard in this model is simply a function of time, and adjusting for additional covariates adds complexity to the model. Additionally, since a hazard function expresses the conditional probability of event occurrence, all records in the person-period data set are assumed to be conditionally independent (Singer and Willett 2003), meaning the model does not have to control for clustering within individuals. This model has a long history of use to examine event histories in discrete time, and is the most appropriate modeling choice for these data.

As in any survey, missing data is an issue with the data set. The amount of data missing for each variable is shown in Appendix Table A1.1. To sustain statistical efficiency, missing data was imputed using multiple imputation (Rubin 1987; van Buuren 2007). Though it is not possible to validate every assumption this strategy entails, it is superior to single imputation. It better handles the uncertainty of generating missing values, and avoids a significant reduction in sample size that results from dropping all observations with any missing data. All imputed variables were treated as either continuous or multinomial. Before imputation, variables with many possible responses (ethnicity and religion for example) were recoded into a smaller number of categories, which allows the imputation to converge. For the same reason, several categorical variables were imputed as continuous variables, and then rounded to the appropriate value. Missing values were imputed for every individual at every time period, and then the data were reshaped so that each individual had four records, each representing one of the survey waves. Individuals who do not experience death by 2010 are right censored. A true person-period data set contains one record for each time period for which an individual is at risk of death. Therefore, at this juncture, any fully imputed records occurring after a respondent had already died were deleted. Furthermore, other fully imputed records were retained only if they occurred between records containing some information available in the original data set. This strategy results in two implicit assumptions. First, it is assumes individuals who were absent when a survey was conducted were indeed still alive. Second, if a respondent was found to be dead in the survey following a unit non-response, it is assumed that their death occurred between the missing wave and the final wave (see Note in Appendix Table A1.1 for more detail on the imputation and exclusion criteria).

The outcome variable in the analysis is death, a dichotomous variable based on the outcome of the survey. Vital status was determined using the survey outcome variable in the data key of the MLSFH. Outcomes other than completed or dead were assumed to be missing, and, if such records were not excluded based on the exclusion criteria, data values were imputed. There is ongoing work within the MLSFH to improve vital status data, which will improve future versions of this analysis. The primary explanatory variable is self-rated health, a categorical variable representing the answer to the question, "In general, would you say your health is excellent, very good, good, fair, or poor?" Secondary analyses substitute direct estimations of mortality risk for self-rated health. Survey respondents were asked to assess their probability of death on one, five and ten year time horizons, with responses scored on a scale of zero to ten. This is done to investigate how predictive self-rated may be as compared to direct mortality perceptions. Other covariates are added to the models to control for the demographic, socioeconomic, and health characteristics of the respondents. These include age at first survey, gender, wealth quintile (taken from calculations by the study team), education, marital status, religion, ethnic group, and region of residence. Descriptive statistics for the individuals contributing to the analysis are located in Table 1.1. Several of these variables are directly related to mortality, self-rated health, or both, and thus must be included, while others serve primarily as controls. HIV status, which was first tested for by the survey team in 2004, and then again in 2006 and 2008, is included due to the relatively high prevalence among this population. A variable for time is included, which assumes linear relationship between time and logit hazard, while also contributing to the baseline hazard function. The linear assumption is appropriate given the outcome of interest and short duration of follow-up. Each regression has two versions: one with only time, age sex, and health included, and then a fuller version with all covariates. Odds ratios are presented for the full sample, and then excluding HIV positive individuals. For models with mortality perceptions, results are reported for simple logistic regressions on death based on different subsamples of the data.

Results

The results for self-rated health are located in Table 1.2, where some general patterns emerge. Risk of mortality does not seem to vary based on ethnicity, marital status, religion, wealth, or education, regardless of model specification. Some regional variation appears, as individuals who live in the Balaka region have significantly elevated odds of death. Age and sex are both unsurprisingly strongly associated with mortality, as older people and males are at greater risk of death. The time covariate is also significant, indicating that probability of death is higher at the end of the period than at the beginning. In the full sample HIV status is included as a covariate, and death is predictably much more likely for individuals who are HIV positive. Finally, and most importantly, those who report fair or poor health have significantly elevated odds of death as compared to those who consider themselves to be in good health. In all the models, fair and poor are combined into one category due to the small size of the 'poor' category. In addition, if we consider only the full sample, those who report excellent health have significantly lower odds of death than those in good health.

The MLSFH provides an interesting opportunity to directly investigate how predictive individual evaluations of mortality risk might be. Table 1.3 displays odds ratios for a similar set of regressions as in Table 1.2, but instead mortality perceptions are substituted for self-rated health. These are simple logistic regressions, in which all covariates are measured only at baseline and mortality is followed up later. The results are generally consistent with those in Table 1.2, save two significantly elevated odds ratios for those of average wealth in the full sample, and less evidence of regional variation. However, when one compares mortality perception to self-rated health, an obvious difference appears. None of the odds ratios for mortality perception are significant, and they mostly hover near a value of one. This is a stark contrast to the self-rated health analysis.

Discussion

Taken together, these results suggest that lower ratings of self-rated health do indeed indicate higher risk of mortality. In addition, those in excellent health have significantly lower odds of death than the reference category 'good'. Furthermore, HIV positive status also imparts its own mortality risk, and as such, one that may distort the effects of poor self-rated health in these models. The general results of the analysis are a confirmation of that which has been found in previous literature, and further validates the evidence regarding self-rated health and HIV found in other parts of Africa (Ardington and Gasealahwe 2014; Dzekedzeke et al. 2008). Malawi represents perhaps the most resource-poor setting in which the predictive ability of self-rated health on mortality has been confirmed, and thus these findings stand apart from the others in the literature. It has been suggested that self-rated health is reported through a cognitive process that is innately subjective and contextual, and that its genesis is the biological and physiological state of the individual, explaining its association with mortality (Jylha 2009).

Previous research into the relationship between self-rated health and mortality can generally be split into three groups. A clear minority fail to find a significant relationship between the two (Bath 2003; Idler and Benyamini 1997). For those that do report a significant finding, some are able to describe it as a dose-response relationship (Bopp et al. 2012; Burstrom and Fredlund 2001). Indeed, the majority of the papers collected by Idler and Benyamini (1997) can be characterized in this manner. Finally, some research is only able to isolate the strong predictive power of the 'poor' health rating on elevated future mortality (Af Sillen et al. 2005; Ben-Ezra and Shmotkin 2006). However, modeling decisions are impactful in this case, as many studies reduce the self-rated health variable from five categories down to as few as two, therefore restricting the ability to uncover any dose-response relationship. The results from the full sample in this analysis could perhaps be categorized as dose-response, given the significant odds ratios for both those in excellent and poor health. Yet dropping the HIV positive individuals, which results in a sample that is at least somewhat more similar to others considered in past research, results in a reduction of the significance of the excellent category. However, the likelihood of reporting individual categories is susceptible to high variation across contexts, especially for the 'fair' category. Translational issues contribute to this variation, and overall semantic issues in making international comparisons should induce caution when attempting to explicitly contextualize one set of findings among many others from different places (Schnohr et al. 2016).

Idler and Benyamini (1997) propose several ways to interpret the relationship between self-rated health and mortality. Some correspond more directly to developed countries, but there are several that speak directly to the context of this analysis. Among them are the fact that self-rated health is a more inclusive and accurate measure of health status and risk factors than other measured covariates, and that self-rated health is a dynamic evaluation that judges both trajectory and level of health. The applicable interpretations undoubtedly vary by context, but evidence presented here suggests that information about self-rated health adds to knowledge regarding mortality risk. The authors also suggested that new research endeavor to explore special populations. This has certainly been the case, as summarized by Jylha (2009). It is possible to think conceptually about this study sample as a special population, one representing an exceedingly rural, impoverished population in sub-Saharan Africa.

In many developed nations, previous work on mortality trends has established the existence of gradients in health and mortality by socioeconomic status. However, in this sample this gradient does not appear, as none of the results for wealth or education are significant in the analysis that controls for other covariates. There may be several reasons that explain this result. It could be the gradient does not exist in this context. Alternatively, it may be that other covariates in the model are more pivotal in characterizing mortality trends in Malawi. Finally, it is possible that the relative poverty of the sample as a whole prevents those gradients from appearing, since there is not enough heterogeneity in the socioeconomic status of the individuals present in the sample, though it should be noted that other work in Africa has reported a gradient (Ardington and Gasealahwe 2014). Whatever the case, this result could be explored further in another analysis.

Even the most conservative reading of these results would include the fact that poor ratings of health in rural Malawi are predictive of mortality. By comparison, it is curious that direct perceptions of mortality were not in any case related to mortality in follow up, when other work has shown anticipated survival reflects actual survival

(Adams et al. 2014). For the one year mortality probabilities these estimates are incredibly conservative, as mortality was assessed two years after data collection, which doubles the amount of time in which deaths could occur. As for the five year mortality probabilities, the follow up occurred after only four years, meaning this analysis may not be able to capture an effect that does exist. Yet overall, it seems to be the case that when asked directly, this population is not able to directly assess mortality accurately, even though the typical finding for self-rated health appears.

This study does have some limitations. The age reporting in this sample is suspect. There is uncertainty as to whether all the individuals actually know their exact ages, but there is also inconsistency of age reporting between surveys. Though the problem remains, confidence can be placed in the significant effect of age on mortality, which matches the demographic expectation. Another problem is missing data, which again is characteristic of data sources like this. The missing information was imputed, but there is a non-negligible amount of missingness. Most studies, when possible, attempt to control for measurable health when investigating this question. The only health measure included here was HIV status, which was undoubtedly important in this setting, but the inclusion of measures of chronic disease or functional limitations would surely improve the analyses. However, the strengths of this analysis included a prospective design with significant follow up and a statistical method that allows the primary independent variable, along with other covariates, to vary with time, adding a significant amount of information to the model.

Conclusion

This paper confirms that fair or poor self-rated health, as assessed by an answer to a single question on a health survey, is predictive of mortality in a sub-Saharan African setting. The results of this analysis encourage the continued use of self-rated health as an indicator of population health. In Malawi, a previously unstudied context, self-rated health exhibits essentially the same effects as it has in developed populations around the world, and also extends the preliminary evidence found in developing areas such as India and Indonesia. There is more work to be done in the assessment of health and mortality in developing contexts, but the evidence presented here suggests that self-rated health, which is present in many health surveys around the world, is indeed an important indicator of health. This very data set can be used in the future to delve more deeply into the mechanics of health perception in sub-Saharan Africa, especially in light of the passage through the HIV epidemic.

References

Adams, J., E. Stamp, D. Nettle, E.M.G. Milne, and C. Jagger. 2014. "Socioeconomic Position and the Association between Anticipated and Actual Survival in Older English Adults." Journal of Epidemiology and Community Health 68(9):818-825.

Af Sillen, U., J.A. Nilsson, N.O. Mansson, and P.M. Nilsson. 2005. "Self-Rated Health in Relation to Age and Gender: Influence on Mortality Risk in the Malmo Preventive Project." Scandinavian Journal of Public Health 33(3):183-189.

Ardington, C.and B. Gasealahwe. 2014. "Mortality in South Africa: Socio-Economic Profile and Association with Self-Reported Health." Development Southern Africa 31(1):127-145.

Bath, P.A. 2003. "Differences between Older Men and Women in the Self-Rated Health Mortality Relationship." Gerontologist 43(3):387-395.

Ben-Ezra, M.and D. Shmotkin. 2006. "Predictors of Mortality in the Old-Old in Israel: The Cross-Sectional and Longitudinal Aging Study." Journal of the American Geriatrics Society 54(6):906-911.

Benyamini, Y.and E.L. Idler. 1999. "Community Studies Reporting Association between Self-Rated Health and Mortality - Additional Studies, 1995 to 1998." Research on Aging 21(3):392-401.

Bopp, M., J. Braun, F. Gutzwiller, D. Faeh, and G. Swiss Natl Cohort Study. 2012. "Health Risk or Resource? Gradual and Independent Association between Self-Rated Health and Mortality Persists over 30 Years." Plos One 7(2):10.

Burstrom, B.and P. Fredlund. 2001. "Self Rated Health: Is It as Good a Predictor of Subsequent Mortality among Adults in Lower as Well as in Higher Social Classes?" Journal of Epidemiology and Community Health 55(11):836-840.

Cramm, J.M.and A.P. Nieboer. 2011. "The Influence of Social Capital and Socio Economic Conditions on Self-Rated Health among Residents of an Economically and Health-Deprived South African Township." International Journal for Equity in Health 10:7.

D'Agostino, R.B., M.L. Lee, A.J. Belanger, L.A. Cupples, K. Anderson, and W.B. Kannel. 1990. "Relation of Pooled Logistic-Regression to Time-Dependent Cox Regression-Analysis - the Framingham Heart-Study." Statistics in Medicine 9(12):1501-1515.

DeSalvo, K.B., N. Bloser, K. Reynolds, J. He, and P. Muntner. 2006. "Mortality Prediction with a Single General Self-Rated Health Question: A Meta-Analysis." Journal of General Internal Medicine 21(3):267-275.

Dowd, J.B.and M. Todd. 2011. "Does Self-Reported Health Bias the Measurement of Health Inequalities in Us Adults? Evidence Using Anchoring Vignettes from the Health and Retirement Study." Journals of Gerontology Series B-Psychological Sciences and Social Sciences 66(4):478-489.

Duncan, T.and E. Frankenberg. 2002. "The Measurement and Interpretation of Health in Social Surveys." in Summary Measures of Population Health: Concepts, Measurement, Ethics, and Applications, edited by C.J.L. Murray, Joshua A. Salomon, Colin D. Mathers, Alan D. Lopez. Geneva: World Health Organization.

Dzekedzeke, K., S. Siziya, and K. Fylkesnes. 2008. "The Impact of Hiv Infection on Adult Mortality in Some Communities in Zambia: A Cohort Study." Tropical Medicine & International Health 13(2):152-161.

Frankenberg, E.and N.R. Jones. 2004. "Self-Rated Health and Mortality: Does the Relationship Extend to a Low Income Setting?" Journal of Health and Social Behavior 45(4):441-452.

Ginneken, J.and W. Groenewold. 2012. "A Single-Vs. Multi-Item Self-Rated Health Status Measure: A 21-Country Study." The Open Public Health Journal 5.

Hirve, S., S. Juvekar, S. Sambhudas, P. Lele, Y. Blomstedt, S. Wall, L. Berkman, S. Tollman, and N. Ng. 2012. "Does Self-Rated Health Predict Death in Adults Aged 50 Years and above in India? Evidence from a Rural Population under Health and Demographic Surveillance." International Journal of Epidemiology 41(6):1719-1727.

Idler, E.L.and R.J. Angel. 1990. "Self-Rated Health and Mortality in the Nhanes-I Epidemiologic Follow-up-Study." American Journal of Public Health 80(4):446-452.

Idler, E.L.and Y. Benyamini. 1997. "Self-Rated Health and Mortality: A Review of Twenty-Seven Community Studies." Journal of Health and Social Behavior 38(1):21-37.

Idler, E.L., L.B. Russell, and D. Davis. 2000. "Survival, Functional Limitations, and Self Rated Health in the Nhanes I Epidemiologic Follow-up Study, 1992." American Journal of Epidemiology 152(9):874-883.

Kohler, H.-P., S.C. Watkins, J.R. Behrman, P. Anglewicz, I.V. Kohler, R.L. Thornton, J. Mkandawire, H. Honde, A. Hawara, and B. Chilima. 2014. "Cohort Profile: The Malawi Longitudinal Study of Families and Health (MLSFH)." International Journal of Epidemiology:dyu049.

Kohler, I.V., P. Anglewicz, H.-P. Kohler, J.F. McCabe, B. Chilima, and B.J. Soldo. 2012. "Evaluating Health and Disease in Sub-Saharan Africa: Minimally Invasive Collection of Plasma in the Malawi Longitudinal Study of Families and Health (MLSFH)."

Malik, K. 2014. "Human Development Report 2014. Sustaining Human Progress: Reducing Vulnerabilities and Building Resilience."

Miilunpalo, S., I. Vuori, P. Oja, M. Pasanen, and H. Urponen. 1997. "Self-Rated Health Status as a Health Measure: The Predictive Value of Self-Reported Health Status on the Use of Physician Services and on Mortality in the Working-Age Population." Journal of Clinical Epidemiology 50(5):517-528.

Mossey, J.M.and E. Shapiro. 1982. "Self-Rated Health - a Predictor of Mortality among the Elderly." American Journal of Public Health 72(8):800-808.

Ng, N., M. Hakimi, A. Santosa, P. Byass, S.A. Wilopo, and S. Wall. 2012. "Is Self-Rated Health an Independent Index for Mortality among Older People in Indonesia?" Plos One 7(4):8.

O'Reilly, D.and M. Rosato. 2010. "Dissonances in Self-Reported Health and Mortality across Denominational Groups in Northern Ireland." Social Science & Medicine 71(5):1011-1017.

Onadja, Y., S. Bignami, C. Rossier, and M.V. Zunzunegui. 2013. "The Components of Self-Rated Health among Adults in Ouagadougou, Burkina Faso." Population Health Metrics 11:12.

Prus, S.G. 2011. "Comparing Social Determinants of Self-Rated Health across the United States and Canada." Social Science & Medicine 73(1):50-59.

Reile, R.and M. Leinsalu. 2013. "Differentiating Positive and Negative Self-Rated Health: Results from a Cross-Sectional Study in Estonia." International Journal of Public Health 58(4):555-564.

Rubin, D.B. 1987. "Multiple Imputation for Nonresponse in Surveys." New York: J. Wiley & Sons.

Schnittker, J.and V. Bacak. 2014. "The Increasing Predictive Validity of Self-Rated Health." Plos One 9(1):11.

Schnohr, C.W., I. Gobina, T. Santos, J. Mazur, M. Alikasifuglu, R. Valimaa, M. Corell, C. Hagquist, P. Dalmasso, Y. Movseyan, F. Cavallo, S. van Dorsselaer, and T. Torsheim. 2016. "Semantics Bias in Cross-National Comparative Analyses: Is It Good or Bad to Have "Fair" Health?" Health and Quality of Life Outcomes 14:4.

Singer, J.D.and J.B. Willett. 2003. Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence: Oxford university press.

van Buuren, S. 2007. "Multiple Imputation of Discrete and Continuous Data by Fully Conditional Specification." Statistical Methods in Medical Research 16(3):219-242.

Xu, J., J.H. Zhang, L.Y. Feng, and J.C. Qiu. 2010. "Self-Rated Health of Population in Southern China: Association with Socio-Demographic Characteristics Measured with Multiple-Item Self-Rated Health Measurement Scale." Bmc Public Health 10:9-20.

Table 1.1 Background Characteristics of the MLSFH Sample, at Baseline

		Percent (%)
Sex	Male	43.4
	Female	56.6
HIV Positive		4.9
Wealth Quantile	Wealthiest	22.5
	Quantile 2	22.5
	Quantile 3	21.9
	Quantile 4	17.8
	Least Wealthy	15.3
Education	No Education	22.9
	Primary Level	65.3
	Secondary Level or Higher	11.8
Region	Mchinji	33.4
	Rhumpi	31.5
	Balaka	35.1
Ethnicity	Yao	25.9
	Tumbuka	31
	Chewa	29.1
	Other	14
Marital Status	Married	76.6
	Formerly Married	9.1
	Never Married	14.3
Religion	Christian	47.3
	Muslim	23.7
	Other	29
N		4429

Source: Author calculations from MLSFH, 2004-2010

Table 1.2 Logistic Regressions of Self-rated Health on Deaths in Malawi, 2004-2010

	Full Sample		Only HIV Negative	
Time	1.44***	1.55***	1.58***	1.62***
	(1.20, 1.72)	(1.28, 1.86)	(1.28, 1.96)	(1.31, 2.01)
Age at first survey	1.04***	1.05***	1.04***	1.04***
	(1.03, 1.05)	(1.04, 1.06)	(1.03, 1.05)	(1.03, 1.05)
Male	1.65***	1.70***	1.81***	1.84***
	(1.26, 2.17)	(1.24, 2.33)	(1.33, 2.48)	(1.29, 2.64)
Self-rated health				
Excellent	.54*	.59*	0.71	0.72
	(.33, .88)	(.36, .97)	(.41, 1.22)	(.42, 1.26)
Very good	0.94	0.93	1.14	1.12
	(.68, 1.30)	(.66, 1.29)	(.79, 1.64)	(.77, 1.63)
Good	-	-	-	-
Fair/poor	2.24***	2.22***	2.00**	2.09**
	(1.49, 3.35)	(1.47, 3.36)	(1.23, 3.26)	(1.28, 3.43)
HIV Positive	-	7.03***		
		(4.91, 10.1)		
		+ COV		+ COV
Observations	9847		9325	
Sample Size	4429		4211	

*p<.05 **p<.01 ***p<.001 95% CI in parentheses

Source: MLSFH 2004-2010 *Note*: Regressions with +COV at the bottom also included wealth, education, region, ethnicity, marital status, and religion as covariates. The only significant odds ratios to appear were for Balaka region (compared to Mchinji; 1.85 for full sample, 1.81 for HIV negative sample).

Table 1.3 Logistic Regressions of Mortality Perceptions on Deaths in Malawi, 2006-2010

	Full Sample						Only HIV Negative					
	One year, 2006		One year, 2008		Five year		One year, 2006		One year, 2008		Five year	
Age at first survey	1.04***	1.05***	1.03***	1.03***	1.04***	1.05***	1.05***	1.05***	1.04***	1.03***	1.05***	1.05***
	(1.03, 1.06)	(1.03, 1.07)	(1.02, 1.04)	(1.02, 1.04)	(1.02, 1.06)	(1.02, 1.07)	(1.03, 1.07)	(1.02, 1.07)	(1.02, 1.05)	(1.02, 1.05)	(1.03, 1.07)	(1.02, 1.07)
Male	1.39	1.73	1.77**	1.99**	1.45	1.80*	2.05*	2.52*	1.86**	2.04**	2.12*	2.62**
	(.82, 2.34)	(.96, 3.11)	(1.23, 2.55)	(1.29, 3.07)	(.86, 2.43)	(1.00, 3.23)	(1.09, 3.85)	(1.26, 5.05)	(1.25, 2.78)	(1.27, 3.27)	(1.13, 3.98)	(1.30, 5.26)
Mortality Expectation	0.92	0.85	1.04	0.997	1.003	0.95	0.89	0.84	0.99	0.95	0.97	0.94
	(.79, 1.07)	(.71, 1.01)	(.94, 1.15)	(.90, 1.11)	(.89, 1.13)	(.83, 1.09)	(.74, 1.07)	(.69, 1.03)	(.88, 1.11)	(.84, 1.08)	(.84, 1.11)	(.81, 1.09)
HIV Positive	-	7.46***	-	5.18***	-	7.25***						
		(3.85, 14.5)		(2.98, 9.0)		(3.74, 14.1)						
		+COV		+COV		+COV		+COV		+COV		+COV
Sample Size	3381		3524		3381		3185		3357		3185	
	*p<.05 **p<.01 ***p<.001						95% CI in parentheses					

Source: MLSFH 2006-2010 *Note*: Regressions with +COV at the bottom also included wealth, education, region, ethnicity, marital status, and religion as covariates. The only significant odds ratios to appear were for Balaka region in the full sample (compared to Mchinji; 2.56 for one year 2008) and average wealth in the full sample (compared to wealthiest; 3.21 for one year 2006, 3.30 for five year).

CHAPTER 2: State Variation in Life Expectancy and Its Relationship to Internal Migration in the United States

Introduction

Mortality inequality in the United States is greater than in many comparable countries, and remains a great source of concern among both researchers and policy makers. Early in 2016, new research showed the widening life expectancy differential between rich and poor in the US since the turn of the century (Chetty et al. 2016). Beyond the overall differential, the most captivating finding in the paper was that though the rich live longer everywhere, the mortality disadvantage of those with lower incomes varied considerably by local area. This is the most recent addition to what is a growing body of research that explores geographical inequality in the United States. Mortality differentials have been growing on a regional level, evidenced by a 30% growth in the disparity between the US South and the rest of the country and widening urban-rural inequality (Fenelon 2013; Singh and Siahpush 2014). In addition, there is an increasing amount of recent evidence of widespread county-level inequality, existing for both whites and blacks. It may also be the case that inequity is rising, and indeed in some counties life expectancy has decreased in the latter part of the 20th century (Cullen, Cummins and Fuchs 2012; Ezzati et al. 2008). Thus, the long-term observed aggregate health gains of the last half century have not been distributed evenly. In fact, when race and geography are considered together, life expectancy differentials between different groups within the United States can be as extensive as twenty years (Murray et al. 2006). Further, the American mortality experience is characterized by a greater degree of inequality than that which appears in nations of comparable standard of living, and there is evidence that geographic mortality differentials are continuing to increase (Fenelon 2013; Kulkarni et al. 2011; Wang et al. 2013; Wilmoth 2010). Though the literature paints a clear picture of the trends, the role internal migration may be playing in shaping these changes has not been firmly established.

Various mechanisms have been explored in order to explain this body of evidence. Previous research has investigated how issues such as race and socioeconomic status, among other considerations, contribute to the formation of inequality in health and mortality (Adler and Rehkopf 2008; Elo 2009; Lantz et al. 1998; Williams and Collins 1995). The sum of these inputs in producing spatial inequality is not trivial. Though research specifically focused on geographic inequality has considered mechanisms like income inequality, labor market conditions, access to medical care, residential segregation, and cultural factors, and the degree to which they are explanatory is variable and generally small. However, consensus has been reached on another factor. Chetty el al. (2016), Ezzati et al. (2008), Fenelon (2013), Kulkarni et al. (2011), and Murray et al. (2006) all suggest that health behaviors are the vital determinant of spatial inequality. This paper is not an endeavor to contradict this consensus, but instead suggests previous scholarship has overlooked internal migration as a potential contributor to spatial mortality inequality. Most of the work cited above either fails to adequately address internal migration or cites the work of Ezzati et al. (2008), which is a strong treatment of the issue, but is subject to the inherent constraints in making measurement choices about geography and migration.

The focus of this paper is to examine the impact of internal migration on geographic mortality differentials. Specifically, this analysis investigates how the movement of United States citizens across state boundaries affects the observed state-level inequalities in mortality in the United States, contributing to scholarship that is divided in its opinion regarding whether migration exacerbates or reduces spatial inequality in mortality. Though the direction of the effect needs not be identical in every setting, the importance of this issue should not be understated. Additionally, not only does this analysis emphasize the United States context, it also employs a lifetime measure of migration, a technique which is rare in previous literature. Using data from the United States National Vital Statistics System and the United States Census (excluding the foreign born), five life tables are produced for each state-sex combination, each of which is representative of a different hypothetical migration stream within the population. These life tables are compared through the utilization of life expectancy at age 15, and important distinctions are drawn by comparing mortality outcomes for different migrant streams. This research attempts to estimate the effect of internal migration by simulating and then comparing life expectancies for different migrant streams.

Background

There are many previous studies that have investigated the relationship between migration and geographical inequalities in health or mortality. Much of this research was motivated by a concern that the casual use of these indicators was not informed by knowledge of the potential bias resulting from migration, and also by the suggestion that selective migration was at least in part driving the formation or exacerbation of geographical disparities in health or mortality. The conclusions from this body of previous research are not uniform. Most studies find that migration intensifies underlying geographical differences in health or mortality (Brimblecombe, Dorling and Shaw 1999; Brimblecombe, Dorling and Shaw 2000; Kibele and Janssen 2013; Martikainen et al. 2008; Norman, Boyle and Rees 2005; O'Reilly and Stevenson 2003; Riva, Curtis and Norman 2011). Others cannot produce definitive evidence to suggest that migration has a strong effect in one way or another (Boyle, Norman and Rees 2002; Connolly and O'Reilly 2007). There are a few cases in which migration is suggested to attenuate geographic differences, in some areas in an age-dependent manner (Connolly, O'Reilly and Rosato 2007; Jongeneel-Grimen et al. 2011; Jongeneel-Grimen et al. 2013). The age-dependent nature of some findings in of note, given that migrant health selection can in some instances also vary with age (Lu 2008). There is even work that produces descriptive results whose interpretation can demonstrate how both effects may exist (Brown and Leyland 2010; Verheij et al. 1998). Indeed, many scholars have discussed the theoretical possibility that multiple effects can exist, and cause of death may play a role (Larson, Bell and Young 2004). Previous discussion also emphasizes that we must consider the time period during which the relationship is examined and the demographic characteristics of the population under study. However, two measurement choices that all of these studies must make have an indelible and complicating effect on their outcomes.

The first component of research regarding geographic disparities is the geographic scale of measurement. The majority of the most recent scholarship considering both geography and migration comes from the United Kingdom, with some from other

European countries like the Netherlands. The size of these nations, along with the availability of superior data as compared to the United States, has undoubtedly impacted the choice of geographic scale. Many of the studies cited in the previous paragraph investigate small areas, usually called wards or postcodes, while more historical research may focus on the urban-rural dynamic of specific cities (see for example Verheij et al. 1998). This urban-rural split is characteristic of research that occurs on a regional level, before more detailed data sources may have been available. Essentially, measuring on different geographic levels has allowed for the exploration of the scale dependence of this relationship, which can be conceptualized in two ways. It is possible to assert that if a relationship can be confirmed on multiple scales, then it must be true. Alternatively, it is perhaps the case that different scales involve different explanatory dynamics, meaning that different processes drive the relationship at different scales (Dunn, Schaub and Ross 2007). Regardless, this choice is important, as it has been directly demonstrated that alteration of the geographic scale of analysis can affect the conclusions about the effect of migration (Brimblecombe et al. 1999).

The second measurement choice of importance is how to define a migrant. Most studies reviewed above utilized short-term measures of migration which are typically less than five years duration (Boyle et al. 2002; Jongeneel-Grimen et al. 2013; Kibele and Janssen 2013; O'Reilly and Stevenson 2003). Some studies are able to use a measurement of medium duration, often ranging from ten to twenty years (Connolly et al. 2007; Norman et al. 2005; Riva et al. 2011). Finally, only a handful of studies have been able to study health outcomes and measure migration with a lifetime measure of migration (Brimblecombe et al. 1999; Brimblecombe et al. 2000). This definition is a specific application of the idea to use early life exposures as represented by birthplace to study health outcomes in later life (Fang, Madhavan and Alderman 1996; Rasulo et al. 2012). Migrations measured over disparate lengths of time will undoubtedly be influenced by different factors. Though the decision may depend heavily on data source, even with full information about the individual, it would be difficult to determine the appropriate boundary (Boyle 2004). In addition, the choice of migration measure determines which past health exposures are expressed in the population categorized as migrants in their new place of residence, which is critical to the question at hand. This definitely occurs if a short-term measure is used, as any discrete cut off leaves essentially identical individuals on either side of the line. However, a short term cut off has the more impactful result of counting relatively recently arrived individuals as part of the native population, though their health exposures over the life course differentiate them from the rest of the native population. Due to data restrictions, short term measures of migration are common in previous research, but it is possible that these measures do not fully capture the migrant population. Indeed, it has been suggested that the findings in previous research investigating the relationship between geographical inequality and migration are related to the length of migration measure used. Specifically, studies utilizing longer time frames for measuring migration often found evidence for exacerbation of area-level inequality (Jongeneel-Grimen et al. 2013).

It has been suggested that health behaviors are a primary determinant of geographic patterns in mortality in the United States. However, given the evidence from other countries concerning health selective migration and geography, it is impossible to ignore the potential bias imparted by the movement of individuals between places. Though there is little research in the United States context that specifically investigates the potential bias internal migration imparts to geographic mortality differentials, there are many studies that explore those differentials in general. Internal migration is usually addressed only in passing, to state whether or not the authors anticipates that it could bias their results. The primary source referenced to explain why internal migration does not change the interpretation of their results comes from Ezzati et al. (2008). However, this study simulates the potential effects of migration using a one-year measure of migration, on the county level, and for only seven years in the 1990s. Given that measurement choices can introduce variability in the effect of migration, it is striking that much of the prominent US research on geographic mortality differentials depends on one treatment of migration.

This scarcity highlights the need for other investigations as to the effect of internal migration on the measurement of geographic mortality variation, but there is ample flexibility in how to address the issue. Much of the scholarship cited above simply treats internal migration as something to be explained away in the context of their specific research question. The strength here is to make internal migration the primary focus, and to broaden the viewpoint of migration beyond only several years. By examining migration over the life course, this paper implicitly questions whether earlier work has taken a too narrow view in confronting the effect of internal migration. The objective here is to demonstrate the effect internal migration may exert if considered in a different manner, which has little precedent especially in research that concentrates on the United States.

Data and Methods

The data for this analysis originates from two sources. Mortality data is taken from the US National Vital Statistics System through the Multiple Cause of Death public-use microdata files. Three years of deaths (1999-2001) are used for the analysis, in order to follow conventional methods in creating state-level life tables. The death records contain information on the state of birth, state of residence at death, sex, and age of the individual. Population denominators for the calculation of death rates are obtained using information from the 5% sample of the 2000 US Census. This data was accessed through IPUMS (Ruggles et al. 2015), and the calculation of population denominators makes use of the provided person weights. Since state of birth and state of death are used to define all deaths and exposure terms, all the analyses exclude the foreign born, as these individuals do not have a valid state of birth under the framework of this analysis. The effect of the removal of the foreign born is displayed in Appendix Figure A2.1 and Table A2.1. Given the generally superior health of international migrants, most of the differences shown in the figure are negative, and generally the largest differences occur in states with sizable migrant populations. As for aggregate inequality, small but negative percentage change in all inequality measures for both men and women further demonstrates the relative good health migrants. More importantly, minor reduction in inequality also reveals that this analysis is still capable of producing significant conclusions even though part of the population is excluded.

Mortality rates are constructed as any demographic rate, with events in the numerator and an exposure term in the denominator. The deaths in the three years are pooled and used for the numerator, and they are classified by state of birth and state of residence, in addition to age and sex. Population denominators are similarly indexed. In order to ensure the numerator and denominator refer to the same interval of time, exposure terms are estimated by multiplying population estimates by three. The mortality rates are then grouped into life tables by state and sex, and five different tables are produced: each state-sex pairing has a residence, inmigrant, nonmover, outmigrant, and nativity life table. People contribute to the population and death counts in multiple tables, according to the following rules. For the residence life table, death and exposure terms are contributed by all people living in a particular state. The inmigrant life table for the same state only includes individuals who lived there in the year 2000, but were born in another state. Each state's nonmover life table only contains individuals who were born in that state and also lived there in 2000. Contributions to each outmigrant life table are made by individuals born in the state of interest, but in the year 2000 lived in any other state. Finally, the nativity life table for each state is composed of all persons who were born in that state, regardless of where they lived during the year 2000. To be clear, the residence life table corresponds to a life table for the merged nonmover and inmigrant populations, and the nativity life table corresponds to a life table for the merged nonmover and outmigrant populations. All life tables ignore multiple moves over the life course, including potential returns to state of birth, a limitation discussed later.

Implicit in the choice to construct life tables in this manner is the assumption that many of the determinants for future mortality are related to early life experiences. This means that taking a life course perspective on migration allows for an alternative but appropriate way to measure health as it relates to future interstate migrations over time. There will of course be cases where this would not be the best way to measure migration; however, this is true of any possible migration measure, and on the aggregate this measure should be effective. Essentially, this research attempts to estimate the effect of internal migration on state-level mortality through a counterfactual thought experiment: How would state level mortality differentials change if we could compare the current population distribution to what it would look like if people stayed in their place of birth?

The main outcome measure is life expectancy at age 15, so as to minimize the effect of early life mortality, which occurs before most people have agency in their migration decisions. The residence life table is the real world, where migration occurs, while the nativity life table represents a hypothetical world without internal migration. By comparing mortality under these different migration regimes, the analysis can examine differences that are potentially attributable to migration. A series of graphs is reported that breaks the overall effect down into its components: changes as a result of outmigration and inmigration. These effects are examined by looking at differences between migrant streams in each state. Residence and nativity life expectancies are also used to calculate four measures of aggregate inequality between states: Gini coefficient, Theil index, squared coefficient of variation, and mean logarithmic variation. These measures are sensitive to different parts of the distribution of life expectancies, and thus provide a full picture of the potential change in inequality. Though these measures are most recognizably used to measure income inequality, they can also be applied to other ratio-level variables (Goesling and Firebaugh 2004). The calculations are similar, in that they all are different transformations of several quantities, including population proportions and life expectancy ratios.

Results

Life expectancies for each state, sex, and migration stream are reported in Appendix Tables A2.2 and A2.3. The following analysis takes these raw numbers and decomposes the overall effect of migration into its component parts. For each state, the overall effect of migration includes both inmigration and outmigration effects. To begin, Figures 2.1-2.4 show the outmigration and inmigration effects for men. The outmigration effects are displayed in Figures 2.1-2.2, calculations which are accomplished by comparing the outmigrant population of each state to the other individuals born in the same state. Figure 2.1 shows health selection, which is the difference between the outmigrant stream and the nonmover stream. For each state, this comparison includes all individuals that share that state as their birthplace, and by taking this difference we can show how those who no longer live there compare to those who never left. For men, this difference is almost uniformly positive, meaning that men who outmigrated from a given state usually have a higher life expectancy than those who remained in the state. Figure 2.2 displays the difference between the nonmigrant and nativity life tables, which isolates the true effect of outmigration by showing what happens to the nativity value when outmigrants leave the state. It is the true effect because this difference is weighted according the proportions of the population in each state that are nonmovers and outmigrants. The values for men are mostly negative, which makes sense given the patterns in the previous graph. The magnitude of the differences are smaller now, since outmigrants are typically a relatively small part of the population, but outmigration generally has a depressing effect on a state's life expectancy. Since outmigrants have higher life expectancies, their removal from the nativity life table results in decreases in life expectancy.

To analyze the effect of inmigration, the exact opposite calculations are performed. Figure 2.3 shows the absolute differential in life expectancy between inmigrants and nonmovers for men in every state. In this figure the migrant stream is composed of people born in many states, whereas in the previous two figures migrants and nonmovers were all born in the same state. The purpose here is simply to compare the mortality of the two groups that make up the residence life table in each state. As nonmover life expectancy increases, migrant mortality is not able to keep up, and we see a downward trend in the differential for men. Though migrant men are in this case generally healthier when compared to nonmovers, this is not only dependent on being selected on their health. Where these migrant men are moving to matters, since it is possible for a man positively selected on his health to move to a state where the nonmover life expectancy is much higher than where he left. However, in most cases inmigrant men have higher life expectancies than nonmovers in the state of destination, especially in states with low nonmover life expectancy. The true healthy migrant effect is shown in Figure 2.4, which displays the difference between the residence and nonmover life tables. Now the differentials incorporate the health of the migrants, but are weighted by the relative proportion of inmigrants in each state, so that the magnitude is a true reflection of the positive or negative influence of inmigration. The downward trend is now flatter than in Figure 2.3, and magnitudes are smaller. Nevertheless, migrant men tend to exert a positive influence on the mortality in their destination states, which means that inmigrants tend to improve state life expectancies, particularly in states with lower life expectancies.

Figures 2.5-2.8 show the exact same calculations as in Figures 2.1-2.4, but now the calculations are for women's life expectancy. Health selection shown in Figure 2.5 is now more mixed than it was for men, as the number of states with positive and negative differentials is roughly equal. This result is logical, given that, historically, fewer of the women contributing to these life tables were career oriented, and instead were more often migrants due to the labor market realities of their husbands. As a result, the true effect of outmigration, shown in Figure 2.6, is again mixed when compared to men, who had mostly negative differentials. Therefore, it cannot be said, as was the case for men, that the outmigration of women has a depressing effect on a state's life expectancy. For women, the effect can be positive and negative, and must be evaluated on a case by case basis. Figures 2.7-2.8 evaluate the effect of inmigration for women. When they are compared to the corresponding figures for men (2.3-2.4) we see that the differentials are more mixed in sign for women, whereas men had a higher proportion of states with positive differentials. However, the figures are similar in that there is a downward slope to the data in all four. Yet again, tied migration is probably one of the main factors that drives the difference in effect between men and women.

Finally, Figures 2.9 and 2.10 show the overall effect of migration on the state level for men and women by combining the two effects. By subtracting the nativity life table from the residence life table, the resulting differential incorporates the effect of outmigration of individuals from a particular state while simultaneously integrating the effect of inmigration from other states. Broadly, this is a comparison of the world as it truly exists to a world in which everyone remains in their state of birth. A positive differential indicates that the net effect of migration is positive, and the reverse is true of a negative differential. There are two critical points that these figures make clear. First, there is significant heterogeneity in the overall effect of migration on the state level for both sexes. Many of the differentials fall within a half year of zero, which in and of itself is a notable differential. Moreover, for both sexes, several state differentials approach or exceed one year, an occurrence which is not limited to only positive or negative differentials. This illustrates that the effect is particularly meaningful for some states, and that migration is capable of producing strong changes in both directions. The other key point to derive from these two figures involves the downward slope of the data points they contain. The horizontal axis in the plots is nativity life expectancy, a measure that represents each state's mortality without the migration effects typically incorporated in such calculations. The downward slope of the figures indicates that migration tends to have a positive effect on life expectancy where nativity life expectancy is lower, and vice versa. The conclusion to be drawn is that migration reduces inequality between states.

Additionally, the other measure of state level life expectancy that does not include migration effects is the nonmover life table. The downward slope present in these figures remains even when substituting in the nonmover value on the horizontal axis (not shown), which instills even more confidence in this conclusion.

To address the effect of migration on mortality inequality in the aggregate, the values for residence and nativity life expectancy are used to compute the inequality measures reported in Table 2.1. Nativity values are reported first, followed by the values for residence, which simulates moving from a world with no migration to the real world of freedom of movement. Percent changes in moving from nativity to residence are reported in the last column. For both men and women, there are slight decreases in all indicators using residence life expectancy as opposed to nativity life expectancy, though the values were small to start. This indicates that there is slightly more inequality in a hypothetical world without migration as compared to traditional state-level mortality estimates. Though all values and differences are small, there is a significant decrease in the measures when assessed in relative terms, through percent change. The results presented in this table further reinforce the overall conclusion from the state level results. Every indication from this analysis suggests that that migration in fact reduces state level mortality inequality.

Discussion

This paper is the first analysis of its kind, and shows that internal migration in the United States reduces state level mortality inequality, which is an important finding in the research on area level mortality inequality. Many other papers conclude that migration was not driving increasing inequality (typically measured on the county level). Often a contributing factor to this pattern of findings was the inability or disinterest to examine migration as a contributor in its own right, as opposed to simply explaining away the potential contribution of migration to the specific research question being addressed. No previous research has definitively refuted the possibility that migration is a significant mechanism through which mortality inequality might be established or maintained. This analysis suggests that this is indeed probable, though measurement choices often influence the ability to detect the influence of migration. There is no indication that internal migration exacerbates inequality. The best understanding of the results is to recognize that on the aggregate level there is reduction in inequality. When inequality is further examined for individual states, the calculations in Figures 2.9 and 2.10 demonstrate that there are both strong positive and strong negative effects of migration. States with low life expectancies tend to improve due to migration, and states with higher life expectancies tend to experience decreases, resulting in the aggregate reduction in inequality.

The strongest assessment of the potential bias imparted on geographic disparities in mortality by internal migration comes from Ezzati et al. (2008). Much of the other work in this area depends on this analysis when considering internal migration. The present analysis does not directly contradict the conclusions of these authors, but instead suggests that measurement of migration and geography as it pertains to this question is a delicate issue. In the future, work with more diverse treatments of these variables is needed to further understand how migration affects geographic mortality trends. This literature also produced consensus as to the primary role of health behaviors in determining geographic inequality, a narrative perhaps bolstered by these results. It is possible that health behaviors are indeed the primary driver of geographic mortality variation, and that migration, especially when measured over long periods of time, is influential in redistributing these behaviors across state lines. Much has been made of how the US South is separating itself from the rest of the country, e.g. Fenelon (2013). This particular trend may be due to health behaviors (specifically smoking), but the overall effects of migration from this analysis also show some geographic clustering. The effect of migration tends to be negative in the Midwest, whereas mostly positive effects appear in the South. Appendix Figures A2.2-A2.9 show the overall effect of migration from this analysis for each Census Region. Within regions they are organized by Census Division, and the overall pattern of the effects suggest that migration effects on geographical mortality inequality may be directly linked to previous findings on health behaviors, a potential connection which deserves further scrutiny in the future.

Other studies that discount the effect of migration usually do not have the capability to measure migration over the life course. When lifetime migration is considered, as it was here, a subtle but meaningful effect on geographic inequalities in health and mortality appears. A general reduction in inequality as a result of migration is a noteworthy finding for those who work in public health or public policy, as it suggests that any state level disparities in mortality observed using the traditional life table would be even greater in the absence of migration. For example, the effect of migration is variable in size, but there are some large positive effects in the southern part of the

United States. This area of the country is typically compared unfavorably to the rest of the country in terms of mortality outcomes. This analysis suggests that native populations in these states may be especially vulnerable in terms of their health, more so than standard calculations would suggest.

Finally, this analysis also reaffirms some findings from other research. The fact that men are more consistently and strongly positively selected on health is not a surprising finding, in light of the historical differences in reasons for migrating by sex. The cohorts contributing the most deaths to this analysis would have had a higher occurrence of women moving as tied movers to their husbands than today, which indicates that the health selection of migration would be more visible for men. Changing patterns of migration will undoubtedly have an effect on where migrants most contribute to the positive health of their destination in the future, as will the fact that mobility in the United States has been lower recently than in much of the previous century. Nevertheless, given that migrants generally tend to be positively selected on health, I would expect the presence of the healthy migrant effect on the state level to endure.

There are important limitations to this work. The first is the use of dual data sources in constructing death rates, which allows for mismatch between numerator and denominator. The more important limitation is the crude measure of migration utilized. Lifetime migration allows for a wide scope of analysis through which certain trends can be discovered, but it also results in lots of missing migrations. This measure is not able to capture the specific effect of health exposures over the life course based on geography, similar in spirit to the data issues normally described as potential explanation for the migrant mortality advantage. In addition, this measure allows for the internal migration version of the salmon bias, in which individuals spend the majority of their lives in a specific state and then move after retiring and before death. However, though much may be missed, a measure of lifetime migration does provide a unique perspective, and allows the analysis to capture the strong effect of early life conditions on eventual mortality.

Conclusion

The preceding analysis provides strong evidence that internal migration in the United States is subject to health selection, and reduces state level inequality in mortality. Researchers interested in geographic disparities in health and mortality should do their best to consider the effect that migration may have on their results, and endeavor to measure that effect whenever possible. However, there are extensions to this work that would strengthen this conclusion. First of all, future work could examine the role of migration in explaining changes over time in mortality inequality. That could start with a simple replication of this analysis in previous years, as well as updating the calculations with the most recent data possible. Further, only one measure of mortality was used in these analyses. However, an exploration of old age mortality, or a measure of temporary life expectancy, would deepen our understanding of the mechanisms at work here. In addition, the nature of the calculations allow for the parsing of the data along many lines of inquiry. Similar inquiries could be carried out by race/ethnicity or education, in order to more thoroughly develop our understanding of what drives these patterns. This would be only one specific example of a way to take these descriptive findings and attempt to further explain how and why these patterns occur. Finally, longitudinal data is a powerful tool in the exploration of migrant mortality advantage. Future research could utilize small subsamples of national populations that are followed over the life course to offer a more nuanced understanding of health selection of migration and its effect on regional mortality variation.

References

Adler, N.E.and D.H. Rehkopf. 2008. "US disparities in health: Descriptions, causes, and mechanisms." Annual Review of Public Health 29:235-252.

Boyle, P., Duke-Williams, Oliver. 2004. "Does Migration Exaggerate the Relationship between Deprivation and Self-reported Limiting Long-term Illness?" Pp. 129-148 in The Geography of Health Inequalities in the Developed World: Views From Britain and North America, edited by P. Boyle, Sarah Curtis, Elspeth Graham, Eric Moore. Burlington, VT: Ashgate Publishing Company.

Boyle, P., P. Norman, and P. Rees. 2002. "Does Migration Exaggerate the Relationship between Deprivation and Limiting Long-Term Illness? A Scottish Analysis." Social Science and Medicine 55(1):21-31.

Brimblecombe, N., D. Dorling, and M. Shaw. 1999. "Mortality and Migration in Britain, First Results from the British Household Panel Survey." Social Science and Medicine 49(7):981-988.

Brimblecombe, N., D. Dorling, and M. Shaw. 2000. "Migration and geographical inequalities in health in Britain." Social Science & Medicine 50(6):861-878.

Brown, D.and A.H. Leyland. 2010. "Scottish mortality rates 2000-2002 by deprivation and small area population mobility." Social Science & Medicine 71(11):1951-1957.

Chetty, R., M. Stepner, S. Abraham, and et al. 2016. "The association between income and life expectancy in the united states, 2001-2014." JAMA.

Connolly, S.and D. O'Reilly. 2007. "The contribution of migration to changes in the distribution of health over time: Five-year follow-up study in Northern Ireland." Social Science & Medicine 65(5):1004-1011.

Connolly, S., D. O'Reilly, and M. Rosato. 2007. "Increasing inequalities in health: Is it an artefact caused by the selective movement of people?" Social Science & Medicine 64(10):2008-2015.

Cullen, M.R., C. Cummins, and V.R. Fuchs. 2012. "Geographic and Racial Variation in Premature Mortality in the US: Analyzing the Disparities." Plos One 7(4):13.

Dunn, J.R., P. Schaub, and N.A. Ross. 2007. "Unpacking income inequality and population health - The peculiar absence of geography." Canadian Journal of Public Health-Revue Canadienne De Sante Publique 98:S10-S17.

Elo, I.T. 2009. "Social Class Differentials in Health and Mortality: Patterns and Explanations in Comparative Perspective." Annual Review of Sociology 35:553-572.

Ezzati, M., A.B. Friedman, S.C. Kulkarni, and C.J.L. Murray. 2008. "The reversal of fortunes: Trends in county mortality and cross-county mortality disparities in the United States." Plos Medicine 5(4):557-568.

Fang, J., S. Madhavan, and M.H. Alderman. 1996. "The association between birthplace and mortality from cardiovascular causes among black and white residents of New York City." New England Journal of Medicine 335(21):1545-1551.

Fenelon, A. 2013. "Geographic Divergence in Mortality in the United States." Population and Development Review 39(4):611-634.

Goesling, B.and G. Firebaugh. 2004. "The trend in international health inequality." Population and Development Review 30(1):131-146.

Jongeneel-Grimen, B., M. Droomers, K. Stronks, and A.E. Kunst. 2011. "Migration does not enlarge inequalities in health between rich and poor neighbourhoods in The Netherlands." Health & Place 17(4):988-995.

Jongeneel-Grimen, B., M. Droomers, K. Stronks, J.A.M. van Oers, and A.E. Kunst. 2013. "Migration and geographical inequalities in health in the Netherlands: an investigation of age patterns." International Journal of Public Health 58(6):845-854.

Kibele, E.and F. Janssen. 2013. "Distortion of regional old-age mortality due to late-life migration in the Netherlands?" Demographic Research 29:105-131.

Kulkarni, S.C., A. Levin-Rector, M. Ezzati, and C.J.L. Murray. 2011. "Falling behind: life expectancy in US counties from 2000 to 2007 in an international context." Population Health Metrics 9:12.

Lantz, P.M., J.S. House, J.M. Lepkowski, D.R. Williams, R.P. Mero, and J.M. Chen. 1998. "Socioeconomic factors, health behaviors, and mortality - Results from a nationally representative prospective study of US adults." Jama-Journal of the American Medical Association 279(21):1703-1708.

Larson, A., M. Bell, and A.F. Young. 2004. "Clarifying the Relationships between Health and Residential Mobility." Social Science & Medicine 59(10):2149-2160.

Lu, Y. 2008. "Test of the 'healthy migrant hypothesis': A longitudinal analysis of health selectivity of internal migration in Indonesia." Social Science & Medicine 67(8):1331-1339.

Martikainen, P., P. Sipila, J. Blomgren, and F.J. van Lenthe. 2008. "The effects of migration on the relationship between area socioeconomic structure and mortality." Health & Place 14(2):361-366.

Murray, C.J.L., S.C. Kulkarni, C. Michaud, N. Tomijima, M.T. Bulzacchelli, T.J. Iandiorio, and M. Ezzati. 2006. "Eight Americas: Investigating mortality disparities across races, counties, and race-counties in the United States." Plos Medicine 3(9):1513-1524.

Norman, P., P. Boyle, and P. Rees. 2005. "Selective migration, health and deprivation: a longitudinal analysis." Social Science & Medicine 60(12):2755-2771.

O'Reilly, D.and M. Stevenson. 2003. "Selective Migration from Deprived Areas in Northern Ireland and the Spatial Distribution of Inequalities: Implications for Monitoring Health and Inequalities in Health." Social Science & Medicine 57(8):1455-1462.

Rasulo, D., T. Spadea, R. Onorati, and G. Costa. 2012. "The impact of migration in all cause mortality: The Turin Longitudinal Study, 1971-2005." Social Science & Medicine 74(6):897-906.

Riva, M., S. Curtis, and P. Norman. 2011. "Residential mobility within England and urban-rural inequalities in mortality." Social Science & Medicine 73(12):1698-1706.

Ruggles, S., K. Genadek, R. Goeken, J. Grover, and M. Sobek. 2015. "Integrated Public Use Microdata Series: Version 6.0 [Machine-readable database]." Minneapolis: University of Minnesota.

Singh, G.K.and M. Siahpush. 2014. "Widening Rural-Urban Disparities in Life Expectancy, US, 1969-2009." American Journal of Preventive Medicine 46(2):E19-E29.

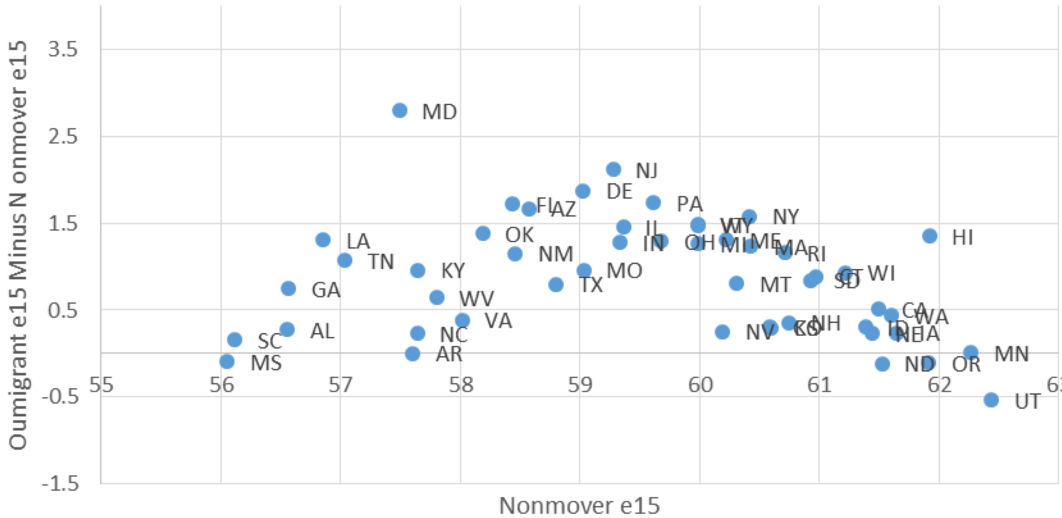
Verheij, R.A., H.D. van de Mheen, D.H. de Bakker, P.P. Groenewegen, and J.P. Mackenbach. 1998. "Urban-rural variations in health in the Netherlands: does selective migration play a part?" Journal of Epidemiology and Community Health 52(8):487-493.

Wang, H.D., A.E. Schumacher, C.E. Levitz, A.H. Mokdad, and C.J.L. Murray. 2013. "Left behind: widening disparities for males and females in US county life expectancy, 1985-2010." Population Health Metrics 11:15.

Williams, D.R.and C. Collins. 1995. "US Socioeconomic and Racial Differences in Health – Patterns and Explanations." Annual Review of Sociology 21:349-386.

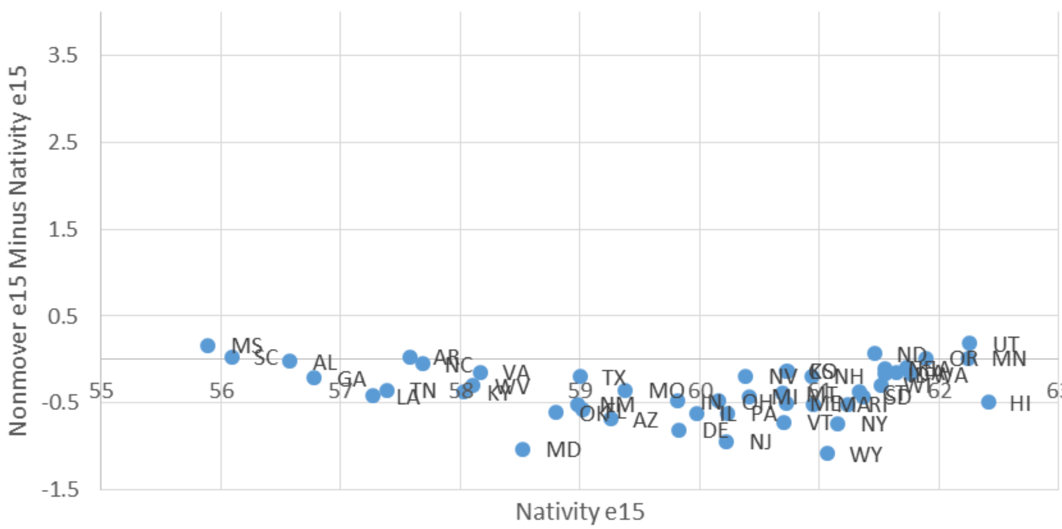
Wilmoth, J.R., Carl Boe, and Magali Barbieri. 2010. "Geographic Differences in Life Expectancy at Age 50 in the United States Compared with Other High-income Countries." Pp. 333-366 in International Differences in Mortality at Older Ages: Dimensions and Sources, edited by S.H. Preston, Eileen Crimmins, Barney Cohen. Washington, DC: National Academies Press.

Figure 2.1 Outmigrant Minus Nonmover Life Expectancy at Age 15 (e15), Men, 2000



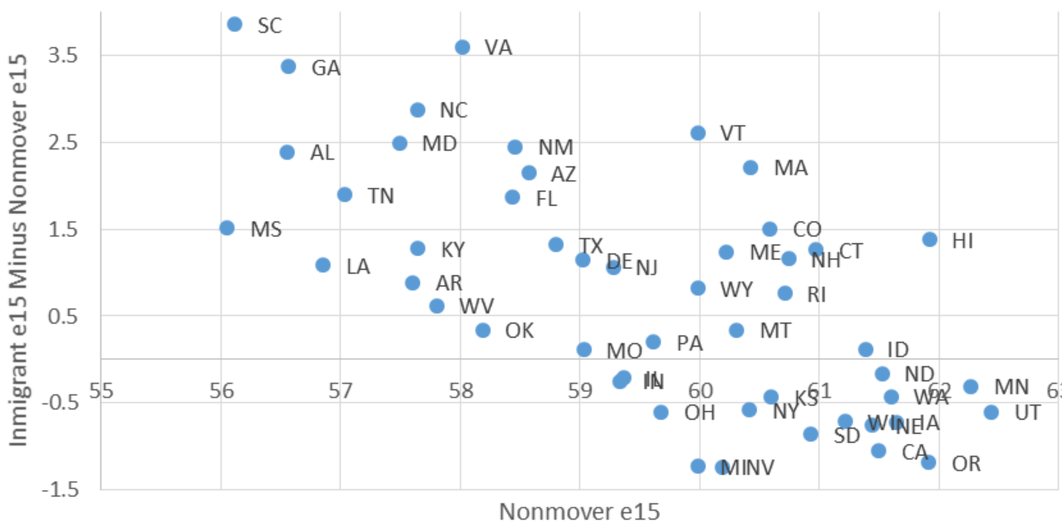
Source: Life expectancy calculations using data from the U.S. Census and National Vital Statistics System

Figure 2.2 Nonmover Minus Nativity Life Expectancy at Age 15 (e15), Men, 2000



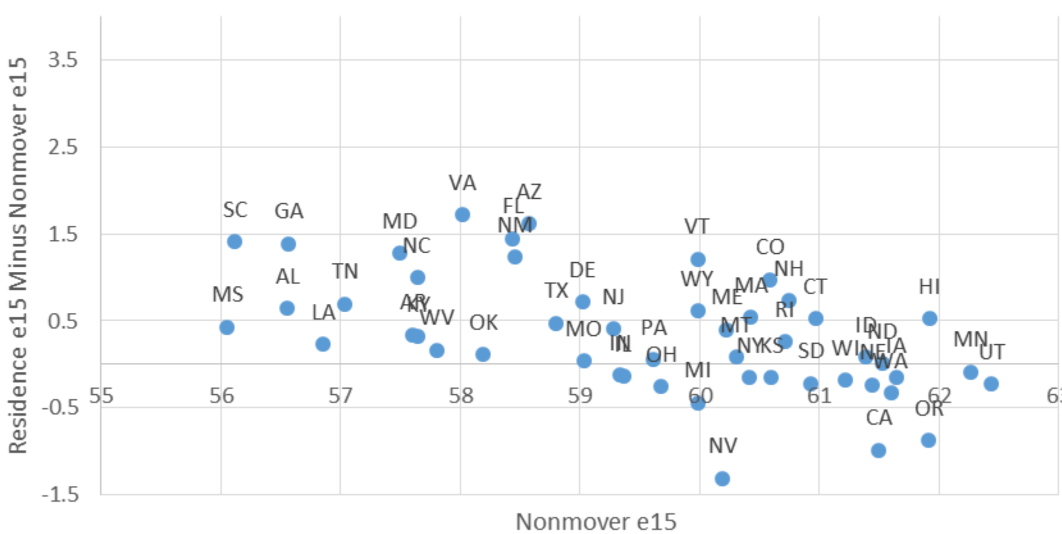
Source: Life expectancy calculations using data from the U.S. Census and National Vital Statistics System

Figure 2.3 Inmigrant Minus Nonmover Life Expectancy at Age 15 (e15), Men, 2000



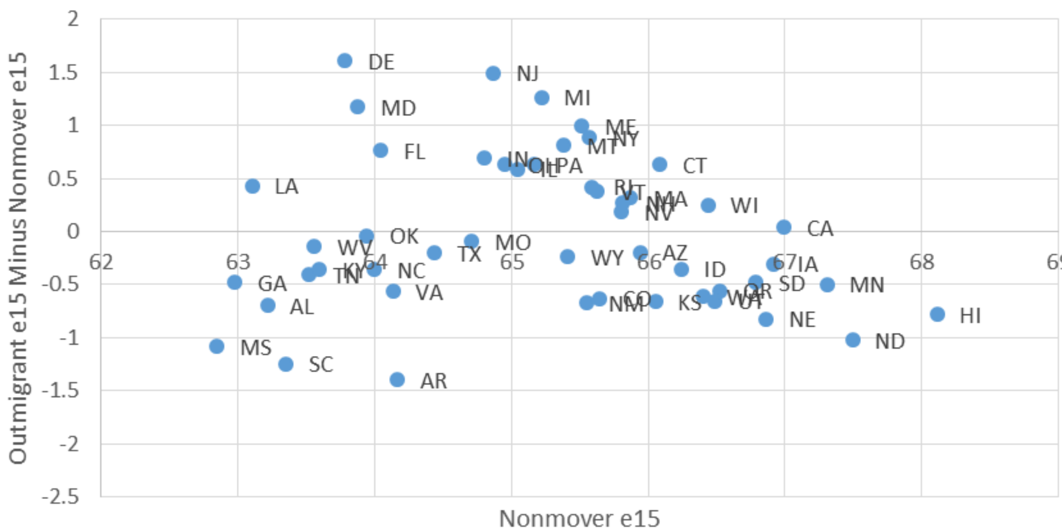
Source: Life expectancy calculations using data from the U.S. Census and National Vital Statistics System

Figure 2.4 Residence Minus Nonmover Life Expectancy at Age 15 (e15), Men, 2000



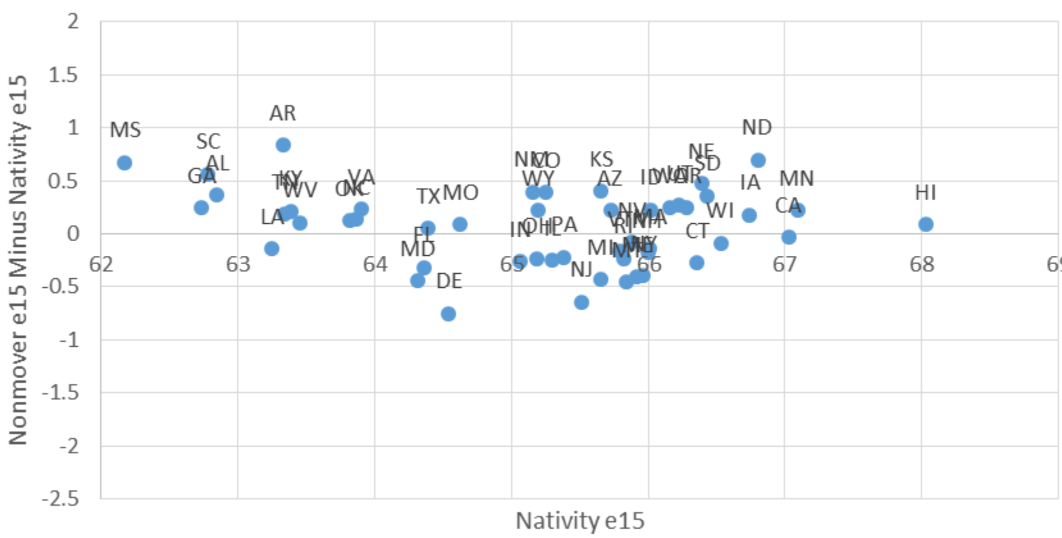
Source: Life expectancy calculations using data from the U.S. Census and National Vital Statistics System

Figure 2.5 Outmigrant Minus Nonmover Life Expectancy at Age 15 (e15), Women, 2000



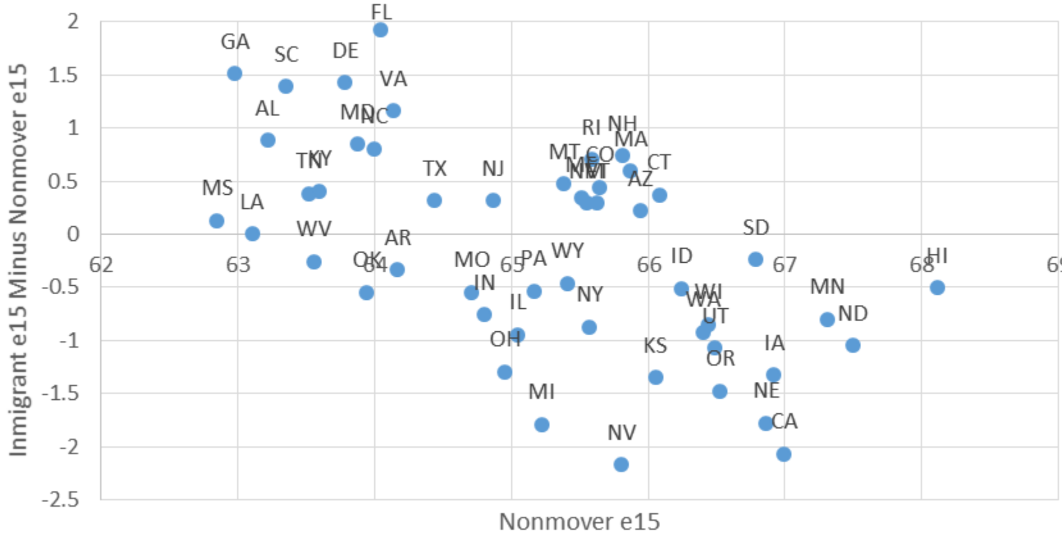
Source: Life expectancy calculations using data from the U.S. Census and National Vital Statistics System

Figure 2.6 Nonmover Minus Nativity Life Expectancy at Age 15 (e15), Women, 2000



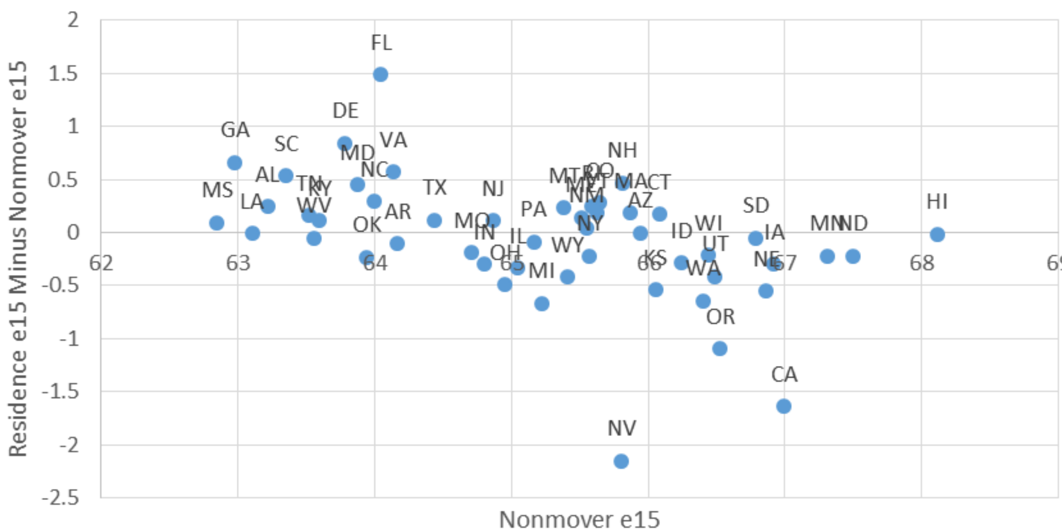
Source: Life expectancy calculations using data from the U.S. Census and National Vital Statistics System

Figure 2.7 Inmigrant Minus Nonmover Life Expectancy at Age 15 (e15), Women, 2000



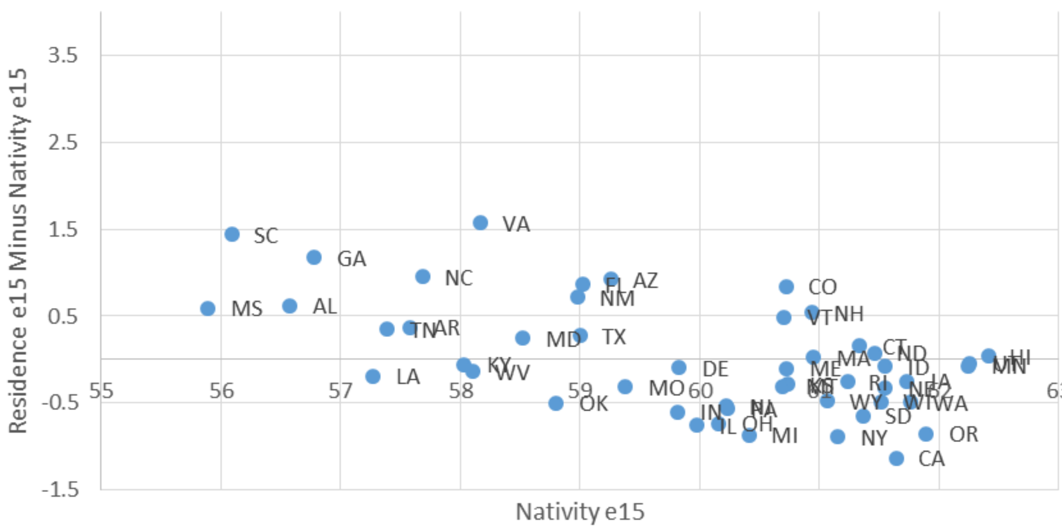
Source: Life expectancy calculations using data from the U.S. Census and National Vital Statistics System

Figure 2.8 Residence Minus Nonmover Life Expectancy at Age 15 (e15), Women, 2000



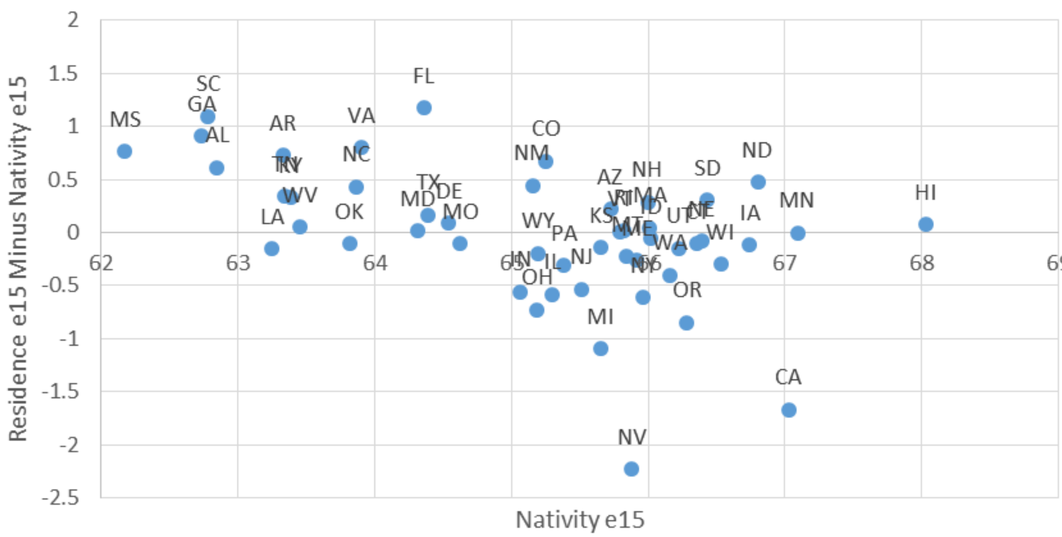
Source: Life expectancy calculations using data from the U.S. Census and National Vital Statistics System

Figure 2.9 Residence Minus Nativity Life Expectancy at Age 15 (e15), Men, 2000



Source: Life expectancy calculations using data from the U.S. Census and National Vital Statistics System

Figure 2.10 Residence Minus Nativity Life Expectancy at Age 15 (e15), Women, 2000



Source: Life expectancy calculations using data from the U.S. Census and National Vital Statistics System

Table 2.1 State Level Inequality in Life Expectancy at Age 15, 2000

	Men			Women		
	Nativity	Residence	Change	Nativity	Residence	Change
Gini coefficient	0.0153	0.0116	-24.1%	0.0112	0.00779	-30.1%
Theil index	0.00037	0.00022	-41.3%	0.0002	0.000098	-49.9%
Squared coefficient of variation	0.00074	0.00044	-41.2%	0.00039	0.00020	-50.0%
Mean logarithmic deviation	0.00038	0.00022	-41.5%	0.000196	0.000098	-50.3%

Source: Life expectancy calculations using data from the U.S. Census and National Vital Statistics System *Note*: Percent change reflects moving from the Nativity to the Residence value for life expectancy at age 15.

CHAPTER 3: Health Selective Internal Migration in the United States

Introduction

Human migration is transformational for both individuals and communities, and research investigating the health selective nature of this process deepens our understanding of the mechanisms that drive it. Perhaps the most prominent work examines the Hispanic population in the United States, and the supposed existence of the Hispanic paradox (Abraido-Lanza et al. 1999; Markides and Coreil 1986). Most often, health status is explored as it relates to migration behavior in an international context, while the dynamics of health selection among internal migrants has been left relatively untouched. Indeed, a general divide exists between the literatures that consider international migration and internal migration, despite the fact that some of the underlying forces could be similar, making a stronger connection between the two literatures valuable (Ellis 2012). The international migration literature provides general support for the assertion that migrants are positively selected on health, though the evidence is largely indirect.

This paper applies this hypothesis to internal migration, asking whether internal migrants in the United States are also positively selected on their health. Previous research scrutinizing the link between health and internal migration has often been conducted in European or Asian contexts. In some places this may be associated with a notable secular increase in the rate of migration, as is the case in China. However, though the United States is a historically mobile population, little research of this sort has been conducted in that context, though one example stands out (Halliday and Kimmitt

2008). Ultimately, “any attempt to build a single overarching theory of migration for all types of migration, for all parts of the world, developed and less developed, and for all periods of time, is illusory” (King and Skeldon 2010). The analysis presented here investigates whether a broadly applied migration theory explains the health patterns among internal migrants in the United States.

Thus, this is a fresh look at the issue, with the additional advantage of using the highest quality data to date. The superiority of these data arises from consistent follow up of the sample, and the inclusion of multiple health measures in the data. This paper uses the most recent data from the Panel Study of Income Dynamics to assess the relationship between health status and interstate migration in the United States. The primary goal is to determine to what extent, if at all, health status influences the likelihood of migration. As there is no established standard for health measurement among the many measures available to social scientists, this research can address which of self-rated health, disability, or health conditions is most related to migration. Finally, due in part to the composition of the sample, a subsample of married couples is examined to determine whether marriage is a moderator of the relationship between health and migration.

Background

Research on the relationship between migration and health or mortality has a long history in the demographic and social sciences literatures, often attempting to characterize the existence of a migrant mortality advantage in the context of international migration. This research has presented several potential reasons for the relatively superior health of migrant populations when compared to non-migrants. Of those, several are not as directly applicable to a similar investigation of the dynamics of internal migration, namely data quality issues and cultural factors.

The two most investigated explanations involve health selection. First, it has been hypothesized that migrants are typically positively selected on their health, meaning that migrant populations represent a particularly healthy group of individuals compared to their country of origin. Positive health selection also usually results in positive comparisons to the populations into which they settle, even if they experience relative socioeconomic disadvantage in their new surroundings. This is the essence of the 'Hispanic Paradox', which health selective migration may explain at least in part (Abraido-Lanza et al. 1999; Markides and Coreil 1986; Palloni and Morenoff 2001). Difficulty in empirically testing for the existence of this selection stems from the complexity of gathering data for migrant populations in both destination countries and a suitable comparison group in the country of origin. While in some studies support for the hypothesis is weak (Rubalcava et al. 2008), there is at least some direct evidence of health selective migration (Jasso 2004), though its importance across time and context is not certain. It is exceedingly likely that migrant advantage in health and mortality is not static (Borrell and Lancet 2012).

A second frequently cited contributor to health differentials between migrants and non-migrants is negative health selection at exit, usually later in life. This is more commonly referred to as 'Salmon bias' (Abraido-Lanza et al. 1999). Essentially, the proposal is that older individuals who are sick are more likely to return to their country of origin, leaving the healthier proportion of the immigrants as the group for which mortality is collected in the immigrant context. This phenomenon has been explicitly demonstrated in previous research, again especially as it pertains to Hispanic populations in the United States (Markides and Eschbach 2005; Palloni and Arias 2004; Turra and Elo 2008). Though it may not be immediately clear why the Salmon bias pertains in any way to internal migration, it has been suggested that health selective migration in general may vary with age. Healthy migrant effects should be stronger at younger ages, whereas Salmon bias type selection should prevail at older ages (Palloni and Arias 2004). Salmon bias is really just negative selection at older ages, and indeed age gradients in selective migration do appear in the literature (Connolly, O'Reilly and Rosato 2007; Jongeneel-Grimen et al. 2013; Markides and Eschbach 2005). Therefore, just as Salmon bias effects are observed in international migration, observing negative health selection at older ages in a study examining internal migration is not an unreasonable expectation.

Previous work exploring health selective internal migration maintains a geographical focus in certain areas, and the context dependent nature of this topic allows for results to differ across many dimensions. However, the healthy migrant hypothesis is generally upheld in internal migration studies. Much research documents the dynamics of health selection in migration in China, especially in the large flows from rural to urban areas. The results mostly support positive health selection (Chen 2011; Tong and Piotrowski 2012), and other studies find that the internal migration equivalent of the ‘Salmon bias’ also occurs in China (Hu, Cook and Salazar 2008; Lu and Qin 2014; Qi and Niu 2013). In the European context, health selective migration is most frequently discussed in terms of how it may shape geographical inequality, especially regarding material deprivation. The typical finding in this literature is the existence of an age gradient in the selection process, in which young people are positively selected and older individuals tend to be negatively selected (Boyle, Norman and Rees 2002; Connolly and O'Reilly 2007; Jongeneel-Grimen et al. 2013). However, other work fails to establish any significant relationship (Popham et al. 2011; Tinghog et al. 2011). There is less work on the U.S. population in the literature, though a few exceptions do exist. A common finding is negative health selection at older ages (Bentham 1988; Findley 1988), whereas only some find positive health selection for younger people (Bentham 1988).

Overall, there is general consensus in the literature concerning age gradients in health selection. However, several factors introduce nuance to this body of work which allow for variation. The settings where this research occurs is variable, and duration of study is dependent on data source. Using longitudinal data, Tong (2012) found that the strength of the relationship between health and migration diminished over time, signifying that length of follow-up can directly affect the ability to detect significant results. In addition, studies utilize different geographical scales on which to evaluate migration. The scale dependence of health selective migration is not the same across spatial units, and it is not advisable to treat this choice as simply incidental to the analysis. There is some work showing that the use of variably sized geographic measures can results in different inferences (Brimblecombe, Dorling and Shaw 1999; Dunn, Schaub and Ross 2007). Most significantly, health measures used in previous research are not uniform, nor are the results associated with them. Some studies had multiple measures available for analysis, and the findings vary. Given the ease with which it can be collected, self-rated health is a commonly available measure, and one upon which migrants are often selected (Chen 2011; Jongeneel-Grimen et al. 2013; Tong and Piotrowski 2012). However, others fail to find a relationship when health is measured this way, but instead point to mental health as the measure upon which people are selected (Larson, Bell and Young 2004). Yet here again, other research explicitly fails to find selection on mental health (Chen 2011). Disability, often measured using activities of daily living, is a measure of physical function shown to relate to geographical mobility (Lu 2008). The large body of literature on geographic inequality and deprivation in Europe mentioned above frequently measures health by limiting long term illness, which somewhat relates to disability. Finally, health conditions, including a host of acute and chronic diseases, are used when they are available; negative health selective migration is evident in some cases (Jongeneel-Grimen et al. 2013; Larson et al. 2004), but in other studies no relationship appears (Lu 2008).

One analysis in particular is relevant to this study, as it uses the same data source as this analysis (Halliday and Kimmitt 2008). Though the authors are interested in the same question that is at issue here, they use data from the Panel Study of Income Dynamics from the period 1984-1993. They find that individuals below age 60 are positively selected on their health, when measuring by self-rated health status. Men above age 60 are more mobile at the top and bottom of the health distribution, whereas there is no relationship between health and mobility for older women. Further, they find no relationship between migration and disability. However, the disability measure is crudely assessed based on the answer to only one question, and the authors acknowledge the high probability of measurement error. In a separate analysis of married individuals, they find that for men own health matters, whereas for women spouse's health is of most importance.

This research contributes to the literature by taking advantage of the quality of the newest data in the Panel Study of Income Dynamics. Health selection is examined longitudinally over a 14 year period following the sample used by Halliday and Kimmitt. The richness of the data allows for the comparison of self-rated health, disability, and chronic conditions as they relate to health selection, and since health is measured before migration the analysis is able to directly test for health selection among migrants. Thus this analysis is capable of uncovering new patterns of selection that the previous analysis with this data could not, while also replicating parts of the previous work to see if the results hold over time. This is the most recent, complete, and thorough evaluation of health selective migration in the United States, a setting where research of this sort is rare. Additionally, the composition of the sample requires the investigation of marriage as it relates to this literature, as a large portion of the sample under consideration is married couples.

Data and Methods

This study uses data from the Panel Study of Income Dynamics (PSID), a nationally representative sample which began in 1968, making it the longest running longitudinal household survey in the world. The data used here come from the years 1999 to 2013. The original sample began with over 18,000 people living in more than

5,000 families, and now follows their descendants as well. The outcome of interest is a binary variable indicating whether or not an individual resides in a different state than in the previous wave. Thus multiple moves between waves, which occur every two years, are not captured by the data. Health measures, which are the control variables of interest, are collected for heads of households and their spouses, so the analyses are limited to only these individuals. For most of its history, the PSID has collected information about self-rated health status, a classic health measures used in the social sciences. However, starting in 1999, the survey also began to include more detailed information about disability and chronic conditions, which is the reason the analysis starts at this point. Activities of daily living (ADL) were reported, and respondents were also asked to state if they had ever been diagnosed with a host of chronic and acute conditions. All the health measures are summarized in Table 3.1.

The study of the question at hand will use health measures before migration, so as not to confound the analysis with health changes that may occur as a part or result of the migration process. In the PSID, self-rated health and ADLs are reported only at the time of interview, and the timing of moves and diagnoses of health conditions are also updated only at this time. Therefore, in order to streamline the analysis and to reduce any error in recalling the timing of events, all variables are treated as discrete.

Due to the clustered nature of the longitudinal data, and the fact that the outcome is repeatable, it is not possible to conduct the analysis simply through a failure model using logistic regression. There are several modeling options, but given the nature of the data, the analysis is conducted via a marginal model, an extension of the generalized linear model that accounts for the lack of independence among repeated measures in longitudinal data. In addition, the calculations are created by using generalized estimating equations (GEE). This analytical strategy is about as precise or efficient as maximum likelihood estimation, and retains consistency even in the face of misspecification of within-subject associations among repeated measures. This model is able to "generalize and extend the usual likelihood equations for a generalized linear model for a univariate response by incorporating the covariance matrix of the vector of responses" (Fitzmaurice, Laird and Ware 2011). Variances are calculated empirically, and the models specify an unstructured variance correlation matrix.

The health measures are handled in the following manner. Self-rated health remains a categorical variable with five possible responses, corresponding to the five possible answers that respondents can give when asked about their health. Good health, the middle category, is treated as the reference group. ADLs and health conditions are indicator variables, where individuals either do or do not have trouble with a particular function or disease. In the regressions, disability and health conditions are treated as sums of the number of reported problems for each individual. In addition to the health measures, the models control for sociodemographic characteristics that could be related to migration. Age in years and sex appear in every model. Race is included, coded as white, black, or other. The models for the entire sample included a married indicator. Finally, a categorical variable for education is incorporated as a measure of socioeconomic status, and is coded as less than high school, high school graduate, some college experience, and college graduate and above. Sample sizes refer to individuals or couples who are included in each regression, while the reported number of observations instead reflects the number of intervals used in each analysis that were contributed by those in the sample size.

There are two broad sets of models included below. The first regressions are for the full sample, and results are reported by sex. The subsequent set is for the subsample of married people, in which self-rated health is dropped from consideration. This is due to the fact that the reporting of data for heads and spouses in the PSID comes from only one individual. Having a person report the general health status of their spouse may nullify the rationale in using the measure, which allows for a subjective consideration of health that is not possible to measure from an outside perspective.

Since married couples generally move together, those migration events are included in both the regressions for men and for women. This masks the fact that those events are counted twice. Therefore, the regressions for the married subsample are combined so that both the husband's and the wife's characteristics appear in the same regressions, meaning that each covariate has a husband and wife version. The exception is the race variables, which are combined into indicators based on the racial composition of the individuals in any specific union. These models are centered on the couple. In order to examine age gradients in selection, the models for the full sample are reported for people less than age 60 and then for individuals above that threshold. For the married models, the age cut is based on husband's age.

Results

Table 3.2 displays the odds ratios from the results for the entire sample, with separate sections by sex and age. Sample sizes vary for each regression, but even the smallest regression contains 919 individual contributing 3260 intervals. On the large end, regressions contain over 6000 individuals contributing upwards of 25,000 intervals. For individuals under the age of sixty, increases in age are associated with a significant decrease in the propensity to move for both sexes. Further, younger men and women are less likely to move if they are black, as compared to whites, a result that weakens for older individuals. Married people at younger ages are also generally less mobile, an effect which again weakens in older age. Individuals aged less than sixty with at least some college experience have higher odds of moving than those with only high school diplomas, and for women this effect intensifies in older age. Finally, and most importantly, the results for the health variables are relatively sparse. Self-rated health does not significantly impact propensity to migrate in any of the regressions, and theory on health selection cannot even explain why the odds ratios are above or below one. Disability and health conditions were tested in individual models, but the results match almost identically the version with both incorporated, which is what is included in the table. The only significant association is disabled older men, who have significantly elevated odds of migration.

In order to examine health selective migration in the context of marriage we turn our attention to couples instead of the individual. In Table 3.3 results are reported for the subsample of married individuals, with covariates and sample sizes now reflecting couples. The sample size for younger couples is 10,220 contributing 39,974 intervals, whereas the regressions for older couples are roughly one quarter that size. Self-rated health is no longer considered, and the regressions are split based on husband's age. Couples in which the husband is less than sixty years of age are significantly less likely to migrate as they age. The strong race effects for blacks do not appear in this set of regressions, but older couples who are not both white or both black have significantly elevated odds of migration. As far as education is concerned, there are slight differences between husbands and wives. Younger couples are significantly more likely to migrate if either the husband or the wife has at least a college degree, and for husbands that effect extends to even some college experience. The effect is pertinent to older couples as well, but only in terms of wife's education, in which case couples where the wife has at least some college experience are significantly more likely to migrate. Lastly, older couples in which the husband did not complete high school are significantly less likely to migrate.

This set of regressions includes disability and health conditions in isolation, and then in combination. The results are simple to interpret. Younger couples are significantly more likely to migrate if the wife is in poor health, regardless of whether health is measured by disability or health conditions. However, when the two are included in a single regression together, only health conditions remain significant. As for older couples, migration increases when husbands are in poor health, but only when measured by disability. This result appears when disability is included on its own, and remains when both measures of health are present in a single regression.

Discussion

To fully understand the health selectivity of internal migration in the United States is to better comprehend a complex process that impacts social, economic, cultural, and political realities on a local level. Any new information gathered from this research might therefore improve he formulation of public policy. However, few previous studies have been able to directly measure health selection among internal migrants in the United States. The results of these analyses show that self-rated health is not related to the probability of migration, whereas increasing levels of disability elevates the odds of migration for older men. No statistically significant evidence appears here to support the assertion that migrants of either sex at any age are selected on a summary measure of health conditions.

When the analysis is restricted only to married couples, younger couples in which the wife is unhealthy are significantly more likely to migrate, especially so when measured using health conditions. Older couples are significantly more likely to migrate as the husband becomes more disabled. These results provide a new perspective on the roles of men's and women's health on migration. Taken together, the findings from this paper represent the most recent and broad examination of health selective internal migration in the United States, on a sample that was constructed to be representative of the country.

Though this inquiry may be the most recent and exhaustive for the United States, Halliday and Kimmitt (2008) also explored similar questions using the same data, though for an earlier time period. Their analysis discovered significant health selective migration using self-rated health status, and the authors are not alone in reporting such results (Chen 2011; Jongeneel-Grimen et al. 2013). Yet no significant positive or negative selection appears in this analysis when self-rated health was the health measure included in the regressions, so it is natural to inquire how this difference between studies may have arisen. There are several possibilities. First, the structure of the PSID is such that the two samples are not identical, though some of the same individuals were included in both time periods. Others dropped out due to death and sample loss, and younger generations of the families were added to the PSID. If differences in age distribution are probable between the two samples, and health selective effects are age graded, the composition of the sample could be impactful. In addition, the passage of time is significantly related to the probability to migrate in one regression from this study, and other studies have reported changes in the strength of health selection over time (Borrell and Lancet 2012; Tong and Piotrowski 2012). Thus this exact sample, if measured at a slightly different time period, may have replicated some of the effects of the earlier investigation. Alternatively, there may simply have been declines in health selectivity between the two time periods. Finally, it may just be the case that self-reported measures of health are less consistent than more concrete measures. Indeed, others have suggested that "reliance on self-assessments of health alone may yield a misleading picture of the health of migrants relative to those who do not move" (Rubalcava et al. 2008). In addition, if there are factors influencing both perception of health and propensity to migrate, then confounding is possible. "For example, those who migrate may have a more optimistic outlook on life, a personality characteristic that is perhaps related to the perception and propensity to report poor self-reported health" (Connolly and O'Reilly 2007). Ultimately, a combination of the preceding factors may have influenced the change in findings for self-rated health.

One of the main conclusions to be drawn from this study is that, in general, when conducting analyses of health selective migration, some health measures are more suitable than others. The variability in results using self-rated health is simply one illustration of this. Of the health measures studied here, health selection on disability seems to be the most significant, especially for men. Selectivity on specific health measures could be reflective of how salient the measures themselves are to the process of migration. Lu (2008) hypothesized that selection would be particularly strong on chronic and severe conditions, as they would relate more directly to one's mobility and ability to adapt after a migration. This study generally supports this assertion. In addition, disability may be an important marker for inability to continue working. This may explain why it is a particularly relevant measure for men in this study, particularly if they are older, as the men of older generations were historically more likely to be the primary income generator of their households.

The final primary conclusion one makes when considering these analyses regards health selectivity as it relates to marriage. From the results, it appears that married couples are more strongly selected on the wife's health in younger ages, but at older ages this reverses to the husband's health. The findings for married couples are especially noteworthy for a couple reasons. A large majority of individuals in the PSID are married, so in general it should be expected that they move together. By conducting the analysis on the couple level, we obtain the clearest picture of health selection in this sample. This is an important analytical decision that can be incorporated into future research in this area. Further, the analysis for couples also makes it clear that a person's own health is not always the factor that leads to a migration. This helps to explain why the results for individuals were relatively weaker. Given the issues with self-rated health previously mentioned, these results may be considered a baseline for future exploration of how marriage now relates to health selective migration.

This is an interesting avenue for future studies, considering the long developing changes to marriage patterns, as well as the full incorporation of women into the workforce throughout all job sectors and educational levels. Health selection on husbands for couples with older husbands makes sense given the historical record of tied migration in the United States. The fact that selection switches to the wife in younger couples perhaps signals that there is a new story to be told, which might involve women balancing the demands of both work and the home, in addition to other possible mechanisms.

The overall pattern of age selectivity based on the chosen health measures revealed here fits well in the literature. Measuring health with ADLs, Lu (2008) found a positive association between poorer health and likelihood of migration for older individuals. Reasons proposed to explain this relationship often center around moving to seek better medical care, or perhaps the support of relatives who could provide general support and also assist in care. However, Lu (2008) also found positive health selective migration among younger individuals when measuring with disability, whereas the present analysis did not uncover such a relationship. As for health conditions, Jongeneel-Grimen et al. (2013) found that the oldest migrants were relatively unhealthy if measuring by the presence of one or more long term illnesses. Larson et al. (2004) showed that moves were linked to having numerous physical symptoms, the presence of at least two chronic diseases, as well as poor mental health.

Beyond the health measures themselves, there are several factors that might explain the degree to which the results from this paper align with those from the other research in the literature on health selective internal migration. First, as mentioned previously, the choices made on which geographic units are used to measure migration can exert a powerful influence on the results. Similarly, it is difficult to determine the appropriate time to use as the migration cutoff in any study, even with full information. It is possible that significant positive health selection would appear if a longer duration were used to define migrants in this study. Alternatively, the rate of internal migration has been slowing over the past several decades, a secular change that is not restricted to particular demographics or geographies (Molloy, Smith and Wozniak 2011). The data from Halliday and Kimmitt (2008) come right at the beginning of this trend, whereas the data here are the most recent. It is possible that this continuing downturn is a result of changes in the nature and purpose of internal migration, and it is an open question whether or not health selection among migrants will change as well. Molloy et al. (2011) also address cyclical housing issues as it may relate to migration patterns. Though they state that the Great Recession cannot be the main driver of the observed migration patterns in the US data, the Great Recession began less than half way through the time period under consideration here. Comparability to other studies on previous migration will surely be affected by such a dramatic economic downturn, especially since housing was a large part of the process. However, the exact nature of potential effects is beyond the scope of this paper.

Though there are many strengths to the analysis, a few limitations should be noted. Non-random attrition from the survey could bias the results. This is a concern in any panel data set. In addition, residence is only assessed during surveys, which occur every two years. Short term moves of a circular nature, as well as multiple moves in a short time period, are thus not captured by the outcome variable, somewhat limiting the generalizability of the findings. Finally, even among the moves detected by the data, distance travelled is incredibly variable, including between states that border each other. The intensity of a move from northern California into Arizona is probably different than from Delaware into New Jersey, but here they are treated identically.

Conclusion

The results show that internal migrants in the United States are not selected on self-rated health, but instead on disability and health conditions. This arises in large part because self-reports of health are generally less valuable than more objective measures when examining health selective migration. Men are negatively selected at older ages, where disability is particularly salient. Married couples are selected for migration on the health of both partners, but at younger ages it is the wife's health that matters, whereas for older couples the husband's health predominates. The historical prevalence of tied migration among older couples drives those results, whereas selection on the wife's health in younger couples shows that this trend has weakened, and that a new story must be told. Moreover, these results demonstrate the value of examining couples separately whenever possible, as it is often the case that one's own health is not the determining factor in migration.

This analysis provides a solid foundation upon which future research should build. The lack of any positive health selection, especially among younger migrants, needs further explanation, particularly as it might pertain to secular changes in internal migration in the United States. Further, a more in depth analysis might be able to discriminate between moves over the life course, investigating health selection that occurs, for example, during the first residential move. In light of the fact that older people were selected on their health, it would be interesting to explore their destinations; perhaps health selection is particularly strong amongst migrants who are moving to be with family. Finally, the PSID itself has some interesting data about the moves in the sample, including self-reported reasons for the move and individual expectations about the likelihood of future migration. All of this information could be utilized to better characterize health selection as it occurs in specific segments of the population.

References

Abraido-Lanza, A.F., B.P. Dohrenwend, D.S. Ng-Mak, and J.B. Turner. 1999. "The Latino Mortality Paradox: A Test of the "Salmon Bias" and Healthy Migrant Hypotheses." American Journal of Public Health 89(10):1543-1548.

Bentham, G. 1988. "Migration and Morbidity - Implications for Geographical Studies of Disease." Social Science & Medicine 26(1):49-54.

Borrell, L.N.and E.A. Lancet. 2012. "Race/Ethnicity and All-Cause Mortality in US Adults: Revisiting the Hispanic Paradox." American Journal of Public Health 102(5):836-843.

Boyle, P., Duke-Williams, Oliver. 2004. "Does Migration Exaggerate the Relationship between Deprivation and Self-Reported Limiting Long-Term Illness?" Pp. 129-148 in The Geography of Health Inequalities in the Developed World: Views from Britain and North America, edited by P. Boyle, Sarah Curtis, Elspeth Graham, Eric Moore. Burlington, VT: Ashgate Publishing Company.

Boyle, P., P. Norman, and P. Rees. 2002. "Does Migration Exaggerate the Relationship Between Deprivation and Limiting Long-Term Illness? A Scottish Analysis." Social Science and Medicine 55(1):21-31.

Brimblecombe, N., D. Dorling, and M. Shaw. 1999. "Mortality and Migration in Britain, First Results from the British Household Panel Survey." Social Science and Medicine 49(7):981-988.

Chen, J. 2011. "Internal Migration and Health: Re-Examining the Healthy Migrant Phenomenon in China." Social Science & Medicine 72(8):1294-1301.

Connolly, S.and D. O'Reilly. 2007. "The Contribution of Migration to Changes in the Distribution of Health over Time: Five-Year Follow-up Study in Northern Ireland." Social Science & Medicine 65(5):1004-1011.

Connolly, S., D. O'Reilly, and M. Rosato. 2007. "Increasing Inequalities in Health: Is It an Artefact Caused by the Selective Movement of People?" Social Science & Medicine 64(10):2008-2015.

Dunn, J.R., P. Schaub, and N.A. Ross. 2007. "Unpacking Income Inequality and Population Health - the Peculiar Absence of Geography." Canadian Journal of Public Health-Revue Canadienne De Sante Publique 98:S10-S17.

Ellis, M. 2012. "Reinventing Us Internal Migration Studies in the Age of International Migration." Population Space and Place 18(2):196-208.

Findley, S.E. 1988. "The Directionality and Age Selectivity of the Health-Migration Relation - Evidence from Sequences of Disability and Mobility in the United-States." International Migration Review 22(3):4-29.

Fitzmaurice, G.M., N. Laird, and J. Ware. 2011. Applied Longitudinal Analysis. Hoboken, NJ: Wiley and Sons.

Halliday, T.J.and M.C. Kimmitt. 2008. "Selective Migration and Health in the USA, 1984-93." Population Studies 62(3):321-334.

Hu, X.U.J., S. Cook, and M.A. Salazar. 2008. "Internal Migration and Health in China." Lancet 372(9651):1717-1719.

Jasso, G., Massey, D. S., Rosenzweig, R. S., & Smith, J. P. 2004. "Immigrant Health, Selectivity, and Acculturation." Pp. 227-266 in Critical Perspectives on Racial and Ethnic Differences in Health in Late Life, edited by R.A.B. N.B. Anderson, & B. Cohen. Washington, DC: National Academy Press.

Jongeneel-Grimen, B., M. Droomers, K. Stronks, J.A.M. van Oers, and A.E. Kunst. 2013. "Migration and Geographical Inequalities in Health in the Netherlands: An Investigation of Age Patterns." International Journal of Public Health 58(6):845-854.

King, R.and R. Skeldon. 2010. "'Mind the Gap!' Integrating Approaches to Internal and International Migration." Journal of Ethnic and Migration Studies 36(10):1619-646.

Larson, A., M. Bell, and A.F. Young. 2004. "Clarifying the Relationships between Health and Residential Mobility." Social Science & Medicine 59(10):2149-2160.

Lu, Y. 2008. "Test of the 'Healthy Migrant Hypothesis': A Longitudinal Analysis of Health Selectivity of Internal Migration in Indonesia." Social Science & Medicine 67(8):1331-1339.

Lu, Y.and L.J. Qin. 2014. "Healthy Migrant and Salmon Bias Hypotheses: A Study of Health and Internal Migration in China." Social Science & Medicine 102:41-48.

Markides, K.S.and J. Coreil. 1986. "The Health of Hispanics in the Southwestern United States - an Epidemiologic Paradox." Public Health Reports 101(3):253-265.

Markides, K.S.and K. Eschbach. 2005. "Aging, Migration, and Mortality: Current Status of Research on the Hispanic Paradox." Journals of Gerontology Series B-Psychological Sciences and Social Sciences 60:68-75.

Molloy, R., C.L. Smith, and A. Wozniak. 2011. "Internal Migration in the United States." Journal of Economic Perspectives 25(3):173-196.

Palloni, A.and E. Arias. 2004. "Paradox Lost: Explaining the Hispanic Adult Mortality Advantage." Demography 41(3):385-415.

Palloni, A.and J.D. Morenoff. 2001. "Interpreting the Paradoxical in the Hispanic Paradox - Demographic and Epidemiologic Approaches." Pp. 140-174 in Population Health and Aging: Strengthening the Dialogue between Epidemiology and Demography, edited by M. Weinstein, A.I. Hermalin, and M.A. Stoto. New York: New York Acad Sciences.

Popham, F., P.J. Boyle, D. O'Reilly, and A.H. Leyland. 2011. "Selective Internal Migration. Does It Explain Glasgow's Worsening Mortality Record?" Health & Place 17(6):1212-1217.

Qi, Y.Q.and J.L. Niu. 2013. "Health Selection Effects in China's Internal Migration." Asian Population Studies 9(2):142-155.

Rubalcava, L.N., G.M. Teruel, D. Thomas, and N. Goldman. 2008. "The Healthy Migrant Effect: New Findings from the Mexican Family Life Survey." American Journal of Public Health 98(1):78-84.

Tinghog, P., J. Carstensen, G. Kaati, S. Edvinsson, M. Sjostrom, and L.O. Bygren. 2011. "Migration and Mortality Trajectories: A Study of Individuals Born in the Rural Community of Overkalix, Sweden." Social Science & Medicine 73(5):744-751.

Tong, Y.Y.and M. Piotrowski. 2012. "Migration and Health Selectivity in the Context of Internal Migration in China, 1997-2009." Population Research and Policy Review 31(4):497-543.

Turra, C.M.and I.T. Elo. 2008. "The Impact of Salmon Bias on the Hispanic MortalityAdvantage: New Evidence from Social Security Data." Population Research and Policy Review 27(5):515-530.

Table 3.1 Health Measures from the Panel Study of Income Dynamics, 1999-2013

Would you say your health in general is:	Excellent Very good Good Fair Poor
Because of a health or physical problem, do you have any difficulty:	Bathing Dressing Eating Getting in or out of a bed or chair Walking Getting outside Using the toilet, including getting to the toilet
Has a doctor ever told you that you have or had:	Stroke High blood pressure or hypertension Diabetes or high blood sugar Cancer or a malignant tumor, excluding skin cancer Chronic lung disease such as bronchitis or emphysema A heart attack Coronary heart disease, angina, or congestive heart failure Any emotional, nervous, or psychiatric problems Arthritis or rheumatism Asthma Permanent loss of memory or mental ability

Table 3.2 Odds Ratios for Interstate Migration, Full Sample

	Men				Women			
	Age<60		Age≥60		Age<60		Age≥60	
Time	1.004	0.99	1.02	0.97	0.995	0.99	.90*	0.94
	(.97, 1.04)	(.96, 1.02)	(.93, 1.12)	(.90, 1.05)	(.97, 1.03)	(.96, 1.01)	(.81, .99)	(.87, 1.02)
Age	.95***	.95***	1.02	1.01	.95***	.95***	1.01	1.01
	(.94, .96)	(.94, .953)	(.99, 1.05)	(.99, 1.04)	(.94, .96)	(.94, .955	(.98, 1.04)	(.99, 1.04)
Race								
White	-	-	-	-	-		-	-
Black	.76**	.85*	0.57	.40*	.70***	.73***	0.57	.55*
	(.62, .93)	(.72, .996)	(.22, 1.50)	(.18, .87)	(.60, .83)	(.63, .84)	(.30, 1.10)	(.30, .99)
Other	0.70	0.87	1.26	1.13	0.83	0.8	1.01	1.55
	(.48, 1.01)	(.67, 1.14)	(.53, 3.02)	(.55, 2.34)	(.61, 1.12)	(.62, 1.03)	(.37, 2.78)	(.77, 3.12)
Married	.71***	.73***	0.68	.62*	.78***	.79***	1.31	1.29
	(.61, .84)	(.64, .84)	(.42, 1.09)	(.40, .96)	(.68, .90)	(.70, .90)	(.82, 2.11)	(.88, 1.90)
Education								
Less than high school	1.05	0.97	1.14	0.85	0.92	0.96	1.42	1.22
	(.79, 1.38)	(.77, 1.21)	(.53, 2.45)	(.47, 1.54)	(.72, 1.16)	(.77, 1.18)	(.70, 2.88)	(.71, 2.11)
High school graduate	-	-	-	-	-	-	-	-
Some college	1.30*	1.41***	1.58	1.12	1.34**	1.41***	2.83**	2.31***
	(1.04, 1.62)	(1.18, 1.68)	(.78, 3.18)	(.62, 2.02)	(1.12, 1.61)	(1.20, 1.66)	(1.48, 5.43)	(1.41, 3.79)
College graduate plus	1.84***	2.16***	1.92*	1.57	1.89***	1.96***	4.05***	2.91***
	(1.48, 2.30)	(1.82, 2.57)	(1.02, 3.61)	(.95, 2.58)	(1.56, 2.29)	(1.67, 2.31)	(2.14, 7.7)	(1.82, 4.66)
Self-rated health								
Excellent	1.08		0.96		0.95		0.94	
	(.89, 1.32)		(.52, 1.76)		(.80, 1.13)		(.45, 1.98)	
Very good	1.01		0.95		0.87		0.85	
	(.85, 1.21)		(.60, 1.52)		(.75, 1.01)		(.50, 1.43)	
Good	-				-		-	
Fair	0.82		1.06		0.99		1.13	
	(.60, 1.14)		(.60, 1.89)		(.79, 1.23)		(.64, 2.00)	
Poor	0.98		0.8		0.81		1.1	
	(.52, 1.85)		(.31, 2.07)		(.48, 1.35)		(.50, 2.43)	
Disability		0.98		1.19**		0.99		1.10
		(.84, 1.14)		(1.05, 1.34)		(.88, 1.11)		(.98, 1.24)
Conditions		1.02		1.02		1.05		0.96
		(.95, 1.10)		(.90, 1.14)		(.98, 1.11)		(.85, 1.09)
Sample Size	4159	6973	919	1515	6114	8113	1268	1769
Observations	16208	28853	3260	5421	25896	35584	4659	6833

*p<.05 **p<.01 ***p<.001 95% CI in parentheses

Source: PSID 1999-2013

Table 3.3 Odds Ratios for Interstate Migration, Married Couples

	Husband age <60	Husband age ≥60

Time	0.98	0.97	0.97	0.94	0.94	0.94
	(.95, 1.003)	(.95, 1.001)	(.95, 1.001)	(.88, 1.01)	(.88, 1.02)	(.88, 1.01)
Husband Age	0.95***	.95***	.95***	1.01	1.02	1.01
	(.94, .97)	(.93, .97)	(.93, .97)	(.98, 1.05)	(.99, 1.05)	(.98, 1.05)
Wife Age	0.99	0.99	0.99	1.03	1.02	1.03
	(.97, 1.01)	(.97, 1.005)	(.97, 1.005)	(.997, 1.06)	(.99, 1.05)	(.996, 1.06)
Black Couple	0.95	0.94	0.94	0.54	0.57	0.53
	(.80, 1.12)	(.80, 1.11)	(.80, 1.11)	(.26, 1.09)	(.28, 1.15)	(.26, 1.08)
Mixed/Other Couple	1.14	1.15	1.15	1.82*	1.88**	1.81*
	(.95, 1.36)	(.96, 1.37)	(.96, 1.37)	(1.14, 2.92)	(1.18, 2.99)	(1.13, 2.89)
Husband Education						
Less than high school	1.08	1.07	1.07	.45**	.46*	.45**
	(.85, 1.37)	(.84, 1.35)	(.84, 1.35)	(.24, .82)	(.25, .84)	(.24, .82)
High school graduate	-	-	-	-	-	-
Some college	1.53***	1.53***	1.53***	0.84	0.83	0.84
	(1.28, 1.82)	(1.28, 1.82)	(1.28, 1.82)	(.50, 1.39)	(.50, 1.38)	(.50, 1.39)
College graduate plus	2.37***	2.39***	2.39***	1.08	1.1	1.09
	(1.97, 2.85)	(1.98, 2.87)	(1.99, 2.87)	(.69, 1.68)	(.70, 1.70)	(.70, 1.70)
Wife Education						
Less than high school	0.86	0.85	0.85	1.22	1.28	1.23
	(.66, 1.12)	(.65, 1.10)	(.65, 1.10)	(.67, 2.22)	(.72, 2.30)	(.68, 2.23)
High school graduate	-	-	-	-	-	-
Some college	1.14	1.14	1.14	2.50***	2.54***	2.51***
	(.96, 1.35)	(.96, 1.35)	(.96, 1.35)	(1.61, 3.90)	(1.63, 3.95)	(1.61, 3.91)
College graduate plus	1.27**	1.29**	1.29**	2.36***	2.39***	2.37***
	(1.06, 1.53)	(1.07, 1.54)	(1.07, 1.54)	(1.49, 3.73)	(1.51, 3.78)	(1.50, 3.76)
Husband Disability	0.97		0.93	1.18**		1.19**
	(.82, 1.15)		(.79, 1.11)	(1.04, 1.33)		(1.04, 1.35)
Wife Disability	1.12*		1.08	1.02		1.02
	(1.00, 1.25)		(.96, 1.21)	(.88, 1.18)		(.88, 1.18)
Husband Conditions		1.06	1.06		1.004	0.97
		(.98, 1.14)	(.99, 1.15)		(.90, 1.11)	(.87, 1.08)
Wife Conditions		1.09*	1.08*		1.03	1.03
		(1.02, 1.16)	(1.006, 1.16)		(.92, 1.16)	(.91, 1.16)
Sample Size		10220			2472	
Observations		39974			8510	
*p<.05 **p<.01 ***p<.001		95% CI in parentheses			Age splits based on husband age	

Source: PSID 1999-2013 *Note*: Individuals are able to contribute to multiple marriages over the observation period. Mixed/other couple refers to any couples that are not both white or both black.

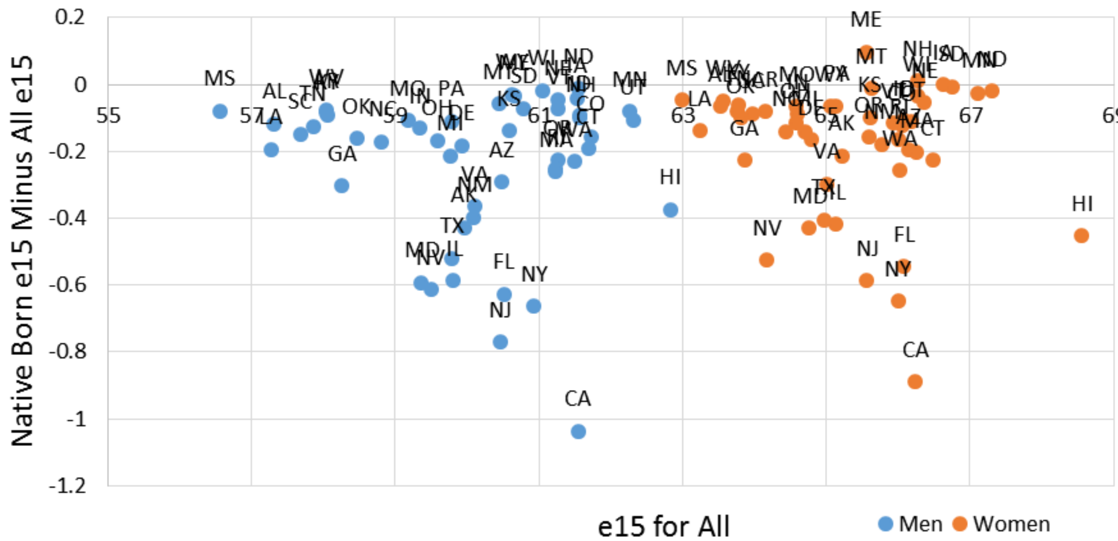
APPENDIX

Table A1.1 Missing Data Imputed via Multiple Imputation for the Malawi Longitudinal Study of Families in Health, 2004-2010

Variable	Cases Imputed	Maximum Number of Records Used in Any Regression	Percent Missing
Age	24	9847	0.24%
Self-rated health	1711	9847	17.38%
Wealth quintile	2032	9847	20.64%
Education	1146	9847	11.64%
HIV Status	2433	9847	24.71%
Ethnicity	59	9847	0.60%
Marital Status	529	9847	5.37%
One year mortality, 2006	646	3381	19.11%
Five year mortality	666	3381	19.70%
One year mortality, 2008	932	3524	26.45%

Note: In Stata, the imputation was carried out using chained equations, specifying the augment option, and producing ten imputed data sets. Gender, region, and religion did not need to be imputed. Ethnicity (recoded into a reduced number of categories to facilitate imputation) and marital status were imputed as categorical variables, and others were treated as continuous. After imputation, the categorical variables that were imputed as continuous (for example, education) were then rounded to the nearest whole number for use in the analysis. Though the model requires data in long format, imputation was only possible while the data was still in wide format, with one record per person. Thus, during imputation missing values are filled in even when they are definitively not needed, such as a survey following a recorded death. However, using survey outcome variables originating in the data key for the MLSFH, all unnecessary values such as these are deleted after reshaping the data. Fully imputed records were only kept if that record occurred between two records for which the survey outcome variable was something other than missing.

Figure A2.1 Changes in Life Expectancy at Age 15 (e15) Due to the Removal of the Foreign Born, 2000



Source: Life expectancy calculations using data from the U.S. Census and National Vital Statistics System

Table A2.1 State Level Inequality in Life Expectancy at Age 15 Due to the Removal of the Foreign Born, 2000

	Men			Women		
Measure of Inequality	All	Native Born	Percent	All	Native Born	Percent
Gini coefficient	0.0121	0.0115	-4.4%	0.0083	0.0077	-6.8%
Theil index	0.000237	0.000217	-8.2%	0.000111	0.000097	-12.8%
Squared coefficient of variation	0.00047	0.00043	-8.1%	0.00022	0.00019	-12.6%
Mean logarithmic deviation	0.000238	0.000218	-8.3%	0.000111	0.000097	-12.9%

Source: Life expectancy calculations using data from the U.S. Census and National Vital Statistics System *Note*: Percent change reflects moving from the All to only Native Born value for life expectancy at age 15.

Table A2.2 Life Expectancy at Age 15 by Migrant Stream, Men, 2000

Residence	Inmigrant	Nonmover	Outmigrant	Nativity

Alabama	57.20	58.95	56.55	56.83	56.58
Alaska	59.53	60.95	55.75	53.39	54.89
Arizona	60.19	60.72	58.57	60.25	59.26
Arkansas	57.95	58.49	57.61	57.61	57.58
California	60.51	60.44	61.50	62.01	61.64
Colorado	61.57	62.10	60.59	60.89	60.72
Connecticut	61.50	62.23	60.97	61.85	61.34
Delaware	59.74	60.17	59.02	60.90	59.83
Washington DC	53.40	57.60	47.32	60.29	58.39
Florida	59.88	60.31	58.44	60.16	59.02
Georgia	57.96	59.95	56.57	57.32	56.78
Hawaii	62.45	63.30	61.92	63.28	62.41
Idaho	61.47	61.51	61.39	61.69	61.55
Illinois	59.22	59.15	59.36	60.82	59.98
Indiana	59.21	59.09	59.34	60.61	59.81
Iowa	61.48	60.92	61.64	61.88	61.73
Kansas	60.44	60.16	60.60	60.89	60.73
Kentucky	57.97	58.93	57.65	58.61	58.03
Louisiana	57.08	57.95	56.86	58.17	57.27
Maine	60.62	61.46	60.22	61.53	60.72
Maryland	58.77	59.99	57.49	60.30	58.53
Massachusetts	60.97	62.65	60.43	61.67	60.95
Michigan	59.55	58.75	59.99	61.26	60.42
Minnesota	62.18	61.96	62.27	62.27	62.25
Mississippi	56.48	57.57	56.05	55.96	55.89
Missouri	59.07	59.14	59.03	59.99	59.38
Montana	60.39	60.65	60.30	61.11	60.69
Nebraska	61.21	60.68	61.44	61.67	61.55
Nevada	58.89	58.95	60.20	60.45	60.39
New Hampshire	61.48	61.91	60.75	61.11	60.94
New Jersey	59.69	60.35	59.28	61.40	60.23
New Mexico	59.70	60.92	58.46	59.61	58.98
New York	60.27	59.83	60.42	62.00	61.16
North Carolina	58.64	60.52	57.64	57.87	57.69
North Dakota	61.54	61.37	61.53	61.41	61.46
Ohio	59.42	59.07	59.68	60.97	60.16
Oklahoma	58.30	58.53	58.19	59.57	58.80
Oregon	61.04	60.73	61.91	61.80	61.90
Pennsylvania	59.67	59.82	59.61	61.35	60.24
Rhode Island	60.98	61.48	60.72	61.88	61.23
South Carolina	57.53	59.99	56.12	56.28	56.09
South Dakota	60.71	60.07	60.93	61.78	61.36
Tennessee	57.74	58.94	57.04	58.11	57.39
Texas	59.27	60.12	58.80	59.60	59.00
Utah	62.21	61.84	62.44	61.90	62.26
Vermont	61.20	62.60	59.98	61.46	60.71
Virginia	59.74	61.63	58.02	58.41	58.17
Washington	61.27	61.17	61.60	62.04	61.76
West Virginia	57.97	58.43	57.80	58.45	58.11
Wisconsin	61.03	60.50	61.22	62.15	61.51
Wyoming	60.59	60.81	59.98	61.47	61.07

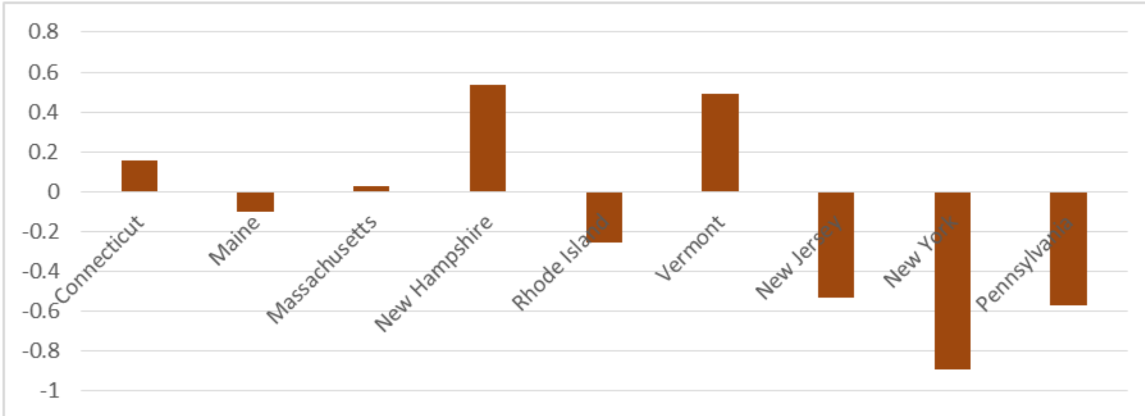
Source: Life expectancy calculations using data from the U.S. Census and National Vital Statistics System

Table A2.3 Life Expectancy at Age 15 by Migrant Stream, Women, 2000

	Residence	Inmigrant	Nonmover	Outmigrant	Nativity
Alabama	63.46	64.10	63.22	62.52	62.85
Alaska	65.00	65.52	63.92	57.88	60.10
Arizona	65.94	66.16	65.95	65.75	65.73
Arkansas	64.06	63.83	64.16	62.76	63.33
California	65.35	64.93	67.00	67.03	67.03
Colorado	65.93	66.09	65.65	65.01	65.26
Connecticut	66.25	66.45	66.08	66.72	66.35
Delaware	64.63	65.21	63.78	65.39	64.54
Washington DC	61.88	64.09	58.64	65.26	64.30
Florida	65.53	65.97	64.04	64.81	64.36
Georgia	63.64	64.49	62.98	62.49	62.73
Hawaii	68.10	67.62	68.12	67.34	68.03
Idaho	65.96	65.74	66.25	65.90	66.02
Illinois	64.72	64.10	65.05	65.64	65.30
Indiana	64.51	64.04	64.80	65.49	65.06
Iowa	66.62	65.59	66.92	66.62	66.74
Kansas	65.51	64.70	66.05	65.39	65.65
Kentucky	63.71	64.01	63.60	63.24	63.39
Louisiana	63.10	63.11	63.10	63.54	63.25
Maine	65.65	65.86	65.51	66.51	65.91
Maryland	64.33	64.73	63.88	65.05	64.32
Massachusetts	66.06	66.46	65.87	66.19	66.01
Michigan	64.56	63.43	65.22	66.49	65.66
Minnesota	67.09	66.51	67.31	66.81	67.09
Mississippi	62.94	62.98	62.85	61.76	62.17
Missouri	64.52	64.17	64.71	64.62	64.63
Montana	65.61	65.86	65.38	66.20	65.84
Nebraska	66.31	65.08	66.86	66.03	66.39
Nevada	63.65	63.63	65.80	65.99	65.88
New Hampshire	66.28	66.56	65.82	66.08	66.00
New Jersey	64.98	65.18	64.87	66.36	65.51
New Mexico	65.60	65.84	65.55	64.87	65.15
New York	65.35	64.70	65.57	66.46	65.97
North Carolina	64.30	64.81	64.00	63.64	63.87
North Dakota	67.28	66.45	67.50	66.47	66.80
Ohio	64.46	63.66	64.95	65.58	65.18
Oklahoma	63.71	63.40	63.94	63.90	63.82
Oregon	65.43	65.05	66.53	65.97	66.28
Pennsylvania	65.07	64.63	65.16	65.79	65.39
Rhode Island	65.84	66.30	65.59	66.01	65.82
South Carolina	63.88	64.74	63.35	62.10	62.78
South Dakota	66.74	66.55	66.79	66.31	66.43
Tennessee	63.69	63.91	63.52	63.12	63.34
Texas	64.56	64.76	64.44	64.24	64.39
Utah	66.07	65.42	66.49	65.84	66.22
Vermont	65.81	65.92	65.63	66.00	65.80
Virginia	64.71	65.30	64.14	63.57	63.91
Washington	65.76	65.48	66.40	65.79	66.16
West Virginia	63.51	63.31	63.56	63.42	63.46
Wisconsin	66.23	65.59	66.44	66.69	66.53
Wyoming	65.00	64.95	65.41	65.18	65.19

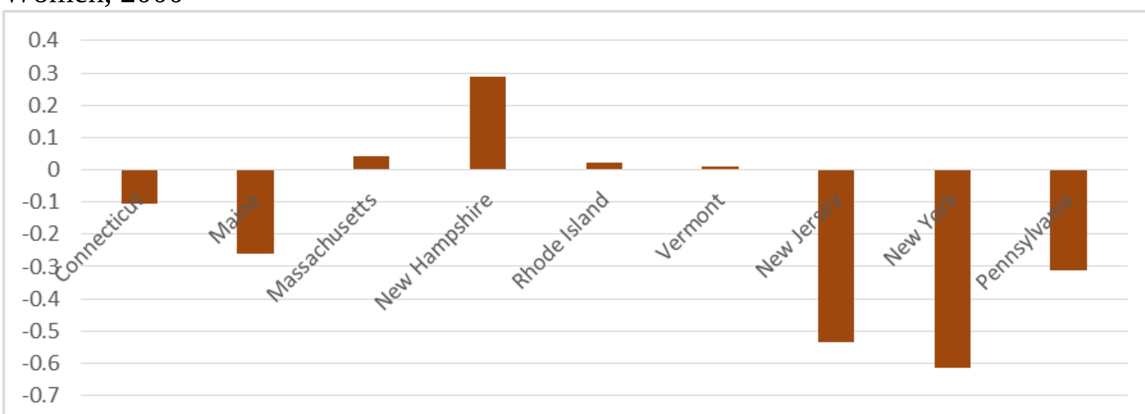
Source: Life expectancy calculations using data from the U.S. Census and National Vital Statistics System

Figure A2.2 Residence Minus Nativity Life Expectancy at Age 15, Northeast Region, Men, 2000



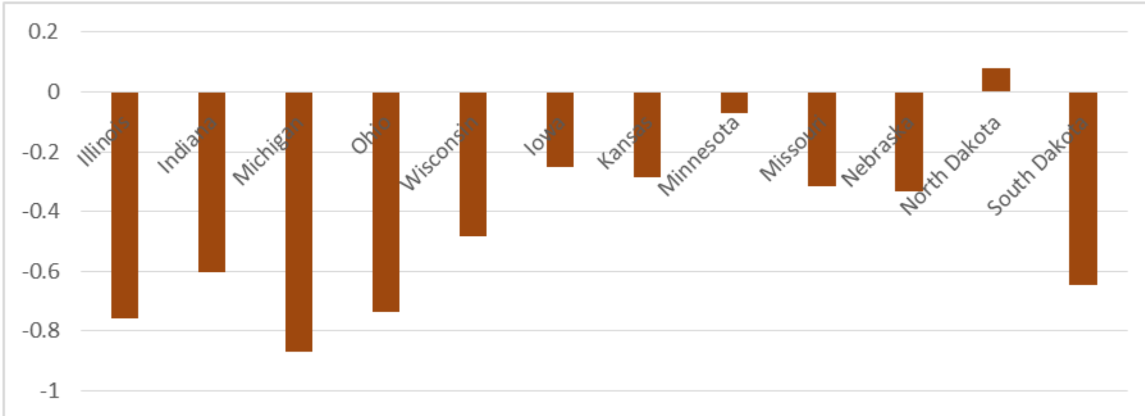
Source: Life expectancy calculations using data from the U.S. Census and National Vital Statistics System

Figure A2.3 Residence Minus Nativity Life Expectancy at Age 15, Northeast Region, Women, 2000



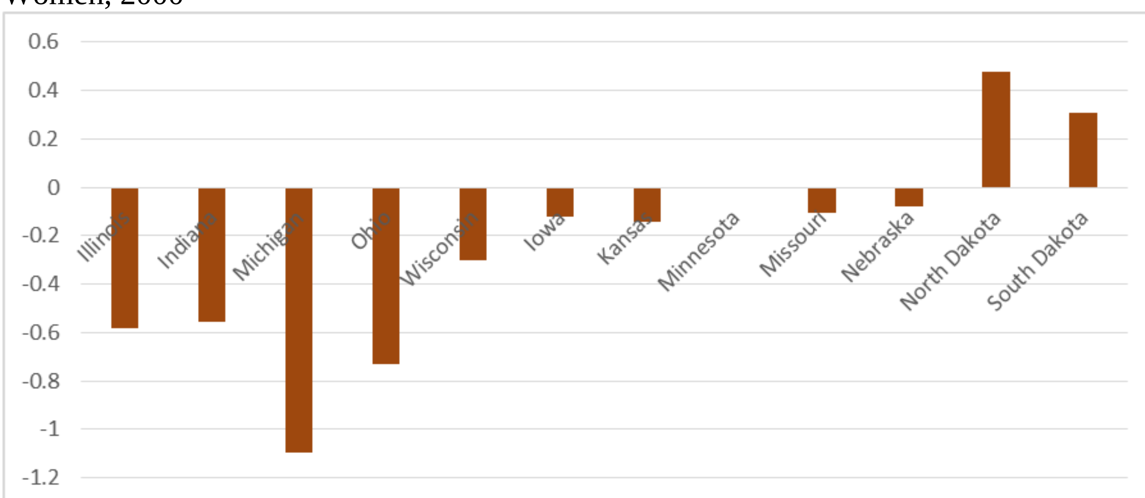
Source: Life expectancy calculations using data from the U.S. Census and National Vital Statistics System

Figure A2.4 Residence Minus Nativity Life Expectancy at Age 15, Midwest Region, Men, 2000



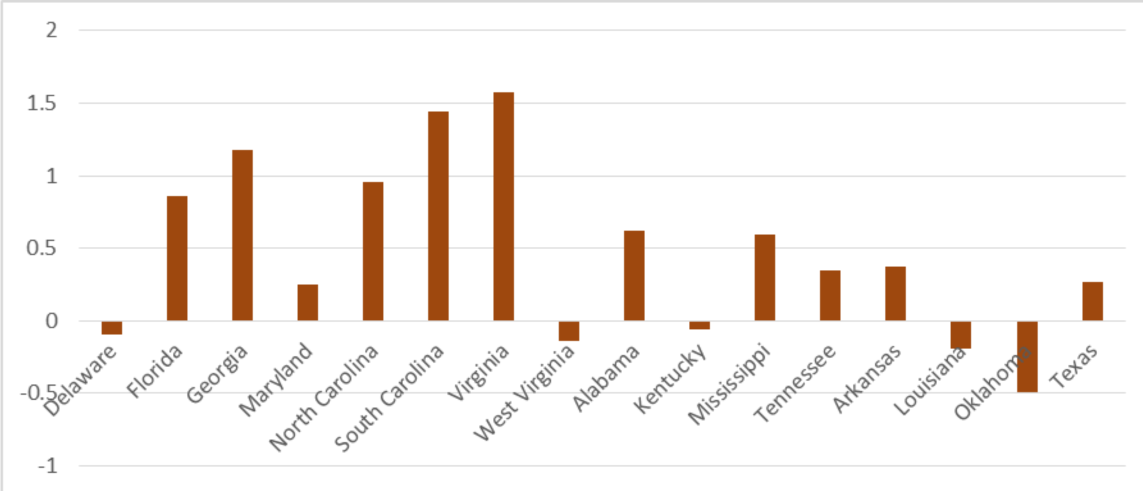
Source: Life expectancy calculations using data from the U.S. Census and National Vital Statistics System

Figure A2.5 Residence Minus Nativity Life Expectancy at Age 15, Midwest Region, Women, 2000



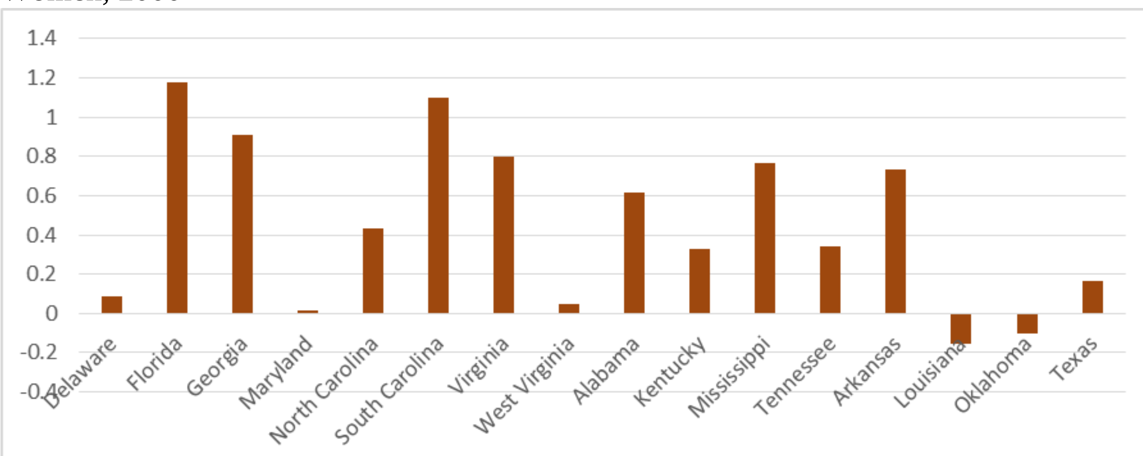
Source: Life expectancy calculations using data from the U.S. Census and National Vital Statistics System

Figure A2.6 Residence Minus Nativity Life Expectancy at Age 15, South Region, Men, 2000



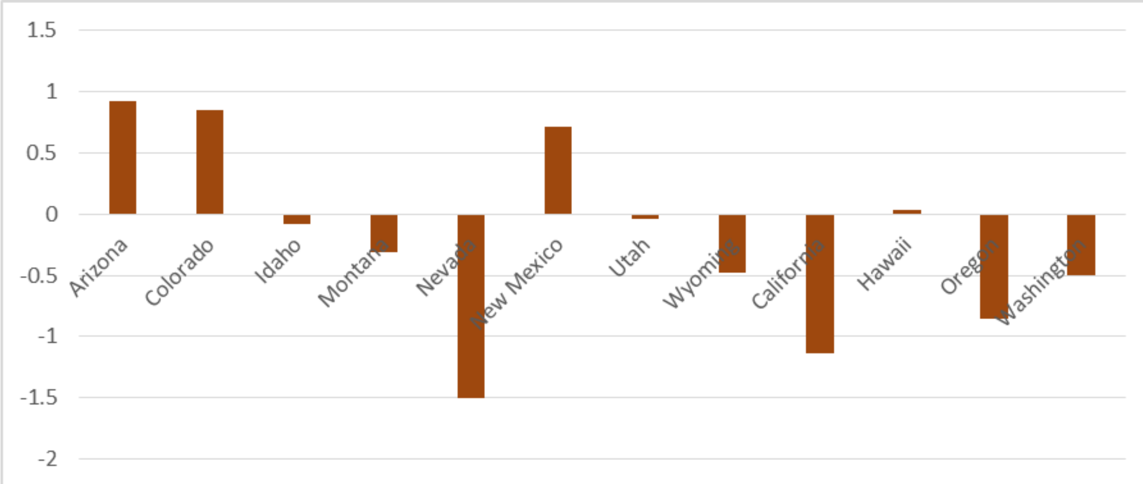
Source: Life expectancy calculations using data from the U.S. Census and National Vital Statistics System

Figure A2.7 Residence Minus Nativity Life Expectancy at Age 15, South Region, Women, 2000



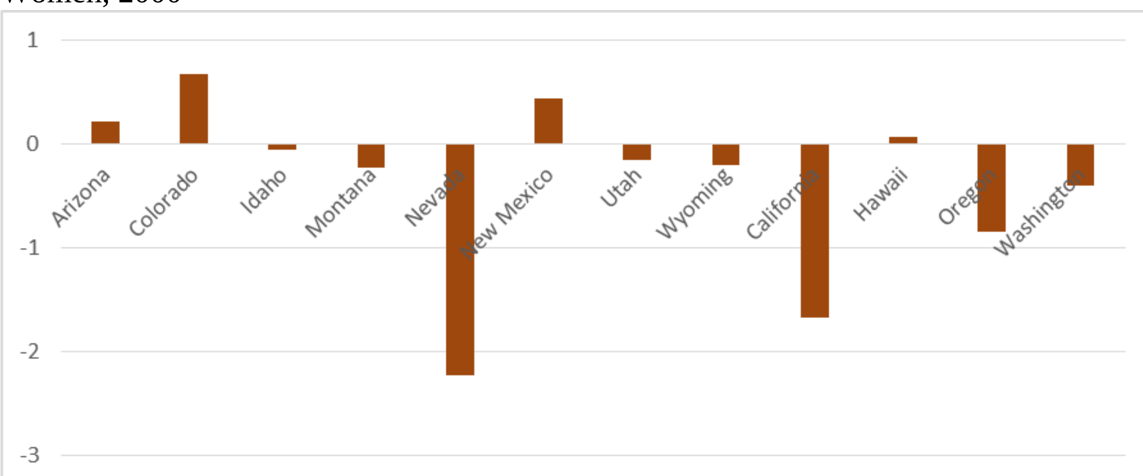
Source: Life expectancy calculations using data from the U.S. Census and National Vital Statistics System

Figure A2.8 Residence Minus Nativity Life Expectancy at Age 15, West Region, Men, 2000



Source: Life expectancy calculations using data from the U.S. Census and National Vital Statistics System

Figure A2.9 Residence Minus Nativity Life Expectancy at Age 15, West Region, Women, 2000



Source: Life expectancy calculations using data from the U.S. Census and National Vital Statistics System