Selective Migration and Health

by Timothy J. Halliday*,

Department of Economics and John A. Burns School of Medicine,

University of Hawaii at Manoa

and Michael Kimmitt,

Department of Economics,

University of Hawaii at Manoa

Working Paper No. 07-20

August 8, 2007

Abstract

We investigate the proposition that the health of migrants does not constitute a random sample of the health of the sending region using data from the Panel Study of Income Dynamics on internal migration within the United States. Panel data is crucial, as it enables us to observe geographic mobility as well as the health of the migrant prior to migration. We find that, for men and women below 60 years of age, a move from the middle to the bottom of the health distribution reduces mobility by 32-40% and 12-18%, respectively. Non-random attrition from the panel implies that these estimates are lower bounds. By contrast, we find evidence that, among older people, there is higher mobility at the top and bottom of the health distribution than there is at the middle. We consider two explanations for this: first that elderly persons may migrate to be closer to a family network once they fall ill, and second that non-random attrition may also be causing an upwards bias in the estimated effect of illness on mobility.

Key Words: Migration, Health, Selection, Attrition

---

* Corresponding author. Department of Economics and John A. Burns School of Medicine. Address: Department of Economics; 2424 Maile Way; Saunders Hall 533; Honolulu, HI 96822. Tele: (808) 956 - 8615. E-mail: halliday@hawaii.edu. This paper was originally the third chapter of my disseration at Princeton University. I would like to thank my advisors, Chris Paxson and Bo Honoré, for comments on an earlier draft. I would also like to thank Randy Akee for useful discussions. Both authors would like to thank the editor of this journal and two anonymous referees for excellent comments. The usual disclaimer applies.

1 Introduction

Social scientists have long recognized that health can impact both the costs and net benefits of migration. Poor health may increase the costs of executing all the activities that are necessary to migrate. In addition, poor health often necessitates an infrastructure to maintain and to facilitate activity, doctors to monitor their illnesses, medications and insurance to pay for it all. On the other hand, however, there are plausible scenarios in which illness may impact, either positively or negatively, the net benefits of migration. For example, sick people may have a greater incentive to migrate to locations where there is a stronger familial support network or better health care.[1] Overall, there are good reasons to expect health to impact a person's probability of migration (although the sign and magnitude of this effect is not *a priori* obvious) and that the health of a migrant will not be a random draw from the distribution of health in the sending region.

Many studies have investigated the idea that migrants do not constitute a random sample of the population from which they were drawn. The majority of these studies have concentrated on

[1] It has been suggested that this is one of the primary motives of return migration of Latinos from the US (Scribner 1994 and 1996).

labor market outcomes. In one of the earliest studies on the topic, Chiswick (1978) observed that in the period immediately after arrival, immigrants in the US tended to earn less than natives; however, after spending ten to fifteen years in the US, migrant earnings overtook native earnings. The explanation that was given for this phenomenon was that migrants are positively selected from their home countries and, hence, tend to be more motivated and ambitious than their native-born counterparts.[2] In a similar piece, Gabriel and Schmitz (1994) tested for positive selection in internal migration within the US. They showed that, prior to moving, migrants earned higher wages than demographically comparable non-migrants. More recently, Chiquiar and Hanson (2005), using Mexican and US Census data, calculated counter-factual wage densities and showed that Mexican immigrants in the US would have occupied the middle to upper parts of the Mexican wage distribution had they remained in Mexico. All of these studies support the claim that migrants are positively selected, at least when it comes to labor market outcomes.

There is also a substantial and related literature in demography and epidemiology that investigates the relationship between health and migration. Much of this literature focuses on international migration and documents a correlation in which migrants tend to be in better health than non-migrants. Interestingly, this literature has failed to produce a consensus about the underlying causal mechanisms which are responsible for this correlation.

Several explanations have been proposed. Pablos-Mendéz (1994) and Scribner (1994 and

[2] However, Chiswick's assertion that migrants are a positively selected group has been contested - most notably by Borjas (1987) who points out that Chiswick's estimates may be biased by omitted cohort effects. Borjas contends that these omitted cohort effects may matter if there has been a progressive deterioration in migrant quality with successive migrant cohorts. Using US census data, he argues that migrants from Western Europe have assimilated quite well and have progressively exhibited an increase in earnings over time. In contrast, he argues that migrants from poorer areas of the world, notably Latin America, have shown the opposite patterns; their rates of assimilation are quite poor and their earnings have steadily declined with successive cohorts. However, Borjas provides no evidence that migrants in the US would have resided in the lower tail of the earnings distribution of their home had they not migrated. Finally, it is important to mention that Borjas' findings have been scrutinized by Jasso and Rosenzweig (1990).

1996) speculated that the observed lower mortality rates among Latinos in the United States, or the “Latino Paradox,” is a consequence of Latino migrants returning to their home countries to die once they fall ill - the so-called “Salmon Bias.” Such return migration would mean that many Latino deaths would never get recorded in the US death records, thereby rendering many immigrants “statistically immortal.” (Pablos-Mendéz 1994) However, the evidence on the validity of this hypothesis is mixed.[3] Others have argued that more favorable cultural and behavioral factors are responsible for better health outcomes among migrants.[4] A third explanation posits that, just as migrants are positively selected on their ability to perform in the labor market, they are also positively selected on their health status.

There is some evidence of positive health selectivity in the demography and epidemiology literatures. For example, Jasso, Massey, Rosenzweig and Smith (2004) (hereafter JMRS) use the National Health Interview Survey and show that international migrants in the United States have a lower prevalence of many chronic conditions upon arrival.[5] In another study, Marmot, Adelstein and Bulusu (1984) compared mortality rates of migrants in the United Kingdom to the corresponding mortality rates of non-migrants in the sending countries and showed a strong tendency for mortality rates to be substantially lower among migrants than non-migrants. Swallen (2002) has found similar findings in the United States.

---

[3] Palloni and Arias (2004) provide evidence that the Salmon Bias is important in explaning the Latino Paradox. However, Abriado-Lanza, *et al.* (1999) provide evidence that mortality differentials exist for Puerto Ricans, whose deaths do get recorded in the National Death Index, and Cuban migrants, who primarily for political reasons do not return home, and thus provide evidence against the hypothesis.

[4] For example, Marmot and Syme (1976) have speculated that the lower rates of coronary heart disease among Japanese migrants to California are the consequence of a lower level of psycho-social stress which results from the way that Japanese society is structured. Others such as Markides and Coreil (1986) have speculated the lower mortality rates among Latinos are the consequence of differing health related behaviors.

[5] Two points are worth noting. First, this may reflect a lower diagnosis of certain chronic conditions among immigrants, many of whom came from countries with less developed health care systems. Second, this result compares the health on migrants and non-migrants in the receiving region, not the sending region.

However, despite this suggestive evidence of positive health selectivity, there is a surprising dearth of papers in the literature that have provided a rigorous investigation into the issue. As pointed out by JMRS, a proper test of health selection involves a comparison of the health of migrants and non-migrants in the sending region, not the receiving region since there are a number of factors that could differ across regions which influence mortality. We believe that one of the primary reasons for the lack of proper tests of health selectivity is inadequate data from sending countries prior to the occurrence of any migration.

In this paper, we attempt to fill this void in the literature by quantifying both the sign and the magnitude of the impact of health on migration using data on internal migration within the United States from the Panel Study of Income Dynamics (PSID). One of the advantages of focusing on internal US migration is that it enables us to observe the migrant's health prior to migration and, thus, provides us with a direct test of selection. One disadvantage of our approach, however, is that it is not clear how much we can extrapolate our results to the case of international migration. Nevertheless, as suggested by JMRS if illness increases intranational migration costs, it is also reasonable to expect similar (if not greater) effects when considering international migration.

The balance of this paper is organized as follows. In section 2, we outline our theoretical framework. In section 3, we describe the data. In section 4, we discuss our empirical methodology. In section 5, we discuss our empirical findings. In section 6, we discuss some sources of bias in our estimates including systematic biases in self-reported health status and non-random attrition from the panel. Section 7 concludes.

2 Theoretical Considerations

We model the migration decision using the standard model of migration as discussed in Borjas (1987). The respective utilities of remaining at home and moving to the destination are given by

$$u_0(h) = \mu_0(h) + \varepsilon_0 \tag{1}$$

and

$$u_1(h) = \mu_1(h) + \varepsilon_1 \tag{2}$$

where h is an index that is increasing in the quality of the individual's health. The values μ and ε represent the systematic and idiosyncratic components of utility. There are costs to migration which are given by $C(h)$. We assume that $C'(h) < 0$ so that the costs of migration are lower for healthier people. These migration costs may be pecuniary costs such as the actual costs of relocation, but they may also capture the "psychic costs" of migration, as well.

Following Borjas, the migration decision is determined by the sign of the index function

$$I = (\mu_1(h) - \mu_0(h) - C(h)) + (\varepsilon_1 - \varepsilon_0)\,. \tag{3}$$

If $I > 0$ ($I < 0$), then the person migrates (does not migrate). Letting $F(.)$ denote the Cumulative Density Function of $\upsilon \equiv \varepsilon_1 - \varepsilon_0$, we obtain that the migration propensity is given by

$$P \equiv 1 - F(z) \tag{4}$$

where $z \equiv -(\mu_1(h) - \mu_0(h) - C(h))$.

We will then have that

$$\frac{\partial P}{\partial h} = -f(z)\left[C'(h) - (\mu_1'(h) - \mu_0'(h))\right] \tag{5}$$

where $f(.)$ is the Probability Density Function of υ. The first term in brackets, $C'(h)$, is the impact of health on the costs of migration and the second term, $\mu_1'(h) - \mu_0'(h)$, is the impact of health on the net benefits of migration. If the second term is zero, then we will unambiguously have that $\frac{\partial P}{\partial h} > 0$ so that good health lowers migration costs and, thus, increases mobility.[6] Understanding the second term is slightly more complicated. While one can tell a variety of stories, we believe that if the term is non-zero, it is most likely negative so that good health would decrease the benefits of migration. For example, sick people with insufficient support networks at home may have to relocate to areas with better health care or stronger familial support networks. In such a scenario, the second term would operate in the opposite direction as the first and, thus, the net effects of health on mobility would be ambiguous. We conclude that, on purely theoretical grounds, there are no *a priori* reasons to expect the sign of $\frac{\partial P}{\partial h}$ to be either positive or negative. Accordingly, we turn to the data.

3 The Data

We use data from the PSID spanning the years 1984 to 1993 on geographic mobility, health status and other control variables which include age, income, gender, education, race and marital

[6]Note that the assumption that the second term is zero does not imply that health has no effect on the systematic component of utility which one would expect if good health raised wages by making it easier to work, but it does require that the systematic returns to good health be "balanced" in the sense that $\mu_1'(h) = \mu_0'(h)$.

status.[7] The PSID only has data on health status for heads of household and their spouses (if the household head is married). Consequently, throughout this analysis, we restrict our attention to these people. Our migration measure, "Moved," is an indicator of whether or not the individual changed states across survey years which is turned on if the individual lived in a different state in the previous time period. This definition of migration is common in the literature on internal migration within the US (*e.g.* Borjas, Bronnars and Trejo 1992; Gabriel and Schmitz 1994). We use Self-Reported Health Status (SRHS) as our measure of health. SRHS is a categorical variable that takes on integer values between one and five with one the most healthy category and five the least healthy. While these measures are subjective, there is an extensive literature that has shown a strong link between SRHS and health outcomes such as mortality and the prevalence of disease (Mossey and Shapiro 1982; Kaplan and Camacho 1983; Idler and Kasl 1995; Smith 2003). We break the SRHS variable into two binary indicator variables: Healthy, which is turned on when SRHS is either one or two, and Unhealthy, which is turned on when SRHS is either four or five. All other variable definitions and descriptive statistics for this sample can be found in Table 1. Finally, the PSID contains an over-sample of economically disadvantaged people called the Survey of Economic Opportunities (SEO). In this paper, we include the SEO. We do so because dropping it would have substantially reduced the number of moves in our data and thus resulted in more inefficient estimates.[8]

[7] We do not use data prior to 1984 because the SRHS question was not asked prior to that year. We not use data past 1993 because data on location are not publicly available from 1994 onward.

[8] There is little consensus within the profession about how one should deal with the SEO. Because it is selected on income and, thus, endogenous, conventional weighting schemes will not work. Accordingly, some people such as Lillard and Willis (1977) simply recommend dropping the SEO due to endogenous selection. Nevertheless, there are others such as Hyslop (1999) and Meghir and Pistaferri (2004) who include the SEO.

4 Empirical Methods

Our empirical analogue to the migration rule in equation (3) is

$$m_{i,t} = 1(g_{i,t-1}\gamma + b_{i,t-1}\beta + \mathbf{X}'_{i,t}\boldsymbol{\theta} + \varepsilon_{i,t} \geq 0) \tag{6}$$

where $m_{i,t}$ is a migration indicator which we defined in the next section, $g_{i,t}$ is the variable "Healthy," an indicator of excellent or very good health, $b_{i,t}$ is the variable "Unhealthy," an indicator of poor or fair health, and $\mathbf{X}_{i,t}$ is a column vector of individual characteristics. Note that in the above specification, the middle SRHS category (*i.e.* SRHS equal to three) is omitted. $\mathbf{X}_{i,t}$ contains age, gender, education, race, state and marital status dummies as well as a quadratic function of (lagged) income. The index function inside of equation (6) is meant to approximate the net benefits of migration as defined in the theoretical model. We assume a Normal distribution for $\varepsilon_{i,t}$ and so generate all results using Probit estimation. This gives us an empirical analogue to equation (4)

$$P\left(m_{i,t} = 1 | \mathbf{Z}_{i,t}\right) = \Phi\left(\mathbf{Z}'_{i,t}\boldsymbol{\lambda}\right) \tag{7}$$

where $\mathbf{Z}_{i,t} \equiv \left(1, g_{i,t-1}, b_{i,t-1}, \mathbf{X}'_{i,t}\right)'$ and $\boldsymbol{\lambda} \equiv (\alpha, \gamma, \beta, \boldsymbol{\theta}')'$. All standard errors are adjusted for correlations within individuals.

A crucial part of our identification strategy is to include lags of the health status variables. We do this because a proper test of health selectivity involves a comparison of the health of movers and stayers prior to the occurrence of any migration as there are a variety of reasons to expect migration to feed back and impact a person's health. For example, if it is the case that a sick person moves to be closer to a family member who is able to take care of them, then we might expect the move to improve the person's health. On the other hand, Kasl and Berkman (1983) have speculated that the stress of moving may induce a deterioration in health. Both of these scenarios suggest that it is crucial to have a measure of health status prior to the occurrence of the move if we are to reliably estimate the impact of health on migration. Fortunately, the panel structure of our data allows us to do this.

5 Empirical Results

Table 2 displays our estimation results for men and women younger than age 60 estimated separately by gender. The first three columns display the results for men and the last three for women. In all six columns, we see substantial evidence that migrants are positively selected on health. The F-tests show that the health variables are always jointly significant. For men, being unhealthy decreases the probability of migration by between 1.2 and 1.5 percentage points. For women, being unhealthy decreases the probability of migration by between 0.4 and 0.6 percentage points. Since the average probabilities of migration for men and women under 60 are 3.7% and 3.3%, respectively, these marginal effects constitute rather large effects in percentage terms. Indeed, moving from good health, which is the omitted SRHS category in these regressions, to worse health lowers the probability of migration by 32-40% for men and 12-18% for women. In columns three and six, we add a comprehensive set of controls including a complete set of state dummies. While the coefficients on the health variables are attenuated somewhat, they still remain jointly significant at levels higher than 95%. The coefficients on the additional control variables all have the expected signs.

In Table 4, we estimate the models using a sub-sample of people older than 60. We consider the same specifications as we considered in Table 3. As can readily be seen, the effects of health on migration are substantially different for older people than they are for younger people. For men, we see that being both healthy *and* unhealthy have positive impacts on migration probabilities and, thus, health appears to have a non-monotonic effect on mobility. By contrast, we do not see any evidence that health impacts the mobility of women over 60.

Taken at face value, these results suggest that both of the effects of health on the incentives to migrate which were summarize in equation (5) may be operating in different parts of the health distribution. The positive effect of being healthy on migration may be indicative of good health reducing migration costs. The positive effect of being unhealthy on migration suggests that, in this part of the health distribution, illness increases the benefits to migration. For example, older people who are exceptionally ill may migrate to be closer to family members who can care for them. Moreover, we might expect this effect to especially large among a population of older people, many of whom are widowed and so may lack an adequate support network to care for them in old age.

One crude way of testing this hypothesis is to see if being married attenuates the impact of poor health on migration. The rationale behind this is that married individuals who are in poor health can be taken care of by their spouses and would, thus, have lower incentives to relocate to a state where they can be cared for by another family member. To test this "family support network" hypothesis, we replicate the regressions from columns 3 and 6 of Table 4 except that now we interact the unhealthy indicator with the marriage indicator. If this hypothesis is true, then we would expect the interaction between bad health and the marriage indicator to be negative. The results are reported in Table 5. We see that the interaction term is negative and significant for men, but is negative and insignificant for women. We conclude that the evidence for this hypothesis is mixed.

6 Potential Sources of Bias

In this section, we explore some factors that may bias our estimates. The next sub-section investigates the role of reporting bias in SRHS. After that, we look into the role of non-random attrition from the panel.

6.1 Systematic Reporting Bias in SRHS

While it true that SRHS has proven to be reliable measure of health, it is still subject to biases and errors which, in some contexts, have been shown to be systematically correlated with socio-economic variables. If similar errors exist in our data that are systematically correlated with migrant status then our results will be biased. For example, if movers are systematically more optimistic about their health than stayers then this would result in estimates that look as if illness raises the costs of migration when, in fact, there may be no actual relationship between health (in the objective sense) and migration.

To investigate this issue, we employ the PSID's mortality file and estimate the relationship between SRHS and mortality using Cox-Proportional Hazard Models while adjusting for age.[9]

[9] The PSID's mortality file is considered sensitive and, thus, not publicly available. The death file contains mortality information on all individuals in the PSID from 1968 to 2003 who were known to have died prior to 2004. Mortality information first comes from interviews with PSID families. PSID then corroborates this information with the National Death Index. Death dates are recorded to the nearest month.

To see if there were any systematic biases across movers and stayers, we split the sample by migration status. One sample contained people who never moved while they were in our sample, and the other contained people who moved at least once. We then estimated the hazard models on both sample broken down by gender and age. Results are reported in Table 5.[10]

We take three points away from these results. First, we observe a statistically significant relationship between SRHS and mortality for both movers and stayers in all specifications except for women under 60 years old. However, it is important to mention that we do see a statistically significant relationship for the sample of all women. Second, the samples of movers can be quite small and, thus, the 95% confidence intervals for the hazard ratios in these samples are wide. Accordingly, the confidence intervals on the SRHS hazard ratios always overlap for movers and stayers. Third, while the hazard ratios for movers and stayers do differ somewhat, it is never systematic. In other words, there is no evidence of a systematically weaker (or stronger) relationship between SRHS and mortality among movers than stayers. Overall, we take this as evidence that there are probably not any systematic biases in our SRHS measures that are impacting our findings.

6.2 Non-Random Attrition

A potentially more important source of bias is non-random attrition. The reason is that two of the most common reasons for attrition are migration and death, the latter being more common for unhealthy people.[11] Unfortunately, non-random attrition is one of the least understood

[10] We used the 1984 wave of the PSID for the estimates. Each cell reports the hazard ratio for the relevant variable and its 95% confidence interval. If the hazard ratio is above (below) unity then that variable has a positive (negative) effect on mortality.

[11] Fitzgerald, Gottschalk and Moffitt (1998) exhaustively examine the reasons for attrition from the PSID. The reason for attrition of approximately 60% of all responders is not known. Of the remainder, approximately

areas in econometrics and, thus, there are few solutions proposed in the literature to address it. Moreover, some solutions which have been proposed are valid only under restrictive assumptions which are almost certainly violated in our data.[12] Consequently, due to a lack of valid solutions to this difficult problem, our approach is to provide a heuristic discussion of the bias that will result from non-random attrition and argue that our estimates constitute a lower bound of the true effect of health on migration. We also hold that non-random attrition can help to make sense of the observed non-monotonicty in Table 3.

With some abuse of notation, we consider a linear version of equation (6):

$$m_{i,t} = g_{i,t-1}\gamma + b_{i,t-1}\beta + \mathbf{X}'_{i,t}\boldsymbol{\theta} + \varepsilon_{i,t}. \tag{8}$$

Working with this linear model greatly facilitates the exposition. It should be noted that the OLS estimates of equation (8) are very similar to the marginal effects of Probit estimation of equation (6).[13] Equation (8) corresponds to the underlying population regression equation.

Due to non-random attrition, our data do not constitute an *i.i.d.* sample from this population. Instead, we only observe observations for individuals who "survive," or do not attrite, across survey years. We let $s_{i,t}$ denote the survival indicator. If the individual has survived from time *0* to time t, then the indicator equals unity; otherwise, it is zero. For the sake of simplicity, we assume that attrition is an absorbing state. As a consequence of panel attrition, the econometrician does not observe the vector $(m_{i,t}, g_{i,t-1}, b_{i,t-1}, \mathbf{X}'_{i,t})$. Instead, she observes

two-thirds attrite due to mortality and one-third attrite due to a move which could not be followed.

[12] One common procdure is Inverse Probability Weighting (IPW) of Moffitt, Fitzgerald and Gottschalk (1999), which is valid when attrition is affected by observable characteristics from the first year of the panel, but is unaffected by anything that occurs subsequently. Clearly, this criterion is not met in our case since mortality and migration are two of the primary causes of attrition in our data.

[13] We do not report the OLS results, but they are available upon request.

$(m^*_{i,t}, g^*_{i,t-1}, b^*_{i,t-1}, \mathbf{X}^{*\prime}_{i,t})$, where we have adopted the notation that $z^*_{i,t} = s_{i,t} z_{i,t}$, and she attempts to consistently estimate the parameters in

$$m^*_{i,t} = g^*_{i,t-1}\gamma + b^*_{i,t-1}\beta + \mathbf{X}^{*\prime}_{i,t}\boldsymbol{\theta} + \varepsilon^*_{i,t}. \tag{9}$$

However, OLS will result in inconsistent parameter estimates if the attrition is systematic since this implies that the residual in equation (3) will be correlated with the right-hand side regressors, particularly the health variables.[14] This, in turn, implies that the orthogonality conditions which are required for identification will be violated and, thus, OLS will not recover the parameters in equations (8) or (9).

The direction of the bias of the OLS estimates of γ and β will depend on the signs of two expectations: $E[s_{i,t} g_{i,t-1} \varepsilon_{i,t}]$ and $E[s_{i,t} b_{i,t-1} \varepsilon_{i,t}]$. We argue that the former is negative and the latter is positive. The reason is that survival in the panel is positively (negatively) correlated with being healthy (unhealthy) and negatively correlated with migration. Accordingly, we expect that the OLS estimate of γ to be biased downwards and the estimate of β to be biased upwards.

In Table 6, we give the reader a sense of how attrition rates in our PSID sample vary by health status. Each cell of the table reports the percentage of a PSID wave that attrites across survey years. We break the calculations down by gender, health status and age. Two points should be taken away from the table. First, attrition rates are substantially higher in the bottom two SRHS categories than they are when calculated for the entire health distribution. Second, attrition rates are substantially larger among people who are 60 years of age or older.

[14] If the attrition is not systematic or is random then $E[g_{i,t-1}\varepsilon_{i,t} | s_{i,t} = 1] = 0$ and, thus, by the law of iterated expectations, we will also have that $E[s_{i,t} g_{i,t-1} \varepsilon_{i,t}] = 0$. A similar argument can be applied to $E[s_{i,t} b_{i,t-1} \varepsilon_{i,t}]$.

This has several implications for our findings. First, it implies that the estimates of the effects of health on migration from Table 2 are conservative. In other words, the true impact of health on migration is probably greater than what we have estimated. Second, it suggests that if the attrition bias is great enough, the OLS estimate of β may actually be positive even if the true parameter is negative. Consequently, non-random attrition may also be responsible for the observed non-monotonicity for older men from Table 4, especially since we would expect the biases from attrition to be higher among older people for whom mortality-induced attrition is higher.

7 Conclusions, Limitations and Broader Implications

In this paper, we test the proposition that the health of migrants does not constitute a random sample of health in the sending region. Our results indicate that among men and women younger than age 60, being healthy (unhealthy) increases (decreases) geographic mobility. Among men older than age 60, the results appear to suggest that there is higher mobility at both the top and bottom of the health distribution. For older women, there is no evidence that health impacts mobility. We argue that, due to the bias induced by non-random attrition, the effects of health on migration constitute a lower bound. In other words, good (bad) health increases (decreases) mobility by more than we estimate.

The primary limitation of this work is that it is not clear how much one can extrapolate our results to other forms of migration such as international migration. Accordingly, our work does not (at least directly) suggest that the better health outcomes that we observe among many international immigrant groups are a consequence of positive selection on health. Nevertheless, we claim that our results are suggestive that positive selection may be an important part of this puzzle. The paucity of evidence on positive selection on health in international migration presumably has much to do with a lack of adequate data sources from the sending country. This suggests an important avenue for future research.

A second avenue for future research is to better understand the relationship between positive migrant selection on both health and labor market outcomes. One would expect these two types of selection to be intimately related, as there is a large literature which has shown that poor health has large causal effects on labor supply (Smith 1999; Rust and Phelan 1997) and educational attainment (Miguel and Kremer 2004; Bobonis, Miguel and Puri-Sharma 2006; Case, Fertig and Paxson 2004). Accordingly, an interesting (and ambitious) topic for future research would be to investigate how much of the observed positive selection on labor market outcomes is the result of the impact of health on labor supply and educational attainment.

References

[1] Abraído-Lanza, A.F., Dohenwend, B.P., Ng-Mak, D.S. and Turner, J.B., (1999), "The Latino Mortality Paradox: A Test of the 'Salmon Bias' and Healthy Migrant Hypotheses," *American Journal of Public Health*, 89: 1543-1548.

[2] Bobonis, G., Miguel, E. and Puri-Charma, C. (2006), "Anemia and School Participation," forthcoming *Journal of Human Resources.*

[3] Borjas, G.J. (1987), "Self-Selection and the Earnings of Immigrants," *American Economic Review*, 77: 531-553.

[4] Borjas, G.J., Bronnars, S.G. and Trejo, S.J. (1992), "Assimilation and the Earnings of Young Internal Migrants," *Review of Economics and Statistics*, 74: 170-175.

[5] Case, A., Fertig, A. and Paxson C. (2005), "The Lasting Impact of Childhood Health and Circumstance," *Journal of Health Economics*, 24: 364-389.

[6] Chiquiar, D. and Hanson, G.H. (2005), "International Migration, Self-Selection, and the Distribution of Wages: Evidence from Mexico and the United States," *Journal of Political Economy*, 113: 239-281.

[7] Chiswick, B.R. (1978), "The Effects of Americanization on the Earnings of Foreign Born Men," *Journal of Political Economy*, 86: 897-921.

[8] Fitzgerald, J., Gottschalk, P. and Moffitt, R. (1998), "An Analysis of Sample Attrition in Panel Data: The Michigan Panel Study of Income Dynamics," *Journal of Human Resources*, 33: 251-299.

[9] Gabriel, P.E. and Schmitz, S. (1995), "Favorable Self-Selection and the Internal Migration of Young White Males in the United States," *Journal of Human Resources*, 30: 460-471.

[10] Hyslop, D. (1999), "State Dependence, Serial Correlation and Heterogeneity in Intertemporal Labor Force Participation of Married Women," *Econometrica,* 67: 1255-1294.

[11] Idler, E.L. and Kasl, S.V. (1995), "Self-ratings of Health: Do They Also Predict Change in Functional Ability?" *Journal of Gerontology*, 50, S344-S353.

[12] Jasso, G and Rosenzweig, M.R. (1990), "Self-Selection and the Earnings of Immigrants: Comment," *American Economic Review*, 80, 298-304.

[13] Jasso, G., Massey, D.S., Rosenzweig, M.R. and Smith, J.P. (2004), "Immigrant Health - Selectivity and Acculturation," unpublished manuscript, Institute for Fiscal Studies.

[14] Kaplan, G.A. and Camacho, T. (1983), "Perceived Health and Mortality: A 9 Year Follow-up of the Human Population Laboratory Cohort," *American Journal of Epidemiology*, 177, 292-304.

[15] Kasl, S.V. and Berkman, L. (1983), "Health Consequences of the Experience of Migration," *Annual Review of Public Health.* 1983. 4: 69-90.

[16] Lillard, L.A. and Willis, R.J. (1978), "Dynamic Aspects of Earning Mobility," *Econometrica*, 46, 985-1012.

[17] Markides, K.S. and Coreil, J. (1986), "The Health of Hispanics in the Southwestern United States: An Epidemiological Paradox," *Public Health Reports*, 101: 253-265.

[18] Marmot, M.G., Adelstein, A.G. and Bulusu, L. (1984), "Lessons from the Study of Immigrant Mortality," *Lancet.* June 30: 1455-1457.

[19] Marmot, M. G. and Syme, L. (1976), "Acculturation and Coronary Heart Disease in Japanese-Americans," *American Journal of Epidemiology*, 104: 225-247.

[20] Meghir, C. and Pistaferri, L. (2004), "Income Variance Dynamics and Heterogeneity," *Econometrica*, 72: 1-32.

[21] Miguel, E. and Kremer, M. (2004), "Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities," *Econometrica*, 72: 159-218.

[22] Moffitt, R., Fitzgerald, J. and Gottschalk, P. (1999), "Sample Attrition in Panel Data: The Role of Selection on Observables," *Annale d'Economie et de Statistique*, 55-56: 129-152.

[23] Mossey, J.M. and Shapiro, E. (1982), "Self-rated Health: A Predictor of Mortality Among the Elderly," *American Journal of Public Health*, 72: 800-808.

[24] Pablos-Mendéz, A. (1994), "Mortality among Hispanics," *Journal of the American Medical Association*, 271: 1237.

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

[26] Scribner, R. (1994), "Mortality among Hispanics," *Journal of the American Medical Association,* 271: 1238-1239.

[27] Scribner, R. (1996), "Paradox as Paradigm - the Health Outcomes of Mexican Americans" (editorial), *Journal of the American Medical Association*, 86: 303-305.

[28] Rust, J. and C. Phelan (1997), "How Social Security and Medicare Affect Retirement Behavior In a World of Incomplete Markets," *Econometrica*, 65, 781-831.

[29] Smith, J. (1999), "Healthy Bodies and Thick Wallets: The Dual Relation between Health and Economic Status," *Journal of Economic Perspectives*, 13, 145-166.

[30] Smith, J. (2003), "SES and Health Over the Life-Course," unpublished manuscript, RAND.

[31] Swallen, K.C. (2002), "Mortality in the US: Comparing Race/Ethnicity and Nativity," unpublished manuscript, University of Wisconsin.

Table 1: Variable Definitions and Statistics

Variable	Definition	Mean (Standard Deviation)
Moved	Indicator of whether or not individual has moved across two survey years	0.03 (0.17)
Self Reported Health Status (SRHS)	1=excellent; 2 = very good; 3 = good 2=fair; 1 = poor	2.49 (1.15)
Healthy	SRHS = 1 or 2	0.53 (0.50)
Unhealthy	SRHS = 4 or 5	0.19 (0.39)
Age	Individual's Age	42.49 (15.91)
Labor Income	Individual's Labor Income in 1982 dollars	11928.21 (15843.05)
Sex	=1 if female	0.54 (0.50)
No College Experience	= 1 if the individual never attended college	0.65 (0.48)
College Degree	= 1 if the individual has a college degree	0.19 (0.39)
White	= 1 if the individual is white	0.66 (0.48)
Black	= 1 if the individual is black	0.29 (0.46)
Married	= 1 if the individual is married	0.71 (0.45)

Table 2: Lagged Period Health - Under Age 60

	(1)	(2)	(3)	(4)	(5)	(6)
	Men			Women		
Unhealthy at t-1[1]	-0.014**	-0.015**	-0.012**	-0.006*	-0.006*	-0.004
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.002)
Healthy at t-1[2]	0.008**	0.008**	0.004	0.007**	0.008**	0.003
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Income at t-1	-	-0.001	0.001	-	0.003*	0.004**
		(0.002)	(0.001)		(0.001)	(0.001)
Income at t-1 Squared	-	0.00008	-0.0002	-	-0.0003*	-0.0005**
		(0.0002)	(0.0001)		(0.0001)	(0.0001)
No College Experience	-	-	-0.012**	-	-	-0.014**
			(0.003)			(0.003)
College Degree	-	-	0.016**	-	-	0.007*
			(0.004)			(0.003)
White	-	-	0.009	-	-	0.004
			(0.005)			(0.004)
Black	-	-	0.002	-	-	-0.005
			(0.006)			(0.004)
Married	-	-	-0.019**	-	-	-0.007**
			(0.003)			(0.002)
State Dummies	No	No	Yes	No	No	Yes
F-test[3]	40.69	41.59	23.18	25.51	27.95	6.39
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.041]
R ˆ 2	0.0361	0.0362	0.0863	0.0352	0.0358	0.0731
NT[4]	39679	39677	39363	45291	45289	44906

[+]This table reports marginal effects of Probit models where the dependent variable is moved. Standard errors of the marginal effect are in parentheses. Standard errors allow for clustering within individuals. All regressions include a complete set of age dummies.

[*] Denotes 95% significance.

[**]Denotes 99% significance.

[1]Refers to SRHS equal to 4 or 5.

[2]Refers to SRHS equals to 1 or 2.

[3]$F-$ test of the null that the health variables are zero. p-values are in brackets.

[4] NT refers to individual/time observations.

Table 3: Lagged Period Health - Age 60 and Over

	(1)	(2)	(3)	(4)	(5)	(6)
	Men			Women		
Unhealthy at t-1[1]	0.014**	0.015**	0.016**	-0.002	-0.002	-0.000
	(0.005)	(0.005)	(0.005)	(0.003)	(0.003)	(0.002)
Healthy at t-1[2]	0.024**	0.022**	0.020**	0.001	0.000	-0.001
	(0.007)	(0.007)	(0.006)	(0.003)	(0.003)	(0.002)
Income at t-1	-	-0.0008	-0.0009	-	-0.0003	-0.0014
		(0.002)	(0.001)		(0.002)	(0.002)
Income at t-1 Squared	-	0.0001	0.0001	-	-0.00004	0.0002
		(0.0002)	(0.0002)		(0.0003)	(0.0002)
No College Experience	-	-	-0.003	-	-	0.002
			(0.006)			(0.003)
College Degree	-	-	0.005	-	-	0.010
			(0.006)			(0.007)
White	-	-	0.004	-	-	0.001
			(0.005)			(0.005)
Black	-	-	-0.004	-	-	-0.006
			(0.005)			(0.004)
Married	-	-	-0.005	-	-	-0.007**
			(0.004)			(0.002)
State Dummies	No	No	Yes	No	No	Yes
F-test[3]	20.37	19.61	24.19	0.59	0.55	0.23
	[0.000]	[0.000]	[0.000]	[0.7436]	[0.7593]	[0.8928]
$R\,\hat{}\,2$	0.0400	0.0408	0.1064	0.0255	0.0255	0.1069
NT[4]	7591	7591	6299	10895	10895	10455

[+]This table reports marginal effects of Probit models where the dependent variable is moved. Standard errors of the marginal effect are in parentheses. Standard errors allow for clustering within individuals. All regressions include a complete set of age dummies.

[*] Denotes 95% significance.

[**]Denotes 99% significance.

[1]Refers to SRHS equal to 4 or 5.

[2]Refers to SRHS equals to 1 or 2.

[3]$F-$ test of the null that the health variables are zero. p-values are in brackets.

[4] NT refers to individual/time observations.

Table 4: The Impact of Marriage on Selective Migration

	(1)	(2)
	Men	Women
Unhealthy at t-1[1]	0.040**	0.002
	(0.015)	(0.002)
Healthy at t-1[2]	0.019**	−0.001
	(0.005)	(0.002)
Unhealthy at t-1 * Married	−0.012*	−0.005
	(0.004)	(0.003)
Married	0.005	−0.004
	(0.004)	(0.002)
F-test on healthy and unhealthy[3]	26.93	1.27
	[0.000]	[0.5290]
F-test on health variables and interaction terms[3]	26.93	2.93
	[0.000]	[0.4018]
R ˆ 2	0.1137	0.1086
NT[6]	6299	10455

[+]This table reports marginal effects of Probit models where the dependent variable is moved. Standard errors of the marginal effects are in parentheses. Standard errors allow for clustering within individuals. All regressions contain the same set of controls as are in columns 3 and 6 of Tables 2 and 3.
[*]Denotes 95% significance.
[*]Denotes 99% significance.
[1]Refers to SRHS equal to 4 or 5.
[2]Refers to SRHS equal to 1 or 2.
[3]p - values are in brackets.
[4]NT refers to individual/time observations.

Table 5: SRHS and Mortality by Migration Status

< 60 Years

	Men		Women	
	Stayers	Movers	Stayers	Movers
Age	1.079**	1.103**	1.087**	1.102**
	[1.067, 1.092]	[1.071, 1.136]	[1.073, 1.100]	[1.062, 1.143]
Unhealthy[1]	2.067**	1.314	1.755**	2.307
	[1.542, 2.772]	[0.498, 3.470]	[1.306, 2.359]	[0.761, 6.993]
Healthy[2]	0.632**	0.361**	0.706**	0.709
	[0.474, 0.843]	[0.180, 0.722]	[0.514, 0.972]	[0.232, 2.169]
N	3329	601	4010	649

All Ages

	Men		Women	
	Stayers	Movers	Stayers	Movers
Age	1.084**	1.092**	1.090**	1.100**
	[1.076, 1.092]	[1.073, 1.111]	[1.084, 1.095]	[1.083, 1.118]
Unhealthy[1]	1.398**	1.260	1.685**	0.924
	[1.159, 1.687]	[0.687, 2.310]	[1.419, 2.002]	[0.517, 1.652]
Healthy[2]	0.611**	0.411**	0.839	0.427**
	[0.507, 0.735]	[0.232, 0.728]	[0.691, 1.020]	[0.234, 0.777]
N	3952	647	4904	731

+This table reports estimates of Cox proportional hazard models. The dependent variable is mortality. Each cell reports the hazard ratio and its 95% confidence interval. All estimates adjust for clustering within individuals. If the hazard ratio is greater (less) than unity then the variable has a positive (negative) effect on mortality.

*Denotes 95% significance.

**Denotes 99% significance.

[1]Refers to SRHS equal to 4 or 5.

[2]Refers to SRHS equal to 1 or 2.

Table 6: Panel Attrition Rates by Age, Gender and Health Status

< 60 Years

Men		Women	
All	SRHS = 4 or 5	All	SRHS = 4 or 5
4.32%	5.87%	4.19%	6.03%

>= 60 Years

Men		Women	
All	SRHS = 4 or 5	All	SRHS = 4 or 5
5.17%	7.35%	5.10%	7.68%