The long run health consequences of rural-urban migration

Janna E. Johnson
Humphrey School of Public Affairs, University of Minnesota

Evan J. Taylor
Center for the Economics of Human Development, University of Chicago

Rural-urban migration is an integral part of the structural transformation as societies move from a traditional agricultural economy to a modern economy. This process has many potential consequences for migrants. Our study focuses on the lifetime health effects of the large mid-20th century migration out of rural U.S. Northern Great Plains states, primarily to urban locations in the West and Midwest. An analysis of marginal treatment effects (MTEs) shows that (a) migrants are positively selected, and (b) the causal impact of migration is decreased longevity. Our evidence suggests that elevated mortality among migrants is linked to increased smoking and alcohol consumption.

Keywords. Rural-urban migration, mortality, marginal treatment effects.

JEL classification. C31, I12, R23.

1. Introduction

Rural-urban migration is a fundamental feature of economic development—a process completed over the past century in developed countries, and still underway in many developing countries. As Lucas (2004) argues, rural-urban migration is key to the "transition from a traditional agricultural society to a society of sustained growth in opportunities, of human and physical capital accumulation. . . . [It is] an irreversible process that every industrializing society undergoes once and only once."

Given that rural-urban migration is ubiquitous, it is valuable to understand the impact of migration on migrants themselves. We undertake one such evaluation, studying an archetypal episode of rural to urban migration—the massive mid-20th century outflow of migrants from the rural Northern Great Plains (North Dakota, South Dakota, and Montana) to urban centers in the Western and Midwestern U.S. For the cohorts we study, born in the Dakotas and Montana, 1916–1927, lifetime migration rates were extremely high; well over half of these individuals migrated out of the area before age 40. We estimate the effect of migration on an important dimension of lifetime well-being: longevity.

Since the inception of the scientific study of migration, scholars have recognized two salient features of migration, both of which are potentially germane in our setting. First,

Janna E. Johnson: jannaj@umn.edu
Evan J. Taylor: evanjtaylor@gmail.com

migrants likely differ from nonmigrants along unobserved dimensions. Second, the impact of migration plausibly varies across individuals. Analyses that ignore self-selection can give misleading results, and failure to account for heterogeneity in effects may mask important information about the causal impact of migration on lifetime health.[1]

We proceed with a research design set up to account for both of these features of migration. In economic models of migration, the migration decision hinges on a comparison of anticipated costs and benefits, requiring the formation of expectations about prospects (employment opportunities, wages, etc.) in distant locations, which in turn depends on the availability of information. Our research design exploits plausibly quasi-random variation among Northern Great Plains residents in access to relevant information about life prospects in potential migration destinations. During the era we study—the early- to mid-20th century—many young people born in the Northern Great Plains lived in remote areas that were quite isolated from the rest of the country, and plausibly had little in the way of salient information about what life would be like in Seattle, Los Angeles, Minneapolis, or other potential urban locations to which they might migrate. Others lived in towns that were well connected to the outside world, where newspapers were widely available, and where many family friends and relatives had previously moved to urban locations. Individuals born in towns that were "information centers" were plausibly able to form assessments about outside opportunities more precisely than were individuals born in towns that were more isolated from information flows; differences in the availability of information likely resulted in variation in town-level out-migration.

We cannot directly access the amount or quality of information about distant locations available to young potential migrants in the 1930s, but we do have two sets of instruments to help us in implementing our design. First, we use instruments based on the location of railway lines in place pre-1900, and the location of railway lines constructed post-1900. At the turn of the 20th century, migration out of the Northern Great Plains was surely shaped in part by the extant railway structure, both in terms of the level of migration, and the direction of the migratory flow.[2] We argue that these early railway towns thereby became "information centers," in which residents would likely be particularly well informed about economic prospects in far-way urban centers, especially in destinations chosen by migrants who left the town in the late 19th and very early 20th centuries.[3] Towns on these train lines would also have had a constant flow of visiting

---

[1] The idea of systematic selection into migration extends back at least to William Farr (1864), who argues that migrants from rural England to London were likely healthier than their nonmigrating counterparts. An early appearance of the idea of heterogeneous effects of migration is found in an 1885 interaction between Roland Hamilton and E. G. Ravenstein (reported in the *Journal of the Statistical Society of London*), in which Hamilton suggests that the economic welfare impacts of internal migration in England differed across the "very mixed" population of migrants. He thought that migration would have relatively larger positive effects among "the more vigorous class" (Ravenstein (1885)).

[2] This observation—that railways play a key role in shaping migration—appears widely in the economic literature on migration, for example, Woodruff and Zenteno (2007), Boustan (2010), and Black, Sanders, Taylor, and Taylor (2015), as well as in the broader historical literature.

[3] In short, we expect some "socially influenced" migration, a phenomenon described in Carrington, Detragiache, and Vishwanath (1996), and documented, for both black and white migrants, during the era we study (Stuart and Taylor (2017)).

travelers from other states, who provided perspectives on distant locations, and trains facilitated travel for families visiting locations outside the Northern Great Plains. Towns near railways built post-1900 likely shared many similarities with towns on pre-1900 railway lines, but plausibly were somewhat less well connected. Towns with no rail services were the least connected.

Our second design is closely related. We argue that geographical patterns of U.S. mail delivery in the early 20th century in the Northern Great Plains provide a reasonably good indication of the extent to which individuals in rural communities—scattered across the Dakotas and Montana—were knowledgeable about the outside world and, therefore, aware of the economic alternatives in other parts of the country. We form instruments based on town-level flows of mail using historical data from Borchert (1987).

As an empirical matter, we find that both sets of instruments affected lifetime migration exactly as expected. Rates of migration were higher for individuals born in towns located on pre-1900 railways than those born in towns located on post-1900 railways, and, in turn, individuals born in nonrailway towns had the lowest rates of out-migration. Similarly, town-level out-migration rates were positively related to town-level postal mail flows. We proceed to estimate marginal treatment effects (MTEs), using the approaches of Heckman and Vytlacil (1999, 2005, 2007) and Brinch, Mogstad, and Wiswall (2017).

To preview results: In our context, there is evidence of positive selection into migration in terms of economic prospects and lifelong health. Our key result is that the "treatment effect" of migration out of the Northern Great Plains is to substantially *reduce* longevity on average. Taken at face value, our MTE estimates suggest that the longevity penalty was greatest for individuals least likely to migrate, but this result is inconclusive (as differences in estimated treatment effects over the "ability" distribution are not statistically significant).

Our empirical analysis contributes to the broader literature on rural-urban migration—a demographic phenomenon that is fueling a rapid and historic global transformation.[4] Despite the substantial literature on rural to urban migration, there is very little systematic work on the causal impact of moving on lifetime outcomes of any sort. One exception is a recent paper by Black et al. (2015) on the Great Migration of African Americans out of the South, work with which this paper shares several similarities. We use the same data (a proprietary match of records from the Social Security Administration and Medicare), similar methods, and find similar surprising results: those who migrated experienced higher mortality in older age compared to those who remained in their origin area. While we acknowledge these parallels, our work both expands on the analysis of Black and coauthors and has broader implications for the study of rural-urban migration.

First, the Great Migration—important as it was—was a peculiar migration unique to the time and place in which it occurred. African Americans in the early 20th cen-

---

[4]From 1950 through 2010, driven in large measure by rural to urban migration, the proportion of people living in urban areas increased from 54 to 77% in developed regions of the world, and from 18 to 46% in less developed areas (United Nations, Department of Economic and Social Affairs, Population Division (2014)). Over these same years, urbanization increased from 12 to 49% in China, from 10 to 49% in Nigeria, from 21 to 83% in South Korea, and from 36 to 84% in Brazil.

tury South faced political disenfranchisement and institutional discrimination, conditions which helped spur their mass movement to the more tolerant North. In contrast, we examine a much more common type of migration, seen worldwide: the emptying of rural agricultural areas brought on by industrialization and urbanization. In addition, unlike the individuals impacted by the Great Migration, the population we study was not generally disadvantaged,[5] nor did they face racial discrimination in their origin or destination areas. These differences could lead to differing effects of migration on longevity for the two populations. Finally, in comparison to Black et al. (2015) we make some methodological progress, by using an estimation procedure that allows for heterogeneity in causal effects of migration.

As for causal mechanisms, we provide suggestive evidence that the lower longevity experienced by migrants is the consequence, at least in part, of increased mortality due to chronic liver disease and cirrhosis, and lung cancer and chronic obstructive pulmonary diseases. These findings point to an important role for such behavioral factors as alcohol abuse and smoking. In this respect, our work provides a useful empirical input to a current policy issue. Researchers who study public health in China have expressed concern that rural to urban migration in that country is fueling a national smoking epidemic (Yang, Wu, Rockett, Abdullah, Beard, and Ye (2009), Liu et al. (2015), Ji, Liu, Zhao, Jiang, Zeng, and Chang (2016)), a fact also acknowledged in a recent report by the World Health Organization (World Health Organization (2017)). Evidence that rural-urban migrants suffer poorer health than their rural counterparts has also surfaced in India (Reddy, Shah, Varghese, and Ramadoss (2005)) and Indonesia (Lu (2010)). Our work suggests that a similar dynamic was at play decades earlier in the United States; migration to cities resulted in premature deaths, plausibly because of behavioral factors such as smoking.

Our paper proceeds as follows: Section 2 provides an historical backdrop. Section 3 describes how information availability plausibly influenced migration and outlines our research design. Section 4 describes our data. Section 5 gives empirical results, and discusses threats to the validity of our results. Section 6 concludes.

2. Migration from the U.S. Great Plains in historical perspective

2.1 *Rural-urban migration and its impact on health*

Rural-urban migration is a widely-documented, intensely-studied demographic phenomenon found in countries across the globe. In the important early economic analyses of Todaro (1969) and Harris and Todaro (1970), and more recent work of Lucas (2004), rural-urban migration is discussed in terms of the natural evolution of societies as their economies become industrialized. As described in this literature, the movement of individuals from rural to urban areas is spurred by higher expected wages in cities relative to rural areas. As the manufacturing and service jobs available in growing cities generally offer higher returns to skill than those available in agriculture-dominated rural

---

[5]White individuals born in the Dakotas and Montana, who were largely descendants of northern Europeans, had on average over 11 years of education (more than whites born in the rest of the country), while the African Americans studied in Black et al. (2015) had less than 8.

areas, higher-skilled individuals tend to concentrate in urban areas. The growth of cities drives economic expansion by increasing human capital, as the spread of knowledge through worker proximity is aided by increased urban density (Glaeser (1999), Glaeser and Resseger (2010)).

The timing and pace of rural-urban population flows is driven by the timing and pace of the transition of a society from a traditional economy to a modern economy. The Industrial Revolution triggered rapid urbanization in England and western Europe as early as the mid-19th century (Davis (1955)). In the United States, this process started in the late 19th century and was largely complete within less than a century.[6]

While migration to cities provided expanded economic opportunities, concerns about health impacts were expressed in the contemporaneous literature even in the mid-19th century (Farr (1864)). These concerns were likely well founded, given the poor sanitary conditions in many urban areas, including contaminated water supplies, improper waste management, and crowded unventilated housing, which enabled the rapid spread of infectious disease (Grob (2002)). Studies suggest that urban mortality was higher than rural mortality into the 20th century, when living conditions in cities improved rapidly (Vlahov, Gibble, Freudenberg, and Galea (2004)). Today, age-adjusted mortality rates in the U.S. are higher in rural areas than urban ones (National Center for Health Statistics (2015)). Modern cities are thought to have features that positively contribute to the health of their populations (Glaeser (2011)).

As we noted in the Introduction, there are continued public health concerns that rural-urban migration creates serious health deficits among migrants. However, there is very little research evaluating the causal impact on health of migration from rural to urban areas; our goal is to provide such evidence for one important episode of rural-urban migration.

2.2 *Rural-urban migration from the Northern Great Plains*

Migration has played a dominant role in shaping the population of the Dakotas and Montana since their first settlement by non-Native peoples. Prior to 1864, few people lived in the three states (then the Dakota and Montana territories). The passage of the Homestead Act in 1864, which allowed settlers to claim 120 acres of federal land for their own, and the construction of the first railroad lines in the area in the 1870s and 1880s, triggered a flood of in-migration from both the Eastern United States and abroad. As a result, the population of the region grew rapidly, to almost 1 million in 1900 and close to 2 million in 1930. By the 1920–1930 period, though, the region was experiencing steady net *out*-migration. Figure 1 shows net population gains/losses per 1000 inhabitants due to net migration between successive decennial censuses.[7] The migration out of the region that began in the 1920–1930 period continued for many successive decades.

---

[6]In 1900, 39% of the U.S. population resided in urban areas, and by 1990 the figure was 75% (Ruggles, Alexander, Genadek, Goeken, Schroeder, and Sobek (2010)).

[7]These rates of net migration are calculated using successive decennial censuses and survival ratios derived from U.S. Census data (Ferrie (2006)).
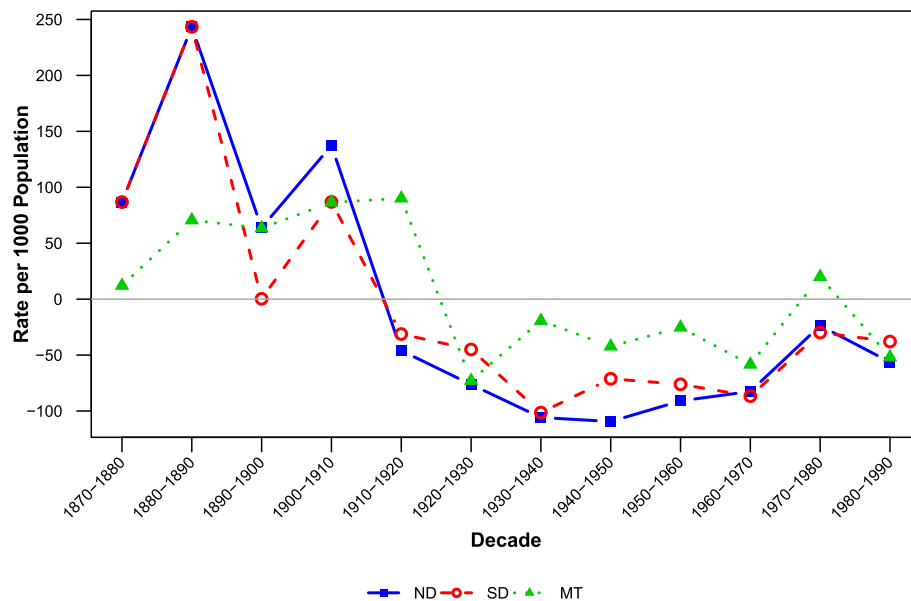
Figure 1. Net intercensal migration, Dakotas and Montana, 1880–1990. Source: "Net intercensal migration, by race, nativity, and state: 1850–1990 [Survival rate method]." Table Ac206-413 in Ferrie (2006).

The cohorts we study were born in the Dakotas and Montana between 1916 and 1927, near the beginning of this sustained period of out-migration. This migratory flow is an exemplary case of the rural to urban migration studied by Lucas (2004). Over the 20th century the U.S. experienced a dramatic structural transformation that reshaped the dispersion of the country's population (Michaels, Rauch, and Redding (2012)). Many regions in which employment was initially predominantly agricultural experienced out-migration; many urban manufacturing centers expanded rapidly.

In 1920, even as net out-migration from the Northern Great Plains began, the region was extremely rural; the population density in the three states was only 6.2 people per square mile (compared to 38.0 for the rest of the country). The least densely populated counties in each state had between one and three individuals per square mile, and the three most densely populated counties, which contain the largest city in each state, had only between 23.5 and 83.1 individuals per square mile.[8] In addition to being entirely rural, Montana, North Dakota, and South Dakota were homogeneous racially; 97.7% were white, 2.0% Native American, and 0.3% other. (We exclude nonwhites from our empirical work.) Employment in the region was highly concentrated in the agricultural sector. According to 1920 Census records, 58% of the male labor force reported employment as farmers, farm laborers, farm managers, and farm advisors (compared to 28% in other states).

---

[8]The least densely populated counties were Powder River, MT (population 3357), Billings, ND (3126), and Washabaugh, SD (1166). The three largest cities were very small: Butte, MT had population 41,611, Fargo, ND, 21,961, and Sioux Falls, SD, 25,202.
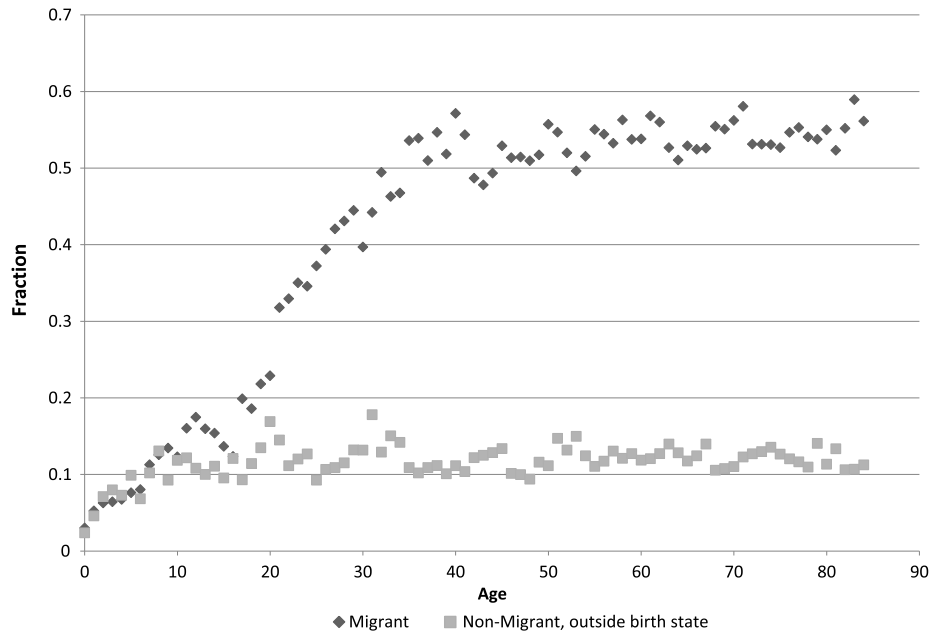
Figure 2. Location by age, white individuals born in the Dakotas and Montana, 1916–1927. Notes: Sample includes non-Hispanic white individuals born in the Dakotas and Montana between 1916 and 1927. Each point is calculated as a mean for each single year of age (all ages between 0 and 84 are covered due to the sample of 11 birth cohorts and decennial censuses). Excluded category is nonmigrant, in birth state. For migrant definition, see text. Source: Authors' calculations using the IPUMS of the 1920–2000 Census (Ruggles et al. (2010)).

In our study of rural-urban migration, we define "nonmigrants" to be individuals born in the Dakotas and Montana who in old age reside in those three states or in the surrounding states of Idaho, Wyoming, Nebraska, and Iowa, or in rural Minnesota. Individuals remaining in this broad region are designated "nonmigrants" because the entire region is rural, with economies dominated by agriculture; people moving within the region were generally *not* rural-urban migrants. "Migrants" are individuals who moved elsewhere in the U.S. (to other states or to the Minneapolis-St. Paul metropolitan area). None of the major destination cities of migrants out of the Dakotas and Montana were in the surrounding states (save Minneapolis-St. Paul), and a relatively small fraction of the individuals we study moved to these surrounding states. This can be seen in Figure 2, in which we use U.S. Census data to trace out the location of residence for individuals born in the Dakotas and Montana, 1916–1927, by age—separately for those who migrate out of the region, and nonmigrants who live outside of their state of birth (with the excluded category being non-migrants still residing in their state of birth). After around age 10, the fraction of nonmigrant individuals outside of their state of birth remains constant at approximately 0.12.

We note three additional features from Figure 2. First, migration occurred primarily when people were young adults; net migration was completed by the time these individuals were in their 30s. Second, there does not appear to be significant net return mi-

TABLE 1. Location of White Dakota and Montana Natives in 1960 and in old age, birth cohorts 1916–1927.

	1960	Old age
Of all surviving, percent		
Migrant	51.3	53.2
Nonmigrant, in birth state	36.8	34.3
Nonmigrant, outside birth state	11.9	12.6
Of migrants, percent in		
California	27.7	24.9
Washington	21.7	20.1
Oregon	9.2	10.4
Minnesota	7.8	5.9
Illinois	4.3	3.0
Wisconsin	3.8	3.0
Colorado	2.2	3.3
Michigan	2.5	2.0
Texas	2.0	2.8
Arizona	1.6	5.7
Florida	0.9	3.0
Other states	16.3	15.9
Of migrants, percent in		
Metropolitan area	67.4	83.2
Nonmetropolitan area	32.6	16.8
Of migrants, percent in		
Seattle, WA		10.8
Los Angeles, CA		9.5
Portland, OR		6.9
Minneapolis-St. Paul, MN		6.1
Other metropolitan areas		50.0
Nonmetropolitan areas		16.8

Note: Sample includes non-Hispanic whites born in the Dakotas and Montana. For migrant definition, see text. The 1960 Census does not identify specific metropolitan areas, just metropolitan status. States shown that have at least 1% of all migrants in one or both time periods, metropolitan areas shown that have at least 5% of all migrants in old age.
Sources: Authors' calculations using the IPUMS of the 1960 census (Ruggles et al. (2010)) and Duke SSA/Medicare data.

gration nor additional out-migration as these cohorts aged (such as at retirement ages), as the fraction migrating remains quite constant at older ages. Third, migration rates are extremely high. By age 40, well over half of individuals resided outside of the Northern Great Plains. For comparison, among these same cohorts the fraction of African Americans who migrated out of the Deep South during the Great Migration was approximately 0.45 (Black et al. (2015)).

Table 1 shows the location in 1960 of white individuals born in the Dakotas and Montana, 1916–1927 (when they were aged 33–44, calculated using Census data) and in old age (aged 65+, calculated using the Duke SSA/Medicare data, which we describe below). Results are consistent with Figure 2. We show also that approximately half of all migrants ended up on the West Coast, but many also migrated to nonrural Minnesota, Illinois, or

Wisconsin. In both periods, the majority of migrants were in metropolitan areas, nearly 70% in 1960 and more than 80% in old age.[9]

2.3 *Migration and selection*

In the Harris–Todaro and Lucas models of rural-urban migration, individuals are motivated to migrate to cities because urban areas specialize in human capital-intensive production. The returns to latent "ability," the underlying capability to acquire and use human capital, are higher in urban areas than rural areas. If there is variation in latent ability, we expect positive selection into migration.[10]

Before turning to our primary empirical analysis, we first provide evidence about the selection of migration in terms of education and earnings—characteristics that are likely correlated with latent ability (Griliches and Mason (1972)). We calculate differences in education for white nonmigrants and migrants born in the Dakotas and Montana 1916–1927, using the 1960 U.S. Census.[11] We find that migrants are better educated than nonmigrants; among men average education is 10.8 for nonmigrants and 12.0 for migrants, while for women, education averages 11.2 for nonmigrants and 11.9 for migrants. Overall, for white men and women born 1916–1927 in the Dakotas and Montana, average educational attainment is slightly higher than whites in the rest of the country (11.1 for both men and women).

Migrants also have higher earnings than nonmigrants. Table 2 shows results from OLS regressions of log earnings on migrant status for men in the labor force, using data the 1960, 1970, and 1980 U.S. Decennial Census. After controlling for education and age, migrants earn on average 40–56% more than nonmigrants. This difference could be due to positive selection of migrants on characteristics that increase earnings (besides education), a causal effect of migration on earnings, or some combination of the two.[12] Also, real earnings differences between migrants and nonmigrants are likely lower, due to regional cost-of-living differences.

In any event, differences in the observed characteristics of migrants and nonmigrants give us reason to suspect there may be differences also along a latent dimension. In turn, this same latent characteristic is plausibly related to lifetime health. Specifically, a very large literature shows a positive correlation between education/earnings and health (see Sorlie, Rogot, Anderson, Johnson, and Backlund (1992) and Elo and Preston (1996) for two examples), giving rise to our concern that migrants may be positively selected on a latent dimension related to health.

[9]Note that the increase over the two periods could be due to migrants moving to metropolitan areas and/or the expansion of metropolitan areas between 1960 and the time the cohorts reached old age (during the 1980s). The 1960 IPUMS does not include the name of the metropolitan area in which one resides; we list the most popular destination cities in the Duke SSA/Medicare data.

[10]More generally, migrants are not always positively selected. For example, Abramitzky, Boustan, and Eriksson (2012) found that Norwegian immigrants to the United States in the late 1800s were negatively selected, in terms of wealth and occupational status.

[11]The source of our main analysis sample, the Duke SSA/Medicare data, does not contain education or earnings, hence our use of census data.

[12]We cannot estimate the causal effect of migration on earnings in our data, because we do not have earnings and one of our instruments in the same data set.

Table 2. Earnings differences between migrants and nonmigrants.

	1960 Ages 33–44 (1)	1970 Ages 43–54 (2)	1980 Ages 53–64 (3)
Migrant	0.447	0.367	0.344
	(0.035)	(0.024)	(0.021)
Percentage effect	0.564	0.443	0.411
R^2	0.140	0.167	0.094
Observations	1701	3274	6436

Note: Sample includes non-Hispanic white males in the labor force born in the Dakotas and Montana 1916–1927. For migrant definition, see text. Dependent variable is the natural log of earnings. Regressions control for education and age using a cell estimator. Robust standard errors in parentheses. Percentage effect is the coefficient on migrant status exponentiated. Source: Authors' calculations using IPUMS of the 1960, 1970, and 1980 U.S. Census (Ruggles et al. (2010)).

3. The role of information in the migration decision

Our intention is to set up an empirical design in which we can plausibly identify effects of migration on lifelong health—a design that credibly deals with potential selection into migration. As discussed in the Introduction, our strategy is based on town-level quasi-random variation in access to relevant information about migration destination locations. Previous studies have discussed the role of information for individuals considering rural-urban migration (see the discussion in Lucas (1977)). In the Appendix, available in the Online Supplementary Material (Johnson and Taylor (2019)), we develop a simple model in which improved information about potential locations increases rural-urban migration, and in which migrants tend to be "high ability" individuals. However, our model only demonstrates one plausible pattern of selection; the role of information access in migration, and the nature of selection into migration, is largely an empirical matter. In this section, we develop some plausible instruments and find that they are strongly correlated with migration out of the Northern Great Plains.

3.1 *Information and migration in the Northern Great Plains*

During the era we study, rapid growth of industrial centers in the Midwest and on the West Coast resulted in rising wages, while at the same time the industrialization of agriculture in the Great Plains resulted in declining economic opportunities in the Dakotas and Montana. Knowledge about the opportunities outside the Great Plains presumably varied substantially among young people growing up in the 1930s in towns scattered across the Dakotas and Montana. Our empirical strategy relies on two sets of instruments that are likely related to town-level information that would have been available to young people born in these towns.

3.1.1 *The role of railroads in the Northern Great Plains*   Our first set of instruments is based on the location of railway lines: those extant in 1900, and those constructed post-1900. Circa 1900, both the level and direction of out-migration of individuals born in the Northern Great Plains were doubtless shaped in part by railway lines. Thus by the 1930s, long-time railway towns were plausibly high information towns, in which many

residents had friends and relatives who had previously migrated to far-way destinations. Towns on these train lines would also have had many travelers from other states, and the ready access to train travel facilitated visits outside the Northern Great Plains. Towns on railways built after 1900 shared the same characteristics as towns on pre-1900 railway lines, but these towns were somewhat less well connected. Towns with no rail services were the least connected.

Empirically, as we show shortly, rates of out-migration were indeed highest in towns located on pre-1900 rail lines, somewhat lower from towns located on post-1900 rail lines, and lower yet on towns with no rail service.

3.1.2 *The railway mail service*   Railways were also important for the flow of postal mail—a crucial means of communication during the 1920s and 1930s. Much of the information about life prospects in distant locations arrived in rural towns via the mail. For example, in a 1922 letter to North Dakota's *Ward County Independent*, three young men from Minot, ND described a journey to California: "Before leaving Minot . . . some folks tried to frighten us by telling us that we would not get a job in California. We were offered $80 a month and feed the next day after we got here. It looks to us now as tho [sic] we might remain here for the next 50 years." In a letter published in *The Producers News* of Plentywood, MT in 1921, Maurice Finn, a former resident of nearby Dooley, MT, extols the virtues of San Diego, calling it "the finest place to live in on this earth," where "the 4th of July and the 25th of December are alike as far as climate is concerned." To readers in Plentywood, where the average low on December 25 is 4°F (-16°C) and average high on July 4 is 82°F (28°C), this description would likely have been quite appealing.

All households in the Northern Great Plains had access to postal mail, but the speed and efficiency of postal deliveries varied substantially across communities. During this era, most mail was transported via the Railway Mail Service (RMS), which served nearly every railway town in the country. Postal transportation clerks sorted mail en route on special Railway Post Office (RPO) cars, for high volume routes, as a way of improving the delivery time of "time-value mail," including magazines and newspapers (Long (1951)). RPO cars were usually part of passenger trains, and the volume of mail carried determined the size of the RPO car assigned to a route. There were three sizes of RPO cars, 15- and 30-foot apartments and full 60-foot cars, with the remainder of the standard 60-foot car length in the smaller versions taken up by passengers.[13] On small routes where the volume did not justify even a 15-foot car, mail was transported via "closed pouch service," meaning trains picked up mail sacks, but there was no dedicated rail car space for mail and definitely no sorting en route.

In the most remote locations in the Great Plains, mail arrived via the pouch service to the nearest railway town, which in some cases would have been quite slow, especially if the local post office adopted a "hold until full" policy of pouch delivery.[14] On the other hand, towns located along the highest-volume RMS mail delivery routes had rapid and

---

[13]The exception to this were all-RPO trains, which only operated on very high volume routes, such as between Pittsburgh and New York.

[14]Another issue for more remote towns is that train lines to those towns often did not operate regularly during the winter, in which case mail delivery was much slower.

reliable mail delivery from even distant cities. In these latter towns, residents (or town libraries) could subscribe to newspapers from San Francisco or Seattle and expect timely delivery.

Fortunately, we can construct town-level measures of the flow of railway mail carried by the RMS through the Dakotas and Montana, which provide us with a good metric for the reliability and timeliness of mail delivery. Figure 3 shows railway mail flows in Montana, North Dakota, South Dakota, Minnesota, and parts of Iowa, Michigan, Nebraska, Wyoming, and Wisconsin in the year 1924. The map indicates the linear rail car feet leased per week by the RMS on the railway routes active in that year, divided into six size categories.[15] These six categories of mail capacity form our set of instruments; we view them as useful proxies for the extent to which towns were connected to the rest of the country via the mail service.[16]

Not surprisingly, larger towns were more likely than smaller towns to be located on a route with higher mail capacity, but town population size was not perfectly correlated with mail-car capacity. For example, one of the region's largest towns, Great Falls, MT (population 24,121 in 1920), was served by a modest line, assigned 270–510 feet of space, while New Rockford, ND (population 2111) was on a line with 600–720 linear feet of mail.

The six categories of mail-car capacity and the fraction of our analysis sample in each category are shown in Table 3, as well as the fraction of the sample migrating out of the Northern Great Plains in each category. Mail-car capacity strongly predicts migration, as fraction migrant increases with mail-car capacity, from 0.49 for the smallest category (town not located on a railroad) to 0.65 for those living in a town in the largest capacity category.

In sum, we believe that we we have two reasonable sets of instruments that differentiate the extent to which towns were connected to the rest of the country in terms of information availability. And as an empirical matter, they do predict out-migration. The validity of these instruments relies on the assumption that these town-level variables have no other influence on older-age mortality outside of their effect on the probability of migration. We discuss threats to this assumption in Section 5.5, after presenting our primary empirical results.

4. Data

We require data that identifies individuals' location of birth and location in old age, *and* provides the age of death (at least among the older population). Fortunately, we have access to an excellent proprietary data source for our purposes.

Our main source of data is the Duke SSA/Medicare data set. These data merge the Master Beneficiary Records from the Supplementary Medical Insurance Program (Medicare Part B) to records from the Numerical Identification Files (NUMIDENT) of the Social Security Administration (SSA). The data contain over 70 million records over the

---

[15]This figure is a direct reproduction of Figure 20 in *America's Northern Heartland*, by John R. Borchert. A geographer specializing in the Upper Midwest, Borchert collected mail-car capacity information as part of his study of the economic and historical geography of the region.

[16]Our mail capacity instrument is correlated with newspaper delivery, independent of town population. Using data from Gentzkow, Shapiro, and Sinkinson (2016), towns on railroads with higher mail-car capacity are more likely to receive an out-of-state newspaper, controlling for population and state.

Figure 20. Flow of Railway Mail, 1924. The great preponderance of intercity communication moved by railway mail. The Twin Cities were the main focus along the northern transcontinental corridor and a powerful competitor of Chicago across northern Iowa and eastern South Dakota. Source: note 40.
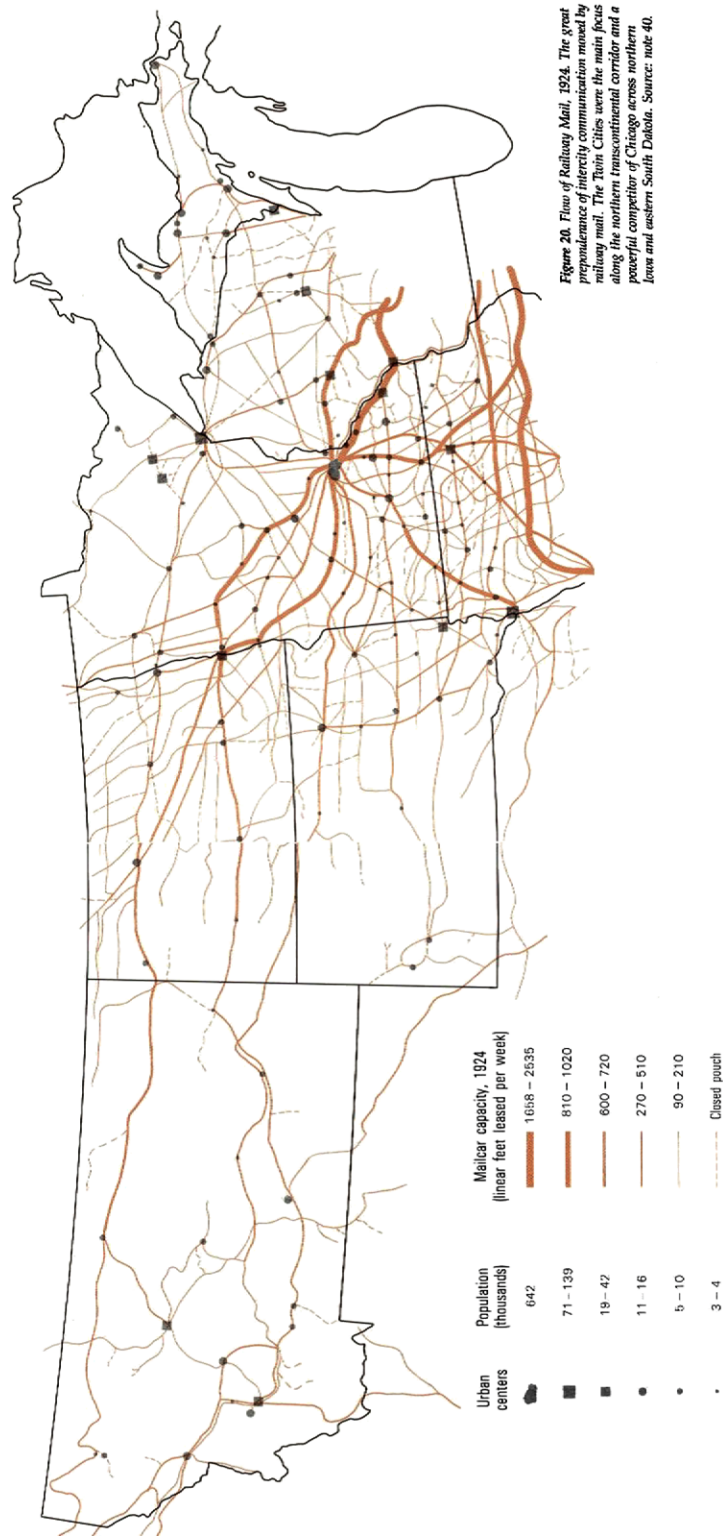
FIGURE 3. Flow of railway mail, 1924. Note: This figure is a direct reproduction of Figure 20 in Borchert (1987). Used by permission of the University of Minnesota Press from America's Northern Heartland: An Economic and Historical Geography of the Upper Midwest by John R. Borchert. Copyright 1987 by the University of Minnesota.

TABLE 3. Distribution of mail-car capacity and fraction migrant.

Mail-car capacity	Observations	Sample percentage	Fraction migrant
Not on railroad	17,640	5.34	0.49
Closed pouch	21,283	6.44	0.52
90–210	136,112	41.19	0.51
270–510	108,320	32.78	0.55
600–720	39,488	11.95	0.58
810–1020	7585	2.30	0.65
Total	330,428	100.00	0.54

Note: Sample includes white individuals born in the Dakotas and Montana 1916–1927. Mail-car capacity measured in linear feet per week. For migrant definition, see text.
Source: Authors' calculations using Duke SSA/Medicare data and mailcar capacity information from Borchert (1987).

period 1976–2001, covering a very high proportion of the population aged 65 years and older. Using counts from the 2000 Census, we estimate that the total coverage rate for our cohorts of interest is 92%. Because enrollment requires proof of age, the age validity of the records is high compared with other data sources for the U.S. elderly population. In addition to race, sex, and age, our data include zip code of the place of residence, exact date of death, and detailed place of birth information. Specifically, the data include either town and state of birth or town, county and state of birth for all U.S.-born respondents.

Our sample consists of white individuals born in the states of North Dakota, South Dakota, and Montana. Due to decreased coverage rates for earlier birth cohorts, we limit our analysis sample to birth cohorts 1916–1927. Our sample has 330,428 individuals. To be in this sample, a record must have nonmissing birth town information, as our instruments are defined at the town level, as well as complete information on place of residence at age 65 to define migrant status. We lose 30,710 individuals because we are missing birth town, and 749 individuals for whom we have missing residence information at age 65. The total match rate for our sample is therefore 91.3%.[17] To be included in our sample, an individual must be a Medicare Part B recipient, which for most individuals, requires reaching the age of 65. Thus we evaluate mortality post age 65. The implications of this age requirement for the validity and interpretation of our results are discussed below.

While the Duke SSA/Medicare data cover a very high percentage of our cohorts of interest and have the necessary information information on town of birth and residence in old age, they do not contain much else in the way of individual characteristics, for example, education, earnings, or marital status. We find it useful to turn to other data sources to undertake additional analyses. We use the Integrated Public Use Samples (IPUMS) of the U.S. Census (Ruggles et al. (2010)) to compare migrants and nonmigrants on characteristics that do not appear in the Duke dataset. We also make use of Vital Statistics Multiple Cause of Death Data from the National Center on Health Statistics to compute

---

[17]Migrants are matched at slightly higher rates than nonmigrants (1-percentage point) and those born in later cohorts are matched at slightly higher rates. For more information on the Duke SSA/Medicare dataset, see Black et al. (2015).

mortality rates for our sample both pre- and post-age 65, and to investigate causes of death.

5. Empirical approach and results

The framework for our estimation approach is a generalized Roy model. As described in Heckman and Vytlacil (1999), the approach combines expressions for potential outcomes with a discrete choice latent variable model for selection into treatment. We briefly describe the model as it applies to our context.[18]

In our case "treatment" is migration out of the Northern Great Plains. We are interested in comparing longevity—measured as an indicator for 10-year survival from age 65 to 75—of migrants to nonmigrants. Let Y_1 be survival if an individual has migrated, and Y_0 if not. We can model these potential outcomes as

$$Y_0 = \mu_0(X) + U_0,$$
$$Y_1 = \mu_1(X) + U_1, \tag{1}$$

where X is a vector of observed characteristics and $E[U_1|X = x] = E[U_0|X = x] = 0$.

An individual migrates if the expected payoff from migration exceeds some threshold. Let this expected gain from migration be D^*, and let D be an indicator for migration. The choice equation can be written,

$$D^* = \mu_D(X, Z) - V, \tag{2}$$

with

$$D = \begin{cases} 1 & \text{if } D^* \geq 0, \\ 0 & \text{otherwise.} \end{cases}$$

The expected gain from migration, $\mu_D(X, Z)$, is determined by the observable variables, X, and instruments, Z, which in our case are town-level variables related to the availably of information about opportunities in potential destination areas. V is an unobserved continuous random variable which denotes an individuals "distaste" for migration (to use the terminology of Cornelissen et al. (2016)). High values of V correspond to characteristics, such as "low latent ability," which make individuals more likely to be "stayers." An individual migrates if the expected gain from migration is greater than the preference for staying, $\mu_D(X, Z) \geq V$.

For implementation of the MTE model, it is helpful to transform this inequality as follows. Define F_V to be the CDF of V, let $F_V(V) = U_D$ and let $F_V(\mu_D(X, Z)) = P(X, Z)$. Then we can express the indicator for migration as

$$D = \mathbb{1}\big(P(X, Z) > U_D\big), \tag{3}$$

---

[18]Cornelissen, Dustmann, Raute, and Schönberg (2016) provide an intellectual history of the development of the MTE framework. Key contributions include Björklund and Moffitt (1987), who introduced the concept of MTEs, Heckman and Vytlacil (1999, 2001, 2005), and Brinch, Mogstad, and Wiswall (2017).

where $P(X, Z)$ is a propensity score and U_D is a variable normalized to be uniformly distributed over $[0, 1]$.[19]

For each value of the stayer index U_D, we can define the marginal treatment effect of migrating out of the Northern Great Plains on longevity as

$$\text{MTE}(X = x, U_D = u_D) = E(Y_1 - Y_0 | X = x, U_D = u_D). \qquad (4)$$

For each $U_D = u_D$, there is some propensity score $P(X, Z) = p$, determined by observable characteristics X and the value of the birth-town instrument Z, at which individuals are indifferent between (i.e., "at the margin of") migrating and staying in the Northern Great Plains. One can therefore replace the expression $U_D = u_D$ in (4) with $U_D = p$. The MTE can then be thought of as the treatment effect for an individual whose unobserved stayer index U_D is exactly equal to their observed propensity score $P(X, Z) = p$.

The standard identifying assumptions for estimation of the MTE in this framework are exogeneity, (U_0, U_1, U_D) is independent of $(Z|X)$, and relevance, Z does not enter trivially into $P(X, Z)$.[20] In the ideal case, the instrument Z would be continuous and contain enough variation to generate common support over the full range of U_D (zero to one) for treated and untreated observations conditional on $X = x$. Without parametric assumptions, the MTE is only identified over the range of the propensity scores for each X. Unfortunately, as is typical in many empirical applications, our instruments do not provide the variation necessary to yield the full support propensity scores, and so we can only recover the MTE over a more limited range of U_D.

We estimate the first stage using a logistic regression. We condition on X in a linear fashion and estimate our propensity scores by regressing an indicator for migration out of the Northern Great Plains on X, a vector of observable characteristics that includes full interactions of birth year, sex, and state of birth, and cubics in county population, county average household size, fraction of county households that are farm households, and fraction of county households that own a radio, and the first degree interaction of the four continuous regressors. Our two town-level instruments generate three possibilities for Z: the railroad instrument only (expressed as dummies for a town being on a pre- or post-1900 railroad, with not on a railroad the excluded category), the mail flow instrument only (a vector of dummy variables for each of the six mail-car capacity categories, again with the excluded category of not on a railroad), and both the railroad and mail flow instruments together. We include county-level control variables.[21] From these logistic regressions, we recover the predicted values for each individual, that is, the estimated propensity scores.

---

[19]Our notation follows Heckman and Vytlacil (1999). Note that U_D is an unobserved random variable that can be interpreted as a "stayer index" in our context (high values indicate a high probability that an individual stays in the Northern Great Plains), and is distinct from the unobserved random variables U_1 and U_2 which appear in (1).

[20]Vytlacil (2002) shows the equivalence of the latent index model shown above and the standard IV assumption of monotonicity (Angrist, Imbens, and Rubin (1996)).

[21]Inclusion is necessary because in tests described by Goldsmith-Pinkham, Sorkin, and Swift (2017) our county-level controls prove to be correlated with average county mail-car capacity. See the Appendix in the Online Supplementary Material.

Figure 4 shows the distribution of propensity scores for migrants and nonmigrants using the railroad instrument (Panel A), the mail-car capacity instrument (Panel B), and both instruments combined (Panel C). The distributions using all three instrument combinations are very similar. Each generates propensity scores ranging from approximately $p = 0.35$ to $p = 0.7$. The area of common support across migrants and nonmigrants spans this entire range, with the nonmigrant distribution lying slightly to the left of that for migrants. However, all distributions have a strong central tendency, with much of the density concentrated around $p = 0.52$. We have very sparse samples towards the upper and lower ends of our propensity score range.

Before proceeding to our MTE estimation, we first report Instrumental Variable (IV) estimates of the effect of migration out of the Northern Great Plains on older-age longevity. We describe in detail the relationship between the MTE and IV estimates in Section 5.3 below.

5.1 *Instrumental variable estimates*

We first estimate the aggregate effect of migration out of the Northern Great Plains on older-age longevity using traditional two-stage least square (2SLS) IV methods.

We estimate

$$Y = \beta_0 + \beta_1 D + X\beta_2 + \varepsilon, \tag{5}$$

where D is an indicator for migration and X is the vector of observable characteristics described above. Table 4 gives OLS estimates in column (1), and in columns (2), (3), and (4) we provide IV estimates using, respectively, our mail-car capacity instruments, our railway town instruments, and both sets of instruments together. The OLS coefficient on D (the migration variable) is -0.012.[22] In all three cases, the IV estimate is negative and statistically significant; the IV estimate using mail-car capacity instruments is approximately -0.09, that using railway town is approximately -0.12, and that using both is also -0.09. (The post-1900 instrument is dropped from this specification as all towns on railroads built after 1900 are closed pouch lines.) Given that the mean 10-year survival probability of approximately 0.81, we are estimating substantial negative effects of migration on longevity.

We also include a fifth column in Table 4 that uses the propensity score as an instrument. This propensity score is estimated using both the railroad and mail-car capacity instruments and the same observable controls as we include in the IV, but in contrast to the linear first stage of the IV, uses a logit. We report the IV results using this propensity score to provide direct comparison to our MTE results. The results using the propensity score instrument are nearly identical to those using 2SLS (approximately -0.09).

As for the first stage estimates, results are exactly as expected. The mail-car capacity variables predict out-migration with increasing strength (given an excluded category,

---

[22]We have excluded those missing birth town data, because we cannot form town-level instruments for these individuals. Including them in the OLS regression leaves the results virtually unchanged (the estimated coefficient is -0.013, and the standard error is 0.001).
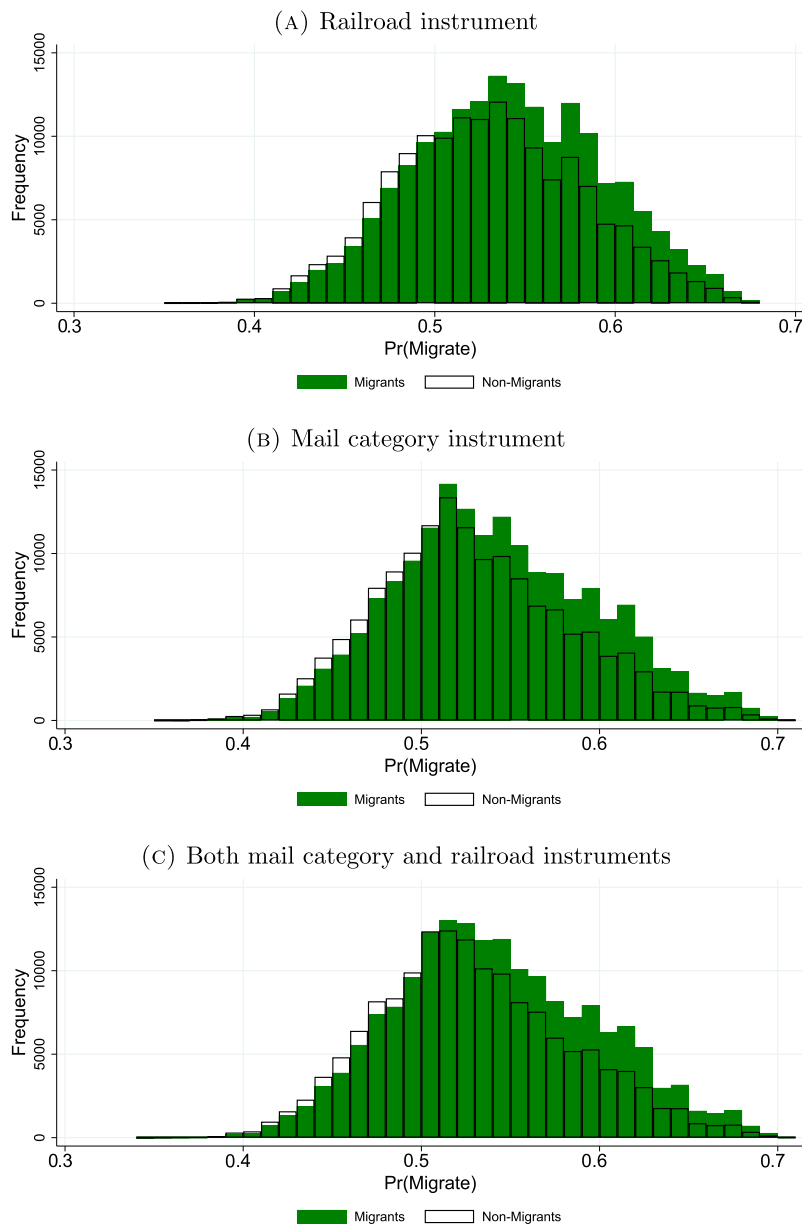
FIGURE 4. Propensity score distributions using different instruments. Notes: Propensity scores estimated using the predicted values of a logit regression of an indicator for migration out of the Northern Great Plains on one of three town-level instruments: six categories of mail capacity (Panel A), three categories of railroad type (Panel B), and both instruments combined (Panel C). Sample includes white individuals born 1916–1927 in the Dakotas and Montana who lived to age 65. Included in all specifications are indicators for birth cohort × sex and state of birth, a cubic in 1930 county population, mean county household size in 1930, fraction of county households that are farm households in 1930, and fraction of county households owing a radio in 1930. Source: Authors' calculations using Duke SSA/Medicare data, mail-car capacity information from Borchert (1987), and county characteristics from the 1930 U.S. Census ((Ruggles et al., 2010)).

TABLE 4. IV estimates of the effect of migration out of the Northern Great Plains on older-age longevity.

	OLS (1)	Railroad instruments (2)	Mailcar capacity instruments (3)	Both instruments (4)	Propensity score as instrument (5)
Migrant	−0.012	−0.118	−0.089	−0.094	−0.093
	(0.001)	(0.048)	(0.033)	(0.033)	(0.033)
Dependent variable mean	0.806	0.806	0.806	0.806	0.806
First Stage					
Closed pouch			0.041	0.039	
			(0.011)	(0.011)	
90–120			0.043	0.040	
			(0.008)	(0.009)	
270–510			0.052	0.045	
			(0.010)	(0.011)	
600–720			0.107	0.098	
			(0.015)	(0.016)	
810–1020			0.190	0.180	
			(0.021)	(0.023)	
Pre-1900 railroad		0.071		0.012	
		(0.009)		(0.008)	
Post-1900 railroad		0.043			
		(0.008)			
Partial *F* Statistic		33.07	24.97	22.89	–
Dependent Variable Mean		0.536	0.536	0.536	0.536
Observations	330,428	330,428	330,428	330,428	330,428

Note: Standard errors clustered on town of birth in parentheses. Sample includes white individuals born 1916–1927 in the Dakotas and Montana who lived to age 65. Dependent variable is indicator for survival to age 75. Included in all specifications are indicators for birth cohort × sex and state of birth, a cubic in 1930 county population, mean county household size in 1930, fraction of county households that are farm households in 1930, and fraction of county households owing a radio in 1930. Partial *F* statistic is that for the vector of instruments in the first stage of each specification.
Source: Authors' calculations using Duke SSA/Medicare data, mail-car capacity information from Borchert (1987), and county characteristics from the 1930 U.S. Census (Ruggles et al. (2010)).

not on a railroad), and the first stage partial *F* statistic is 24.97. Similarly, birth in a pre-1900 railway town is strongly associated with out-migration, and birth in a post-1900 railway town somewhat less so (with again the excluded category, not on a railroad), and again we have a large partial *F* statistic (33.07). The partial *F* statistic using both instruments is 22.89.

Table 5 repeats this analysis separately for men and women, using both instruments. Results are similar if we use the railroad instrument and mail capacity instrument separately. The estimated IV effect for both male and female migrants is negative, and they are similar in magnitude, but the estimate for men is not statistically significant.

TABLE 5. IV estimates by sex.

	Men (1)	Women (2)
Migrant	−0.063	−0.128
	(0.051)	(0.032)
Dependent variable mean	0.753	0.854
First Stage		
Closed pouch	0.039	0.061
	(0.014)	(0.013)
90–120	0.048	0.056
	(0.012)	(0.011)
270–510	0.055	0.059
	(0.011)	(0.011)
600–720	0.111	0.109
	(0.016)	(0.015)
810–1020	0.197	0.187
	(0.023)	(0.021)
Pre-1900 railroad	0.012	0.013
	(0.009)	(0.008)
Partial *F* Statistic	18.44	20.54
Dependent Variable Mean	0.524	0.546
Observations	158,132	172,296

Note: Specifications estimated using 2SLS and mail capacity instrument. Standard errors clustered on town of birth in parentheses. Sample includes white individuals born 1916–1927 in the Dakotas and Montana who lived to age 65. Dependent variable is indicator for survival to age 75. Included in all specifications are indicators for birth cohort × sex and state of birth, a cubic in 1930 county population, mean county household size in 1930, fraction of county households that are farm households in 1930, and fraction of county households owing a radio in 1930. Partial *F* statistic is that for the vector of instruments in the first stage of each specification.
Source: Authors' calculations using Duke SSA/Medicare data, mail-car capacity information from Borchert (1987), and county characteristics from the 1930 U.S. Census (Ruggles et al. (2010)).

5.2 *MTE definition and estimation*

We use two different approaches to estimate MTEs of migration out of the Northern Great Plains on longevity: the Local Instrumental Variables (LIV) approach of Heckman and Vytlacil (1999) and the "separate approach" introduced by Heckman and Vytlacil (2007) and further developed by Brinch, Mogstad, and Wiswall (2017). In our main specification, the first stage is estimated using both the railroad and mail capacity instruments.

As is common in the MTE literature, we assume additive separability between the unobserved and observed components in the outcome equations for Y_0 and Y_1.[23] We also assume linearity of the outcome equations in X:

$$
\begin{aligned}
E(Y_0|X = x, U_D = u_D) &= x\beta_0 + E(U_0|U_D = u_D), \\
E(Y_1|X = x, U_D = u_D) &= x\beta_1 + E(U_1|U_D = u_D).
\end{aligned}
\tag{6}
$$

---

[23]As noted by Brinch, Mogstad, and Wiswall (2017), this condition is implied by, but does not imply, (U_0, U_1, U_D) independent of (X, Z).

This assumption means that the relationship between the error terms in the outcome equation and error term in the choice equation does not depend on X. This is useful for estimation as it would be empirically difficult to estimate a more general model without a huge number of observations. The MTE is therefore also composed of additively separable unobserved and observed components:

$$\text{MTE}(X = x, U_D = u_D) = x(\beta_1 - \beta_0) + E(U_1 - U_0 | U_D = u_D). \tag{7}$$

In our estimation, both the LIV and separate estimation approaches rely on the assumptions of linearity in X and additive separability;[24] the difference between them lies in the method employed to estimate the unobserved and observed components of the MTE.

In combination with the conditional independence assumption of IV, the assumption of additive separability restricts the shape of the MTE curve to be independent of X, with the exception of the intercept, which can vary with X.

5.2.1 *Local instrumental variables*　Given the assumptions of linearity in X, conditional independence, and additive separability, we have the following outcome equation that follows from (6):

$$E[Y | X = x, P(Z) = p] = x\beta_0 + px(\beta_1 - \beta_0) + K(p), \tag{8}$$

where $K(p)$ is a nonlinear function of the propensity score.[25] Heckman and Vytlacil (1999) show the MTE is identified as the derivative of the outcome equation with respect to p:

$$\text{MTE}(X = x, U_D = p) = x(\beta_1 - \beta_0) + \frac{\partial K(p)}{\partial p}. \tag{9}$$

This forms the foundation of the LIV method. We employ two different methods to estimate $K(p)$. The first specifies $K(p)$ in a parametric fashion as a polynomial in p (the "parametric LIV" estimator), and the second involves using local polynomial regression. To implement the parametric LIV, we estimate (8), recover the parameter estimates, and calculate the MTE using (9). The steps we employ to estimate the semiparametric LIV are outlined in Appendix B.1 of Cornelissen et al. (2016).

Our LIV MTE results are shown in Figure 5. We show results using three different specifications for the parametric and semiparametric approaches: polynomials of degree 1, 2, and 3 for the parametric (these are the degrees of $K(p)$, the degree of the MTE in p is one lower), and bandwidths of 0.05, 0.075, and 0.1 for the semiparametric. All six approaches (with the exception of the flat degree 1 parametric) are downward sloping, implying that migrants of lower ability (and higher U_D) experienced a higher longevity penalty to migration out of the Northern Great Plains. While we show parametric results

---

[24]The additive separability assumption is weaker in the Local IV framework, where only the MTE needs to be additively separable. In the separate estimation approach, both conditional outcome functions are assumed to be additively separable.

[25]In particular, $K(p) = pE(U_1 | U_D < p) + (1 - p)E(U_0 | U_D \geq p)$. It can be shown that $\frac{\partial K(p)}{\partial p} = E(U_1 | U_D = p) - E(U_0 | U_D = p)$.
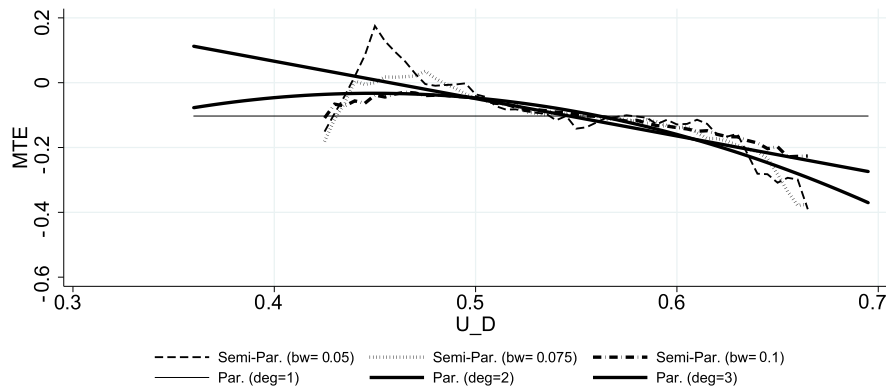
Figure 5. MTE estimates, local instrumental variables approach. Note: MTEs estimated using LIV approach and both instruments. "Par" (parametric) specifications use a polynomial of indicated degree in p for $K(p)$; "Semi-Par" (semiparametric) estimates $K(p)$ using a local quadratic regression with the indicated bandwidth. Sample includes white individuals born 1916–1927 in the Dakotas and Montana who lived to age 65. Included in all specifications are indicators for birth cohort $\times$ sex and state of birth, a cubic in 1930 county population, mean county household size in 1930, fraction of county households that are farm households in 1930, and fraction of county households owing a radio in 1930. Source: Authors' calculations using Duke SSA/Medicare data, mail-car capacity information from Borchert (1987), and county characteristics from the 1930 U.S. Census (Ruggles et al. (2010)).

over our full area of common support, we do not report semiparametric results at the edges of this support area (below approximately $p = 0.42$ and above $p = 0.67$) as the low data support in these regions leads to noisy estimates. In general, the parametric and semiparametric results are very similar, with the parametric of degree 3 and semiparametric of bandwidth 0.01 yielding nearly identical results. All approaches produce very similar results in the area of highest data density (around $p = 0.53$); for individuals with propensity scores in this region, the effect of migration out of the Northern Great Plains on survival to age 75 is approximately -0.08.

5.2.2 *Separate approach*   Define the following to be the conditional expectations of U_0 and U_1:[26]

$$k_0(p) = E[U_0|U_D = p],$$
$$k_1(p) = E[U_1|U_D = p].$$
(10)

The MTE can therefore be expressed as

$$\text{MTE}(X = x, U_D = p) = \mu_1(x) - \mu_0(x) + k_1(p) - k_0(p). \tag{11}$$

The key difference between the separate approach and LIV is we estimate each of the conditional outcome equations in (6) using treated and untreated observations sepa-

---

[26]We continue to assume additive separability between X and U, these expectations are functions of p only, not x.

rately. We again use $\mu_0(x) = x\beta_0$ and $\mu_1(x) = x\beta_1$, and estimate the following outcome equations:

$$E[Y_0|X = x, U_D = p] = x\beta_0 + K_0(p),$$
$$E[Y_1|X = x, U_D = p] = x\beta_1 + K_1(p). \tag{12}$$

Brinch, Mogstad, and Wiswall (2017) show that taking the derivatives of $K_1(p)$ and $K_0(p)$ and rearranging yields

$$k_1(p) = p\frac{\partial K_1(p)}{\partial p} + K_1(p) \tag{13}$$

and

$$k_0(p) = -(1 - p)\frac{\partial K_0(p)}{\partial p} + K_0(p). \tag{14}$$

Using these expressions, we can recover estimates of $k_1(p)$ and $k_0(p)$ and form the MTE.

In sum, the estimates of β_0 and the parameters of $k_0(p)$ are identified from regression of Y_0 on X and some function of the propensity score for untreated individuals (nonmigrants), and β_1 and $k_1(p)$ from a similar regression of Y_1 using treated observations (migrants) only.

As with LIV, the main difference between the parametric and semiparametric variants of the separate approach lies in the specification of $K_0(p)$ and $K_1(p)$. The parametric separate approach uses degree 1, 2, and 3 polynomials in p for $K_0(p)$ and $K_1(p)$. As noted by Brinch, Mogstad, and Wiswall (2017), when estimating the separate approach parametrically the degree of $K_d(p)$ is equal to the degree of p in the MTE, rather than one higher as in the local IV. The semiparametric estimation employs a local polynomial regression using the same bandwidths as before. We describe the steps to implement each approach in detail in the Appendix available in the Online Supplementary Material.

Separate approach MTE estimates are shown in Figure 6. Results are nearly identical to those using the LIV approach.

Standard errors for select specifications are shown in Figure 7. As expected, standard errors are narrower for the parametric estimation, especially near the tails where the support of the propensity score distribution is thin. In this application, the separate approach also produces smaller standard errors than the local IV. This may be due to the stronger parametric assumptions associated with the separate approach. In the separate approach, both conditional outcome functions, as shown in (6), are assumed to take a specific parametric structure. In contrast, in the local IV, we only assume that the *difference* between the two conditional outcome functions (the MTE) has a specific parametric structure. In general, standard errors are fairly wide, so we should interpret any heterogeneity in treatment effects as suggestive. In the parametric separate approach, we can reject that treatment effects are zero only for individuals with higher U_D.
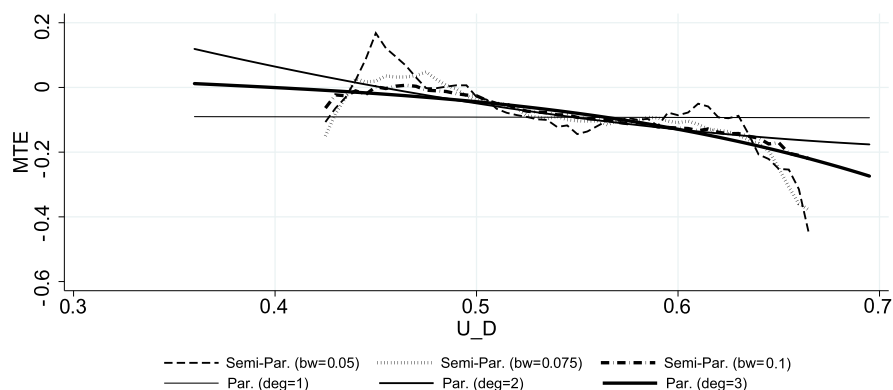
Figure 6. MTE estimates, separate approach. Notes: MTEs estimated using separate approach and both instruments. "Par" (parametric) specifications use a polynomial of indicated degree in p for $K_0(p)$ and $K_1(p)$; "Semi-Par" (semiparametric) estimates $K_0(p)$ and $K_1(p)$ using a local quadratic regression with the indicated bandwidth. Sample includes white individuals born 1916–1927 in the Dakotas and Montana who lived to age 65. Included in all specifications are indicators for birth cohort $\times$ sex and state of birth, a cubic in 1930 county population, mean county household size in 1930, fraction of county households that are farm households in 1930, and fraction of county households owing a radio in 1930. Source: Authors' calculations using Duke SSA/Medicare data, mail-car capacity information from Borchert (1987), and county characteristics from the 1930 U.S. Census (Ruggles et al. (2010)).

5.3 Relationship between IV and MTE

Heckman, Urzua, and Vytlacil (2006) shows that the estimates from Instrumental Variables are a weighted average of the Marginal Treatment Effects which depend on the choice of the instrument used in the first stage. The IV weights for the railroad and mail category instruments, as well as both instruments and the propensity score, are shown in Figure 8.

For any choice of instrument most of the weight is placed near the center of the U_D distribution. Weights are nearly identical using the mail instrument, using both instruments or using the propensity score. This makes sense intuitively, as the railroad instrument has only a small impact when using both instruments. The propensity score is calculated over a range of propensity scores for which the logistic regression closely approximates a linear regression.[27]

---

[27]When the railroad instrument is used in the first stage, the weights are shifted slightly to the left. Based on the MTE results, this should move the IV estimates towards zero, which does not appear to be the case. The difference between the IV estimates when using the rail and rail instruments does not appear to be explained by different MTE weights. However, this difference is not statistically significant.

(A) MTE with Standard Errors: Local IV Parametric
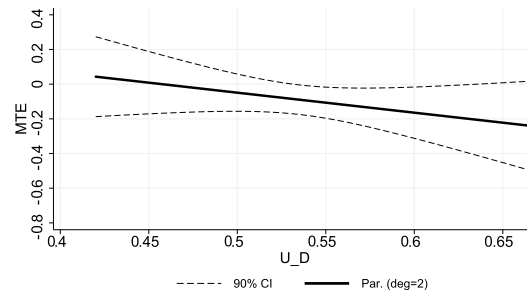
(B) MTE with Standard Errors: Local IV Non-parametric
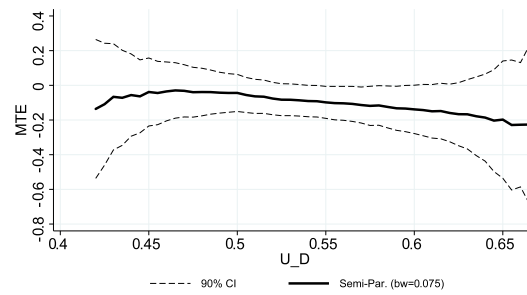
(C) MTE with Standard Errors: Separate Approach Parametric
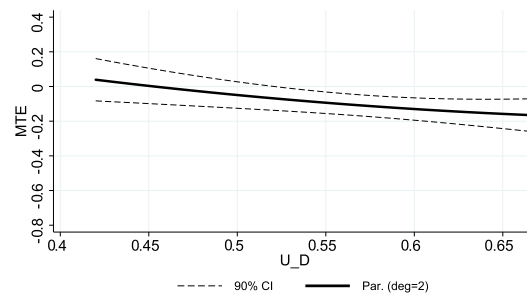
(D) MTE with Standard Errors: Separate Approach Non-parametric
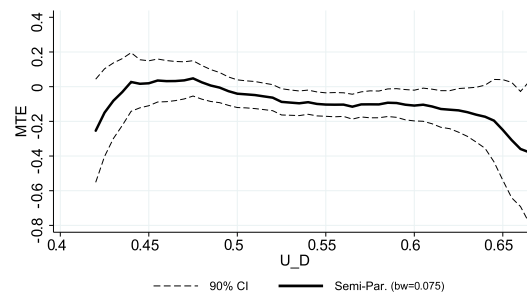
Figure 7. Standard errors on MTE estimates. Note: Standard errors calculated using 100 bootstrap replications with draws taken at the town of birth level. Parametric estimates use degree 2 polynomials in estimation. Semiparametric estimates use $bw = 0.075$. Source: Authors' calculations using Duke SSA/Medicare data, mail-car capacity information from Borchert (1987), and county characteristics from the 1930 U.S. Census (Ruggles et al. (2010)).
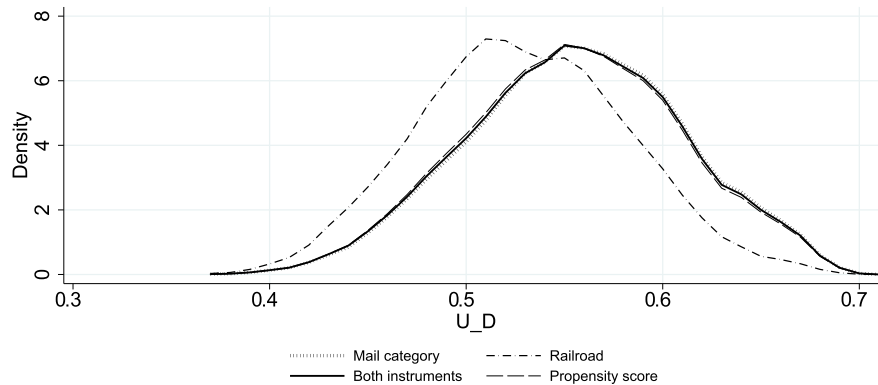
FIGURE 8. IV weights. Notes: Figure shows the MTE weights for each IV estimates using the four different instruments. Derivation and formulas of the weights are given in Heckman and Vytlacil (2005). Also see Appendix C of Cornelissen et al. (2016) for instructions on calculating the IV weights. Sample includes white individuals born 1916–1927 in the Dakotas and Montana who lived to age 65. Included in all specifications are indicators for birth cohort × sex and state of birth, a cubic in 1930 county population, mean county household size in 1930, fraction of county households that are farm households in 1930, and fraction of county households owing a radio in 1930. Source: Authors' calculations using Duke SSA/Medicare data, mail-car capacity information from Borchert (1987), and county characteristics from the 1930 U.S.Census (Ruggles et al. (2010)).

5.4 *Bounding the average treatment effect and heterogeneity in observables*

We estimate the average treatment effect (ATE) over the range of common support (approximately $0.35 < p < 0.71$) by taking the average MTE for all individuals.[28] When we do so, using a degree-3 parametric specification, we estimate an ATE of -0.095.

As for the ATE over the full unit interval, we can make some make progress by estimating *bounds*, if we are willing to make some assumptions about values of Y_1 and Y_0 outside of our region of support. Consider nonmigrants whose propensity score is the highest value we observe ($p_{\max} = 0.71$). As these individuals are not treated, they must have values of U_D greater than $p_{\max}$. Similarly, migrants with a propensity score at the bottom of our observed range ($p_{\min} = 0.35$) must have values of U_D below 0.35. We can therefore identify:

$$
\begin{aligned}
E[Y_0|X = x, U_D > p_{\max}] &= E\big[Y|X = x, \hat{P}(Z) = p_{\max}, D = 0\big], \\
E[Y_1|X = x, U_D < p_{\min}] &= E\big[Y|X = x, \hat{P}(Z) = p_{\min}, D = 1\big].
\end{aligned}
\tag{15}
$$

These are comparable to estimating the average conditional outcome for "always" and "never takers" in a LATE framework.[29] We now have estimates of Y_0 for U_D above our

---

[28]This is equivalent to integrating the MTE with respect to u_D from 0.35 to 0.71 for the population, and dividing by $(0.71 - 0.35)$.

[29]We are not in a LATE framework as our instruments are nonbinary. In our case, "never takers" are individuals who would not migrate at any *observed* value of the propensity score in our data, and "always takers" will migrate at any propensity score we observe.

TABLE 6. Average treatment effect bounds.

	$U_D < 0.35$	$0.35 < U_D < 0.71$	$U_D > 0.71$	Lower bound	Upper bound
A. Theoretical bounds					
$E(Y_1)$	0.825	0.730	[0, 1]	0.552	0.842
$E(Y_0)$	[0, 1]	0.825	0.782	0.874	0.524
Average treatment effect				-0.322	0.318
B. Bounds using never smokers and heavy smokers					
$E(Y_1)$	0.825	0.730	[0.697, 0.873]	0.754	0.805
$E(Y_0)$	[0.697, 0.873]	0.825	0.782	0.829	0.768
Average treatment effect				-0.076	0.037

Note: Outcomes estimated using cubic parametric separate approach and both instruments; ATE is the difference between $E(Y_1)$ and $E(Y_0)$ for all individuals. Upper and lower bounds on longevity for never and heavy smokers from Rogers, Hummer, Krueger, and Pampel (2005). Dependent variable is indicator for survival to age 75. Included in all specifications are indicators for birth cohort × sex and state of birth, a cubic in 1930 county population, mean county household size in 1930, fraction of county households that are farm households in 1930, and fraction of county households owing a radio in 1930.
Source: Authors' calculations using Duke SSA/Medicare data, mail-car capacity information from Borchert (1987), and county characteristics from the 1930 U.S. Census (Ruggles et al. (2010)).

$p_{\max}$ (0.71) and of Y_1 below our $p_{\min}$ (0.35). However, we still need estimates of the counterfactuals we cannot identify:

$$E[Y_0|X = x, U_D < p_{\min}],$$
$$E[Y_1|X = x, U_D > p_{\max}]. \tag{16}$$

As our outcome is binary, the weakest bounds are [0, 1]. As Panel A of Table 6 shows, the ATE bounds are very wide, approximately $[-0.32, 0.32]$. Given the relatively narrow range of our common support, this is not surprising.[30]

We can tighten these bounds somewhat if we make additional assumptions about the potential values of the missing counterfactuals. It strikes us unlikely that the counterfactual mortality rate for either of these two groups would have been zero or one. So to make tighter bounds, we propose the following. Using results from Rogers et al. (2005), we calculate that the 10-year survival rate for a never smoker from ages 65–75 in 2000 was 0.873 and the 10-year survival rate for a heavy smoker was 0.697.[31] Given that smoking is a very strong predictor of mortality,[32] we will take these values as lower and upper bounds for our two groups. Our assumption is that these group would not have been more healthy than an average 65–75 year-old who has never smoked and would not have been less healthy than an average 65–75 year-old who is a heavy smoker.

---

[30]For this exercise, we use the degree-3 parametric specification. Results are very similar if we instead use degree-1 or degree-2 parametric specifications.

[31]These results are for an "average" U.S. adult in terms of Socioeconomic and Health Status, and includes both men and women. The yearly mortality rate for never smokers is 0.011 for ages 65–69 and 0.016 for ages 70–74. For current smokers who smoke two or more packs per day, the yearly mortality rate is 0.029 for ages 65–69 and 0.042 for ages 70–74. Our bounds are calculated as follows: $(1 - 0.011)^5 \times (1 - 0.016)^5 = 0.873$ and $(1 - 0.029)^5 \times (1 - 0.042)^5 = 0.697$.

[32]Rogers et al. (2005) showed smoking has the largest mortality effects of any risk factor other than age (these factors included education, sex, race, employment status, drinking, seat-belt use, stress, and obesity).

Even with these more restrictive choices, our ATE bounds, $[-0.076, 0.037]$, include 0, as shown in Panel B of Table 6. This panel provides us with an excellent summary of what can be learned from our empirical exercise: First, by comparing $E(Y_1)$ for low-U_D and medium-U_D individuals we document positive selection into migration in terms of Y_1 (survival among those who move). Second, by comparing $E(Y_0)$ for medium-U_D and high-U_D we document positive selection into migration in terms of Y_0 (survival among those who do not move). Finally, of course we cannot estimate $E(Y_0)$ for individuals who migrate at any observed propensity score, nor estimate $E(Y_1)$ among those who fail to migrate at any observed propensity score. In our application, we would need to place fairly narrow bounds on those objects to obtain narrow bounds on the ATE for the full population.

Using the approach just described might also provide bounds for two additional treatment effects common in the literature: the average treatment effect for treated individuals (ATET), and average treatment effect for untreated individuals (ATUT).[33] Again using the degree-3 polynomial approach, we find a bound of $[-0.44, 0.070]$ for the ATET and $[-0.112, -0.002]$ for the ATUT.

We explore the heterogeneity in the MTE with respect to sex.[34] Results estimated using the LIV approach are shown in Figure 9. The MTE curve for men slopes downward while that for women is relatively flat. While the standard errors are too wide to draw any definitive conclusions, this pattern suggests that while the effect of migration out of the Northern Great Plains varied by underlying ability for men, the longevity penalty of migration for women was large and negative regardless of their U_D. Note that the 90% upper confidence bound barely includes zero for low ability (high-U_D) men, but this upper bound for women is well below zero over nearly our entire observable range of U_D.

5.5 *Concerns about the validity of our research design*

Our results show that migrants out of the Northern Great Plains experienced significantly lower longevity in old age than those who did not migrate, despite documented positive selection of migrants on characteristics thought to positively influence health. The causal interpretation of our IV and MTE results relies in particular on the conditional independence assumption: controlling for individual and place-specific observables, our birth-town-level railroad, and mail flow instruments are independent of an individual's longevity after age 65. We consider two major potential threats to this assumption.

First, all towns in the Northern Great Plains were small, but towns on major rail lines and on rail lines with higher postal mail flows were generally the most populous. Conceivably, living in a large town might have contributed to poorer health post-age 65. For instance, maybe children growing up in larger towns had higher levels of exposure to communicable diseases, which could lead to decreased life expectancy in old age. To

---

[33]Now instead of integrating the MTE with respect to u_D from 0.35 to 0.71 for the all individuals (as with the ATE), the idea is to do so instead for migrants (for the ATET), or nonmigrants (for the ATUT).

[34]We found no other interesting heterogeneity with respect to the other observables X.
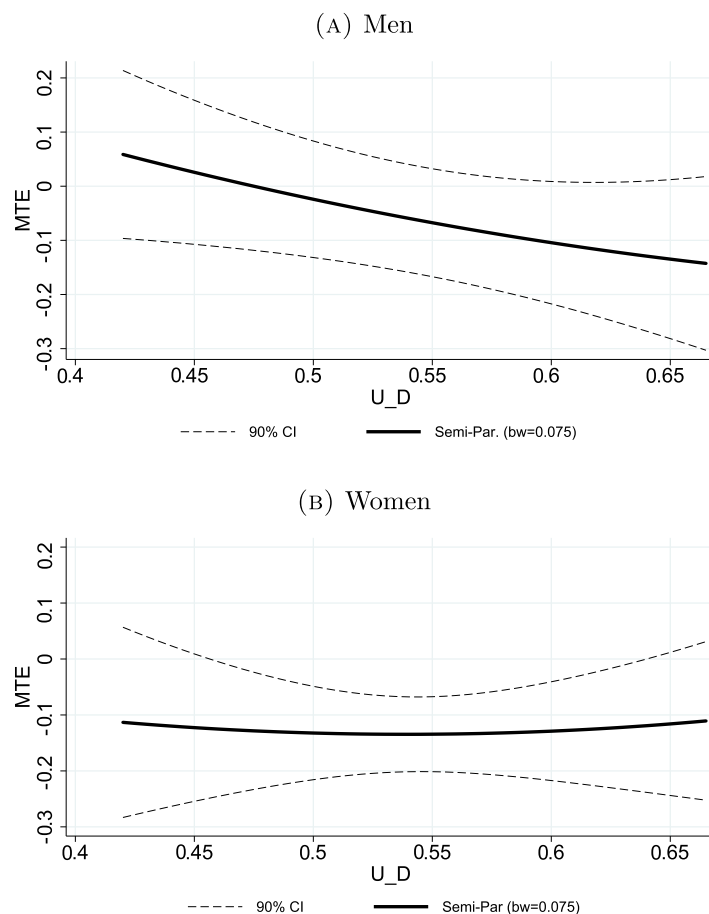
(A) Men

(B) Women

FIGURE 9. MTE estimates by sex. Notes: MTEs estimated using parametric LIV approach (degree=2) and both instruments. Standard errors calculated using 100 bootstrap replications with draws taken at the town of birth level. Sample includes white individuals born 1916–1927 in the Dakotas and Montana who lived to age 65. Included in all specifications are indicators for birth cohort × sex and state of birth, a cubic in 1930 county population, mean county household size in 1930, fraction of county households that are farm households in 1930, and fraction of county households owing a radio in 1930. Source: Authors' calculations using Duke SSA/Medicare data, mail-car capacity information from Borchert (1987), and county characteristics from the 1930 U.S. Census (Ruggles et al. (2010)).

explore this issue, we exclude all towns with population greater than 10,000 in 1940.[35] We find that the mail flow first stage works as with the larger sample (partial F statistic 29.39). Our IV estimate is little changed (though the standard error is larger); it is -0.124

---

[35]Those thereby excluded are Sioux Falls, SD (population in 1940, 40,832), Butte, MT (37,081), Fargo, ND (32,580), Great Falls, MT (28,928), Billings, MT (23,316), Grand Forks, ND (20,228), Aberdeen, SD (17,015), Bismarck, ND (15,496), and Minot, ND (10,476), resulting in a sample of 288,909.

(s.e. 0.058). Clearly, our results are not being driven exclusively by the larger towns in our sample.[36]

A second concern is that towns on major rail lines, and those with high postal mail flows, are more prosperous than other towns. If so, children growing up in these towns might be better educated and wealthier, and thus healthier in old age. In the Appendix, available in the Online Supplementary Material, we explore this issue in some detail, for the mail flow instruments, using U.S. Census data from 1930. We find that areas with higher levels of mail flow, relative to other areas, have (1) larger mean household sizes, (2) lower fraction of households that are farm households, and (3) higher rates of radio ownership. On the other hand, school attendance for both boys and girls were similar across towns. In sum, children in these towns were not better educated than other children, but their parents were more prosperous. Controlling for these variables may not account for all potential sources of bias in our MTE and IV estimates. There could be additional correlation between our mail-car capacity instrument and town-level unobserved characteristics that affect longevity. However, as the correlation between the observables above and mail-car capacity shows a positive relationship between local area prosperity and mail flow (which likely led to lower mortality in old age for migrants), we suspect that any remaining correlation between our instrument and unobservable town characteristics also leads to an upward bias in our estimates of the effect of migration on longevity. If so, our estimates might actually understate our paper's key conclusion— that migration has a negative impact on health.

Another issue, unrelated to the conditional independence assumption, concerns the ages over which we have measured mortality; we use a 10-year survival variable, conditional on survival to age 65. It is theoretically possible that the impact of migration on mortality is substantially different at younger ages than at these older ages. A worst case scenario for us would be a "mortality crossover," which could occur as follows: Suppose mortality is *lower* for migrants than nonmigrants at young ages, so among migrants a higher fraction of "high frailty" individuals remain post-age 65 (to use terminology of demographers such as Vaupel (1997)). Mortality post-age 65 might be higher for migrants than nonmigrants for this reason alone.[37] This would be a concern in our study if migrants had higher survival rates at younger ages than nonmigrants. To see if this is the case, we compare survival rates for migrants and nonmigrants pre-age 65 constructed using Vital Statistics death records and Census population estimates. We calculate 1-year survival rates by migrant status, sex, state of birth, and birth cohort for the years 1959–1962 and 1980–1992, when the cohorts in our sample were aged 32–46 and 53–64.[38] To determine if survival rates differed by migrant status, we regress the 1-year

---

[36]Excluding these towns from specifications using the two instruments separately leads to the same conclusion.

[37]Some scholars have presented evidence indicating such mortality "crossovers" in human populations. For example, there are papers suggesting that in the U.S. mortality is higher for blacks than for whites for ages less than approximately 85, but lower at older ages. Black, Hsu, Sanders, Schofield, and Taylor (2017) argued that the extant evidence about the black-white mortality crossover (and other similar crossovers) is likely due to measurement error in mortality. Nonetheless, such a crossover is a theoretical possibility in our case.

[38]We are limited to these years due to the Vital Statistics death records only being available starting in 1959, and the lack of state of birth on these records between 1963 and 1979.

TABLE 7. Comparison of migrant and nonmigrant pre-65 survival rates.

	Full sample (1)	Men (2)	Women (3)
Migrant	−0.0008	−0.0008	−0.0009
	(0.0002)	(0.0003)	(0.0001)
Mean one-year survival rate	0.9913	0.9886	0.9940
Observations	1510	756	754

Note: Sample includes white individuals born in the Dakotas and Montana 1916–1927. Dependent variable is the natural log of the 1-year survival rate, constructed from Vital Statistics death records and Census data for census years 1959–1962 and 1980–1992. Regressions include controls for sex, birth year, census year, and state of birth. Standard errors clustered on birth year in parentheses.
Sources: Authors' calculations using U.S. Vital Statistics and Census IPUMS data (Ruggles et al. (2010)).

survival rate on migrant status, controlling for birth cohort, sex, and year. Results of this analysis are presented in Table 7. For both the combined sample and men and women separately, migrant survival rates pre-age 65 are *lower* than those for nonmigrants. This evidence suggests that there is no crossover in the survival rates for migrants and non-migrants prior to age 65.

To explore this issue further, we repeat our IV analysis splitting our 10-year dependent variable into two 5-year periods: survival to age 70, given one has lived to age 65, and survival to age 75 given survival to age 70. If the results for these two different time periods showed a positive migrant effect on survival to age 70, but a negative effect after age 70, we would potentially be concerned about a mortality crossover driving our results. IV estimates for survival between age 65 and 70 are shown in Table 8 and those for ages 70–75 are in Table 9. Interestingly, it appears the longevity penalty for migrants only appears after age 70, as the difference in survival to age 70 is not significantly different between migrants and nonmigrants (though the point estimate is negative). However, it does not appear migrants have a higher probability than nonmigrants to survive to age 70. This, taken together with our results using Vital Statistics data, lead us to conclude it is unlikely a mortality crossover between migrants and nonmigrants explains our results.

5.6 *Potential causal mechanisms*

Migrants out of the Dakotas and Montana do not live as long as comparable individuals who remained behind. Why do these migrants experience a longevity penalty?

To identify potential reasons, we consider the causes of death reported for migrants and nonmigrants in Vital Statistics data.[39] Table 10 shows the dissimilarity index—the ratio of the proportion of deaths attributed to particular causes for migrants to that same proportion for nonmigrants. This statistic equals to 1.00 if an equal fraction of deaths for

[39]Ideally, we would investigate cause of death differences between migrants and nonmigrants using our instrument, but the Duke SSA/Medicare data do not contain cause of death. Two-sample IV is also not possible as the Vital Statistics data do not contain town of birth, which we need to form our town-level instruments.

TABLE 8. IV estimates for survival to age 70.

	OLS (1)	Mail-car capacity instruments (2)	Railroad instruments (3)	Both instruments (4)	Propensity score as instrument (5)
Migrant	−0.007	−0.023	−0.030	−0.025	−0.025
	(0.001)	(0.017)	(0.030)	(0.017)	(0.017)
Dependent variable mean	0.918	0.918	0.918	0.918	0.918
First Stage					
Closed pouch		0.041		0.039	
		(0.011)		(0.011)	
90–120		0.043		0.040	
		(0.008)		(0.009)	
270–510		0.052		0.045	
		(0.010)		(0.011)	
600–720		0.107		0.098	
		(0.015)		(0.016)	
810–1020		0.190		0.180	
		(0.021)		(0.023)	
Pre-1900 railroad			0.071	0.012	
			(0.009)	(0.008)	
Post-1900 railroad			0.043		
			(0.008)		
Partial *F* statistic		24.97	33.07	22.89	–
Dependent variable mean		0.536	0.536	0.536	0.536
Observations	330,428	330,428	330,428	330,428	330,428

Note: Standard errors clustered on town of birth in parentheses. Sample includes white individuals born 1916–1927 in the Dakotas and Montana who lived to age 65. Dependent variable is indicator for survival to age 70. Included in all specifications are indicators for birth cohort × sex and state of birth, a cubic in 1930 county population, mean county household size in 1930, fraction of county households that are farm households in 1930, and fraction of county households owing a radio in 1930. Partial *F* statistic is that for the vector of instruments in the first stage of each specification.
Source: Authors' calculations using Duke SSA/Medicare data, mail-car capacity information from Borchert (1987), and county characteristics from the 1930 U.S. Census (Ruggles et al. (2010)).

migrants and nonmigrants is attributed to the same cause of death; values greater than 1.00 indicate that a larger fraction of deaths among migrants is attributable to the cause, and values less than 1.00 indicate the opposite. Results are shown for deaths occurring to men and women in our cohorts of interest when they were between the ages of 65 and 75.

Four causes stand out as having a significantly higher fraction of deaths among migrants: cirrhosis, pneumonia, and influenza (for men), chronic obstructive pulmonary disease (COPD, for women), and lung cancer. These causes, which are often the consequence of excessive alcohol consumption and tobacco smoking, are known as "preventable causes of death."[40]

---

[40]Of course, not all such deaths are preventable, but many are. The connection between alcohol consumption and cirrhosis is well known, as is the connection between smoking and both lung cancer and COPD. As for pneumonia and influenza, there is also a likely link to smoking: The 2004 Surgeon General's

TABLE 9. IV estimates for survival to age 75, given survival to age 70.

	OLS (1)	Mail-car capacity instruments (2)	Railroad instruments (3)	Both instruments (4)	Propensity score as instrument (5)
Migrant	−0.007	−0.075	−0.100	−0.079	−0.078
	(0.001)	(0.030)	(0.042)	(0.030)	(0.030)
Dependent variable mean	0.878	0.878	0.878	0.878	0.878
First Stage					
Closed pouch		0.040		0.038	
		(0.011)		(0.011)	
90–120		0.042		0.039	
		(0.008)		(0.009)	
270–510		0.052		0.045	
		(0.010)		(0.011)	
600–720		0.105		0.096	
		(0.015)		(0.017)	
810–1020		0.190		0.180	
		(0.021)		(0.022)	
Pre-1900 railroad			0.070	0.012	
			(0.009)	(0.008)	
Post-1900 railroad			0.042		
			(0.009)		
Partial *F* statistic		25.28	31.41	23.33	–
Dependent variable mean		0.534	0.534	0.534	0.534
Observations	303,315	303,315	303,315	303,315	303,315

Note: Standard errors clustered on town of birth in parentheses. Sample includes white individuals born 1916–1927 in the Dakotas and Montana who lived to age 70. Dependent variable is indicator for survival to age 75. Included in all specifications are indicators for birth cohort × sex and state of birth, a cubic in 1930 county population, mean county household size in 1930, fraction of county households that are farm households in 1930, and fraction of county households owing a radio in 1930. Partial *F* statistic is that for the vector of instruments in the first stage of each specification.
Source: Authors' calculations using Duke SSA/Medicare data, mail-car capacity information from Borchert (1987), and county characteristics from the 1930 U.S. Census (Ruggles et al. (2010)).

We have some additional indirect evidence consistent with an important role for smoking and drinking, drawn from data collected by the Center for Disease Control's Behavioral Risk Factor Surveillance System (BRFSS). These data include information on smoking and alcohol consumption for U.S. adults. The BRFSS does not collect state of birth (so we cannot identify migrants and nonmigrants), but we can compare smoking behaviors among residents in states where our migrants and nonmigrants live. Columns (1) and (2) of Table 11 show this comparison for two measures of smoking using data from 1990 through 2000 (when our birth cohorts were between the ages of 63 and 84): the percentage who smoked 100 cigarettes in their life, and the percentage who are cur-

report states, "The evidence is sufficient to infer a causal relationship between smoking and acute respiratory illnesses, including pneumonia, in persons without underlying smoking-related chronic obstructive lung disease" (U.S. Department of Health and Human Services (2004)). Individuals with COPD are also more likely to experience flu-related complications, including pneumonia and bronchitis (Centers for Disease Control and Prevention (2015)).

TABLE 10. Dissimilarity of cause of death between migrants and non-migrants, white individuals born in the Dakotas and Montana.

Primary cause of death	Dissimiliarity	Deaths	Percent of all deaths	Rank
A. Men				
Chronic liver disease and cirrhosis	1.47*	722	1.35	11
Pneumonia and influenza	1.42*	1037	1.94	8
Cancer: respiratory and intrathoracic organs	1.07*	6628	12.39	2
Cancer: other	1.05	3273	6.12	6
All other diseases and internal causes	1.04	4316	8.07	3
Chronic obstructive pulmonary diseases	1.04	3875	7.24	5
Symptoms, signs, and ill-defined conditions	1.00	333	0.62	17
Cancer: stomach, colon, pancreas, peritoneum	1.00	4071	7.61	4
Major cardiovascular diseases	0.96*	22,332	41.74	1
All other and unspecified accidents and adverse effects	0.96	700	1.31	12
Intentional self-harm (suicide)	0.92	610	1.14	14
Cancer: leukemia	0.92	631	1.18	13
Motor vehicle accidents	0.91	413	0.77	15
Cancer: genital organs	0.90*	1992	3.72	7
Cancer: urinary organs	0.87*	890	1.66	10
Diabetes mellitus	0.85*	1027	1.92	9
Nephritis, nephrotic syndrome, and nephrosis	0.75*	350	0.65	16
B. Women				
Chronic liver disease and cirrhosis	1.46*	465	1.27	11
Chronic obstructive pulmonary diseases	1.36*	3029	8.25	4
Cancer: respiratory and intrathoracic organs	1.32*	3767	10.26	2
Symptoms, signs, and ill-defined conditions	1.15	223	0.61	17
Pneumonia and influenza	1.09	744	2.03	10
Cancer: leukemia	0.99	398	1.08	13
Cancer: genital organs	0.98	1442	3.93	8
Cancer: urinary organs	0.97	408	1.11	12
All other diseases and internal causes	0.96	3710	10.11	3
All other and unspecified accidents and adverse effects	0.95	364	0.99	14
Major cardiovascular diseases	0.94*	12,661	34.49	1
Cancer: other	0.91*	2630	7.16	6
Cancer: breast	0.90*	2150	5.86	7
Nephritis, nephrotic syndrome, and nephrosis	0.89	241	0.66	16
Cancer: stomach, colon, pancreas, peritoneum	0.89*	2890	7.87	5
Motor vehicle accidents	0.73*	302	0.82	15
Diabetes mellitus	0.71*	944	2.57	9

Note: Sample includes white individuals born in the Dakotas and Montana 1916–1927 who died between the ages of 65 and 75. Causes of death shown that account for at least 0.5% of deaths. Causes not shown include ulcer of stomach and duodenum (men and women), congenital malformations, deformations and chromosomal abnormalities (men and women), assault (homicide) (men and women), tuberculosis (men and women), All other external causes (men and women), cancer: breast (men), and intentional self-harm (suicide) (women). * indicates the 95% confidence interval does not include 1 (using a bootstrap procedure with 1000 replications).
Source: Authors' calculations using 1981–2001 Vital Statistics Mortality Data (National Center for Health Statistics (1968–2011)).

rent smokers.[41] For men, there are no statistically significant differences in these rates between the two sets of states. In contrast, women in migrant states have substantially

---

[41]We show weighted state averages, using the fraction of our sample living in each state in old age.

TABLE 11. Comparison of smoking and drinking behavior, white residents of migrant and non-migrant states born 1916–1927, BRFSS 1990–2000.

State of residence	Percent smoked 10 cigarettes in life	Percent current smokers	Percent binge drank in last month	Percent drank alcohol in last month
A. Low Education Men				
Migrant state	74.9	14.6	16.0	46.0
Nonmigrant State	73.4	14.8	12.3	42.4
Difference	1.4	−0.3	3.7	3.6
N	25,637	25,602	6479	18,379
B. High Education Men				
Migrant state	66.7	9.9	10.6	61.4
Nonmigrant state	67.8	10.0	9.9	52.7
Difference	−1.1	−0.2	0.7	8.7
N	19,168	19,153	7260	13,403
C. Low Education Women				
Migrant state	44.3	13.4	4.5	31.5
Nonmigrant state	33.5	10.4	5.1	22.3
Difference	10.8	3.0	−0.6	9.3
N	52,340	52,327	7324	36,281
D. High Education Women				
Migrant state	45.0	11.5	3.2	47.4
Nonmigrant state	35.3	9.9	2.6	34.7
Difference	9.6	1.6	0.6	12.7
N	28,860	28,852	7737	19,793

Note: Sample includes white individuals born 1916–1927. Percentages are computed by weighting responses from each state by fraction of our estimation sample (individuals born in the Dakotas and Montana) residing in that state in old age. High education denotes at least some college. Binge drinking is defined as consuming at least five alcoholic drinks in one occasion in the last month.
Source: Authors' calculations using Behavioral Risk Factor Surveillance System (BRFSS).

higher rates of tobacco use than women in nonmigrant states. These relationships are consistent with results presented in Table 10: female migrants were at particular risk of elevated death due to lung cancer and COPD. More broadly, they are also consistent with our finding that the average impact of migration on longevity is more adverse for women than men (Table 5).[42]

---

[42]Migrants tended to live in cities, in which they would have had increased exposure to air pollution, which may have further contributed to deaths due to cardiorespiratory causes (Ebenstein, Fan, Greenstone, He, Yin, and Zhou (2015), Pope, Arden, Ezzati, and Dockery (2009), Chen, Ebenstein, Greenstone, and Li (2013)). Given the gender differences we have just discussed, and the critical role of smoking identified in the literature, it is likely that smoking is playing the more important role in increased deaths due to COPD and lung cancer among migrants. We note one more complicating factor: many men in the earliest cohorts we study served in the military during World War II, and veterans were also more likely to migrate out of the Northern Great Plains. Work by Bedard and Deschenes (2006) shows that cohorts with higher rates of military service in WWII experienced higher mortality rates from cardiorespiratory causes over the ages 40–75, which they attribute to military-induced smoking. However, very few women served in the military, and our evidence for a negative effect of migration on longevity is especially strong for women, as is the evidence for a role of smoking for migrating women.

For both men and women, rates of alcohol use, as measured as the fraction of individuals reporting having an alcoholic drink in the last month, are higher in migrant states than nonmigrant states. The difference is consistent with results reported in Table 10 concerning excess deaths due to chronic liver disease and cirrhosis found among both male and female migrants. Rates of binge drinking, defined as five or more alcoholic drinks in one occasion and thought to be particularly harmful to health (Stahre, Roeber, Kanny, Brewer, and Zhang (2014)), were significantly higher only for men without at least some college education.[43]

Given the available evidence, we find it quite likely that behavioral factors—smoking and excess alcohol consumption—were at least part of the story for the increased mortality experienced by migrants. Recall that in our study of MTEs we do not have clear results concerning heterogeneity in treatment effects, but there is some evidence that the negative impact of migration on health may be particularly pronounced for high-U_D individuals, that is, migrants who are drawn from the lower end of the "ability" distribution. We are thus interested to see if we can provide any additional evidence that the excess levels of "preventable deaths" are concentrated among low-ability migrants. We focus our attention on smoking-related deaths, because these are far more common than alcohol-related deaths.[44] Specifically, we look to see if the excess mortality among migrants due to these diseases is concentrated among individuals with relatively low levels of education.

Education is reported on the death certificate in at least some years for certain states; we use all available data.[45] We combine COPD and lung cancer to form a "smoking-related deaths" category, and then calculate our dissimilarity index for those with education less than high school, and those with high school and above (recalling that for this generation the overall average education among whites is slightly over 11). For the relatively poorly-educated male migrants, the "smoking-related dissimilarity index" is 1.14 (with a 0.95 confidence interval of 1.08 to 1.20), while for well-educated male migrants it is 0.99 (c.i., 0.89 to 1.09). For the relatively poorly-educated female migrants the "smoking-related dissimilarity index" is 1.26 (c.i., 1.18 to 1.35), while for better-educated female migrants it is a nearly identical 1.29 (c.i., 1.14 to 1.44).

We thus find that for men, excess mortality among migrants due to smoking-related causes is higher among those with relatively low levels of education. This is consistent with the MTE pattern we estimate (though this is merely suggestive). However, among women, both poorly-educated and well-educated migrants tend to have increased death due to smoking. Table 11 shows that differences in smoking and drinking behaviors between migrant and nonmigrant states are similar by education group for women. This is also consistent with our MTE results for women, which show equally large longevity penalties across the ability distribution.

---

[43]This may help explain the patterns found for men in the MTE in Figure 9, which show larger treatment effects for high U_D, although rates of alcohol use point in the other direction.

[44]From Table 6, we note that approximately 19% of deaths have COPD or lung cancer listed as a cause, while only 1% list liver disease and cirrhosis.

[45]Sample sizes are relatively small because of missing years and states. Also, we note that education reports on death certificates are likely subject to substantial misreporting; education reported on the death certificate tends to be upward biased (Sorlie and Johnson (1996)).

Future researchers may discover that the unhealthy behaviors and poor health outcomes among migrants are related to the loss of family and social ties, as some individuals experience personal distress when they leave families and communities behind. There is sociological research suggesting that personal networks in rural areas are stronger and more supportive than those in urban areas (Beggs, Haines, and Hurlbert (1996)), and other research has shown that increased social interaction and decreased feelings of loneliness are associated with decreased mortality risk, especially among the elderly (Steinbach (1992), Penninx, van Tilburg, Kriegsman, Deeg, Boeke, and van Eijk (1997), Patterson and Veenstra (2010)).

6. Conclusion

Individuals born in the rural U.S. Northern Great Plains in the early 20th century had much to gain by migrating to growing cities in the West and Midwest—in terms of opportunities to enjoy high-amenity urban centers, expanded labor market prospects, and increased lifetime wealth. However, our study of these migrants suggests that they paid a price in terms of older-age longevity. This result is important. Economic prosperity is an important aspect of lifetime welfare. But health and longevity are also key dimensions of human well-being.[46]

Our analysis suggests that a causal mechanism for migrants' reduced longevity was the adoption of risky health-related behaviors, including smoking and excess alcohol use. We hope this finding motivates additional scholarship on the health effects of rural to urban migration. Research in this area can improve our understanding about how migration affects lives, and can potentially provide a useful input for public health policy in the many countries where rural-urban migration continues to be an important phenomenon.

References

Abramitzky, R., L.P. Boustan, and K. Eriksson (2012), "Europe's tired, poor, huddled masses: Self-selection and economic outcomes in the age of mass migration." *American Economic Review*, 102, 1832–1856. [573]

Angrist, J. D., G. W. Imbens, and D. B. Rubin (1996), "Identification of causal effects using instrumental variables." *Journal of the American Statistical Association*, 91, 444–455. [580]

Bedard, K. and O. Deschenes (2006), "The long-term impact of military service on health: Evidence from World War II and Korean War veterans." *The American Economic Review*, 96, 176–194. [599]

Beggs, J. J., V. A. Haines, and J. S. Hurlbert (1996), "Revisiting the rural-urban contrast: Personal networks in nonmetropolitan and metropolitan settings." *Rural Sociology*, 61, 306–325. [601]

---

[46]Many scholars have made this latter point. For a particularly cogent expression of this idea, see Amartya Sen's Innocenti Lecture (Sen (1998)).

Björklund, A. and R. Moffitt (1987), "The estimation of wage gains and welfare gains in self-selection models." *The Review of Economics and Statistics*, 42–49. [579]

Black, D. A., Y.-C. Hsu, S. G. Sanders, L. S. Schofield, and L. J. Taylor (2017), "The Methuselah effect: The pernicious impact of unreported deaths on old-age mortality estimates." *Demography*, 54, 2001–2024. [594]

Black, D. A., S. G. Sanders, E. J. Taylor, and L. J. Taylor (2015), "The impact of the great migration on mortality of African Americans: Evidence from the deep South." *American Economic Review*, 105, 477–503. [566, 567, 568, 572, 578]

Borchert, J. R. (1987), *America's Northern Heartland*. University of Minnesota Press. [567, 577, 578, 582, 583, 584, 586, 588, 589, 590, 591, 593, 596, 597]

Boustan, L. P. (2010), "Was postwar suburbanization 'white flight'? Evidence from the black migration." *The Quarterly Journal of Economics*, 125, 417–443. [566]

Brinch, C. N., M. Mogstad, and M. Wiswall (2017), "Beyond LATE with a discrete instrument." *Journal of Political Economy*, 125, 985–1039. [567, 579, 584, 587]

Carrington, W. J., E. Detragiache, and T. Vishwanath (1996), "Migration with endogenous moving costs." *The American Economic Review*, 909–930. [566]

Centers for Disease Control and Prevention (2015), "People at high risk of developing flu-related complications." [597]

Chen, Y., A. Ebenstein, M. Greenstone, and H. Li (2013), "Evidence on the impact of sustained exposure to air pollution on life expectancy from China's Huai River policy." *Proceedings of the National Academy of Sciences*, 110, 12936–12941. [599]

Cornelissen, T., C. Dustmann, A. Raute, and U. Schönberg (2016), "From LATE to MTE: Alternative methods for the evaluation of policy interventions." *Labour Economics*, 41, 47–60. [579, 585, 590]

Davis, K. (1955), "The origin and growth of urbanization in the world." *American Journal of Sociology*, 429–437. [569]

Ebenstein, A., M. Fan, M. Greenstone, G. He, P. Yin, and M. Zhou (2015), "Growth, pollution, and life expectancy: China from 1991–2012." *American Economic Review*, 105, 226–231. [599]

Elo, I. T. and S. H. Preston (1996), "Educational differentials in mortality: United States 1979–1985." *Social Science & Medicine*, 42, 47–57. [573]

Farr, W. (1864), *Supplement to the 25th Annual Report of the Registrar General*. HMSO, London. [566, 569]

Ferrie, J. P. (2006), "Internal migration." In *Historical Statistics of the United States, Earliest Times to the Present: Millennial Edition* (S. B. Carter, S. S. Gartner, M. R. Haines, A. L. Olmstead, R. Sutch, and G. Wright, eds.), Cambridge University Press. chapter Ac, 1-489–1-494. [569, 570]

Gentzkow, M., J. Shapiro, and M. Sinkinson (2016), *Circulation of US Daily Newspapers, 1924, Audit Bureau of Circulations [Data Set]*. Inter-University Consortium for Political and Social Research, MI. [576]

Glaeser, E. L. (1999), "Learning in cities." *Journal of Urban Economics*, 46, 254–277. [569]

Glaeser, E. L. (2011), *Triumph of the City: How Our Greatest Invention Makes Us Richer, Smarter, Greener, Healthier, and Happier*. Penguin Press, New York, NY. [569]

Glaeser, E. L. and M. G. Resseger (2010), "The complementarity between cities and skills." *Journal of Regional Science*, 50, 221–244. [569]

Goldsmith-Pinkham, P., I. Sorkin, and H. Swift (2017), "Bartik instruments: What, when, why and how." Working Paper. [580]

Griliches, Z. and W. M. Mason (1972), "Education, income, and ability." *The Journal of Political Economy*, S74–S103. [573]

Grob, G. N. (2002), *The Deadly Truth: A History of Disease in America*. Harvard University Press. [569]

Harris, J. R. and M. P. Todaro (1970), "Migration, unemployment and development: A two-sector analysis." *The American Economic Review*, 60, 126–142. [568]

Heckman, J. J., S. Urzua, and E. Vytlacil (2006), "Understanding instrumental variables in models with essential heterogeneity." *The Review of Economics and Statistics*, 88, 389–432. [588]

Heckman, J. J. and E. J. Vytlacil (1999), "Local instrumental variables and latent variable models for identifying and bounding treatment effects." *Proceedings of the National Academy of Sciences*, 96, 4730–4734. [567, 579, 580, 584, 585]

Heckman, J. J. and E. J. Vytlacil (2001), "Instrumental variables, selection models, and tight bounds on the average treatment effect." *Econometric Evaluation of Labour Market Policies*, 1–15. [579]

Heckman, J. J. and E. J. Vytlacil (2005), "Structural equations, treatment effects, and econometric policy evaluation." *Econometrica*, 73, 669–738. [567, 579, 590]

Heckman, J. J. and E. J. Vytlacil (2007), "Econometric evaluation of social programs, part I: Causal models, structural models and econometric policy evaluation." *Handbook of Econometrics*, 6, 4779–4874. [567, 584]

Ji, Y., S. Liu, X. Zhao, Y. Jiang, Q. Zeng, and C. Chang (2016), "Smoking and its determinants in Chinese internal migrants: Nationally representative cross-sectional data analyses." *Nicotine and Tobacco Research*, 18, 1719. [568]

Johnson, J. E. and E. J. Taylor, (2019), "Supplement to 'The long run health consequences of rural-urban migration'." *Quantitative Economics Supplemental Material*, 10, https://doi.org/10.3982/QE962. [574]

Liu, Y., H. Song, T. Wang, T. Wang, H. Yang, J. Gong, Y. Shen, W. Dai, J. Zhou, S. Zhu, and Z. Pan (2015), "Determinants of tobacco smoking among rural-to-urban migrant workers: A cross-sectional survey in Shanghai." *BMC Public Health*, 15. [568]

Long, B. A. (1951), *Mail by Rail: The Story of the Postal Transportation Service*. Simmons-Boardman. [575]

Lu, Y. (2010), "Rural-urban migration and health: Evidence from longitudinal data in Indonesia." *Social Science & Medicine*, 70, 412–419. [568]

Lucas, R. E. (1977), "Internal migration and economic development: An overview." In *Internal Migration: A Comparative Perspective* (A. A. Brown and E. Neuberger, eds.), 37–60, Academic Press, New York, NY. Chapter 3. [574]

Lucas, R. E. (2004), "Life earnings and rural-urban migration." *Journal of Political Economy*, 112, S29–S59. [565, 568, 570]

Michaels, G., F. Rauch, and S. J. Redding (2012), "Urbanization and structural transformation." *The Quarterly Journal of Economics*, 127, 535–586. [570]

National Center for Health Statistics (1968–2011), "Mortality multiple cause of death, public use data file and documentation." [598]

National Center for Health Statistics (2015), "Health, United States, 2014: With special feature on adults aged 55–64." [569]

Patterson, A. C. and G. Veenstra (2010), "Loneliness and risk of mortality: A longitudinal investigation in Alameda county, California." *Social Science & Medicine*, 71, 181–186. [601]

Penninx, B. W. J. H., T. van Tilburg, D. M. W. Kriegsman, D. J. H. Deeg, A. J. P. Boeke, and J. T. M. van Eijk (1997), "Effects of social support and personal coping resources on mortality in older age: The longitudinal aging study Amsterdam." *American Journal of Epidemiology*, 146, 510–519. [601]

Pope III, C. Arden, M. Ezzati, and D. W. Dockery (2009), "Fine-particulate air pollution and life expectancy in the United States." *New England Journal of Medicine*, 360, 376–386. [599]

Ravenstein, E. G. (1885), "The laws of migration." *Journal of the Statistical Society of London*, 48, 167–235. [566]

Reddy, K. S., B. Shah, C. Varghese, and A. Ramadoss (2005), "Responding to the threat of chronic diseases in India." *The Lancet*, 366, 1744–1749. [568]

Rogers, R. G., R. A. Hummer, P. M. Krueger, and F. C. Pampel (2005), "Mortality attributable to cigarette smoking in the United States." *Population and Development Review*, 31, 259–292. [591]

Ruggles, S., J. T. Alexander, K. Genadek, R. Goeken, M. B. Schroeder, and M. Sobek (2010), *Integrated Public Use Microdata Series: Version 5.0 [Machine-Readable Database]*. University of Minnesota, Minneapolis. [569, 571, 572, 574, 578, 582, 583, 584, 586, 588, 589, 590, 591, 593, 595, 596, 597]

Sen, A. (1998), "Mortality as an indicator of economic success and failure." *The Economic Journal*, 108, 1–25. [601]

Sorlie, P. D. and N. J. Johnson (1996), "Validity of education information on the death certificate." *Epidemiology*, 437–439. [600]

Sorlie, P. D., E. Rogot, R. Anderson, N. J. Johnson, and E. Backlund (1992), "Black-white mortality differences by family income." *The Lancet*, 340, 346–350. [573]

Stahre, M., J. Roeber, D. Kanny, R. D. Brewer, and X. Zhang (2014), "Contribution of excessive alcohol consumption on deaths and years of potential life lost in the United States." *Preventing Chronic Disease*, 11. [600]

Steinbach, U. (1992), "Social networks, institutionalization, and mortality among elderly people in the United States." *Journal of Gerontology*, 47, S183–S190. [601]

Stuart, B. and E. J. Taylor (2017), "Social interactions and location decisions: Evidence from U.S. mass migration." Working Paper. Available at https://sites.google.com/site/evanjtaylor. [566]

Todaro, M. P. (1969), "A model of labor migration and urban unemployment in less developed countries." *The American Economic Review*, 138–148. [568]

U.S. Department of Health and Human Services (2004), "The health consequences of smoking: A report of the surgeon general." Technical report, U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health, Atlanta, GA. [597]

United Nations, Department of Economic and Social Affairs, Population Division (2014), *World Urbanization Prospects: The 2014 Revision, Highlights (ST/ESA/SER.A/352)*. United Nations Publications. [567]

Vaupel, J. W. (1997), "Trajectories of mortality at advanced ages." In *Between Zeus and the Salmon: The Biodemography of Longevity*, Vol. 4, Chapter 2 (K. W. Wachter and C. E. Finch, eds.), 17–37, National Acadamies Press. [594]

Vlahov, D., E. Gibble, N. Freudenberg, and S. Galea (2004), "Cities and health: History, approaches, and key questions." *Academic Medicine*, 79, 1133–1138. [569]

Vytlacil, E. (2002), "Independence, monotonicity, and latent index models: An equivalence result." *Econometrica*, 70, 331–341. [580]

Woodruff, C. and R. Zenteno (2007), "Migration networks and microenterprises in Mexico." *Journal of Development Economics*, 82, 509–528. [566]

World Health Organization (2017), "The bill China cannot afford: Health, economic, and social costs of China's tobacco epidemic." Technical report, World Health Organization Western Pacific Region, United Nations Development Program. [568]

Yang, T., J. Wu, I. R. H. Rockett, A. S. Abdullah, J. Beard, and J. Ye (2009), "Smoking patterns among Chinese rural–urban migrant workers." *Public Health*, 123, 743–749. [568]