NBER WORKING PAPER SERIES

TRANSPORTATION AND HEALTH IN A DEVELOPING COUNTRY: THE UNITED STATES, 1820–1847

Ariell Zimran

Working Paper 24943
http://www.nber.org/papers/w24943

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
August 2018

I am indebted to Joel Mokyr, Joseph Ferrie, and Matthew Notowidigdo for encouragement and guidance. I also thank Jeremy Atack, Hoyt Bleakley, Natalia Cantet, William Collins, Price Fishback, Bernard Harris, Richard Hornbeck, Robert Margo, Yannay Spitzer, and John Wallis for helpful suggestions and insightful comments. Thanks are also due to Timothy Cuff for sharing his data on Pennsylvania recruits to the Union Army; to Noelle Yetter for assistance at the National Archives; to Ashish Aggarwal and Danielle Williamson for excellent research assistance; to seminar participants at Vanderbilt University, Tel Aviv University, the Hebrew University of Jerusalem, and Ben Gurion University of the Negev; and to participants in the 2016 Social Science History Association Conference, the 2017 NBER DAE Summer Institute, the 2018 H2D2 Research Day at the University of Michigan, and the 2018 Midwest International Economic Development Conference. This project was supported by an Economic History Association Dissertation Fellowship, by the Northwestern University Center for Economic History, and by the Balzan Foundation. This project, by virtue of its use of the Union Army Data, was supported by Award Number P01 AG10120 from the National Institute on Aging. The content is solely the responsibility of the author and does not necessarily represent the official views of the National Institute on Aging or the National Institute of Health. All errors are my own. The views expressed herein are those of the author and do not necessarily reflect the views of the National Bureau of Economic Research.

Transportation and Health in a Developing Country: The United States, 1820–1847
Ariell Zimran
NBER Working Paper No. 24943
August 2018
JEL No. I15,N31,N71,O18

ABSTRACT

I study the impact of transportation on health in the rural US, 1820–1847. Measuring health by average stature and using within-county panel analysis and a straight-line instrument, I find that greater transportation linkage, as measured by market access, in a cohort's county-year of birth had an adverse impact on its health. A one-standard deviation increase in market access reduced average stature by 0.10 to 0.29 inches. These results explain 26 to 65 percent of the decline in average stature in the study period. I find evidence that transportation affected health by increasing population density, leading to a worse epidemiological environment.

Ariell Zimran
Department of Economics
Vanderbilt University
2301 Vanderbilt Place
Nashville, TN 37235
and NBER
ariell.zimran@vanderbilt.edu

1 Introduction

In the four decades prior to the Civil War, the United States experienced a "transportation revolution" (Taylor 1951) that was in part responsible for the prodigious growth of the antebellum American economy (e.g., Atack et al. 2010). This experience is often cited as evidence that transportation improvements are crucial to spurring and supporting economic growth in modern developing countries (e.g., Banerjee, Duflo, and Qian 2012)—a view that has inspired massive investment in transportation infrastructure in the developing world (World Bank 2007) with largely positive impacts (see Donaldson 2015).

Despite the well known benefits of economic growth, transportation projects that induce it may not be unambiguously welfare-improving. In the antebellum United States the early phases of modern economic growth were accompanied by declining health as measured by life expectancy and average stature (Floud et al. 2011). Similar patterns have been documented in nineteenth-century England and in modern developing countries such as China and India (Deaton 2007; Floud et al. 2011; Floud, Wachter, and Gregory 1990; Jayachandran and Pande 2017; Trivedi 2017),[1] all of which have experienced transportation revolutions of their own. If transportation improvements were in any way responsible for these deteriorations in health, then this impact must be weighed against the benefits of growth in assessing the welfare effects of infrastructure development. Little empirical evidence exists, however, on whether and how transportation improvements affect health, and economic theory shows that the impact may be positive or negative.

In this paper, I provide such evidence by studying the effect of transportation improvements on health in the rural United States in the period 1820–1847. Besides being of interest for its own sake, the antebellum United States is a particularly good setting in which to study this relationship. The transportation improvements of this period—consisting mostly of canal construction and improvements in river navigability—were transformative of the American continent and economy, involving the expansion of transportation infrastructure into large areas that were previously isolated and undeveloped. Moreover, the time horizon of data available due to the historical setting permits the observation of permanent and long-term health effects.

My analysis is based on two main data sources. To describe the development of the transportation network in the antebellum United States, I use GIS shape files that have recently been made available by Atack (2015, 2016, 2017). This source provides the location and opening date of all canals, railroads, and navigable waterways in the antebellum United States. I use these data to compute Donaldson and Hornbeck's (2016) market access statistic, which is my main measure of transportation linkage, for all counties east of

---

[1] This does not refer to the commonly cited counter-cyclicality of health (e.g., Ruhm 2000). Unlike this cyclical relationship, I am referring to a relationship between economic growth and health over a few decades.

the Mississippi River for each year 1820-1847. As is common in studying developing and historical contexts, I measure health by average stature, which reflects net nutritional status in childhood and adolescence (Floud et al. 2011; Steckel 1995). I use stature data from the records of enlisters in the Union Army (Records of the Adjutant General's Office 1861–1865), providing data on the heights, counties of birth, and birth cohorts of 25,567 native-born white men in the birth cohorts of 1820–1847 in the Northeast and Midwest regions of the United States. I limit the study period to 1820–1847 because these birth cohorts are the only ones (in the antebellum period) for which there exists a sample of health data that is reasonably representative of the population (Zimran 2018). The combination of data from these sources enables the construction of a panel data set of county average stature and transportation linkage.[2]

The main empirical challenge of this paper is to determine the impact of transportation improvements on health while addressing the possibility that any correlation between the two might be driven by omitted and potentially unobservable variables. For instance, local characteristics might spur economic growth, attracting transportation, affecting health, and creating a spurious relationship between the two. To address this possibility, I use two empirical approaches. First, I exploit the panel structure of the data to estimate specifications that include county fixed effects. Second, I use an instrumental variables approach. I construct an instrumental variable based on the principle that transportation improvements were intended to connect major watersheds (i.e., the Mississippi, Great Lakes, and Atlantic) to one another and to major cities. In particular, I augment the 1820 transportation network with the shortest straight lines creating these connections, and treat these lines as canals built incrementally over a period of 15 years. I then compute market access based only on the 1820 network and these straight-line connections, and use this alternative measure as the instrument for market access. This instrument builds on and shares an interpretation with the straight-line instruments commonly used in the literature on the effects of transportation (e.g., Atack et al. 2010; Banerjee, Duflo, and Qian 2012; Chandra and Thompson 2000; Ghani, Goswami, and Kerr 2016). It also adds to this set of instruments by introducing a temporal component to them (see also Hornung 2015).

Using each of these identification strategies, I find a negative relationship between transportation linkage as measured by market access and health as measured by average stature. The magnitude of this relationship is large. According to my estimates, a one-standard deviation increase in market access was associated with a 0.10 to 0.25 inch decline in average stature, depending on the identification strategy. To put this figure in perspective, Zimran (2018) estimates that urbanites during this period suffered a height penalty of 0.29 inches relative to ruralists, and Deaton and Arora (2009) estimate that college graduates enjoy a 0.7 inch

---

[2] In practice, I do not use the average observed stature. Instead, I use the individual observations of stature linked to individuals' birth years, clustering standard errors by county of birth.

height premium over high school graduates in the modern United States. This negative relationship is robust to the inclusion of numerous controls and a variety of time trends.

I also investigate the hypothesis that improved market access reduced health by generating increases in population density. In combination with insufficient sanitation and public health infrastructure in the antebellum period, such concentration of population would have made previously undeveloped locations less healthy (Costa 1993; Floud et al. 2011; Steckel 1995). In support of this mechanism and consistent with other studies of the effects of transportation construction in the antebellum United States (Atack et al. 2010), I find evidence of rising population density in a county in response to increases in market access. I also find that the effects of market access on increasing population density were stronger in counties where the suitability for wheat and corn production was greater (according to the Food and Agriculture Organization 2002), and that the negative impact of market access on stature was stronger in these same counties. That is, counties where population density increased the most in response to rising market access were those where the deleterious effect on average stature was the greatest.

This paper contributes to a number of literatures. Narrowly, it adds to the understanding of the deterioration in health experienced in the United States at the onset of modern economic growth—a phenomenon known as the "Antebellum Puzzle." This pattern is a fundamental stylized fact of American economic history that bears on the evaluation of the welfare effects of economic growth in developing countries; but its cause has remained poorly understood due to a lack of well identified empirical investigations. In this paper, I provide perhaps the first piece of direct and plausibly causal evidence as to a potential explanation for this puzzle by showing that the effect of market access on average stature, combined with the rise in market access over the antebellum period, can explain up to 65 percent of the decline in stature. Also in the specific context of the antebellum United States, this paper adds to the literature studying the effects of canal construction (e.g., Ransom 1970). Despite the recognized importance of these projects, the bulk of recent scholarly attention has accrued to the later rail construction (e.g., Donaldson and Hornbeck 2016).

More broadly, this paper adds to the literature on the impacts of transportation improvements. Although there is a large literature describing these impacts on a variety of economic outcomes,[3] the effects on health have received far less empirical attention and are not understood as well. Previous findings of a negative relationship between transportation presence and average stature in the antebellum United States (Cuff 2005; Haines, Craig, and Weiss 2003; Yoo 2012) have largely been constrained by data availability and

---

[3] Specific case studies of the impacts of transportation improvements on a variety of economics outcomes are given by Atack et al. (2010), Baum-Snow et al. (2018), Chandra and Thompson (2000), Donaldson and Hornbeck (2016), Duranton and Turner (2012), Emran and Hou (2013), Ghani, Goswami, and Kerr (2016), Jacoby (2000), Jacoby and Minten (2009), Jaworski and Kitchens (2016), Storeygard (2016), and Tang (2014), among others.

methodological limitations to documenting correlations. To my knowledge, only a few studies (Burgess and Donaldson 2012; Tang 2017) exist determining the causal effect of transportation on health in specific cases. This paper contributes to this literature by providing an analysis of the effect of transportation on health in the context of a large and historically important infrastructure project, and showing, with attention to causality, that this project, despite its well known economic benefits, had a negative impact on health.

2 Background

2.1 The Economics of Transportation and Health

Economic theory proposes a number of mechanisms by which transportation might impact health. The most direct are its potential epidemiological effects. For instance, transportation can carry disease along with freight and passengers, and might bring infection to places that it had once been unable to reach (e.g., Tang 2017). In the antebellum United States in particular, this mechanism might have acted by linking relatively healthy rural areas to urban areas, where disease was prevalent, and carrying this disease from the latter to the former (Floud et al. 2011). Conversely, transportation linkages might provide previously isolated areas with better health care by reducing pecuniary and non-pecuniary access costs, though this mechanism is unlikely to apply to the antebellum United States, in which the available health care was primitive at best.

Transportation can also affect health indirectly through its effects on income and development. Transportation linkages are often found to increase economic activity in newly linked areas (e.g., Duranton and Turner 2012; Ghani, Goswami, and Kerr 2016), and the antebellum period is no exception (Atack et al. 2010). The resulting rise in income would lead to improved health through consumption of more and better health-improving goods, such as food and medicine (Emran and Hou 2013; Fogel 2004). This growth can also generate increases in population density or urbanization in newly linked areas. This effect has the potential to harm health by increasing exposure to disease, both through increased contact between individuals and through the sanitation consequences of greater concentrations of population (Costa 1993; Steckel 1995). This mechanism is particularly relevant in the antebellum American context given the lack of adequate sanitation infrastructure and technology and absence of public health projects. Besides its impact on the level of income, transportation integrates newly linked areas to the larger economy, potentially affecting the volatility of income, with theoretically ambiguous impacts on health (Burgess and Donaldson 2012).

Finally, transportation can affect health through its impact on relative prices. In the antebellum United States, the areas being linked to the transportation network were largely food-producing. Transportation linkages in this setting would tend to increase the relative farm-gate price of food: access to larger markets would increase the price that farmers could command for their output, while linkage to manufacturing centers in urban areas would reduce the price of manufactures produced there (Komlos 1987; Komlos and Coclanis 1997). On the other hand, rising relative food prices would bolster the incomes of food producers. If the net effect of these changing relative prices was to reduce the consumption of health-improving goods, then health could have deteriorated in response.

Combining all of these theoretical mechanisms, the bottom-line prediction of the impact of transportation on health is ambiguous in sign, and is thus an empirical question. Yet there is relatively little empirical work to enlighten this theoretical puzzle. A considerable fraction of the work that does exist focuses on the antebellum United States as part of efforts to understand the progress of health during industrialization. Cuff (2005), Haines, Craig, and Weiss (2003), and Yoo (2012) show that areas linked to the transportation network or with better market access due to relative proximity to cities had worse health, as measured by average stature and death rates, than other areas.[4] However, limited information on the historical transportation network available at the time that these studies were conducted (Atack 2013), together with limited methods available to quantify transportation linkages, constrained these authors to describing correlations, often with only a single year of observation of transportation presence. Thus, these results are at best suggestive of the causal impact of transportation on health.

Studies of the transportation-health relationship in other contexts (e.g., Ali et al. 2015; Banerjee and Sachdeva 2015; Bell and van Dillen 2018; Blimpo, Harding, and Wantchekon 2013; Stifel and Minten 2015) are also largely suggestive, as they either are correlational, report effects on indices including health but not on health separately, study very small regions, or focus on inputs to health rather than on health outcomes.[5] These studies are also often constrained to study only the short-term effects of transportation.

There are two notable exceptions that provide causal estimates of the effect of transportation infrastructure on health. Tang (2017) studies the mortality effects of the construction of the railroad network in late-nineteenth century Japan. His difference-in-differences approach reveals an increase in mortality coming from new rail linkage that is generated by the spread of communicable diseases. Burgess and Donaldson (2012) give causal evidence of beneficial impacts of transportation on health by showing that transportation

---

[4] Yoo's (2012) result is more subtle, showing a positive effect of transportation in the Northeast and a negative effect in the Midwest. His analysis, however, does not exclude urban areas and is based on only a single year of observation of the transportation network.

[5] It is particularly important not to consider an improvement in health inputs, such as improved consumption or access to health care, as necessarily generating improvements in health outcomes. In the antebellum United States in particular, and also in many developing contexts, apparent improvements in health inputs (such as greater income and consumption) are in fact accompanied by declining health. It is thus important to study health outcomes in order to determine the true net effect on health.

improvements in colonial India reduced the increases in mortality in response to negative agricultural yield shocks. Notwithstanding these papers, economists' understanding of the health effects of transportation remains limited, in part because these studies find opposing effects and in part because of the limited number of case studies. Additional studies are necessary to better understand these impacts and the mechanisms that generate them, especially given the potential value to policy makers in determining the welfare effects of transportation improvements.

2.2 Transportation Improvements in the Antebellum United States

The quintessential transportation improvement of the antebellum United States was the railroad. As a result, the impacts of this mode of transportation have been the subject of considerable and notable scrutiny (e.g., Atack et al. 2010; Donaldson and Hornbeck 2016; Fishlow 1965; Fogel 1964). Although some rail construction occurred in the 1830s and 1840s, the bulk of antebellum railroad construction did not occur until the 1850s (especially in the Midwest)—after the study period for this paper. Instead, the improvement of water transportation was the key component of the transportation revolution in the period on which this paper focuses. This included the construction of the canal network in the Northeast and Midwest. It also included expansions in navigability of the Mississippi River system and its major and minor tributaries through improvements in steamboat technology and the clearance of hazards to navigation.[6]

The impacts of these canals and improvements in navigability have received less attention in the recent economic history literature on transportation in the United States than have those of railroads. Earlier work attributes considerable economic benefits to canal construction, all of which are likely to have contributed to the ultimate health impacts of these transportation improvements on health. The most notable success stories were the Erie Canal (Segal 1961) and the Ohio and Erie Canal (Ransom 1970). Despite the notable financial failure of the latter, these canals contributed considerably to economic growth and development in the areas through which they passed, to the development of manufacturing and commerce in these same areas, and to the broader economic development of the Midwest (Niemi 1970; Ransom 1967, 1971).

2.3 The Antebellum Puzzle

Any study of health in the antebellum United States is inextricably linked to the "Antebellum Puzzle." Despite an improvement in the standard of living according to conventional economic measures, such as income per capita and real wages (Costa and Steckel 1997), the antebellum period was characterized by a

---

[6] The geographical development of all of these systems is described graphically in Figure A.1.

precipitous decline in health. Between the first and second quarters of the nineteenth century, life expectancy at age 10 for males declined by about 3 years (Fogel 1986). Moreover, the average height of native-born white males in the United States—the tallest in the world at the start of the 19th century (Steckel 1995, p. 1920)—declined by between 0.65 and 1.25 inches (depending on the estimate) between the birth cohorts of 1830 and 1860 (A'Hearn 1998; Floud et al. 2011; Komlos 1987; Zimran 2018). It was not until nearly the birth cohort of 1900 that average stature would begin to rise again (Fogel 1986; Steckel 1995; Zehetmayer 2011). This pattern is generally interpreted as indicating that the early stages of modern economic growth in the United States were not unambiguously welfare-improving.

Despite a large body of research devoted to describing the decline in health during the antebellum period, a definitive explanation has not been identified. Recent scholarship favors a combination of two mechanisms (Floud et al. 2011).[7] One, the *disease explanation*, holds that a variety of forces led to an increased exposure to disease during the antebellum period (Costa 1993; Fogel 1986; Steckel 1995). The second, the *food price explanation*, holds that the decline in height was the result of a rise in the relative price of food that led individuals to substitute away from food consumption towards the consumption of manufactures (Komlos and Coclanis 1997). Although these explanations hypothesize that forces beyond the expansion of transportation infrastructure in the period played a role in spreading disease and changing relative prices, both also posit a strong role for transportation, which, as discussed in section 2.1 above, can have such effects.

The empirical evidence underlying both of these explanations is limited, largely due to limited data availability in this period. Like the evidence described above on the relationship between transportation and health in this period, much of the evidence that has been marshaled in support of either or both of these explanations is suggestive, based either on cross-sectional correlations or on national time series. To my knowledge, there does not exist any work that shows directly that any particular force caused declines in height.[8] This paper, building on recent improvements in data availability and in methodological approaches to studying transportation, contributes to addressing this limitation by showing that the antebellum decline in health may be in part the product of the adverse consequences of transportation improvements. This result alone cannot distinguish between the food price and disease explanations, but it does make progress towards understanding the phenomenon, and can shed light on these two canonical explanations through investigation of the mechanism by which transportation acts on health.

---

[7] In fact, there are at least fifteen distinct explanations, some of which are summarized by Bodenhorn, Guinnane, and Mroz (2017, p. 175). Almost all, however, can be grouped into one of these two larger categories.

[8] Indirect evidence is provided by Costa (1993), Haines, Craig, and Weiss (2003), Hong (2007), Komlos (1987), Sunder (2011), Sunder and Woitek (2005), and Woitek (2003), among others.

3 Empirical Approach

3.1 Empirical Specification

The basic specification that I use to investigate the relationship between transportation and health is

$$h_{ijt} = \gamma_t + \delta_a + \beta T_{jt} + \mathbf{z}_j'\tau + \varepsilon_{ijt}, \tag{1}$$

where h_{ijt} is the height of individual i born in county j in year t (my measure of health), γ_t are birth cohort-specific intercepts, δ_a are indicators for each measurement age below 21 to address cases in which individuals are observed before reaching terminal height, T_{jt} is a measure of transportation linkage in the birth year, and $\mathbf{z}_j$ is a vector of various county-level control variables to be introduced in section 5 below.[9] Because the outcome of interest is observed at the individual level but the regressor of interest is observed only at the county level, I cluster standard errors throughout the analysis at the county level. My initial analysis estimates this equation by ordinary least squares. This specification is comparable to that used by prior studies of the transportation-health relationship in the antebellum United States, especially Haines, Craig, and Weiss (2003).

This framework assumes that the effect of transportation on height is described fully by the relationship of terminal height with transportation linkage in the birth year. While previous studies suggest that transportation in the birth year is likely to be more important than in any other year of life (e.g., Steckel 1995; Woitek 2003), it is possible to determine the consequences of relaxing this assumption. I do this in Appendix B, where I find that transportation linkage around the year of birth is more strongly associated with terminal stature than is transportation linkage in other phases of life.

A key concern with specification (1) is that any relationship that it uncovers between transportation and height may be spurious. For instance, a particular county may have been densely populated or highly urbanized for some reason besides transportation linkage, such as a favorable geographic location. When transportation infrastructure was constructed, the fact that this county was already developed would make it more likely to become linked to the network. Moreover, the sanitation consequences of population concentration might make this area unhealthy. This hypothetical relationship would produce a negative β in

---

[9] I do not include individual-level controls (e.g., occupation) for two reasons. First, the Union Army data, which are my source of all individual-level information, suffer from a large degree of missing data. Limiting the sample to observations with data on all fields of interest would have serious implications for statistical power. This limitation is exacerbated by the fact that successful census linkage is required to observe many variables of interest, and requiring such linkage would further reduce sample size. Second, any individual-specific variables are more properly considered outcomes of the presence of transportation and are therefore "bad controls."

specification (1) even if the true β were zero.[10]

One approach that I take to address such concerns is to augment specification (1) with the addition of county fixed effects α_j (requiring the omission of the county-specific controls $\mathbf{z}_j$) so that it becomes

$$h_{ijt} = \alpha_j + \gamma_t + \delta_a + \beta T_{jt} + \varepsilon_{ijt}. \tag{2}$$

This specification captures time-invariant county characteristics and exploits the panel structure of the data. It also improves on studies of the transportation-health relationship in the antebellum United States, in which panel data have not previously been available.

A concern that remains in equation (2) is that faster economic growth in a county-year driven by a force other than transportation might both affect health and attract transportation. The concern is similar to that expressed above, except that it applies to a county over only part of the sample period rather than the whole, and would thus not be captured by the county fixed effects α_j. One approach that I will use to address this concern is to include county-decade fixed effects rather than simply county fixed effects.

3.2 Measures of Transportation Linkage

I use two measures of transportation linkage in the empirical analysis. The first is a simple measure that takes a value of one in years in which a county was linked to water or rail transportation, and a zero otherwise. While it is a straightforward measure, it faces some important drawbacks. First, it does not capture the impacts of new forms of transportation entering already linked areas. This is exacerbated by the fact that all coastal counties are defined as having always been linked to the transportation network. Moreover, this binary measure does not capture changes in the transportation network that affect a county but take place far away from it in the network. Perhaps the most important such change in the study period is the construction of the Erie Canal, which had profound effects on the Midwest's ability to access markets despite all of the construction being located in the Northeast.

To address these shortcomings, I use Donaldson and Hornbeck's (2016) market access measure. Following an algorithm described in Appendix C, I compute approximate iceberg transportation costs, $\tau_{ijt} \geq 1$, between each county pair ij in each year $t \in \{1820, \ldots, 1847\}$. Market access in county i for year t is then defined as

$$m_{it} = \sum_j p_j \tau_{ijt}^{\theta}, \tag{3}$$

---

[10] Not all confounds must be in this direction. For instance, if better agricultural land attracted transport construction and raised incomes and health, a spurious positive β would arise.

where p_j is the population of county j in 1820. The choice to use 1820 population rather than year t population is a is made because allowing population to change over time would cause market access to capture both improvements in transportation linkage and population growth, which would have its own impacts on health.[11]

There are two issues that must be addressed before the market access measure can be used. The first concerns the choice of θ. In choosing the value of this parameter, I follow the example of Donaldson and Hornbeck (2016) and estimate equation (2) by nonlinear least squares, taking the logarithm of market access as defined in expression (3) as the variable T_{jt}. This estimation gives an estimate of $\hat{\theta} = -3.82$ (s.e. = 0.48), which I use throughout the analysis.[12]

The second issue concerns the interpretation of the coefficient β when T_{jt} is the logarithm of market access. Interpreting specific changes in this regressor (e.g., a ten percent increase in the logarithm of market access) is not informative, as the range of market access is affected by the choice of θ, which in turn impacts the estimate of β. Instead, the parameters β and θ must be interpreted jointly. I focus on the impact of a one-standard deviation increase in market access (0.30 log points).

3.3 Instrumental Variables

As an alternative identification strategy, I develop an instrument for market access that builds on the straight-line instruments commonly used in studying the economic impacts of transportation improvements (e.g., Atack et al. 2010; Banerjee, Duflo, and Qian 2012; Ghani, Goswami, and Kerr 2016; Hornung 2015).[13] It is based on the principle that antebellum internal improvements were intended to link major watersheds (the Atlantic, Great Lakes, and Mississippi) to one another and to major cities (Taylor 1951, p. 37).

Specifically, I draw a series of straight lines, depicted in Figure 1. The first set of lines, depicted in panel 1(a), are the shortest connections between the major watersheds, based on the steamboat-navigability of rivers in 1820.[14] The next set of lines, depicted in panels 1(b)–1(d), identifies the 25 largest cities over

---

[11] I have repeated the analysis with population fixed at 1840 and with year t population. Results in each case are similar to those using 1820 population, though the interpretation is different with year t population.

[12] Ultimately, the choice of θ is not very important. Any change in the value of θ used will be largely offset by changes in the estimated value of β (Donaldson and Hornbeck 2016, pp. 831–832). Indeed, when θ is set to -1, the estimates of β are qualitatively almost identical: the numerical estimates differ, but their interpretation is nearly identical.

[13] The precise methods generated in previous studies are not suitable for use in the context of the antebellum United States prior to the construction of railroads. For example, a Euclidean network of the type used by Banerjee, Duflo, and Qian (2012) is based on the existence of major cities that must be linked by transportation. However, in the United States, the major cities were all on the East Coast while construction of transportation was designed to link the East to the West. Similarly, Atack et al.'s (2010) survey cities instrument is better suited to the denser rail construction of the 1850s than to the earlier, geographically dispersed canal construction of earlier years.

[14] I group rivers with the major body of water that they flow into. For instance, the Hudson River is part of the Atlantic watershed and the Ohio River is part of the Mississippi watershed. Panel 1(a) treats Lake Ontario as a separate watershed, as it was not connected to the other Great Lakes by a navigable waterway until 1829.

10,000 population in each census year 1820–1840 (though it was not until 1840 that there were at least 25 such cities) and draws the shortest lines between these cities and the three major watersheds (Atlantic, Great Lakes, and Mississippi), provided that these lines are not more than 300 miles in length nor originate in the South (except for Virginia, Maryland or Washington, DC).[15] The repetition of lines between panels 1(b), 1(c), and 1(d) is not concerning, as the construction of a second line overlapping a first will have no impact.

I then compute market access as above and in Appendix C, with the following changes: (1) I begin with the transportation network in its 1820 state; (2) I treat the lines of Figure 1 as canals; (3) I augment the 1820 network by letting each line develop—beginning in 1820 for the lines in panel (a) of Figure 1 and from the decadal year for those in other panels—over a period of 15 years in equal increments, beginning at the originating city or at the easternmost watershed.[16] This alternative measure of market access is the instrumental variable, which I use to estimate equations (1) and (2) by instrumental variables.

As with any candidate instrument, the key concerns are relevance and excludability. Relevance will be formally established in estimation of the first-stage equations but is already suggested by Figures 1 and 2(a). Figure 1 (and comparison to Figure A.1) reveals that the location of these lines is a good approximation of actual construction. For instance, the line linking the Atlantic and Great Lakes watersheds in panel 1(a) is close to the Erie Canal; the lines in Pennsylvania in panel 1(b) closely approximate the construction of Pennsylvania's Main Line; and the lines in Ohio, Indiana, and Illinois in panels 1(a), 1(c), and 1(d) are also close approximations to actual construction. Because these lines are used to compute an alternative measure of market access, they also affect counties away from where they are constructed, as the Erie Canal did. Moreover, as shown in Figure 2(a) for the whole sample and in Figure 2(b) for the specific case of Montgomery County, Ohio (an arbitrary example), the temporal development of the market access implied by the instrument tracks well with that of the actual measure.

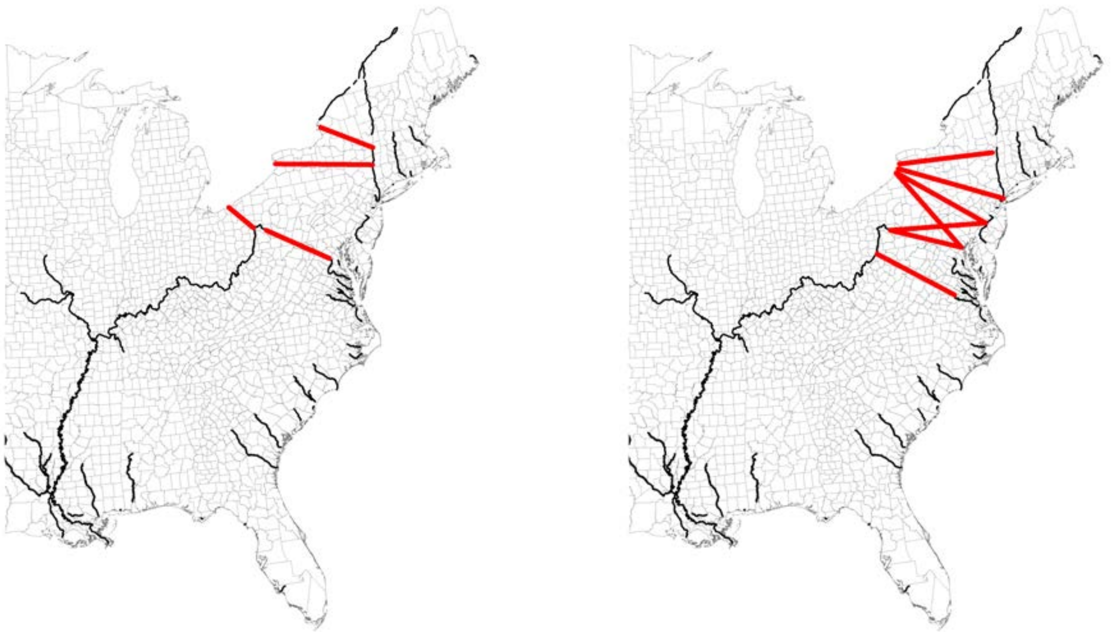
Excludability of the instrument requires the following assumptions. In the cross-section, the identification assumption is comparable to that of other straight-line instruments. It is that, after excluding counties from which the lines in panels 1(b)–1(d) originate, counties on or near the straight lines of Figure 1 are similar to those further from the lines except in their likelihood to receive beneficial surges in market access. The identification assumption in the second dimension—the time series—has fewer analogs in the literature.[17] It

---

[15] These are actually based on the urban population of counties, rather than city populations. Southern cities are excluded to better capture the true lack of internal improvements there.

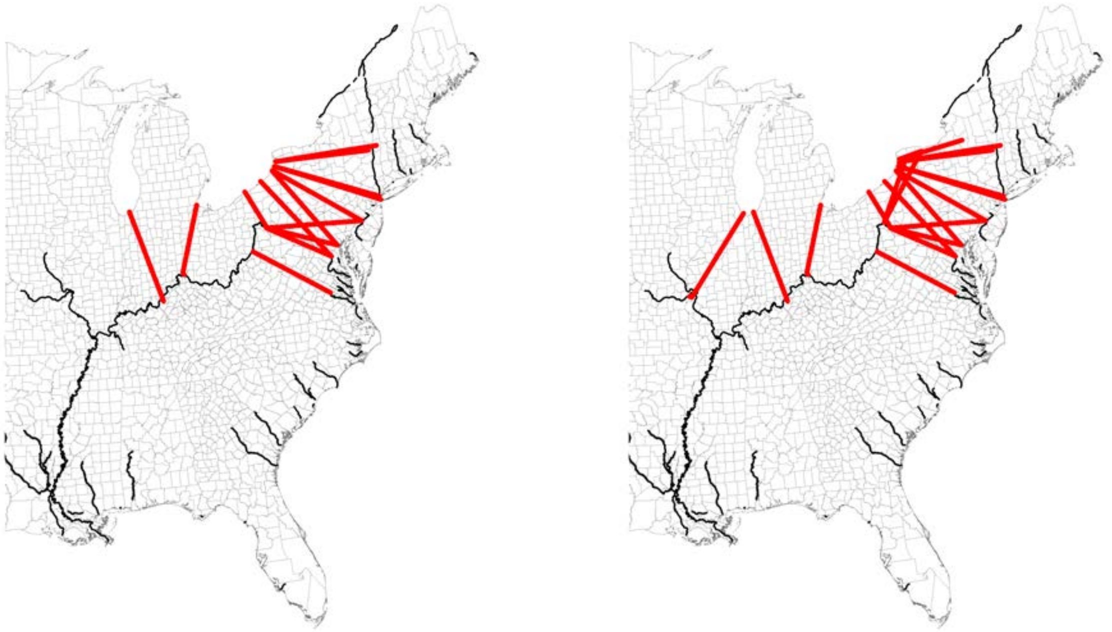
[16] An example of the evolution of one such line is shown in Figure A.2. I have also used a 10 year development period, but the variable generated in this way does not satisfy the relevance condition for instrumental variables, whereas the variable generated with a 15 year development period does.

[17] An exception is Hornung (2015), who creates a dynamic straight-line instrument based on the principle that future construction is likely to link ends of existing lines to target destinations along the shortest possible route. My approach differs from this by not being based on actual construction.



(a) Watersheds

(b) 1820

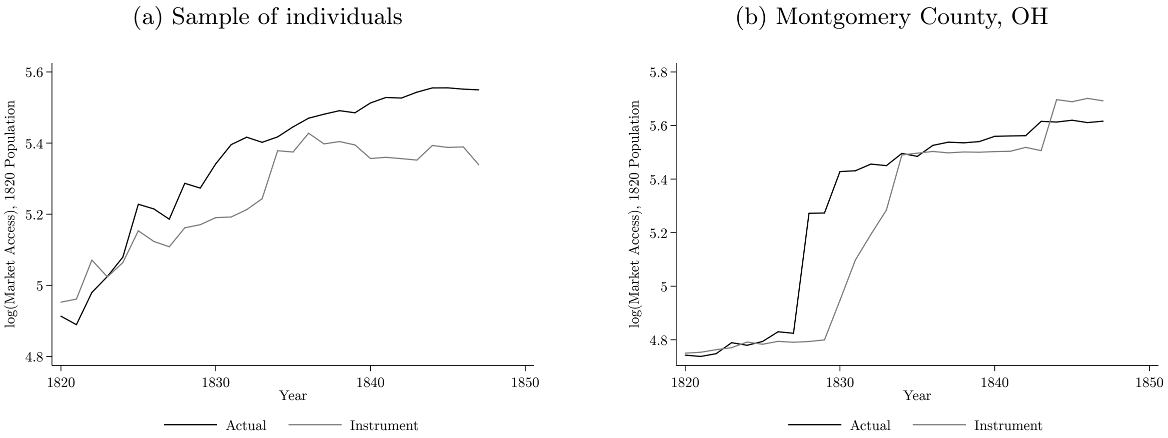


(c) 1830

(d) 1840

Figure 1: Straight lines for instrumentation

Note: All maps include the 1820 transportation network. In panel 1(a) the lines presented are those linking the major watersheds to one another. The lines presented in panels 1(b)–1(d) link the top 25 cities with over 10,000 population (usually there are fewer than 25) to the major watersheds with lines of 300 miles or less outside of the South, except for Virginia, Maryland and Washington, DC.



(a) Sample of individuals

(b) Montgomery County, OH

Figure 2: Actual and hypothetical market access

Note: The line labeled "Actual" plots the average log market access. The line labeled "Instrument" plots instrument calculated using the straight lines of Figure 1. Panel 2(a) covers the benchmark sample of individuals using market access and the instrument in the year of birth. Panel 2(b) covers the example of Montgomery County, Ohio.

is that counties closer to the origin of a straight line in Figure 1 are not fundamentally different from those further from the origins, except that they are likely to be linked to the transportation network sooner. A clear concern is that the origins of the lines represent points of interest, such as cities; but given the high costs of wagon transportation, excluding the terminus counties should render the remaining counties equally isolated.[18] I provide some empirical support for these assumptions in section 5 below.

One concern with this instrument can be easily dismissed. Although the evolution of the straight lines is based on a fixed annual expansion, the instrument is not a time trend (indeed, year-specific indicators are included in all specifications). Instead, the instrument, like the measure of market access, evolves discontinuously in response to a new transport link. An example of the evolution of the instrument and of market access in a single county in shown in Figure 2(b), which describes the experience of Montgomery County, Ohio. The rapid increases in market access in the 1820s come from the construction of the Miami and Erie Canal, which passed through the county and linked it to the Ohio River. The rapid increase in the instrument in the 1830s comes from the passage of the straight line linking Hamilton County, Ohio to the Great Lakes through the county linking it to the Ohio River. The smaller increase in the 1840s comes from the completion of that line, completing the hypothetical linkage to Lake Erie.

---

[18] This view is supported by Donaldson and Hornbeck's (2016) finding that Fogel's (1964) proposed canals were not good substitutes for railroads because of the value of railroads in reducing wagon haul distances. This implies that the reduction of wagon haul distances necessary to reach transportation infrastructure is particularly important, and supports the notion that areas even a short wagon haul away from a city would be relatively isolated—a view supported by the poor roads of the antebellum period.

4 Data

4.1 Sources

Information on transportation infrastructure is given by GIS shape files produced by Atack (2015, 2016, 2017). These files, which also form the basis for Donaldson and Hornbeck's (2016) market access calculations, provide the location of all steamboat-navigable rivers, canals, and railroads in the continental United States constructed or opened prior to 1914.[19] These files do not provide information on the location of turnpikes, but this omission is unlikely to have a major effect on results because of the high costs of wagon transportation (Donaldson and Hornbeck 2016; Taylor 1951). Until 1850, these shape files also provide the year in which any particular form of transportation first became operational (or navigable); after 1850, these are known yearly for water transportation, but only every two years for railroads until 1860. Together with the categorization of all coastal counties (either on the Atlantic, the Gulf, or the Great Lakes) as having always had access to water transport, it is thus possible to determine whether a particular county was linked to the transportation network in any year in the sample period (1820–1847),[20] and to perform the cost calculations necessary for the market access measure for each year in the sample period.

The information on transportation that this source provides improves on that available in prior studies of the transportation-health relationship in the antebellum United States. As discussed by Atack (2013), earlier studies of this period relied on potentially inaccurate information on the location of transport infrastructure and did not have information on the opening dates of this infrastructure. For this reason, the measure of transport linkage used by Haines, Craig, and Weiss (2003) and Yoo (2012) was an indicator for having water transport in 1840. The new shape files of Atack (2015, 2016, 2017) enable me to improve on this measure, both through the improved accuracy of the locations of infrastructure and by providing a temporal component to the evolution of the transport network.

I measure health using adult height. This measure, which is commonly used as an indicator of health in historical and developing contexts (e.g., Deaton 2007; Floud et al. 2011), is unique in the antebellum United States in that it is perhaps the only measure of health that can provide insights into health for the bulk of the population for a number of years.[21] Average stature is increased by greater calorie consumption and a better sanitary environment, while strenuous physical labor, malnutrition, and chronic disease tend to

---

[19] I have supplemented these files with the canals and rivers of the St. Lawrence and Champlain waterways.

[20] The "year of transportation arrival" refers to the year in which non-wagon transportation first became possible. The development of the transportation network divided by mode is presented in Figure A.1.

[21] An alternative measure, the crude death rate, is available in the antebellum period, but only for a single year (1850). It is therefore not possible to exploit changes over time in the transport network, as I do below in studying the impacts on height. Time series of life expectancy are also available, but cover only specific subsets of the population.

decrease average stature (Deaton 2007; Floud et al. 2011; Steckel 1995).[22]

Data on the heights of men born in the United States in the years 1820–1847 are available from the records of enlistments in the Union Army during the Civil War (Records of the Adjutant General's Office 1861–1865). This widely used source is informative of height, place of birth, age, year of enlistment, and place of enlistment. I combine three random samples of this source. The first comes from the Union Army Project (Fogel et al. 2000), which provides information on a random sample of approximately 40,000 individual observations from the original records. The second is provided by Cuff (2005), yielding approximately 12,000 additional observations of men born in the state of Pennsylvania and serving in Pennsylvania regiments. Finally, I collected and digitized approximately 3,000 additional observations from the original records.

As is standard in uses of these data, I restrict the sample to white men born in the Northeast or the Midwest. I also exclude individuals measured before age 18, which, due to the timing of the Civil War, implies that the youngest birth cohort that is systematically observed is that of 1847, as this cohort would have turned 18 in 1865, the last year of the Civil War. I also exclude birth cohorts older than 1820 because of the relative lack of representation of these older cohorts in the military. Finally, I limit the sample to those for whom county of birth could be determined and for whom height, birth year, and age of measurement are known.[23] After imposing these restrictions, 31,403 observations remained for all counties (rural and otherwise). For a subset of these observations, the county of enlistment could also be determined.

For two reasons I restrict attention to individuals born in counties that had no urban population in 1820, which reduces the sample to 25,567 individuals.[24] First, there is little variation over time in the transportation linkage of the excluded counties, as they are nearly all on major transport routes in 1820. Second, there are many forces that may have affected health in cities that would be difficult to disentangle.

A key question regarding the enlistment data is whether they are representative of the broader population of interest—native-born white males in the birth cohorts of 1820–1847. The over-sampling of Pennsylvanians

---

[22] Although declining height is generally understood to imply deteriorating health in historical contexts (e.g., Fogel 1986; Steckel 1995), it is also possible that declining height might be an indication of a shift from selection to scarring. That is, declining average height might actually indicate better health if it allowed individuals who would have died in infancy to survive but to reach shorter average terminal height than those who would have survived to adulthood in the absence of improved health (Deaton 2007). Unfortunately, the data necessary to determine whether changing height is the result of selection or scarring in the context of this paper are not available. There exist data on mortality (Haines, Craig, and Weiss 2003), but these are available only for 1850 and thus do not permit the same panel analysis as do the height data. As a result, I rely on the standard interpretation of the historical heights literature, on the negative correlation between terminal height and those mortality rates that are observed in this period (e.g., Floud et al. 2011; Fogel 1986; Haines, Craig, and Weiss 2003; Steckel 1995), and on the results presented in Table A.1 showing that death rates were greater in counties with greater market access, to interpret declines in average stature as deteriorations in health, and vice versa.

[23] In most cases, a county of birth is directly reported, and the individual is assigned to that county. In some cases, a city or town of birth was reported instead. These were manually assigned to the appropriate county. In cases where a state of birth is reported but no county is reported, and in which the individual was linked to a census in 1850 or 1860 (linkage was only performed for observations collected by Fogel et al. 2000), the individual is assigned to his county of residence in the first census in which he is observed.

[24] Figure A.3 indicates the counties removed from the sample by this restriction.

is one obvious concern, which I address by re-weighting so that the distribution of states of residence matches that of the 1860 census. A more nuanced concern is that selection into military service was non-random (Bodenhorn, Guinnane, and Mroz 2017). While this is theoretically a valid concern, its potential severity is mitigated by the fact that nearly half of the population at risk for observation and military service enlisted (Zimran 2018). For this reason, the Union Army data are considered to be representative of the white male population of the Northern states (Fogel et al. 2000). This view is reinforced by Zimran's (2018) formal investigation of bias in historical height data sources, which finds that the height data provided by the Union Army records suffer from relatively little bias.[25]

Another concern is that entrance into the Union Army was subject to a minimum height requirement. Although this requirement was not stringently enforced, the left tail of the height distribution was under-represented.[26] The common approach in the historical heights literature is to use a reduced-sample maximum likelihood estimator that omits any observations below the cutoff point and assumes normality of the stature distribution (A'Hearn 1998). In the present context, however, the omission of data is undesirable because of the considerable loss of degrees of freedom through the inclusion of county fixed effects in the main specifications and because of the subsequent introduction of instrumental variables. As a result, the results reported below do not use such a truncation-corrected regression.

Finally, I gather county-level data from the decennial United States censuses of 1820–1850 (Manson et al. 2017). This source provides county-level population, urban population (which, following the standard census definition, is the number of people living in places of population 2,500 or greater), and data on agricultural and manufacturing production and employment. I supplement these data with Craig, Copland, and Weiss's (2012) data on the nutritional value of agricultural production for 1840 and 1850 and with data on suitability for wheat and corn production from the Food and Agriculture Organization (2002).

I standardize all data—including the transportation linkage indicator, market access computations, assignment of counties of birth, and the county-specific data described above—to 1860 county boundaries. I focus on 1860 counties because the counties of birth of enlisters are reported in the years 1861–1865, and enlisters are likely to have reported their place of birth based on the boundaries existing at the time of the report. Where necessary, I standardize variables to 1860 county boundaries using Hornbeck's (2010) method.

---

[25] The issues of selection bias also inform my choice to focus on the birth cohorts of 1820–1847. While height data are available from military records for cohorts throughout the later antebellum period and nineteenth century, Zimran (2018) shows that combining data from the Civil War and from later periods can lead to strong selection bias, generated in part by the fact that after the end of the Civil War, only a small fraction of the population entered the military and had its height observed.

[26] This is shown in Figure A.4.

4.2 Summary Statistics

Using the sources described above, I create and merge two data sets. The first is a panel data set with observations at the county-year level on transportation linkage and market access. The second provides individual-level data from the Union Army on native-born white males with known height, county of birth, year of birth, and age of measurement, born between 1820 and 1847. Merging these two data sets links each individual in Union Army data to the characteristics of his county of birth in his year of birth.

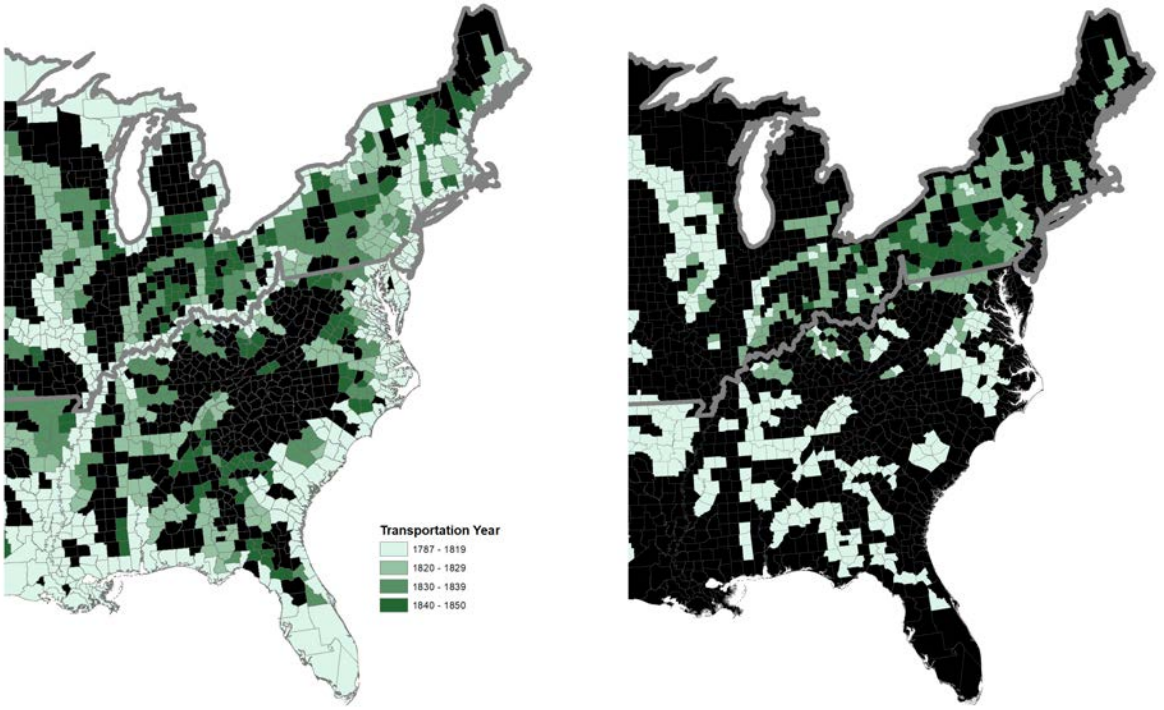
Table 1 summarizes the county-level measures of transportation linkage, divided by region and decadal year, and weighted by population. There is a clear pattern of growth over time in the fraction of the population living in a county linked to the transportation network. In the entire sample region, less than 40 percent of the population lived in a county that was linked to the transportation network in 1820. By 1850, this fraction had risen to over 80 percent. The Northeast and the Midwest viewed separately exhibit similar patterns, although the population of the Northeast is consistently more linked than is that of the Midwest.

Table 1: Summary statistics for county-level data

	All				Midwest				Northeast			
Variable	(1) 1820	(2) 1830	(3) 1840	(4) 1850	(5) 1820	(6) 1830	(7) 1840	(8) 1850	(9) 1820	(10) 1830	(11) 1840	(12) 1850
Transportation Present	0.396	0.562	0.711	0.805	0.316	0.476	0.599	0.704	0.420	0.600	0.796	0.907
	(0.489)	(0.496)	(0.454)	(0.397)	(0.465)	(0.500)	(0.491)	(0.457)	(0.495)	(0.491)	(0.404)	(0.291)
log(Market Access), 1820 Pop.	5.005	5.420	5.563	5.650	4.835	5.273	5.429	5.519	5.056	5.484	5.664	5.782
	(0.495)	(0.296)	(0.261)	(0.252)	(0.370)	(0.240)	(0.225)	(0.220)	(0.517)	(0.296)	(0.240)	(0.211)
Counties	945	945	945	945	774	774	774	774	171	171	171	171

Notes: The sample in columns (1)–(4) includes all counties with no urban population in 1820. Columns (5)–(12) divide this sample by region. Means presented with standard deviations in parentheses. Observations weighted by population.

Figure 3 provides a graphical summary of the spread of transportation linkage over this period. Panel 3(a) shows that the transportation network gradually spread inland during this period. The sample period began with only the coasts and the counties bordering the major internal waterways being linked to the network, and concluded with much of the interior being linked. However, as discussed in section 3.2 above, this binary measure is problematic. Beyond the conceptual difficulties that it poses, there simply are not many observations of height data in counties experiencing changes in transport linkage. This is shown in panel 3(b), which isolates the counties in which there was a change in transportation linkage between the years 1820 and 1847 and divides them into three groups. The first (shaded in the lightest color), which represents many of the counties in the South or westernmost Midwest are not represented by any individual height observations, or all the representation comes from before or after the change in transportation linkage. The second group (shaded somewhat darker) has individual height observations from both before and after the change in linkage, but has only a small number of observations of stature in at least one of these groups. Only the third group (the darkest shade, besides the black background), consisting of 31 counties, mostly in Pennsylvania, has at least 25 observations of individual heights both before and after the change in transportation linkage.



(a) Year of arrival

(b) Changes in linkage

Figure 3: Counties by transportation change and sample coverage

Note: Panel 3(a) presents the year in which each county received a transport link, treating coastal counties and counties with an always navigable river as being linked in 1787. Panel 3(b) marks counties experiencing a change in transport linkage in 1820–1847. Counties in black experienced no change in transportation linkage between 1820 and 1847. The lightest colored counties experienced a change in transportation linkage in this period, but have no observations either before or after the change. The darker counties have observations both before and after the transportation change, but only the darkest counties have at least 25 observations both before and after the change. Sample region indicated by thick boundary.

Fortunately, the market access measure helps to address this concern. In particular, it generates variation in the magnitude of transportation linkages and allows new linkages to affect counties other than only those through which the infrastructure passes. The set of "treated" counties can thus be considered larger and there is more variation in the treatment. This measure is also summarized in Table 1. As with the linkage measure, this measure shows patterns of growth over the study period, and of greater market access in the Northeast than in the Midwest. Directly interpreting the magnitude of the market access measure is not possible given the discussion of section 3.2 above, but it is still possible to compare the changes over time and differences over regions in other terms. For instance, the increase between 1820 and 1850 is equal to about two and a half standard deviations of the measure in 1850.

The development of the market access measure over time is described graphically in Figure 4. This Figure depicts the change in market access in each decade, shading counties with greater increases darker.[27] It shows that market access captures changes that transportation linkage does not. For instance, the counties in the sample region with greatest increase in market access between 1820 and 1830 are those bordering the upper Mississippi and the Great Lakes, as well as those in western New York. These changes reflect the opening of the Erie Canal and of the upper Mississippi. Between 1830 and 1840, large increases are observed in central Pennsylvania and in Indiana and Ohio, reflecting canal construction. Finally, between 1840 and 1850, large increases are again observed in Indiana and Ohio, also reflecting canal construction.

Table 2 provides summary statistics at the individual level for heights and for other variables for the complete sample and for various subsamples. Column (1) represents the benchmark sample of analysis—native-born white males whose counties of birth had no urban population in 1820. Columns (2) and (3) divide the sample by region, and columns (4) and (5) divide the sample by whether the individual's county of birth was linked or unlinked to the transportation network in the individual's year of birth.

A majority of the sample was born in the Northeast (even after adjusting for the Pennsylvania oversample)—a mechanical consequence of weighting the data to reflect state population in 1860. Figure 5 delves into the geographic distribution of data in further detail. It presents the number of individual height observations by county, separating Pennsylvania from the rest of the country as a result of its over-representation in the sample. On the whole, the sample tends to draw from the more populous areas of the country. Importantly, it includes almost all counties in the Northeast and the Midwest.[28]

Table 2 also shows that the benchmark sample was 68.1 inches tall on average, and columns (2) and (3) reveal that the Northeast suffered a height disadvantage of about half an inch relative to the Midwest. A height disadvantage of about 0.4 inches is present for those born in transportation-linked counties.[29]

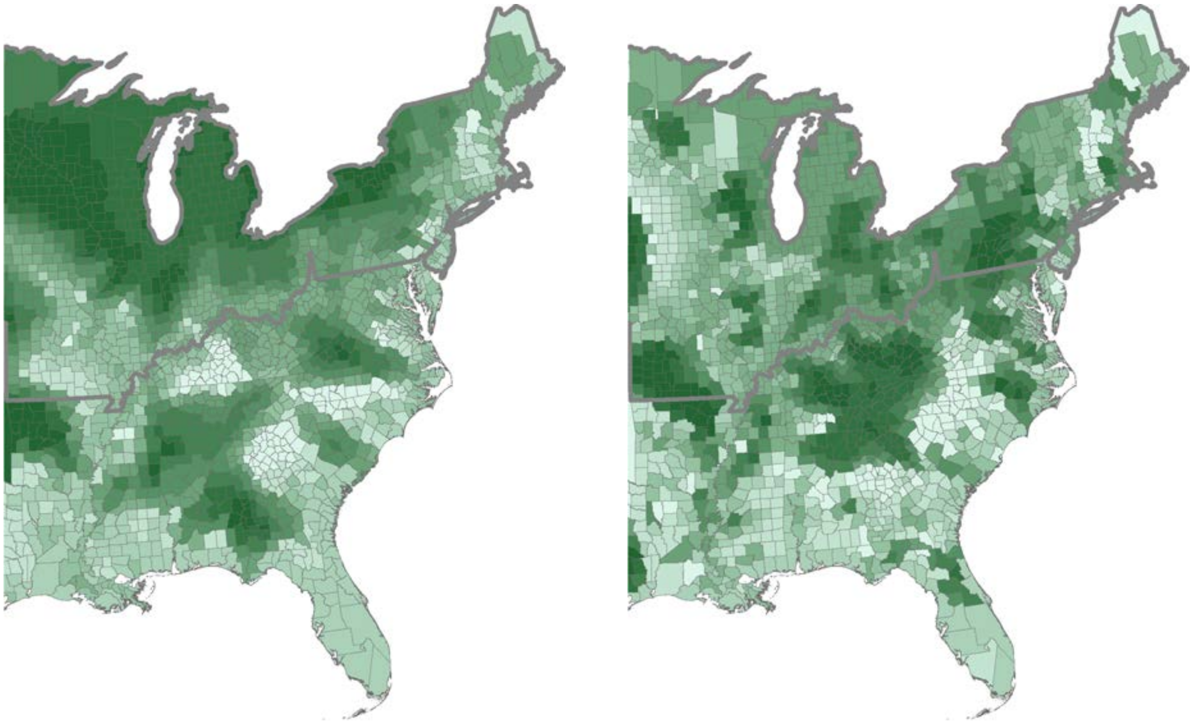
There are also differences between regions and between linked and unlinked counties in measures of population concentration. Consistent with the expected effects of transportation linkage (and with a variety of endogeneity concerns), there is a considerable advantage in urbanization and population density at birth

---

[27] The scale in each panel is different, dividing counties by deciles of the increase in market access. The levels of market access in each year are presented in Figure A.5.

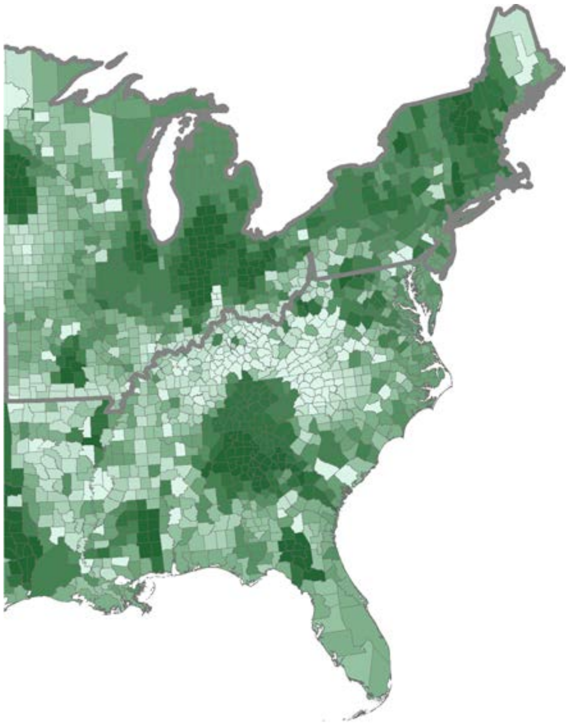
[28] The number of observations by birth cohort is given in Figure A.6. The number of observations is increasing in the birth cohorts from 1820 to the early 1840s, consistent with the idea that younger individuals would be more likely to join the military. The number of observations then falls sharply among the birth cohorts of the mid 1840s, consistent with the requirement to be at least 18 years of age to enlist.

[29] Figure A.4 presents a histogram describing the distribution of individual height observations. It shows the tendency to heap on whole inches and to exhibit shortfall below the minimum height requirement of 64 inches, but is otherwise regular.



(a) 1820–1830

(b) 1830–1840



(c) 1840–1850

Figure 4: Changes in market access by decade.

Note: Each panel shows the change in market access over the listed decade. For example, the panel labeled "1820–1830" shows the change in market access from 1820 to 1830. The scales are not comparable across years; instead, they depict deciles of the change in market access for that decade. Darker counties experienced a greater increase in market access. Sample region indicated by thick boundary.

for individuals born in linked counties.[30] There is also an advantage in population density at birth for Northeasterners, though the level of urbanization at birth was similar for the Northeast and the Midwest (recall that any county with an urban population in 1820 is omitted). While there is a premium in agricultural suitability for the Midwest, there does not appear to be a meaningful difference in agricultural suitability of the birth county for individuals born in linked and unlinked counties.

Table 2: Summary statistics for individual-level data

Variable	(1) All	(2) MW	(3) NE	(4) Linked	(5) Unlinked
Individual-level data					
Height	68.064	68.326	67.843	67.916	68.343
Inches	(2.640)	(2.632)	(2.626)	(2.631)	(2.636)
Birthyear	1838.262	1839.100	1837.555	1839.039	1836.787
	(6.231)	(5.729)	(6.542)	(5.616)	(7.023)
Age of Enlistment	24.277	23.484	24.946	23.511	25.731
	(6.228)	(5.666)	(6.591)	(5.572)	(7.088)
Enlisted in Different State	0.280	0.315	0.251	0.266	0.308
	(0.449)	(0.464)	(0.434)	(0.442)	(0.462)
Enlisted in Different County	0.631	0.721	0.563	0.604	0.686
	(0.482)	(0.448)	(0.496)	(0.489)	(0.464)
County-year-level data					
Urbanization at Birth	0.017	0.015	0.018	0.025	0.002
	(0.060)	(0.062)	(0.058)	(0.072)	(0.013)
log(Population Density) at Birth	3.274	2.802	3.629	3.505	2.809
	(1.103)	(1.206)	(0.862)	(0.992)	(1.167)
Transportation Linkage at Birth	0.655	0.567	0.729		
	(0.475)	(0.495)	(0.445)		
log(Market Access) at Birth, 1820 Pop.	5.462	5.378	5.532	5.603	5.192
	(0.300)	(0.265)	(0.310)	(0.192)	(0.284)
County-level data					
Midwest	0.457			0.396	0.573
	(0.498)			(0.489)	(0.495)
Northeast	0.543			0.604	0.427
	(0.498)			(0.489)	(0.495)
log(Wheat Suitability)	8.693	8.914	8.508	8.702	8.678
	(0.316)	(0.166)	(0.292)	(0.268)	(0.389)
log(Corn Suitability)	8.548	8.787	8.348	8.563	8.522
	(0.404)	(0.214)	(0.417)	(0.354)	(0.483)
Observations	25,567	10,210	15,357	16,875	8,692

Notes: Sample includes all height observations of native-born white males born in the Northeast or Midwest in counties with no urban population in 1820. Means presented with standard deviations in parentheses. Observations weighted to correct for oversampling. Linked indicates individuals born in linked counties; unlinked denotes the opposite. MW denotes Midwest; NE denotes Northeast. The number of observations refers to the number of individuals in the sample with known height, year of enlistment, age of enlistment, and county of birth.

Finally, about 27 percent of the sample enlisted in a state other than the state of birth (state of enlistment is determined by the state of the regiment in which an individual enlisted), while nearly 63 percent enlisted in a county other than the county of birth (limiting the sample to those enlisting in the state of their regiment).[31]

---

[30] For intercensal years, the urban and total populations are imputed by assuming constant growth rates between censuses. These imputations are not used in analysis below, but are useful for developing a sense of the divisions of the sample by urbanization and population density.

[31] In some cases, individuals enlisted while the regiment was in the field. As I do not wish to consider military deployment

The probability of enlisting in a county or state other than that of birth was greater for Midwesterners but smaller for individuals born in counties linked to the transportation network.

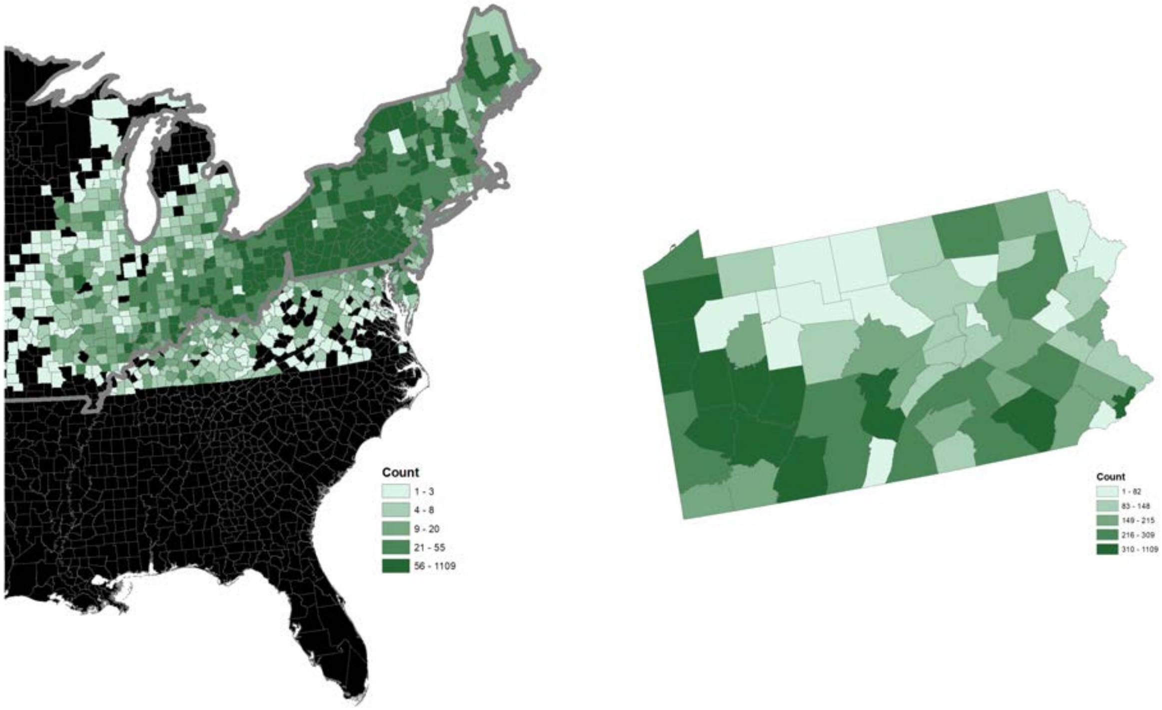


Figure 5: Number of observations by county

Note: This Figure includes both rural and non-rural counties and indicates the number of native-born observations of stature listing a birth place in each county with information on height and age of enlistment. Pennsylvania is displayed separately because of the oversample caused by the incorporation of the Cuff (2005) data. Sample region indicated by thick boundary.

5 Results

5.1 OLS Results

I begin the analysis by estimating equation (1) by ordinary least squares using the binary indicator of transportation linkage as the explanatory variable of interest T_{jt}. Results of this estimation are presented in columns (1)–(5) of Table 3. The regression of column (1), which includes only birth year indicators, age-of-measurement indicators, and no other controls, yields a negative and statistically significant relationship between transportation presence in the birth year and average stature. This relationship is robust to the inclusion of state-specific fixed effects in column (2), though this addition reduces the magnitude of the

---

as a form of migration, I exclude these individuals when considering county-level migration.

estimated coefficient by about half. This latter estimate indicates that individuals whose counties of birth had some sort of transportation linkage in their birth year were 0.17 inches shorter than those whose counties of birth were unlinked in their year of birth. This magnitude is large compared to the contemporaneous urban height penalty of 0.29 inches (Zimran 2018). It is also roughly comparable in magnitude to the estimates of Haines, Craig, and Weiss (2003), whose benchmark results indicate that transportation linkage in the county of birth (though not necessarily in the year of birth) was associated with a height penalty of about 0.25 inches.[32] This similarity of results is not trivial, as my transportation measure, due to the availability of Atack's (2015, 2016, 2017) data, is more refined, as discussed above.

Table 3: OLS regressions

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Transport	-0.336^{***} (0.073)	-0.168^{***} (0.057)	-0.069 (0.067)	-0.071 (0.067)	-0.069 (0.070)					
log(Market Access), 1820 Pop.						-0.994^{***} (0.117)	-0.564^{***} (0.119)	-0.389^{***} (0.150)	-0.367^{**} (0.151)	-0.370^{**} (0.163)
Observations	25,567	25,567	23,567	23,567	23,567	25,567	25,567	23,567	23,567	23,567
R-squared	0.055	0.073	0.077	0.079	0.105	0.061	0.074	0.077	0.079	0.106
State FE	No	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes
Controls	No	No	Yes	Yes	Yes	No	No	Yes	Yes	Yes
Birth Year × Region FE	No	No	No	Yes	No	No	No	No	Yes	No
Birth Year × State FE	No	No	No	No	Yes	No	No	No	No	Yes

Significance levels: *** p<0.01, ** p<0.05, * p<0.1
Notes: Dependent variable is height in inches. Sample includes individuals born in the Northeast or Midwest in counties with no urban population in 1820. All specifications include birth year and measurement age fixed effects. Standard errors clustered at the county level. Observations weighted to correct for oversampling.

In column (3), I repeat the specification of column (2) with the addition of a variety of county-level controls. Some of these control variables are those that Haines, Craig, and Weiss (2003) include in their analysis—1840 calorie and protein production, Herfindahl indices for calorie and protein production, and 1850 values of farms and capital in manufacturing. I add several other variables that may have impacted health. These include area and 1820 population (to capture population concentration in 1820); 1840 cattle and swine stocks; 1840 employment by sector and values of agricultural and manufacturing output. All of these variables are included in log form and I also include the log of population in 1840 and 1850 in order to make the other measures effectively per-capita. I also include third-degree polynomials in the logarithm of distance from New York and Cincinnati. These controls are intended to capture a variety of county characteristics, such as agricultural productivity, density, and geography, that might generate health differences even in the absence of a transport link. The post-1820 values are included with full recognition that their 1820 values would be preferable (later values may be "bad controls"). However, due to the limited

---

[32] Haines, Craig, and Weiss (2003) also do not limit the sample to only rural areas, as I have done.

data availability of the antebellum period the inclusion of data on, (for example) agricultural production is not possible prior to 1840, and I err on the side of controlling for the features that these measures capture rather than not doing so.

The inclusion of these controls in column (3) reduces the magnitude of the estimated coefficient on transportation linkage and renders the estimated coefficient statistically insignificant. While the magnitude of the resultant coefficient is non-negligible, it is considerably smaller than the estimates of columns (1) and (2). This indicates that the relationship in columns (1) and (2) may be the product of omitted variables bias. The addition in column (4) of interactions of birth year and region fixed effects, or of the interaction of state and birth year fixed effects in column (5) has little impact on the estimates.

To determine whether the lack of a meaningful relationship between transportation linkage and height in the presence of controls is the product of deficiencies in the binary measure or a true absence of a relationship, columns (6)–(10) of Table 3 repeat the analysis of columns (1)–(5), but replace the binary measure of linkage with the logarithm of market access as the explanatory variable of interest T_{jt}. Columns (6) and (7) estimate equation (1) without the additional county-level controls, without and with the inclusion of state fixed effects, respectively. As was the case with the binary measure of transportation linkage, a large, negative, and statistically significant coefficient is present on the measure of transportation linkage, and is nearly halved (but is otherwise robust) when state fixed effects are included. In particular, the estimates of column (7), which include the state fixed effects, indicate that a one-standard deviation increase in market access (0.30 log points, as shown in Table 2) is associated with a reduction in average height of 0.17 inches.

Columns (8)–(10) repeat this estimation, including the various county-level control variables, and the region-by-birth year or state-by-birth year indicators. Unlike their analogs in columns (3)–(5), the addition of these controls to regressions of height on market access does not eliminate the statistical significance of the negative relationship between market access and height. Moreover, the impact of the inclusion of the controls on the magnitude of the coefficient is smaller than it was for the transport indicator. In particular, the estimates of column (10), which includes the state-by-birth year indicators, imply that a one-standard deviation increase in market access is associated with a decline in average stature of 0.11 inches, or about 1.6 times the implied impact of a transportation linkage in its analog, column (5).

On the whole, these estimates suggest that there is a negative correlation between transportation linkage and health as implied by average stature, and that the elimination of this relationship in columns (3)–(5) of Table 3 is the product of deficiencies of the transport indicator rather than of omitted variables bias.

5.2 Fixed Effects Results

Like the conclusions of existing work on health in the antebellum United States, the estimates of Table 3 do not address concerns of endogeneity such as those discussed in section 3 above. Indeed, these are merely correlations, and may be driven by transportation construction in areas that were unhealthy for reasons unrelated to transportation. The structure of my data, in particular the ability to describe the evolution of the transportation network over time, enables the estimation of equation (2) to partially address these concerns.[33] Results of this estimation are presented in columns (1)–(5) of Table 4. I begin in column (1) by estimating specification (2) with the transport linkage indicator as the regressor of interest. Given the binary regressor, this coefficient can be interpreted as a generalized difference-in-differences coefficient. The estimated coefficient is -0.037, which is smaller than the estimates including controls in Table 3, and is statistically insignificant. Given the limitations of the transport linkage indicator, as discussed above, the absence of a meaningful transport-health relationship using this regressor is not surprising.

Table 4: County fixed-effects regressions

Variables	(1)	(2)	(3)	(4)	(5)
Transport	-0.037 (0.114)				
log(Market Access), 1820 Pop.		-0.519^{***} (0.193)	-0.657^{**} (0.280)	-0.441^{**} (0.187)	-0.323 (0.217)
Observations	25,567	25,567	25,567	25,567	25,567
R-squared	0.124	0.124	0.171	0.127	0.154
Birth Year × Region FE	No	No	No	Yes	No
Birth Year × State FE	No	No	No	No	Yes
County × Decade FE	No	No	Yes	No	No

Significance levels: *** p<0.01, ** p<0.05, * p<0.1
Notes: Dependent variable is height in inches. Sample includes individuals born in the Northeast or Midwest in counties with no urban population in 1820. All specifications include birth year, measurement age, and county fixed effects. Standard errors clustered at the county level. Observations weighted to correct for oversampling.

Column (2) estimates the same specification with the logarithm of market access as the regressor of interest. Unlike specification (1), this estimation approach relates within-county changes in market access to within-county changes in average stature, making no cross-county comparisons. This column reveals that the negative and statistically significant relationship between market access and height is robust to the inclusion of the fixed effects, and thus to the concerns that they address over endogeneity. Moreover, at -0.519, the magnitude of the coefficient is comparable to the estimates of Table 3.[34]

---

[33] The county-fixed effects approach has the added benefit of not requiring the inclusion of potentially endogenous controls such as the 1840 and 1850 controls above. Instead, the county-specific characteristics that these are meant to capture will be captured by the fixed effects.

[34] The specification of column (2) is the one estimated by nonlinear least squares. The estimates are $\hat{\beta} = -0.519$ (s.e. = 0.201)

This result and approximate magnitude is robust to the inclusion, in column (3), of county-decade fixed effects, rather than simply county fixed effects, in order to more flexibly address county-specific characteristics that may be time variant. Columns (4) and (5) supplement the county fixed effects with region- and state-by-birth year indicators, respectively. While the inclusion of these indicators reduces the magnitude of the estimated coefficients, and in the case of column (5) it is reduced to the point of statistical insignificance ($p = 0.138$), the rough magnitude and sign of the coefficient is retained, supporting the conclusion that transportation improvements generated declines in stature-implied health.

5.3 Instrumental Variables Results

As an alternative approach to addressing the endogeneity issues facing the estimates of Table 3, I implement the straight-line instrument strategy introduced in section 3.3 above. Before delving into the estimates, I briefly explore, in Table 5, the evidence in support of excludability of the instrument. In particular, I relate the characteristics of counties that are observed in 1820 to the lines of Figure 1. Given the sparsity of data available in the early censuses, the only measures available are population density and the measures of agricultural suitability.[35]

Table 5: Correlates of instrumental variables line placement

Variables	(1) Wheat	(2) Corn	(3) Dens.	(4) Wheat	(5) Corn	(6) Dens.	(7) Wheat	(8) Corn	(9) Dens.
On IV Line	0.029 (0.019)	0.022 (0.026)	0.032 (0.213)						
log(IV Market Access) in 1850				0.015 (0.022)	−0.049 (0.030)	−0.415 (0.638)			
IV Line Year							0.001 (0.005)	0.005 (0.006)	−0.076 (0.077)
Observations	942	941	87	941	940	87	119	119	35
R-squared	0.605	0.583	0.464	0.617	0.611	0.627	0.571	0.368	0.312

Significance levels: *** p<0.01, ** p<0.05, * p<0.1
Notes: Dependent variable in column header. Sample includes counties with no urban population in 1820 that are not origins of straight lines of instrumentation. Sample for regressions of population density restricted to counties that had achieved 1860 boundaries by 1820. All specifications include state fixed effects and cubics in the logarithm of distance from Cincinnati and New York. Specifications with the 1850 market access instrument as a regressor also condition on the 1820 market access instrument. Robust standard errors in parentheses.

In column (1), I regress the logarithm of the wheat suitability measure of a county on an indicator for being on one of the lines presented in Figure 1. This regression includes state fixed effects and the same functions of distance from New York and Cincinnati as included above. The resulting coefficient is

---

and $\hat{\theta} = -3.822$ (s.e. = 0.476). The estimated D-statistic is 1.410 (s.e. = 0.727). The standard error for $\hat{\beta}$ is larger than the one in Table 4 because of the additional uncertainty coming from the need to jointly estimate θ.

[35] The measures of agricultural suitability are not from 1820, but are innate, and so can be considered representative of the conditions in 1820.

statistically insignificant and small, indicating that it is not possible to reject the null hypothesis that counties on the lines were ex ante different from others. The regression in column (2) of corn suitability shows similar results. In both of these cases, even if the coefficients were of larger magnitude and statistically significant, the bias induced by the positive coefficients would tend to mute the negative relationships of the transport-health relationship that I have found. Construction targeting more potentially agriculturally productive areas would tend to be associated with greater average height if agricultural suitability supported better health. The regression in column (3) of the logarithm of population density on the same regressor (limiting the sample to counties that had achieved their 1860 boundaries by 1820) shows similar results.[36]

Columns (4)–(6) repeat the same estimation with the value of the instrument in 1850 (approximately the end of the study period) as the regressor. This is the value of the instrument generated by the "construction" of the hypothetical links. In these regressions I also control for the level of the instrument in 1820 in order to isolate the effects on the instrument of the addition of lines. These regressions yield similar results. Finally, in columns (7)–(9), I regress the same outcomes on the year in which the lines of instrumentation reach a particular county, restricting to counties through which a line passes. Little relationship if any is found. Thus, these results support the identification assumptions that counties on and off of the lines are ex ante similar, and that counties closer and farther from the origins of the line are ex ante similar.

Table 6 presents the coefficient from the estimation of equation (1) by instrumental variables with state-specific indicators and no other controls; it is analogous to column (7) of Table 3. The first feature of note in this column is that the first-stage estimation—that is, the estimation of specification (1) with the logarithm of market access as the dependent variable and the logarithm of the instrumental variables-implied market access as the regressor of interest—shows a positive and strongly statistically significant relationship between the instrument and the potentially endogenous regressor of interest, indicating that the instrument satisfies the relevance condition. This satisfaction of the relevance criterion remains robust throughout the various specifications in this Table.

The relationship between market access and health as estimated by this instrumental variables approach in column (1) is negative and statistically significant.[37] Its magnitude is comparable to the ordinary least squares estimate of Table 3 and to the fixed effects estimates of columns (2)–(5) of Table 4. Column (2) of Table 6 adds the county-specific controls discussed above. Unlike the ordinary least squares regressions of Table 3, the introduction of these controls increases rather than decreases the magnitude of the coefficient,

---

[36] This sample limitation is made in order to avoid changes in population density coming from changing boundaries.

[37] The results of Table 6 include individuals born in counties that have no urban population in 1820 but that are origin points of a line in panels (c) or (d) of Figure 1. Omission of these individuals, who number 303, or 158 in birth years after the decadal year in which the line first appears, yields results that are virtually identical to those of Table 6.

Table 6: Instrumental variables regressions

Variables	(1)	(2)	(3)	(4)	(5)
log(Market Access), 1820 Pop.	-0.541^{***} (0.190)	-0.744^{**} (0.369)	-0.655^{*} (0.368)	-0.832^{*} (0.452)	-0.331 (0.494)
Observations	25,567	23,567	23,567	23,567	25,567
R-squared	0.074	0.077	0.079	0.105	0.055
State FE	Yes	Yes	Yes	Yes	No
Controls	No	Yes	Yes	Yes	No
Birth Year × Region FE	No	No	Yes	No	No
Birth Year × State FE	No	No	No	Yes	No
County FE	No	No	No	No	Yes
First Stage	0.398^{***} (0.033)	0.264^{***} (0.027)	0.269^{***} (0.027)	0.237^{***} (0.027)	0.355^{***} (0.033)

Significance levels: *** p<0.01, ** p<0.05, * p<0.1
Notes: Dependent variable is height in inches. Sample includes individuals born in the Northeast or Midwest in counties with no urban population in 1820. All specifications include birth year and measurement age fixed effects. Standard errors clustered at the county level. Observations weighted to correct for oversampling. The column with the header FE includes both county fixed effects and an instrumentation approach.

which, at −0.744 remains negative and statistically significant, though less precisely estimated. Columns (3) and (4) add region- and state-by-birth year indicators to the instrumental variables specification with controls. The negative and statistically significant coefficient is robust to these controls (though the statistical significance is marginal, with p values of 0.075 and 0.066, respectively), as is its approximate magnitude.

That the instrumental variables estimate is more negative than the analogous ordinary least squares estimate suggests that, in fact, the direction of the bias addressed by the instrumental variables approach is the opposite of the bias hypothesized in section 3 above. This pattern is consistent with transportation being constructed towards areas with agricultural potential, which would have improved health, all else equal. However, this conclusion must be taken carefully given the presence of local average treatment effects and the possibility of measurement errors.

Finally, column (5) combines the two empirical approaches by estimating equation (2) by instrumental variables. The first stage estimate is strong, indicating that prior first-stage estimates are robust to the inclusion of county fixed effects. The second-stage coefficient of interest remains negative, and the magnitude is comparable to estimates of Tables 4 and 6.[38] However, the standard error of this coefficient is more than doubled by the demands of this estimation (relative to the non-instrumental variables analog), making it impossible to reject the null hypothesis of no effect.

Overall, based on the results of Tables 4 and 6, I conclude that the data provide strong and robust

---

[38] The difference between the estimates with and without instrumental variables, though small, tends to support transportation targeting less healthy areas.

evidence of a negative relationship between stature and market access in the county-year of birth.[39] These estimates are consistent with previous descriptions of correlations in the antebellum United States, though unlike those estimates, these can plausibly be interpreted causally.[40]

5.4 Robustness Checks

Table 7 presents a variety of robustness checks of the main results presented above. Columns (1)–(3) verify the robustness of the results of the county-fixed effects regressions of column (2) of Table 4. Column (1) adds the transport indicator into this regression, which already includes market access. This approach, developed by Donaldson and Hornbeck (2016), has the benefit of identifying the impacts of market access while holding constant a county's transportation linkage. Identification is then based on construction elsewhere in the transportation network. Concerns that transportation construction targeted areas that were more or less healthy are thus reduced.[41] Column (1) reveals that the negative and statistically significant coefficient on market access is robust to this alternate source of identification. Column (2) generalizes this approach by controlling separately for railroad, canal, and river linkages, with similar results. Finally, column (3) includes year-specific quadratic functions in latitude and longitude. Although the coefficient is less precisely estimated ($p = 0.159$), it retains its negative sign and approximate magnitude.

Table 7: Robustness checks

Variables	(1) FE	(2) FE	(3) FE	(4) IV	(5) IV
log(Market Access), 1820 Pop.	-0.623^{***} (0.209)	-0.624^{***} (0.203)	-0.302 (0.214)	-1.611^{**} (0.634)	-1.306^{*} (0.757)
Observations	25,567	25,567	25,567	23,567	23,567
R-squared	0.124	0.125	0.136	0.074	0.087
Added Control	Transport Indicator	Transport Mode	Geo. by Yr.	Starting MA	Geo. by Yr.

Significance levels: *** p<0.01, ** p<0.05, * p<0.1
Notes: Dependent variable is height in inches. Sample includes individuals born in the Northeast or Midwest in counties with no urban population in 1820. Standard errors clustered at the county level. Observations weighted to correct for oversampling. Added controls explained in text. Columns headed FE include county fixed effects. Columns headed IV estimated using the straight-line-based instrument and include all controls described in text.

Columns (4) and (5) of Table 7 test robustness of the instrumental variables regression of column (2) of Table 6. Column (4) controls for the level of market access in a county in 1820, in order to more

---

[39] In Table A.1, I use the single year of data on death rates (1850) to study the relationship between transportation linkage as measured by market access and health as measured by death rates. In general, the results of this Table are supportive of the conclusion that there was a negative relationship between transportation and health.

[40] One potential concern is that migration out of the county of birth in response to transportation linkage might be responsible for the reduction in stature, rather than any effect within the county of birth. This concern is addressed in Table A.2, where I show that migration patterns in response to market access are of the opposite sign to be consistent with this concern.

[41] The concerns are not totally alleviated, however, as construction might take place away from a county in order to increase its market access. The Erie Canal is an example of such construction.

effectively isolate changes over time in market access, rather than its level, which may be endogenous even after instrumentation because the instrument is based on the (potentially endogenous) 1820 network. The negative and statistically significant coefficient of market access is robust to this control, and its magnitude is increased. Finally, column (5) includes year-specific quadratics in latitude and longitude, and the result is again robust.

5.5 The Local Development Channel

Understanding the channel through which transportation operates on health is important for at least two reasons. First, understanding the mechanism of the effect would help to understand whether these results apply in other settings, such as in modern developing countries. Second, a negative effect of market access on stature is consistent with both explanations for the Antebellum Puzzle, as discussed in section 2 above. Investigation of the channel through which the effect operates can help to better understand this phenomenon and potentially to distinguish between these two explanations.

Given the limited data availability for this period, it is not possible to evaluate all of the mechanisms described in section 2. Instead, I focus on determining whether there is empirical support for the disease explanation. I concentrate specifically on the part of this explanation that claims that transportation improvements generated a worse epidemiological environment by creating growth in newly linked areas, which generated worse sanitation conditions and thus worse health.[42]

I begin by testing whether the arrival of transportation infrastructure generated local development in the form of greater population density by estimating the specification

$$\log(d_{jt}) = \alpha_j + \gamma_t + \beta \log(\mathrm{MA}_{jt}) + \varepsilon_{jt}$$

for census years 1820–1850, where d_{jt} is population density and MA_{jt} is market access. I estimate this equation by ordinary least squares and by instrumental variables, presenting the results in column (1) of Table 8. Following Atack et al. (2010), I limit the sample to county-years in which counties had already achieved their 1860 boundaries so that results are not driven by, for instance, changes in population density caused by changes in county boundaries. Although the magnitude of the estimated relationship between market access and population density is impacted by whether or not an instrumental variables method is used, the general qualitative result is not. In particular, column (1) of Table 8 shows a large, positive,

---

[42] Unfortunately, a paucity of data prevent an effective test of the other mechanisms.

and statistically significant impact of market access on population density. The estimated coefficients imply that a one-standard deviation increase in market access (0.407 in the data set with counties as the unit of observation, as shown in Table 1) is associated with a 0.181 log point increase in population density according to the fixed effects estimates, or a 0.431 log point increase according to the instrumental variables estimates.

Table 8: The local development mechanism

Variable	(1) Dens.	(2) Dens.	(3) Dens.	(4) Height	(5) Height
Panel A: Fixed Effects					
log(Market Access), 1820 Pop.	0.445^{***}	0.470^{***}	0.436^{***}	-0.392^{**}	-0.392^{**}
	(0.129)	(0.123)	(0.126)	(0.180)	(0.183)
log(MA) × log(Wheat Suit.)		0.649^{***}		-0.844^{*}	
		(0.210)		(0.475)	
log(MA) × log(Corn Suit.)			0.490^{***}		-0.636
			(0.146)		(0.428)
Observations	1,166	1,166	1,166	25,567	25,567
R-squared	0.921	0.926	0.926	0.124	0.124
Panel B: Fixed Effects and IV					
log(Market Access), 1820 Pop.	1.059^{***}	1.038^{***}	1.015^{***}	-0.215	-0.174
	(0.188)	(0.182)	(0.186)	(0.535)	(0.555)
log(MA) × log(Wheat Suit.)		0.677^{**}		-0.400	
		(0.266)		(0.699)	
log(MA) × log(Corn Suit.)			0.442^{*}		-0.423
			(0.232)		(0.674)
Observations	1,122	1,122	1,122	25,567	25,567
R-squared	0.525	0.542	0.543	0.056	0.056

Significance levels: *** p<0.01, ** p<0.05, * p<0.1

Notes: Dependent variable listed in the column header. Sample in columns (1)–(3) includes all county-years with borders fixed to 1860, with no urban population in 1820, and in the Midwest or Northeast. Sample in columns (4) and (5) includes individuals born in the Northeast or Midwest in counties with no urban population in 1820. All specifications include year and county fixed effects. Columns (4) and (5) also include measurement age fixed effects. Observations in columns (1)–(3) weighted by the ratio of county population in each year to total population in that year. Observations in columns (4) and (5) weighted to correct for oversampling. Standard errors clustered at the county level.

In columns (2) and (3) of Table 8, I investigate whether the effect of market access on population density varies by a county's potential agricultural productivity. I interact market access with the log of a measure of a county's average suitability for wheat or corn production. The results of this estimation reveal that the effects of market access on increasing population density were stronger in counties with greater agricultural suitability. I demean the measures of suitability so that, for example, the fixed effects estimates of column (2) can be interpreted as implying that a county with average wheat suitability experienced an increase in population density of 0.470 percent in response to a one percent increase in market access, and that a county with wheat suitability one percent above the mean had a 0.649 percentage point stronger reaction. Similar results are evident for corn suitability, and comparable results (though with less precision and larger coefficients) are found when using instrumental variables. Thus, counties with greater agricultural suitability tended to experience greater increases in population density in response to the same improvements in transport linkages, likely reflecting immigration from other areas of individuals seeking to establish farms.[43]

Given the difference in the unit of observation available for the analysis of population density on the one hand (one observation per county-decade) and transportation and height on the other (annual county-level observations), it is not possible to directly relate changing population density to changing stature. Instead, to determine whether there was a relationship between growth in population density and declines in average stature, I investigate whether the responsiveness of height to changes in market access also differed by crop suitability. To this end, I repeat the interaction approach in estimating equation (2) in columns (4) and (5), with height as the dependent variable. When estimating by ordinary least squares (with county fixed effects) in Panel A, I find that the negative relationship of market access with stature is stronger in the more agriculturally suitable counties, though the interaction coefficients are only marginally statistically significant ($p = 0.076$ for wheat suitability and $p = 0.137$ for corn suitability). The magnitudes of the estimated interaction coefficients are large. When these estimates are repeated with the combination of instrumental variables and county fixed effects, as in column (5) of Table 6, results of the same sign are found, but as with column (5) of Table 6, the coefficients are smaller and imprecisely estimated.

Together, these results indicate that, in the counties where agricultural productivity caused population density to grow more in response to rises in market access, the effects of market access on reducing height were stronger. This is consistent with the contention that the effect of transportation on market access passed through the channel of increasing local development that worsened the local disease environment.

6 The Antebellum Puzzle

In addition to the contribution made by this paper to understanding the health impacts of transportation improvements in developing countries, this paper also helps to understand the causes of the deterioration in health during the early phases of modern economic growth in the United States. Having shown that transportation linkages were responsible for declining health in this period, this paper provides an empirical basis for a potential explanation of this trend. This is the first evidence of an explanation for the Antebellum Puzzle with estimates that can plausibly be given a causal interpretation, and the first that relates declines in height over time to change in local circumstances (by virtue of the county fixed effects estimation). In this section, I use the results presented in section 5 above to determine how much of the deterioration in average stature can be attributed to the growing transportation network of the antebellum period.

---

[43] This view is further supported by the results of Table A.3, which shows an increase in the acreage devoted to farming in response to rising market access.

The empirical pattern to be explained is a 0.82 inch decline in average stature that is present in the benchmark sample.[44] To determine the fraction of this decline that is attributable to rising market access, I determine the estimated impact of the rise in market access over the study period on health as implied by the estimates above. Table 1 shows that the logarithm of market access increased by 0.645 from 1820 to 1850. Thus, the largest coefficient in Tables 4 and 6, -0.832, can explain a decline in stature of 0.54 inches, or about 65 percent of the total decline of 0.82 inches. As a lower bound, the coefficient of -0.323 can explain a decline in stature of 0.21 inches, or about 26 percent of the total decline.

The finding that transportation was responsible for a large fraction of the decline in stature of the antebellum period is important in documenting a definitive cause for the decline in health, which has heretofore eluded researchers. It is not, however, helpful on its own in choosing between the disease and the food price explanations. Discriminating between these explanations is important to understanding the potential impacts of economic growth beyond that induced by transportation expansions, and in determining whether, as Komlos (1987) has argued, that the decline in stature is not indicative of a negative impact of industrialization because it was the product of utility-maximizing choice. In demonstrating that the increased concentration generated by transportation was in part responsible for the negative impact of transportation, this paper also contributes by providing evidence supporting the disease explanation.

7 Conclusion

The beneficial effects of transportation improvements for economic development are well known (Donaldson 2015). Less is known, however, about the impact of transportation on health, where the theoretical impact is ambiguous and only a small number of empirical investigations exist to study this impact in specific settings (Burgess and Donaldson 2012; Tang 2017). In this paper, I study the impacts on health of the internal improvements made in the antebellum United States, focusing specifically on the impacts on the average stature of the birth cohorts of 1820–1847. Besides being important for its own sake, this setting enables the study of the impact of the construction of large infrastructure projects in previously unlinked and undeveloped areas with long-term outcomes. It also provides valuable insights for the possible cause or causes of one of the fundamental and least understood stylized facts in American economic history—the decline in health during early modern growth known as the Antebellum Puzzle. Existing evidence (Haines,

---

[44] This is calculated by regressing heights on birth year and measurement age indicators and then using a local polynomial regression to smooth the estimated coefficients on the birth year indicators. Reassuringly, this decline is comparable to other estimates of the decline in the period, many of which are based on the same or similar data (A'Hearn 1998; Floud et al. 2011; Fogel 1986; Steckel 1995). This includes Zimran's (2018) calculation of a 0.94 inch decline that adjusts for potential representativeness issues in the sample, but which is not limited to counties with no urban population in 1820.

Craig, and Weiss 2003) reveals a negative correlation between transport and health in this period, but does not address the possibility that the correlation may be spurious.

I study this relationship using data on stature from the records of enlistments in the Union Army (Records of the Adjutant General's Office 1861–1865) combined with GIS data on the development of the antebellum transportation network (Atack 2015, 2016, 2017). Using a within-county fixed effects approach and an instrumental variables approach based on straight lines connecting major watersheds to one another and to major cities, I find evidence of a negative effect of market access on average stature. The estimated effect is large enough to explain up to 65 percent of the decline in average stature behind the Antebellum Puzzle. Moreover, I find evidence consistent with the effect operating through a channel in which transportation leads to increased population density, which in turn led to worse health in a period of poor sanitation technology.

These results provide an interesting and important warning. The Antebellum Puzzle cautions that economic growth may not be unambiguously welfare-improving. The results of this paper show that infrastructure improvements with well known economic benefits may also have had an unintended negative impact on at least some aspect of welfare and therefore were also not unambiguously welfare-improving.

References

A'Hearn, Brian (1998). "The Antebellum Puzzle Revisited: A New Look at the Physical Stature of Union Army Recruits during the Civil War." In *The Biological Standard of Living in Comparative Perspective*. John Komlos and Jörg Baten (ed.). Stuttgart: Franz Steiner Verlag, pp. 250–267.

Ali, Rubaba, Alvaro Federico Barra, Claudia N. Berg, Richard Damania, John Nash, and Jason Russ (2015). "Transport Infrastructure and Welfare: An Application to Nigeria." World Bank Policy Research Working Paper 7271.

Atack, Jeremy (2013). "On the Use of Geographic Information Systems in Economic History: The American Transportation Revolution Revisited." *Journal of Economic History* 73:2, pp. 313–338.

——— (2015). *Historical Geographic Information Systems (GIS) Database of Steamboat-Navigated Rivers during the Nineteenth Century in the United States* [machine-readable database].

——— (2016). *Historical Geographic Information Systems (GIS) Database of US Railroads* [machine-readable database].

——— (2017). *Historical Geographic Information Systems (GIS) Database of Nineteenth Century US Canals* [machine-readable database].

Atack, Jeremy, Fred Bateman, Michael R. Haines, and Robert A. Margo (2010). "Did Railroads Induce or Follow Economic Growth? Urbanization and Population Growth in the American Midwest, 1850–1860." *Social Science History* 34:2, pp. 171–197.

Atack, Jeremy and Peter Passell (1994). *A New Economic View of American History from Colonial Times to 1940*. 2nd ed. New York: W W Norton & Co.

Banerjee, Abhijit, Esther Duflo, and Nancy Qian (2012). "On the Road: Access to Transportation Infrastructure and Economic Growth in China." NBER Working Paper 17897.

Banerjee, Rakesh and Ashish Sachdeva (2015). "Pathways to Preventive Health, Evidence from India's Rural Road Program." USC Dornsife Institute for New Economic Thinking Working Paper No. 15-19.

Baum-Snow, Nathaniel, J. Vernon Henderson, Matthew A. Turner, Qinghua Zhang, and Loren Brandt (2018). "Does Investment in National Highways Help or Hurt Hinterland City Growth?" NBER Working Paper 24596.

Bell, Clive and Susanne van Dillen (2018). "On the Way to Good Health? Rural Roads and Morbidity in Upland Orissa." *Journal of Transport and Health* Forthcoming.

Blimpo, Moussa P., Robin Harding, and Leonard Wantchekon (2013). "Public Investment in Rural Infrastructure: Some Political Economy Considerations." *Journal of African Economies* 22, pp. 57–83.

Bodenhorn, Howard, Timothy W. Guinnane, and Thomas A. Mroz (2017). "Sample-Selection Biases and the Industrialization Puzzle." *Journal of Economic History* 77:1, pp. 171–207.

Burgess, Robin and Dave Donaldson (2012). "Railroads and the Demise of Famine in Colonial India." Mimeo., MIT.

Chandra, Amitabh and Eric Thompson (2000). "Does Public Infrastructure Affect Economic Activity? Evidence from the Rural Interstate Highway System." *Regional Science and Urban Economics* 30, pp. 457–490.

Costa, Dora L. (1993). "Height, Wealth, and Disease among the Native-Born in the Rural, Antebellum North." *Social Science History* 17:3, pp. 355–383.

Costa, Dora L. and Richard H. Steckel (1997). "Long-Term Trends in Health, Welfare, and Economic Growth in the United States." In *Health and Welfare During Industrialization*. Richard H. Steckel and Roderick Floud (ed.). Chicago: University of Chicago Press. Chap. 2, pp. 47–90.

Craig, Lee A., Andrew Copland, and Thomas Weiss (2012). *Agricultural Production by County, Quantities and Nutrients, 1840 to 1880* [computer files]. Raleigh: North Carolina State University.

Cuff, Timothy (2005). *The Hidden Cost of Economic Development: The Biological Standard of Living in Antebellum Pennsylvania*. Burlington, Vermont: Ashgate.

Deaton, Angus (2007). "Height, Health, and Development." *Proceedings of the National Academy of Sciences* 104:33, pp. 13232–13237.

Deaton, Angus and Raksha Arora (2009). "Life at the Top: The Benefits of Height." *Economics and Human Biology* 7, pp. 133–136.

Donaldson, Dave (2015). "The Gains from Market Integration." *Annual Review of Economics* 7, pp. 619–647.

Donaldson, Dave and Richard Hornbeck (2016). "Railroads and American Economic Growth: A 'Market Access' Approach." *Quarterly Journal of Economics* 131:2, pp. 799–858.

Duranton, Gilles and Matthew A. Turner (2012). "Urban Growth and Transportation." *Review of Economic Studies* 79:4, pp. 1407–1440.

Emran, M. Shahe and Zhaoyang Hou (2013). "Access to Markets and Rural Poverty: Evidence from Household Consumption in China." *Review of Economics and Statistics* 95:2, pp. 682–697.

Fishlow, Albert (1965). *American Railroads and the Transformation of the Ante-Bellum Economy*. Cambridge: Harvard University Press.

Floud, Roderick, Robert W. Fogel, Bernard Harris, and Sok Chul Hong (2011). *The Changing Body: Health, Nutrition, and Human Development in the Western World since 1700*. New York: Cambridge University Press.

Floud, Roderick, Kenneth W. Wachter, and Anabel S. Gregory (1990). *Height, Health and History: Nutritional Status in the United Kingdom, 1750–1980*. Cambridge: Cambridge University Press.

Fogel, Robert W. (1964). *Railroads and American Economic Growth: Essays in Econometric History*. Baltimore: The Johns Hopkins Press.

——— (1986). "Nutrition and the Decline in Mortality since 1700: Some Preliminary Findings." In *Long-Term Factors in American Economic Growth*. Stanley L. Engerman and Robert E. Gallman (ed.). Chicago: University of Chicago Press. Chap. 9, pp. 439–556.

——— (2004). *The Escape from Hunger and Premature Death, 1700–2100*. New York: Cambridge University Press.

Fogel, Robert W., Dora L. Costa, Michael R. Haines, Chulhee Lee, Louis Nguyen, Clayne Pope, Irvin Rosenberg, Nevin Scrimshaw, James Trussell, Sven Wilson, Larry T. Wimmer, John Kim, Julene Bassett, Joseph Burton, and Noelle Yetter (2000). *Aging of Veterans of the Union Army: Version M-5*. Chicago: Center For Population Economics, University of Chicago Graduate School of Business, Department of Economics, Brigham Young University, and the National Bureau of Economic Research.

Food and Agriculture Organization (2002). *Global Agro-Ecological Zones* [machine-readable database]. New York: United Nations.

Ghani, Ejaz, Arti Grover Goswami, and William R. Kerr (2016). "Highway to Success: The Impact of the Golden Quadrilateral Project for the Location and Performance of Indian Manufacturing." *Economic Journal* 126, pp. 317–357.

Haines, Michael R., Lee A. Craig, and Thomas Weiss (2003). "The Short and the Dead: Nutrition, Mortality, and the 'Antebellum Puzzle' in the United States." *Journal of Economic History* 53:2, pp. 382–413.

Hong, Sok Chul (2007). "The Burden of Early Exposure to Malaria in the United States, 1850–1860: Malnutrition and Immune Disorders." *Journal of Economic History* 67:4, pp. 1001–1035.

Hornbeck, Richard (2010). "Barbed Wire: Property Rights and Agricultural Development." *Quarterly Journal of Economics* 125, pp. 767–810.

Hornung, Erik (2015). "Railroads and Growth in Prussia." *Journal of the European Economic Association* 13:4, pp. 699–736.

Jacoby, Hanan G. (2000). "Access to Markets and the Benefits of Rural Roads." *Economic Journal* 110, pp. 713–737.

Jacoby, Hanan G. and Bart Minten (2009). "On Measuring the Benefits of Lower Transport Costs." *Journal of Development Economics* 89, pp. 28–38.

Jaworski, Taylor and Carl T. Kitchens (2016). "National Policy for Regional Development: Evidence from Appalachian Highways." NBER Working Paper 22073.

Jayachandran, Seema and Rohini Pande (2017). "Why are Indian Children so Short? The Role of Birth Order and Son Preference." *American Economic Review* 107:9, pp. 2600–2629.

Komlos, John (1987). "The Height and Weight of West Point Cadets: Dietary Change in Antebellum America." *Journal of Economic History* 47:4, pp. 897–927.

Komlos, John and Peter Coclanis (1997). "On the Puzzling Cycle in the Biological Standard of Living: The Case of Antebellum Georgia." *Explorations in Economic History* 34, pp. 433–459.

Manson, Steven, Jonathan Schroeder, David Van Riper, and Steven Ruggles (2017). *IPUMS National Historical Geographic Information System: Version 12.0* [Database]. Minneapolis: University of Minnesota.

Niemi, Albert W. (1970). "A Further Look at Interregional Canals and Economic Specialization: 1820–1840." *Explorations in Economic History* 7:4, pp. 499–520.

Ransom, Roger L. (1967). "Interregional Canals and Economic Specialization in the Antebellum United States." *Explorations in Economic History* 5, pp. 12–35.

——— (1970). "Social Returns from Public Transport Investment: A Case Study of the Ohio Canal." *Journal of Political Economy* 78:5, pp. 1041–1060.

——— (1971). "A Closer Look at Canals and Western Manufacturing." *Explorations in Economic History* 8, pp. 501–508.

Records of the Adjutant General's Office (1861–1865). Regimental records, including descriptive rolls, order, and letter books, and morning reports, of volunteer organizations, Civil War, 1861–65. Records relating to volunteers and volunteer organizations. Record Group 94.2.4. Washington, DC: National Archives Building.

Ruhm, Christopher J. (2000). "Are Recessions Good for Your Health?" *Quarterly Journal of Economics* 115:2, pp. 617–650.

Segal, Harvey H. (1961). "Canals and Economic Development." In *Canals and American Economic Development.* Carter Goodrich (ed.). New York: Columbia University Press, pp. 216–248.

Steckel, Richard H. (1995). "Stature and the Standard of Living." *Journal of Economic Literature* 33:4, pp. 1903–1940.

Stifel, David and Bart Minten (2015). "Market Access, Welfare, and Nutrition: Evidence from Ethiopia." Ethiopia Strategy Support Program Working Paper 77.

Storeygard, Adam (2016). "Farther on down the Road: Transport Costs, Trade and Urban Growth in Sub-Saharan Africa." *Review of Economic Studies* 83, pp. 1263–1295.

Sunder, Marco (2011). "Upward and Onward: High-Society American Women Eluded the Antebellum Puzzle." *Economics and Human Biology* 9, pp. 165–171.

Sunder, Marco and Ulrich Woitek (2005). "Boom, Bust, and the Human Body: Further Evidence on the Relationship Between Height and Business Cycles." *Economics and Human Biology* 3, pp. 450–466.

Tang, John P. (2014). "Railroad Expansion and Industrialization: Evidence from Meiji Japan." *Journal of Economic History* 74:3, pp. 863–886.

——— (2017). "The Engine and the Reaper: Industrialization and Mortality in Late Nineteenth Century Japan." *Journal of Health Economics* 56, pp. 145–162.

Taylor, George Rogers (1951). *The Transportation Revolution: 1815–1860.* Vol. IV. The Economic History of the United States. New York: Reinhart & Company, Inc.

Trivedi, Anjani (2017). "Why Chinese Men Are Dying Despite Rising Income." *Wall Street Journal* February 25, 2017, B.10.

Woitek, Ulrich (2003). "Height Cycles in the 18th and 19th Centuries." *Economics and Human Biology* 1, pp. 243–257.

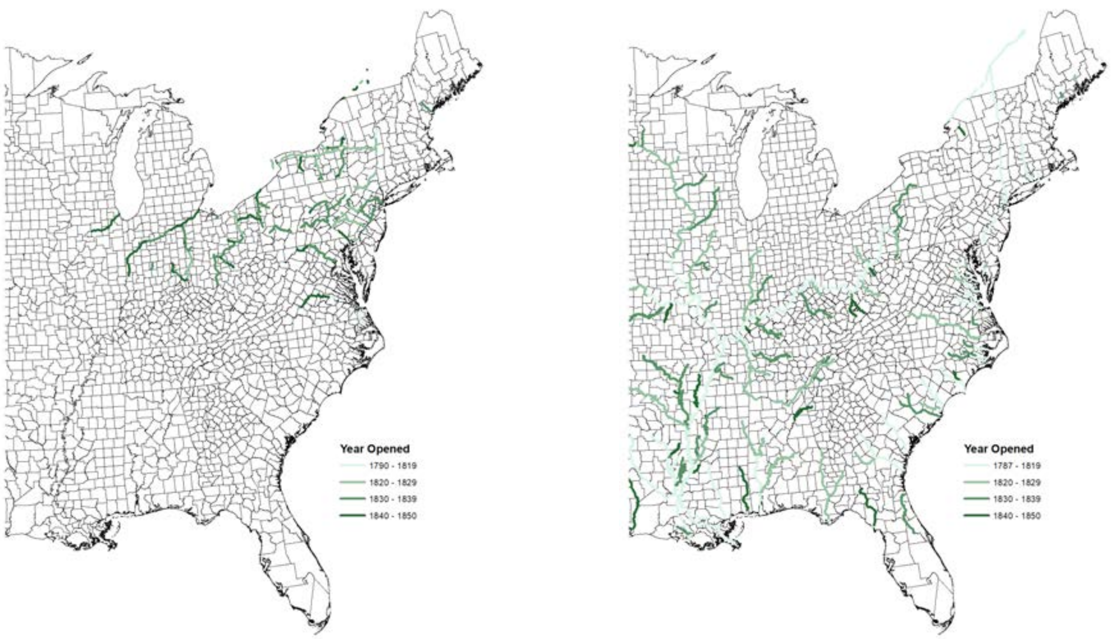
World Bank (2007). *Evaluation of World Bank Support to Transportation Infrastructure*. Washington: World Bank.

Yoo, Dongwoo (2012). "Height and Death in the Antebellum United States: A View through the Lens of Geographically Weighted Regression." *Economics and Human Biology* 10, pp. 43–53.

Zehetmayer, Matthias (2011). "The Continuation of the Antebellum Puzzle: Stature in the US, 1847–1894." *European Review of Economic History* 15, pp. 313–327.

Zimran, Ariell (2018). "Sample-Selection Bias and Height Trends in the Nineteenth-Century United States." *Journal of Economic History* Forthcoming.

A Additional Tables and Figures



(a) Canals

(b) Steamboat-Navigable Rivers

(c) Railroads

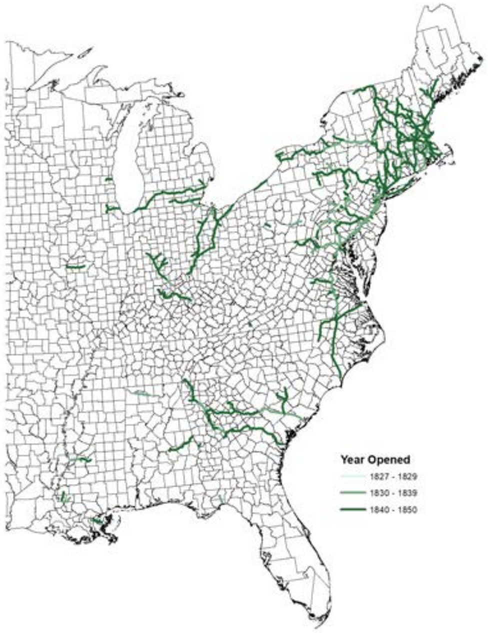


Figure A.1: Spread of transportation infrastructure

Note: Always-navigable rivers, lakes, and oceans are assigned an opening date of 1787.

Source: Atack (2015, 2016, 2017)

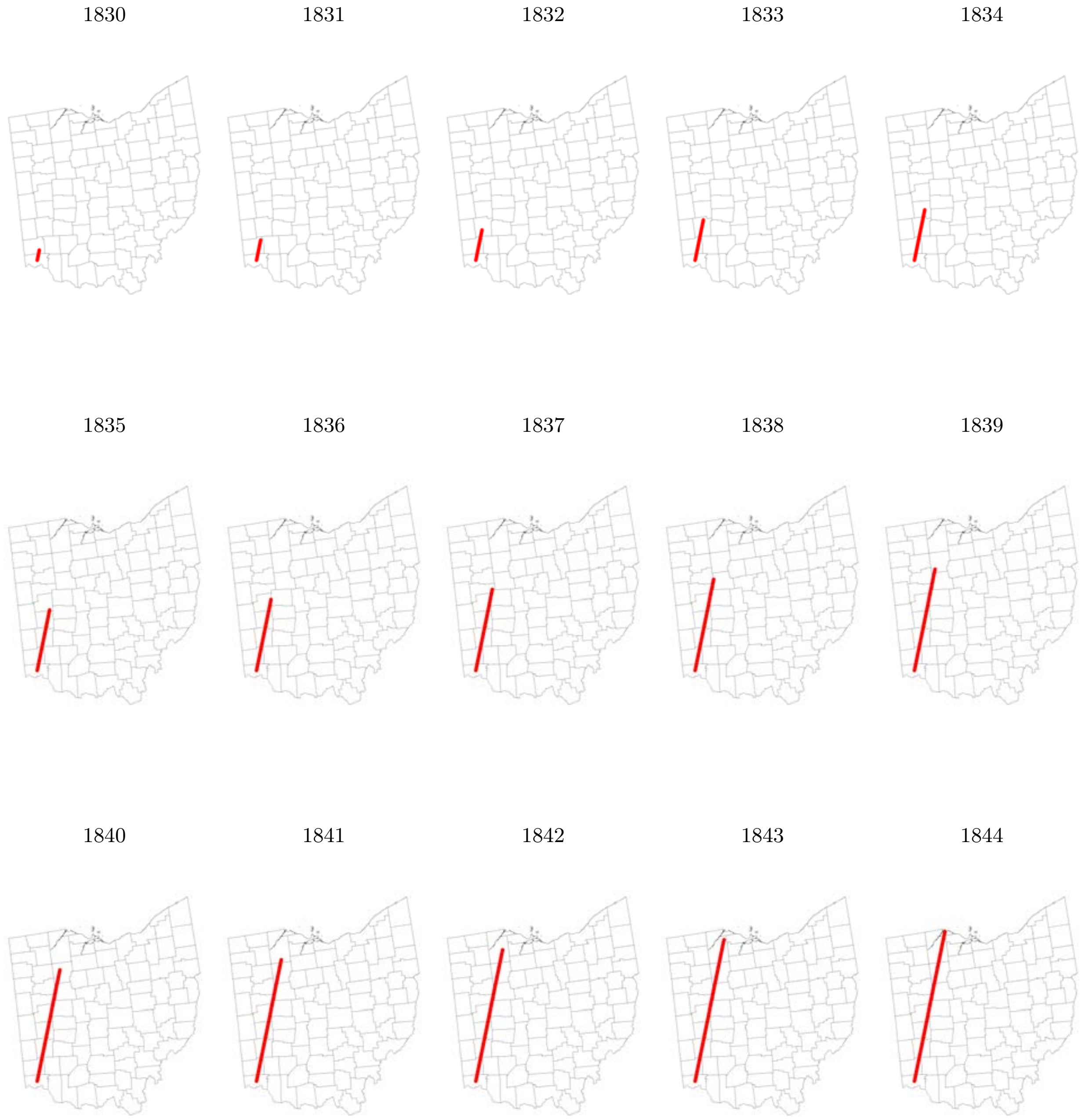


Figure A.2: Evolution of the instrument line linking Hamilton County, Ohio to the Great Lakes

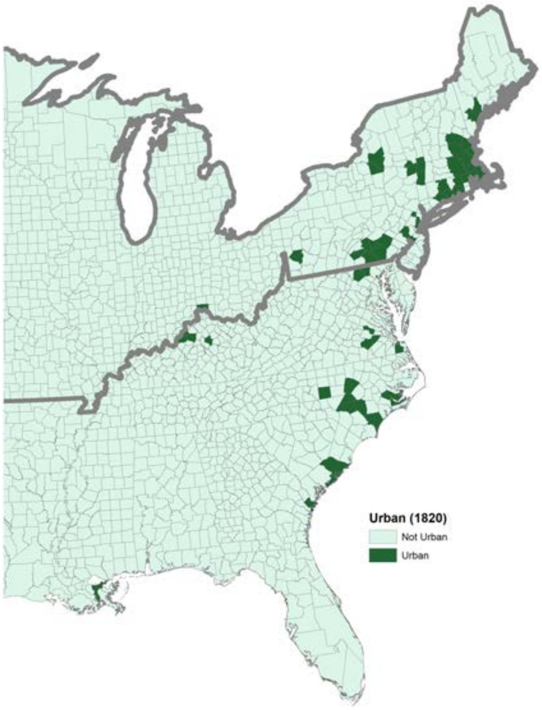


Figure A.3: Urbanization in 1820

Note: Urban counties are defined as those with any urban population in 1820. Sample region indicated by thick boundary.

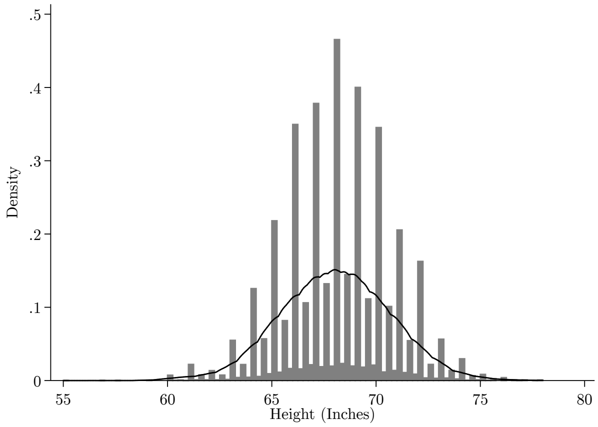
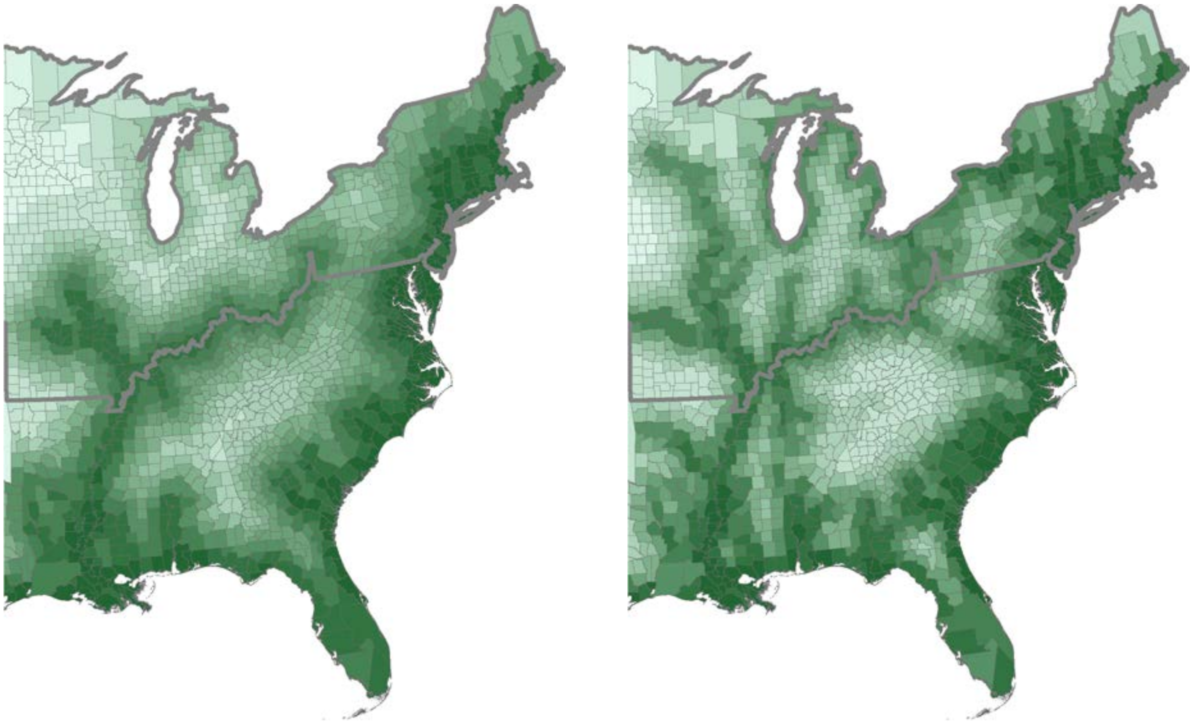


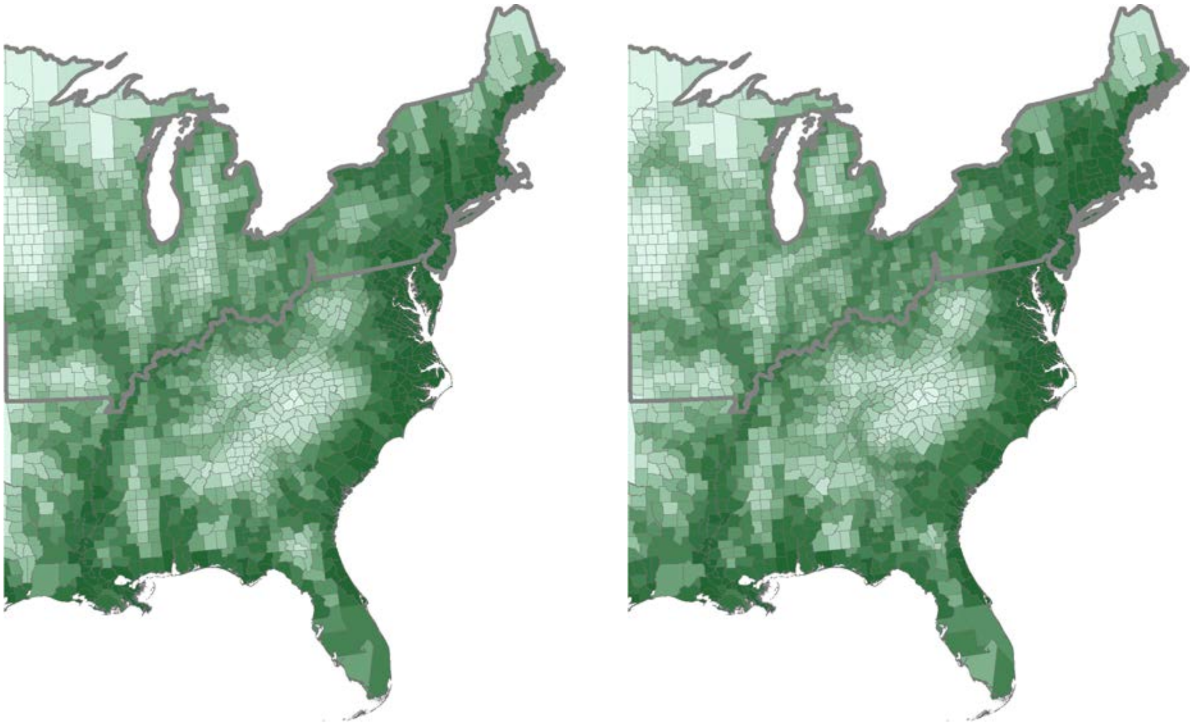
Figure A.4: Distribution of heights in the original data

Note: Sample includes individuals born in the Northeast or the Midwest in counties with no urban population in 1820. The histogram is divided into quarter-inch bins.



(a) 1820

(b) 1830



(c) 1840

(d) 1850

Figure A.5: Market access by decade.

Note: Each panel divides counties into deciles of market access for that year, with darker counties having greater market access. The scales are not comparable across years. Sample region indicated by thick boundary.

Figure A.6: Number of individual height observations by birth cohort

Note: Sample includes individuals born in the Northeast or Midwest in counties with no urban population in 1820 for whom height and county of birth are known.

Table A.1: Regressions of 1850 death rates

Variables	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) IV	(6) IV	(7) IV	(8) IV
Panel A: Weighted by 1850 population								
log(Market Access), 1820 Pop.	0.657***	0.555***	0.174	0.118	1.078***	1.234***	0.571**	0.479
	(0.060)	(0.070)	(0.109)	(0.114)	(0.122)	(0.231)	(0.274)	(0.385)
Initial Market Access		0.083***		0.080		−0.062		0.032
		(0.029)		(0.050)		(0.056)		(0.069)
Observations	790	790	472	472	790	790	472	472
R-squared	0.239	0.247	0.466	0.469	0.190	0.154	0.449	0.457
Panel B: Unweighted								
log(Market Access), 1820 Pop.	0.407***	0.331***	0.397***	0.365***	0.876***	1.219**	0.599***	0.607**
	(0.055)	(0.060)	(0.099)	(0.106)	(0.156)	(0.585)	(0.197)	(0.299)
Initial Market Access		0.113***		0.042		−0.105		−0.003
		(0.036)		(0.054)		(0.148)		(0.074)
Observations	790	790	472	472	790	790	472	472
R-squared	0.300	0.309	0.335	0.336	0.234	0.111	0.329	0.328
Controls	No	No	Yes	Yes	No	No	Yes	Yes

Significance levels: *** p<0.01, ** p<0.05, * p<0.1

Notes: Dependent variable is log of deaths per 1850 population. Sample includes counties in the Northeast or Midwest with no urban population in 1820. All specifications include state fixed effects. Robust standard errors in parentheses.

Table A.2: Migration regressions

Variables	(1) FE	(2) FE	(3) FE	(4) FE	(5) IV	(6) IV	(7) IV	(8) IV	(9) IV/FE
Panel A: State-level migration indicator									
log(Market Access), 1820 Pop.	−0.025	0.105*	−0.018	0.005	−0.067	−0.067	−0.059	−0.028	−0.173*
	(0.039)	(0.056)	(0.038)	(0.045)	(0.065)	(0.128)	(0.129)	(0.151)	(0.094)
Observations	25,567	25,567	25,567	25,567	25,567	23,567	23,567	23,567	25,567
R-squared	0.234	0.286	0.236	0.259	0.051	0.071	0.073	0.107	0.012
Panel B: County-level migration indicator									
log(Market Access), 1820 Pop.	−0.078**	0.005	−0.080**	−0.024	0.033	−0.105	−0.089	−0.014	−0.216**
	(0.039)	(0.055)	(0.039)	(0.042)	(0.101)	(0.181)	(0.184)	(0.218)	(0.093)
Observations	21,095	21,095	21,095	21,095	21,095	19,429	19,429	19,429	21,095
R-squared	0.349	0.381	0.350	0.373	0.082	0.118	0.120	0.153	0.017
Panel C: Migration to denser county indicator									
log(Market Access), 1820 Pop.	−0.103***	−0.037	−0.111***	−0.087**	−0.103	0.107	0.120	0.166	−0.136
	(0.040)	(0.048)	(0.038)	(0.038)	(0.074)	(0.126)	(0.126)	(0.149)	(0.085)
Observations	21,095	21,095	21,095	21,095	21,095	19,429	19,429	19,429	21,095
R-squared	0.306	0.357	0.308	0.330	0.065	0.130	0.132	0.165	0.019
Panel D: log(Population density) difference of enlistment and birth county									
log(Market Access), 1820 Pop.	−0.247**	−0.057	−0.260**	−0.244**	−0.153	0.110	0.141	0.176	−0.105
	(0.123)	(0.128)	(0.113)	(0.113)	(0.146)	(0.209)	(0.207)	(0.241)	(0.220)
Observations	21,088	21,088	21,088	21,088	21,088	19,425	19,425	19,425	21,088
R-squared	0.228	0.297	0.231	0.264	0.040	0.102	0.105	0.144	0.011
State FE	No	No	No	No	Yes	Yes	Yes	Yes	No
Controls	No	No	No	No	No	Yes	Yes	Yes	No
Birth Year × Region FE	No	No	Yes	No	No	No	Yes	No	No
Birth Year × State FE	No	No	No	Yes	No	No	No	Yes	No
County × Decade FE	No	Yes	No	No	No	No	No	No	No

Significance levels: *** p<0.01, ** p<0.05, * p<0.1

Notes: Dependent variable indicated in the panel header. Sample includes individuals born in the Northeast or Midwest in counties with no urban population in 1820. All specifications include birth year and measurement age fixed effects. Standard errors clustered at the county level. Observations weighted to correct for oversampling. Columns with the header FE include county fixed effects, except those with county-decade-specific fixed effects. Columns with the header IV use an instrumental variables approach to address endogeneity. The column with the header IV/FE includes both county fixed effects and an instrumentation approach.

Table A.3: Regressions of improved acreage

Variable	(1) Acreage	(2) Acreage	(3) Acreage
Panel A: Fixed Effects			
log(Market Access), 1820 Pop.	0.247^{***}	0.122^{**}	0.092
	(0.064)	(0.058)	(0.063)
log(MA) × log(Wheat Suit.)		0.875^{***}	
		(0.158)	
log(MA) × log(Corn Suit.)			0.763^{***}
			(0.133)
Observations	1,427	1,427	1,427
R-squared	0.955	0.961	0.962
Panel B: Fixed Effects and IV			
log(Market Access), 1820 Pop.	4.029	2.657^{*}	5.302
	(2.703)	(1.388)	(4.859)
log(MA) × log(Wheat Suit.)		1.856	
		(1.582)	
log(MA) × log(Corn Suit.)			−1.114
			(3.033)
Observations	1,427	1,427	1,427
R-squared	−7.881	−3.639	−13.552

Significance levels: *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Notes: Dependent variable is the ratio of improved acres to total acreage. Sample includes all county-years with borders fixed to 1860, with no urban population in 1820, and in the Midwest or Northeast. All specifications include year and county fixed effects. Observations weighted by the ratio of county population in each year to total population in that year. Standard errors clustered at the county level.

B Effects of Market Access Over the Life Cycle

In this appendix, I determine whether there are impacts of market access or transportation linkage in years other than the birth year on terminal height. This relaxes the restriction in the main text that the impact of transportation linkage on health comes from the state of the transportation network in the birth year.

One strategy is to simply repeat the analysis of the main text, but to include the measures of transportation linkage in years other than the year of birth in the regressions. However, the resultant loss of power from this approach would be too severe to yield any meaningful results. I therefore use the more restrictive specifications

$$h_{ijt} = \gamma_t + \delta_a + \mathbf{z}_j'\tau + \beta_1 \bar{X}_{jt}^{[-6,-2]} + \beta_2 \bar{X}_{jt}^{[-1,3]} + \beta_3 \bar{X}_{jt}^{[4,11]} + \beta_4 \bar{X}_{jt}^{[12,18]} + \varepsilon_{ijt} \tag{B.4}$$

and

$$h_{ijt} = \alpha_j + \gamma_t + \delta_a + \beta_1 \bar{X}_{jt}^{[-6,-2]} + \beta_2 \bar{X}_{jt}^{[-1,3]} + \beta_3 \bar{X}_{jt}^{[4,11]} + \beta_4 \bar{X}_{jt}^{[12,18]} + \varepsilon_{ijt}, \tag{B.5}$$

where $\bar{X}_{jt}^{[a,b]}$ is the average of the logarithm of market access in county j over ages a to b of the cohort born in year t,[45] and the other notation is as in the main text. These are analogs of equations (1) and (2), respectively, with the substitution of the four explanatory variables of interest for the one. The four divisions in equations (B.4) and (B.5) are intended to denote the period before conception, the *in utero* and infancy period, childhood, and adolescence, respectively.

Results of estimation of equations (B.4) and (B.5) are presented in Figure B.1. The results of four specifications are presented—with and without county fixed effects and with and without region-specific birth year fixed effects. The results of all four specifications are similar. In all four cases, estimates of β_1, β_3, and β_4—that is, of the coefficients on market access in years other than infancy and the *in utero* period—are statistically insignificant and in most cases effectively zero.[46] Only β_2—the coefficient on the average of market access for ages -1–3—is consistently of one sign and nearly statistically significant. This result suggests that the previous analyses' focus on market access in the year of birth did not overlook important effects.

The absence of an effect of market access on height outside of the years surrounding the birth year need not indicate that this is the only point in the life cycle in which there is an effect, as I interpret it above. Instead, it could be the case that migration in later life could generate measurement error for the later-in-life measures of market access.

An alternative approach is to relate terminal stature to the number of years to which an individual was linked to the transportation network. Table B.1 presents estimates of the equation

$$h_{ijt} = \gamma_t + \delta_a + \beta E_{jt} + \varepsilon_{ijt}, \tag{B.6}$$

where E_{jt} is the exposure time to transportation of cohort t born in county j and all other notation is as above. Exposure time is computed as follows: for those who are born into a county that is already transportation-linked, the exposure time is set to 23 years; for individuals for whom transportation arrives at age a, I set $E_{jt} = \max\{0, 23 - a\}$. I estimate equation (B.6) by OLS in columns (1)–(4) and with county fixed effects in columns (5)–(7). In no case is a statistically significant relationship between stature and exposure time present, and in all cases the estimates are small.

---

[45] I compute market access for years 1810–1860. This implies that for the later cohorts, $\bar{X}_{jt}^{[12,18]}$ may be the average of a shorter span of ages. For instance, for the 1847 cohort, it is only the average for ages 12 and 13. The lower frequency of rail information in the 1850s also reduces its accuracy.

[46] The fixed effects estimates of β_4 are positive, but they are not statistically significant and are inconsistent with the OLS estimates.

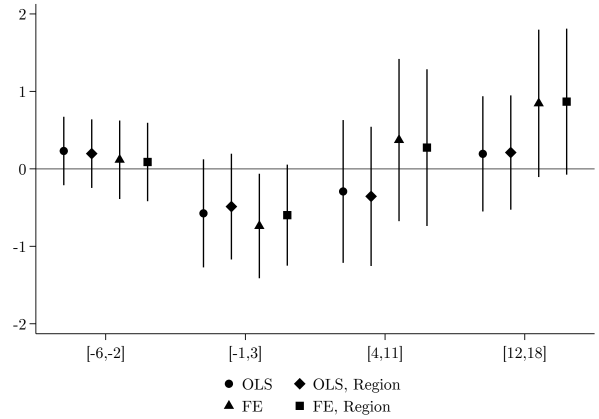


Figure B.1: Market access and stature over the life cycle

Note: This Figure presents coefficients and 95% confidence intervals from estimates of equations (B.4) and (B.5). The x-axis indicates the range of ages over which market access is averaged. The estimates marked "OLS" and "OLS, Region" are estimates of equation (B.4) without and with region-specific measurement age and birth year fixed effects, respectively. Estimates marked "FE" and "FE, Region" are estimates of equation (B.5) without and with region-specific measurement age and birth year fixed effects, respectively.

Table B.1: Exposure time regressions

Variables	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) OLS	(6) FE	(7) FE	(8) FE	(9) FE
Exposure Time	−0.008 (0.006)	0.004 (0.006)	0.007 (0.007)	0.006 (0.007)	0.003 (0.007)	0.018 (0.024)	−0.022 (0.040)	0.022 (0.025)	−0.006 (0.027)
Observations	13,807	13,807	12,722	12,722	12,722	13,807	13,807	13,807	13,807
R-squared	0.043	0.060	0.066	0.072	0.134	0.135	0.191	0.141	0.199
State FE	No	Yes	Yes	Yes	Yes	No	No	No	No
Controls	No	No	Yes	Yes	Yes	No	No	No	No
Birth Year × Region FE	No	No	No	Yes	No	No	No	Yes	No
Birth Year × State FE	No	No	No	No	Yes	No	No	No	Yes
County × Decade FE	No	No	No	No	No	No	Yes	No	No

Significance levels: *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Notes: Dependent variable is height in inches. Sample includes individuals born in the Northeast or the Midwest in counties with no urban population in 1820. All specifications include birth year and measurement age fixed effects. Standard errors clustered at the county level. Observations weighted to correct for oversampling. Controls include the logs of the following variables: 1820 population; area; 1840 population, cattle, pigs, calories and protein; Herfindahl indices for protein and calorie production; 1840 employment by sector and values of agricultural and manufacturing output; 1850 population and values of farms and capital in manufacturing; and distance from New York and Cincinnati. Columns titled FE include either county fixed effects or county-decade-specific fixed effects, as indicated in the column.

C Market Access Computation Algorithm

Procedure C.1. The procedure for computation of transportation costs in a particular year t is as follows. This approach is based on that of Donaldson and Hornbeck (2016).

1. A map of US counties with 1860 boundaries and of all transportation infrastructure present in year t were loaded. The transportation infrastructure includes Donaldson and Hornbeck's (2016) maps of seas, lakes, and the intercoastal waterway linking the Atlantic and Pacific, revised to accurately reflect the state of linkages between the various Great Lakes in the antebellum period.

2. For each mode, linkages are made between county centroids and the several nearest forms of transportation of each mode. In addition, direct linkages between county centroids within 300 kilometers are made. For linkages of county centroids to modes in another county, distance is taken as geographic distance. For linkages between a county centroid and modes within the county, linkages are given the distance of an average of the distances of 200 randomly selected points within the county to that mode, as in Donaldson and Hornbeck (2016).

3. Transshipment links are created between modes of transportation.

4. Transportation rates are assigned using Taylor's (1951) rates, as reported in Atack and Passell (1994). These are as reported in Table C.1. Transshipment is assigned a cost of 50 cents per ton per transshipment.

5. An origin-destination cost matrix calculation is performed, computing transport costs c_{ijt} between each county pair.

6. Following Donaldson and Hornbeck (2016), I compute the iceberg cost as

$$\tau_{ijt} = 1 + \frac{c_{ijt}}{35}.^{47}$$

Table C.1: Transportation costs

Mode	Cost (cents per ton mile)
New York Canals	0.99
Ohio, Indiana, and Illinois Canals	1.60
Other Canals	2.40
Mississippi and Ohio Rivers	0.37
Other Rivers	1.20
Great Lakes	0.10
Oceans	0.049
Railroads	1.95
Wagon Haul	21.00

Notes: Rates per ton mile are taken from Taylor (1951), as reported by Atack and Passell (1994). Transshipment is assigned a cost of 50 cents per ton per transshipment.

---

[47] Given that the trans-shipment cost and the value of goods are taken from Fogel (1964), they may not "match" with the freight rates taken from Taylor (1951), as the former are for the postbellum period and the latter are for the antebellum period. However, given that transportation linkage is the single strongest factor affecting market access, the impacts of changing these figures is likely to be second-order relative to the impact of a changing transport linkage.