NBER WORKING PAPER SERIES

TRANSPORTATION AND HEALTH IN THE ANTEBELLUM 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, Revised November 2019

Previously circulated as "Transportation and Health in a Developing Country: The United States, 1820–1847." 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 the Antebellum 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, 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 0.14 inches, and rising market access over the study period can explain 37 percent of the contemporaneous decline in average stature, known as the Antebellum Puzzle. 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

A online appendix is available at http://www.nber.org/data-appendix/w24943

1 Introduction

Economic histories of the US identify a period of rapid economic growth beginning in the 1820s and lasting to the outbreak of the Civil War (e.g., David 1967; Davis 2004), during which per-capita income, real wages, and most other economic indicators of the standard of living increased considerably (e.g., Costa and Steckel 1997; Goldin and Margo 1992; Margo 2000). Despite this progress, the US entered the Civil War having taken a step backwards in another important facet of the standard of living—health. The best known manifestation of this pattern was the decline in average stature observed in the native-born Northern white male birth cohorts of the 1830s and 1840s (A'Hearn 1998; Floud et al. 2011; Fogel 1986; Komlos 1987; Margo and Steckel 1983; Zimran 2019).[1] Known as the *Antebellum Puzzle* because rising incomes would typically be expected to improve health (Fogel 1994; McKeown 1976), this discrepancy between trends in economic and biological indicators of living standards has complicated the interpretation of the welfare effects of early modern economic growth in the US. Despite the importance of this pattern to US economic history, questions remain regarding its causes.

The economic growth of the antebellum period was caused in part by a contemporaneous *Transportation Revolution* that transformed the American continent and economy by expanding transportation infrastructure into large areas that were previously unlinked and undeveloped (Taylor 1951). Although economic theory shows that the impact of transportation on health may be positive or negative, it has long been suspected that the antebellum transportation expansion may also have contributed to the contemporaneous decline in health (Cuff 2005; Haines, Craig, and Weiss 2003; McGuire and Coelho 2011; Yoo 2012). But the effect of transportation on health in the antebellum US remains unclear because of data limitations (described by Atack 2013) and methodological constraints faced by prior research.

In this paper, I exploit recent data and methodological advances to determine the effect of transportation improvements on average stature in the rural US in the period 1820–1847. This period encompasses the decline in average stature that constitutes the Antebellum Puzzle (Craig 2016; Zehetmayer 2011). It also captures infrastructure expansion on a massive scale into previously

[1] There is also evidence of a decline in life expectancy during this period (Fogel 1986; Pope 1992).

isolated areas, consisting largely of construction of the canal network in the Northeast and Midwest and of 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. I focus on the rural population because this is the sector in which transportation is hypothesized to have contributed to declining health, and because this was the sector in which the bulk of the decline in stature occurred (Steckel 1995).[2]

My analysis is based on two main data sources that together enable me to link individuals' adult height—a common measure of health in historical settings that reflects net nutritional status in childhood and adolescence (Floud et al. 2011; Steckel 1995)—to the transportation linkage status of their county-year of birth.[3] In particular, I use height data from the records of enlisters in the Union Army (Records of the Adjutant General's Office 1861–1865). This source provides information on the heights, counties of birth, and birth cohorts of 25,566 native-born rural white men in the birth cohorts of 1820–1847 in the Northeast and Midwest regions of the US. To describe the development of the transportation network in the antebellum US, I use GIS shape files created by Atack (2015, 2016, 2017), which provide the location and opening date of all canals, railroads, and steamboat-navigable waterways in the antebellum US. This source marks an improvement over the transportation data available in previous investigations of the link between transport and health in the antebellum US (Haines, Craig, and Weiss 2003; Yoo 2012), for which only approximations of the state of the transportation network in a single year were possible (Atack 2013). I use these data to compute Donaldson and Hornbeck's (2016) market access statistic for all counties east of the Mississippi River for each year 1820-1847, which is my main measure of transportation linkage.

The main empirical challenge of this paper is to determine the impact of transportation improvements on height 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. Two aspects of my analysis, made possible by the new data and

[2]Steckel (1995, p. 1927) points out that "the search for understanding should recognize that most of the mid-nineteenth century height decline occurred within the rural population."

[3]I focus on the birth year because there is evidence that conditions in early life have the strongest impact on adult stature (e.g., Steckel 1995; Woitek 2003). I explore the effects at different ages in Online Appendix B.

methods that I use to measure transportation linkage, enable me to address these concerns to an extent not possible in prior studies of the link between transportation and health in the antebellum US (Cuff 2005; Haines, Craig, and Weiss 2003; Yoo 2012). First, the panel structure of the data on a county's transport linkage enable me to estimate specifications that include county fixed effects, capturing location-specific characteristics that might create a spurious relationship between transportation and height. Second, my use of the market access statistic to measure transportation linkage makes it possible to control directly for potentially endogenous transportation linkage and to identify the effects of market access on health using variation generated by construction elsewhere in the network, which is less likely to be related to confounding factors that might relate construction in a county to its health (Donaldson and Hornbeck 2016; Hornbeck and Rotemberg 2019).

I find a negative effect of transportation linkage, as measured by market access in the county-year of birth, on health, as measured by adult stature. The magnitude of this relationship is large: in my preferred specification, I find that a one-standard deviation increase in market access reduced average stature by 0.14 inches. To put this figure in perspective, average stature in my data declined by 0.82 inches from 1820 to 1847, and Zimran (2019) estimates that urbanites during this period suffered a height penalty of 0.29 inches relative to ruralists.

I also investigate one possible mechanism for the negative effect of transportation on average stature, which attributes the deleterious impact of transportation to the increases in population density that transportation generates. I refer to this as the *local development mechanism*. In combination with insufficient sanitation and public health infrastructure in this period, such concentration of population would have made previously undeveloped locations less healthy (Costa 1993; Floud et al. 2011; Steckel 1995). Consistent with this mechanism, I find a positive impact of market access on population density. Heterogeneity in the impact of market access on population density and stature provides further support for this mechanism. In particular, I 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 (Food and Agriculture Organization 2002). This result is consistent with the importance of these goods in long-distance trade (Atack and Passell 1994, pp. 292–294): counties that were better able to produce these grains would have benefited more from improved linkage to markets, leading to greater population growth in these counties.[4] I also find that the negative impact of market access on stature was stronger in counties with greater wheat and corn suitability. Thus, counties where population density increased the most in response to rising market access were those where the deleterious effect of market access on average stature was the greatest. This is consistent with rising population density contributing to the decline in stature.

This paper contributes to the literature seeking to understand the causes of the deterioration of average stature in the antebellum US (e.g., Costa 1993; Haines, Craig, and Weiss 2003; Komlos 1998; Steckel 1995). Several theories have been proposed to explain this decline (Floud et al. 2011), and much evidence consistent with these theories has been marshaled. But well identified causal evidence that any particular factor caused declines in height in the antebellum US is lacking. In this paper, I provide 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 37 percent of the decline in stature.[5] Moreover, by providing evidence consistent with the local development mechanism, which attributes the decline in health to epidemiological factors, the paper contributes to the debate over whether the changing epidemiological environment was a contributor to the antebellum decline in health (e.g., Costa 1993; Komlos 2012).

This paper also brings the impacts of the antebellum period's transportation improvements into sharper focus, adding to the literature on the impacts of transportation improvements in the nineteenth-century US. Recent data improvements have enabled scholars to better determine the causal effects of transportation improvements in this context, but the focus has typically been on the impact of railroads on economic development (e.g., Atack et al. 2010; Donaldson and Hornbeck 2016; Hornbeck and Rotemberg 2019). This paper broadens the scope of analysis of such work to health impacts, and also adds to the literature studying the effects of canal construction and expansions in river navigability,[6] which were the main transportation improvements in the period

[4]This population growth might come from a Malthusian response to rising income (Ahmad 2019) or from migration to these areas.

[5]Studies commonly find a negative cross-sectional relationship between height and some factor, and then attribute falling heights to an increase in that factor. The panel data in this paper enable me to find factors directly associated with within-county height declines, which I refer to as "direct evidence."

[6]The main studies of the effects of canals are Niemi (1970), Ransom (1967, 1970, 1971), and Segal (1961).

that I study.

Outside the nineteenth-century US, although there is a large literature describing the impacts of transportation improvements on a variety of economic outcomes,[7] such as urbanization and industrialization, the effects on health have received far less empirical attention and are not understood as well. To my knowledge, only a few studies (Ahmad 2019; Burgess and Donaldson 2012; Tang 2017) exist determining the causal effect of transportation on health in specific cases.[8] 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 expansion. It shows, with attention to causality, that despite the well known economic benefits of this expansion, it had a negative impact on health.

2 Transportation and Health in the Antebellum US

2.1 The Antebellum Puzzle

Modern economic growth in the US began around 1820 (David 1967). This was the beginning of a long process that eventually brought the US to world economic leadership and led to a drastic long-run improvement in living standards, including health (Fogel 1994). In the shorter run, however, the welfare effects of modern economic growth are less clear. Whereas income per capita and wages rose in the first few decades of US modern economic growth (Costa and Steckel 1997; Goldin and Margo 1992; Margo 2000), health declined. The average height of native-born Northern white males in the US 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 2019). It was not

[7]Specific case studies of the impacts of transportation improvements on a variety of economic outcomes are given by Atack et al. (2010), Baum-Snow et al. (2018), Bogart et al. (2019), Donaldson (2018), 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 (2019), Storeygard (2016), and Tang (2014), among others.

[8]Indeed, Bogart (2018) discusses only a single economic history paper on the effects of transportation on health (Tang 2017). There exist a number of studies that are suggestive of the effects of transport on health (e.g., Ali et al. 2015; Banerjee and Sachdeva 2015; Bell and van Dillen 2018; Blimpo, Harding, and Wantchekon 2013; Stifel and Minten 2015). But these studies 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. They are also generally constrained to study only short-term effects. More research is therefore needed to understand the impact of transportation on health outcomes.

until nearly the birth cohort of 1900 that average stature would begin to rise again (Fogel 1986; Steckel 1995; Zehetmayer 2011).[9] Moreover, between the first and second quarters of the nineteenth century, life expectancy at age 10 for males declined by about 3 years (Fogel 1986).

Despite a large body of research devoted to describing and debating the decline in health in the antebellum US, a definitive explanation for it has not been identified. Recent scholarship favors a combination of two mechanisms (Floud et al. 2011).[10] The first, 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*, argues 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 1987; Komlos and Coclanis 1997). Distinguishing between these explanations is important in determining the welfare effects of early modern economic growth. Under the disease explanation, the decline in health was the product of a negative externality of economic growth, the welfare effects of which were thus ambiguous; but under the food price explanation, it was the product of utility-maximizing choice, implying that welfare improved unambiguously during this period.

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 role for transportation. Under the disease explanation, transportation may have carried disease along with freight and passengers, bringing infection to places that it had once been unable to reach (e.g., Tang 2017). Indeed, disease was prevalent in urban areas in the antebellum US (Floud et al. 2011), and transportation linkages in rural areas may have facilitated the spread of disease from urban areas to previously salubrious rural ones. Transportation may also have generated increases in local development (e.g., Atack et al. 2010; Donaldson and Hornbeck 2016). These

[9] Bodenhorn, Guinnane, and Mroz (2017) have argued that the observed decline in stature may have been an artifact of sample-selection bias rather than of true trends in height, sparking a debate over true trends in population stature (e.g., Bodenhorn, Guinnane, and Mroz 2019; Komlos and A'Hearn 2019). Zimran (2019) shows that although there is some evidence of sample-selection bias in the data used to establish the antebellum decline in stature, there was a true decline. Gallman (1996) and Komlos (1996) also discuss issues of the veracity of the observed trends in average stature.

[10] 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. Komlos (1998) also surveys a number of possible mechanisms.

gains may have reduced health as greater population density or urbanization increased exposure to disease by increasing contact between individuals and through the sanitation consequences of greater concentrations of population (Costa 1993; Steckel 1995). Under the food price explanation, transportation may have affected health through its impact on relative prices. The originally rural areas being linked to the transportation network in this period were largely food producing. Transportation linkage in this setting would tend to increase the farm-gate price of food relative to that of manufactures. Under certain price and income elasticities, this would induce individuals to consume less food (or reduce expenditure on food), reducing health (Komlos 1987; Komlos and Coclanis 1997).

Importantly, transportation need not have had a negative effect on health at all. Indeed, the rising incomes generated by new transport links (e.g., Donaldson and Hornbeck 2016) might have contributed to improving health, just as the rising incomes of the period were expected to improve health. Indeed, Baten and Fertig (2009) find evidence of such a positive effect in Germany in the late nineteenth century.

The empirical evidence underlying both of these explanations is limited. Largely as a result of limited data availability for the antebellum period, much of the evidence that has been marshaled in support of these explanations is suggestive, based either on cross-sectional correlations or on national time series. Indeed, to my knowledge, there does not exist strong causal evidence showing that any particular force caused declines in height.[11]

Among this prior research is the study of Haines, Craig, and Weiss (2003), to which the present paper is most closely related. Haines, Craig, and Weiss (2003) study the relationship between height (also using data from the Union Army records) and mortality, on the one hand, with transportation and a variety of measures of local agricultural production on the other. They find that transportation presence was associated with average stature being lower by about 0.25–0.30 inches. The data and methods available at the time that their study was carried out prevent the results from being given a strong causal interpretation. First, Haines, Craig, and Weiss's (2003, pp. 391–392) measure of transportation linkage was an indicator of whether a county was on a navigable waterway in

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

1840—the most detailed data available at the time. Besides the obvious limitations coming from the lack of time variation in these data, Atack (2013, p. 318) describes the difficulty in ascertaining the precise location of these waterways. Donaldson and Hornbeck (2016) also point out the limitations of a binary indicator of transportation linkage, which does not capture the impacts of construction elsewhere in the network. Moreover, Haines, Craig, and Weiss (2003) do not effectively address issues of endogeneity in the relationship between transport and health. A host of factors may have affected both transport linkage and health, creating a spurious relationship between the two. Thus, while Haines, Craig, and Weiss's (2003) results are consistent with transportation contributing to the antebellum decline in health, convincing causal evidence does not yet exist. Similar concerns constrain the related results of Cuff (2005) and Yoo (2012). Moreover, no prior evidence exists regarding the mechanism by which transportation may have affected health in this context. This paper is able to advance on all of these fronts, providing a clearer view of the role of the Transportation Revolution in shaping the evolution of health in the antebellum US.

2.2 Transportation in the Antebellum US

The first major internal improvement in the US was the Erie Canal, completed in 1825. Though coastal shipping was well developed by this point and some turnpikes had emerged, the interior of the continent was virtually isolated from the Atlantic Coast. The Erie Canal was thus revolutionary, dramatically easing interregional trade between the Northeast and the (upper) Midwest. The financial success of the Erie Canal sparked a surge in canal construction in the Northeast and Midwest, with the greatest investment in canal construction occurring in the decade ending 1845. Although many of these canals were financially unsuccessful, they had important impacts on the economic development of the areas through which they passed and on the counties that they linked to the broader market, increasing land values and reducing transport costs. The new canals also reshaped trade patterns, with long-distance trade being rerouted from the Mississippi and Ohio rivers to the Great Lakes and Erie Canal. Another main transportation innovation of the antebellum period was the steamboat, with improvements in steamboat construction technology and the clearance of hazards to navigation gradually opening first the main rivers and then the smaller tributary rivers to steam navigation. The first railroads were also built in the antebellum period, but largely in the 1850s—after the study period of this paper (Atack and Passell 1994, pp. 150–159).

The impact of improved water transportation on agriculture is well documented. Canals and steamboats made shipments of meat and grain from the Midwest to the markets of the Northeast and Europe possible for the first time. Indeed, shipments from western states "grew at an average annual rate of over 10 percent from the 1830s as production was increasingly oriented toward more distant markets" (Atack and Passell 1994, p. 162). Competition from Midwestern farms also caused a shift in the production of Northeastern farms towards more locally oriented goods, such as dairy (Atack and Passell 1994, pp. 160–171).

The impacts of canal construction and improvements in river navigability on non-agricultural outcomes, such as industrialization, have received less attention. However, some of the recent research on the impacts of railroads in the antebellum US suggests that expanded water transportation may also have had an effect on other forms of development. For instance, Hornbeck and Rotemberg (2019) find that the railroad increased manufacturing, using data from the 1860 Census of Manufactures. Such a potential impact of transportation is particularly important in the case of the Northeast, where farming became less important after expanding water transportation increased Midwestern competition, as discussed above. Atack et al. (2010) also find that the railroad caused an increase in urbanization in the Midwest during the 1850s. All of these impacts of transportation provide potential channels for transportation to have affected health. Among many other possible outcomes, the income gains coming from greater ability to produce agricultural and industrial products for larger markets may have generated population growth, both due to Malthusian responses (Ahmad 2019) and due to in-migration. This greater population concentration in turn may have led to a greater prevalence of disease.

3 Data and Empirical Approach

3.1 Height Data

The main outcome on which I focus is 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 US in that it is perhaps the only measure that can provide insights into health for the bulk of the population for a number of years.[12] Average stature is increased by greater calorie and protein consumption and by a better sanitary environment, while strenuous physical labor, malnutrition, and chronic disease tend to decrease it (Deaton 2007; Floud et al. 2011; Steckel 1995).[13]

Data on the heights of men born in the US 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

[12]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. Nevertheless, I do provide a cross-sectional analysis of the correlation between transportation and mortality in Table A.1 in Online Appendix A, finding additional evidence of deleterious health effects of transportation. Time series of life expectancy are also available, but cover only specific subsets of the population. It is also possible to link the records of Union Army soldiers who survived the war to their death records (e.g., Hong 2019), but this is only possible for a limited portion of my sample (i.e., those who survived the war), and has only been performed for an even more limited sample (i.e., war survivors from the Fogel et al. 2000 dataset), making this analysis impractical. There are also concerns regarding selection into war survival as well as the effects of wartime experience on life expectancy to consider in such an analysis.

[13]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. 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 in Online Appendix A showing that death rates were greater in counties with greater market access, to interpret declines in average stature as deteriorations (rather than as imporovements) in health, and vice versa.

Pennsylvania and serving in Pennsylvania regiments. Finally, I collected and digitized approximately 3,000 additional observations from the original records.

As is standard in the use of these data, I restrict the sample to white men born in the Northeast or the Midwest. Coverage of the South is sparse because of the very limited representation of southerners in the Union Army. I also exclude individuals measured before age 18, which implies that the youngest birth cohort included in the sample is that of 1847. 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, height, birth year, and age of measurement are known.[14] After imposing these restrictions, 31,402 observations remained. For a subset of these observations, the county of enlistment could also be determined.

I restrict attention to individuals born in counties that had no urban population in 1820, reducing the sample to 25,566 individuals.[15] I make this restriction for three main reasons. First, Steckel (1995, p. 1927) points out that understanding patterns in height in the antebellum US requires a focus on the rural sector. This sector encompasses the bulk of the population, and drove the bulk of the decline in stature. Moreover, the decline in stature of the urban sector is less puzzling than the decline of the rural sector, as the growing cities of the period were noted as unhealthy (e.g., Floud et al. 2011; Haines 2004); the challenge is to explain the health declines of rural areas. Second, transport linkages are not hypothesized as explanatory factors for declining stature in urban areas—indeed, if anything better transport linkages between cities and food-producing areas would tend to improve health (Zehetmayer 2013). It is the effect of transport on rural areas that is important in searching for an explanation for the Antebellum Puzzle. Finally, many transport links were built to link the interior to specific urban areas (e.g., the Erie Canal, Pennsylvania's Main Line, the C&O Canal), implying that endogeneity of transport is particularly severe for these areas.

A key question regarding the enlistment data is whether they are representative of the broader

[14]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. One observation was removed because it listed an enlistment year of 1860, which is impossible given the creation of the Union Army in 1861.

[15]Figure A.1 in Online Appendix A indicates the counties removed from the sample by this restriction.

population of interest—native-born rural white males in the Northeast and Midwest in the birth cohorts of 1820–1847. The over-sampling of Pennsylvanians is one obvious concern, which I address by re-weighting so that the distribution of states of residence matches that of the 1860 census. 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.[16] 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, there were a number of individuals who were able to enlist despite the restriction, suggesting that this is not a completely proper modeling of the role of the minimum height requirement on enlistment. As a result, the results reported below do not use such a truncation-corrected regression. Online Appendix C repeats the main results with this method of analysis, showing that the results are robust to it.

A more nuanced concern is that selection into military service was non-random (Bodenhorn, Guinnane, and Mroz 2017). This concern is mitigated by the fact that nearly half of the population at risk for military service enlisted (Zimran 2019). 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 (2019) formal investigation of bias in historical height data sources, which finds that the height data in the Union Army records suffer from little bias.[17] Moreover, my inclusion of county, birth year, and enlistment year fixed effects in all main specifications addresses many of the possible forms of this bias, such as changing incentives for enlistment over time (Bodenhorn, Guinnane, and Mroz 2017) and differences over counties in ideology (Zimran 2019, 2020). But in the end I am not able to rule out the possibility that selection may have varied by county-birth cohort such that selection changed in response to transportation creating a relationship between *observed* height and transportation that does not reflect the relationship between

[16]This is shown in Figure A.2 in Online Appendix A.

[17]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 (2019) shows that combining data from the Union Army and from the records of enlistments for the later period 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.

actual height and transportation.[18] Despite this limitation, the Union Army data provide the best available view of trends in health during the antebellum period.

3.2 Transportation Data

Information on transportation infrastructure is given by GIS shape files created by Atack (2015, 2016, 2017). These files provide the location of all steamboat-navigable rivers, canals, and railroads in the continental US 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). These shape files also provide the year in which any particular segment of canal, rail, or river first became available (and, if applicable, when they ceased to be).

I use these data to create two measures of transportation linkage. 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 zero otherwise.[20] While it is a straightforward measure and is used in prior research (e.g., Atack et al. 2010; Haines, Craig, and Weiss 2003), this faces some important limitations. First, it does not capture the impacts of new forms of transportation entering already linked areas. Relatedly, it treats all transportation links as equivalent, though it is clear that they would not be. More importantly given the transportation history of the period, 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 New York. An alternative measure that allows heterogeneity in the value of a transport link and that captures the effects of distant construction is therefore preferable.

Donaldson and Hornbeck's (2016) market access measure addresses these shortcomings, and is therefore my second and preferred measure of transport linkage. Formally, the measure is created

[18] Ruling this out would require identifying a county-year-varying force affecting entry into the military but not height.

[19] I have supplemented these files with my own hand-traced shape files describing the canals and rivers of the St. Lawrence and Champlain waterways in Canada.

[20] I treat all coastal counties as being always linked. This simplification, as well as, for instance, treating all ports as providing equal access to the ocean, is a coarse but necessary approximation.

as follows. Following an algorithm described in Online Appendix D, I compute approximate iceberg transportation costs, $\tau_{jkt} \geq 1$, between each county pair jk in each year $t \in \{1820, \ldots, 1847\}$. These costs are based on the least-cost shipping path between each county pair according to state of the transportation network in each year and estimates of mode-specific transport costs by distance. Market access in county j for year t is then defined as

$$\mathrm{MA}_{jt} = 10^{-4} \sum_k p_k \tau_{jkt}^{\theta}, \tag{1}$$

where p_k is the population of county k in 1820. The choice to use 1820 population rather than year t population is made because allowing population to change over time would cause market access to capture both improvements in transportation linkage and population growth.[21] My interest is in the effects of the former; the latter would have its own impacts on health. I compute equation (1) using a value of $\theta = -3.51$,[22] though ultimately the choice of θ, which governs the rate at which transport costs decrease connectedness, has little impact on the qualitative results.[23]

Less formally, market access is a measure that increases as a county becomes closer—in the sense of reduced transportation costs—to population in other parts of the country. As new transportation links are constructed, they reduce transportation costs from one county to others (and differentially so based on their location in space or in the network), increasing a county's market access. In many ways, market access provides similar information to the binary measure of linkage—given the high cost of land transportation, the arrival of a transport link in an unlinked county drastically reduces the transport cost between this county and all others, raising market access considerably. But it is also a richer measure than this. Donaldson and Hornbeck (2016) explain that it can capture the impacts of improvements throughout the transport network on a county. This is most clearly illustrated by the construction of the Erie Canal, through it applies to all other transport construction. The Erie Canal was built in upstate New York, and in the sense of the binary transport

[21] I have repeated the analysis with population fixed at 1840 and with year t population (based on interpolation between censuses). Results in each case are similar to those using 1820 population.

[22] I arrive at this value by estimating equation (3) by nonlinear least squares.

[23] 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) such that the estimated relationship between a one-standard deviation change in market access and height is virtually unchanged. Indeed, when θ is set to -1, the estimates of β are qualitatively almost identical.

indicator would appear to affect only those counties through with it passed. But it is well known that the Erie Canal had a profound impact on the upper Midwest. The market access measure captures this effect—the transport cost between Midwestern counties and the population centers of the Northeast would be considerably reduced by the opening of a water route linking the Great Lakes and the Atlantic, leading to an increase in the market access of the midwestern counties.

Market access also captures the arrival of new transport to already linked areas, and allows heterogeneity in the impact of transport linkages—their impact depends on the ability of new transport to reduce transport costs between a county and others in the US. For instance, construction that linked one county only to another sparsely populated county would lead to a smaller increase in market access than construction that linked that county to a major city. Similarly, the first construction linking a county to a major city would generate a greater increase in that county's market access than a second link that only slightly reduced travel costs.

3.3 Other Data

I supplement the height and transportation data with county-level data from the decennial US 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, as implemented by Perlman (2014).

3.4 Summary Statistics

Using the sources described above, I created and merged 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 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 matches each individual in the Union Army data to the transportation linkage and market access of his county-year of birth.

Table 1 summarizes the county-level measures of transportation linkage, divided by region and decadal year, and weighted by population. There was 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 exhibited similar patterns, although the population of the Northeast was consistently more linked than was that of the Midwest.

Figure 1 provides a graphical summary of the spread of transportation infrastructure over this period. Panels (a)–(c) show the spread of each mode of transportation. Panel (d) uses this information to determine the year that a county was first linked to the (non-wagon) transport network. It shows that the transportation network gradually spread inland. 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, particularly in the North.

Market access 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. The development of the market access measure over time and space is described graphically in Figure 2. This Figure depicts the change in market access in each decade, shading counties with greater increases darker.[24] It shows that, as intended, 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

[24]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 and on a constant scale are presented in Figure A.3 in Online Appendix A.

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—a mechanical consequence of weighting the data to reflect state population in 1860. Figure 3 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. 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.[25] Nonetheless, some counties are represented by only a small number of observations. To ensure that limited representation of some counties does not create spurious results, Online Appendix E repeats the main analysis limiting the sample either to the better represented states of New York and Pennsylvania or to counties with at least 100 observations, with similar results.

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 relative to those whose birth counties were unlinked in the birth year.[26]

There are also differences between regions and between linked and unlinked counties in measures

[25]The number of observations by birth cohort is given in Figure A.4 in Online Appendix A. 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.

[26]Figure A.2 in Online Appendix A 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.

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 for individuals born in linked counties.[27] 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.

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.[28] 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.

3.5 Empirical Specification

The basic specification that I use to study the relationship between transportation and height is

$$h_{ijt} = \gamma_t + \delta_a + \delta_e + \beta x_{jt} + \mathbf{z}_j' \tau + \varepsilon_{ijt}. \tag{2}$$

In equation (2), h_{ijt} is the height of individual i born in county j in year t; 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; δ_e are enlistment-year fixed effects designed to address the possibility of changing composition of military enlisters over the course of the Civil

[27]Recall that counties with an urban population in1820 are omitted, but other counties included in the sample may have developed an urban population after 1820. 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.

[28]In computing this figure, I limited the sample to those who enlisted in the state with which their regiment was associated (e.g., members of a Massachusetts regiment enlisting in Massachusetts). In some cases, individuals enlisted while the regiment was in the field. As I do not wish to consider military deployment as a form of migration, I exclude these individuals when considering county-level migration.

War;[29] x_{jt} is a measure of transportation linkage in the birth year (either the binary measure or the logarithm of market access), and $\mathbf{z}_j$ is a vector of various county-level control variables to be introduced in section 4 below.[30] I cluster standard errors throughout the analysis at the county level.[31] 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 US (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 conditions in the birth year are 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 Online 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.

Although equation (2) is useful in establishing a correlation, 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.[32] Moreover, the sanitation consequences of population concentration might make this area unhealthy. This hypothetical relationship would produce a negative β in specification (2) even if the true β were zero. Not all confounds must be in this direction. For instance, if better agricultural land

[29] It is possible to include all of δ_a, δ_e, and γ_t because the measurement-age fixed effects are not for all possible measurement ages, but only for ages below 21.

[30] 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."

[31] Online Appendix F repeats the main results with standard errors clustered at both the county and the birth year level, with similar results.

[32] Given the exogenous location of rivers, this is less a concern for steamboats than for canals; but the presence of a river might have its own independent impact on health (such as through the transmission of disease, whether the river is navigable or not), and the decision to clear navigation hazards might have depended on local development.

attracted transport construction and raised incomes and health, a spurious positive β would arise.

The improved data available from Atack's (2015, 2016, 2017) shape files and the advantages of market access as a measure of transport linkage enable me to address these endogeneity concerns. First, the panel structure of the transportation data enable me to augment equation (2) with the addition of county fixed effects α_j (requiring the omission of the county-specific controls $\mathbf{z}_j$) so that it becomes

$$h_{ijt} = \alpha_j + \gamma_t + \delta_a + \delta_e + \beta x_{jt} + \varepsilon_{ijt}. \tag{3}$$

This specification captures time-invariant county characteristics (such as a favorable location), identifying the effect of transportation on height on the basis only of local changes.

A concern that remains in equation (3) is that a force other than transportation that varies over time within a county might both affect health and attract transportation. For instance, a favorable location might only become important as the economy grew. This concern would 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.[33] Another approach is made possible by the use of market access to measure transportation linkage. Specifically, I can estimate

$$h_{ijt} = \alpha_j + \gamma_t + \delta_a + \delta_e + \beta \log(\mathrm{MA}_{jt}) + \pi T_{jt} + \varepsilon_{ijt}, \tag{4}$$

where MA_{jt} is market access computed as in equation (1) and T_{jt} is a binary indicator of transportation linkage. The coefficient of interest in this equation is β. Based on the identification strategy used by Donaldson and Hornbeck (2016) and Hornbeck and Rotemberg (2019), this specification identifies the effect of market access on height on the basis only of variations in market access induced by construction elsewhere in the network. That is, any endogeneity is likely to be capture by directly controlling for local transport linkage. The remaining variation in market access, which drives the estimate of β, is then relatively free of this endogeneity.

Online Appendix G presents an additional identification strategy, based on the straight-line instruments commonly used in the transport literature (e.g., Atack et al. 2010; Banerjee, Duflo,

[33] My inclusion of region- and state-specific time trends in the analyses below also serves to partially address this concern, as does my analysis in Table A.2, which includes year-specific functions of latitude and longitude.

and Qian 2012; Ghani, Goswami, and Kerr 2016; Hornung 2015). In particular, I compute an alternative measure of market access based on the 1820 transport network augmented with canals built to connect major watersheds (i.e., the Atlantic, Great Lakes, and Mississippi) to one another and to major cities. I then use this alternative measure of market access to instrument for actual market access. This analysis, which also finds evidence of a negative effect of market access on height, provides supporting evidence based on an entirely different source of identifying variation.

4 The Effect of Transportation on Height

4.1 OLS Results

I begin the analysis by estimating equation (2) by ordinary least squares using the binary indicator of transportation linkage as the regressor of interest. This estimation most closely matches that performed by Haines, Craig, and Weiss (2003). Results of this estimation are presented in Panel A of Table 3. The regression of column (1), which includes only birth-year indicators, age-of-measurement indicators, enlistment-year 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 fixed effects in column (2), though this addition reduces the magnitude of the estimated coefficient by about half. This latter estimate indicates that individuals whose counties of birth had a transport link 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 2019). 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.

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 also 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 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. For the main results, I will capture these characteristics with county fixed effects.

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). The addition in column (4) of interactions of birth year and region fixed effects, or of the interactions of state and birth year fixed effects in column (5), has little additional impact on the estimates.

One potential reason for the lack of a statistically significant relationship between transport and height in columns (3)–(5) of Panel A of Table 3 is that the transport indicator captures only a limited amount of variation in linkage.[34] Besides the conceptual benefits of the market access measure, discussed above, it also has the technical benefit of providing more treatment variation in

[34]This is shown in Figure A.5 in Online Appendix A, 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 (which is oversampled), has at least 25 observations of individual heights both before and after the change in transportation linkage.

a setting where data are limited. Panel B of Table 3 repeats the analysis of Panel A, but replaces the binary measure of linkage with the logarithm of market access as the explanatory variable of interest. Columns (1) and (2) estimate equation (2) 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 (2), which include the state fixed effects, indicate that a one-standard deviation increase in market access (0.28 log points, as shown in Table 2) is associated with a reduction in average height of 0.171 inches.

Columns (3)–(5) repeat this estimation, including the various county-level control variables, and the region-by-birth year or state-by-birth year indicators. 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 (5), 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 0.115 inches, or about 1.7 times the implied impact of a transportation linkage in its analog in Panel A.

4.2 Fixed Effects Results

As discussed above, the results of Table 3 are potentially spurious. To begin addressing endogeneity concerns, Table 4 presents estimates of equation (3), which includes county fixed effects. I begin in column (1) by using 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 -0.042, which is smaller than the estimates including controls in Table 3, and is statistically insignificant. Given the conceptual and technical limitations of the transport linkage indicator, as discussed above, the absence of a meaningful transport-health relationship using this regressor is not surprising.

Column (2) estimates the same specification with the logarithm of market access as the regressor of interest. This column reveals that the negative and statistically significant relationship between market access and height is robust to the inclusion of the county fixed effects, and thus to the concerns that they address over endogeneity. Moreover, at -0.597, the magnitude of the coefficient is comparable to the estimates of Table 3.

This result and its approximate magnitude are robust to the inclusion, in column (3), of county-decade 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 being marginally statistically insignificant ($p = 0.102$), the rough magnitude and sign of the coefficient is retained, supporting the conclusion that transportation improvements generated declines in stature-implied health.

Table 5 investigates the extent to which the height-market access relationship uncovered in Table 4 is driven by connections to nearby or more distant counties. In particular, I adjust equation (1) to compute market access as

$$\widetilde{\text{MA}}_{jt} = 10^{-4} \sum_k p_k \tau_{jkt}^{\theta} \mathbb{1}\left\{d_{jk} > \bar{d}\right\}, \tag{5}$$

where d_{jk} is the distance between counties j and k, $\mathbb{1}\{\cdot\}$ is the indicator function, and $\bar{d}$ is a cutoff value. I then re-estimate equation (3) using this alternative measure of market access as the regressor of interest. Panel A of Table 5 uses a 40-mile cutoff, following Fogel's (1964) use of this distance as an important cutoff; Panel B uses a 100-mile cutoff. The results of Table 5 indicate that the impact of market access on height largely comes from the reduction of transport costs to more distant markets—the results are virtually identical to those of Table 4 even after replacing MA_{jt} with $\widetilde{\text{MA}}_{jt}$ computed as in equation (5).

To further address potential endogeneity of transport linkage, I estimate equation (4). By controlling directly for transport linkage, these specifications identify the effect of market access on height using only variation in linkage coming from construction elsewhere in the network, which is less likely to be endogenous to local development than is the directly controlled-for local construction. Reassuringly, the negative effect of market access on height documented in Table 4 is made even stronger by the results of Table 6, which provides plausibly causal estimates consistent with this effect.[35] Overall, based on the results of Tables 4–6, I conclude that the data provide strong and robust evidence of a negative relationship between stature and market access in the county-year of birth.

Having shown that transportation linkages had a deleterious impact on health in the antebellum period, these results provide an empirical basis for a potential explanation of the Antebellum Puzzle—that it was caused in part by the expansion of the transportation network in the antebellum period. 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 to changes in local circumstances (by virtue of the county fixed effects included in the estimation).

How much of the deterioration in average stature can be attributed to the growth of the transportation network? The empirical pattern to be explained is a 0.82 inch decline in average stature that is present in the benchmark sample.[36] Table 1 shows that log market access increased by 0.603 from 1820 to 1850. My preferred estimate of the effect of market access on height is given by column (4) of Table 6, which uses the Donaldson and Hornbeck (2016) and Hornbeck and Rotemberg (2019) method to address endogeneity and which includes the most conservative time trends.[37] The coefficient of -0.501 thus implies that changing market access can explain a decline in average stature of 0.30 inches, or 37 percent of the observed decline.

It is also possible that a negative relationship between market access in the county-year of birth and health as measured by adult stature may be the result of individuals leaving newly linked locations to move to less healthy urban environments, implying that the new transport linkages did

[35]Table A.3 in Online Appendix A replaces the transport-linkage indicator with indicators for linkage to rail, canal, and other water transport. Results are similar.

[36]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 (2019) 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.

[37]Table A.2 in Online Appendix A repeats the main results with year-specific functions of latitude and longitude, providing even more conservative estimates. Though somewhat less precise, the results are largely similar.

not reduce the health of the place of birth, but instead pushed individuals to move to other, less healthy locations. Alternatively, the results may be the product of less healthy individuals moving to newly linked areas and having children there who have lower adult stature. Neither of these cases would inhibit the interpretation of the results above as the causal effect of transportation linkages on the terminal heights of individuals born in an affected county-year; but if these were the mechanisms by which the effect operated, the interpretation of the results as explaining an overall decline in average stature in the whole economy would be more tenuous. In Online Appendix H, I show that, to the extent that it is possible to evaluate these mechanisms, they are not consistent with the data.

5 The Local Development Mechanism

As discussed in section 2, there are at least three mechanisms by which market access may have negatively impacted stature-implied health—disease transmission, local development, and relative prices. Distinguishing between these various theoretical mechanisms is important both in better understanding the effects of transportation, as well as in understanding the causes of the antebellum decline in health and ultimately the welfare effects of early modern economic growth in the US. Unfortunately, the limited data availability for the antebellum US makes it impossible to evaluate all of these possible mechanisms. But there are sufficient data to determine whether there is empirical support for the mechanism of transportation increasing population density, thereby decreasing health, which I call the local development mechanism.

A necessary condition for this mechanism to be valid is that there should be a positive impact of increased market access on population density. I test this by estimating the equation

$$\log(y_{jt}) = \alpha_j + \gamma_t + \beta \log(\text{MA}_{jt}) + \pi T_{jt} + \varepsilon_{jt}$$

for years 1820–1850, where y_{jt} is the population density of county j in year t, and the coefficient of interest is β. This is an adapted version of equation (4), and like this equation it identifies the effect of market access by observing the response to improvements in market access coming from construction elsewhere in the network. I present the results in column (1) of Table 7. 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 changes in population density caused by changes in county boundaries. I weight observations according to the ratio of county population in each year to total population in that year. I find a large, positive, and statistically significant impact of market access on population density. The estimated coefficients imply that the increase in market access from 1820 to 1850 (0.603 in the data set with counties as the unit of observation, as shown in Table 1) is associated with a 0.362 log point increase in population density.

This result is an important piece of evidence in favor of the local development mechanism—it is well known that greater population density was associated with worse health (Floud et al. 2011), and it stands to reason that by inducing greater population density, transportation may have harmed health. More evidence, however, is available from heterogeneity in the effects of increased market access. I focus specifically on heterogeneity with respect to a county's suitability for wheat and corn production.[38] Given the importance of these products in the long-distance trade that market access measures (as indicated in part by Table 5), I expect a greater increase in population density in response to increases in market access in more agriculturally suitable counties, as the benefits of greater market access are likely to be larger there.

Columns (2) and (3) of Table 7 test this prediction by interacting market access with the logarithm of a measure of a county's average suitability for wheat or corn production (Food and Agriculture Organization 2002). The results are consistent with this prediction, revealing 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 estimates of column (2) can be interpreted as implying that a county with average wheat suitability experienced an increase in population density of 0.616 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 0.654 percentage point stronger reaction. Similar results are evident for corn suitability. Thus, counties with greater

[38]Any source of heterogeneity in the impact of market access on population density and health could be used here. I focus on these measures because they are some of the few measures that can be observed that are exogenous to the arrival of transportation, given the limited data availability prior to 1840.

agricultural suitability tended to experience greater increases in population density in response to improvements in transport linkages, likely reflecting immigration of individuals seeking to establish farms. This view is further supported by Table A.4 in Online Appendix A, which shows an increase in the acreage devoted to farming in response to rising market access, with greater increases in counties with greater suitability for wheat and corn production.

This heterogeneity in the effect of market access on population provides an opportunity to test the local development mechanism. If population density increased more in response to rising market access in more agriculturally suitable counties, then average stature should have decreased more in response to rising market access in these counties.[39] To test this, I repeat the interaction approach in estimating equation (4) in columns (4) and (5) of Table 7, with height as the dependent variable. I find that, as predicted, the negative relationship of market access with stature was stronger in the more agriculturally suitable counties, though the interaction coefficients are only marginally statistically significant ($p = 0.095$ for wheat suitability and $p = 0.159$ for corn suitability).

Altogether, the results of columns (2)–(5) of Table 7 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, in a setting of poor sanitation and public health, worsened the local disease environment.

This evidence thus adds to the debate over the causes of the antebellum deterioration in health. As discussed above, the local development mechanism is a component of the disease explanation for the Antebellum Puzzle. By finding evidence consistent with this mechanism, I thus add to the evidentiary basis of the disease explanation.[40]

[39] 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.

[40] As with the analysis of the effect of transportation on height, I verify the robustness of these results to a variety of other sample restrictions and specifications in Online Appendices C, E, and F.

6 Conclusion

Why did health in the US deteriorate in the antebellum period, just as modern economic growth began? This is an important question of American economic history—one that bears centrally on understanding the evolution of living standards on the American path to industrial leadership. Since this Antebellum Puzzle was first discovered by Fogel et al. (1979), economic historians have sought to explain why health deteriorated at the onset of modern economic growth, which should have improved it. But evidence as to possible answers to this question is limited. In this paper, I test whether the transportation improvements of the antebellum period contributed to the decline in average stature.

Specifically, I study the relationship between transportation and height in the antebellum US using data on stature from the records of enlistments in the Union Army combined with GIS data on the development of the antebellum transportation network. I find evidence of a negative effect of market access on average stature that is large enough to explain 37 percent of the decline in average stature in my data. Evidence supporting such a relationship had previously been found (Cuff 2005; Haines, Craig, and Weiss 2003; Yoo 2012), but improvements in data availability (Atack 2013) and methods for the study of transportation (Donaldson and Hornbeck 2016) enable me to provide plausibly causal estimates where only suggestive correlations had previously been available. I also go beyond this existing research by providing evidence consistent with the contention that a mechanism for this effect was the role of transportation in increasing population density, which would have led to a deterioration in the epidemiological environment in a period of poor sanitation technology.

This paper thus helps to better understand the impacts of transportation infrastructure construction in the historical US, and in developing countries more generally. In such settings the beneficial effects of transportation improvements for economic development are well known (Bogart 2018; 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). This paper thus sheds light on an aspect of the literature that is only beginning to come into focus.

On the whole, the results of this paper provide an interesting and important warning. The Antebellum Puzzle cautions that the early phases of modern 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 an unintended negative impact on at least some aspect of welfare and therefore may also not be 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.

Ahmad, Shameel (2019). "Endogenous Demography and Welfare Gains from Trade: Evidence from British India." Mimeo., Brandeis University.

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.

Baten, Joerg and Georg Fertig (2009). "Did the Railway Increase Inequality? A Micro-Regional Analysis of Heights in the Hinterland of the Booming Ruhr Area During the Late Nineteenth Century." *Journal of European Economic History* 38:2, pp. 263–299.

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?" *Journal of Urban Economics* 103124.

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* 10, pp. 369–380.

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.

——— (2019). "Diagnosing Sample-Selection Bias in Historical Heights: A Reply to Komlos and A'Hearn." *Journal of Economic History* 79:4, pp. 1154–1175.

Bogart, Dan (2018). "Clio on Speed: A Survey of Economic History Research on Transport." In *Handbook of Cliometrics*. Claude Diebolt and Michael Haupert (ed.). Springer.

Bogart, Dan, Xuesheng You, Eduard Alvarez, Max Satchell, and Leigh Shaw-Taylor (2019). "Railway Endowments and Population Change in 19th Century England and Wales." Mimeo., UC Irvine.

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

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. (2016). "Antebellum Puzzle: The Decline in Heights at the Onset of Modern Economic Growth." In *Handbook of Economics and Human Biology*. John Komlos and Inas Kelly (ed.). Oxford: Oxford University Press.

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.

David, Paul A. (1967). "The Growth of Real Product in the United States before 1840: New Evidence, Controlled Conjectures." *Journal of Economic History* 27:2, pp. 151–197.

Davis, Joseph H. (2004). "An Annual Index of US Industrial Production, 1790–1915." *Quarterly Journal of Economics* 119:4, pp. 1177–1215.

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

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

——— (2018). "Railroads of the Raj: Estimating the Impact of Transportation Infrastructure." *American Economic Review* 108:4–5, pp. 899–934.

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.

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.

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.

——— (1994). "Economic Growth, Population Theory, and Physiology: The Bearing of Long-Term Processes on the Making of Economic Policy." *American Economic Review* 84:3, pp. 369–395.

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.

Fogel, Robert W., Stanley L. Engerman, Roderick Floud, Richard H. Steckel, T. James Trussell, Kenneth W. Wachter, Robert A. Margo, Kenneth Sokoloff, and Georgia C. Villaflor (1979). "The Economic and

Demographic Significance of Secular Changes in Human Stature: The US 1750–1960." NBER Working Paper.
Food and Agriculture Organization (2002). *Global Agro-Ecological Zones* [machine-readable database]. New York: United Nations.
Gallman, Robert E. (1996). "Dietary Change in Antebellum America." *Journal of Economic History* 56:1, pp. 193–201.
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.
Goldin, Claudia and Robert A. Margo (1992). "Wages, Prices, and Labor Markets before the Civil War." In *Strategic Factors in Nineteenth Century American Economic History*. Claudia Goldin and Hugh Rockoff (ed.). Chicago: University of Chicago Press. Chap. 2, pp. 67–104.
Haines, Michael R. (2004). "Growing Incomes, Shrinking People—Can Economic Development Be Hazardous to Your Health?" *Social Science History* 28:2, pp. 249–270.
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.
——— (2019). "Shortened Lifespan: A Legacy of Exposure to Malaria Risk in Early Life." Mimeo., Seoul National University.
Hornbeck, Richard (2010). "Barbed Wire: Property Rights and Agricultural Development." *Quarterly Journal of Economics* 125, pp. 767–810.
Hornbeck, Richard and Martin Rotemberg (2019). "Railroads, Reallocation, and the Rise of American Manufacturing." Mimeo., University of Chicago.
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 (2019). "National Policy for Regional Development: Evidence from Appalachian Highways." *Review of Economics and Statistics* Forthcoming.
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.
——— (1996). "Anomalies in Economic History: Toward a Resolution of the 'Antebellum Puzzle'." *Journal of Economic History* 56:1, pp. 202–214.
——— (1998). "Shrinking in a Growing Economy? The Mystery of Physical Stature during the Industrial Revolution." *Journal of Economic History* 58:3, pp. 779–802.
——— (2012). "A Three-Decade History of the Antebellum Puzzle: Explaining the Shrinking of the U.S. Population at the Onset of Modern Economic Growth." *Journal of the Historical Society* 12:4, pp. 395–445.
Komlos, John and Brian A'Hearn (2019). "Clarifications of a Puzzle: The Decline in Nutritional Status at the Onset of Modern Economic Growth in the United States." *Journal of Economic History* 79:4, pp. 1129–1153.
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.
Margo, Robert A. (2000). *Wages and Labor Markets in the United States, 1820–1860*. Chicago: University of Chicago Press.

Margo, Robert A. and Richard H. Steckel (1983). "Heights of Native-Born Whites During the Antebellum Period." *Journal of Economic History* 43:1, pp. 167–174.

McGuire, Robert A. and Philip R. P. Coelho (2011). *Parasites, Pathogens, and Progress: Diseases and Economic Development*. Cambridge: MIT Press.

McKeown, Thomas (1976). *The Modern Rise of Population*. London: Arnold.

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

Perlman, Elisabeth (2014). *Tools for Harmonizing County Boundaries* [computer software].

Pope, Clayne (1992). "Adult Mortality in America before 1900: A View from Family Histories." In *Strategic Factors in Nineteenth Century American Economic History: A Volume to Honor Robert W. Fogel*. Claudia Goldin and Hugh Rockoff (ed.). Chicago: University of Chicago Press. Chap. 9, pp. 267–296.

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.

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.

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

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.

——— (2013). "Health, Market Integration, and the Urban Height Penalty in the US, 1847–1894." *Cliometrica* 7, pp. 161–187.

Zimran, Ariell (2019). "Sample-Selection Bias and Height Trends in the Nineteenth-Century United States." *Journal of Economic History* 79:1, pp. 99–138.

——— (2020). "Recognizing Sample-Selection Bias in Historical Data." *Social Science History* Forthcoming.

Tables

Table 1: Summary statistics for county-level data

Variable	(1) 1820	(2) 1830	(3) 1840	(4) 1850
Panel A: All Counties				
Transportation Present	0.396	0.562	0.711	0.805
	(0.489)	(0.496)	(0.454)	(0.397)
log(Market Access), 1820 Pop.	5.142	5.531	5.664	5.745
	(0.457)	(0.272)	(0.239)	(0.231)
Counties	945	945	945	945
Panel B: Midwest				
Transportation Present	0.316	0.476	0.599	0.704
	(0.465)	(0.500)	(0.491)	(0.457)
log(Market Access), 1820 Pop.	4.988	5.396	5.542	5.626
	(0.341)	(0.220)	(0.207)	(0.202)
Counties	774	774	774	774
Panel C: Northeast				
Transportation Present	0.420	0.600	0.796	0.907
	(0.495)	(0.491)	(0.404)	(0.291)
log(Market Access), 1820 Pop.	5.187	5.589	5.757	5.866
	(0.478)	(0.272)	(0.221)	(0.194)
Counties	171	171	171	171

Notes: The sample in Panel A includes all counties with no urban population in 1820. Panels B and C divide this sample by region. Means presented with standard deviations in parentheses. Observations weighted by population in each decade.

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.	5.571	5.495	5.634	5.701	5.323
	(0.277)	(0.244)	(0.286)	(0.177)	(0.262)
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	25,566	10,209	15,357	16,874	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 counties with a transport link in the birth year; 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.

Table 3: OLS regressions

Variables	(1)	(2)	(3)	(4)	(5)
Panel A: Binary transport indicator					
Transport	−0.335***	−0.169***	−0.068	−0.071	−0.067
	(0.072)	(0.057)	(0.067)	(0.067)	(0.070)
Observations	25,556	25,556	23,557	23,557	23,557
R-squared	0.057	0.075	0.079	0.081	0.108
Panel B: Market access					
log(Market Access), 1820 Pop.	−1.087***	−0.612***	−0.432***	−0.410**	−0.410**
	(0.127)	(0.127)	(0.161)	(0.162)	(0.175)
Observations	25,556	25,556	23,557	23,557	23,557
R-squared	0.064	0.077	0.080	0.082	0.108
State FE	No	Yes	Yes	Yes	Yes
Controls	No	No	Yes	Yes	Yes
Birth Year × Region FE	No	No	No	Yes	No
Birth Year × State FE	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, measurement age, and enlistment year fixed effects. Standard errors clustered at the county level. Observations weighted to correct for oversampling.

Table 4: County fixed effects regressions

Variables	(1)	(2)	(3)	(4)	(5)
Transport	−0.042				
	(0.116)				
log(Market Access), 1820 Pop.		-0.597^{***}	-0.744^{**}	-0.517^{**}	−0.386
		(0.210)	(0.305)	(0.204)	(0.236)
Observations	25,556	25,556	25,556	25,556	25,556
R-squared	0.126	0.126	0.174	0.129	0.157
County × Decade FE	No	No	Yes	No	No
Birth Year × Region FE	No	No	No	Yes	No
Birth Year × State FE	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, measurement age, enlistment year, and county fixed effects. Standard errors clustered at the county level. Observations weighted to correct for oversampling.

Table 5: Market access omitting nearby counties

Variables	(1)	(2)	(3)	(4)
Panel A: 40 mile radius				
log(Market Access), 1820 Pop.	−0.583***	−0.719**	−0.509**	−0.387*
	(0.205)	(0.300)	(0.200)	(0.233)
Observations	25,556	25,556	25,556	25,556
R-squared	0.126	0.174	0.129	0.157
Panel B: 100 mile radius				
log(Market Access), 1820 Pop.	−0.560***	−0.694**	−0.499***	−0.390*
	(0.196)	(0.287)	(0.190)	(0.223)
Observations	25,556	25,556	25,556	25,556
R-squared	0.126	0.174	0.129	0.157
County × Decade FE	No	Yes	No	No
Birth Year × Region FE	No	No	Yes	No
Birth Year × State FE	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, measurement age, enlistment year, and county fixed effects. Panel A computes a county's market access omitting counties within 40 miles. Panel B computes a county's market access omitting counties within 100 miles. Standard errors clustered at the county level. Observations weighted to correct for oversampling.

Table 6: County fixed effects regressions controlling for transport linkage

Variables	(1)	(2)	(3)	(4)
log(Market Access), 1820 Pop.	−0.712***	−0.754**	−0.606***	−0.501**
	(0.226)	(0.320)	(0.221)	(0.249)
Observations	25,556	25,556	25,556	25,556
R-squared	0.127	0.174	0.129	0.157
County × Decade FE	No	Yes	No	No
Birth Year × Region FE	No	No	Yes	No
Birth Year × State FE	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, measurement age, enlistment year, and county fixed effects, as well as an indicator for having a transport link in the birth year in the county of birth. Standard errors clustered at the county level. Observations weighted to correct for oversampling.

Table 7: The local development mechanism

Variable	(1) Dens.	(2) Dens.	(3) Dens.	(4) Height	(5) Height
log(Market Access), 1820 Pop.	0.600***	0.616***	0.580***	−0.579***	−0.580***
	(0.150)	(0.143)	(0.145)	(0.212)	(0.211)
log(MA) × log(Wheat Suit.)		0.654***		−0.842*	
		(0.203)		(0.509)	
log(MA) × log(Corn Suit.)			0.490***		−0.646
			(0.144)		(0.460)
Observations	1,166	1,166	1,166	25,556	25,556
R-squared	0.925	0.929	0.929	0.127	0.127

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 and control for transportation linkage. Columns (4) and (5) also include measurement age and enlistment year 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.

Figures

(a) Canals | (b) Steamboat-Navigable Rivers | (c) Railroads | (d) Year of arrival

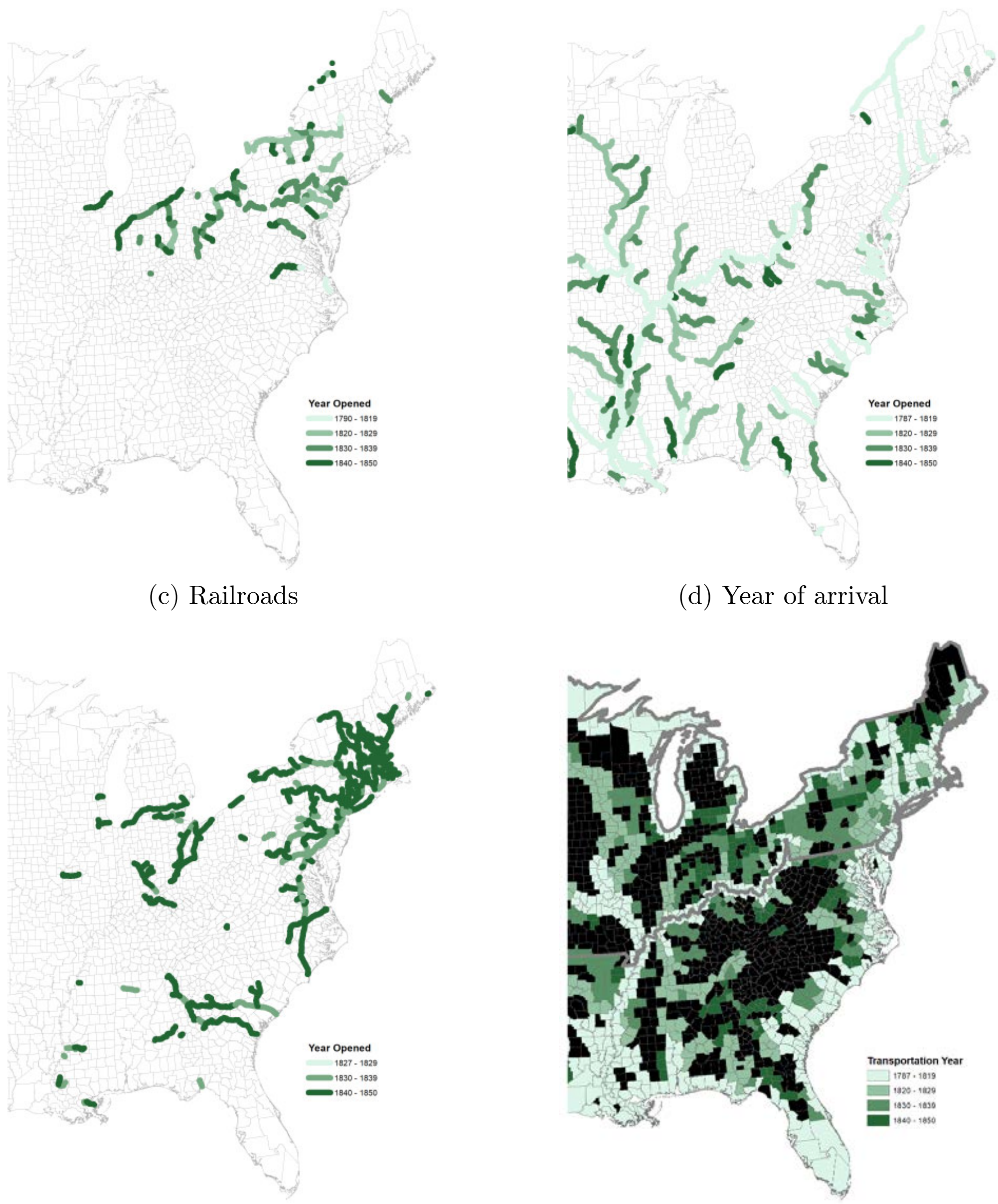

Figure 1: Spread of transportation infrastructure

Note: Always-navigable rivers, lakes, and oceans are assigned an opening date of 1787. Panel (d) shows the earliest year of arrival of one of these transportation modes for arrivals by 1850.

Source: Atack (2015, 2016, 2017)

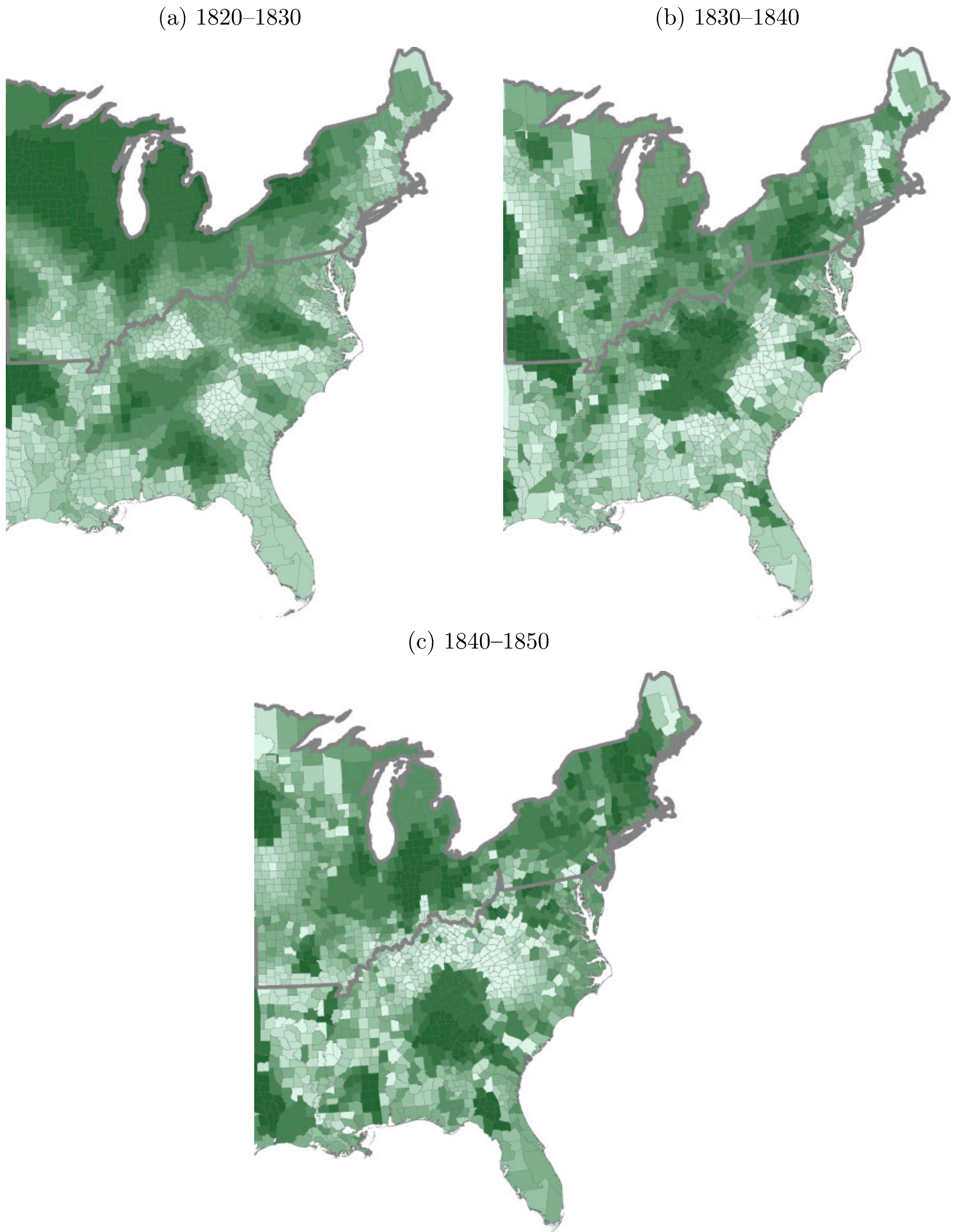

Figure 2: Changes in market access by decade.

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.

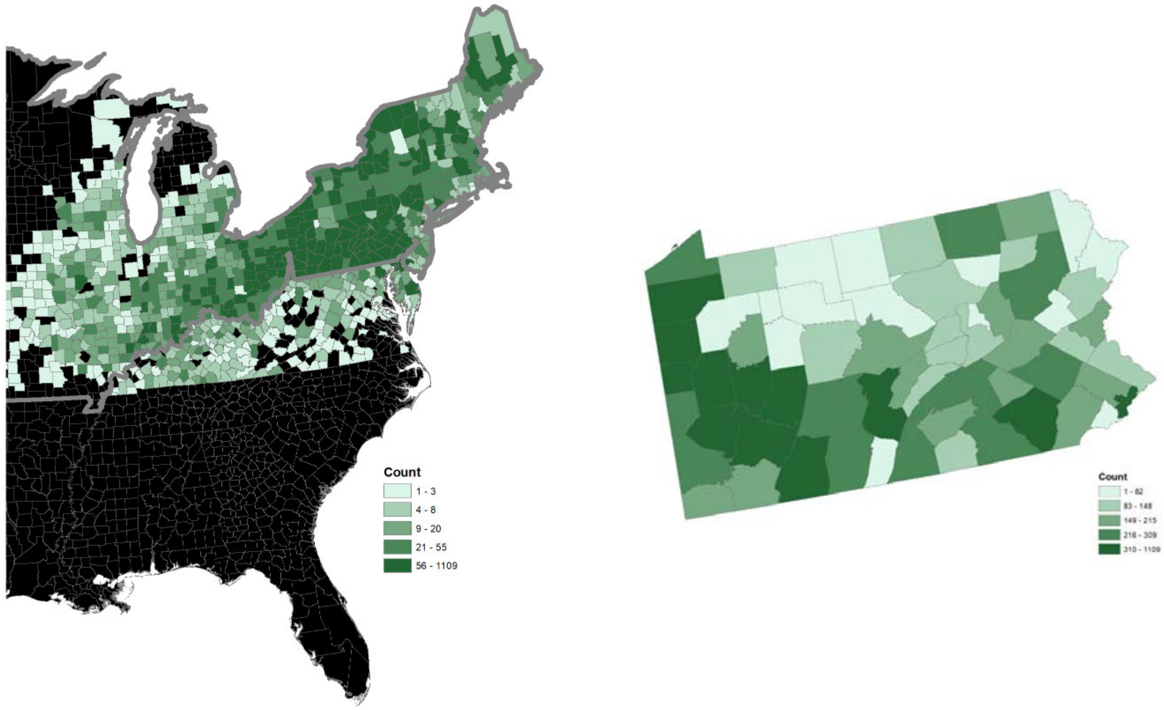

Figure 3: Number of observations by county

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.