Online Appendix Materials for

Transportation and Health in the Antebellum United States 1820–1847

Journal of Economic History

Ariell Zimran

Vanderbilt University and NBER

A Additional Tables and Figures

Table A.1: Regressions of 1850 death rates

Variables	(1)	(2)	(3)	(4)
Panel A: Weighted by 1850 population				
log(Market Access), 1820 Pop.	0.684***	0.544***	0.330**	0.266*
	(0.082)	(0.094)	(0.145)	(0.150)
Initial Market Access		0.095***		0.083
		(0.032)		(0.054)
Observations	790	790	472	472
R-squared	0.239	0.248	0.469	0.472
Panel B: Unweighted				
log(Market Access), 1820 Pop.	0.439***	0.356***	0.417***	0.376**
	(0.076)	(0.080)	(0.137)	(0.146)
Initial Market Access		0.122***		0.047
		(0.039)		(0.058)
Observations	790	790	472	472
R-squared	0.300	0.309	0.335	0.336
Controls	No	No	Yes	Yes

Significance levels: *** p<0.01, ** p<0.05, * p<0.1
Notes: Dependent variable is log of deaths per 1850 population. Sample includes counties in the Northeast or Midwest with no urban population in 1820. All specifications include state fixed effects and an indicator for 1850 transport linkage. Robust standard errors in parentheses.

Table A.2: Results with year-specific functions of latitude and longitude

Variables	(1)	(2)	(3)	(4)	(5)
log(Market Access), 1820 Pop.	-0.452^{**}	-0.373	-0.372	-0.375^{*}	-0.489^{*}
	(0.196)	(0.233)	(0.229)	(0.219)	(0.262)
Observations	23,557	25,556	25,556	25,556	25,556
R-squared	0.117	0.138	0.138	0.138	0.138
State FE	Yes	No	No	No	No
County FE	No	Yes	Yes	Yes	Yes
Controls	Yes	No	No	No	No
Transport	No	No	No	No	Yes
Radius			40 miles	100 miles	

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 fixed effects, and year-specific functions of latitude and longitude. Standard errors clustered at the county level. Observations weighted to correct for oversampling. Radius indicates the exclusion of counties within the listed distance in calculating market access.

Table A.3: County fixed effects regressions controlling for transport linkage by mode

Variables	(1)	(2)	(3)	(4)
log(Market Access), 1820 Pop.	-0.708^{***}	-0.824^{**}	-0.618^{***}	-0.497^{**}
	(0.220)	(0.320)	(0.216)	(0.242)
Observations	25,556	25,556	25,556	25,556
R-squared	0.127	0.174	0.129	0.157
Birth Year × Region FE	No	No	Yes	No
Birth Year × State FE	No	No	No	Yes
County × Decade FE	No	Yes	No	No

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

Table A.4: Regressions of improved acreage

Variable	(1) Acreage	(2) Acreage	(3) Acreage
log(Market Access), 1820 Pop.	0.201**	0.121	0.082
	(0.095)	(0.093)	(0.094)
log(MA) × log(Wheat Suit.)		0.943***	
		(0.182)	
log(MA) × log(Corn Suit.)			0.820***
			(0.145)
Observations	1,427	1,427	1,427
R-squared	0.955	0.961	0.962

Significance levels: *** p<0.01, ** p<0.05, * p<0.1
Notes: Dependent variable is the ratio of improved acres to total acreage. Sample includes all county-years with borders fixed to 1860, with no urban population in 1820, and in the Midwest or Northeast. All specifications include year and county fixed effects as well as an indicator for transport linkage. Observations weighted by the ratio of county population in each year to total population in that year. Standard errors clustered at the county level.

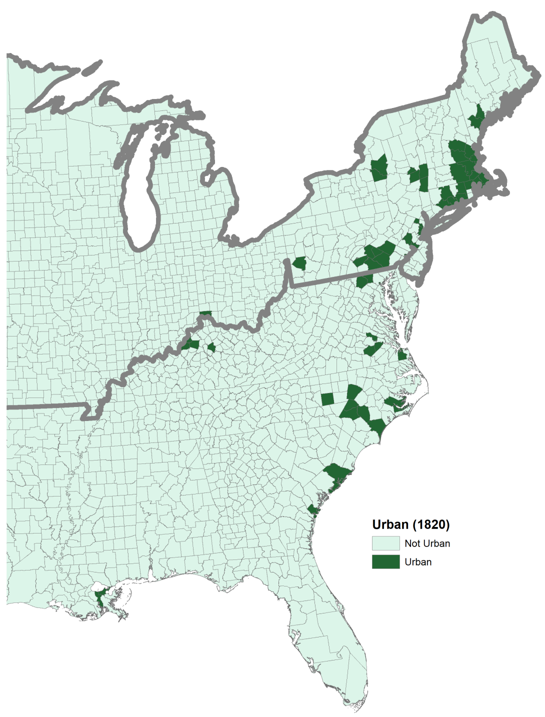

Figure A.1: Urbanization in 1820

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

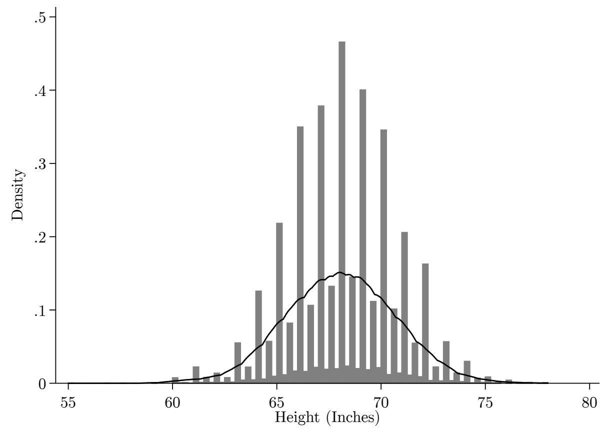

Figure A.2: Distribution of heights in the original data

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

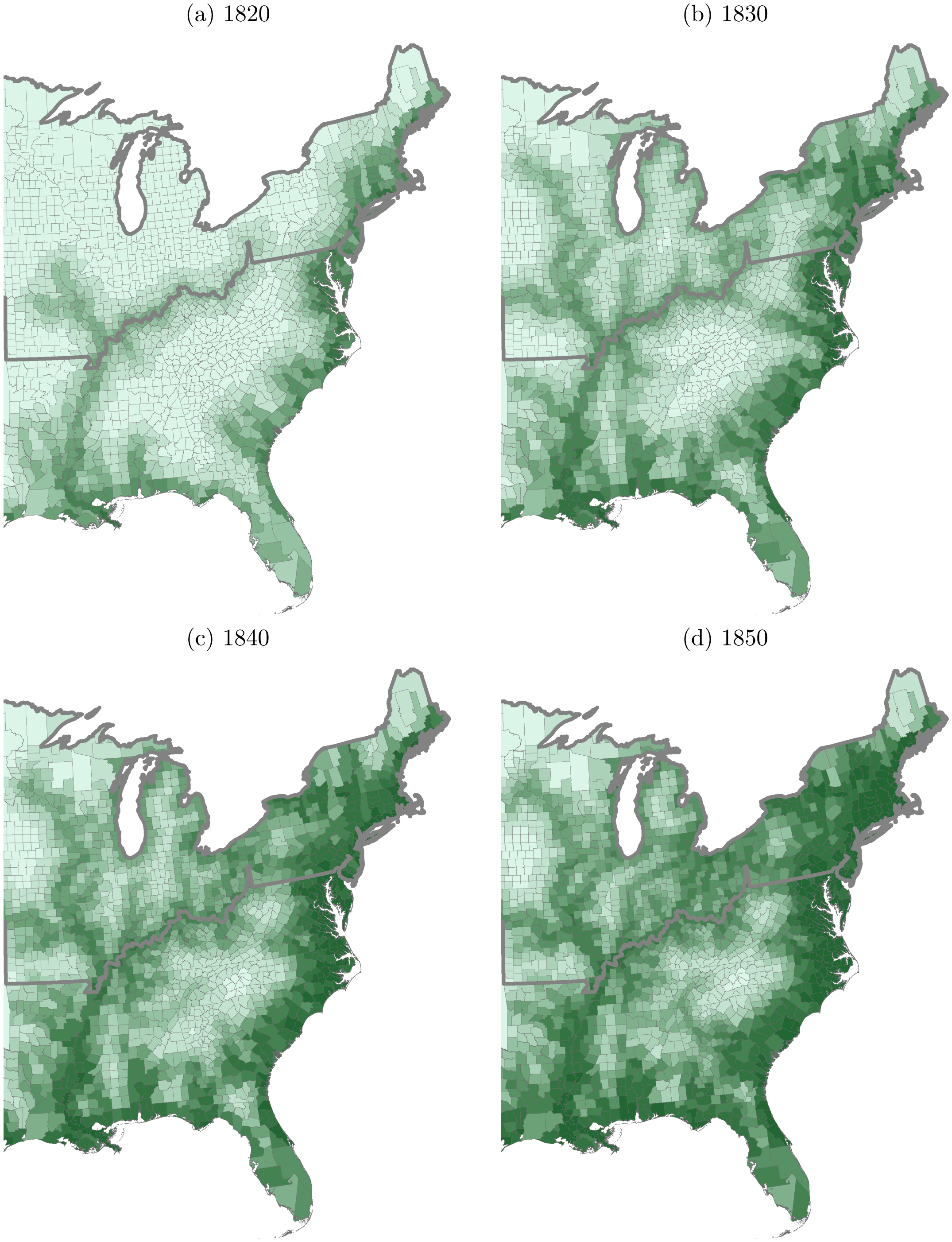

Figure A.3: Market access by decade.

Note: Each panel divides counties into deciles of market access for that year, with darker counties having greater market access. The scale is comparable across years, and is based on deciles of the market access measure in 1850. Sample region indicated by thick boundary.

Figure A.4: Number of individual height observations by birth cohort

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

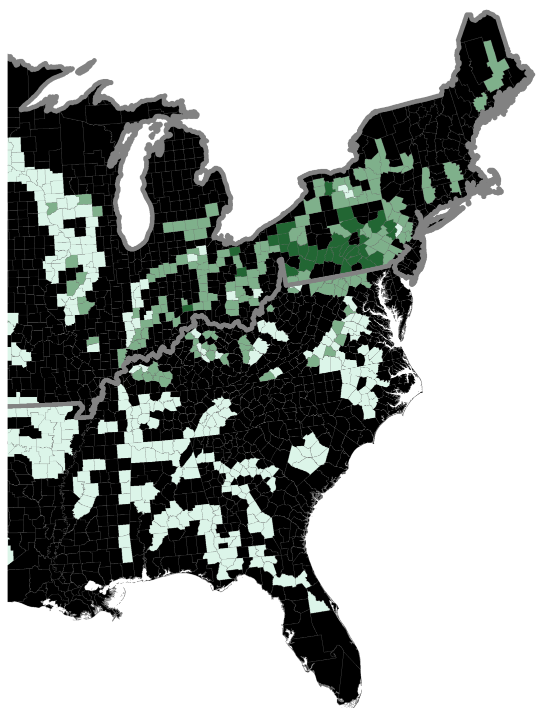

Figure A.5: Changes in linkage

Note: This Figure marks counties experiencing a change in transport linkage in 1820–1847. Counties in black experienced no change in transportation linkage between 1820 and 1847. The lightest colored counties experienced a change in transportation linkage in this period, but have no observations either before or after the change. The darker counties have observations both before and after the transportation change, but only the darkest (non-black) counties have at least 25 observations both before and after the change. Sample region indicated by thick boundary.

B Life-Cycle Effects of Transportation

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

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

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

and

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

where $\bar{X}_{jt}^{[a,b]}$ is the average of the logarithm of market access in county j over ages a to b of the cohort born in year t,[1] and the other notation is as in the main text. These are analogs of equations (2) and (3), respectively, with the substitution of the four explanatory variables of interest for the one. The four divisions in equations (B.1) and (B.2) are intended to denote the period before conception, the *in utero* and infancy period, childhood, and adolescence, respectively.

Results of estimation of equations (B.1) and (B.2) are presented in Figure B.1. The results of four specifications are presented—with and without county fixed effects and with and without region-specific birth year fixed effects. The results of all four specifications are similar. In all four cases, estimates of β_1, β_3, and β_4—that is, of the coefficients on market access in years other than infancy and the *in utero* period—are statistically insignificant and in most cases effectively zero.[2]

[1] I compute market access for years 1810–1860. This implies that for the later cohorts, $\bar{X}_{jt}^{[12,18]}$ may be the average of a shorter span of ages. For instance, for the 1847 cohort, it is only the average for ages 12 and 13. The lower frequency of rail information in the 1850s (every 2 years instead of annually) also reduces its accuracy.

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

Only β_2—the coefficient on the average of market access for ages -1–3—is consistently of one sign and nearly statistically significant. This result suggests that the previous analyses' focus on market access in the year of birth did not overlook important effects.

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

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

$$h_{ijt} = \gamma_t + \delta_a + \delta_e + \beta E_{jt} + \varepsilon_{ijt}, \tag{B.3}$$

where E_{jt} is the exposure time to transportation of cohort t born in county j and all other notation is as above. Exposure time is computed as follows: for those who are born into a county that is already transportation-linked, the exposure time is set to 23 years; for individuals for whom transportation arrives at age a, I set $E_{jt} = \max\{0, 23 - a\}$. I estimate equation (B.3) by OLS in columns (1)–(4) and with county fixed effects in columns (5)–(7). In no case is a statistically significant relationship between stature and exposure time present, and in all cases the estimates are small.

Table B.1: Exposure time regressions

Variables	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) OLS	(6) FE	(7) FE	(8) FE	(9) FE
Exposure Time	−0.007	0.002	0.004	0.005	0.001	0.022	−0.005	0.026	−0.004
	(0.005)	(0.005)	(0.006)	(0.006)	(0.006)	(0.018)	(0.029)	(0.019)	(0.020)
Observations	13,800	13,800	12,716	12,716	12,716	13,800	13,800	13,800	13,800
R-squared	0.048	0.061	0.066	0.071	0.119	0.128	0.177	0.132	0.179
State FE	No	Yes	Yes	Yes	Yes	No	No	No	No
Controls	No	No	Yes	Yes	Yes	No	No	No	No
Birth Year × Region FE	No	No	No	Yes	No	No	No	Yes	No
Birth Year × State FE	No	No	No	No	Yes	No	No	No	Yes
County × Decade FE	No	No	No	No	No	No	Yes	No	No

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

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

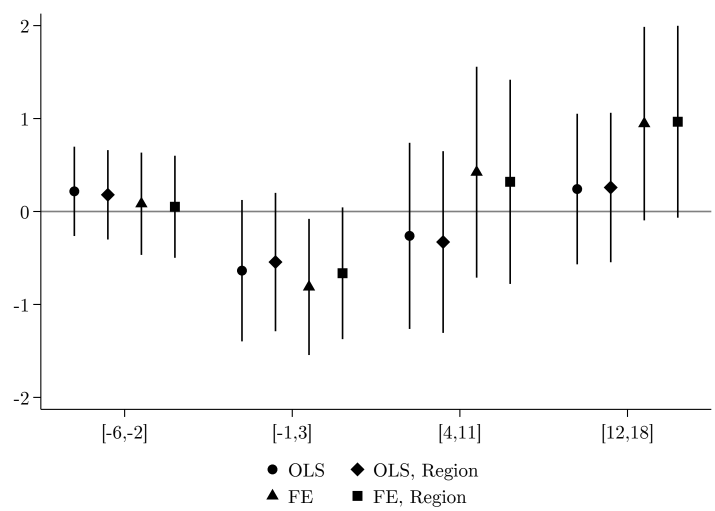

Figure B.1: Market access and stature over the life cycle

Note: This Figure presents coefficients and 95% confidence intervals from estimates of equations (B.1) and (B.2). The x-axis indicates the range of ages over which market access is averaged. The estimates marked "OLS" and "OLS, Region" are estimates of equation (B.1) without and with region-specific measurement age and birth year fixed effects, respectively. Estimates marked "FE" and "FE, Region" are estimates of equation (B.2) without and with region-specific measurement age and birth year fixed effects, respectively.

C Results with Truncated Regression

This appendix tests the robustness of the main results to the use of truncated regression. The Union Army was subject to a minimum height requirement, though this appears to not have been strictly enforced (Figure A.2 in Online Appendix A). Nonetheless, it is possible to test whether implementing the literature-standard method of addressing shortfall in height regressions (Komlos 2004) affects the results. I do so with a truncation point of 64 inches (A'Hearn 1998; Komlos 1998; Zimran 2019), which requires omitting any observation with height less than 64 inches. I present results in Tables C.1–C.3. In terms of sign and statistical significance, the results are similar to those presented in the main text, though the magnitudes of the estimated coefficients are slightly smaller.

Table C.1: County fixed effects regressions

Variables	(1)	(2)	(3)	(4)	(5)
Transport	−0.088				
	(0.108)				
log(Market Access), 1820 Pop.		−0.529**	−0.723**	−0.475**	−0.295
		(0.221)	(0.284)	(0.214)	(0.220)
Observations	23,786	23,786	23,786	23,786	23,786
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. All specifications estimated by truncated regression with a lower truncation point of 64 inches. Sample includes individuals at least 64 inches tall who were 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 C.2: County fixed effects regressions controlling for transport linkage

Variables	(1)	(2)	(3)	(4)
log(Market Access), 1820 Pop.	-0.572^{**}	-0.740^{**}	-0.493^{**}	-0.311
	(0.241)	(0.295)	(0.232)	(0.243)
Observations	23,786	23,786	23,786	23,786
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. All specifications estimated by truncated regression with a lower truncation point of 64 inches. Sample includes individuals at least 64 inches tall who were 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 C.3: The local development mechanism

Variable	(1) Dens.	(2) Dens.	(3) Dens.	(4) Height	(5) Height
log(Market Access), 1820 Pop.	0.617^{***}	0.631^{***}	0.612^{***}	-0.505^{**}	-0.510^{**}
	(0.157)	(0.159)	(0.160)	(0.228)	(0.228)
log(MA) × log(Wheat Suit.)		0.626^{**}		-0.413	
		(0.278)		(0.510)	
log(MA) × log(Corn Suit.)			0.481^{**}		-0.296
			(0.220)		(0.455)
Observations	1,122	1,122	1,122	23,786	23,786

Significance levels: *** p<0.01, ** p<0.05, * p<0.1
Notes: Dependent variable listed in the column header. Specifications in columns (4) and (5) estimated by truncated regression with a lower truncation point of 64 inches. 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 at least 64 inches tall who were 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.

D Market Access Computation Algorithm

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

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

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

3. Transshipment links are created between modes of transportation.

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

5. An origin-destination cost matrix calculation is performed. This entails computing the minimum transport cost c_{jkt} between each county pair jk that is possible given the transportation network in year t.

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

$$\tau_{jkt} = 1 + \frac{c_{jkt}}{35}.$$

Table D.1: Transportation costs

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

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

E Results with Limited Samples

This appendix checks whether the main results are robust to limiting the sample to counties where height data are relatively more plentiful. These limitations address a potential concern that small samples of height data in some counties may have led to unrepresentativeness of true heights in these counties, and thus to spurious results. Specifically, Tables E.1–E.3 repeat the main results of the paper with two different sample limitations. The first limitation restricts the sample to New York and Pennsylvania, where the data are relatively more abundant, particularly given the over-sampling of Pennsylvania. The second limitation restricts the sample to counties with at least 100 observations of individual height. In both cases, there is no indication that the sparsity of height data in some counties is responsible for results.

There are two limitations to this robustness. The first is that the results of Panel A of Table E.1 generally do not enable the rejection of the null hypothesis of no effect of market access on stature, likely because the number of counties (on which standard errors are clustered) is severely limited by this restriction. Nonetheless, implementing Donaldson and Hornbeck's (2016) and Hornbeck and Rotemberg's (2019) identification strategy in Table E.2 restores the statistical significance of the results. The second is that in Panel A of Table E.3, the interaction with wheat and corn suitability in columns (4) and (5) does not have the same result as in the main results or as in Panel B; however, given that this applies also in columns (1)–(3), where sample size of heights is not an issue, this is likely a result of the relative lack of variation in these factors when restricting the sample to New York and Pennsylvania. Indeed, in the case of Table E.3, the estimates of the interactions of market access with suitability measures in columns (2) and (3) are too imprecise to draw any conclusion regarding heterogeneous effects of market access on population density in the limited sample of counties, likely due in part to the small number of observations in this analysis. However, since the concern that this sample limitation is intended to address is the lack of representativeness induced by small height samples in some counties, it is the results in columns (4) and (5) of Table E.3 that are of particular interest; the results in columns (1)–(3) of Table 7 are not in danger of being driven by potentially small samples of height in some counties.

Table E.1: County fixed effects regressions

Variables	(1)	(2)	(3)	(4)	(5)
Panel A: NY and PA Only					
Transport	−0.015				
	(0.159)				
log(Market Access), 1820 Pop.		−0.505*	−0.607	−0.505*	−0.478
		(0.271)	(0.369)	(0.271)	(0.339)
Observations	13,886	13,886	13,886	13,886	13,886
R-squared	0.092	0.093	0.123	0.093	0.095
Panel B: Counties with at least 100 observations					
Transport	−0.178				
	(0.136)				
log(Market Access), 1820 Pop.		−0.855***	−1.257***	−0.961***	−0.875***
		(0.279)	(0.335)	(0.274)	(0.304)
Observations	14,053	14,053	14,053	14,053	14,053
R-squared	0.099	0.099	0.116	0.106	0.118
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 E.2: County fixed effects regressions controlling for transport linkage

Variables	(1)	(2)	(3)	(4)
Panel A: NY and PA Only				
log(Market Access), 1820 Pop.	-0.586^{**}	-0.682^{*}	-0.586^{**}	-0.621^{*}
	(0.276)	(0.389)	(0.276)	(0.331)
Observations	13,886	13,886	13,886	13,886
R-squared	0.093	0.123	0.093	0.095
Panel B: Counties with at least 100 observations				
log(Market Access), 1820 Pop.	-0.837^{***}	-1.241^{***}	-0.962^{***}	-0.879^{**}
	(0.274)	(0.340)	(0.276)	(0.333)
Observations	14,053	14,053	14,053	14,053
R-squared	0.099	0.116	0.106	0.118
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 E.3: The local development mechanism

Variable	(1) Dens.	(2) Dens.	(3) Dens.	(4) Height	(5) Height
Panel A: NY and PA Only					
log(Market Access), 1820 Pop.	0.625***	0.611***	0.617***	−0.612**	−0.592**
	(0.164)	(0.162)	(0.165)	(0.267)	(0.272)
log(MA) × log(Wheat Suit.)		−0.394		0.523	
		(0.594)		(1.277)	
log(MA) × log(Corn Suit.)			−0.140		0.201
			(0.494)		(1.001)
Observations	265	265	265	13,886	13,886
R-squared	0.943	0.944	0.943	0.093	0.093
Panel B: Counties with at least 100 observations					
log(Market Access), 1820 Pop.	0.648***	0.640***	0.642***	−0.881***	−0.866***
	(0.185)	(0.185)	(0.188)	(0.261)	(0.254)
log(MA) × log(Wheat Suit.)		−0.278		−1.412**	
		(0.404)		(0.583)	
log(MA) × log(Corn Suit.)			−0.139		−1.314**
			(0.366)		(0.527)
Observations	158	158	158	14,053	14,053
R-squared	0.941	0.942	0.941	0.100	0.100

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 enlistment year and measurement age fixed effects. Observations in columns (1)–(3) weighted by the ratio of county population in each year to total population in that year. Observations in columns (4) and (5) weighted to correct for oversampling. Standard errors clustered at the county level.

F Results with Two-Way Clustered Standard Errors

This appendix repeats the main results with standard errors clustered on the county and birth year level rather than only on the county level, presenting results in Tables F.1–F.3. The results in this appendix are slightly differences in the estimates and sample sizes because of the omission of counties with a single observation from the results.[3] Even with the adjusted standard errors, the statistical significance of the results is largely unchanged, though naturally the estimates are less precise.

Table F.1: County fixed effects regressions

Variables	(1)	(2)	(3)	(4)	(5)
Transport	−0.042				
	(0.109)				
log(Market Access), 1820 Pop.		−0.597***	−0.744***	−0.517**	−0.386
		(0.194)	(0.252)	(0.193)	(0.264)
Observations	25,492	25,492	25,330	25,492	25,492
R-squared	0.121	0.121	0.158	0.124	0.152
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 and the birth cohort levels. Observations weighted to correct for oversampling.

[3] This is necessitated by the use of the `reghdfe` (Correia 2017) command to compute these standard errors.

Table F.2: 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^{*}
	(0.206)	(0.235)	(0.209)	(0.270)
Observations	25,492	25,330	25,492	25,492
R-squared	0.122	0.158	0.124	0.152
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 and the birth cohort levels. Observations weighted to correct for oversampling.

Table F.3: The local development mechanism

Variable	(1) Dens.	(2) Dens.	(3) Dens.	(4) Height	(5) Height
log(Market Access), 1820 Pop.	0.617^{***}	0.629^{***}	0.610^{***}	-0.579^{***}	-0.579^{***}
	(0.131)	(0.133)	(0.136)	(0.190)	(0.176)
log(MA) × log(Wheat Suit.)		0.626^{**}		-0.842	
		(0.308)		(0.558)	
log(MA) × log(Corn Suit.)			0.481^{**}		-0.646
			(0.217)		(0.445)
Observations	900	900	900	25,492	25,492
R-squared	0.916	0.919	0.919	0.122	0.122

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 enlistment year and measurement age fixed effects. Observations in columns (1)–(3) weighted by the ratio of county population in each year to total population in that year. Observations in columns (4) and (5) weighted to correct for oversampling. Standard errors clustered at the county and year levels.

G Instrumental Variables Estimation

In this appendix, I develop an alternative identification strategy based on an instrument for market access that builds on the straight-line instruments commonly used in studying the economic impacts of transportation improvements (e.g., Atack et al. 2010; Banerjee, Duflo, and Qian 2012; Ghani, Goswami, and Kerr 2016; Hornung 2015). It is based on the principle (Taylor 1951, p. 37) that antebellum internal improvements were intended to link major watersheds (the Atlantic, Great Lakes, and Mississippi) to one another (easing interregional trade, as the Erie Canal did by linking the Atlantic and Great Lakes watersheds) and to major cities (as was the target of the various canals built from the east coast after the Erie).[4]

Specifically, I draw a series of straight lines, depicted in Figure G.1. The first set of lines, depicted in panel G.1(a), are the shortest connections between the major watersheds, based on the steamboat-navigability of rivers in 1820.[5] The next set of lines, depicted in panels G.1(b)–G.1(d), identifies the 25 largest cities over 10,000 population in each census year 1820–1840 (though it was not until 1840 that there were at least 25 such cities) and draws the shortest lines between these cities and the three major watersheds (Atlantic, Great Lakes, and Mississippi), provided that these lines are not more than 300 miles in length nor originate in the South (except for Virginia, Maryland or Washington, DC).[6] The repetition of lines between panels G.1(b), G.1(c), and G.1(d) is not concerning, as the construction of a second line overlapping a first will have no impact.

I then compute market access as described in the main text and in Online Appendix D, with the following changes: (1) I begin with the transportation network in 1820; (2) I treat the lines of Figure G.1 as canals; (3) I augment the 1820 network by letting each line develop—beginning in 1820 for the lines in panel (a) of Figure G.1 and from the decadal year for those in other panels—

[4] As Taylor (1951, p. 37) explains, "three types of major canals were built: (1) those designed to improve transportation between the upcoming and tidewater in states bordering the Atlantic from Maine to Virginia; (2) those, like the Erie, designed to link the Atlantic states with the Ohio River Valley; and (3) those in the West which were planned to connect the Ohio-Mississippi system with the Great Lakes."

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

[6] This is a simple way to avoid lines completely outside of practicality (the distance restriction) or actuality (the regional restriction). The definition of cities is actually based on the urban population of counties, rather than city populations. Southern cities are excluded to better capture the true lack of internal improvements there.

over a period of 15 years in equal increments, beginning at the originating city or at the easternmost watershed. An example of the evolution of one such line is shown in Figure G.2.[7] This alternative measure of market access is the instrumental variable, which I use to estimate equations (2) and (3) by instrumental variables.

Relevance of the instrument will be formally established in estimation of the first-stage equations but is already suggested by Figure G.1. This Figure (and comparison to Figure 1) reveals that the location of these lines is a good approximation of actual construction. For instance, the line linking the Atlantic and Great Lakes watersheds in panel G.1(a) is close to the actual location of the Erie Canal; the lines in Pennsylvania in panel G.1(b) closely approximate the construction of Pennsylvania's Main Line; and the lines in Ohio, Indiana, and Illinois in panels G.1(a), G.1(c), and G.1(d) are also close approximations to actual construction. Because these lines are used to compute an alternative measure of market access, they also affect counties away from where they are constructed, as the Erie Canal did. Moreover, as shown in Figure G.3, the temporal development of the market access implied by the instrument tracks well with that of the actual measure.[8]

Excludability of the instrument requires the following assumptions. In the cross-section, the identification assumption is comparable to that of other straight-line instruments. It is that, after excluding counties from which the lines in panels G.1(b)–G.1(d) originate, counties on or near the straight lines of Figure G.1 are similar to those further from the lines except in their likelihood to receive beneficial surges in market access. The identification assumption in the second dimension—the time series—has fewer analogs in the literature. It is that counties closer to the origin of a straight line in Figure G.1 are not fundamentally different from those further from the origins, except that they are likely to be linked to the transportation network sooner. A clear concern is that

[7] I have also used a 10 year development period, but the variable generated in this way does not satisfy the relevance condition for instrumental variables, whereas the variable generated with a 15 year development period does. Although the evolution of the straight lines is based on a fixed annual expansion, the instrument is not a time trend (indeed, year-specific indicators are included in all specifications). Instead, the instrument, like the measure of market access, evolves discontinuously in response to a new transport link.

[8] An example of the evolution of the instrument and of market access in a single county in shown in Figure G.4, which describes the experience of Montgomery County, Ohio. The rapid increases in market access in the 1820s come from the construction of the Miami and Erie Canal, which passed through the county and linked it to the Ohio River. The rapid increase in the instrument in the 1830s comes from the passage of the straight line linking Hamilton County, Ohio to the Great Lakes through the county linking it to the Ohio River. The smaller increase in the 1840s comes from the completion of that line, completing the hypothetical linkage to Lake Erie.

the origins of the lines represent points of interest; but given the high costs of wagon transportation, excluding the terminus counties should render the remaining counties equally isolated.[9]

In Table G.1 I briefly explore the evidence in support of excludability of the instrument. In particular, I relate the characteristics of counties that are observed in 1820 to the lines of Figure G.1. Given the sparsity of data available in the early censuses, the only measures available are population density and agricultural suitability.

In column (1), I regress the logarithm of the wheat suitability measure of a county on an indicator for being on one of the lines presented in Figure G.1. This regression includes state fixed effects and the same functions of distance from New York and Cincinnati as described in the main text. The resulting coefficient is statistically insignificant and small, indicating that it is not possible to reject the null hypothesis that counties on the lines were ex ante the same as others. The regression in column (2) of corn suitability shows similar results. In both of these cases, even if the coefficients were of larger magnitude and statistically significant, the bias induced by the positive coefficients would tend to mute the negative relationships of the transport-health relationship that I have found. Construction targeting more potentially agriculturally productive areas would tend to be associated with greater average height if agricultural suitability supported better health. The regression in column (3) of the logarithm of population density on the same regressor (limiting the sample to counties that had achieved their 1860 boundaries by 1820) shows similar results.[10]

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

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

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

assumptions that counties on and off of the lines are ex ante similar, and that counties closer and farther from the origins of the line are ex ante similar.

I implement this strategy in Table G.2, which presents the coefficient from the estimation of equation (2) by instrumental variables with state-specific indicators and no other controls. The last row of this Table shows the coefficient on the instrument from the first-stage estimation—that is, the estimation of specification (2) with the logarithm of market access as the dependent variable and the logarithm of the instrumental variables-implied market access as the regressor of interest. It shows a positive and strongly statistically significant relationship between the instrument and the potentially endogenous regressor of interest, indicating that the instrument satisfies the relevance condition. This satisfaction of the relevance criterion remains robust throughout the various specifications in this Table.

The relationship between market access and health as estimated by this instrumental variables approach in column (1) is negative and statistically significant.[11] Its magnitude is comparable to the ordinary least squares estimate of Table 3 and to the fixed effects estimates of columns (2)–(5) of Table 4. Column (2) of Table G.2 adds the county-specific controls discussed in the main text. Unlike the ordinary least squares regressions of Table 3, the introduction of these controls increases rather than decreases the magnitude of the coefficient, which, at -0.763, remains negative and statistically significant, though less precisely estimated. Column (3) controls also for 1820 market access, as in columns (4)–(6) of Table G.1. This approach more effectively isolates changes over time in market access, rather than its level, which may be endogenous even after instrumentation because the instrument is based on the (potentially endogenous) 1820 network. Columns (4) and (5) add to the specification of column (2) region- and state-by-birth year indicators to the instrumental variables specification with controls. The negative and (marginally) statistically significant coefficient is robust to these controls.

Finally, column (6) estimates equation (3) by instrumental variables. The first stage estimate is strong, indicating that prior first-stage estimates are robust to the inclusion of county fixed

[11] The results of Table G.2 include individuals born in counties that have no urban population in 1820 but that are origin points of a line in panels (c) or (d) of Figure G.1. Omission of these individuals, who number 303, or 158 in birth years after the decadal year in which the line first appears, yields results that are virtually identical to those of Table G.2. I include them in the analysis of this appendix to maintain comparability with results in the main text.

effects. The second-stage coefficient of interest remains negative, and the magnitude is comparable to estimates of Tables 4 and G.2. However, the standard error of this coefficient is increased considerably by the demands of this estimation (relative to the non-instrumental variables analog in Table 4), making it impossible to reject the null hypothesis of no effect.

Table G.1: Correlates of instrumental variables line placement

Variables	(1) Wheat	(2) Corn	(3) Dens.	(4) Wheat	(5) Corn	(6) Dens.	(7) Wheat	(8) Corn	(9) Dens.
On IV Line	0.029	0.022	0.032						
	(0.019)	(0.026)	(0.213)						
log(IV Market Access) in 1850				0.016	−0.054	−0.447			
				(0.024)	(0.033)	(0.695)			
IV Line Year							0.001	0.005	−0.076
							(0.005)	(0.006)	(0.077)
Observations	942	941	87	941	940	87	119	119	35
R-squared	0.605	0.583	0.464	0.618	0.611	0.626	0.571	0.368	0.312

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

Table G.2: Instrumental variables regressions

Variables	(1)	(2)	(3)	(4)	(5)	(6)
log(Market Access), 1820 Pop.	-0.584^{***}	-0.763^{*}	-1.664^{**}	-0.675^{*}	-0.867^{*}	-0.418
	(0.202)	(0.395)	(0.663)	(0.394)	(0.483)	(0.537)
Observations	25,556	23,557	23,557	23,557	23,557	25,556
R-squared	0.077	0.079	0.077	0.081	0.108	0.058
State FE	Yes	Yes	Yes	Yes	Yes	No
Controls	No	Yes	Yes	Yes	Yes	No
Initial MA	No	No	Yes	No	No	No
Birth Year × Region FE	No	No	No	Yes	No	No
Birth Year × State FE	No	No	No	No	Yes	No
County FE	No	No	No	No	No	Yes
First Stage	0.399^{***}	0.265^{***}	0.200^{***}	0.270^{***}	0.238^{***}	0.357^{***}
	(0.033)	(0.027)	(0.027)	(0.027)	(0.027)	(0.033)

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

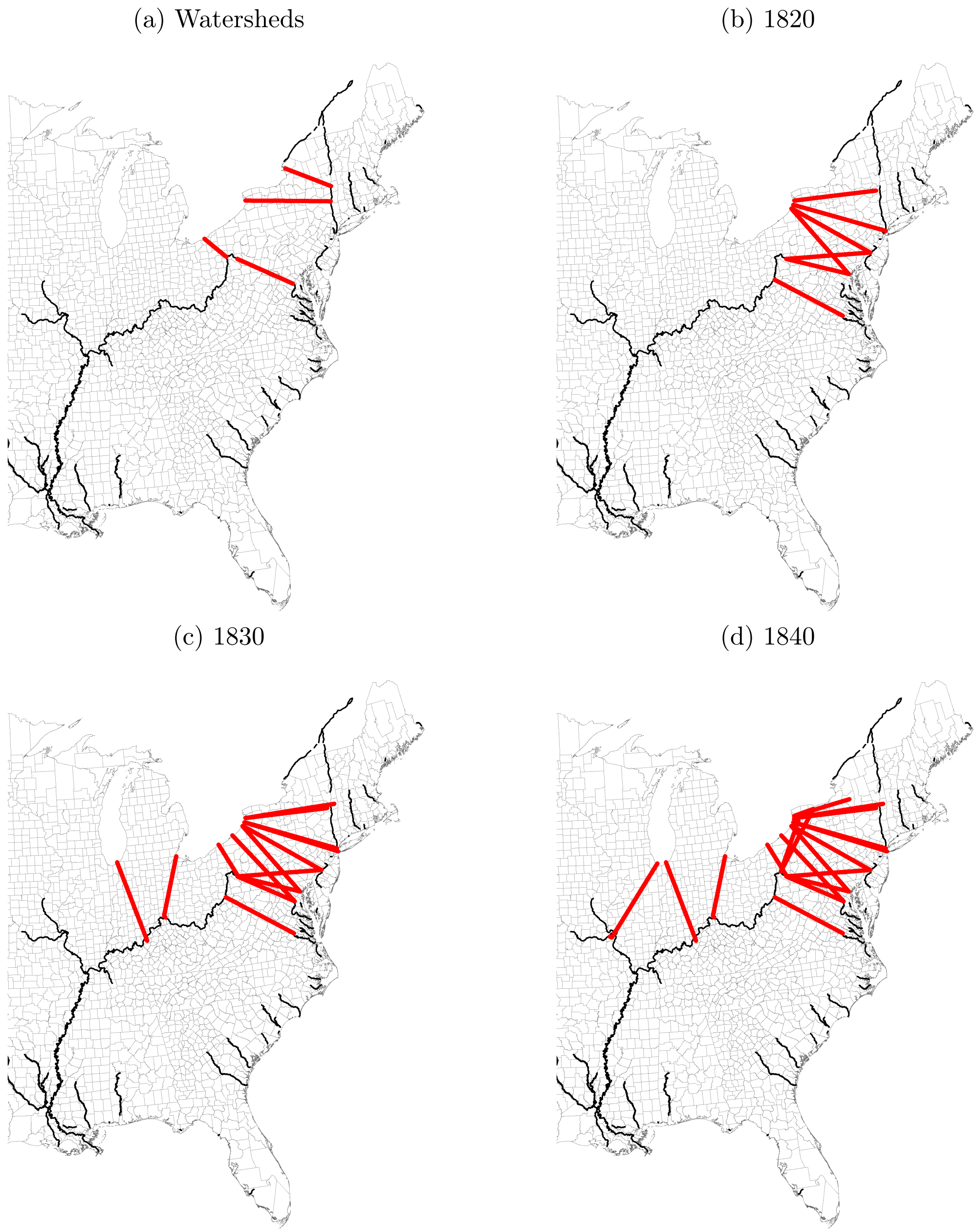

Figure G.1: Straight lines for instrumentation

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

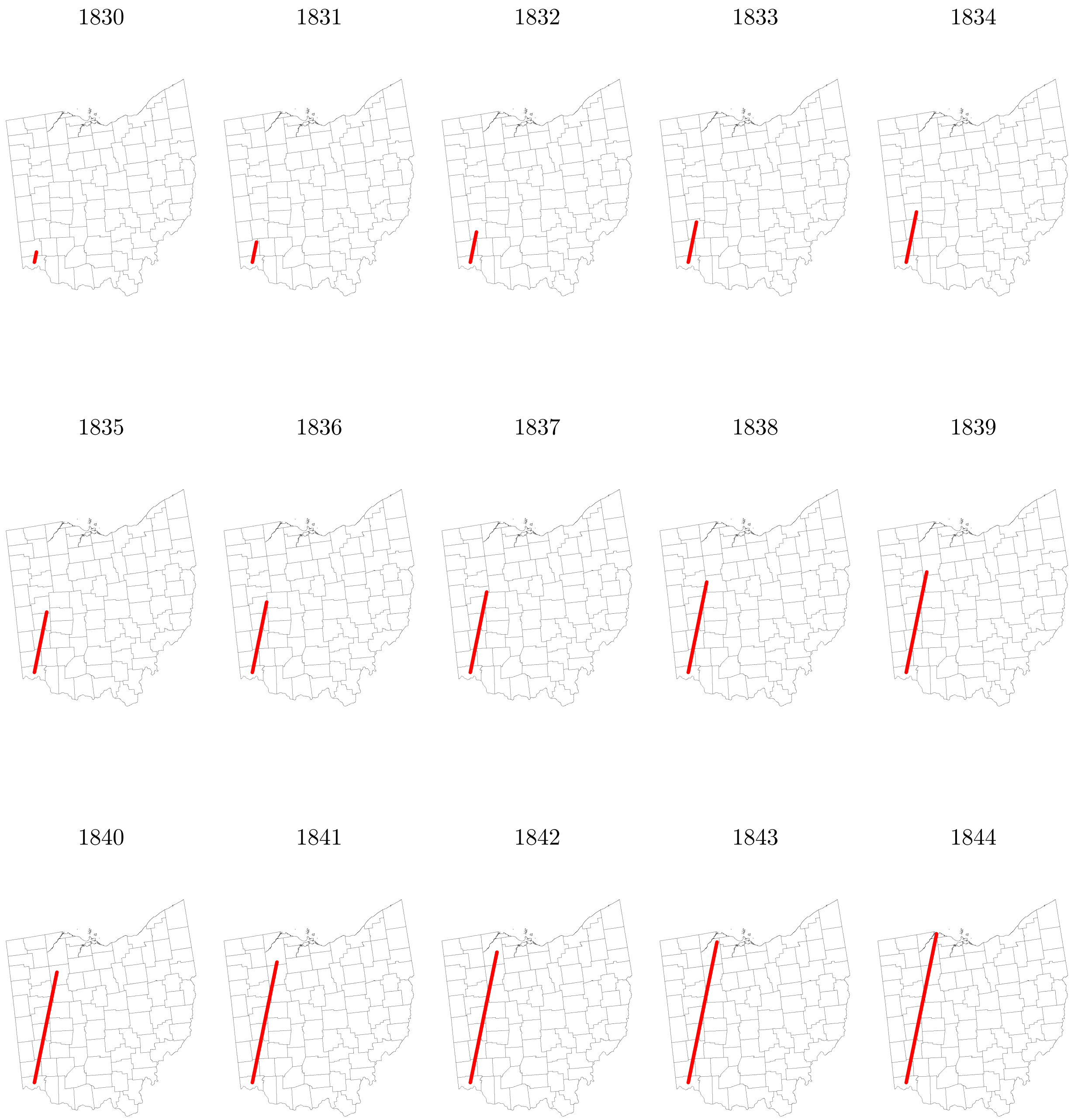

Figure G.2: Evolution of the instrument line linking Hamilton County, Ohio to the Great Lakes

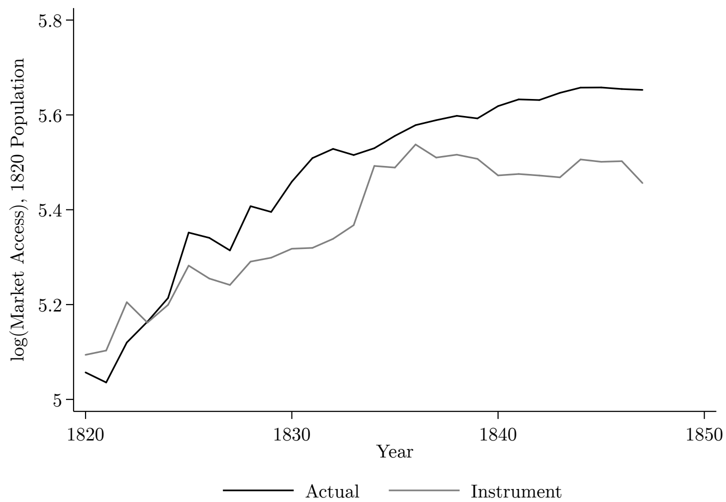

Figure G.3: Actual and hypothetical market access, whole sample of individuals

Note: The line labeled "Actual" plots the average log market access. The line labeled "Instrument" plots instrument calculated using the straight lines of Figure G.1.

Figure G.4: Actual and hypothetical market access, Montgomery County, Ohio

Note: The line labeled "Actual" plots the average log market access. The line labeled "Instrument" plots instrument calculated using the straight lines of Figure G.1.

H Migration Responses to Transportation

In this appendix, I determine whether the deleterious effect of transportation linkage in an individual's county-year of birth on terminal height may be the product of migration responses to the construction of new transportation linkages other than those that might contribute to rising population density. These responses do not jeopardize the interpretation of the results as being the effect of market access on height; but they would affect their interpretation as potentially explaining the nationwide decline in average stature.

The first mechanism that I seek to rule out is that transportation linkages caused individuals from an affected county-year of birth to be more likely to migrate to other, potentially less healthy, areas than the place of birth, and that their health was harmed by residence in this new location rather than some impact of the new transportation on the place of birth. Table H.1 addresses this concern by testing whether increased market access caused individuals to be more likely to migrate, using a version of equation (4). Four outcomes are considered in this Table, and all are based on whether an individual enlisted in a different state than the state of birth, or in a different county than the county of birth. As no other migration information are available, it must be kept in mind that enlistment in a different county than the county of birth does not necessarily imply migration (individuals may have simply enlisted in a nearby county). Regardless of the measure of migration used, the results of Table H.1 largely support the view that individuals whose county-year of birth had greater market access were less likely to migrate, rather than more likely. This is the opposite result from that which would be consistent with the concern that market access caused individuals to move to less healthy areas and become shorter for that reason, rather than to experience worse health in the place of birth.

Another concern is that rising market access in a county might have attracted less healthy individuals to move there and have shorter children than would have been born to individuals who lived in the county prior to the transport linkage. The theoretical basis for this concern is somewhat weak—although Ferrie (1997) and Stewart (2006) find evidence of negative selection into migration to the frontier in the 1850s and 1860s on the basis of unobservables associated with wealth accumulation, Logan (2009) finds positive selection into internal migration on the basis of health.

Nonetheless, the concern merits consideration. The data at hand do not enable a test of whether immigrants to a newly linked county might have been negatively selected on health (the health of their children, not their own health, is observed). However, it is possible to use the migration of individuals in the data to test whether the observed patterns of migration are consistent with the migration of sicker types to newly linked areas, and I present such a test in Table H.2.[12]

Panel A of Table H.2 estimates equation (4) using only the sample of individuals for which migration status can be determined. The results are broadly similar to those of Table 6. Panels B and C of Table H.2 test whether individuals who moved states (Panel B) or counties (Panel C) tended to be shorter than those who did not. This is precisely the mechanism that would have had to operate among the parents of the individuals in the sample in order for the results to be driven by the migration of sicker types to newly linked regions. Specifically, I add an indicator for having moved to the regressions with county fixed effects. The coefficient on this indicator shows whether individuals who moved tended to be taller or shorter, conditional on county fixed effects, enlistment-year fixed effects, year-of-birth fixed effects, and the market access of their county-year of birth. The results show that, if anything, these individuals tended to be taller, rather than shorter, meaning that it was the healthier types, rather than the sicker types, who migrated. These controls do not have an appreciable effect on the coefficient on market access.

Panel D of Table H.2 adds to the baseline specification the log of market access in 1860 of the individual's county of enlistment (still using 1820 population to calculate market access). If sicker types were likely to move to newly linked areas, then we would expect to find a negative relationship between height and the connectedness of the destination county. It should be noted, however, that such a negative correlation would also arise if migration during childhood or adolescence to a more connected county itself had an effect on health. The regressions of Panel D of Table H.2 (which also add state-of-enlistment fixed effects) do find a small and statistically insignificant negative relationship between the market access of the county of enlistment and height. While this negative coefficient may be an indicator of migration of sicker types to more connected regions, the fact that the coefficient is small and statistically insignificant, the fact that the effect of birth county market

[12]That is, we would like to observe migration by the parents of sample individuals, but can observe it only for the individuals themselves.

access remains strong, and the fact that this result may simply indicate an effect of destination county market access on health, leads me to conclude that the evidence supporting the migration of sicker types to newly linked regions being solely responsible for the documented negative effect of transportation on health is limited at best.

Table H.1: Migration regressions

Variables	(1)	(2)	(3)	(4)
Panel A: State-level migration indicator				
log(Market Access), 1820 Pop.	−0.044	0.111*	−0.038	−0.007
	(0.047)	(0.064)	(0.047)	(0.054)
Observations	25,556	25,556	25,556	25,556
R-squared	0.235	0.287	0.237	0.260
Panel B: County-level migration indicator				
log(Market Access), 1820 Pop.	−0.072	0.040	−0.072	−0.017
	(0.045)	(0.058)	(0.046)	(0.045)
Observations	21,138	21,138	21,138	21,138
R-squared	0.361	0.393	0.363	0.385
Panel C: Migration to denser county indicator				
log(Market Access), 1820 Pop.	−0.074	−0.006	−0.080*	−0.065
	(0.049)	(0.054)	(0.046)	(0.045)
Observations	21,138	21,138	21,138	21,138
R-squared	0.323	0.371	0.325	0.347
Panel D: log(Population density) difference of enlistment and birth county				
log(Market Access), 1820 Pop.	−0.259*	−0.072	−0.271**	−0.291**
	(0.135)	(0.142)	(0.124)	(0.132)
Observations	21,131	21,131	21,131	21,131
R-squared	0.232	0.297	0.235	0.266
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 indicated in the panel header. Sample includes individuals born in the Northeast or Midwest in counties with no urban population in 1820. All specifications include birth year, enlistment year, measurement age, and county fixed effects, as well as an indicator for transport linkage in the birth year.

Table H.2: Results conditional on migration status

Variables	(1)	(2)	(3)	(4)
Panel A: Results with limited sample				
log(Market Access), 1820 Pop.	-0.628^{***}	-0.715^{**}	-0.560^{**}	-0.483^{*}
	(0.238)	(0.351)	(0.234)	(0.267)
Observations	21,138	21,138	21,138	21,138
R-squared	0.135	0.186	0.137	0.167
Panel B: State-level migration				
log(Market Access), 1820 Pop.	-0.619^{***}	-0.747^{**}	-0.552^{**}	-0.488^{*}
	(0.236)	(0.350)	(0.233)	(0.266)
Moved State	0.223^{***}	0.235^{***}	0.226^{***}	0.220^{***}
	(0.068)	(0.070)	(0.068)	(0.066)
Observations	21,138	21,138	21,138	21,138
R-squared	0.136	0.187	0.138	0.168
Panel C: County-level migration				
log(Market Access), 1820 Pop.	-0.625^{***}	-0.717^{**}	-0.557^{**}	-0.483^{*}
	(0.238)	(0.352)	(0.234)	(0.267)
Moved County	0.049	0.047	0.049	0.038
	(0.065)	(0.066)	(0.065)	(0.062)
Observations	21,138	21,138	21,138	21,138
R-squared	0.135	0.186	0.137	0.167
Panel D: Enlistment county market access				
log(Market Access), 1820 Pop.	-0.598^{**}	-0.710^{**}	-0.524^{**}	-0.482^{*}
	(0.235)	(0.348)	(0.231)	(0.266)
Enlistment County MA	−0.030	−0.014	−0.019	−0.015
	(0.224)	(0.244)	(0.222)	(0.216)
Observations	21,138	21,138	21,138	21,138
R-squared	0.140	0.191	0.142	0.172
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 and with data available on county of enlistment. All specifications include birth year, measurement age, enlistment year, and county fixed-effects, as well as an indicator for transport linkage in the birth year. Enlistment County MA is the log of market access based on 1820 population for the enlistment county in 1860. Panel D also includes state-of-enlistment fixed effects. Standard errors clustered at the county level. Observations weighted to correct for oversampling.

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.

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.

Correia, Sergio (2017). "Linear Models with High-Dimensional Fixed Effects: An Efficient and Feasible Estimator." Working Paper `http://scorreia.com/research/hdfe.pdf`.

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

Ferrie, Joseph P. (1997). "Migration to the Frontier in Mid-Nineteenth Century America: A Re-Examination of Turner's 'Safety Valve'." Mimeo., Northwestern University.

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

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.

Hornbeck, Richard and Martin Rotemberg (2019). "Railroads, Reallocation, and the Rise of American Manufacturing." NBER Working Paper 26594.

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

Komlos, John (1998). "Shrinking in a Growing Economy? The Mystery of Physical Stature during the Industrial Revolution." *Journal of Economic History* 58:3, pp. 779–802.

——— (2004). "How to (and How Not to) Analyze Deficient Height Samples." *Historical Methods* 37:4, pp. 160–173.

Logan, Trevon D. (2009). "Health, Human Capital, and African-American Migration before 1910." *Explorations in Economic History* 46, pp. 169–185.

Stewart, James I. (2006). "Migration to the Agricultural Frontier and Wealth Accumulation." *Explorations in Economic History* 43, pp. 547–577.

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

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