

Sample-Selection Biases and the Industrialization Puzzle

HOWARD BODENHORN, TIMOTHY W. GUINNANE,
AND THOMAS A. MROZ

Understanding long-term changes in human well-being is central to understanding the consequences of economic development. An extensive anthropometric literature purports to show that heights in the United States declined between the 1830s and the 1890s, which is when the U.S. economy modernized. Most anthropometric research contends that declining heights reflect the negative health consequences of industrialization and urbanization. This interpretation, however, relies on sources subject to selection bias. Our meta-analysis shows that the declining height during industrialization emerges primarily in selected samples. We also develop a parsimonious diagnostic test that reveals, but does not correct for, selection bias in height samples. When applied to four representative height samples, the diagnostic provides compelling evidence of selection.

Introduced into economic history just over three decades ago, anthropometric history is now an important component of the cliometrician's standard toolkit. Roderick Floud et al. (2011, p. 374) note that anthropometric history originated with the limited objective of establishing

The Journal of Economic History, Vol. 77, No. 1 (March 2017). © The Economic History Association. All rights reserved. doi: 10.1017/S0022050717000031

Howard Bodenhorn is Professor, John E. Walker Department of Economics, Clemson University, Clemson, SC 29634. E-mail: bodenhorn@gmail.com. Timothy W. Guinnane is Philip Golden Bartlett Professor, Department of Economics, Yale University, Box 208269, New Haven, CT 06520-8269. E-mail: timothy.guinnane@yale.edu. Thomas A. Mroz is Professor and Bernard B. and Eugenia A. Ramsey Chair of Private Enterprise, Department of Economics, Andrew Young School of Policy Studies, Georgia State University, 14 Marietta Street NW, Atlanta, GA 30303 and Research Fellow at the Federal Reserve Bank of Atlanta. E-mail: tmroz@clemson.edu.

For comments and suggestions we thank Shameel Ahmad, Cihan Artunç, Gerard van den Berg, Claire Brennecke, Raymond Cohn, Thomas Cvrcek, Jeremy Edwards, James Fenske, Amanda Gregg, Farly Grubb, Sukjin Han, Brian A'Hearn, Philip Hoffmann, Sriya Iyer, John Komlos, John Murray, Sheilagh Ogilvie, Jonathan Pritchett, Paul Rhode, Mark Rosenzweig, Gabrielle Santangelo, Richard Steckel, Jochen Streb, William Sundstrom, Werner Troesken, James Trussell, Christopher Udry, Marianne Wannamaker, John Warner, David Weir, anonymous referees, and participants in seminars at the University of Michigan, the University of Nuremberg, Queen's University (Ontario), the Rhein-Westfälisches Wirtschaftsintitut, Tulane University, and the 2012 Cliometrics meetings. Work with the Union Army data analysis was supported by Award Number P01 AG10120 from the National Institute on Aging. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute on Aging or the National Institute of Health. We thank Emilia Arcaleni, John Murray, and Richard Steckel for sharing data. We acknowledge financial support from the Yale University Economic Growth Center. Meng Liu, Yiming Ma, Michael Sikivie, and Adèle Rossouw provided excellent research assistance. The views expressed here are the authors' and not necessarily those of the Federal Reserve Bank of Atlanta, or the Federal Reserve System.

the heights and health of the inhabitants of North America and Western Europe in the eighteenth century. Since then, the study of heights has “mushroomed into a study of the long-term development of human society,” drawing on the tools and insights of economics, statistics, and medicine, among other fields. It is fair to say that no other branch of cliometric history has had as many resources devoted to it in the past two decades as historical anthropometrics (Voth and Leunig 1996, p. 541).

Physiological improvement takes a central place in the list of factors that influence economic development in the long run and cliometricians have developed a systematic approach to the study of the connection between changes in the human body and economic development labeled the techno-physical approach (Deaton 2013; Floud et al. 2011). But as Floud et al. (2011) note in their summary of the British and American experiences, measures of physiological well-being, including average height, did not grow monotonically over the long run. Much of the available evidence from these two countries reveals that the average measured height of populations stagnated or declined in the early phase of modern economic growth. In the United States, in particular, measured mean height declined for cohorts born between (approximately) the 1830s and the 1890s. This apparent anomaly, now widely labeled the “industrialization puzzle,” is one of the most-studied issues in the subfield of anthropometric history (Floud et al. 2011, p. 298). Similar substantial, long-term declines have also been identified for Sweden and the Habsburg monarchy, although the declines in measured heights occurred at different times (Floud, Wachter, and Gregory 1990; Sandberg and Steckel 1987; Komlos 1989). John Komlos (1994, p. 493) called the decline in heights in the early phase of economic modernization in these countries “the most amazing discovery” of anthropometric history. Although researchers discuss several potential causes, most focus on the failure of food supplies (measured by food’s quantity, quality, or both) and public health, broadly conceived, to keep up with population growth and urbanization (Komlos 2012). “Economic growth in the nineteenth century,” write Floud et al. (2011, p. 348) “was very costly because economic booms caused rapid population growth, internal and external migration, urbanization, sanitation problems, and rampant diseases; all these reduced people’s productivity and their ability to accumulate human capital.”

The apparent decline in heights in the United States, Great Britain, Sweden, and Habsburg-era central Europe is indeed interesting, yet we question the reliability of the evidence adduced for this apparent decline. These countries had fundamentally different economies at the time of their height reversals, but they shared an important feature: they filled their

military ranks with volunteers rather than conscripts. A volunteer sample, which is the predominant type of sample in the literature, is selected in the sense that such samples contain only individuals who chose to enlist in the military. Elsewhere we have shown that the problem of inferring changes in population heights from a selected sample of volunteers can be grave (Bodenhorn, Guinnane, and Mroz 2014). The implications of selection bias render the observed “shrinking in a growing economy” less of an anomaly (Komlos 1998a). As the economy grows, the outside option of military service becomes less attractive, especially to the productive and the tall. Military heights declined because tall people increasingly chose non-military employment. Thus, we cannot really say whether population heights declined; we can only be confident that the average height of those willing to enlist in the military declined. Later we draw on published studies to document an important feature of the data, namely that height reversals or “puzzles” are less often observed in countries that filled their ranks through conscription with nearly universal examination and measurement.

Sample selection can take two forms that the heights literature sometimes confuses. The first type, selection on observables, pertains to sampling on an observable characteristic such as race or sex. Modern surveys often deliberately over-sample on such characteristics. If we know the proportions of such groups in the population, we can construct and apply weights to obtain unbiased estimates of the population’s characteristics. The second type of selection, selection on unobservables, on the other hand, refers to a situation in which an individual enters the sample, in part, due to the unmeasured characteristics that are related to the outcome of interest. A criminal in a prison sample, for example, enters the sample because his non-criminal opportunities are less attractive than criminal activities (Bodenhorn, Moehling, and Price 2012). If the choice to engage in crime was driven by unobservable characteristics that are correlated with height, say childhood poverty, sampling weights based only on observed characteristics will not correct for the selection. The necessary weights would be individual-specific and depend on height or the set of unobserved characteristics correlated with height.

Some cliometricians contend that the issues surrounding selection and selection bias are now well understood and that most historical heights researchers account for (or at least qualify their conclusions based on) any potential selection issues. We disagree. Sample selection problems of the selection-on-unobservables type tend to be minimized when they are discussed at all. A recent study of Portuguese heights is typical in this regard (Stolz, Baten, and Reis 2013). Other researchers make a virtue

of the oversampling of the poor and working classes, but this argument confuses the selection-on-observables type with the selection-on-unobservables type. It also misses the consequences of selection on unobservables. Military and prison samples, for example, over-represent the poor and working classes not because the poor were randomly over-sampled to achieve that result, but because poor and working class men had unobserved characteristics that made them more likely to find soldiering or criminal activity to be their best option. Similarly, the poor men who entered the military or the prison were not a random sample of poor men. As we argue later, there is good reason to believe that even for those of a given socioeconomic class, men in the army and prison were shorter than those who did not enter. And, we reiterate, it is likely that selection on unobservables changes over time and in response to economic conditions.

We are not the first to recognize the problem of selection on unobservables in the historical heights literature. Komlos' (1993, p. 119) reanalysis of lower-class, eighteenth-century English boys leads him to conclude that the "unreasonably abrupt" changes in average heights are likely the result of changing recruitment practices. Paul Johnson and Stephen Nicholas (1995, p. 471), too, attribute large, short-term changes in heights to selection bias. Recently, Ariel Zimran (2015) develops a two-step semi-parametric estimator with potential selection on a single feature, and finds that selection is important in height data derived from nineteenth-century U. S. military samples but the bias is not sufficiently large to overturn the decline in heights observed for Union Army soldiers. Thomas Mroz (2015), however, shows that Zimran's approach may not accurately correct for selection bias if selection reflects multiple selection conditions that depend on more than one unobserved factor. Thus, while we are not the first or the only researchers to discuss the effects of selected samples in the historical heights literature, we may be the first to systematically explore whether it can explain the apparent reversals in average heights in growing economies or the so-called industrialization puzzle in both an American and cross-country context.

THE INDUSTRIALIZATION PUZZLE IN AN AMERICAN FRAME

From very nearly the beginning of the anthropometric history, scholars report an apparent anomaly: the average height of adult male native-born Americans began a decline in height among those males born in the 1830s (we follow the convention in the literature that dates refer to birth years rather than observation years) (Margo and Steckel 1983; Fogel 1986). The downward trend persisted through the 1870s and 1880s, after which average adult male heights increased at the relatively rapid rate of 1.8 cm per decade

between 1902 and 1931 (Fogel 1986, p. 511). The pattern of declining heights at mid-century is “puzzling because according to the conventional indicators the American economy was expanding rapidly during the antebellum decades” (Komlos 1987, p. 898). Ordinarily, economists expect periods of economic growth to translate into rising living standards.

The American incarnation of the industrialization puzzle is all the more puzzling because, among the early industrializers, only England, Sweden, and Austria-Hungary appear to have experienced a similar confluence of economic growth and declining average male stature. Figure 1 reproduces Robert Fogel’s (1986) graph of U.S. adult male heights on which we superimpose mean heights reported for four continental European countries that span the nineteenth century. Plots of the mean height of the Dutch, Swedes, Italians, and French trace out long-run secular growth paths that, while country specific, demonstrate no large, persistent reversal like that observed for the United States. Plots of mean heights (not shown) of Russians (Mironov 1995; Wheatcroft 2009), Bulgarians (Popoff 1926), Spaniards (Ayuda and Puche-Gil 2014), and Japanese (Shay 1994) drawn from large, representative conscript samples all trend upward after 1800 without large or persistent reversals. Despite Komlos’ (1998b, p. 236) contention that the industrialization puzzle is not a “statistical artefact [sic],” its regular appearance in selected samples (e.g., military volunteers, prisoners, students) and its failure to appear in representative conscript samples raises questions about whether the puzzle is an artifact of selected samples.

Convinced that it was not an artifact, historical cliometricians turned their attention to uncovering the industrialization puzzle’s sources. Explanations for the puzzle focus on declines in available foodstuffs (or a rise in their relative price) that led to a long-run decline in net nutrition; to increases in the disease load due to urbanization and a widening of the transportation network; to increased income inequality, which negatively affected the heights of the lower classes more than secular growth positively affected the height of the middle and upper classes; to increased work intensity, which would have more negatively affected the height of factory workers more than farmers (who already worked hard) and white-collar workers; and to increased immigration around mid-century, which would reduce mean adult height if immigrants were shorter than native-born Americans and if the children and grandchildren of immigrants carried the immigrant height disadvantage across generations (Sunder 2004, pp. 76–77; Komlos 2012). The puzzle was a welcome addition to the literature, in part, because it bolstered the case of pessimists who argue that early industrialization diminished aggregate well-being (Feinstein 1998).

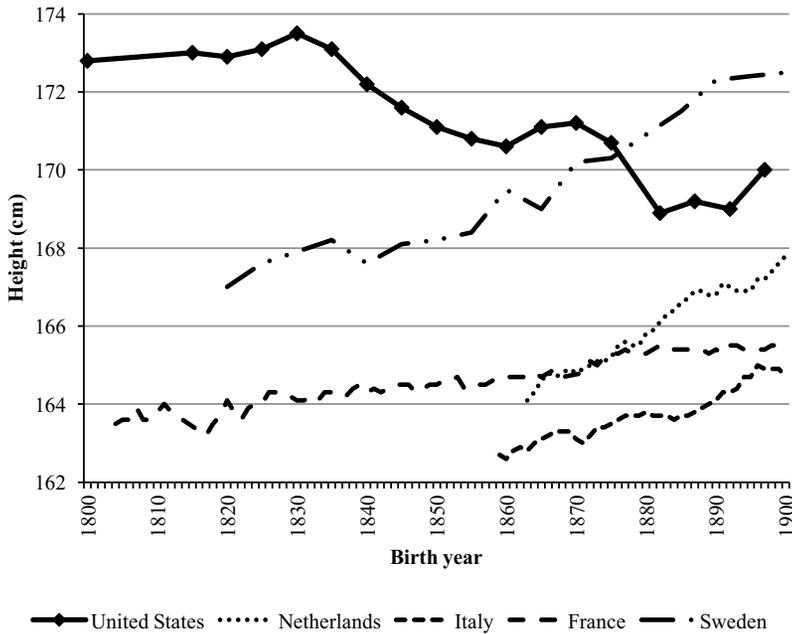


FIGURE 1
MEAN HEIGHTS OF UNITED STATES MILITARY VOLUNTEERS
AND EUROPEAN CONSCRIPTS

Sources: United States: Fogel (1986); Netherlands: Mandemaker (1993); Italy: Federico (2003); France: Weir (1997); Sweden: Sandberg and Steckel (1997).

THE EMPIRICAL BASIS OF THE PUZZLE IN THE UNITED STATES

Despite its emergence as a stylized fact the industrialization puzzle has a surprisingly ambiguous evidentiary basis. In an early review of the literature, Komlos (1998a, p. 782) discusses 21 height samples, 14 of which report declining mean height during one or more decades between 1820 and 1890. In three other samples, mean height increased; one sample reports no mean change in height. Komlos (1998b, p. 236) also reports the approximate date at which heights initially turned down (when they did), with dates ranging between the 1780s and the 1840s. Excluding the three studies that report increasing average height, the modal downturn began in the 1830s; the median downturn occurs in the 1820s. Komlos' dating of the puzzle as beginning circa 1830 and ending circa 1890 accords with Fogel's (1986) account, which he develops by splicing together several independently constructed series. Fogel's (1986, p. 465) graph, in Figure 1, has achieved nearly iconic status (e.g., Steckel 1995; Costa and Steckel 1997; Floud et al. 2011, p. 299).

Many of the most-cited articles in the literature provide evidence that spans only a part of the 1830 to 1890 era; some end before or at

approximately the onset of the posited decline in height. These studies are thus of limited use in documenting or understanding the phenomenon (e.g., Margo and Steckel 1982; Margo and Steckel 1983; A'Hearn 1998; Bodenhorn 1999). Since Komlos' earliest reviews of the literature, several studies have appeared that document or further discuss the puzzle and, importantly, span all or most of the period between 1830 and 1890. Table 1 summarizes the results of 21 samples reported in 12 separate studies. The 21 samples include more than 230,000 individual observations of mostly native-born North Americans. Several studies report results for distinct groups—blacks and whites, men and women, or adults and youth—that are treated here as independent samples drawn from an underlying North American population.

The third column reports the first feature of studies addressing the industrialization puzzle in North America. Most of the evidence on the puzzle is inferred from potentially unrepresentative samples of prisoners, soldiers, students, or passport applicants. Robert E. Gallman (1996) notes that there is no reason to believe that people entering these types of samples are representative of Americans. Moreover, he notes that there is no reason to believe that the characteristics of these self-selected samples remained constant over time, which makes inference highly problematic (see Howard Bodenhorn, Timothy Guinnane, and Thomas A. Mroz 2014 for a formal analysis of the consequences of changing selection on unobservables over time).

The second feature of note in the table appears in the final column, which reports the changes in average height between the first observation date and the final observation date for each sample. At one extreme, Scott Alan Carson's (2009) sample of white male adult prisoners implies that average height declined by 1.68 percent for cohorts born between 1830 and 1889. At the other extreme, Carson's (2011) sample of white female adult prisoners implies that average height increased by 2.00 percent over nearly the same birth cohorts. The other samples' estimates report results between these two bounds, six of which report increasing height and two report no change in height (rounded to the first decimal point in centimeters). The average change in height for the 21 samples is +0.015 percent, which indicates that average height did not change much over the nineteenth century. If, alternatively, we weight the individual sample average changes by sample size, average height changed between 1830 and 1890 by -0.12 percent. If we take the weighted average to approximate the true change in average adult male height circa 1830 to 1890 and that average adult male height in 1830 was 172 cm, average height declined by about 0.2 cm (or one-twelfth of an inch). Compared to the 7 cm standard deviation in a cross section in adult male heights or the 1 cm to 2 cm standard

TABLE 1
SUMMARIES OF NORTH AMERICAN INDUSTRIALIZATION PUZZLE STUDIES

Author	Year of Study	Type	Race	Sex	Age	Sample Size	Years Covered	Height Change (Percent)
Komlos	1987	Military	W	M	Youth	953	1820–79	–0.35
Komlos	1987	Military	W	M	Youth	829	1820–79	–0.23
Steckel/Haurin	1994	Military	W	M	Adult	4,354	1845–95	–1.31
Komlos	1996	Military	W	M	Youth	1,614	1835–79	1.01
Murray	1997	Student	W	M	Youth	2,499	1830–79	–0.23
Cranfield/Inwood	2007	Prison	W	M	Adult	1,827	1830–73	0
Maloney/Carson	2008	Prison	W	M	Adult	17,337	1830–90	–0.8
Maloney/Carson	2008	Prison	B	M	Adult	3,520	1830–90	0
Maloney/Carson	2008	Prison	W	M	Youth	7,152	1830–95	–0.26
Maloney/Carson	2008	Prison	B	M	Youth	1,551	1830–95	–0.31
Carson	2008	Prison	B	M	Youth	17,954	1820–89	0.16
Carson	2008	Prison	B	M	Adult	57,290	1820–89	–0.35
Carson	2009	Prison	W	M	Youth	4,468	1850–94	–0.46
Carson	2009	Prison	B	M	Youth	8,755	1850–94	–0.98
Carson	2009	Prison	W	M	Adult	11,245	1820–94	–1.68
Carson	2009	Prison	B	M	Adult	17,798	1820–94	1.59
Carson	2011	Prison	W	F	Adult	2,244	1830–89	2.00
Carson	2011	Prison	B	F	Adult	2,794	1830–89	0.37
Zehetmayer	2011	Military	W	M	Adult	41,713	1847–89	–0.56
Sunder	2011	Passport	N/A	F	Adult	5,952	1825–94	1.70
Sunder	2013	Passport	N/A	M	Adult	19,893	1830–89	1.01
Average								0.015
Sample size weighted average								–0.118

Notes: These studies were chosen because they report estimates of heights by year of birth, they span all or most of the period between 1825 and 1890, and are based on reasonably large samples. The sample types of military, students, and prisoners are representative of samples used in the literature. The Komlos (1987) are for military academy (West Point) students ages 18 and 19 years and are classified as military, but could be considered students.

Sources: Authors' calculations from listed studies.

deviation in the time series, a 0.2 cm change is substantively trivial. It is hardly evidence, moreover, of a meaningful mid-century North American decline in physiological well-being.

A META-ANALYSIS OF THE HISTORICAL HEIGHTS LITERATURE

Our analysis of 12 studies of the North American industrialization puzzle casts doubt on whether the mid-century decline in heights was genuine, but these studies alone cannot resolve whether the posited decline in height was a consequence of the selected nature of the samples. Ideally, we would need to compare selected samples to non-selected samples drawn from the same population to determine whether the

observed decline in some samples is driven by selection on observables, selection on unobservables, or a genuine decline in average height. There is, however, no known representative sample of American-born men or women that allows for such a comparison. Figure 1 points to an alternative approach: it compares evidence from North America, where, except for three brief periods in the nineteenth and twentieth centuries, armies were filled with volunteers, with evidence from four Western European countries. These four countries recruited their armies differently than the United States: they employed near-universal conscription, a system under which nearly all young men eligible for military service were called for medical examination to determine their fitness.

In theory, all men of a certain age were eligible to serve; in practice, of course, some young men were able to avoid service for a variety of reasons. Unless a significant fraction of the eligible young men were able to avoid appearing for examination and measurement and the ability to avoid was correlated either positively or negatively with the young men's height, conscript samples should provide reliable estimates of the true mean height-at-age of each cohort. It is important to emphasize as well, that being measured and serving were separate events. Not all measured men served, partly because only a fraction of men were called for active service and partly because some young men were able to avoid being called after being measured. The conscript samples, then, should not be plagued as severely by the selection-on-unobservables problem inherent in samples of volunteer armies, prisons, and schools. We exploit this feature of the data to better understand whether the substantial declines in height observed in nonrandom samples in the eighteenth and nineteenth centuries were more likely to have been the consequence of changes in selection or changes in average height driven, presumably, by changes in physical well-being.

Conscript data have additional appealing features. One advantage is that because men were typically examined at a given age (usually around ages 18 to 20), and age-at-measurement did not change much over time, conscript data create snapshots of same-age men across several decades. This feature makes inter-temporal comparisons relatively straightforward, unlike volunteer military or prison samples in which age-at-measurement includes teenagers to 60-year olds. The entrance of several different cohorts makes sorting out selection effects especially problematic because the environmental and economic factors that lead to enlistment in a given year for a 20-year old and a 30-year old are not the same, even though they face the same contemporaneous macro (but not the same micro) phenomena. It is this issue that requires the inclusion of individual-specific selection variables in a model capable of capturing

and correcting for selection on unobservables. A second advantage of conscript data is that many young men were measured by the time they had or had nearly achieved their terminal adult height. Third, because the samples are self-weighted, the clinometrician does not have to rely on a census to determine the sampling weights necessary to reconstruct a population mean height estimate.

Figure 1 provides a handful of examples of the differences that emerge between volunteer and conscript samples. We expanded the range of examples by drawing on 101 historical height studies, either books or articles, that report information on 169 separate samples, which may be separated by race, sex, age groups (adults and youth), and so on.

We employed the following procedure to identify studies for inclusion. First, we identified historical height studies that relied primarily on data assembled from military conscripts and included all that we could locate. Second, we searched for studies that relied primarily on data assembled from volunteers and published between 1995 and 2014 in the principal outlets for historical height studies, which included economic history and human biology journals. Third, we identified nonmilitary studies published in the same sources and during the same period. Fourth, we read articles identified in the first, second, and third steps and continued reading articles until we reached a target total of either 100 articles and/or 150 separate samples. The final set of studies included in the meta-analysis consists of 50 conscript samples, 39 volunteer samples, and 80 samples we label “Other.” Other includes an eclectic mix of runaway and manumitted slaves, indentured servants, immigrant workers, voters, passport applicants, government employees, enrollees in health and life insurance programs, unclaimed Korean corpses, and U.S. baseball players. Only a handful of the nonconscript samples are national in scope. Some samples are very small; one has only 151 observations over 34 years, though the median number of observations in the “Other” group is 3,900 spread over a median period of 40 years. The database of height studies includes samples from 40 countries, including 36 entries for the United States.

The fifth step in reading the historical heights articles was to construct a database that includes the first and last year a height is reported, the average height at first and last year, the age of those measured, whether the study corrected the reported heights for so-called left-tail shortfall, which means that the observed distribution of heights is truncated at or near the military’s minimum height standard, the number of observations and, most importantly, whether the data reveal a substantial decline in reported heights that persists over one or more decades. We label these declines as “Reversals” if the mean (raw data) or estimated height (based on regression coefficients) decline by at least 1 cm during a particular

TABLE 2
SAMPLE TYPE AND FRACTION DEMONSTRATING A 1 CM OR MORE DECLINE
IN MEASURED AVERAGE HEIGHT (PERCENT)

	One Reversal	Two Reversals
Conscripts	20	4
Volunteers	56	33
Students	57	15
Prisoners	70	35
Other	34	7

Notes: See text for explanation of sample types.

Sources: Authors' calculations from 167 samples discussed in text.

decade. Although a 1 cm reversal appears to provide a low threshold for a meaningful height decline, it is one that occurs so infrequently that it will not fall into what Komlos (1993, p. 130) calls the trap of attaching “too much significance to slight deviations from the main trend” in heights. Moreover, Joerg Baten (2009, p. 172) and Baten and Komlos (1998) contend that a 1 to 1.2 cm change in a decade is biologically and economically significant change in mean male adult height.

Table 2 reports the proportion of samples that document a height reversal (or decline of 1 cm or more) in any decade (Column 1) or a decline in two or more decades (Column 2). Twenty percent of studies based on conscripts identify at least one reversal during an average of five decades (median of six decades). The other 80 percent of conscript samples reveal no evidence of a substantive height reversal in any decade. Just 4 percent of conscript samples identify two or more reversals. More than one-half of volunteer samples, on the other hand, report one reversal during a mean sample period of four decades (median of three decades). One-third of volunteer samples identify two or more reversals. Similarly more than one-half of student samples identify one reversal; 70 percent of prisoner studies report a reversal; and one-third of the “Other” samples report reversals.

A simple comparison of the likelihood of observing a reversal in conscript, volunteer, and other common types of samples included in the literature shows that the puzzle appears more often in samples subject to potential selection bias. We sharpen this conclusion with the assistance of a series of probit regressions in which the dependent variable equals one if a sample reports at least one reversal (or, alternatively, two or more reversals) and zero otherwise. A regression analysis allows us to control for effects other than sample type. Column 1 of Table 3 reports the across-sample means for each variable coded from our reading of the 101 studies. The sample is made up of approximately 25 percent each of

TABLE 3
ESTIMATED PROBABILITIES (MARGINAL EFFECTS) OF OBSERVING A 1 CM DECLINE IN ONE OR MORE DECADES

	(1) Summary Statistics	(2) At Least One 1 cm Decline	(3) At Least One 1 cm Decline	(4) At Least One 1 cm Decline	(5) Two or More 1 cm Declines
Conscript	0.255	Reference	Reference	Reference	Reference
Volunteer	0.242	0.372 (0.092)**	0.448 (0.074)**	0.403 (0.073)**	0.453 (0.138)**
Prisoner	0.172	0.257 (0.114)*	0.367 (0.097)**	0.396 (0.081)**	0.413 (0.164)*
Student	0.089	0.207 (0.123)†	0.311 (0.105)**	0.361 (0.077)**	0.239 (0.208)
Other	0.242	0.105 (0.102)	0.228 (0.095)*	Excluded	Excluded
ln(years)	3.809	0.327 (0.071)**	0.305 (0.077)**	0.264 (0.078)**	0.331 (0.091)**
ln(obs)	8.915	-0.070 (0.023)**	-0.063 (0.028)*	-0.047 (0.035)	-0.029 (0.030)
National	0.223	-0.094 (0.091)	-0.039 (0.101)	-0.142 (0.116)	0.252 (0.145)†
Correction	0.166	-0.103 (0.116)	-0.228 (0.094)*	-0.287 (0.089)**	-0.063 (0.105)
North America	0.287		-0.209 (0.087)*	-0.237 (0.088)**	-0.272 (0.081)**
Europe core	0.191		Reference	Reference	Reference
Europe Periphery	0.274		-0.053 (0.106)	-0.016 (0.114)	-0.148 (0.090)†
Asia	0.127		-0.176 (0.136)	-0.443 (0.069)**	-0.179 (0.079)*

South Hemisphere	0.121		−0.323 (0.096)**	−0.321 (0.106)**	−0.139 (0.086)
before 1750	0.115		−0.118 (0.132)	0.021 (0.185)	0.030 (0.157)
1750–1799	0.108		−0.138 (0.108)	−0.243 (0.107)*	−0.227 (0.063)**
1800–1829	0.197		−0.122 (0.092)	−0.099 (0.103)	−0.033 (0.097)
1830–1889	0.363		Reference	Reference	Reference
After 1890	0.216		−0.277 (0.099)**	−0.402 (0.091)**	Excluded
N	157	157	157	119	100
Log likelihood		−84.4	−77.9	−51.3	−38.6
Pseudo R-sq		0.22	0.28	0.38	0.30

Notes: Probit estimations; columns 2 through 5 report estimated average marginal effects, using the the margins dydx command in Stata. Robust standard errors reported in parentheses. ** implies p-value < 0.01; * implies p-value < 0.05; † implies p-value < 0.10. The dependent variable in columns 2 and 3 equals 1 if sample reports a 1 cm or greater decline in average height in any decade, and equals 0 otherwise. The dependent variable in columns 4 and 5 equals 1 if the sample reports a 1 cm or greater decline in average height in two or more decades. The “Other” category includes indentured servants, slaves, migrant workers, baseball players and sundry other groups of individuals. Variable ln(years) calculated as difference between first and last year included in the sample. If the author reported dates such as 1820s–1880s, the first year is taken to be 1820 and the last year as 1889. National equals 1 if sample relies on a nationally representative sample. Correction variable equals one if author reports mean heights after correcting for left-tail shortfall using either reduced sample maximum likelihood estimators (RSMLE), quantile bend estimator (QBE), or Komlos-Kim correction methods. Geographic dummies: Europe Core includes western Europe, Scandinavia and UK, excluding Ireland. Europe Periphery includes southern and eastern Europe countries. Asia includes India. Southern Hemisphere includes South America, Australia and New Zealand. The excluded category is North America. The Time period controls take a value of 1 if the first year of the sample starts in the relevant period. The excluded category is the era commonly attributed to demonstrate the industrialization puzzle in North America (1830–1889).

Sources: Authors’ calculations.

conscripts, volunteers, and others; students (8.8 percent) and prisoners (17.1 percent) make up the remaining observations. The studies differ in that some cover less than two complete decades; others span a century or more. They also vary widely in sample size, from less than 200 to more than 10 million persons. And about two-thirds of the volunteer sample results were reported after making a correction for left-tail truncation.

Column 2 reports estimated average marginal effects for a given type of data (volunteer, prisoner, etc.), as well as the number of years covered, sample size, whether the sample is nationally representative, and whether the author corrects for left-tail truncation on the probability of observing one 1 cm reversal. Column 3 adds broad geographic and time period controls based on the first year height is observed in a sample. The results in Column 2 show that, relative to a sample of conscripts, the probability of observing at least one 1 cm or more decline in average height in at least one decade is 37.2 percentage points greater for samples of volunteer soldiers, 25.7 percentage points greater for prisoners, 20.7 percentage points higher for students, and 10.5 percentage points for “Other” samples, though the last is not statistically significant at standard levels. Estimated average marginal effects on the additional features are consistent with prior expectations. We are more likely to observe a reversal in samples that span more decades, and less likely to observe a reversal as sample size increases, when the sample is nationally representative, and when the study controls for left-tail truncation.

Columns 3 and 4 repeat the same exercise when the regression includes the broad geographic and time period controls, including and excluding the “Other” category, which includes such diverse sources that it is difficult to interpret the effect (Komlos and A’Hearn 2016, pp. 40–41). Given that the samples were drawn from around the world and at different times, there is concern that the increased likelihood of observing a reversal in a nonconscript sample might be driven, in part, by omitted variable bias. But when the geographic and period controls are included, the estimated likelihoods of observing a reversal actually increase. When we include a richer set of controls in Column 3, the likelihood of observing a reversal in a sample of volunteers is 45 percentage points higher than it is for conscript samples; it is more than 37 percentage points higher for prisoners than conscripts, which is one of the most common type of data used in studies that identify an industrialization puzzle. Column 4 excludes the “Other” category and the estimated likelihoods of observing a reversal in volunteer and prisoner samples are 40 percentage points higher than conscripts for each.

In Column 5 the dependent variable equals one only if a sample included in the meta-analysis reports two or more reversals. Whereas about 45

percent of studies report a single reversal, less than 25 percent of samples report two or more reversals. The estimated marginal effects for volunteers and prisoners suggest that observing multiple reversals are about 40 to 45 percentage points higher in volunteer and prisoner samples than in conscript samples. An interesting result in this and in the regressions reported in Columns 3 and 4, is that, after controlling for other features of the sample, the likelihood of observing a reversal in North American samples is about 20 to 30 percentage points lower than observing a reversal in Western Europe. The regression results support our contention that the often-observed result of declining heights in nineteenth-century North America is an artifact of the samples used, not necessarily the consequence of a notable decline in health or nutritional status.

The meta-analysis focuses on the role of sample selection in height reversals. So while it includes studies not concerned with the industrialization puzzle (1830–1890), it provides important insights into the literature. The United States did not have military conscription until the twentieth century, except for a brief period during the American Civil War during which most men could avoid it through various means (A'Hearn 1998). Americans' reliance on local and state militias and federal volunteers complicates the historical cliometrician's task because U. S. height data prior to the twentieth century comes almost exclusively from selected samples. Thus, the limitations on U.S. sources render definitive statements about American incarnations of the industrialization puzzle tenuous.

HISTORICAL HEIGHTS AND SELECTION ON UNOBSERVABLES

Selection on unobservables arises when individuals appear in a sample only because they (or possibly someone else) make decisions that reflect unmeasured individual constraints and preferences that are related to the outcomes of interest. The most general case for our purposes is soldiers joining a volunteer army, but the relevant decision can also be made by someone other than the person whose height is measured, for example, runaway or transported slaves. To make this problem concrete, we summarize the Roy-style model in the spirit of James Heckman and Guilherme Sedlacek's (1985) two-sector occupational choice model that is described in detail in Bodenhorn, Guinnane, and Mroz (2014). The A. D. Roy (1951) model is a workhorse tool in labor economics and other areas, applied to situations where individuals make a binary choice. The model assumes that each individual decides whether to work in the civilian or military sector (i.e., join the army), and only the latter appear in the height sample.

Each individual has a prospective civilian and military wage; at least one of these wages is correlated, in part, with their height. Wages could be correlated with height under either of two conditions. The first is that the army or some civilian occupations might reward height itself. Promotion might come faster to taller soldiers because of their height, and in some military and civilian occupations tall people might be more productive because they are tall. A second possibility seems more consistent with the basic tenets of the heights literature. Height is correlated with a person's (mostly unobserved) health human capital and thus their productivity (Schultz 2002). The model also assumes that each person receives individual-specific shocks to their civilian and military wages, and has individual-specific preferences over civilian versus military life.

The decision to join the army reflects individual shocks and preferences, as well as individual height and the return to height. The distribution of height for men who join the army can then be written as the product of the population height distribution (in our notation, $f(h)$) and a second expression ($Z(h)$) that summarizes the decisions made by individuals of different heights. $Z(h)$ is a function of all model parameters, most importantly, the return to height or its unobserved correlates in the civilian sector (β_C) and the military sector (β_M). Selection on unobservables arises when $\beta_C \neq \beta_M$, or, selection arises when the two sectors differentially reward any individual characteristics correlated with height. If soldiers are selected on a characteristic correlated with height, the height distribution of soldiers will not be the same as the height distribution of the population from which the soldiers are drawn.

If, as seems reasonable for many characteristics, $\beta_C > \beta_M$ then the heights of volunteer soldiers will understate the population height. The degree of height understatement in the selected sample depends critically on the relative sizes of β_C and β_M as well as the other parameters; as the parameters change over time (or space), the magnitude of the selection bias will change. Thus, if β_M is less responsive to fluctuations in the labor market than β_C because, perhaps, military rewards are macroeconomically acyclical, selection into the military will vary in ways that make it treacherous to infer trends in population heights from military samples. The simulations reported in Bodenhorn, Guinnane, and Mroz (2014) demonstrate that small variations in the incentives to join the army can generate empirically important variations in the height of the resulting sample, even when the heights of the underlying population are held constant.

The argument distinguishes sharply between a sample selected on unobserved factors and one selected on observable characteristics. Consider, for instance, that a sample has proportionally more common laborers

than are known to exist in the population. This might reflect selection on observables; that is, the sampling procedure took a random sample of workers, but for whatever reason over-sampled the laborers without consideration of any unobserved characteristics. If so, then it is simple to re-weight the sample to obtain correct population height estimates. If, alternatively, the over-representation of laborers reflects the fact that laborers' more preferred employment option is the military, the sample cannot simply be re-weighted to obtain population estimates. In this case the decision to enter the military would likely depend on unobserved characteristics that are correlated with height. The correct weights would be individual-specific and are unknown without constructing a model of the decision to enter the military. The necessary weights would need to take into account the individual's height, unless one could somehow control for all the unobserved productivity and taste parameters correlated with height that influence enlistment.

The heights literature sometimes posits that a sample made up of a disproportionate number of lower- or working-class individuals will accurately reflect, if not population trends, trends among the working classes (e.g., Carson 2008, p. 214). This would be a reasonable inference if the disproportionate representation of the working class was due to a conscious decision on the part of the researcher to oversample that group. But selection on unobservables correlated with height is as likely to occur among the oversampled (working poor) as the undersampled (wealthy) groups. If remuneration is even partially dependent on characteristics correlated with height within the nonrandomly oversampled group, the Roy model still implies that the laborers who join the army will be, on average, of different heights than the laborers who did not. Most importantly, the extent of the difference will vary with civilian-sector opportunities as they change over time.

Numerous examples in the historical heights literature suffer from some type of selection on unobservables. Komlos (1994), for example, compares trends in the heights of upper-class French students and French conscripts. His sample of students enrolling in the *École Polytechnique* between the 1780s and 1860s shows sharp declines in average height in the 1820s (Parisian students) and 1830s (provincial students). He argues that the student sample reflects a type of selection on observables only, and interprets his results to show a height reversal even for the comparatively well-off portions of French society. By comparison, the French series in Figure 1 comes from David Weir (1997), who shows that the heights of French conscripts rose slowly, but continuously throughout the nineteenth century. Komlos' findings more likely reflect changes in selection on unobservables for French university students. His sample of 18-year

old students are always taller than the general population, but over time they become relatively shorter compared to the population of 20-year old French men called up for military medical examination. This is one of the few historical cases where data exist to compare a selected sample to one that is more nationally representative. The selected sample's misrepresentation of the population trends closely tracks the predictions of the Roy model. Conclusions based on the sample of selected students lead to incorrect conclusions about trends in France's economy or the overall well-being of French men.

A similar feature of disparate interpretation based on sample type is evident in twentieth-century U.S. data. Komlos (2008) compares U.S. military volunteers born in 1940 and 1950 to a random sample of the U.S. population. The relevant enlistment period includes the period of the increasingly unpopular Vietnam War. The sample of military volunteers shows a decline and then recovery in average heights of soldiers, neither of which is evident in the representative random sample. Komlos (2008, p. 477) concludes that "these [military] data have their own limitations insofar as they pertain to those selected into the US military at moderate wages." It is difficult to make valid inferences about height trends when samples are subject to selection on unobservables.

Volunteer soldiers, militia men, National Guardsmen, prisoners, runaway servants, or manumitted slaves are among the most commonly used historical height data source, and they share a common trait, namely that everyone who enters these samples were measured only because of a choice they or someone else made. We are not the first to recognize the difference between selection on observables and selection on unobservables. It is discussed, but its implications for inference not detailed, by Lance Brennan, John McDonald, and Ralph Shlomowitz (1994a, 1994b, 1997), Komlos (1993, p. 122), Ricardo Salvatore (1998, p. 105), and Stephen Nicholas and Richard Steckel (1991, p. 949) in their studies of immigrants, poor London boys, Argentine soldiers, and transported prisoners. But the clearest early recognition of selection on unobservables appears in Joel Mokyr and Cormac Ó Gráda's (1994, 1996) two articles investigating the heights of Irish recruits into the Royal Navy and the English East India Company. They find that Irish recruits were (surprisingly) taller than English recruits and attribute the finding to relative absence of economic opportunities in Ireland. A man of given height in Ireland had fewer non-military opportunities than a similarly situated Englishman of equal stature and was, therefore, more inclined to join the military. (In the Roy model, the finding is consistent with the idea that the Irish and English have identical population height distributions, but that $\beta_C - \beta_M$ was smaller in Ireland than in England.) Remarkably, many

heights articles report Mokyr and Ó Gráda's "tall but poor Irish" result without any discussion of the fact that the authors attribute the apparent anomaly to differential selection on height (see, e.g., Nicholas and Steckel 1991; A'Hearn 1998; Salvatore 2004; Morgan 2004; Cranfield and Inwood 2011; Riggs and Cuff 2013).

SELECTION BIAS AND THE PROBLEM OF IDENTIFICATION

Few historical data sets containing height measures were collected as random samples from the populations of birth cohorts, and for some important countries such as the United States, we have none at all. Can we estimate trends in historical heights from biased samples? Uncovering temporal trends in mean height requires the cliometrician to use multiple birth cohort samples, where changing economic and military conditions could lead to across cohort samples exhibiting non-constant degrees of selection biases. To address the general problem, we separate the height distribution in a birth-cohort from the process of selection into an observed sample. The height distribution is the object of interest, but absent correction for selection, it is a convolution of two functions; that is, the distribution of observed (sampled) heights depends on both the parameters of the birth-cohort height distribution and the parameters of the selection function.

Consider a simple world where individuals born at date b grow until just before they turn age w . Before date $b+w$, all have reached their height potential and, upon obtaining age w they make a choice that determines whether they enter the observed sample. Any minimum or maximum height restrictions are assumed to be built into the selection function. The distribution of observed height in the sample depends on two distinct types of factors: the factors that influence growth, between time b to time $b+w-1$, and those that influence the decision to join the sample, observed at time $b+w$. Compare this to the distribution of height for the entire birth cohort b . This depends only on the factors that affect growth between time b and $b+w-1$. Denote these factors by $e(b)$, and denote $v(t_b)$ as the factors that influence whether a member of birth cohort b enters the observed sample. The mean height for birth cohort b would then be a function of $e(b)$ alone, while the mean height for those selected into the sample would depend on both $e(b)$ and $v(t_b)$. Denote the expected height of birth cohort b by $E_b^T(H)=f(e(b))$. Then let the expected height among those selected into the observable sample be $E_b^0(H) = h(e(b), v(t_b)) = f(e(b)) + g(e(b), v(t_b))$, respectively. This expression equals the true height of the birth cohort plus a bias term represented by $g(.,.)$. It is possible to obtain a consistent estimator of the mean height in the selected sample,

but without further assumptions it is not possible to decompose the estimate into the part of the observed mean height due to the cohort's growth environment ($e(b)$, which is what we want) and the part due to selection into the sample (which is bias).

Consider, further, a comparison of individuals born in different years: those born in year b (as above) and those born in year c . The question is whether we can reliably estimate the trend in heights from selected samples drawn from cohort b and cohort c . The mean height of people born in year c and those born in year c who selected into the sample at date t_c are $E_c^T(H) = f(e(c))$ and $E_c^0(H) = h(e(c), v(t_c)) = f(e(c)) + g(e(c), v(t_c))$. To determine whether or the extent to which average height changes over time requires estimates of $f(e(c)) - f(e(b))$, but a reliable calculation depends on having random samples from the two populations. Researchers tend not have these random samples. Many studies instead estimate temporal changes in heights from samples of individuals who selected in, so that the expected difference in observed heights used to proxy for the change in height across birth cohorts will be: $E_c^0(H) - E_b^0(H) = [f(e(c)) - f(e(b))] + [g(e(c), v(t_c)) - (g(e(b), v(t_b)))]$. The second term in braces (the difference in the $g(.,.)$ functions) represents the bias in the estimated change in heights from using the selected samples instead of random samples from the underlying population of heights.

Thus, the cliometrician observes a change in average height that is the sum of two separate effects: the change in the cohort-specific height and the change in the bias terms. This is a classic identification problem. There are three avenues to proceed. First, find true random samples from each of the birth cohorts so that each component of the bias term is zero. But if random samples were available, there would be no need to have relied on selected samples in the first place. Second, assume that the biases in each cohort are identical. This requires, however, the unlikely claim that selection does not depend on any temporal changes in macroeconomic or environmental conditions between dates $b+w$ and $c+w$. Or, third, assume as in Komlos and Joo Han Kim (1990) that height-based selection involves mostly the very tall or very short and that considerations that lead men in two successive cohorts to enter a sample not change over time. The assumptions underlying the second or third options are strong and potentially incorrect.

AN INDIRECT TEST OF SELECTION BIAS

Our discussion of the identification problem naturally leads to a regression approach to test for the presence of selection. Consider the following thought experiment: Suppose we have two dates when members of a

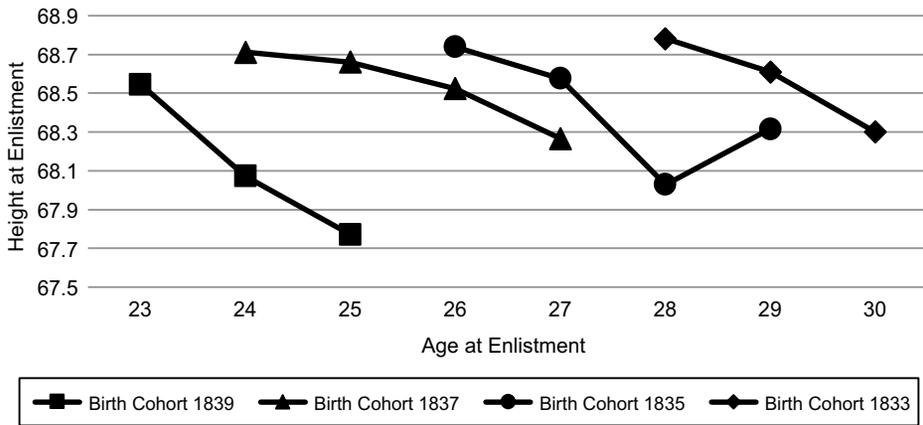


FIGURE 2
 HEIGHT BY AGE AT ENLISTMENT
 FOUR BIRTH COHORTS, UNION ARMY
 (Minimum Height Requirement 63": 9/1863–11/1864)

Sources: Authors' calculation from ICPSR series # ICPSR06837-v6.

birth cohort might select into a sample, such as two ages at which individuals can enlist. Call these ages t_b and $t_b + 1$. In the absence of height-based selection, the distribution of height for those born in cohort b and observed at date t_b should be identical to those born in the same year and “observed” at date $t_b + 1$. If there is no height-related selection, then the average height of individuals from a given cohort should not differ because they joined at one date rather than another. A rejection of the null hypothesis of equal mean height in different observation years for the same birth cohort, provided all members are sufficiently old enough to have reached full height, would be evidence of height-based selection. An ordinary least squares (OLS) regression model can easily carry out this test.

Figure 2 makes this point concrete by presenting mean height by age at enlistment for four birth cohorts using a subset of the Union Army data (Fogel et al. 2000) analyzed in more detail later. An analogous graph for other birth cohorts in the sample looks similar. We focus on only those time periods when the Union Army’s minimum height restriction was exactly 63 inches (160.02 cm) (September 1861–December 1864), and we only consider individuals who reported their age at enlistment as age 23 to age 30. Given the lower age cutoff, it is reasonable to assume that all individuals had reached their terminal height by the time at enlistment.

In the absence of height-related selection, each line in Figure 2 should be horizontal, that is the average height of 23-year olds born in 1839 and measured in 1862 should be the same as a 24-year old born in 1839 and

measured in 1863. Instead, we see declines with age (or equivalently with the duration of the Civil War) over the years observed. The declines range from about one-half inch (1.27 cm) to nearly a full inch (2.54 cm) within these birth cohorts are substantively large height differentials. This is unmistakable evidence of height-related selection. To draw valid inferences about across cohort variations in true mean population height, one would need to specify precise conditions under which the height-related selection biases would cancel out. Further, the researcher would need to provide detailed empirical evidence that such conditions are actually satisfied. That would require an explicit model of the individual level, dynamic decision making process for deciding when to enter the military.

Evidence of this type of height-based selection, however, still faces an identification problem due to the perfect collinearity of birth year (cohort), age, and calendar date (i.e., $\text{age} = \text{date} - \text{birth year}$). What might appear to be selection biases due to calendar year effects estimated from the simple regression, could instead be attributed to age and cohort effects. Consider an admittedly artificial hypothetical height-based selection mechanism. Suppose that 21-year olds from any birth cohort who join the army are always on average 0.5 cm taller than 20-year olds from the same cohort who join the army. Then the difference in the average height between birth cohorts, holding age constant, would reflect true differences in mean height in the birth cohorts' population mean height. The simple regression-based test described earlier, however, would incorrectly indicate biases due to height-based selection.

This example demonstrates that not all samples exhibiting height-based selection will provide biased estimates of how mean height varies in the population across birth cohorts. The cliometrician could argue that there are no calendar year effects consistent with age-based selection, and instead attribute all variation around birth cohort specific means to age effects. But this argument becomes less defensible in comparing cross-cohort differences in the posited age-specific mean height differentials. That is, suppose the difference in height between age 20 and age 21 for birth cohort b is different than it is for birth cohort $(b+c)$, then the cliometrician would need to make a compelling argument for why focusing on height differentials at a particular age, say age 21 rather than age 20, would capture the true difference in mean population height across birth cohorts.

One solution to this knotty problem is to ignore it, by assuming that a random sample from each birth cohort is observed at each calendar date at which the cohort could enter an observed sample. This is a very strong assumption, and any violation of it would be exacerbated as a cohort ages. In military samples, for example, height is typically measured at age of enlistment; entering the military at age 20 typically precludes one

from being in the population at risk to enter the military at age 21 or later. Consequently, if there was height-dependent selection into the sample at younger ages, this would affect the at-risk population's distribution of height at older ages.

RECRUITMENT-YEAR EFFECTS IN THE HISTORICAL HEIGHTS LITERATURE

Most heights studies use cohort dummies or similar controls. The justification for including birth cohorts is well understood, and is cogently articulated by Boris N. Mironov (1999, pp. 3–4). The forgoing discussion, however, suggests the inclusion of a set of recruitment-year (or observation-year) effects as well. This has been much less common in the literature. Brennan, McDonald, and Shlomowitz (1997, pp. 199–200) recognize that the inclusion of recruitment-year effects will capture “varying recruitment conditions ... [and] is the preferred specification because it ... allow[s] for variation in the recruiting environment.” They document instances in which the inclusion of recruitment-year effects, in addition to birth-year effects, reduces the magnitude and statistical significance of birth-year effects for Indian workers emigrating to Mauritius. In his study of early-modern French volunteers, Komlos (2003, p. 167) reports the results of some preliminary regressions that include only enlistment-decade effects; he finds statistically significant coefficients of relatively large magnitude (–0.88 cm to 2.52 cm). But because he observes an inconsistent pattern between youth (less than 23 years) and adults (23 to 49 years) during a given decade he is reluctant to attribute the estimated effects to either changes in measurement techniques or to changes in the supply or demand for height. “While it is imaginable,” Komlos (2003a, p. 167) writes, “that somewhat taller men entered the army during economic downturns, the inconsistencies across age groups lead us not to attribute much significance to this result.” His finding, however, points to a form of selection on unobservables.

Only a handful of heights studies control for both birth-year and recruitment-year effects. They include some research on Union Army soldiers (Margo and Steckel 1983; A’Hearn 1998; Haines 1998). In each article, the inclusion of recruitment-year effects reduces the magnitude and statistical significance of the birth-year effects. Brian A’Hearn (1998) recognizes that the recruitment-year effects are probably indicative of height-based selection. Note, however, that the perfect collinearity of birth year, age, and calendar year implies that one cannot interpret recruitment year effects as capturing the effects of variations in the recruiting environment without assuming the absence of age effects.

TESTING FOR SELECTION IN FOUR SAMPLES

In this section we apply the selection diagnostic described earlier to some of the height samples that form the backbone of the industrialization puzzle. The puzzle first appeared in volunteer military samples and as Komlos (1996, p. 202) notes, “subsequent research has reproduced these results many times over: among the free blacks of Maryland [and Virginia], among Georgia convicts, ... among Amherst students ... and among Pennsylvanian [Union Army] soldiers.” We use these, or closely related samples to investigate whether these samples exhibit selection and whether selection might be the source of the observed decline in heights in the mid-nineteenth-century United States and Britain.

We implement the idea of including both cohort and age or observation year effects in regressions to test for the presence of height-based selection in two different ways. We first ask whether age at recruitment appears to affect the heights of men born in a given year. This test can (falsely) yield evidence of selection. For example, suppose that men who join the army at age 21 are always taller than those who join at age 20, and by the same amount across all birth cohorts. The simple version of the test will incorrectly conclude there is height-based selection at the birth cohort level under this assumption, even when there is not. But we can almost entirely rule this case out by using a more exacting test that allows all age-at-recruitment effects to vary by birth cohort. This test includes in the regression all possible interactions of birth-year and age-at-recruitment dummies. We then test the null hypothesis that any height variations by age at recruitment are constant across birth cohorts. A rejection of this null implies height-based selection that cannot be corrected for in a simple fashion.

The same intuition suggests an alternative specification of the more exacting test; we interact dummies for all birth years with dummies for all calendar years, and test the null that the calendar-year effects do not vary by birth cohort. The exacting, interaction test avoids the shortcomings of just including calendar-year effects in the standard birth-year effects regression, which, without interactions, implies that all individuals entering the sample in a given year are either taller or shorter than the true birth-cohort-specific mean height by the same magnitude. Once again, rejection of this null generally implies the height-based selection implied by the Roy model, as individuals from different cohorts are responding differently to the calendar year influences.

These tests ask whether there is a type of homogeneity in the age- or calendar-year patterns of recruitment across birth cohorts. While it is possible to imagine selection that would be homogeneous in this

sense, the absence of this homogeneity implies selection patterns that are complex and not easily averaged out over successive cohorts. Regardless of which of the two versions of the more exacting test one might use, any behavioral model of enlistment decisions that is consistent with its null hypothesis requires implausibly rigid behavior, namely, that deviations from each true birth cohort's mean height at each age are identical across all cohorts or that the deviations from each true birth cohort's mean height are identical in each calendar year across all birth cohorts.

It is important to recognize that whenever there are multiple reasons for the non-homogeneity of observed heights within birth cohorts, then simple tests like the ones we use need not necessarily reveal whether average heights for birth cohorts would or would not capture the true across-cohort changes in population height. It could be the case, for example, that non-homogeneous selection biases at different ages within birth cohorts would cancel out when one averages across age at enlistment within each birth cohort. Even if that were not the case, any bias that remains after such averaging could potentially be constant across birth cohorts, implying that trends in height by birth cohorts could be identified. (It is also theoretically possible for there to be homogeneity for all ages (years) except for that associated with the excluded age (year) category, but non-homogeneous selection for that excluded category. In this instance one would fail to reject the null hypothesis and attribute incorrectly the non-homogeneous selection to true birth cohort effects.) While one cannot rule out such extreme cases, they do depend upon the true selection effects exhibiting exceptionally restrictive relationships across ages, years, and cohorts. Such singular offsetting effects should be considered unlikely without a moderately realistic behavioral model suggesting that cancellations like these would be likely. Without a detailed model of how observations enter a sample, any variation in observed height within a cohort by age (or, equivalently, enlistment year) for ages after the attainment of terminal height should make one suspicious of the representativeness of the sample.

We report parallel tests of the null hypothesis of no height-related selection for two samples of British soldiers, for three subsets of the Union Army data, for free people of color, and for Pennsylvania prisoners. To restrict the estimation sample to those who have reached full height, the models only use observations for men who are likely to have achieved their terminal stature (age ranges are specified in the tables). To carry out these tests we use OLS with heteroscedasticity-consistent standard errors for the free slave and prisoner samples, and reduced sample maximum likelihood estimators (RSMLE), with the truncation points reported in the tables) for the military samples.

Table 4 reports the results for the two British Army samples. The data reject the homogeneity hypothesis for every model. Table 5 reports the results for three subsets of the Union Army (Fogel et al. 2006). These subsets differ by how they measure the soldier's age at enlistment. The first relies on an integer age variable directly available in the Union Army data set; here, birth cohort is defined as enlistment year minus the reported age at enlistment. The second constructs an (integer) age at enlistment from the difference between the date of enlistment and the birth date. The third subset discards all observations for which these two integer age measures do not agree; the sample size with this restriction is fairly small.

As Table 5 shows, when age at enlistment is taken as that reported in the data, the exacting interaction-based tests (Models 3 and 4) fail to reject the null of no selection. When the age at enlistment variable is calculated from the difference between enlistment and reported birth year, the exacting tests reject the null of no selection at conventional levels. The final column, which includes only the smaller set of soldiers whose reported age is consistent with the constructed age, also fails to reject the null of no selection. The Union Army results may reflect the fact that the recruitment period spans just five years, that the samples are relatively small, or that a relatively large fraction of the recruits and/or recruiters misrepresented the recruits' ages.

Figure 3 reports a graphical warning based on the Union Army results. A model with birth-year dummies only implies that soldiers born in the 1840s were indeed shorter than those born in the late 1830s. But this result is less pronounced in models that control for recruitment year as well. The standard error bands imply that we cannot reject the null of no change in height for the cohort born in 1831 relative to those born in the early 1840s. The figure underscores results such as A'Hearn (1998) and Michael Haines (1998), which also report small cohort effects once recruitment year is controlled for in a regression.

Of the sources that have figured heavily in the industrialization puzzle literature, prison samples number second only to military samples.¹ We ask whether there is evidence of selection in a typical sample of convicts incarcerated during the era of early industrialization, drawing on data from the Pennsylvania penitentiary system between the late 1820s and the late 1870s. Prisoners, especially those confined to state penitentiaries in the nineteenth century, were unlikely to represent random draws from the wider population. Prisoners might not even be representative

¹ For studies of U. S. prisons, see Komlos and Coclanis (1997); Carson (2006, 2008, 2009); Maloney and Carson (2008); Tatarek (2006); Sunder (2004). Nicholas and Steckel (1991) and Nicholas and Oxley (1993) investigate heights using prison records from Great Britain. Twrdek and Manzel (2010) use heights from Peruvian prisons.

TABLE 4
BRITISH ARMY AND AMD

Diagnostic likelihood ratio tests, RSMLE using cohorts and observation years.			
	British Army		AMD
Ages included	23–27		23–25
Observations	5,879		147,882
Model 1: Adding Age Dummies to a Model with Only Birth-Cohort Dummies			
Chi Square Statistic: 4 age dummies	16.04	2 dummies	184.05
P-value	0.0030		<.00001
Degrees of freedom	[4]		[2]
Model 2: Adding Observation Year Dummies to a Model with Only Birth-Cohort Dummies			
Chi Square Statistic: 92 year dummies	194.68		509.19
P-value	<0.0001	34 dummies	<.00001
Degrees of freedom	[92]		[34]
Model 3: Adding Age by Observation Year Dummies to a Model with Birth Cohort and Age Dummies			
Chi Square Statistic: 228 age by observation dummies	379.75	66 dummies	406.90
P-value	<0.0001		<.00001
Degrees of freedom	[228]		[66]
Model 4: Adding Birth by Observation Year Dummies to a Model with Birth-Cohort and Observation Year Dummies			
Chi Square Statistic: 140 year by observation dummies	201.11	34 dummies	81.75
P-value	0.0005		<.00001
Degrees of freedom	[140]		[34]

Notes: Following Komlos (1993) we restricted this sample to the Army only, discarding Marines. “AMD” refers to the annual reports of the Army Medical Department for the years 1864–1913. Floud, Wachter, and Gregory (1990) used these reports but did not include them in the public use sample. Army data: Always lower truncated at 66.99999 inches from below and above 80 inches. AMD Minimum Height Restrictions: 1879–1882: 65”; 1883–1888:63”; 1889–1897: 64”; 1898–1908: 63”; 1909:64”; 1910–1913: 63”. Upper truncation at 73” imposed. Similar rejections if one imposes a 66” minimum height restriction across all years. Heights in AMD dataset are only in one inch ranges, so we use a RSMLE approach that adjusts for heights in ranges.

Sources: British Army comes from the public-use sample associated with Floud, Wachter, and Gregory (1990) (available at the U.K. data archive, SN 2131-2134).

of criminals. The imprisoned arrived after traversing a criminal process that required several decisions by different agents: individuals chose to (allegedly) commit a crime; the police chose whether to arrest and charge the suspect; the prosecutor chose whether to prosecute the case; a judge and jury chose to convict and to impose a sentence of more than one year of incarceration. Ultimately, men committed to the state prisons

TABLE 5
UNION ARMY

Tests for Non-Homogenous Heights RSMLE Estimates: Ages 23–30 With 3 Different Age Definitions for the Union Army Data			
	Reported age at enlistment	Constructed age at enlistment	Reported and constructed ages agree
Sample Size	7,458	3,245	1,695
Model 1: Adding Age Dummies to a Model with Only Birth Cohort Dummies			
Chi-Square Statistic: 7 age dummies	55.94	18.06	12.14
P-value	<0.0001	0.0117	0.0959
Degrees of freedom	[7]	[7]	[7]
Model 2: Adding Enlistment Year Dummies to a Model with Only Birth Cohort Dummies			
Chi-Square Statistic: 4 year dummies	52.27	17.53	9.13
P-value	<0.0001	0.0015	0.0579
Degrees of freedom	[4]	[4]	[4]
Model 3: Adding Age by Enlistment Year Dummies to Birth and Age at Enlistment Dummies			
Chi-Square Statistic: 21 unique dummies	26.91	39.74	27.89
P-value	0.1739	0.0080	0.1432
Degrees of freedom	[21]	[21]	[21]
Model 4: Adding Birth by Enlistment Year Dummies to Birth Cohort and Enlistment Years			
Chi-Square Statistic: 24 unique dummies	30.57	40.27	30.91
P-value	0.1666	0.0200	0.1565
Degrees of freedom	[24]	[24]	[24]

Sources: Authors' calculations from Union Army Project data (Fogel et al. 2006). RSMLE estimates of Likelihood Ratio Test Statistics. Minimum height restrictions: 64.5" August 1861 and earlier; 63" September 1861–December 1864; 60" post 1864.

were those who were convicted of relatively serious crimes. Bodenhorn, Carolyn Moehling, and Gregory Price (2012), in fact, show criminals were short relative to their contemporaries and that shorter men entered prison at younger ages. The mean age at admission into the Eastern and Western penitentiaries was 28.5 years, and ages ranged from 11 to 89 years. Criminologists identify the prime offending ages from the mid-teens to the mid-twenties, which is consistent with the historical data as well. Because less-privileged individuals tended to not reach their terminal adult heights until age 20 or later and, because immigrants faced different childhood environments, we limit the sample to native-born men between 23 and 50 years.

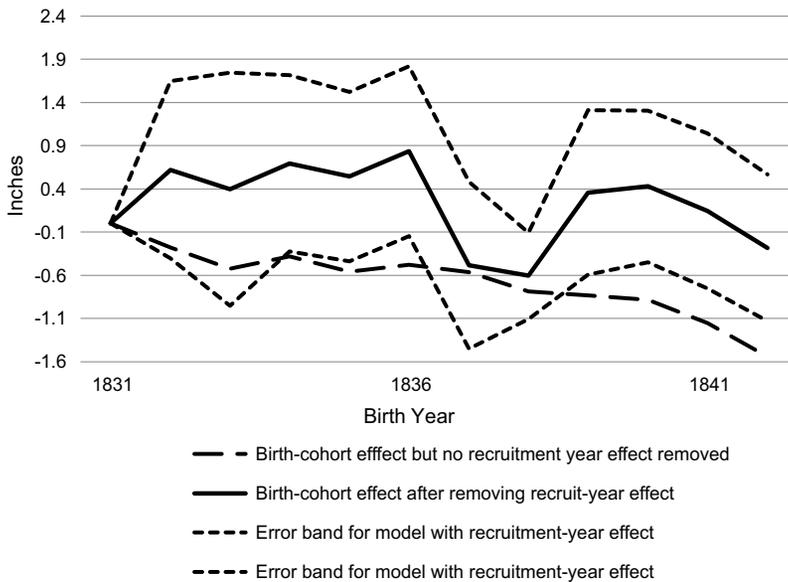


FIGURE 3
 BIRTH-COHORT AND RECRUIT-YEAR CONTROLS: RSMLE ESTIMATES
 Union Army, Recruits Aged 23–30

Sources: Authors’ calculation from ICPSR series # ICPSR06837-v6.

The first column of Table 6 reports results of our four models, including the more exacting specifications (Models 3 and 4). When the regressions include age by observation year dummies or birth cohort by observation year dummies they reject the homogeneity assumption for models based on enlistment year at p-values of 0.02 and 0.06 respectively.

Another group that has figured prominently in the puzzle literature is free blacks from the Chesapeake region (Komlos 1992; Bodenhorn 1999). Virginia’s and Maryland’s “black code” required all free and manumitted African Americans to register with the local county clerk. Any noncompliant free person risked arrest and jailor’s fees, which might be expected to have encouraged near universal registration because the law was enforced, even if unevenly. But only a fraction of African Americans actually registered perhaps as few as one-third (Bodenhorn 1999, p. 980). The literature on manumission gives good reason to think that freed slaves themselves would not represent a random draw from the population of African-Americans born into slavery, so we have several reasons to think that samples of free blacks would suffer selection on unobservables. A second feature of Virginia’s 1793 act imposed a \$5 fine (per act) on any employer who hired a free person of color without a proper registration. This provision might lead to selective registration;

TABLE 6
PENNSYLVANIA PRISONERS AND CHESAPEAKE FREE BLACKS

Tests for Non-Homogenous Heights: Ages 23–30		
PA Prisoners and Free Black Samples		
	PA Prisoners	Free Blacks
Sample Size	1,800	4,797
Model 1: Adding Age at Observation Dummies to a Model with Only Birth Cohort Dummies		
F-Statistic: 7 age dummies	1.81	1.39
P-value	0.0812	0.2032
Degrees of freedom	[7,1744]	[7,4725]
Model 2: Adding Observation Year Dummies to a Model with Only Birth Cohort Dummies		
F-Statistic: 44 year dummies	0.94	0.86
P-value	0.5929	0.7563
Degrees of freedom	[44,1707]	[57,4675]
Model 3: Adding Age by Observation Year Dummies to Birth Cohort and Age Dummies		
F-Statistic: 157 age by observation dummies	1.25	1.19
P-value	0.0242	0.015
Degrees of freedom	[157,1587]	[308,4417]
Model 4: Adding Birth Cohort by Observation Year Dummies to Cohort and Observation Year Dummies		
F-Statistic: 120 year by observation dummies	1.22	1.2
P-value	0.0606	0.0182
Degrees of freedom	[120,1587]	[258,4417]

Sources: Authors' calculations from data used in Komlos (1992) and Bodenhorn (2009, 2016).

most free-born registrants appear in the records between the ages of 17 and 25, when young adults typically enter the paid workforce so the unobservable might be paid-labor prospects in the local market, a prospect not unrelated to productivity and, therefore, health and height.

The second column of Table 6 reports tests of homogeneity for free blacks. The exacting, fully interactive specifications reported as Models 3 and 4 reject homogeneity in age by observation and birth-year by observation year specifications, but not for the simpler models that include only age, birth-year, or observation year dummies without the interaction terms (Models 1 and 2). In general, the types of samples on which the industrialization puzzle is based all reveal some evidence of selection on observables. This finding calls into question the stylized fact that industrialization was associated with declining heights.

CONCLUSION

For several decades now, cliometricians have discussed the industrialization puzzle: the apparent finding that human heights declined during periods of rising real incomes. The industrialization puzzle has achieved the status of stylized fact in its depiction of the nature of economic growth, modernization, and urbanization in the mid- to late nineteenth-century United States. The findings for the United States have led heights researchers to look for similar patterns in other countries. The core issues this literature discusses are central to understanding the process of modern economic growth; the “standard of living debate” gets to the heart of how economic growth affects human welfare.

Unfortunately, the heights literature has relied heavily on sources that likely reflect various forms of selective sampling. Volunteer militaries are the most common source. The decision to join the army reflects an individual’s evaluation of his best prospects in life. Those prospects depend on unobserved, individual-specific factors that are a function of the individual’s human capital, and thus likely correlated with height. Thus the heights of recruits at any one time cannot yield unbiased estimates of population heights. In addition, the heights of those in the choice-based sample will react to changing economic conditions in complex ways. The Roy model reported in Bodenhorn, Guinnane, and Mroz (2014) implies that improvements in civilian labor markets will lead to a shorter army. That is, the industrialization need not be a puzzle once one appreciates the consequences of the selection process underlying many heights sources. Given the evidence supporting the existence of selection on unobservables correlated with height, we cannot draw firm conclusions about long-run trends in U.S. or British heights. Height might have declined over time, but that decline is likely much smaller than what has been reported. An increase in heights is also consistent with the available evidence once we consider the impact of selection.

Direct testing of the selection hypothesis requires matching military records, for example, to some other source, a complex process that can induce its own selection problems. Controlling for selection biases when enlistment decisions take place across many points within the lifecycle will be a complex task. The heights data sources, however, do contain internal evidence of sample-selection bias. We develop and report a series of tests that rely on the idea that the decision to join the military, for example, reflects conditions at the time one joins, while, under the basic idea of the heights literature, the forces that determine adult height

reflect events that occurred long before the age people join the military. The tests look for selection that need not “average out” over a cohort’s lifetime; all show that the sources underlying the industrialization puzzle findings cannot yield unbiased estimates of heights or trends in heights.

These diagnostic tests reveal a troubling pattern: in almost all samples, individuals within the same birth cohort who enter the military at different ages have different mean height by age at enlistment (or, equivalently, by enlistment year). Since we only examine ages after reaching full-height, this is *prima facie* evidence of height-based selection into the military. There is something correlated with height that influences the timing of one’s decision to enter the military. Furthermore, our more exacting tests reveal that the magnitudes of these correlations likely differ across birth cohorts. This lack of uniformity across birth cohorts in the deviations of height by age/date at enlistment makes any simple comparison of averaged observed heights across birth cohorts problematic.

Despite the claims made by this literature, the direct evidence for the puzzle is less robust than one would want. A meta-analysis of some 169 samples drawn from 101 studies that document historical heights demonstrates that most findings of a height reversal rely on selected or small samples. In the vast majority of cases where conscription provided something close to a random sample of young men, or the population of young men, heights grew monotonically throughout the nineteenth century. The United States, which is the core example for the puzzle literature, did not have a meaningful draft during the nineteenth century and thus lacks comparable random samples.

Is the industrialization puzzle real? Scholars who believe it is typically point to a range of evidence other than heights to support the findings based on heights alone. Mortality rates remained stubbornly high through the early decades of industrialization, for example, and in some cases actually increased, as cities became larger and less healthy places to live. Real wages rarely fell, but there is reason to doubt that feeble nominal-wage growth protected the lowest strata from the consequences of food-price shocks. The contentious standard of living debate continues for the simple reason that there is evidence of both improvement and deterioration of living standards as part of the process of economic growth. If anthropometric evidence is to contribute to this debate, scholars must bear in mind how sample-selection bias affects their results and colors their interpretations.

REFERENCES

- A'Hearn, Brian. "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*, 250–67. Edited by John Komlos and Joerg Baten. Stuttgart: Steiner, 1998.
- Ayuda, María-Isabel, and Javier Puche-Gil. "Determinants of Height and Biological Inequality in Mediterranean Spain, 1859–1967." *Economics and Human Biology* 15, no. 1 (2014): 101–19.
- Baten, Joerg. "Protein Supply and Nutritional Status in Nineteenth Century Bavaria, Prussia and France." *Economics and Human Biology* 7, no. 2 (2009): 165–80.
- Baten, Joerg, and John Komlos. "Height and the Standard of Living." *The Journal of Economic History* 58, no. 3 (1998): 866–70.
- Bodenhorn, Howard. "A Troublesome Caste: Height and Nutrition of Antebellum Virginia's Rural Free Blacks." *Journal of Economic History* 59, no. 4 (1999): 972–96.
- . "Criminal Sentencing in Nineteenth-Century Pennsylvania." *Explorations in Economic History* 46, no. 3 (2009): 287–98.
- . *The Color Factor: African-American Well-Being in the Nineteenth-Century South*. New York: Oxford University Press, 2016.
- Bodenhorn, Howard, Timothy Guinane, and Thomas A. Mroz. "Sample Selection Bias in the Historical Heights Literature." Cowles Foundation Working Paper, Yale University, 2014.
- Bodenhorn, Howard, Carolyn Moehling, and Gregory N. Price. "Short Criminals: Stature and Crime in Early America." *Journal of Law and Economics* 55, no. 2 (2012): 393–419.
- Brennan, Lance, John McDonald, and Ralph Shlomowitz. "The Heights and Economic Well-Being of North Indians under British Rule." *Social Science History* 18, no. 2 (1994a): 271–307.
- . "Trends in the Economic Well-Being of South Indians under British Rule: The Anthropometric Evidence." *Explorations in Economic History* 31, no. 2 (1994b): 225–60.
- . "Toward an Anthropometric History of Indians under British Rule." *Research in Economic History* 17 (1997): 185–246.
- Carson, Scott Alan. "The Biological Living Conditions of Nineteenth-Century Chinese Males in America." *Journal of Interdisciplinary History* 37, no. 2 (2006): 201–17.
- . "The Effect of Geography and Vitamin D on African American Stature in the Nineteenth Century: Evidence from Prison Records." *Journal of Economic History* 68, no. 3 (2008): 812–31.
- . "African-American and White Inequality in the Nineteenth Century American South: A Biological Comparison." *Journal of Population Economics* 22, no. 3 (2009): 739–55.
- . "Height of Female Americans in the 19th Century and the Antebellum Puzzle." *Economics and Human Biology* 9, no. 2 (2011): 157–64.

- Costa, Dora, and Richard H. Steckel. "Long-Term Trends in Health, Welfare, and Economic Growth in the United States." In *Health and Welfare during Industrialization*, 47–90. Edited by Richard H. Steckel and Roderick Floud. Chicago: University of Chicago Press, 1997.
- Cranfield, John, and Kris Inwood. "The Great Transformation: A Long-Run Perspective on Physical Well-Being in Canada." *Economics and Human Biology* 5, no. 2 (2007): 204–28.
- . "Stayers and Leavers, Diggers and Canucks: The 1914–1918 War in Comparative Perspective." Working Paper, University of Guelph, 2011.
- Deaton, Angus. *The Great Escape: Health, Wealth and the Origins of Inequality*. Princeton: Princeton University Press, 2013.
- Federico, Giovanni. "Heights, Calories and Welfare: A New Perspective on Italian Industrialization, 1854–1913." *Economics and Human Biology* 1 (2003): 289–308.
- Feinstein, Charles. "Pessimism Perpetuated: Real Wages and the Standard of Living in Britain during and after the Industrial Revolution." *Journal of Economic History* 58, no. 3 (1998): 625–58.
- Floud, Roderick, Robert W. Fogel, Bernard Harris, et al. *The Changing Body: Health, Nutrition, and Human Development in the Western World since 1700*. New York: Cambridge University Press, 2011.
- Floud, Roderick, Kenneth Wachter, and Annabel Gregory. *Height, Health and History: Nutritional Status in the United Kingdom, 1750–1980*. New York: Cambridge University Press, 1990.
- Fogel, Robert William. "Nutrition and the Decline in Mortality Since 1700: Some Preliminary Findings." In *Long-Term Factors in American Economic Growth*, 439–556. Edited by Stanley L. Engerman and Robert E. Gallman. Chicago: University of Chicago Press, 1986.
- Fogel, Robert W., Dora L. Costa, Michael Haines, et al. *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, 2000.
- Fogel, Robert W., et al. *Aging of Veterans of the Union Army: Military, Pension, and Medical Records, 1820–1940* [Computer file]. ICPSR 06837-v6. Chicago, IL: Robert W. Fogel et al., University of Chicago, Center for Population Economics [producers], 2000.
- Gallman, Robert E. "Dietary Change in Antebellum America." *Journal of Economic History* 56, No. 1 (1996): 193–201.
- Haines, Michael R. "Height, Nutrition and Mortality." In *The Biological Standard of Living in Comparative Perspective*, 155–80. Edited by John Komlos and Joerg Baten. Stuttgart: Steiner, 1998.
- Heckman, James J., and Guilherme Sedlacek. "Heterogeneity, Aggregation and Market Wage Functions: An Empirical Model of Self-Selection in the Labor Market." *Journal of Political Economy* 93, no. 6 (1985): 1077–25.
- Johnson, Paul, and Stephen Nicholas. "Male and Female Living Standards in England and Wales, 1812–1857: Evidence from Criminal Height Records." *Economic History Review* 48, no. 3 (1995): 470–81.
- Komlos, John. "The Height and Weight of West Point Cadets: Dietary Change in Antebellum America." *Journal of Economic History* 47, no. 4 (1987): 897–927.

Sample-Selection Biases and the Industrialization Puzzle 205

- . *Nutrition and Economic Development in the Eighteenth-Century Habsburg Monarchy*. Princeton: Princeton University Press, 1989.
- . “Toward an Anthropometric History of African-Americans: The Case of Free Blacks in Antebellum Maryland.” In *Strategic Factors in Nineteenth American Economic History: A Volume to Honor Robert W. Fogel*, pp. 297–329. Edited by Claudia Goldin and Hugh Rockoff. Chicago: University of Chicago Press, 1992.
- . “The Secular Trend in the Biological Standard of Living in the United Kingdom, 1730–1860.” *Economic History Review* 46, no. 1 (1993): 115–44.
- . “The Nutritional Status of French Students.” *Journal of Interdisciplinary History* 24, no. 3 (1994): 493–508.
- . “Anomalies in Economic History: Toward a Resolution of the ‘Antebellum Puzzle’.” *Journal of Economic History* 56, no. 1 (1996): 202–14.
- . “Shrinking in a Growing Economy? The Mystery of Physical Stature during the Industrial Revolution.” *Journal of Economic History* 58, no. 3 (1998a): 779–802.
- . “On the Biological Standard of Living of African Americans: The Case of the Civil War Soldiers.” In *The Biological Standard of Living in Comparative Perspective*, 236–49 Edited by John Komlos and Joerg Baten. Stuttgart: Steiner, 1998b.
- . “An Anthropometric History of Early-Modern France.” *European Review of Economic History* 7, no. 2 (2003) 159–89.
- . “Stagnation of Height among Second-Generation US-Born Army Personnel.” *Social Science Quarterly* 89, no. 2 (2008): 445–55.
- . “A Three Decade ‘Kuhnian’ History of the Antebellum Puzzle: Explaining the Shrinking of the U.S. Population at the Onset of Modern Economic Growth.” Working Paper, University of Munich, 2012.
- Komlos, John, and Brian A’Hearn. “The Decline in the Nutritional Status of the U.S. Antebellum Population at the Outset of Modern Economic Growth.” NBER Working Paper No. 21845, Cambridge, MA, 2016.
- Komlos, John, and Joo Han Kim. “Estimating Trends in Historical Heights.” *Historical Methods: A Journal of Quantitative and Interdisciplinary History* 23, no. 3 (1990): 116–20.
- Komlos, John, and Peter Coclanis. “On the Puzzling Cycle in the Biological Standard of Living: The Case of Antebellum Georgia.” *Explorations in Economic History* 34, no. 4 (1997): 433–59.
- Maloney, Thomas N., and Scott Alan Carson. “Living Standards in Black and White: Evidence from the Heights of Ohio Prison Inmates, 1829–1913.” *Economics and Human Biology* 6, no. 2 (2008): 237–51.
- Mandemakers, C. A., and J. L. Van Zanden. “The Height of Conscripts and National Income: Apparent Relations and Misconceptions.” *Explorations in Economic History* 30, no. 1 (1993): 81–97.
- Margo, Robert A., and Richard H. Steckel. “The Heights of American Slaves.” *Social Science History* 6, no. 4 (1982): 516–38.
- . “Heights of Native-Born Whites during the Antebellum Period.” *Journal of Economic History* 43, no. 1 (1983): 167–174.
- Mironov, Boris N. “Diet, Health, and Stature of the Russian Population from the Mid-Nineteenth to the Beginning of the Twentieth Century.” In *The Biological Standard of Living on Three Continents: Further Explorations in Anthropometric History*, 59–79. Edited by John Komlos. Boulder: Westview Press, 1995.

- . “New Approaches to Old Problems: The Well-Being of the Population of Russia from 1821 to 1910 as Measured by Physical Stature.” *Slavic Review* 58, no. 1 (1999): 1–26.
- Mokyr, Joel, and Cormac Ó Gráda. “The Heights of the British and the Irish c. 1800–1815: Evidence from Recruits to the East India Company’s Army.” In *Stature, Living Standards, and Economic Development*, edited by John Komlos. Essays in Anthropometric History, 39–59. Chicago: The University of Chicago Press, 1994.
- . “Height and Health in the United Kingdom 1815–1860: Evidence from the East India Company Army.” *Explorations in Economic History* 33, no. 2 (1996): 141–68.
- Morgan, Stephen L. “Economic Growth and the Biological Standard of Living in China, 1880–1930.” *Economics and Human Biology* 2, no. 2 (2004): 197–218.
- Mroz, Thomas A. “Sample Selection with Multiple Selection Indicators in Two-Step Estimators.” Mimeo, Georgia State University. May 2015.
- Nicholas, Stephen, and Deborah Oxley. “The Living Standards of Women during the Industrial Revolution, 1795–1820.” *Economic History Review* 46, no. 4 (1993): 723–49.
- Nicholas, Stephen, and Richard H. Steckel. “Heights and Living Standards of English Workers during the Early Years of Industrialization, 1770–1815.” *Journal of Economic History* 51, no. 4 (1991): 937–57.
- Popoff, Kiril G. “Résultats des Mensurations Anthropométriques des Conscrits en Bulgarie.” *Bulletin de L’Institut International de Statistique* 22, no. 2 (1926): 383–420.
- Riggs, Paul, and Timothy Cuff. “Ladies from Hell, Aberdeen Free Gardeners, and the Russian Influenza: An Anthropometric Analysis of WWI-Era Scottish Soldiers and Civilians.” *Economics and Human Biology* 11, no. 1 (2013): 69–77.
- Roy, A. D. “Some Thoughts on the Distribution of Earnings.” *Oxford Economic Papers* 3, no. 2 (1951): 135–46.
- Salvatore, Ricardo D. “Heights and Welfare in Late-Colonial and Post-Independence Argentina.” In *The Biological Standard of Living in Comparative Perspective*, 97–121. Edited by John Komlos and Joerg Baten. Stuttgart: Franz Steiner Verlag, 1998.
- . “Status Decline and Recovery in a Food-Rich Export Economy: Argentina, 1900–1934.” *Explorations in Economic History* 41, no. 3 (2004): 233–55.
- Sandberg, Lars G., and Richard H. Steckel. “Heights and Economic History: The Swedish Case.” *Annals of Human Biology* 14, no. 2 (1987): 101–10.
- . “Was Industrialization Hazardous to Your Health? Not in Sweden!” In *Health and Welfare during Industrialization*, 127–59. Edited by Richard H. Steckel and Roderick Floud. Chicago: University of Chicago Press, 1997.
- Schultz, T. Paul. “Wage Gains Associated with Height as a Form of Health Human Capital.” *American Economic Review* 92, no. 2 (2002): 349–53.
- Shay, Ted. “The Level of Living in Japan, 1885–1938: New Evidence.” In *Stature, Living Standards, and Economic Development: Essays in Anthropometric History*, 173–201. Edited by John Komlos. Chicago and London: University of Chicago Press, 1994.
- Steckel, Richard H. “Stature and the Standard of Living.” *Journal of Economic Literature* 33, no. 4 (1995): 1903–40.

- Steckel, Richard H., and Donald R. Haurin. "Health and Nutrition in the American Midwest: Evidence from the Height of Ohio National Guardsmen, 1850–1910." In *Stature, Living Standards, and Economic Development: Essays in Anthropometric History*, 117–128. Edited by John Komlos. Chicago: University of Chicago Press, 1994.
- Stolz, Yvonne, Joerg Baten, and Jaime Reis. "Portuguese Living Standards, 1720–1980, in European Comparison: Heights, Income, and Human Capital." *Economic History Review* 66, no. 2 (2013): 545–78.
- Sunder, Marco. "The Height of Tennessee Convicts: Another Piece of the 'Antebellum Puzzle'." *Economics and Human Biology* 2, no. 1 (2004): 75–86.
- . "Upward and Onward: High-Society American Women Eluded the Antebellum Puzzle." *Economics and Human Biology* 9, no. 2 (2011): 165–71.
- . "The Height Gap in 19th-century American: Net Nutritional Advantage of the Elite Increased at the Onset of Modern Economic Growth." *Economics and Human Biology* 11, no. 3 (2013): 245–58.
- Tatarek, Nancy E. "Geographical Height Variation among Ohio Caucasian Male Convicts Born 1780–1849." *Economics and Human Biology* 4, no. 2 (2006): 222–36.
- Twrdek, Linda, and Kerstin Manzel. "The Seed of Abundance and Misery: Peruvian Living Standards from the Early Republican Period to the End of the Guano Era (1820–1880)." *Economics & Human Biology* 8, no. 2 (2010): 145–52.
- Voth, Hans-Joachim, and Tim Leunig. "Did Smallpox Reduce Height? Stature and the Standard of Living in London, 1770–1873." *Economic History Review* 49, no. 3 (1996): 541–60.
- Weir, David. "Economic Welfare and Physical Well-Being in France, 1750–1990." In *Health and Welfare during Industrialization*, 161–200. Chicago: University of Chicago Press, 1997.
- Wheatcroft, Stephen G. "The First 35 Years of the Soviet Living Standards: Secular Growth and Conjunctural Crises in a Time of Famines." *Explorations in Economic History* 46, no. 1 (2009): 24–52.
- Zehetmayer, Matthias. "The Continuation of the Antebellum Puzzle: Stature in the U.S., 1847–1894." *European Review of Economic History* 15, no. 2 (2011): 313–27.
- Zimran, Ariel. "Does Sample Selection Bias Explain the Industrialization Puzzle? Evidence from Military Enlistment in the Nineteenth Century United States. Working Paper, Northwestern University, 2015.