



THE LONDON SCHOOL
OF ECONOMICS AND
POLITICAL SCIENCE ■

Economic History Working Papers

No: 273/2018

Sample Selection Biases And The Historical Growth Pattern Of Children

Eric B. Schneider
London School of Economics

LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE

DEPARTMENT OF ECONOMIC HISTORY

WORKING PAPERS

NO. 273 – JANUARY 2018

Dr Eric B. Schneider
London School of Economics
e.b.schneider@lse.ac.uk

Abstract

Bodenhorn et al. (2017) have recently sparked considerable controversy by arguing that the fall in adult stature observed in military samples in the United States and Britain during industrialisation was a figment of sample selection bias. While subsequent papers have questioned the extent of the bias (Komlos and A'Hearn 2016; Zimran 2017), there is renewed concern about selection bias in historical anthropometric datasets. This paper extends Bodenhorn et al.'s discussion of selection bias on unobservables to sources of children's growth, specifically focussing on biases that could distort the age pattern of growth. Understanding how the growth pattern of children has changed is important since these changes underpinned the secular increase in adult stature and are related to child stunting observed in developing countries today. However, there is potential for selection on unobservables in historical datasets containing children's and adolescents' height, so scholars must be aware of these biases before analysing these sources. This paper highlights, among others, three common sources of bias: 1) positive selection of children into secondary school in the late nineteenth and early twentieth centuries; 2) distorted height by age profiles created by age thresholds for enlistment in the military; and 3) changing institutional ecology which determines to which institutions children are sent. Accounting for these biases weakens the evidence of a strong pubertal growth spurt in the nineteenth century and raises doubts on some long run analyses of changes in children's growth, especially for Japan.

Keywords: selection bias, child growth, anthropometrics, health history

JEL Codes: I00, N3, O15, J13, C52, C81

Acknowledgments

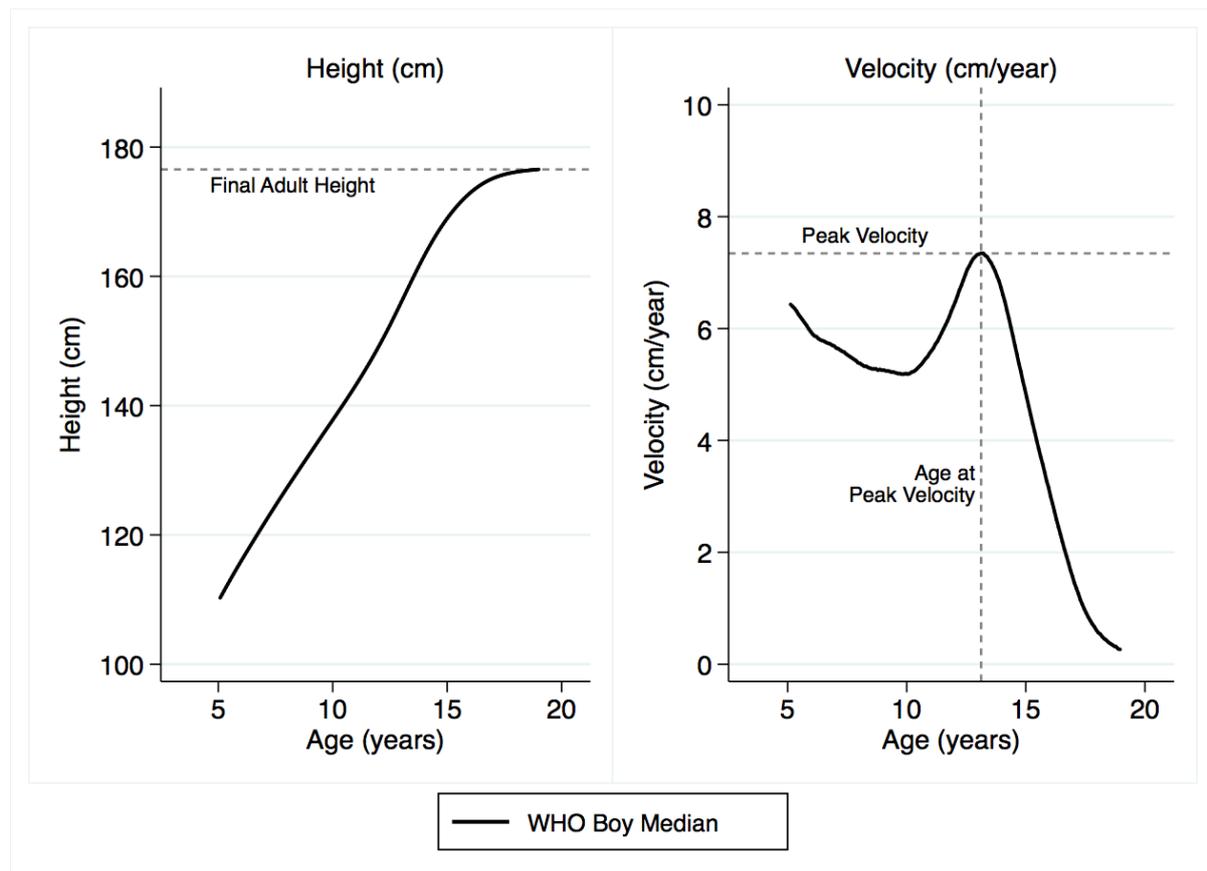
The author wishes to thank Hamish Maxwell-Stewart and Kota Ogasawara for sharing data and Ewout Depauw, Kris Inwood, Hamish Maxwell-Stewart, Kota Ogasawara and Ariell Zimran along with participants at the EHES Conference 2017 for helpful comments on the paper. The usual disclaimer applies. The author gratefully acknowledges funding from ESRC grant ES/L010267/1.

From its birth forty years ago, anthropometric history, the study of human welfare through the analysis of body measurements, has grown and developed into a strong subfield within economic history (Komlos and Baten 2004; Steckel 1995; 2009). It took the tireless effort of many scholars to convince the wider discipline that heights proxied people's welfare. This was especially true when the trends in heights departed from trends in other measures of living standards such as real wages and GDP per capita during industrialisation in the United States and Britain (Floud et al. 1990; Komlos 1993, 1998; Margo and Steckel 1983). In both countries, heights declined during periods when real wages and GDP per capita were growing strongly. This divergence in welfare measures has been explained by the negative health effects of urbanisation and by the cost and availability of food. However, Bodenhorn et al. (2017) have recently called these trends into question. They argue that because the height trends come from data based on military recruits rather than conscripts, unobserved factors leading certain types of people to join the military are influencing the trends. When opportunities were difficult in the civilian labour market, the military was attractive to a wider range of men. However, when economic conditions improved in the civilian labour market, higher quality, taller individuals would join the military at lower rates. This mechanism could explain a decline in heights when economic conditions were improving. Importantly, Bodenhorn et al. argue that this selection mechanism is unobservable, i.e. not correlated with or captured by the typical controls included in height regressions.

Bodenhorn et al.'s working papers (2013, 2014) and final published paper (2017) sparked tremendous debate among anthropometric historians about selection bias and what had and could be done to address their concerns. However, these discussions of selection bias have not been as readily translated to sources of children's growth. There has been extensive discussion about selection bias in the slave manifests used to study the growth pattern of slave children in the US South (Bodenhorn et al. 2013; Calomiris and Pritchett 2009; Pritchett and Freudenberger 1992, 2016; Pritchett 1997; Steckel and Ziebarth 2016). However, selection bias in other sources of children's growth has been relatively understudied. As in all cases, the importance of different forms of selection bias is fundamentally related to the question an author is asking of the data. Thus, if one were examining how children's heights changed over time by analysing the children at one age entering a particular institution over a long period of time, it would be important to consider the types of forces that Bodenhorn *et al.* (2017) discuss. This application would be relatively straight forward, so rather than discussing that in detail, this paper will assess how unobservable selection could lead to biased inferences about the growth pattern of children.

The growth pattern is the age pattern in height and velocity of height across the growing years. It is defined by four key characteristics (see Figure 1): the final adult height, the age at peak velocity during the pubertal growth spurt, the growth velocity during the pubertal growth spurt and the age when growth stops occurring. In general, auxologists and anthropometric historians have found that the growth pattern has changed in four key ways over the past century: adult height has increased (Hatton and Bray 2010; NCD Risk Factor Collaboration 2016); the pubertal growth spurt has occurred at earlier ages; the velocity of growth during puberty has increased; and the growing years have shortened with people reaching their final adult heights at earlier ages (Steckel 1987; Schneider 2017: 23; Tanner 1962:143-55). This pattern seems fairly universal although there are exceptions and the timing and causes of the shift in the growth pattern are not clear (Gao and Schneider 2018).

Figure 1: Characteristics of the Growth Pattern of Boys



Notes: The growth pattern of girls is different than for boys with girls experiencing an earlier and less pronounced pubertal growth spurt, lower velocity and adult height and earlier age when growth stops.

Sources: de Onis et al. (2007); data drawn from <http://www.who.int/growthref/en/>.

The historical research that traced these changes in the growth pattern has generally relied upon sources that provide the heights of children at various ages measured at the same time. Thus, it is important to note that the vast majority of research on children's growth in the past is not based on longitudinal height measures of the same children across their growing years but on the change in height between different groups of children at adjacent ages (though c.f. Arthi and Schneider 2017; Gao and Schneider 2018; Komlos et al. 1992; Schneider 2016). These are known as cross-sectional or period growth curves because they measure an average growth curve across individuals at one particular point in time (McMurray 1996). Because the height profile is then strongly influenced by the children who are measured at each age, most of the potential selection bias is a result of selection on unobservables that changes at different ages.

The paper will proceed as follows. First, I will discuss three sources, school records, prison/criminal records and army records, that have been or could be used to analyse the growth pattern of children and highlight some of the potential forms of selection on unobservables that could influence these types of data. This discussion will be largely hypothetical though there will be some specific examples along the way. In the second part of the paper, I examine how to detect and manage selection bias on unobservables and other sources of bias and measurement error in actual datasets. I first discuss the selection bias diagnostics suggested by

Bodenhorn et al. (2017) and show that these are not appropriate for use in sources of children's height. Then I present various strategies for detecting unobservable selection in sources of children's growth from Japan, Tasmania, Boston, Massachusetts and Britain. I also discuss other sources of bias and measurement error that would influence the growth pattern and make it more difficult to determine whether selection on unobservables was present in the data.

On the whole this paper shows that selection bias can be a problem in sources of children's heights and needs to be considered when analysing an earlier literature that did not account for this as carefully as might be desired. However, selection bias on unobservable characteristics does not render all of these sources unusable. With careful attention to selection processes and analysis of the data, it is possible to determine which datasets are most problematic and in some cases determine ways to use parts of the data while excluding data subject to bias.

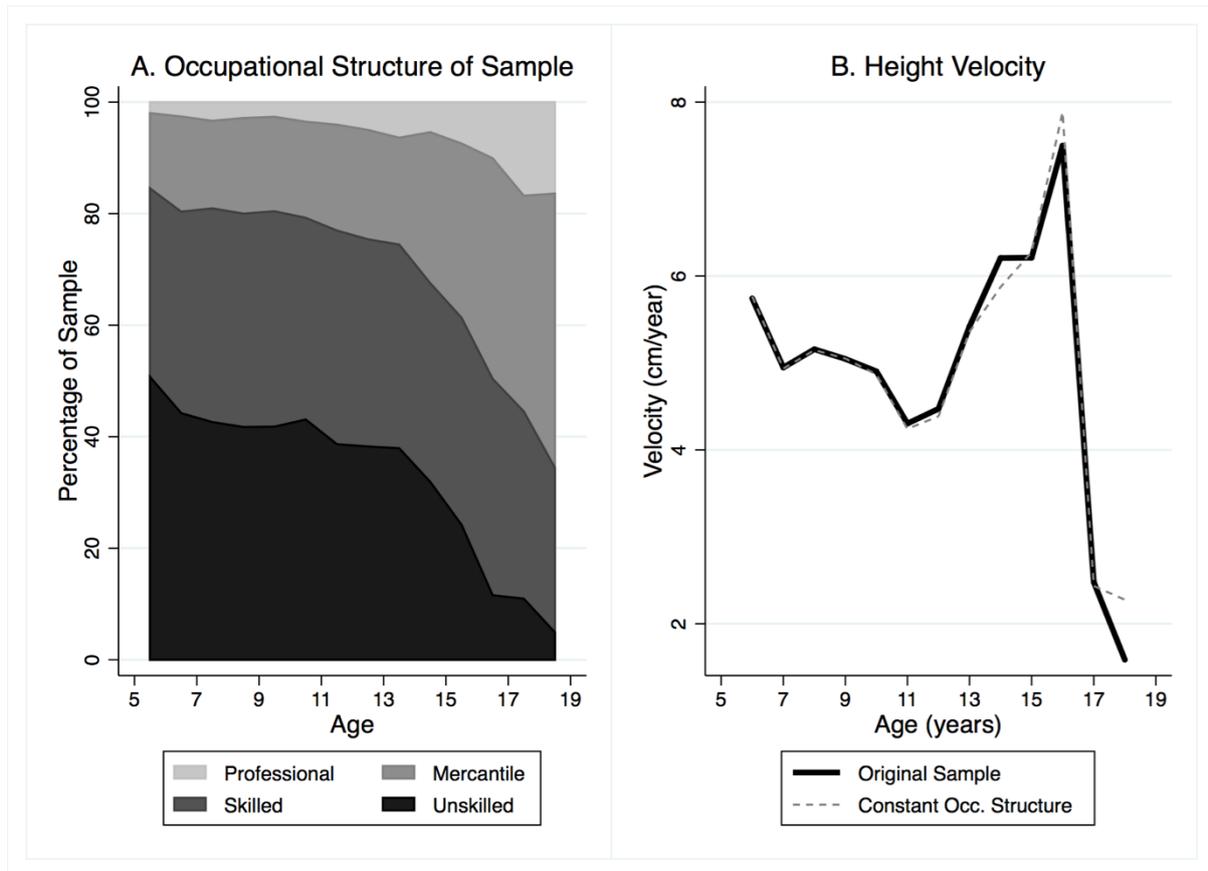
1. Historical Sources of Children's Growth and Potential Selection Bias

Before discussing how to detect selection bias on unobservables with specific data, it is perhaps helpful to discuss important sources of potential selection bias in the typical sources used to reconstruct the growth pattern of children: school records, prison registers and military enlistment or conscription records.

The most prominent set of sources used to reconstruct the historical growth pattern of children are school records. School records have been used to study growth since 1870s when Roberts in the UK and Bowditch in Boston collaborated with schools to collect cross-sectional data on children's heights (Bowditch 1877; 1879; Roberts 1874). These types of studies were replicated across Europe and North America at the end of the nineteenth and in the early twentieth centuries (Burk 1898; Tanner 1981). In addition, state and local bureaucracies also began collecting school children's height information in a number of countries including the UK and Japan in the early twentieth century (Floud et al. 1990: 175-82; Harris 1994; Saito 2003). These early efforts at data collection have served as important sources for long run studies on changes in children's growth (Cameron 1979; Steckel 1987; Tanner 1981).

Unfortunately, the individual-level data underpinning these records are almost entirely lost (though cf. Roberts and Warren 2016), so most of the time the only data available is the average heights and weights of boys and girls at one-year age intervals. At times, the nineteenth-century auxologists also broke down their data according to some characteristics of the children such as their ethnic background and their fathers' occupational status. This is especially important considering that working class children were far less likely to continue in school at later ages than their more privileged counterparts. We can see this in the Bowditch data in the 1870s, which was collected largely from public schools in Boston. In his data, the percentage of children whose father had an unskilled occupation fell from 50 per cent at age five to less than 5 per cent at age eighteen (Figure 2A) (Bowditch 1879: 38-43). Because the average heights are given for each subgroup, it is possible to generate a growth profile and velocity curve that reweights the series for compositional change across observed categories (Figure 2B). As we can see, the composition-adjusted velocity curve is not substantially different from the original curve. However, the bias from compositional change may be small relative to selection on unobserved characteristics.

Figure 2: Occupational Structure and the Growth Pattern of Boys in Bowditch’s 1870s Boston Data



Notes: The constant occupational structure line in panel B was calculated by weighting the occupational group growth profiles with the occupational structure at age 8 at all ages. This approach assumes that the occupational structure for 8-year-olds matched the true population occupational structure, which seems plausible since enrolment rates in Boston were highest at age 8 at nearly 80%.

Sources: Bowditch (1879: 38-43).

The more troubling question is whether the unskilled working-class children remaining in the sample at age eighteen were still representative of the unskilled working class more generally. If the working-class children remaining in school tended to be of higher status or if they were healthier than others in their class, then they would be taller than their average counterpart at the same age. This would lead to upward bias in the growth profile and would tend to overestimate the height interval between adjacent ages as the sample became more positively selected. On the other hand, Bleakley et al. (2013) argue that in the nineteenth-century United States there was not a strong positive relationship between height and human capital. There was no relationship because the opportunity cost of schooling was high since physical labour was important in the economy and men were paid a premium for their brawn. This would suggest that taller and stronger working-class children would be less likely to stay in school, especially as other job opportunities opened up for them and they were no longer legally required to be in school. If this were true, then the children remaining in school at later ages would be negatively selected and the velocity (height intervals at adjacent ages) would be underestimated as the children aged. Thus, the effect of children dropping out of school on the growth pattern is ambiguous and needs to be analysed empirically. The next section will do this using the Bowditch data along with data from Japan.

The second set of records that could be used to analyse children's growth would be prison or criminal records. Somewhat surprisingly to our modern sensibility, nineteenth-century prisons housed numerous children; for instance, in the Wandsworth House of Correction in London in 1860s-1880s, the youngest prisoner was a 7-year-old English boy, and 10 per cent of female and 26 per cent of male prisoners were aged 18 or under (Horrell et al. 2009: 99). In Carson's (2009) penitentiary sample for the United States covering 14 states, 13.5 and 19.1 per cent of white and black male prisoners respectively were in their teens with the youngest being 12 years old. There were also substantial numbers of adolescent convicts in Europe and in Commonwealth countries (Depauw 2012; Inwood and Maxwell-Stewart 2017). Thus, one could reconstruct the growth pattern of children from this data, and historians have started to do this.

However, there are a number of sources of selection that could be problematic for this kind of exercise. Obviously the first would be whether there were changes in the selection into committing crime from childhood through to adolescence. This could be tested on observable characteristics, but we could never entirely rule out all selection processes, though that does not mean that these would lead to significant bias. Another problematic source of selection relates to the variety of institutions available to which children could be sent. In a nineteenth-century British context children were sometimes sent to adult prisons after committing crimes though they were kept separate from the general adult population under the 1823 Gaol Act. However, there were a host of other institutions that juvenile offenders might find themselves sent to depending on the severity of their offence including workhouses, poor law schools, industrial and reformatory schools, and juvenile detention centres (Godfrey et al. 2017: 24-36). Understanding this ecosystem of institutions is particularly important when analysing individual-level information from one institution because the researcher cannot capture the individuals who were sent to other institutions. Changes in the *de jure* or *de facto* rules regulating where children were sent could lead to selection bias that might not be clear by looking at observable characteristics. Thus, it is crucial for historians to learn the finer details of the wider institutional setting and also be wary of signs of changing institutional structures.

Finally, historians studying criminal records have to be cautious about changes in the treatment of children by courts at specific age cut-off points. In the UK the judicial process for children changed substantially across the nineteenth century. In the early nineteenth century children under the age of 14 were tried in the same manner as adults with full jury trials for indictable (serious) offenses. However, after the 1847 Juvenile Offenders Act and the 1879 Summary Jurisdiction Act, children under 12 were tried in a summary court without a jury for all offenses except murder and manslaughter and children under 16 could opt for the summary option as well. These changes in the age cut off and method of trial may have influenced the types of children that ended up in a particular prison and therefore could be an important source of unobserved selection at different ages (Godfrey et al. 2017: 27).

The final set of sources that could be analysed to understand the growth pattern of children are army records. While most enlistment or conscription records would not include many adolescents under the age of 16 or 17, these records could be used to understand the growth pattern after the pubertal growth spurt and notably when individuals stopped growing. A'Hearn et al. (2009) use Italian military registration records to trace changes in the growth pattern from age 17 onward for birth cohorts from 1855-1910. In addition, many studies have noted that military recruits appeared to be growing into their early to mid-twenties suggesting a much longer growing period than is typical of modern, healthy populations (Beekink and Kok 2017; Cinnirella 2008; Floud et al. 1990: 153-54). The literature on military recruits has

always been concerned about selection into their samples with particular focus on minimum height requirements (truncation), the rejection of medically unfit individuals and the selection of recruits into different units and military services (Cinnirella 2008; Floud et al. 1990, 30-83, 118-27). However, there are a few ways in which selection by age might influence the reconstruction of the growth pattern. First, as we will see below, any restrictions placed at specific ages create incentives for men to lie about their age. For instance, if there were a minimum age requirement for enlistment during a period when patriotism was strong, then there would be incentives for men to misrepresent their age in order to enlist at earlier ages.

In addition, as Bodenhorn et al. (2017: 191-93) point out, there could be bias in the age pattern of growth in military samples since recruits at earlier ages are removed from the population at risk of enlistment at subsequent ages. If taller troops are more likely to be removed in the first rounds of recruitment either because of a binding minimum height requirement or another selection mechanism that draws healthier men, then the population at risk of being recruited at later ages would be of lower quality. This effect is very difficult to test with actual data, but a simulation exercise below shows that the bias could be significant.

2. Existing Approaches for Detecting Selection Bias on Unobservables

Having described the hypothetical potential for selection bias on unobservables in sources of children's growth, we can now turn to trying to detect selection bias. Bodenhorn et al. (2017: 190-200) propose a series of tests to determine whether selection on unobservables has biased the inferences from a dataset. These stem from the intuition that in the typical height regression birth year, age and year of measurement are perfectly collinear. In the absence of any selection on unobservables, the predicted height of a single birth cohort should be the same at each age between 23 and 50, the period when adult height is stable, controlling for all other observed characteristics. If the predicted heights in that age range were significantly different from one another, then that would suggest that there was unobserved selection that could bias the results. Likewise, the predicted final adult height (aged 23-50) of a single birth cohort should not vary based on the year in which the cohort was measured (year of recruitment, imprisonment, etc.) controlling for all other observable characteristics. If the predicted height of the cohort changed over measurement years, this would suggest that short run conditions such as the demands of the army, the business cycle or other unobserved factors influenced individuals' probability of being observed within a cohort. Bodenhorn et al. (2017) propose a weak and strong test of this selection. The weak test simply introduces one-year age or measured year dummies and tests their joint significance. The stronger test interacts these one-year age or measured year dummies with all birth cohort dummies, testing whether the relationship between age or measured year and height differs significantly across birth cohorts. Bodenhorn et al. (2017) find that many of the standard datasets used to analyse trends in adult heights in the US and UK fail these tests of selection on unobservables.

These diagnostic tests are interesting and useful, but they are also limited in a few respects. First, although they can provide evidence that selection on unobservables may be present in a sample, this does not necessarily mean that the selection would reverse established trends in the data. Zimran (2017) analysed one potential selection mechanism into the Union Army and found that while selection on unobservables was present in the data, it did not explain away the industrial puzzle or regional pattern in heights in the United States. Komlos and A'Hearn (2016) also analyse the Union Army records and show that the Bodenhorn et al. selection mechanism was not at play in the data. Thus, although some datasets may be subject

to selection on unobservables, that does not mean that all inferences from the data are wrong, especially when trends and patterns are corroborated in other sources that have different selection mechanisms.

Second, these tests will be less helpful when there is age heaping or other measurement error in ages in the data.² Measurement error from age heaping could have two potentially contradictory effects. The measurement error in age could lead to attenuation bias in both the birth cohort and age dummy coefficients in the height regressions. This would therefore increase the probability of making type II errors, i.e. ruling out selection bias on unobservables when it was actually important. On the other hand, if people within a birth cohort with heaped ages were systematically of lower human capital and health ‘quality’ than those who remembered their accurate ages, this could create the illusion of selection on unobservables since people with rounded ages would be shorter than their non-rounded counterparts even within the same cohort. Thus, the effect of age heaping on the Bodenhorn et al. selection bias tests is ambiguous. In the end, it is possible that many sources of adult height would fail the strict selection bias tests that Bodenhorn et al. (2017) propose. However, historians do not have the luxury of returning to the past to collect random samples. Thus, we need to develop ways of working with data that may have some potential bias rather than simply scrapping datasets that fail the Bodenhorn et al. tests and rejecting all findings based on these datasets.

The Bodenhorn et al. (2017) selection bias diagnostics are also unfortunately unable to assess selection on unobservables in sources of children’s growth. This is because we must always include age dummies in the regressions to capture height differences across ages. Finding a changing age pattern of growth across birth cohorts could reflect selection on unobservables similar to what Bodenhorn et al. (2017) suggest, but it could also simply be the result of the well-known change in the growth pattern mentioned above. Converting the children’s heights to height-for-age Z-scores of modern WHO standards would seem to be an easy way to eliminate the need to include age dummies in the height regression. However, the pubertal growth spurt occurred at later ages in historical populations than in modern, healthy populations. Thus, one tends to observe a decline in height-for-age around age 12 for boys and age 10 for girls as the modern children enter their pubertal growth spurt with recovery as the historical children enter their own growth spurt and the modern children grow at lower velocities. This difference in the growth pattern means that the WHO reference produces a distorted height-for-age profile for historical populations and would require age dummies anyway to account for this (Schneider 2016). Thus, we cannot rely on the diagnostic tests developed for adult height datasets to determine whether samples of children’s heights may be problematic.

3. Discovering Selection Bias on Unobservables

Because more precise statistical tests of selection on unobservables do not work with sources of children’s heights, we are left with fewer options in attempting to understand selection on unobservables in these samples. Fundamental for understanding selection bias is to take the time and effort to study the selection mechanisms into a sample very carefully. This

² Age heaping occurs when individuals round their age to the nearest number ending in 0 or 5 rather than reporting their true age. A substantial fraction of individuals in the past heaped their ages, and the extent of age heaping has been used as a proxy for the numeracy and level of human capital of populations in the past (A’Hearn et al. 2009; Hippe and Baten 2012).

section walks through a series of datasets attempting to understand whether and to what extent there is selection bias in each case. Each dataset presents a different method for assessing whether selection bias on unobservables distorts the age pattern of growth. Unfortunately, this does not make for a simple checklist of steps to conduct in order rule out bias. However, I hope that the examples provided will raise potential issues so that researchers are more aware of these in the future. I will first discuss examples of potential selection bias in school sources (which to an extent would also apply to prison samples) before moving on to military sources.

3.1 School Sources

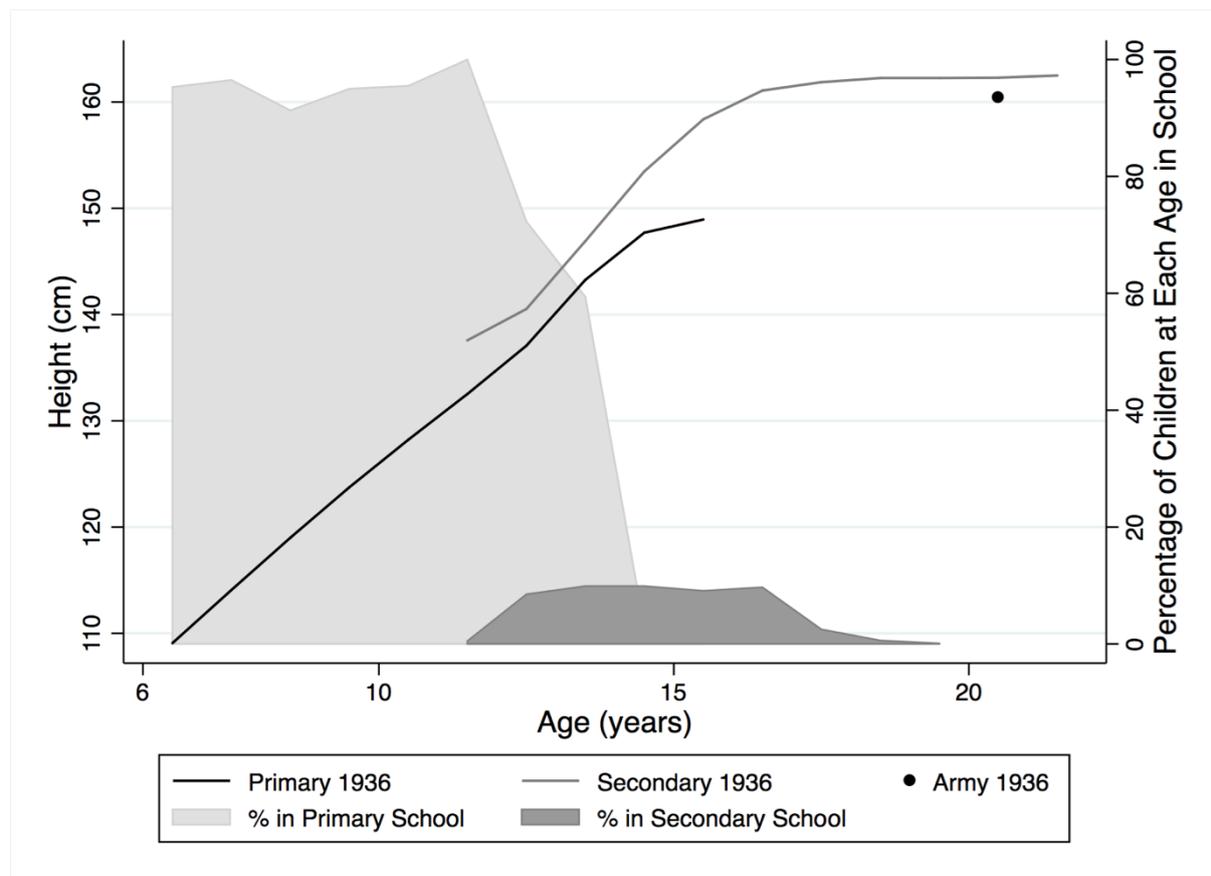
As mentioned above, the most important source of selection bias in school sources is likely to be the selection into secondary school. Typically it is very difficult to assess the extent of selection bias present from this positive selection because we do not have population parameters with which to compare. However, in the case of Japan, the population parameters exist to make the comparison. In the early twentieth century, the Japanese government began recording the heights and weights of schoolchildren in all schools in the country and reporting national averages of heights and weights for boys and girls at one-year age intervals. These national-level period growth curves were reported annually from 1900 to the present and therefore serve as an incredibly detailed set of information on changes in the growth pattern over time (Ali et al. 2000; Mosk 1996). However, because the survey simply measured children in school over time, there is potential for the selection bias described above because the enrolment rate varied across ages. For children ages 6 to 11 in primary school, enrolment was universal capturing around 95 per cent of the population. However, enrolment fell dramatically after age 12 and was never above 10 per cent for secondary schools in the early twentieth century. The increase in secondary school enrolment from 10 per cent to near universal coverage across the twentieth century, then, could substantially distort the observed changes in the pattern of growth.

There is some straightforward evidence that positive selection occurred since the 20-year-old male heights in the school data are substantially higher than the average heights of 20-year-old men conscripted into the military from the same birth cohort: the gap was 1.9 cm for soldiers and students measured in 1936 (see Figure 3). These military heights again covered approximately 95 per cent of the population. However, we can see this process more clearly by looking at children in primary and secondary schools separately. It is not possible to view children in the different schools in the national data, but from 1929 to 1939, the average heights and weights of boys and girls at each age are listed for primary and secondary schools separately at the prefecture level. Thus, it is possible to aggregate up the prefecture level results to the national level using population size as a weight. Figure 3 presents the results for boys measured at various ages in 1936. Clearly, children in secondary school were positively selected since they are taller at all observable ages than children in primary school. At the two ages where the largest number of children are in both schools (12 and 13) the gap in the mean heights is 3.55 cm or 0.48 standard deviations relative to the WHO reference. The growth profile of children in each school is also influenced by selection at different ages. The first very small group of boys to enter secondary school at age 11 were tall for their age even relative to boys who entered secondary school later. We also see that the children remaining in primary school at ages 14 and 15 became more and more negatively selected as the percentage of children in primary schools declined.

Although there is strong evidence of selection bias in the Japanese data, the data highlights the fact that some signs of selection bias do not necessarily require that an entire dataset be discarded. The information on children's heights in primary school before the age

of 14 is of reasonably high quality and could be used to analyse children’s heights. However, adjusting the secondary school data is much more difficult. The fact that the height gap is much larger at ages 12 and 13 than at age 20 suggests that the secondary children were not just taller children growing on the same growth curve as the average child. Instead, the elite secondary school group had an earlier pubertal growth spurt and likely grew at higher velocities across the growing years than the average group. This means that shifting the level of the secondary curve downward to match the heights of the general population at ages 12 or 13 or age 20 would not account for differences in the tempo and velocity of growth between the two groups. Thus, until better methods are developed or other corroboratory data is found, it may not be possible to accurately adjust the secondary school growth profile.

Figure 3: Evidence of positive selection of children into secondary school in Japan, 1936



Notes: My thanks to Kota Ogasawara for his help in extracting this data from Japanese archival sources. Prefecture-level data were aggregated using prefecture population as a weight.

Sources: see Schneider and Ogasawara (2017: Appendix B).

Another important source of selection on unobservables for school and prison data is the institutional ecosystem in which a particular system exists. This may influence which children enter the sample overall, but it also may lead to changing bias over time as the institutional ecosystem changes. One example where the institutional ecosystem influenced the growth profile constructed for children comes from the Marcella Street Home, which was a residential school in late-nineteenth-century Boston. Generally, the Marcella Street Home served pauper children and children whose parents had neglected them and had been sent to the home by the courts. However, Schneider (2016: 292) found that between October 1895 and June 1896 a large number of boys and a smaller number of girls entered the home to serve

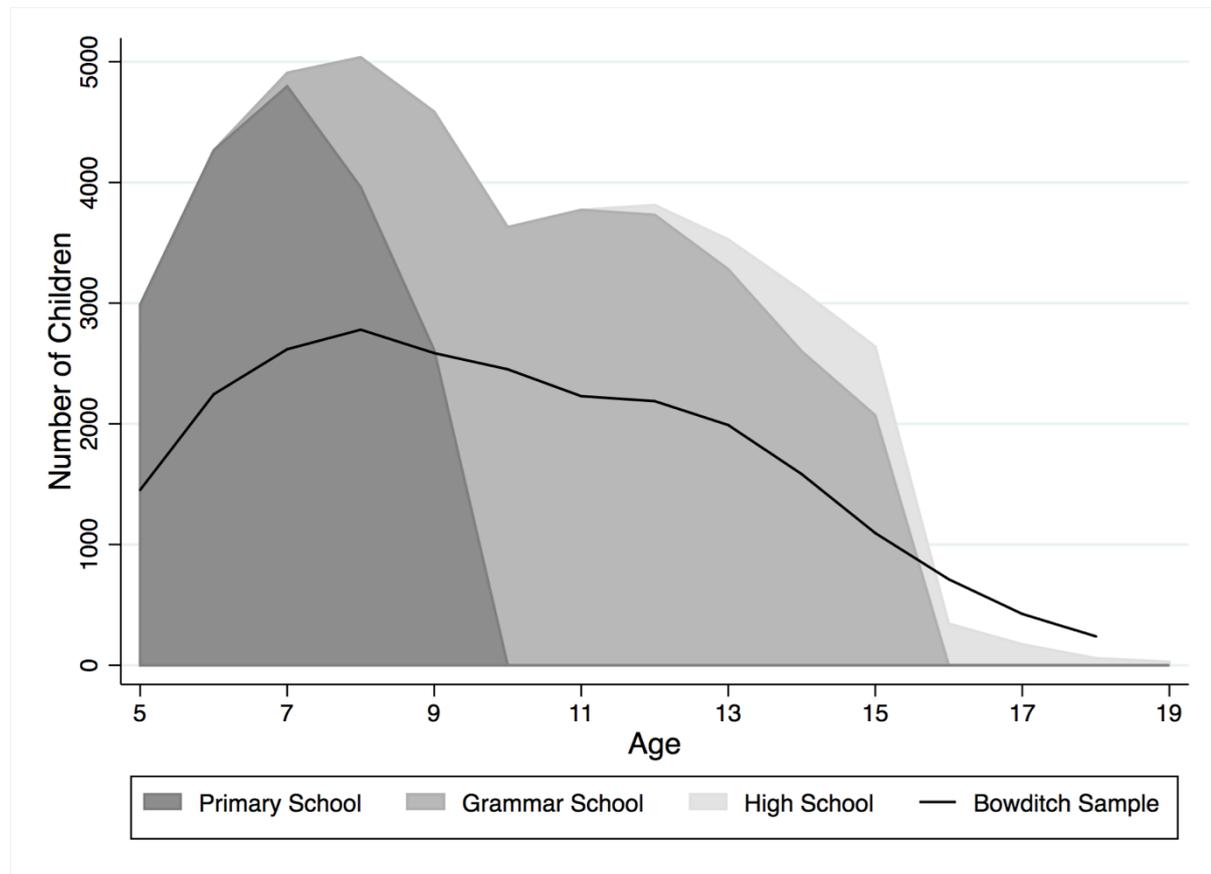
sentences for truancy. These children appear to have been sent to the Marcella Street Home because the city of Boston was attempting to move the location of the truant school. Until 1895, the Boston Truant School was co-located on Deer Island with the House of Industry, a jail for people convicted of minor crimes, and the House of Reformation, which was a juvenile jail. However, the city decided that the truant children should be located elsewhere to reduce the negative spillovers from the other institutions, and so they funded the building of a new school called the Parental School, which opened in late 1895 (Public Institutions Department 1895: 9-19; Public Institutions Department 1897: 14, 18). However, the transfer was drawn out because of construction problems, and the Parental School was overcrowded from the moment it was founded (Public Institutions Department 1895: 16; Institutions Commissioner 1898: 18). Thus, it seems likely that the truant children sent to the Marcella Street Home were sent there because there was not space for them elsewhere in the system of public institutions. The truant children were substantially older than the typical child in the Marcella Street Home and also taller for their age (Schneider 2016: 337). Fortunately, there was an indication that these children were different in the records and the analysis controlled for the truants as a group. Otherwise this influx of a separate population of children could have substantially biased the analysis.

The institutional ecosystem is also important for understanding the reliability of the Bowditch data collected in Boston public schools introduced above. One way to assess whether the children were more positively selected at later ages is to look at how the total enrolment in public schools changed across the school ages and compare this to Bowditch's sample, which was largely drawn from public schools (Bowditch 1877: 7). Figure 4 shows the number of Boston children at each age in public primary, grammar and high schools in 1875. The number of children in public school peaked at age eight and declined sharply between ages 15 and 16 when most children left grammar schools and far fewer entered high schools. In fact, there are more children in the Bowditch sample at ages 16, 17 and 18 than were in all public high schools in Boston, highlighting the fact that Bowditch relied quite heavily on private high schools at these later ages. The sharp drop in enrolment between ages 15 and 16 is doubly suspicious because this is the exact point at which the male velocity curve increases dramatically (Figure 2B above). Thus, the positive selection may mislead researchers about the timing of the pubertal growth spurt.

Staying with the Bowditch data, it is often possible to discover signs of selection on unobservables by looking carefully at the age pattern of growth across the observable categories. For instance, boys whose parents were born in America of the professional and mercantile classes were consistently taller than boys of the skilled and unskilled working classes before the age of 14 (Figure 5A). However, from age 16 onward, the unskilled working-class children were either taller than or equal in height-for-age Z-scores to the mercantile and professional classes. This same pattern is present though less clear for girls whose parents were born in America (Figure 5B). While this is not incontrovertible proof of selection on unobservables, it is highly suspicious given the nature of the selection process discussed above. This potential for unobserved selection bias also makes it difficult to determine whether there is catch-up growth between mid-childhood and adulthood. For both boys and girls the average height-for-age Z-score is higher in adulthood than in childhood before the pubertal growth spurt (age 10 and lower). This could reflect catch-up growth as children in the past had a longer growing period than children in modern populations. However, it could also merely reflect the selection bias on unobservables in the sample where the population of children remaining in school at high ages was positively selected on health. This example suggests that cliometricians should look at the age pattern of growth across observable categories either as raw averages or

by introducing interactions into regression models to check for any suspicious patterns that might not be readily visible in the data.

Figure 4: Number of children in public schools in Boston and in the Bowditch sample, 1875

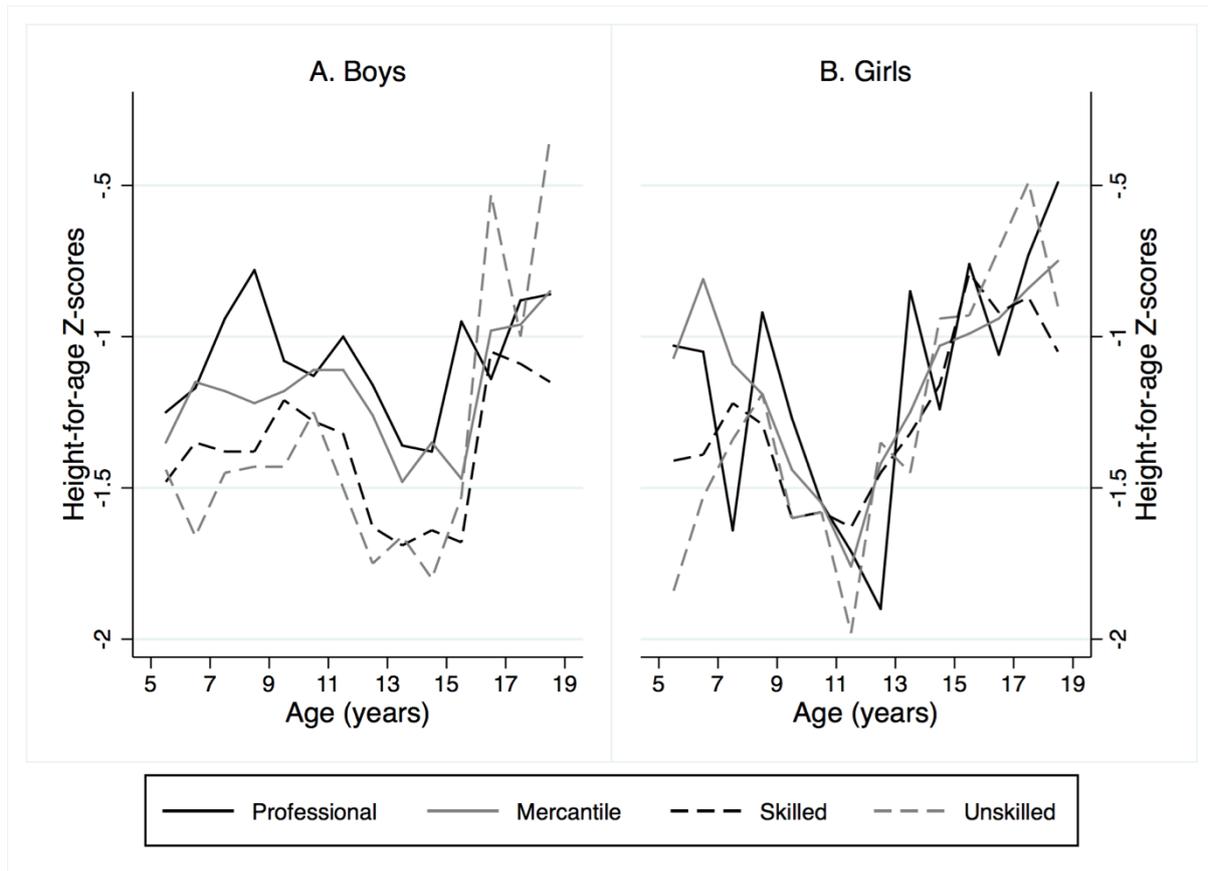


Notes: The sum of primary and grammar school enrolment at age 9 listed in the 1875 Annual Report of the School Committee was larger than the number of nine-year-olds reported in the 1875 census. Thus, the committee seems to have double counted nine-year-olds in both primary and grammar schools. To adjust for this, the enrolment rate at age 8 was used to predict the total number of children enrolled at age 9, and the children were assigned to primary or grammar school proportionately to the figures reported by the committee. This obviously introduces some error, but does not affect the overall trends discussed.

Sources: School Committee (1876: 112-20, 123, 131, 139); The Census of Massachusetts 1875 (1876: 223); Bowditch (1877: 41, 45).

Overall, the evidence presented here for Japan and Boston suggests that there was positive selection on unobservables into secondary school and even into remaining in primary school rather than the opposite effect as might be predicted by Bleakley et al. (2013). Thus, cliometricians and human biologists need to be very careful in using the findings of the nineteenth-century anthropometricians based on school datasets to make claims about the growth pattern of children. The positive selection into secondary school gives the appearance of a more accentuated pubertal growth spurt in these datasets and may distort the age at peak growth velocity during puberty.

Figure 5: Mean height-for-age Z-scores of children from four occupational groups in the Bowditch data from Boston, 1875



Notes: The WHO 2007 growth reference was used to calculate height-for-age Z-scores.

Sources: Bowditch (1879: 38-43).

3.2 Military Sources

When looking at military sources, other sources of selection on unobservables may be more important. As described in general terms above, minimum age requirements may lead individuals to systematically misreport their age in order to enlist at earlier ages than allowed. This type of selection was present in the Australian Imperial Force (AIF) during World War I. Beginning in June 1915, the AIF had strict age restrictions for enlistment. They did not allow soldiers under the age of 18 to enlist and soldiers enlisting under the age of 21 needed permission from their parents. Thus, there were incentives for those under the age of 18 to pretend to be older. These men would likely be taller than the average 17-year-old so that they could pass as 18 more easily but shorter than the average 18-year-old. The same may be true of those under 21. This selection bias, combined with systematic measurement error created by misreporting of ages, would lead to an underestimate of the average height of 18-year-olds, accentuating the growth still occurring after that age.

Inwood and Maxwell-Stewart addressed this by linking the AIF enlistees for Tasmania to their birth records where precise birth date information was available.³ Thus, it is possible to see which soldiers were lying about their age and what the overall impact of their deception would be on the age pattern of height for the soldiers. Across all ages, the reported age and true age were identical for 75.0 per cent of enlistees. There was also considerable random noise in the reporting of age with 17.8 per cent of recruits having a random error, i.e. a difference in reported and true age that was not consistent with the soldier lying about his age to avoid the age requirements. 7.2 per cent of the sample did systematically misrepresent their age so that they could join the force before reaching the required age. This 7.2 per cent may seem too low to influence the results, but the percentage of people systematically lying about their age at certain critical ages was much higher: 35.7 per cent of 18-year-olds and 18.8 per cent of 21-year-olds. Thus, the results could be altered dramatically at those two ages with important effects on inferences for the growth pattern.

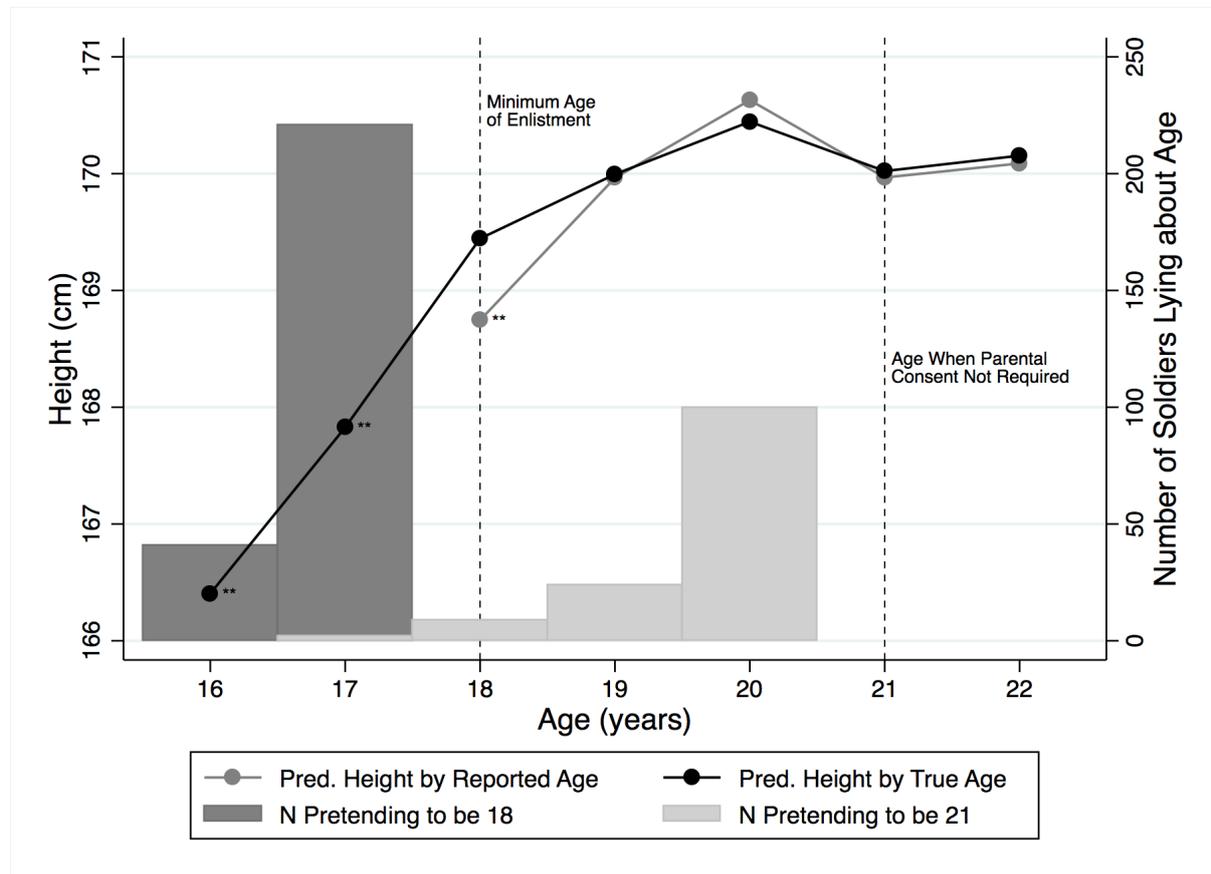
In order to test the effect of this selection bias/measurement error, I conduct two sets of truncated maximum likelihood regressions with the dependent variable (height) truncated at the minimum height requirement, which was 64 inches for the AIF between June 1915 and April 1917 (Whitwell et al. 1997: 415). The regressions include dummy variables controlling for the soldier's father's occupation in twelve HISCLASS categories, the soldier's birth place within Tasmania and the unique month that the soldier enlisted allowing for unobserved changes in recruitment patterns. The variables of interest, however, are the dummies related to age. The first regression specification includes dummies for reported age whereas the second reports dummies for the soldiers' true age. The reference categories for the dummy variables were held constant so that the age pattern of height would be consistent across the two estimations. Finally, because there was considerable random error in the reporting of ages (17.8 per cent of the sample), I did not want this random error to cloud the influence of the systematic selection and error created by people trying to cheat the system. Therefore, in the regressions I only included the 75.0 per cent of soldiers whose age was accurately reported along with the 7.2 per cent who systematically lied about their age. Thus, the differences in height between the reported and true ages reflect solely the influence of the systematic error.

Figure 6 presents the results graphically with the predicted height of soldiers from the regressions shown by reported age and true age. Clearly, once the 16- and 17-year-olds pretending to be 18 are given the correct age in the sample, the predicted height of 18-year-olds increases dramatically by 0.69 cm. 16- and 17-year-olds are much shorter than their 18-year-old counterparts, but their predicted heights were still far above the minimum height requirement of 162.56 cm, so it is plausible that they could pass as 18. The effect of men pretending to be 21 in order to enlist without the permission of their parents is ambiguous. This may be because the majority of these men were 20 years old, and there were relatively small differences in height between 20- and 21-year-olds. However, overall the selection bias from people pretending to be 18 drastically changes the way one would interpret the growth pattern. Based on the reported data, one would argue that there was still substantial growth of greater than one centimetre (1.22 cm to be precise) between ages 18 and 19. After the correction, we see that this growth after age 18 is much smaller (0.55 cm) than we might have otherwise thought. In addition, the height of 18-year-olds is no longer statistically different than the height of 21-year-olds. This suggests that growth was slowing at earlier ages with final adult height reached between ages 18 and 19. The estimates for height at ages 16 and 17 may also be overestimates since presumably taller and more developed 16- and 17-year-olds could more

³ Many thanks to Hamish Maxwell-Stewart for providing the data and the idea to look at this particular selection mechanism.

easily pass for being 18. Thus, this case highlights that whenever there are age thresholds that could encourage individuals to misrepresent their age, it is possible to get biased estimates of the growth pattern.⁴

Figure 6: Predicted heights of soldiers enlisting in the Australian Imperial Force (AIF) by reported age and true age along with the number of men misreporting their age above the given threshold at each age.



Notes: The predicted heights in the regression are predicted from truncated maximum likelihood regressions controlling for father’s HISCLASS, birth location and enlistment month. The omitted category for the regression which applies to the height profile drawn relates to soldiers whose fathers were unskilled labourers, who were born in Hobart and who enlisted in August 1915. ** denotes a point estimate that is statistically significant from age 21 at the 1 per cent level. See text for more detail.

Sources: Hamish Maxwell-Stewart, personal communication.

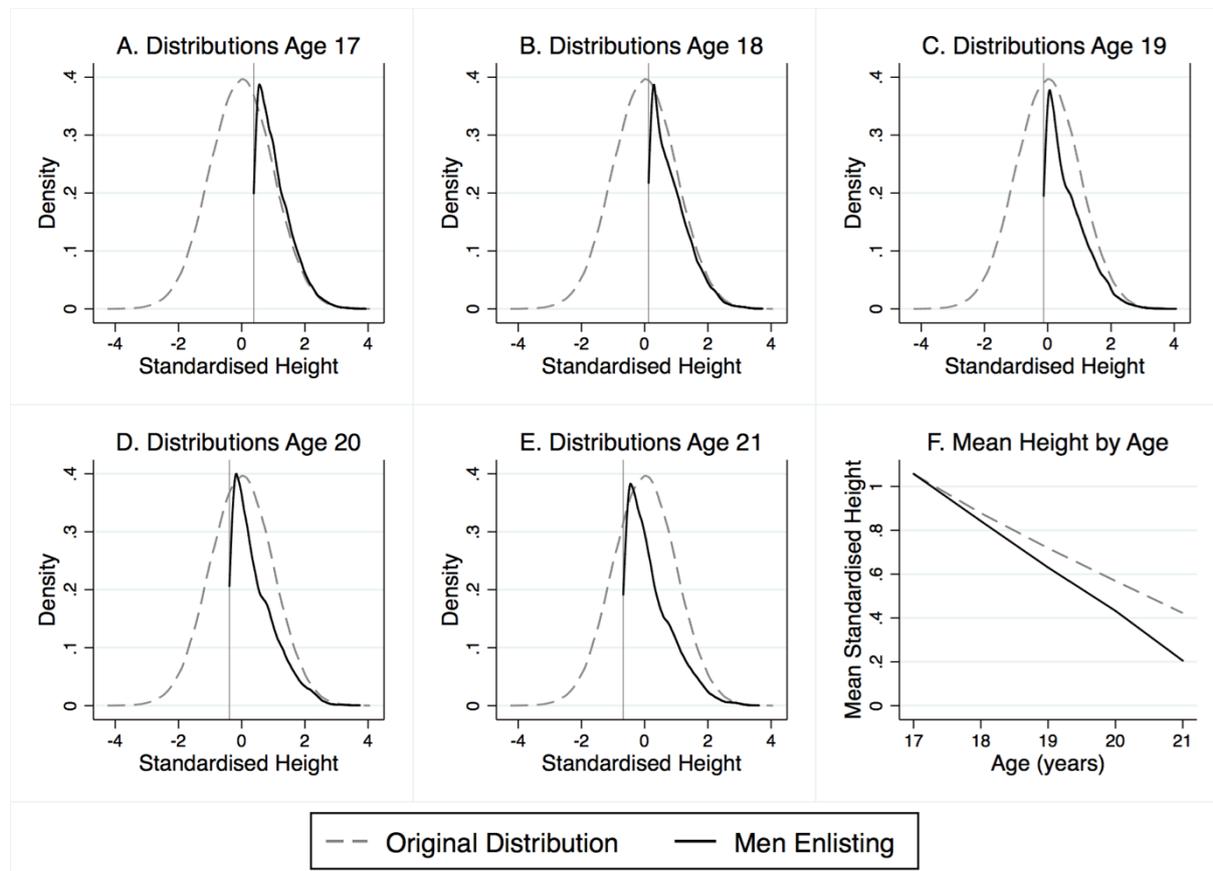
Another potential source of selection on unobservables in military data highlighted by Bodenhorn et al. (2017: 191-93) stems from the fact that recruits say at age 18 are removed from the population at risk of recruitment for a given cohort at age 19 or later ages since soldiers could only enlist once. The removal of ‘high quality’ recruits at younger ages could produce real bias in the age pattern of growth. It is extremely difficult to show this pattern with real data, so I will present simulation evidence to show how this effect works. For the sake of argument, assume that men become eligible to enlist in the army at age 17 and there is one standard and binding minimum height requirement across all ages. The minimum height requirement is at the 65th percentile of height of 17-year-olds in the population and drops 10

⁴ Horrell and Oxley (2016: Appendix A-C) include an extensive discussion of age thresholds and their influence on the growth pattern of factory children in the 1830s. They do not believe that the Horner (1837) data is biased.

percentiles for each year the cohort ages as it grows so that the minimum height requirement is at the 25th percentile of height for 21-year-olds in the population. We assume that the experiment is taking place during a major war where 7.5 per cent of men in a given cohort enlist at each age from 17 to 21. Thus, 37.5 per cent of the total cohort population enlists, which is somewhat less than the 46 per cent of males age 15-49 who enlisted in the army during the first world war in England and Wales (Bailey et al. 2016: 43). Enlistment is determined randomly from men with heights above the minimum height requirement. Once the men have enlisted, they are no longer at risk for enlisting at subsequent ages and are removed from the population distribution at risk for enlistment at the next age. With this information, we can run a simple simulation to see the influence of men enlisting at earlier ages on the growth pattern.

Figures 7A-7E show the standardised height distribution of the population with a mean of zero and standard deviation of one at each age, the dashed grey curve. The vertical grey line is the minimum height requirement. The black distribution is the height distribution of the men enlisting at each age. Figure 7A shows that when the first members of the cohort enlist at age 17, the height distribution of recruits looks very similar to a truncated version of the population distribution. However, as more and more men are removed from the population at risk as the cohort ages, the distribution of recruits shifts to the left of the population distribution and no longer matches the right tail of the population distribution. Although the population mean height is zero across all ages, Figure 7F plots the mean standardised height for the population

Figure 7: Influence of recruits being removed from the population at risk of being recruited at subsequent ages



Notes: Vertical grey lines show the minimum height requirement.

Sources: author's calculations; see text for details.

above the minimum height requirement at each age (dashed grey line). This mean could be corrected to the population mean using truncated maximum likelihood regressions. The mean standardised heights of recruits are given as the black line. Both lines fall across ages as the cohort grows and the minimum height requirement becomes less binding. The two means are the same at age 17 when the first draw occurs, but by age 21, they are 0.24 standard deviations apart which is a substantial difference. Thus, the fact that army recruits are removed from the population at risk at later ages could lead to downward bias in the heights of recruits as the cohort ages, underestimating the growth occurring at later ages.

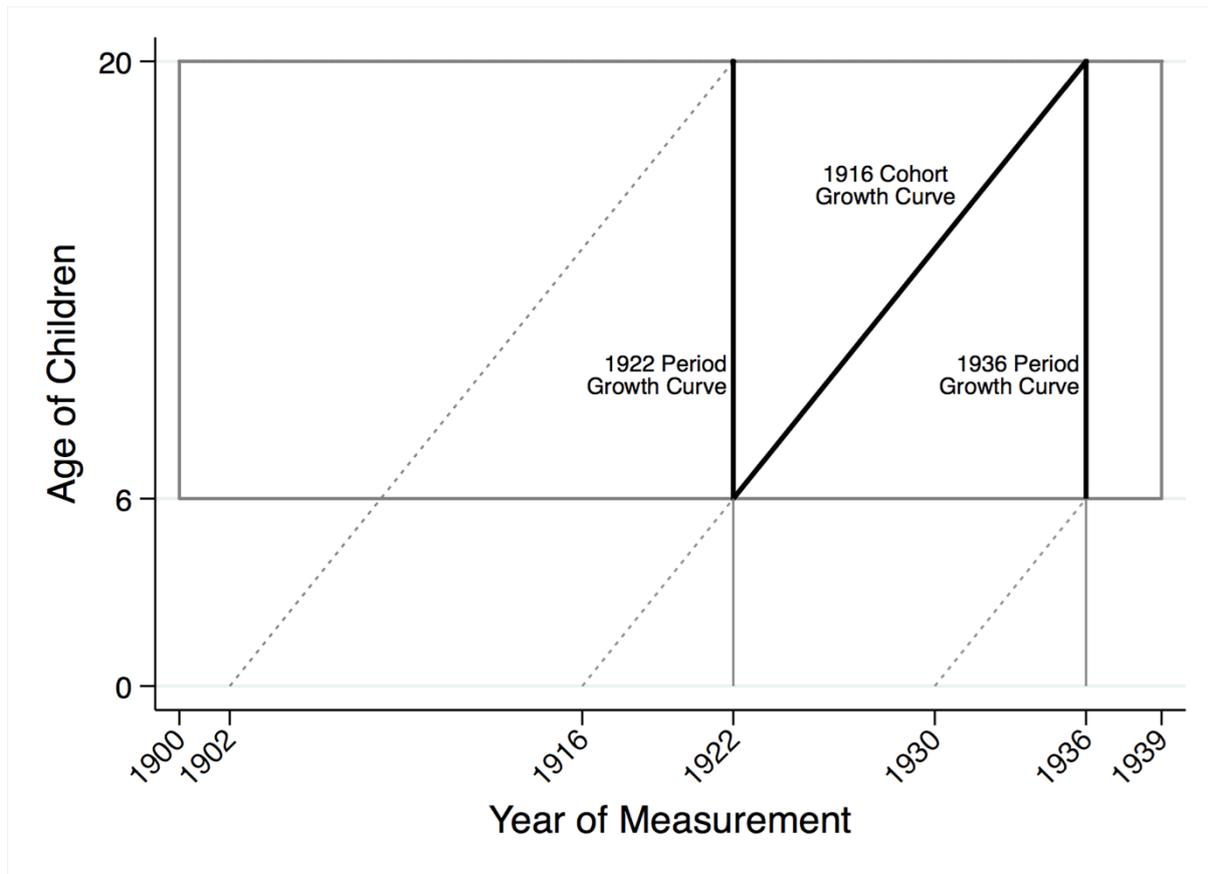
This simulation shows the potential for bias, but it is extremely difficult to adjust for this in real data for a number of reasons. First, the percentage of a cohort that enlisted would not be uniform across ages and would often be lower than 7.5 per cent at each age. If this cohort enlistment rate were very low, then the biases would be much smaller. However, if say 20 per cent of the cohort was recruited at age 18, this would mean the bias would be even more pronounced at ages 19 and onward. Second, minimum height requirements were very rarely as binding as the simulation assumes with many people below the minimum height being allowed into the military. Minimum height requirements were also substantially relaxed during the first world war, which had the high enlistment rates that could produce bias (Bailey et al. 2016: 41; Whitwell et al. 1997: 415). Third, in some cases a large number of men were rejected for service, so there was selection out of the cohort for rejected service as well which would push the mean standardised height upward closer to those for the population. The biggest problem of testing this with real data, though, is that we do not have the population height distribution, so it is extremely difficult to truly understand whether the minimum height requirements were binding enough to produce the stark results from the simulation. Even if the minimum height requirements were not binding, this same bias could exist if selection into enlisting favoured those in the upper part of the height distribution. Thus, this source of selection bias is one that researchers working with military data should consider seriously.

4. Other Biases and Measurement Error

In addition to the issues related to selection bias above, it is also worth noting a few sources of measurement error that can both hide unobserved selection bias and create bias in the growth pattern in their own right. This section discusses bias introduced by period versus cohort growth curves, truncated samples of children's growth and measurement error in ages.

The first measurement related issue relates to the bias associated with mixing many birth cohorts in the typical period (cross-sectional) growth curve used to study children's growth in the past. To illustrate this point, it is easiest to use a real-world example, so here I use the national records of children's growth in Japan mentioned above for the period before World War II. The records report the average heights of boys and girls at ages 6 to 20 in every year from 1900 to 1939. This data can be represented in a lexis diagram to better capture its period-cohort nature. The grey box in Figure 8 shows the age and date range of data that is available. The accented black lines represent different growth curves that could be taken from the data. The vertical lines represent the typical period or cross-sectional growth curves available for historical periods. They reflect children at different ages measured in the same year, in this case 1922 or 1936. The diagonal lines reflect a birth cohort moving through the various ages. The black accented diagonal line marks the 1916 cohort growth curve, which is different than the period growth curve because it follows the same children, those born in 1916, across their childhood and adolescence.

Figure 8: Lexis diagram showing the difference between period and cohort growth curves

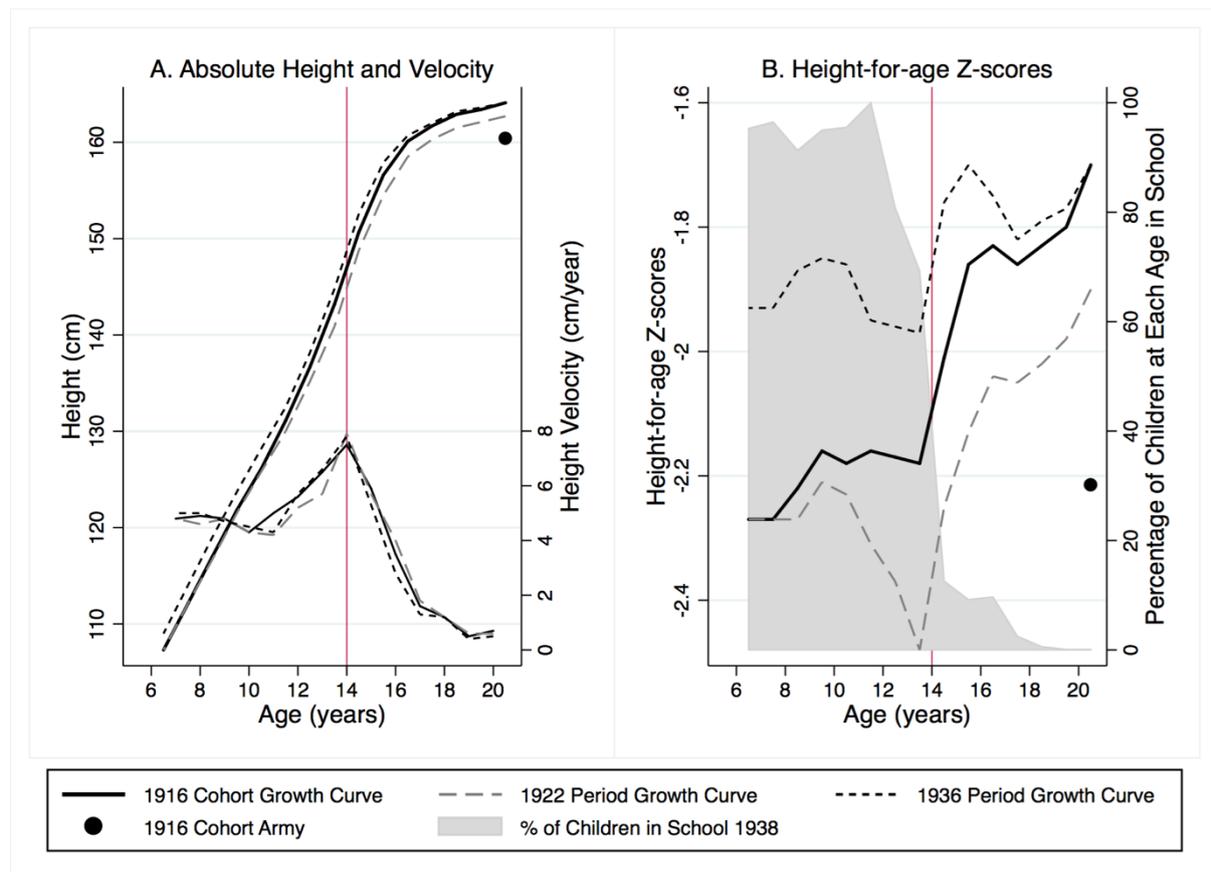


The lexis diagram should immediately highlight one of the main problems with period growth curves: they include children from a very large number of cohorts who may have faced very different conditions in early life. The oldest children in the 1922 period growth curve were born in 1902 whereas the youngest were born in 1916. In a period where heights were increasing over time, this also leads to distortion in period growth curve. It will tend to flatten the height profile since the oldest children are relatively shorter for their age than the youngest children. Figure 9A shows the male period growth curves for 1922 and 1936 and the corresponding male cohort growth curve for 1916 as shown in the lexis diagram above (Figure 8). The same figure could be produced for girls, but the patterns are very similar and it is easier to gauge the magnitude of the selection bias for boys since boys can be compared with male conscripts in the army. The period profiles are clearly flatter relative to the cohort profile. All of the growth curves have heights at age 20 above the level for army conscripts measured in 1936 from the 1916 birth cohort, showing the positive selection of children remaining in school as described above.

The period distortion becomes clearer when looking at Figure 9B. Here the heights have been expressed as Z-scores of the WHO modern growth reference including the heights of conscripts at age 20 in 1936. Looking at the two period growth curves, there is some evidence of positive selection of students into secondary school because the height-for-age Z-scores of children are lower in primary school than at the end of secondary school. However, the full extent of the selection effect is attenuated by the mixing of birth cohorts since heights were increasing during this period. Looking at the 1916 cohort growth curve, we can see that the height-for-age Z-scores of children in primary school (up to age 13) are very similar to the height-for-age Z-score of military conscripts at age 20 for the same birth cohort. However,

beginning around the time that most children left primary school, the height-for-age Z-scores of the 1916 cohort increase dramatically. The effect of children leaving school at these ages also accentuates the pubertal growth spurt more fully than would be the case in the absence of the selection bias. In the end, the actual level of selection bias is far greater in the cohort growth curve (0.5 standard deviations of the WHO reference) than in either of the period growth curves. This is especially problematic for historical research because historians often rely on period growth curves, which may hide the kind of selection bias observed in the Japanese and Boston data in this paper.

Figure 9: Differences between cohort and period growth curves



Notes: My thanks to Kota Ogasawara for his help in extracting this data from Japanese archival sources. The WHO 2007 growth reference was used to calculate height-for-age Z-scores.

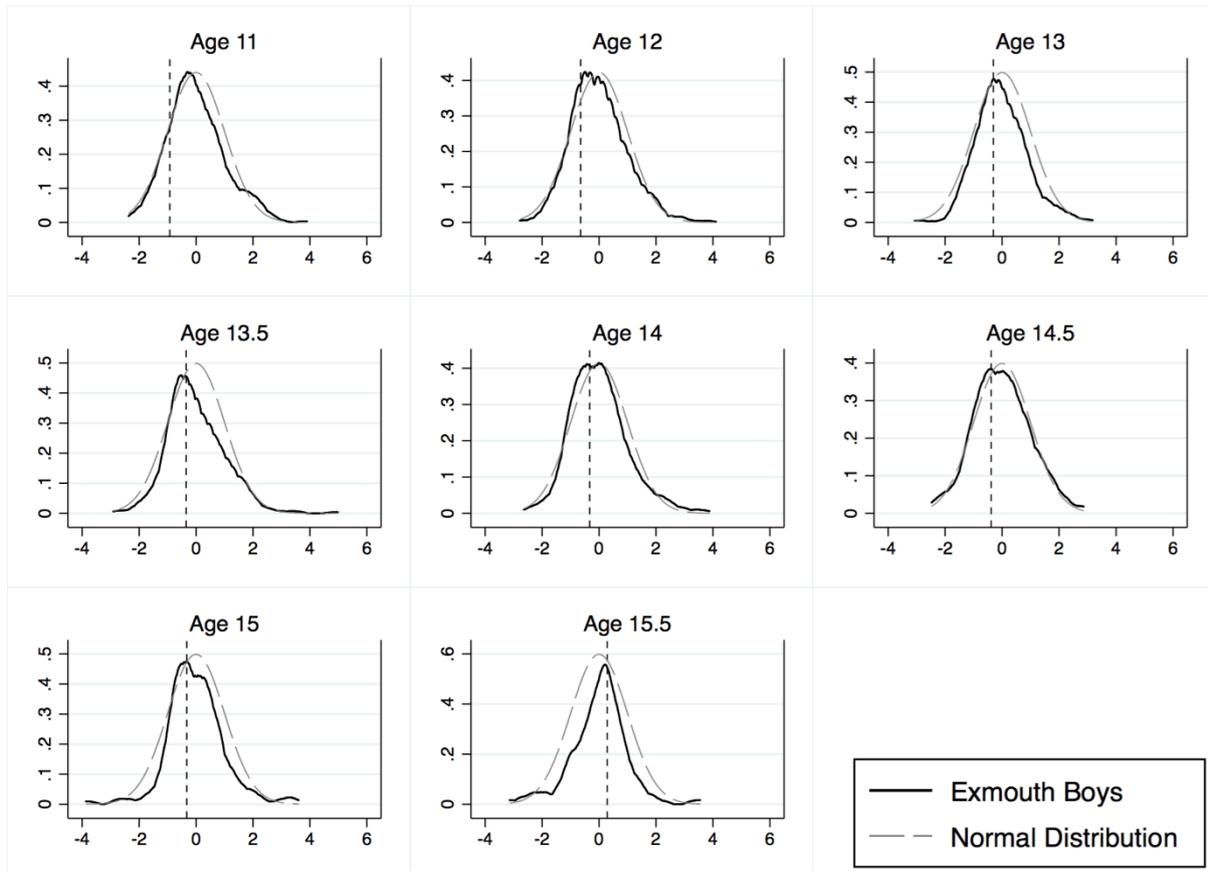
Sources: see Schneider and Ogasawara (2017: Appendix B).

Another frequent issue with data on children's growth is that some institutions that captured children's heights also had minimum height requirements like those discussed for the AIF above. Minimum height requirements were common for training programs for the navy and merchant marine. For instance, in their study of the Marine Society from the mid-eighteenth to mid-nineteenth centuries, Floud et al. (1990: 164-5) found that the society changed its minimum height requirement 13 times. The Marine Society set a minimum height requirement irrespective to age, but the opposite was the case for the training ship *Exmouth*, which varied its minimum height requirements by the age of the children entering. It even produced minimum height requirements by half age from ages 13 to 16 (Thirty-Fifth Annual Report 1911: 5). However, in both the Marine Society and the training ship *Exmouth*, the minimum height requirements were not strictly enforced.

To deal with these minimum height requirements, Floud et al. (1990: 164-5) used the quantile bend estimator (QBE) developed by Wachter and Trussell (1982) to adjust for the left-tail truncation. The QBE assumes that the data should be normally distributed and compares the shape of the sample distribution with a normal distribution in a quantile-quantile plot, finding the point at which the sample diverges from the normal distribution and the sample becomes incomplete. The QBE method works better than the truncated maximum likelihood estimator when the truncation point is undefined or not strictly enforced, but it does not easily allow for multivariate analysis, which might be necessary if controlling for observables were important (A'Hearn et al. 2009: 2). Another crucial problem with the QBE estimator is that we would not actually expect heights to be normally distributed at each age across adolescence because of individual-level variation in developmental tempo. At ages 13 and 14, children developing more quickly would experience the pubertal growth spurt and fill out the right tail of the distribution. At the mid-point of the pubertal growth spurt, the distribution should look more or less normal again. Then, toward the end of the pubertal growth spurt, the distribution should look left skewed as the late developers now stretch the left tail of the distribution. This natural pattern of changing skewness is still present among healthy modern children but was exaggerated in the past because of the greater dispersion in the age at peak velocity during the pubertal growth spurt (A'Hearn et al. 2009: 2; Gao and Schneider 2018). The fact that we would not expect the distributions to be normal also affects the truncated maximum likelihood estimation technique which relies on a normal distribution even though it does a much better job of producing smooth trends and plausible height levels in the Marine Society data (Komlos 2004: 168).

We can see an example of how problematic the normality assumption is by examining data from the training ship *Exmouth*. The data include all boys enrolled on the training ship between 1903 and 1915 when the minimum height requirements described in detail in 1910 were in place (Twenty-Ninth Annual Report 1905: 5; Thirty-Fifth Annual Report 1911: 5). Birthdates and admission dates were available for all boys so we can be reasonably certain that their ages are precisely measured. Figure 10 presents the standardized height distributions with the mean height at zero and a one standard deviation change in the distribution being equal to a one unit increase of the horizontal axis (Z-scores). These are compared with gray dashed normal distributions, and the minimum height requirements are marked by the black dashed vertical lines. Clearly, the minimum height requirements were not binding since at some ages the mean height was actually below the minimum height requirement. We can see some evidence of truncation at ages 11 and 12 before the pubertal growth spurt has begun, but at ages 13 to 14, how is it possible to distinguish between the truncation effect and the expected right skew as the earliest developers experience their growth spurt? Some left skew is noticeable by age 15.5 despite the truncation. Clearly, any estimation strategy that imposes normality on these distributions, such as the QBE or the truncated maximum likelihood estimator, will have difficulty in matching the untruncated distribution. A'Hearn et al. (2009) develop a semiparametric approach that does not require the distributions to be normal, which might be of greater use when dealing with truncated distributions of children's heights, if it could be adapted to account for the truncation points and expected shifting skewness of the distribution.

Figure 10: Standardised height distributions of boys on the training ship *Exmouth* compared with a normal distribution and the minimum height requirement.



Notes: The y-axis shows the kernel density and the x-axis shows the standardised height. The dashed vertical line is the minimum height requirement. Data relate to boys admitted to the ship between 1903 and 1915 when the minimum height requirements were clear and consistent.

Sources: Boys' Record Books, (1876-1915) Training Ship Exmouth, MS MAB/2512, London Metropolitan Archives (LMA), London. Minimum height requirements from Thirty-Fifth Annual Report of the training ship *Exmouth* 1910 (1911) MAB/2554, LMA, London: 5.

A final source of bias is the misreporting of ages. We have seen above that age thresholds can produce substantial bias in the height profile, but any misreporting of ages could affect the height profile if it were prevalent enough. Even random error in ages can significantly affect the growth profile of children. Thus, historical data must be consistently accurate in reporting ages for the data to be of value when looking at long run changes in the growth pattern.

Measurement error in reported age is important because the link between height and age has long been known. In fact, some of the first systematic and large samples of height data collected for British children were collected in an effort to enforce age restrictions on child labour introduced by the Factory Act of 1833. Factory inspectors tasked with enforcing the law needed a way of verifying the ages of children so that factory and mill owners could be held accountable for their labour practices. Therefore, several surgeons began measuring height and tooth eruption as a way of predicting a child's age (Horrell and Oxley 2016: 52-3; Kirby 2013: 99-110). Horrell and Oxley (2016: Appendixes A-C) discuss how the legislation may have created age thresholds in the data but generally find that these were not too problematic.

However, height and age continued to be powerfully linked in the minds of doctors and medical officers of the period and could introduce bias into growth profiles measured in this data.

The training ship *Exmouth*, again, provides useful evidence to test these problems with real data. For the first few years after the *Exmouth* opened (1877-1881), the officers recorded up to two ages for each boy. They always included the age reported by the boy, but for a subsample of boys, the officers also provided their own estimation of the boy's age which they called the 'supposed age'. The officers provided estimated ages for 26.7 per cent of boys entering the *Exmouth* during those years almost always giving the boys a lower estimated age than their reported age.⁵ Thus, the officers clearly thought that some children were far too short to be the age that they reported. However, what is interesting is that the officers did not base their estimations on a simple height threshold. Figure 11A shows the distribution of height-for-age Z-scores of boys whose age was accepted by the medical officer as reliable versus those who were given an estimated age. While the boys who were given an estimated age were shorter than their counterparts whose age was accepted, the distribution of height-for-age for boys given an estimated age was quite wide extending beyond the mean height-for-age Z-score of boys whose age was accepted. Thus, it is not clear how the officers decided who was misreporting (or misinformed about) their age.

The pattern of estimated ages also does not suggest that children were lying in order to enrol on the ship at earlier ages. Although the admission criteria on the *Exmouth* are far from clear in these early days, the ship's administrators took children with reported ages under 12 and the percentage who were given an estimated age increased with the age reported (Figure 11B). Thus, it does not seem that children were misrepresenting their ages to enrol on the ship at earlier ages than regulations allowed. There were no minimum height requirements in place during this period, so that could not have influenced decisions about age reporting. The only criteria might be related to the time spent on the ship before being sent to the navy or merchant marine. Boys entering the ship at later ages spent substantially fewer months on the *Exmouth*, and so would be able to join a merchant ship or the navy more quickly than children entering at younger ages. However, when we compare the time children with reliable ages spent on the ship with those where an estimated age was given (Figure 11C), we see that the ship required the children with estimated ages to remain on the ship for much longer. Thus, there were few benefits for the children from misreporting their ages.

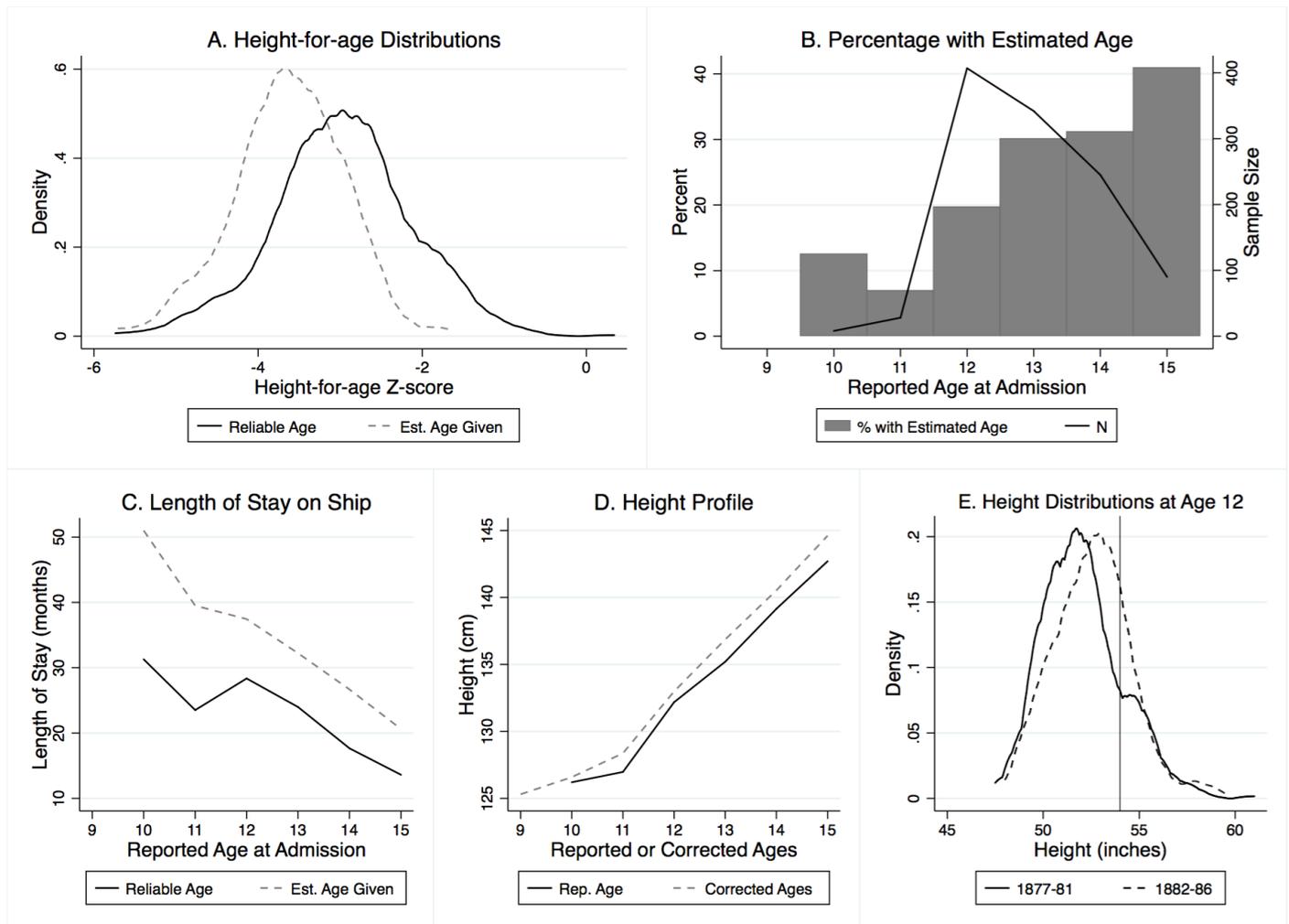
However, this does raise real issues about what a researcher should do with this data. Should the reported ages or the estimated ages be used when comparing these children over time or to other children? Figure 11D shows two growth curves: one based on the reported ages given to all children and another that corrects the reported age to the estimated age for children with an estimated age. Clearly, the average heights of the children increased substantially when conducting this age correction. The average height-for-age Z-score increases from -3.14 to -2.89. So which age measure is correct? On what basis would a researcher make that decision?

This issue also raises questions about how the institution's record keeping practice changed over time. The captain-superintendent of the *Exmouth* first recommended that the ship introduce minimum height requirements in the annual report of 1880 in part because of the supposed error in the reported ages on the ship (Fifth Annual Report 1881: 24-25). If we look at the height distribution of boys reported at age 12 during the period where estimated ages were listed (1877-81) and the period shortly thereafter (1882-86), the height distribution in the

⁵ The officers gave a higher estimated age than reported age in only 0.6 per cent of cases where an estimated age was given.

later period shifts substantially to the right (Figure 11E). This could have been driven in part by the introduction of minimum height requirements, which would have encouraged all involved to report ages correctly. However, these minimum height requirements were not remotely followed during the later period. Figure 11E shows that the minimum height requirement (vertical line) was substantially above the mean and median of the height distribution (1882-86). Thus, it is unclear whether the reported ages became more accurate after 1881 or the person measuring the children's height simply started listing their estimated age as their reported age in the records. This problem becomes less of an issue when birth dates are recorded for the *Exmouth* later in the nineteenth century, but it does raise serious questions about how to interpret changes in mean height and the height distribution between the two periods.

Figure 11: Figures highlighting the effects of misreported ages on the training ship *Exmouth*, 1877-81



Notes: All data is for 1877-81 unless otherwise noted. In panels A and C, reliable age refers to children whose reported age was considered reliable by *Exmouth* officers. Estimated age given refers to children whose age was not considered reliable and were given an estimated age. Panel D shows the height profile for *Exmouth* boys by their reported age and by their corrected age which assumes that all of the estimated ages given by the *Exmouth* officers were correct. Panel E compares the height distribution of 12-year-olds in the early period when estimated ages were given (1877-81) with the heights of 12-year-olds just after that period (1882-86). The vertical line marks the minimum height requirement suggested by the captain-superintendent in 1880, which apparently was not enforced.

Sources: Boys' Record Books, (1876-1915) Training Ship *Exmouth*, MS MAB/2512, LMA, London.; Fifth Annual Report of the training ship *Exmouth* 1880 (1981) MAB/2524, LMA, London: 24-25.

Although this case was specific to the *Exmouth*, this example of potentially misreported ages suggests how important it is to take the time to understand historical height datasets very carefully. It is often possible to detect how the records are subtly changing over time, but this is not possible if all datasets are transcribed by research assistants with little input from the principal investigator. Measurement error in ages may not be detectable at all in some cases, highlighting the importance of triangulating key results with multiple datasets.

In sum, there are a number of potential sources of bias and measurement error not directly related to selection on unobservables that could still influence the growth pattern and diminish a researcher's ability to detect selection bias. Researchers need to consider the potential biases introduced by using period growth curves, truncated samples and samples with measurement error in ages and weigh the strengths of the dataset relative to the potential biases. Unfortunately, this will mean that some datasets are very difficult to work with and may need to be abandoned.

5. Conclusion

This paper has shown how sample selection bias and other sources of measurement error and bias could substantially distort inferences about the growth pattern of children in historical sources. The most important sources of bias raised here are the positive selection of children at later ages into remaining in secondary school, individuals lying about their age around age thresholds, the institutional ecosystem that determines the institutions children end up in and soldiers enlisting at younger ages falling out of the population at risk for enlistment at subsequent ages. Measurement error can also bias the growth pattern, so researchers need to understand the biases that arise from period versus cohort growth curves, truncation created by minimum height requirements and systematic measurement error in ages.

However, the selection biases discussed in this paper do not require substantial changes to the current state of the field for two reasons. First, the changing growth pattern of children is relatively understudied compared with trends in adult stature, so there are fewer studies that could have run afoul. Second, anthropometric historians tend to be fairly careful in their research, so many of the potential issues with existing datasets have already been discussed at length (Horrell and Oxley 2016; Schneider 2016). Instead, I hope that this paper can serve as a guide for those approaching the topic in the future, which should help prevent larger critiques from accumulating as in the industrialisation puzzle.

Having said this, there are three ways in which this paper should lead to revisions or at least inquisitive scepticism toward the existing literature. First, the data that Roberts (1874) and Bowditch (1877, 1879) along with the other anthropometricians at the end of the nineteenth century used to establish a strong pubertal growth spurt are often flawed. Because they drew mostly from public schools, it is very difficult to rule out the positive selection discussed at length in this paper. This does not mean that the pubertal growth spurt did not exist in this earlier period, but we must be incredibly cautious about how we interpret their data. As Tanner (1981: 177-78) noted of Roberts's work, his sample up to age 15 was based on school children, but from 16 to 25 it was based on boys accepted to the army or navy. Tanner argues that taking the modal height for these groups rather than the mean or median would deal with this bias, but that is optimistic. When looking at the growth velocity curve, it is precisely between ages 15 and 16 that the velocity peaks to produce a typical pubertal growth spurt. Thus, further

analysis using longitudinal microdata is necessary to truly understand the growth pattern of children in the nineteenth century (Gao and Schneider, 2017).

In addition, studies of the change in the growth pattern of Japanese children using the height data collected by the Ministry of Education over the twentieth century suffer from serious bias as discussed above (Ali et al. 2000; Mosk 1996). Because the percentage of the population attending secondary school rose from less than 10 per cent to near universal rates, the amount of positive selection into secondary school has changed dramatically over time. Most studies using this data have not taken this into account, though those looking at the post-war period (Bassino and Kato 2010) or using the National Nutritional Surveys (Cole and Mori, 2017) may not suffer to the same extent. Finally, papers that have used truncated maximum likelihood or quantile bend estimation to deal with minimum height requirements for boys during the adolescent years are very likely to produce biased results since we would not expect boys' heights to be normally distributed at these ages (Floud et al. 1990: 164-5; Komlos 2004: 168). Thus, it seems at the moment that anthropometric historians are ill-equipped to deal with truncated samples of adolescent heights. We will need new statistical techniques to overcome these challenges.

References

- A'Hearn, Brian, Franco Peracchi, and Giovanni Vecchi (2009) "Height and the Normal Distribution: Evidence from Italian Military Data." *Demography* 46 (1): 1–25. doi:10.1353/dem.0.0049.
- A'Hearn, Brian, Joerg Baten, and Dorothee Crayen (2009) "Quantifying Quantitative Literacy: Age Heaping and the History of Human Capital." *The Journal of Economic History* 69 (3): 783–808. doi:10.2307/40263943.
- Ali, Md Ayub, Pete E. Lestrel, and Fumio Ohtsuki (2000) "Secular Trends for Takeoff and Maximum Adolescent Growth for Eight Decades of Japanese Cohort Data." *American Journal of Human Biology* 12 (5): 702–12.
- Arthi, Vellore, and Eric B. Schneider (2017) "Infant Feeding and Cohort Health: Evidence from the London Foundling Hospital." CEPR Discussion Paper. No. 12165.
- Bailey, Roy E, Timothy J. Hatton, and Kris Inwood (2016) "Health, Height, and the Household at the Turn of the Twentieth Century." *The Economic History Review* 69 (1): 35–53. doi:10.1111/ehr.12099.
- Bassino, Jean-Pascal, and Noriko Kato. 2010. "Rich and Slim, but Relatively Short Explaining the Halt in the Secular Trend in Japan." Center for Economic Institutions Working Paper Series. No. 2010-5.
- Beekink, Erik, and Jan Kok (2017) "Temporary and Lasting Effects of Childhood Deprivation on Male Stature. Late Adolescent Stature and Catch-Up Growth in Woerden (the Netherlands) in the First Half of the Nineteenth Century." *The History of the Family* 22 (2-3): 196–213. doi:10.1080/1081602X.2016.1212722.
- Bleakley, Hoyt, Dora Costa, and Adriana Lleras-Muney (2013) "Health, Education and Income in the United States, 1820-2000." NBER Working Paper. No. 19162.
- Bodenhorn, Howard, T. W. Guinnane, and T. Mroz (2013) "Problems of Sample-Selection Bias in the Historical Heights Literature: A Theoretical and Econometric Analysis." Economic Growth Center Discussion Paper: No. 1023. Yale University.
- Bodenhorn, Howard, Timothy W. Guinnane, and Thomas A. Mroz (2014) "Caveat Lector: Sample Selection Bias in Historical Heights and the Interpretation of Early Industrializing Economies." NBER Working Paper: No. 19955.
- Bodenhorn, Howard, Timothy W. Guinnane, and Thomas A. Mroz (2017) "Sample-Selection Biases and the Industrialization Puzzle." *Journal of Economic History* 77 (1): 171–207. doi:10.1017/s0022050717000031.
- Bowditch, H. P. (1877) *The Growth of Children*. Boston: Albert J. Wright.
- Bowditch, H. P. (1879) *The Growth of Children: A Supplementary Investigation*. Boston: Rand, Avery & Co.
- "Boys Record Books." (1876-1915) Training Ship Exmouth. MS MAB/2512, London Metropolitan Archives, London.

- Burk, Frederic (1898) "Growth of Children in Height and Weight." *The American Journal of Psychology* 9 (3): 253–326.
- Calomiris, Charles W., and Jonathan B. Pritchett (2009) "Preserving Slave Families for Profit: Traders' Incentives and Pricing in the New Orleans Slave Market." *The Journal of Economic History* 69 (4): 986–1011.
- Cameron, Noel (1979) "The Growth of London Schoolchildren 1904-1966: An Analysis of Secular Trend and Intra-County Variation." *Annals of Human Biology* 6 (6): 505–25. doi:10.1080/03014467900003921.
- Carson, Scott Alan (2009) "Geography, Insolation, and Vitamin D in Nineteenth Century US African-American and White Statures." *Explorations in Economic History* 46 (1): 149–59. doi:10.1016/j.eeh.2008.09.002.
- "The Census of Massachusetts: 1875" (1876) Boston: Albert J. Wright.
- Cinnirella, F. (2008) "Optimists or Pessimists? a Reconsideration of Nutritional Status in Britain, 1740–1865." *European Review of Economic History* 12 (3): 325–54.
- Cole, T J, and H Mori. 2017. "Fifty Years of Child Height and Weight in Japan and South Korea: Contrasting Secular Trend Patterns Analyzed by SITAR." *American Journal of Human Biology* (12): e23054–13. doi:10.1002/ajhb.23054.
- Depauw, Ewout (2012) "Grote Gangsters of Klein Gespuis? De Lichaamslengte in De Gentse Gevangenis in De Negentiende Eeuw." *Handelingen Van De Maatschappij Voor Geschiedenis en Oudheidkunde Te Gent* (66): 145–73.
- "Fifth Annual Report of the Committee for the Training Ship Exmouth 1880." (1881) MS MAB/2524, London Metropolitan Archives, London.
- Floud, Roderick, Kenneth Wachter and Annabel Gregory (1990) *Height, Health and History: Nutritional Status in the United Kingdom, 1750-1980*. Cambridge: Cambridge University Press.
- Gao, Pei and Eric B. Schneider (2018) "The Economics and Biology of Shifting Growth: The Growth Pattern of British Children, 1850-1975." Working Paper.
- Godfrey, Barry, Pamela Cox, Heather Shore and Zoe Alker (2017) *Young Criminal Lives: Life Courses and Life Changes from 1850*. Oxford: Oxford University Press.
- Harris, Bernard (1994) "The Height of Schoolchildren in Britain, 1900-1950," in John Komlos (ed.) *Stature, Living Standards, and Economic Development: Essays in Anthropometric History*. Chicago: University of Chicago Press: 25-38.
- Hatton, Timothy J., and Bernice E. Bray (2010) "Long Run Trends in the Heights of European Men, 19th-20th Centuries." *Economics and Human Biology* 8 (3): 405–13. doi:10.1016/j.ehb.2010.03.001.
- Hippe, Ralph, and Joerg Baten (2012) "Regional Inequality in Human Capital Formation in Europe, 1790–1880." *Scandinavian Economic History Review* 60 (3): 254–89. doi:10.1080/03585522.2012.727763.

- Horrell, Sara, and Deborah Oxley (2016) "Gender Bias in Nineteenth-Century England: Evidence from Factory Children." *Economics and Human Biology* (22): 47–64. doi:10.1016/j.ehb.2016.03.006.
- Horrell, Sara, David Meredith, and Deborah Oxley (2009) "Measuring Misery: Body Mass, Ageing and Gender Inequality in Victorian London." *Explorations in Economic History* 46 (1): 93–119. doi:10.1016/j.eeh.2007.12.001.
- Institutions Commissioner. (1898) "Annual report of the institutions commissioner for the year 1896-7." In *Documents of the city of Boston for the year 1896-7*. City of Boston Archives, Boston, MA, USA.
- Kirby, Peter (2013) *Child Workers and Industrial Health in Britain, 1780-1850*. Woodbridge: The Boydell Press.
- Komlos, John (1993) "The Secular Trend in the Biological Standard of Living in the United Kingdom, 1730-1860." *The Economic History Review* 46 (1): 115–44.
- Komlos, John (1998) "Shrinking in a Growing Economy? the Mystery of Physical Stature During the Industrial Revolution." *The Journal of Economic History* 58 (3): 779–802. doi:10.2307/2566624.
- Komlos, John (2004) "How to (and How Not to) Analyze Deficient Height Samples." *Historical Methods: a Journal of Quantitative and Interdisciplinary History* 37 (4): 160–73. doi:10.3200/HMTS.37.4.160-173.
- Komlos, John, and Brian A'Hearn (2016) "The Decline in the Nutritional Status of the U.S. Antebellum Population at the Onset of Modern Economic Growth." NBER Working Paper. No. 21845.
- Komlos, John, and Joerg Baten (2004) "Looking Backward and Looking Forward: Anthropometric Research and the Development of Social Science History." *Social Science History* 28 (2): 191–210. doi:10.2307/40267839.
- Komlos, J., J. M. Tanner, P. S. W. Davies, and T. Cole (1992) "The Growth of Boys in the Stuttgart Carlschule, 1771–93." *Annals of Human Biology* 19 (2): 139–52. doi:10.1080/03014469200002022.
- Margo, R. A., and R. H. Steckel (1983) "Heights of Native-Born Whites During the Antebellum Period." *Journal of Economic History* 43 (1): 167–74.
- McMurray, C. (1996) "Cross-Sectional Anthropometry: What Can It Tell Us About the Health of Young Children?." *Health Transition Review* 6 (2): 147–68.
- Mosk, Carl (1996) *Making Health Work: Human Growth in Modern Japan*. Berkeley: University of California Press.
- NCD Risk Factor Collaboration (2016) "A Century of Trends in Adult Human Height." *eLife* (5): 1–29. doi:10.7554/eLife.13410.001.
- de Onis, Mercedes, A. W. Onyango, E. Borghi, Amani Siyam, Chizuru Nishida, and Jonathan Siekmann (2007) "Development of a WHO Growth Reference for School-Aged

- Children and Adolescents.” *Bulletin of the World Health Organization* 85 (9): 660–67. doi:10.2471/BLT.07.043497.
- Pritchett, J. B. (1997) “The Interregional Slave Trade and the Selection of Slaves for the New Orleans Market.” *Journal of Interdisciplinary History* 28 (1): 57–85.
- Pritchett, J. B., and H. Freudenberger (1992) “A Peculiar Sample: The Selection of Slaves for the New Orleans Market.” *Journal of Economic History* 52 (1): 109–27.
- Pritchett, Jonathan B., and H. Freudenberger (2016) “A Peculiar Sample: A Reply to Steckel and Ziebarth.” *The Journal of Economic History* 76 (1): 139–62.
- Public Institutions Department (1895) “Annual report of the public institutions department, for the year 1894.” In *Documents of the city of Boston for the year 1894*. City of Boston Archives, Boston, MA, USA.
- Public Institutions Department (1897) “Annual report of the public institutions department, for the year 1896.” In *Documents of the city of Boston for the year 1896*. City of Boston Archives, Boston, MA, USA.
- Roberts, C. (1874) “The physical development and the proportions of the human body.” *St George’s Hospital Reports* (8): 1-48.
- Roberts, Evan, and John Robert Warren (2017) “Family Structure and Childhood Anthropometry in Saint Paul, Minnesota in 1918.” *The History of the Family* 22 (2-3): 258–90. doi:10.1080/1081602X.2016.1224729.
- Saito, O. (2003) “Human growth and economic development: an examination of school physical records, Yamanashi prefecture, Meiji Japan [in Japanese].” *Keizai Kenkyu*, 54 (1), 19–32.
- Schneider, Eric B. (2016) “Health, Gender and the Household: Children’s Growth in the Marcella Street Home, Boston, MA, and the Ashford School, London, UK.” *Research in Economic History* (32): 277–361. doi:10.1108/S0363-326820160000032005.
- Schneider, Eric B. (2017) “Children's Growth in an Adaptive Framework: Explaining the Growth Patterns of American Slaves and Other Historical Populations.” *The Economic History Review* 70 (1): 3–29. doi:10.1111/ehr.12484.
- Schneider, Eric B., and Kota Ogasawara (2017) “Disease and Child Growth in Industrialising Japan: Assessing Instantaneous Changes in Growth and Changes in the Growth Pattern, 1911-39.” LSE Economic History Working Paper. No. 265.
- School Committee (1876) *Annual Report of the School Committee of Boston 1875*. Boston: Rockwell and Churchill.
- Steckel, R. H. (1987) “Growth Depression and Recovery: The Remarkable Case of American Slaves.” *Annals of Human Biology* 14 (2): 111–32. doi:10.1080/03014468700006852.
- Steckel, Richard H. (1995) “Stature and the Standard of Living.” *Journal of Economic Literature* 33 (4): 1903–40. doi:10.2307/2729317.

- Steckel, Richard H. (2009) "Heights and Human Welfare: Recent Developments and New Directions." *Explorations in Economic History* 46 (1): 1–23.
doi:10.1016/j.eeh.2008.12.001.
- Steckel, Richard H., and Nicolas Ziebarth (2016) "Trader Selectivity and Measured Catch-Up Growth of American Slaves." *Journal of Economic History* 76 (1): 109–38.
doi:10.1017/S0022050716000437.
- Tanner, J. M. (1962) *Growth at Adolescence*. Oxford: Blackwell Scientific Publications.
- Tanner, J. M. (1981) *A History of the Study of Human Growth*. Cambridge: Cambridge University Press.
- "Thirty-Fifth Annual Report of the Committee for the Training Ship Exmouth 1910." (1911) MS MAB/2554, London Metropolitan Archives, London.
- "Twenty-Ninth Annual Report of the Committee for the Training Ship Exmouth 1904." (1905) MS MAB/2548, London Metropolitan Archives, London.
- Wachter, Kenneth W., and James Trussell (1982) "Estimating Historical Heights." *Journal of the American Statistical Association* 77 (378): 279–93. doi:10.2307/2287231.
- Whitwell, Greg, Christine de Souza, and Stephen Nicholas (1997) "Height, Health, and Economic Growth in Australia, 1860-1940," in Richard H. Steckel and Roderick Floud (eds.) *Health and Welfare During Industrialization*. Chicago: University of Chicago Press: 379–422.
- Zimran, Ariell (2017) "Does Sample-Selection Bias Explain the Antebellum Puzzle? Evidence from Military Enlistment in the Nineteenth-Century United States." Working Paper.