DISEASE AND CHILD GROWTH IN INDUSTRIALISING JAPAN: ASSESSING INSTANTANEOUS CHANGES IN GROWTH AND CHANGES IN THE GROWTH PATTERN, 1911-39[1]

Dr Eric B. Schneider

Assistant Professor
Economic History Department, LSE

and

Dr Kota Ogasawara

Assistant Professor
Graduate School of Social Sciences, Chiba University

7 June 2017

Preliminary: please do not cite or circulate without permission.

Abstract

This paper assesses how the disease environment in interwar Japan influenced children's growth and health. The data is drawn from government records from 1929 to 1939 which report the average heights of boys and girls in school at each age (6-21) for each of Japan's 47 prefectures. We test the influence of disease in two ways. First, we test the influence of the disease environment at birth, proxied by the infant mortality rate, on the cohort growth pattern of children using the SITAR model to parameterise the growth pattern. In addition, we use a bilateral-specific fixed effects model to understand how disease instantaneously influenced growth controlling for prefecture-birth cohort effects. Our results suggest that health conditions in early life did not have a strong influence on the growth pattern of children in Japan. However, we do find a significant and economically meaningful instantaneous effect of the infant mortality rate on child height at ages 6-11 for both boys and girls. This suggests that child morbidity was very important to the increase in stature during interwar Japan, but it also suggests that the emphasis placed on preventing child stunting in the first thousand days in the modern development literature may be misplaced. The secular increase in height in interwar Japan was more strongly influenced by cumulative responses to the health environment across child development rather than being simply the outcome of improving cohort health.

[1] The authors wish to thank Tim Cole, Yuki Kawakubo and Yukitoshi Matsushita, participants at the CEPR-NYUAD Economic History Workshop 2017 and seminar participants at Hitotsubashi University and the University of Barcelona for helpful comments on the paper. The usual disclaimer applies. Eric Schneider gratefully acknowledges funding from ESRC grant ES/L010267/1 and the Tokyo Institute of Technology. Kota Ogasawara received funding from JSPS KAKENHI Grant Number 16K17153.

Children's growth is an important measure of health and nutritional status that has been employed in the historical literature to track economic development and health (Schneider, 2016, 2017a; Hatton, 2011; Harris, 1995). In addition, stunting rates (the percentage of children below a certain height for age) are also one of the most important indicators of malnutrition used to measure progress in the Millennium Development Goals and Sustainable Development Goals around the world today (de Onis and Blössner, 2003). Reducing stunting rates is in fact manufacturing a change in the growth pattern of children (adult height, timing of the pubertal growth spurt and the velocity of growth). Thus, historical examples of how the growth pattern changed may reveal the kinds of policies and interventions that might help relieve malnutrition in the developing world today. Between the birth cohorts of the 1880s and 1980s, the average height of Japanese adult men increased by 13.9 cm (Baten and Blum, 2012). The average WHO height-for-age Z-score of Japanese six year olds in the period 1929-39 was -2.0, suggesting a stunting rate of around 50 per cent whereas the average in 2016 was -0.46 with a much lower stunting rate. Thus, there was a radical change in the growth pattern in Japan across the twentieth century, making it an ideal case for studying changes in children's growth. Thus, this paper uses Japan as a case study to analyse changes in children's growth pattern and in their instantaneous growth in the interwar period to determine how exposure to disease may have caused these changes controlling for a number of confounding factors.

Drawing on data from Japanese government reports, we have constructed a panel dataset containing the average heights of boys and girls for all 47 prefectures from 1929-39. The data list the heights of primary students and secondary students separately. The primary school averages reflect around 95 per cent of the population from ages 6 to 11 and 50-75 percent at ages 12-13 whereas the secondary school averages capture a much smaller group of children (around 10 per cent). Thus, our data for primary children is largely representative of children in each prefecture in Japan during our time period.

To test the influence of disease on child growth, we implement two study designs: one intended to measure how environmental conditions affect the growth pattern of children and another that analyses instantaneous changes in growth from year to year. In order to measure the growth pattern of the children, we are the first to use the SITAR growth model developed by Cole *et al.* (2008) and Cole *et al.* (2010) to parameterise the growth pattern of historical populations. SITAR predicts a parameter that measures the size, tempo (timing of the pubertal growth spurt) and velocity of each cohort growth curve in our dataset. We then take the predicted random effects for the three SITAR parameters and use these as dependent variables in fixed effects regressions to understand how changes in health conditions in the year of birth affected each SITAR parameter. There are implicit lags in these regressions since we are regressing growth in ages 6-18 on birth year conditions, eliminating potential endogeneity problems. In our second study design, we attempt to understand how child morbidity, proxied by the infant mortality rate, influenced the instantaneous growth of children at ages 6 to 11. We use a bilateral-specific fixed effects model, which includes fixed effects for prefecture interacted with birth year, controlling for any differences in initial birth conditions of the children across prefectures and cohorts. There is greater potential for endogeneity in these instantaneous regressions, but by focusing on mortality of younger children, especially infant mortality, in the same year, we eliminate this potential bias.

Briefly, we find that health conditions in early life did not have a strong influence on the growth pattern of children in Japan. However, we do find a significant and economically meaningful effect of the infant mortality rate on child height in the current year for both boys and girls. This suggests that child morbidity was very important to the increase in stature during

interwar Japan, but it also suggests that the emphasis placed on preventing child stunting in the first thousand days in the modern development literature may be misplaced. The secular increase in height in interwar Japan was more strongly influenced by cumulative responses to the health environment at all ages across development rather than being simply the outcome of improving cohort health. The rest of the paper discusses the historical and biological background of the study, present the data, methods and results, and concludes linking our findings to the wider historical and development literature.

1. Background

Before getting into the specifics of our study, it will be helpful to briefly discuss the secular increase in height and changes in children's growth in Japan. Figure 1 shows the mean male adult height of Japanese men compared with other East Asian countries and the United Kingdom. In the late nineteenth century, Japanese men were shorter than their other East Asian counterparts and 10 cm shorter than men in the UK. However, there was a strong secular increase in height in Japan so that men born in 1980 were nearly 14 cm taller than those born in 1880. This led to considerable convergence with China and Taiwan although South Korea now has the tallest men in East Asia.

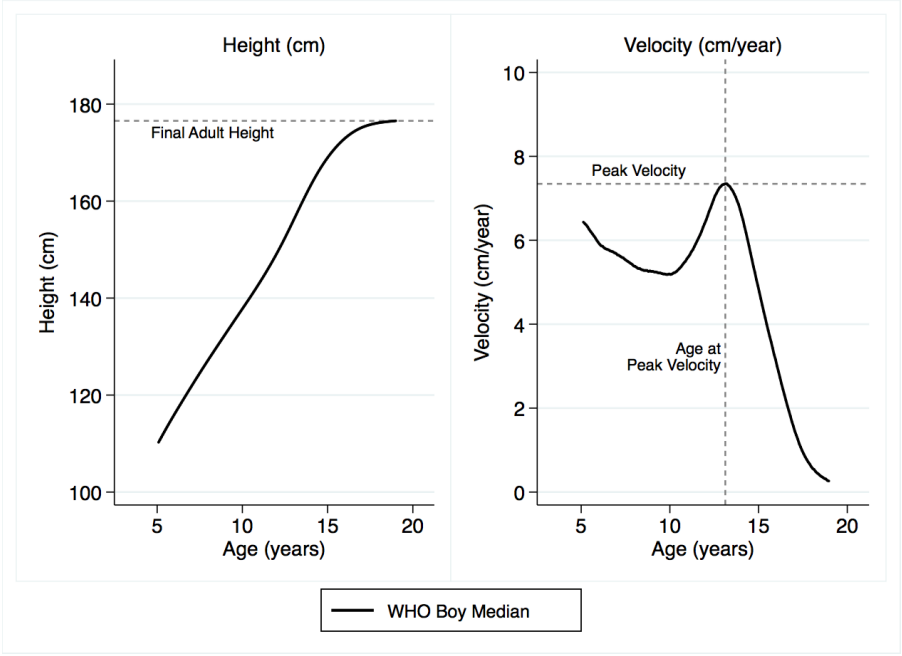
Figure 1: Trends in mean adult male stature in Japan compared with other East Asian countries and the UK.

Source: Blum and Baten (2012).

This secular increase in height also coincided with changes in the growth pattern of children. Figure 2 presents a typical growth curve for a boy, based on the 2007 WHO growth reference, on the left with its corresponding growth velocity curve on the right. The growth pattern in childhood and adolescence, then, is defined by three key characteristics: 1) the timing of the pubertal growth spurt or the age at which a child reaches peak growth velocity in puberty;

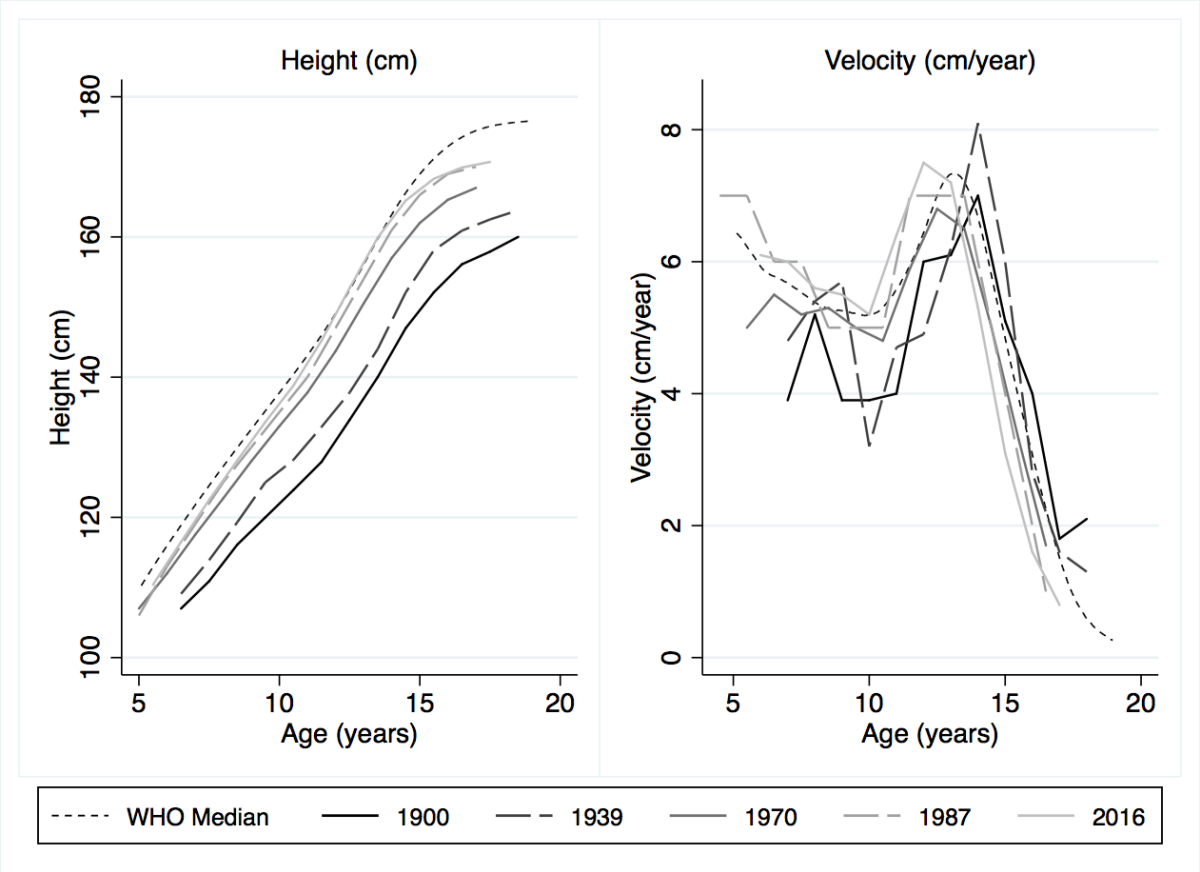
2) the level of velocity at the peak and surrounding the peak; and 3) the final adult height. Figure 3 shows average growth and velocity curves for boys measured cross-sectionally from 1900 to 2016 based on national data collected by the Japanese government. These curves show a definitive change in the growth pattern of Japanese children. Not only did adult stature increase over time, but the age at peak velocity shifted to earlier ages from age 14 in 1900 to age 12 in 2016. The peak velocity during the growth spurt also increased over time. These changes in the three characteristics of the growth pattern are similar to those observed in other countries (Cameron, 1979). Previous historical studies have almost exclusively focused on only one characteristic of the growth pattern, final adult height, so one of the goals of this paper is to understand how socioeconomic and environmental factors influenced all three characteristics of the growth pattern.

Figure 2: Characteristics of a typical growth curve for boys.



We are not the first to study the anthropometric history of Japan (c.f. Mosk, 1996; Saito, 2004; Bassino, 2006) or use the Ministry of Education data on children's heights. Saito (1989) first studied these records finding a significant correlation between changes in the infant mortality rate and children's heights at various ages in 1938. Later, he conducted a more specific study of Yamanashi prefecture in 1902-1906 to look at the differences in children's growth in rural and urban areas, finding small difference in children's heights between towns and villages exposed and not exposed to *Schistosoma japonicum* (Saito, 2003). Finally, Bassino and Kato (2010) have analysed prefecture-level data on children's growth from the post-war period in an attempt to understand why the secular increase in height stopped in Japan in the 1980s. Bassino's (2006) article is also relevant as it shows the importance of income and access to health services for final adult heights in Japan. Finally, Ogasawara (2017a) recently showed that children born during the 1918-20 flu pandemic experienced a long-run height penalty relative to cohorts born shortly before or after the pandemic.

Figure 3: Growth pattern of boys measured in Japan at the following years.



Sources: See data appendix.

These articles are very interesting, but the more direct inspiration for our current study derives from work being conducted on the influence of infant mortality on child growth in Europe. Hatton (2011) pioneered this field by analysing the influence of infant mortality on the growth of children in 40 British towns in the first half of the twentieth century. He noted that the effect of infant mortality in the year of birth on child height at later ages would be a balance of selection, the culling of weak, shorter children, and scarring, the long-lasting negative health effects of higher morbidity on the survivors. He found that the selection and scarring effects balanced each other out when analysing the influence of infant mortality in the birth year on later growth. However, he also looked at the effect of the infant mortality rate that children faced at ages 2-4 on later growth to single out the scarring effect. This alternative measure of infant mortality had strong negative effects on children's heights at ages 6 to 13, suggesting that child morbidity at early ages was very important for growth outcomes in childhood and pre-adolescence. These results were further strengthened by Hatton's (2014) finding that reductions in infant mortality were very important for increases in mean adult stature at the country level in Europe and by Bailey *et al.*'s (2016a) study which showed infant mortality in 1890s Britain significantly shaped the heights of conscripts during the first world war. Stolz *et al.* (2013) did not find a significant effect of infant mortality on adult male stature in a European cross-country panel covering the period 1720 to 1910, but this may be partly due to the fact that they exclude the twentieth century when most of the infant mortality decline and secular increase in height occurred. In any case, no one has been able to test whether these same effects existed outside of Europe or influenced all three characteristics of the growth pattern rather

than height at a specific age. Thus, there are many benefits to analysing this question using the very rich Japanese data that are available.

2. Data

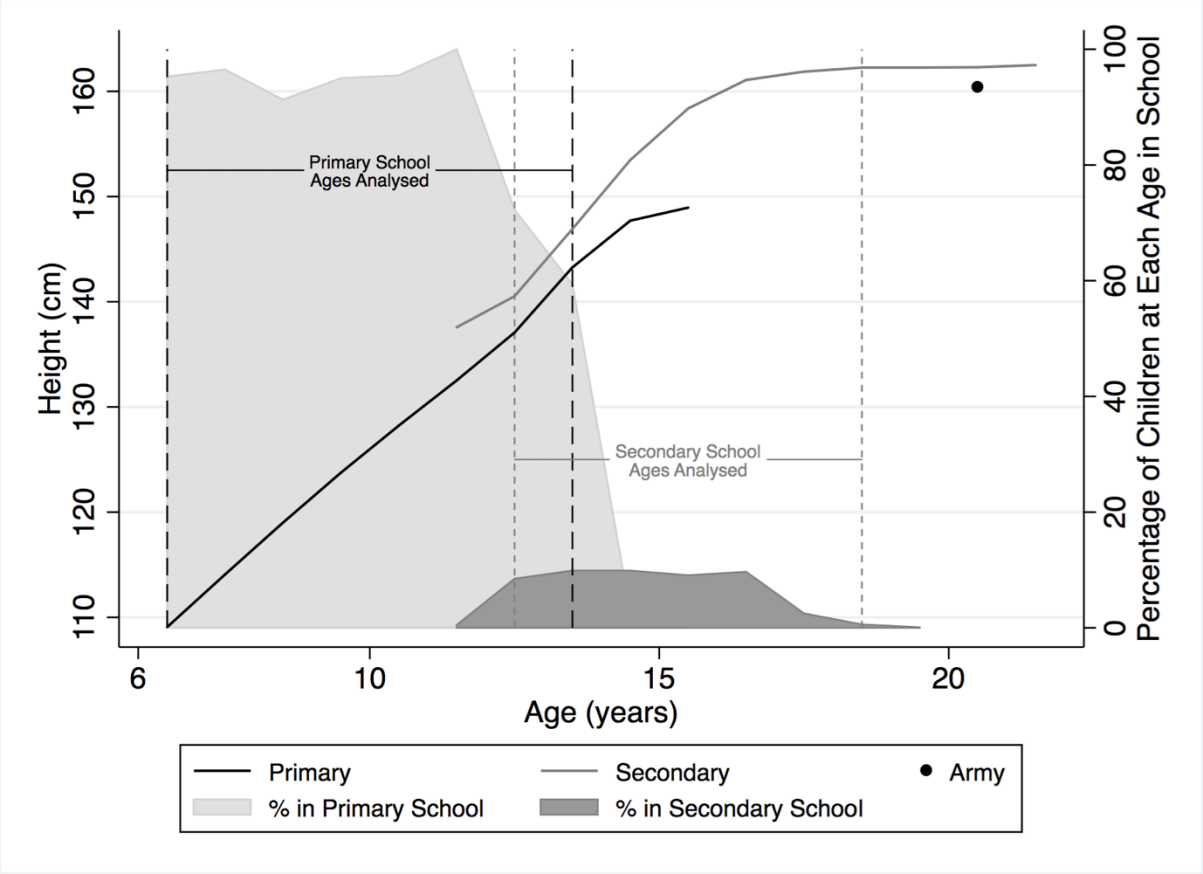
The main sources of data for this paper are the Statistics of School Physical Examination published by the Physical Bureau of the Ministry of Education between 1931 and 1943. These reports record the average heights of boys and girls aged 6 to 21 for each of Japan's 47 prefectures measured annually between 1929 and 1939. These average figures were compiled from physical examinations that occurred in each school every April for all primary schools (*shogakko*), higher primary schools (*koto shogakko*), secondary schools (*chugakko*), and girls' high schools (*koto jyogakko*). The averages reported therefore reflect the children in school at each age. Unfortunately, the individual-level data originally collected has not survived. The data list the heights of primary school and secondary school children separately, and the two school types overlapped between ages 11 and 15. Thus, we have a 47-prefecture panel of children's heights at a large range of ages measured each year between 1929 and 1939. Figure A1 in the web appendix presents maps that display the large variation in the average height of children at age six across the prefectures. Okinawa is clearly an outlier with especially short children. The maps also warn of potential spatial autocorrelation, but our results are robust to clustering at larger regions above the prefecture level.

Given the recent debate over the importance of selection bias in anthropometric sources (Bodenhorn *et al.*, 2017), it is worth discussing potential selection biases in the data presented here. As mentioned above, the Statistics of School Physical Examination report the average heights of children in school. Thus, changes in the percentage of the population enrolled in school could lead to selection bias in our sample. Figure 4 shows the mean height of boys in Japan measured in 1936 separating primary and secondary schools. On the right-hand axis, it also shows the percentage of boys enrolled in 1938 at each age in primary and secondary school. For boys (and girls which are not reported but very similar) the share of children in primary school hovered around 95 per cent of the population from ages 6 to 11 declining to 60 per cent by age 13. The share in primary school then dropped to very low shares of 2.8 and 0.1 per cent at ages 14 and 15 respectively. Secondary schools, on the other hand, were the reserve of a privileged few in the interwar period with only around 9 per cent of the population enrolled between the ages of 12 and 16 and lower share at other ages. The elite, selected nature of secondary schools shows up in the data with secondary students around 3.4 cm taller than primary students at ages 12 and 13 when the data are reliable for both. We can also see this positive selection when comparing secondary school children at age 20 with army conscripts representing again 95 per cent of the population: the secondary school children were 1.3 cm taller than army conscripts. Thus, given the selection of secondary school students, we conduct the bulk of our analysis on the primary school children, only using the secondary school children to complete the growth curve.

In addition to these selection issues across school types, there are also selection problems across ages during the transition from primary to secondary school. Children who enter secondary school at age 11 were slightly taller than expected based on secondary children's heights at later ages because only the very elite entered secondary school at such an early age. Likewise, primary school children at ages 14 and 15 became more and more negatively selected as the proportion of all children in primary school at these ages fell. After the age of 18, there was also a huge amount of measurement error in the average heights of

secondary school children because of small sample sizes. The error in the estimates at these ages is even more pronounced when looking at the individual cohort growth curves (figure A2). Thus, we only include primary school children aged 6 to 13 and secondary school children aged 12 to 18 in our analysis below.

Figure 4: Weighted mean height of boys in primary and secondary schools measured in 1936 across the 47 prefectures compared with the heights of conscripts.



Sources: See data appendix.

Another feature of our panel is that we can express the time dimension in two ways: the year in which the children were measured (period), 1929-39, or the year in which the children were born (cohort), 1911-1933. Thus, we can relate the heights of children either to conditions they were instantaneously facing in a given year (period effects) or to the conditions they faced in early life (cohort effects). Accordingly, we have constructed a prefecture panel of control variables that capture socioeconomic and environmental conditions in each prefecture from 1911 to 1939. These controls and their sources and method of construction are discussed at length in supplementary appendix B. Descriptive statistics for all covariates are also included in the web appendix (Tables A1 and A2). However, it is worth briefly considering the infant mortality rate since this is the most important independent variable in our analysis. Figure 5 presents the infant mortality rate in Japan as a whole, in major cities with more than 100,000 inhabitants and in other areas aside from the major cities. Clearly, there was a major decline in the infant mortality rate across our period from over 180 infant deaths per 1,000 births to less than 100 with an annual average rate of decline of 3.17 per cent. Thus, if infant

mortality had a strong influence on child growth, we would expect to find it in our Japanese case study.

Figure 5: Infant mortality rates in Japan, 1918-40

Notes: Major cities have populations larger than 100,000 inhabitants, other areas are the rest.

Sources: Statistics Bureau of the Cabinet (1924).

3. Methodology

Having explained the data, we can now discuss the two study designs that we use to understand how the disease environment influenced the growth of children in the interwar period. The first study design aims to understand how health conditions *in utero* and shortly after birth affect the growth of children later in childhood and in adolescence. The second analyses how health conditions during the growing years influenced child growth immediately, controlling for shared cohort effects. These designs are based on observational data and do not employ any instrumental variables or natural experiments to precisely identify causal channels. While there are exogenous shocks in our period which might provide helpful internal validity, for instance the 1918-20 flu pandemic and 1923 Kanto earthquake (c.f. Ogasawara, 2017a, 2017b), it is not clear that these would help us understand the secular increase in height. Flu pandemics and earthquakes did not happen every year, and therefore the external validity of these natural experiments may be low in helping us understand a long-term process like the secular change in growth (Deaton and Cartwright, 2016). Likewise, we have not found policy shocks that could be employed to sort out the causation.[2] Thus, we proceed with our

[2] Saito (2008a; 2008b) analyses the effect of the Aiiku-son project in 1936, which was a government project that trained midwives in rural areas to improve sanitary conditions and reduce infant mortality rates. We do not use this project to investigate children's growth because the number of designated villages was quite small at the beginning (for example, only 5 villages in 1936), and the project only became widespread from 1939, which is out of the period when data is available.

observational approach with the caveat that our estimates are not as precisely identified as might be ideal.

However, our study improves substantially on existing observational work in two respects. First, our prefecture panel allows us to include prefecture (and sometimes prefecture-birth year) fixed effects; we have a much wider range of control variables than earlier studies (Hatton, 2011); and we have data for annual birth cohorts rather than data for five or ten-year birth cohorts (Bailey *et al.*, 2016a; Hatton, 2014; Stolz *et al.*, 2013). These precise specifications drastically reduce the potential for omitted variable bias in our regressions relative to earlier work. Second, our observational regressions are designed to minimise potential endogeneity issues. When analysing the influence of health conditions in the birth year on the cohort pattern of growth at ages 6 through 18, there are implicit lags in the regression that limit endogeneity. There is greater potential for issues when we test the instantaneous effects of health conditions on children's heights, but in these regressions we focus principally on the instantaneous influence of infant mortality on children's heights at ages 6 to 11. It is difficult to invent a reverse causality story for this variable. Thus, despite the fact that we do not have a quasi-experimental design, we can make some causal inferences.

Another methodological improvement of our study is to analyse the growth pattern of children rather than focusing solely on mean adult stature. As mentioned above, the growth pattern of children is defined by three characteristics that have varied dramatically over time (see Figure 2): the final adult stature, the timing of the pubertal growth spurt (age at peak velocity) and the velocity of growth at the adolescent peak. We analyse each of these components separately to begin to understand what was driving changes in the growth pattern of children over time. Generally, in order to study the growth pattern, one needs longitudinal data so that individuals' growth can be tracked across various ages. Unfortunately, we do not have individual-level data, only mean heights at various ages for each prefecture. However, we are able to study the growth pattern because we can create pseudo-longitudinal growth curves for each prefecture-birth cohort by grouping the heights of children born in the same year but measured at different ages (and in different years) later in life. For example, we observe the heights of children born in Tokyo prefecture in 1923 at age six in 1929, at age seven in 1930, at age eight in 1931, and so on. Therefore, we can create a cohort growth curve from which to analyse the growth pattern of children born in various prefectures and years (see Figure A3).

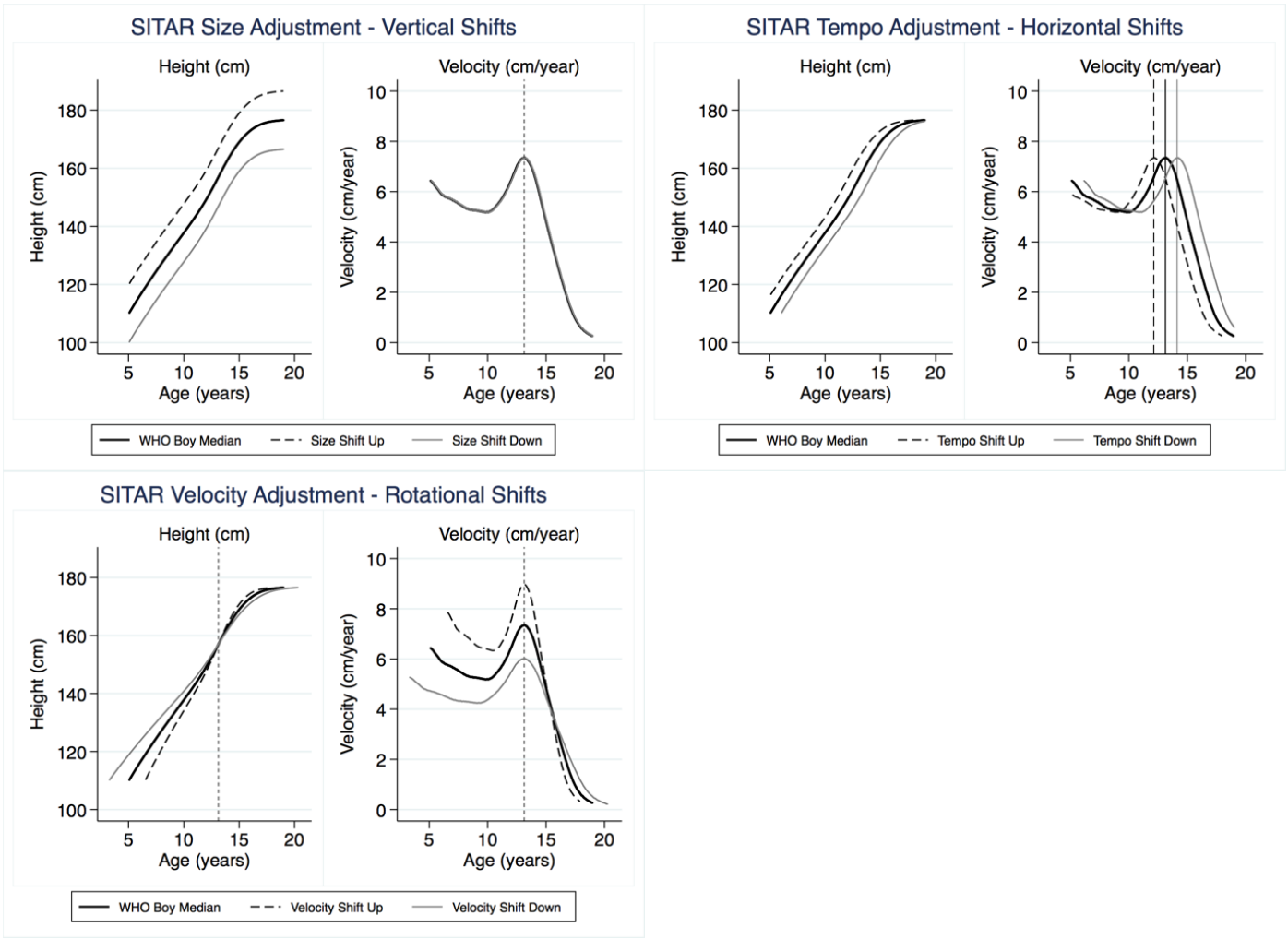
3.1 Early life health conditions and growth

Our first study design analyses how health conditions *in utero* and the first year of life influenced children's growth in childhood and adolescence. However, in order to understand the influence of early life health conditions on the growth pattern, we must parameterise the growth pattern, i.e. produce several indicators for each growth curve that capture the growth pattern. We do this using the SITAR (SuperImposition by Translation and Rotation) growth model developed by Tim Cole and co-authors (Cole *et al.*, 2008; Cole *et al.*, 2010). SITAR fits a non-parametric, random effects model to longitudinal growth curves in an attempt to simplify a vast array of individual-specific variation into a single median growth curve (R Core Team, 2016; Cole, 2016). It is defined by the following equation:

$$y_{i,a} = \alpha_i + h\left(\frac{a-\beta_i}{\exp(-\gamma_i)}\right) \quad (1)$$

where $y_{i,a}$ is the height of prefecture-cohort i at age a, $h(a)$ is a natural cubic spline of height versus age and α_i, β_i, and γ_i are prefecture-cohort-specific random effects. These random effects will be referred to in the rest of the paper as the SITAR parameters. α_i adjusts for differences in mean height between the various growth curves and will be referred to as the size parameter. β_i is the tempo parameter that adjusts for differences in the timing of the pubertal growth spurt between the different growth curves. Finally, γ_i is the velocity parameter that alters the duration and velocity of growth during the pubertal growth spurt. More technically, this parameter stretches or shrinks the age scale adjusting for differences in developmental age and calendar age across individuals. Thus, SITAR attempts to place each individual growth curve onto the median curve predicted by the spline by adjusting the three parameters.

Figure 6: Height and velocity charts showing the various SITAR parameter adjustments to the growth curve.



Notes: In these figures, the median male growth curve from the WHO school-aged and adolescent growth reference (de Onis *et al.*, 2007) is adjusted to show how each SITAR parameter influences the growth pattern of children.

Sources: See data appendix.

The three parameters and the way they influence the growth and velocity curves are presented in the panels of figure 6. The size parameter shifts the growth curve up or down and does not influence the velocity curve. The tempo parameter shifts the growth curve left or right

and also shifts the velocity curve left or right but does not change the levels or trends in velocity. The velocity parameter stretches or shrinks the age scale and thus changes the velocity of growth and the peak velocity during the pubertal growth spurt. Thus, using the SITAR growth model, we are able to estimate size, tempo and velocity parameters for each prefecture-birth cohort in our data.

However, estimating the SITAR parameters is more complicated than it would first appear because the secondary school children in our data were positively selected. As mentioned above and described in figure 4, our data for primary school children are largely representative of the population as a whole between ages 6 and 13, but the children measured in secondary school were always an elite group of children, never including more than 13 per cent of the population of a given prefecture. We could use the primary school data from ages 6 to 11, take some kind of weighted average of the primary and secondary school series for the ages that overlap and then continue using the secondary school data from age 14 onward, but this exercise would clearly produce a biased growth curve since the early ages would be representative and the later ages would be biased towards elites. Therefore, we treat the primary and secondary prefecture-birth cohorts as separate individuals in the SITAR specifications. Thus, the growth curve for primary school students born in Tokyo prefecture in 1922 is entered separately from the growth curve for secondary school students born in Tokyo prefecture in 1922. This allows us to use the secondary data to draw the median spline but estimate the SITAR parameters for the representative primary school data and biased secondary school data separately.[3] We exclude the SITAR parameters produced from the secondary school data in further analysis since any pattern that we might find would be confounded by the changing selection into secondary schools across prefectures and over time. We estimate the parameters for boys and girls separately at all times.

Having estimated the SITAR parameters, we then use them as dependent variables in a series of fixed effects regressions that try to explain variation in the growth pattern using a host of covariates from the birth year. We estimate the following equation:

$$SITAR_{i,t} = \alpha + \beta_1 d_{i,t} + \beta_2 temp_{i,t} + \boldsymbol{\chi}'_{i,t}\boldsymbol{\gamma} + \nu_i + \phi_t + t\lambda_i + \varepsilon_{i,t} \qquad (2)$$

where $SITAR_{i,t}$ is the estimated value of each SITAR parameter (estimated separately) for primary school children in prefecture i and birth year t, $d_{i,t}$ is the infant mortality rate, $temp_{i,t}$ is the average annual temperature and $\boldsymbol{\chi}_{i,t}$ is vector of prefecture and birth year variant controls. Controls include those accounting for population (population density and the crude birth rate), health infrastructure (density of doctors, midwives and hospitals) and income (rice yields, soy yields, milk production per capita and percentage peasants). We also include prefecture fixed effects (ν_i), birth year dummies (ϕ_t) and prefecture-specific trends ($t\lambda_i$). We are particularly interested in the effects of infant mortality and temperature, along with other socioeconomic variables, because these variables reflect the nutritional and disease environment of the children in early life, which theoretically can have strong effects on the growth pattern (Schneider, 2017a). Statistically significant and economically meaningful results would suggest that early life health conditions were very important to the growth pattern of an entire cohort.

In addition to our baseline specification above, we also conduct further analysis to separate the influence of scarring and selection of infants on the growth pattern. As Deaton (2007) noted, the effect of the infant mortality rate in the birth year on later growth or height

[3] Web appendix C discusses some other robustness checks that we performed to test these assumptions.

is the sum of selection and scarring effects of mortality. If the coefficient on infant mortality is positive, then the selection effect, culling of the weak, dominated. If the coefficient is negative, then the scarring effect, the physiological costs of being exposed to a poorer disease environment, dominated. Hatton (2011) separates these effects by comparing the effect of infant mortality rates that children were exposed to in the birth year and at ages 2-4 to proxy for the disease environment in early childhood. We follow a similar methodology by lagging and leading the infant mortality rate in the regressions. Thus, we estimate the following equation:

$$SITAR_{i,t} = \alpha + \beta_1 d_{i,t+n} + \beta_2 temp_{i,t} + \boldsymbol{\chi}'_{i,t}\boldsymbol{\gamma} + \nu_i + \phi_t + t\lambda_i + \varepsilon_{i,t} \qquad (3)$$

which is identical to equation 2 except that we now allow for lags and leads in the infant mortality rate, $d_{i,t+n}$, allowing n to vary from -2 to 4. We do not include multiple lags or leads in the same specification because of very high collinearity between the lags and leads. We include a lag of one year before the birth year to capture potential effects *in utero* and a two-year lag as a placebo test since this should not influence the growth pattern. Infant mortality in the year of birth will show the net effect of selection and scarring, but infant mortality in subsequent years ($n > 0$) will reflect the scarring mechanism. Testing for the scarring mechanism is especially important given the focus in the current paediatric and development literature on preventing stunting in the first thousand days of a child's life (Victora *et al*., 2010). These leads allow us to test how important these early years were to children's growth pattern in interwar Japan and compare with other historical studies (Hatton, 2011).

3.2 Instantaneous influences on growth

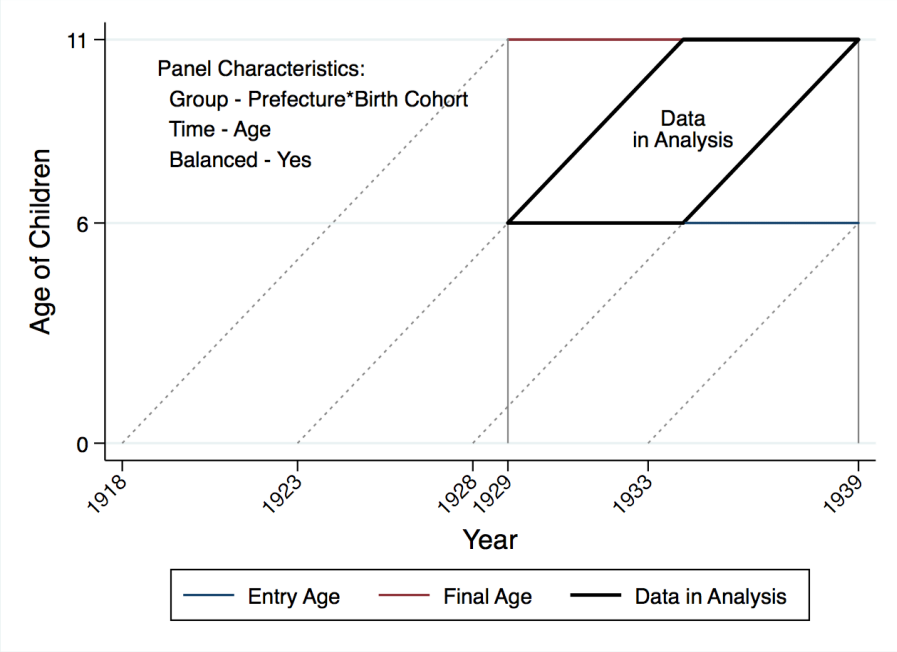
In addition to studying how birth conditions affect the growth pattern of children, we also analyse how health and socioeconomic conditions during the growing years influenced the children's instantaneous growth controlling for conditions that might have affected each cohort at birth. We focus our analysis on the heights of children ages 6 to 11 where the mean heights of children in our dataset cover over 90 per cent of the population. Figure 7 presents a lexis diagram that explains the data included in the analysis. Prefecture-birth cohorts are treated as the group variable in the panel with age providing the time variable. Thus, to have a balanced panel, we can only include birth cohorts with the complete set of ages (6 to 11), i.e. children born from 1923 to 1928. The variation that we analyse then is how health conditions in the years in which the children were measured (period years from 1929 to 1939) influenced their growth. To do this, we use the bilateral-specific fixed effects model presented below:

$$h_{i,t,a} = \alpha + \beta_1 d_{i,t} + \beta_2 cd_{i,t} + \beta_3 temp_{i,t} + \boldsymbol{\chi}'_{i,t}\boldsymbol{\gamma} + \nu_{i,t-a} + \phi_a + \varepsilon_{i,t,a} \qquad (4)$$

where $h_{i,t,a}$ is the mean height of primary school children in prefecture i measured in year t at age a. Our primary variables of interest are the infant mortality rate ($d_{i,t}$), the child mortality rate between ages one and five ($cd_{i,t}$) and the annual average temperature ($temp_{i,t}$). $\boldsymbol{\chi}_{i,t}$ is a vector of prefecture and measured year variant controls including those accounting for population (population density and the crude birth rate), health infrastructure (number of water taps per 100 people, density of doctors, midwives and hospitals) and income (rice yields, soy yields, milk production per capita and percentage peasants). To control for conditions in the first year of life that would have been the same for all children born in a given year and prefecture, we include a fixed effect for each prefecture-birth cohort ($\nu_{i,t-a}$). We also include age fixed effects (ϕ_a) to capture the differences in height at various ages. We could include the

water tap and child mortality variables in these instantaneous height regressions and not in the birth conditions regressions presented above because these data were only available from the mid 1920s onward.

Figure 7: Lexis diagram showing data used in the instantaneous growth regressions.



Thus, these regressions capture the effects of health conditions on growth in childhood controlling for any specific conditions that are common to children in the same prefecture-birth cohort. Note that there are no selection issues related to the mortality variables because the deaths do not relate to children in the same cohort as the height data. For instance, we are looking at the effect of the infant mortality rate of infants born in 1929 on the heights of children age 6 in 1929 who were born in 1923. The same is true for the child mortality rate. Thus, these variables are a proxy for the disease environment and child morbidity that could shape children's immediate growth. In addition, because we are looking at growth at ages 6 to 11, these regressions test whether changes in the health environment later in childhood can have a positive or negative effect on children's growth. This is another test of the importance of the first thousand days to see whether recovery or further stunting is possible after age 3.

4. Results

4.1 Results of SITAR models

Before presenting results from the regressions, we must first discuss the SITAR models and how well they performed in summarising the growth pattern of the prefecture cohorts. In general, the models performed well. Table 1 presents the standard deviations for each of the three SITAR parameters along with standard deviation of the residuals, which at between 7 and 8 mm are of the same order of magnitude as other height datasets analysed using SITAR (Cole *et al.*, 2010). The standard deviations of the three SITAR parameters are substantially lower than those estimated for individual-level datasets because as averages, they are missing most of the individual-level variation. We see that size and tempo are negatively correlated and

size and velocity are positively correlated, though these correlations are fairly low. There is a strong negative correlation between tempo and velocity, which suggests that these two parameters were highly related in the model and should not be taken as entirely independent of one another. The models were estimated with varying levels of degrees of freedom taking the estimations with the lowest BIC as the best fit. Thus, the SITAR models did a good job of describing the growth of Japanese children, and therefore parameters predicted from the model can be used in subsequent regressions to describe the growth pattern of the children.

Table 1: Summaries of SITAR models for boys and girls

	Boys			Girls		
	Standard Deviations	Correlations		Standard Deviations	Correlations	
		Size	Tempo		Size	Tempo
Size (cm)	0.828			0.822		
Tempo (years)	0.235	-0.117		0.318	-0.099	
Velocity (fractional)	0.031	0.245	-0.990	0.035	0.244	-0.917
Residual	0.790			0.713		
Degrees of Freedom	5			4		
BIC of Model	22782.74			19883.53		

Notes: SITAR models were estimated with varying degrees of freedom. We used the degrees of freedom that provided the lowest BIC score in line with Cole *et al.* (2010) as our baseline reported here.

Sources: See data appendix. Statistical citations: R Core Team (2016) and Cole (2017).

The predicted SITAR parameters for primary school boys and girls in each prefecture and birth year are presented in figure A4 in the web appendix. We have excluded the SITAR parameters for the first and last birth cohort, 1916 and 1933, because the parameters are estimated from only one height at one age making them unreliable. The size and velocity parameters tended to increase over time whereas the tempo parameter fell in line with our expectations of the change in the growth pattern during the secular increase in height. There is also a clear outlier in all of the graphs, Okinawa, which highlights the importance of including prefecture fixed effects in all of the regressions. However, in general it is clear that there is enough annual variation across prefectures to be able to test the influence of the health environment at birth on the growth pattern of children.

4.2 Results of early life health conditions on the growth pattern

Table 2 presents the results for the estimation of equation (2) above showing the influence of health conditions at birth on the growth pattern measured through the SITAR parameters. Columns 1-3 show each SITAR parameter, size, tempo and velocity, for boys and 4-6 the same for girls. A quick scan of the table shows no significant coefficients for girls and although there are a few significant variables for boys, namely population density and soy yields, the size of the effect for soy yields is relatively low. Population density has a larger effect on boys' growth pattern with higher population density leading to lower size, an earlier pubertal growth spurt and a more rapid velocity of growth. These results are interesting because they do not strictly conform with a simple view of how improvements in health might influence

the growth pattern, i.e. the long-run trend of increasing size, decreasing tempo and increasing velocity. Rather increasing population density led to improvements in tempo and velocity but a perverse decrease in the growth pattern for size. This does fit within a theory of predictive adaptive responses on growth, which would hold that children with poorer health conditions *in utero* and in infancy (in more densely populated prefectures) would end up shorter but would experience earlier maturation with an earlier pubertal growth spurt and faster growth so that they could reproduce at an earlier age (Schneider, 2017a; Gluckman and Hanson, 2006).

Table 2: Regression of birth health conditions on the growth pattern (SITAR parameters).

	(1)	(2)	(3)	(4)	(5)	(6)
Sex	Boys	Boys	Boys	Girls	Girls	Girls
Dep. Var.	Size	Tempo	Velocity	Size	Tempo	Velocity
Infant Mortality Rate	-0.002	-0.000	0.003	-0.000	0.000	-0.002
	(0.001)	(0.001)	(0.007)	(0.001)	(0.001)	(0.007)
Average Temperature	-0.000	0.018	-0.231	0.018	-0.010	0.212
	(0.048)	(0.017)	(0.202)	(0.026)	(0.019)	(0.250)
Population Density	-0.002**	-0.001***	0.008***	0.000	0.000	-0.002
	(0.001)	(0.000)	(0.002)	(0.000)	(0.000)	(0.003)
Crude Birth Rate	-0.003	0.004	-0.055	0.002	0.004	-0.053
	(0.011)	(0.004)	(0.044)	(0.008)	(0.005)	(0.065)
% Peasants	-0.000	-0.001	0.015	0.014	-0.004	0.017
	(0.016)	(0.006)	(0.073)	(0.008)	(0.005)	(0.067)
Rice Yield	0.001	-0.001	0.010	-0.000	-0.001	0.010
	(0.002)	(0.001)	(0.010)	(0.002)	(0.001)	(0.011)
Soy Yield	0.015**	0.007**	-0.086**	0.001	0.001	-0.010
	(0.007)	(0.003)	(0.036)	(0.005)	(0.003)	(0.035)
Milk Production per capita	-0.025	0.001	-0.023	0.000	-0.003	0.034
	(0.018)	(0.012)	(0.143)	(0.014)	(0.007)	(0.099)
Doctors per 100 people	-3.906	-0.761	7.936	-0.791	0.768	-2.873
	(2.444)	(1.039)	(12.556)	(2.105)	(1.031)	(14.179)
Midwives per 100 people	-2.599	-1.295	15.458	0.111	0.547	-5.605
	(2.770)	(1.233)	(15.112)	(1.426)	(1.233)	(15.101)
Hospitals per 100 people	-11.096	-6.749	81.619	26.790	-9.763*	95.009
	(22.258)	(5.120)	(57.352)	(19.402)	(5.047)	(80.357)
Prefecture Fixed Effects	yes	yes	yes	yes	yes	yes
Birth Year Fixed Effects	yes	yes	yes	yes	yes	yes
Prefecture-specific Time Trend	yes	yes	yes	yes	yes	yes
N	750	750	750	750	750	750
R-square	0.8450	0.7369	0.7905	0.9014	0.8382	0.7795

Notes: Standard errors in parentheses clustered at the prefecture level. * denotes significance at the 10% level. ** denotes significance at the 5% level. *** denotes significance at the 1% level.

Sources: See data appendix.

However, one must be careful not to over-interpret these results, which are not stable across different specifications (not reported) and do not appear for girls. On the whole, Table 2 suggests that conditions *in utero* and in infancy do not seem to have a strong influence on the growth pattern of children in interwar Japan. This lack of strong effects holds if we replace the SITAR size parameter with simply the heights of children at age 6 or 10 and run the same specification (see appendix table A3). Thus, we do not believe that our null result stems from problems relating to the fitted SITAR models. We just do not find a strong relationship between birth conditions and the growth pattern in this period.

For the infant mortality variable, one could argue that we have not found a significant result because the selection and scarring effects were more or less equal and cancelled each other out. To test this and also align our methods with Hatton's (2011) estimations of the influence of infant mortality on child growth in Britain, we estimate equation (3) above allowing the infant mortality rate variable to shift from two years before birth to four years after birth while holding all other variables at the birth year. Table A4 presents the coefficients on the infant mortality variables in these various specifications. The infant mortality rate never had a strong influence on any aspect of the growth pattern in any of the lags or leads. In addition, the coefficients did not have a consistent sign in the years following the birth year, suggesting that taking an average of multiple years would be unlikely to produce a significant result. This result is incongruent with Hatton's earlier work that showed that the infant mortality rate between ages 2 and 4 strongly and significantly reduced children's heights at later ages. Again, to test whether our results are being driven by potential problems with the SITAR models, we ran the same specification using the heights of children at age 6 and age 10 as the dependent variable in separate specifications. These specifications more accurately reproduce Hatton's estimation strategy for Britain. These results are reported in table A5 and again show no significant relationship between the infant mortality rate in the years surrounding birth and heights at later ages for either boys or girls. We also added and subtracted many different control variables to see if that would influence the infant mortality results, but they were consistently insignificant (not reported). Thus, it appears that the scarring effects of exposure to a poor disease environment in early life, proxied by infant mortality, that Hatton and his co-authors found to be so prominent in late nineteenth and early twentieth century Britain were not as important in interwar Japan (Hatton, 2011; Bailey *et al*., 2016a).

4.3 Results of instantaneous influences on growth

If health conditions in the birth year do not strongly influence the growth pattern, the opposite is true of the instantaneous effects of health conditions on children's heights aged 6-11. Table 3 presents the baseline results estimated from equation 4. As a reminder, we include prefecture-birth cohort specific fixed effects to control for differences in birth conditions for each prefecture-cohort and age fixed effects to account for differences in height by age. Thus, we are looking at how health conditions in each prefecture and year of measurement influenced the average height of children in each prefecture and year of measurement controlling for birth cohort conditions in each prefecture. The infant mortality rate had a strong, negative effect on the heights of children with a one standard deviation increase in infant mortality leading to a decrease in height of 0.12 and 0.22 cm for boys and girls respectively. This may seem like a small amount, but these effects would accumulate as children aged, leading to sharp reductions in height. On the other hand, we do not find a significant instantaneous effect of the child mortality rate aged 1-5 on children's heights. The child mortality rate is never significant in any of our specifications and the size of its effect on height is also an order of magnitude lower

than that of the infant mortality rate. Finally, we find a negative relationship between the average temperature in a year and the average height with a one standard deviation increase in average temperature reducing height by 0.20 and 0.26 cm for boys and girls respectively. Interestingly, girls are more responsive to both of these variables.

Table 3: Baseline estimation of instantaneous influences on child height ages 6-11 measured 1929-38.

	(1) Boys	(2) Girls	(3) Boys	(4) Girls
Infant Mortality Rate	-0.005**	-0.009***	-0.007***	-0.015***
	(0.002)	(0.002)	(0.003)	(0.003)
Child Mortality Rate	0.002	-0.007	0.006	0.007
	(0.006)	(0.007)	(0.006)	(0.007)
Average Temperature	-0.087**	-0.112**	-0.097**	-0.110**
	(0.042)	(0.051)	(0.044)	(0.049)
Population Density			-0.001	0.002***
			(0.000)	(0.001)
Crude Birth Rate			-0.014	-0.048***
			(0.017)	(0.016)
Milk Production per capita			0.033	0.071***
			(0.031)	(0.023)
Coverage of Tap Water			0.078***	0.063**
			(0.029)	(0.029)
Doctors per 100 people			8.061***	5.823*
			(2.966)	(2.975)
Additional Controls	no	no	yes	yes
Age Fixed Effects	yes	yes	yes	yes
Prefecture-Birth Year FE	yes	yes	yes	yes
N	1645	1645	1645	1645
R-square	0.9972	0.9963	0.9972	0.9964
F-statistics	2.599*	11.198***	3.229***	6.669***

Notes: Standard errors (in parentheses) are clustered at the prefecture-birth year level. * denotes significance at the 10% level. ** denotes significance at the 5% level. *** denotes significance at the 1% level. Additional controls include the proportion of peasants, rice yield, soy yield, coverage of midwives, and coverage of hospitals.

Sources: See data appendix.

Specifications three and four of Table 3 introduce additional controls but the coefficients on infant mortality rates and average temperature remain largely unchanged. The controls also provide some interesting but tentative secondary results. Population factors are associated with girls' heights with higher population density leading to taller girls and more births leading to shorter girls. Girls also benefited in regions with higher milk production per capita. Both boys' and girls' heights were positively associated with the coverage of tap water in their prefecture and the number of doctors per hundred people, suggesting that health infrastructure could help improve children's heights as well. All of these results are also robust

to clustering the standard errors around 16 larger macro-regions that capture any internal migration, though internal migration was not substantial in this period of Japanese history (Nakagawa, 2001, pp. 38-45). To test the gender differences in the effects, we also ran a regression pooling both boys and girls and interacting all variables with a boy dummy. The results suggest that gender differences in the coefficients are only statistically significant for the infant mortality rate and population density variables (not reported).

Thus, the results for the instantaneous height regressions are substantially different from those of the birth year regressions; we have found strong and significant relationships between current conditions and child growth but no strong relationships between birth conditions and the characteristics of a cohort's growth pattern. The following section discusses these results in detail and provides further robustness checks to try to understand the mechanisms underlying the relationship between childhood morbidity and growth.

5. Discussion and Robustness Checks

5.1 Why does infant mortality shape growth at later ages?

One of our most striking findings was that although infant mortality rates in the year of birth did not strongly influence the growth pattern of children, there was a strong, significant instantaneous effect of infant mortality on children's heights between the ages of 6 and 11. The effect was specific for infant mortality and was never significant or of the same order of magnitude for child mortality. This result is interesting and begs further exploration as to why the infant mortality rate has this instantaneous effect on child growth at ages 6 and above. To better understand this, we added a number of other types of mortality to the instantaneous growth regression, equation (4), to see whether the infant mortality variable is merely capturing a sum of mortality from other causes. Thus, we have included the fetal death rate as a proxy of conditions *in utero* and also as a potential measure of gender bias since there is evidence of some bias against girls in infanticide (Drixler, 2016, p. 680).[4] We also include deaths from congenital infirmity per 1000 people, though most of these deaths occurred among infants. The typhoid death rate and diarrhoeal death rate proxy water borne diseases that others have argued have large spill-over effects for other types of mortality (Ferrie and Troesken, 2008). The tuberculosis death rate captures the influence of tuberculosis on growth though in Japan tuberculosis tended to be an adult rather than child disease, so we would not expect to see a strong result (Kawakami, 1982, p. 336). The Beriberi death rate is also included as a proxy for nutrition related diseases. Beriberi is caused by thiamine deficiency from a diet too heavily based on white rice (Meade, 1993). Finally, we include two infectious diseases that are prevalent among children, measles and whooping cough, to see how these highly infectious diseases that give immunity after infection affected growth.

Table 4 reports these results. Specifications one and two indicate that including the additional mortality variables only strengthens the effect of infant mortality rather than reducing it. We see that the fetal death rate negatively influences the heights of girls, which may capture potential gender bias against girls in these years. There is also a positive effect for the tuberculosis death rate on girls' heights which is somewhat puzzling. However, a joint test

[4] As Drixler (2016) has recently argued, there were large problems with the mis-reporting of stillbirths in Imperial Japan, but by the time we use the fetal death rate in our regressions, 1929-39, this reporting error was largely corrected (Drixler, 2016, pp. 654, 657). However, the fetal death rate in our period still includes some infanticides, though they were less than 0.5 per cent of total births by 1929 (Drixler, 2016, p. 677).

of significance shows that these additional disease variables are insignificant when taken together. In specifications three and four we remove the infant and child mortality rates from the regression to see whether collinearity between these variables and the other cause-specific mortality rates were driving the lack of significance in the first specifications. Removing the

Table 4: Instantaneous effects of the infant mortality rate on heights of children age 6-11 controlling for other types of mortality.

	(1) Boys	(2) Girls	(3) Boys	(4) Girls
Infant Mortality Rate	-0.008**	-0.017***		
	(0.003)	(0.004)		
Child Mortality Rate	0.003	0.008		
	(0.011)	(0.013)		
Fetal Death Rate	-0.010	-0.018*	-0.008	-0.014
	(0.007)	(0.009)	(0.007)	(0.009)
Death Rates from:				
Congenital Infirmity	-0.002	0.191	-0.165	-0.127
	(0.176)	(0.232)	(0.171)	(0.220)
Typhoid	-0.046	-0.056	0.008	0.051
	(0.230)	(0.254)	(0.233)	(0.256)
Diarrhoea	0.030	0.098	-0.040	-0.031
	(0.107)	(0.104)	(0.068)	(0.062)
Tuberculosis	0.281	0.459**	0.193	0.287
	(0.208)	(0.229)	(0.208)	(0.226)
Beriberi	0.080	0.308	-0.272	-0.396
	(0.254)	(0.292)	(0.220)	(0.257)
Measles	0.268	-0.020	0.090	-0.345**
	(0.245)	(0.275)	(0.128)	(0.142)
Whooping Cough	-0.081	0.185	-0.251	-0.135
	(0.233)	(0.279)	(0.190)	(0.201)
Additional Controls	yes	yes	yes	yes
Age Fixed Effects	yes	yes	yes	yes
Prefecture-Birth Year FE	yes	yes	yes	yes
N	1410	1410	1410	1410
R-square	0.9979	0.9978	0.9979	0.9978
F-stat Disease Covariates	0.970	1.208	0.894	2.090**

Notes: Standard errors (in parentheses) are clustered at the prefecture-birth year level. * denotes significance at the 10% level. ** denotes significance at the 5% level. *** denotes significance at the 1% level. Additional controls include the average temperature, population density, crude birth rate, proportion of peasants, rice yield, soy yield, milk production per capita, coverage of doctors, coverage of midwives, and coverage of hospitals.

Sources: See data appendix.

infant and child mortality, however, has very little influence on the regressions with these specifications also having very few significant variables. Only the measles death rate in girls was significant and the magnitude of the coefficient was very small. Thus, it seems that the infant mortality rate is capturing aspects of the disease environment that are different from these other mortality variables and are important for child growth.

These results suggest, then, that the causes of mortality among children and from specific causes at all ages may be very different from the causes of chronic child morbidity, which could potentially have a stronger influence on child growth. Measuring morbidity is very difficult because vital registration has always focused on the most important causes of death, which tended to be infectious diseases in the late nineteenth and early twentieth centuries. However, there is not a strong case that many of these infectious diseases would necessarily influence growth. Bleakley and Lange (2009) have argued that chronic diseases like malaria and hookworm would have more substantial impacts on child health than infectious diseases. Many infectious diseases would solely make a child sick for a month or two and then kill them or leave them immune rather than placing a chronic burden on the child's energy and nutrients. Oxley (2006) also found that smallpox did not have a consistent negative effect on growth across nineteenth century Britain. Thus, it seems likely that infant mortality captures the types of diseases that make children chronically ill in a way that these other mortality rates do not. This morbidity could include diseases like diarrhoea, which are not prominent killers of children over the age of five, but could still affect growth. The infant mortality effect cannot solely be related to water-borne illnesses, however, because although the expansion of water taps is associated with taller children, this additional control does not explain away the infant mortality effects in specifications 3 and 4 of table 3. We could also be capturing the effects of coal smoke pollution, which was a prominent cause of death among infants and influenced adult heights in nineteenth-century Britain (Beach and Hanlon, 2017; Bailey *et al*., 2016b). However, the infant mortality effect is not solely a pollution story either since adding the percentage of people employed in industry to the regressions does not make the infant mortality result insignificant (not reported). Unfortunately, it is difficult to be more precise about potential mechanisms, but our evidence still shows that the disease environment mattered for children's instantaneous growth in interwar Japan.

5.2 Cohort versus period influences on growth

The most puzzling finding in our results was that conditions in the year of birth did not seem to be as important in determining growth and the growth pattern at later ages as instantaneous conditions. These results are puzzling for a number of reasons. First, historical studies looking at the twentieth century have generally found that the infant mortality rate in the years following birth was strongly related to height at later ages (Hatton, 2011, 2014; Bailey *et al*., 2016a). Second, there is a growing medical and development literature that highlights the importance of the first thousand days of a child's life for subsequent growth, health, human capital and labour market outcomes (Victora *et al.*, 2010). Finally, Ogasawara (2017a) found that cohorts *in utero* and born during the 1918-20 flu pandemic were shorter later in childhood than cohorts born before or after the pandemic, providing one example about how these early life health conditions did affect children's growth in Japan. Thus, it is worth considering the robustness of our results, why our results differ from other historical studies and finally their external validity to be applied in other contexts. We will deal with each of these in turn.

To begin, our null result that birth conditions did not influence subsequent growth seems fairly robust. We tried a number of specifications with various controls, fixed effects,

trends, clustering, etc., and we never produced a specification that produced significant and economically meaningful coefficients. Even if we happened upon one, the instability of the results would be very troubling. As mentioned above, we tried using the heights of children at various ages rather than the SITAR parameters to ensure that error in the SITAR parameters was not driving our results. We also included various leads and lags on the infant mortality rate to see if it mattered more in years after the birth year than in the birth year itself. Thus, our null result appears to be stable.

Our results could differ from the existing historical literature for a number of reasons. First, our regression specifications are far more precise than many of the existing studies. Hatton (2014) uses five-year average infant mortality rates at the country level to explain the average height of five-year birth cohorts. Bailey *et al.* (2016a) relate the average infant mortality from 1891-1900 in particular registration districts to the heights of adults conscripted during the first world war. We used prefecture data and related children's growth to the annual infant mortality rate in the year of their birth. We also include prefecture fixed effects, which Bailey *et al.* (2016a) could not, though they were able to consider household characteristics that we cannot. However, it is possible that in these less precise specifications, the infant mortality rate incorporates aspects of a locality other than the exact conditions that are leading infant mortality to rise or fall. Or it may be that the infant mortality rate over a five or ten year birth cohort is actually capturing the cumulative effects of the instantaneous influence of childhood morbidity on growth rather than an early life effect, though this would not be true of Hatton (2011). The strong effect of infant mortality on child height that Hatton (2011) finds in 40 cities across Britain may also be stronger because of the urban focus which may not represent the effect in the surrounding countryside. None of these issues invalidate these earlier studies; they just pose slight differences in the study design that may influence the results.

However, our own study is not without problems. Our prefecture aggregate data hide a tremendous amount of individual-level and sub-prefecture-level variation in health conditions and growth. There were also small differences in the enrolment rate of primary school children in different prefectures over time, though these enrolment rates were always above 90 per cent between 1929 and 1939 and including the enrolment rate as a control in the regressions above did not influence our results (not reported). These potential flaws are not unique to our data though. Hatton (2011) and (2014) also use aggregated data with representativeness issues and find strong results, so the weaknesses with our study cannot solely explain our results.

Finally, it is important to consider how relevant our findings for Japan are for historical comparisons to other parts of the world and to a wider discussion of influences on children's growth today. One important threat to external validity would be if Japanese children's growth was different from children's growth in other parts of the world. Eveleth and Tanner (1976, 1990) analysed the growth patterns of children in various parts of the world and found that Asian children, including the Japanese, had a slightly different growth pattern than Africans and Europeans, experiencing earlier maturation, shorter adult stature and shorter leg lengths. This is born out in Figure 3 above, which shows that the secular increase in stature has largely stopped since the 1980s despite other aspects of health continuing to improve (Bassino and Kato, 2010). However, despite these minor changes in the growth pattern, Japanese men experienced a greater increase in mean adult stature since the mid-nineteenth century than British men. Thus, it is difficult to argue that our null relationship between conditions in early life and later growth pattern was purely driven by relatively minor differences in the growth pattern of Asian children, especially since the instantaneous effects are so strong.

Another potential concern in comparing our analysis with those conducted in Britain and Europe is that the changes in infant mortality were different. As mentioned above, infant mortality rates fell substantially at a national level from over 180 to under 100 infant deaths per 1,000 births between 1917 and 1939 in Japan. During the time period covered by Hatton's (2011) article, infant mortality rates fell in Britain by a similar absolute amount but started at a lower level around 150 infant deaths per 1,000 births (Woods, 2001, p. 253), and this was generally true of all of Europe (Hatton, 2014). Thus, the absolute magnitude of decline seems fairly similar, but the infant mortality decline happened much more rapidly in Japan at least at the national level (Ito 1998, pp. 726-727).

Table 5: Causes of infant deaths in England and Japan during the infant mortality decline.

	England (1889-91)		England & Wales (1919)		Japan (1921)		Japan (1938)	
	Urban	Rural	Urban	Rural	Major Cities	Other Areas	Major Cities	Other Areas
% of Infant Deaths								
Respiratory	17.0%	21.7%	21.4%	20.2%	21.7%	20.0%	26.6%	21.9%
Diarrhoeal	20.4%	6.2%	10.5%	6.3%	21.0%	17.9%	13.7%	17.4%
Infant Mortality Rate by Cause (infant deaths per 1,000 births)								
Respiratory	37.0	21.1	19.5	16.3	38.1	33.1	26.2	26.0
Diarrhoeal	44.6	6.0	9.6	5.1	36.9	29.6	13.5	20.7
All Causes	218.0	97.2	91.3	80.5	175.7	167.5	98.4	118.9
Average Annual Growth Rate of IMR Between Two Periods								
Respiratory			-2.2%	-0.9%			-2.2%	-1.4%
Diarrhoeal			-5.2%	-0.6%			-5.7%	-2.1%
All Causes			-3.0%	-0.6%			-3.4%	-2.0%

Notes: Respiratory deaths include deaths from bronchitis and pneumonia. Diarrhoeal deaths include deaths from both diarrhoea and enteritis. England (1889-91) - urban refers to the causes of death in three towns (Blackburn, Leicester and Preston) with over 100,000 inhabitants and rural refers to the causes of death in three rural counties (Dorset, Hertfordshire and Wiltshire); England and Wales (1919) - urban and rural districts as defined by the Registrar General; Japan - major cities have populations larger than 100,000 inhabitants, other areas are the rest.

Sources: England (1889-91) - Woods (2001, pp. 258-9); England and Wales (1919) - Eighty Second Annual Report of the Registrar General (1920, pp. 45-9); Japan - Statistics Bureau of the Cabinet (1924).

It is also possible that the causes of infant mortality in Japan were different from those in Europe. If the causes of death were different, this may explain why the morbidity from infant mortality diseases affected children's growth differently in the two contexts. The classification of causes of death among infants is notoriously problematic for both the UK and Japan (Woods, 1997, p. 84; Reid, 2002, p. 158; Ito 1998, p. 731), but here we focus on two fairly easily classifiable causes of death among infants related to respiratory diseases and diarrhoeal diseases respectively: deaths from pneumonia and bronchitis and deaths from diarrhoea and enteritis. Table 5 compares the percentage of infant deaths, the cause-specific infant mortality rates and the annual average rate of infant mortality decline for urban and rural areas in England and in Japan for various years during the infant mortality decline. We will discuss urban areas before moving onto rural areas. Toward the beginning of the infant mortality decline in

England (1889-91) and in Japan (1921), infant mortality rates were very high in urban areas and the two causes highlighted here made up around 40 per cent of infant deaths. Infant death rates from respiratory diseases in the UK and Japan were very similar, but the infant death rate from diarrhoea and enteritis was 20 per cent higher in the UK. The annual average rates of decline in infant mortality in urban areas over the two sub-periods in each country were relatively similar with Japan experiencing slightly more rapid decline.

However, there were much larger differences in infant mortality in rural areas. Part of these differences have to do with the way that rural areas are defined. The three rural counties that were used to proxy the rural mortality rate for England (1889-91) had very few towns whereas the Japanese 'other areas' category includes all mortality outside cities with more than 100,000 inhabitants. Nevertheless, infant mortality rates were much higher in rural areas in Japan than in the UK with especially higher infant death rates from diarrhoeal diseases. Japan did experience rapid declines in infant mortality rates between 1921 and 1938, but it still had higher rates of infant mortality from respiratory diseases and diarrhoea in 1938 than England and Wales did in 1919. Thus, when comparing the characteristics of infant mortality decline in the two countries, the rate of decline if not the level seems to have been fairly similar. In addition, the two leading environmental causes of infant death were the same in both countries and both countries experienced fairly sharp declines in mortality from the types of diseases that would have scarred children in early life and instantaneously affected their growth later in childhood. Therefore, differences in infant mortality do not explain the different results that we find relative to the literature on Britain.

In conclusion, we cannot easily dismiss our finding that although infant mortality mattered for instantaneous growth, it did not have strong scarring effects in early life as found in other countries. Our estimation strategy is more precise than others employed in the literature. In addition, although there are slight differences in the growth pattern of Japanese and British children and they did not face precisely the same conditions in the first half of the twentieth century, our results cannot be easily explained away by these differences. In the end, it seems that in Japan year-to-year conditions even outside of the critical window of the first thousand days played a more important role in shaping children's growth than their initial conditions at birth.

6. Conclusion

This paper has tested the influence of childhood morbidity, proxied through the infant mortality rate, on the growth of children in interwar Japan. It is the first paper to use the SITAR growth model to parameterise the growth pattern of historical children and attempt to explain the parameters. We found that in Japan the infant mortality rate and other health conditions around the time of birth did not have a strong influence on the growth pattern of children from age six onwards. This is a striking contrast to recent literature that found strong scarring effects of exposure to higher childhood morbidity in early life on later growth in Britain and Europe (Hatton, 2011; Hatton, 2014; Bailey *et al.*, 2016a). However, infant mortality did have an important instantaneous effect on children's growth in Japan, suggesting that annual variation in health conditions across the growing years may have been more important in shaping the growth pattern than early-life scarring.

These results are important in two respects. First, they help provide context to explain the secular increase in height in Japan and elsewhere. We can definitively say that reductions

in child morbidity were important for increasing stature during the interwar period in Japan. We also have tentative evidence that the expansion of clean water and other health infrastructure could improve child health. Although these results are only associations, there is substantial evidence that the rollout of sanitation and clean water and the introduction of social workers who gave poorer people access to medical treatment reduced mortality in interwar Tokyo (Ogasawara and Kobayashi, 2014; Ogasawara *et al.*, 2016). However, most of the income and nutrition related variables were insignificant, suggesting that the disease environment was more important for the secular increase in height in interwar Japan than nutrition. This balance of disease versus nutrition may have been different after the second world war, though, when economic growth increased and milk consumption expanded (Takahashi, 1984). Our results also suggest that health conditions in early life were not good predictors of the growth pattern of children or simply their heights at later ages. Thus, we find that children's growth was plastic across the growing years with children catching up or falling behind relative to modern standards as they developed and grew based on annual health conditions. The observed cohort growth pattern, then, was the cumulative effect of these annual fluctuations in health conditions rather than a predetermined pattern set *in utero* and in early life. This view of the secular increase in height places a greater emphasis on growth in childhood and adolescence. Essentially, we find that the period effects on growth are stronger than the cohort ones.

Second, the strong instantaneous effects of infant mortality on children's heights at ages 6 to 11 present challenges to the growing orthodoxy in the modern development literature that places enormous importance on growth in the first thousand days of life (Victora *et al.*, 2010). Our evidence from interwar Japan shows that conditions outside of this early life critical window can influence children's growth and may be more important for their completed growth pattern. This concurs with evidence from the Consortium of Health Oriented Research in Transitioning Societies (COHORTS) longitudinal studies that showed that catch-up growth was possible between 24 months and mid-childhood and between mid-childhood and adulthood (Prentice *et al.*, 2013). This matters for the current development debate on child stunting for two reasons. First, it highlights that, at least as far as growth is concerned, interventions outside the thousand-day critical window can influence children's growth pattern. However, we are cautious about this finding since studies have found that growth faltering can have detrimental cognitive effects for children, and we cannot measure the cognitive effects of interventions at later ages. Second, if one takes a long-run perspective on reducing child stunting, the goal is in effect to create the conditions that led to the historical secular increase in height. Recent evidence on birth weights (Schneider, 2017b) suggests that birth weights have not changed very much over the past one hundred years in Western Europe and North America. Thus, the secular increase was likely created by a combination of reducing growth faltering early in life and by strengthening positive interventions later in childhood and adolescence that could make up for slower growth at earlier ages.

Finally, our paper raises several important issues for future research. It highlights the importance of child morbidity for child growth but unfortunately was not able to test precisely how exposure to any given type of illness influenced children's growth. Infant mortality represents a fairly wide range of potential exposures so more detailed research on particular diseases would be helpful. In addition, more studies are needed that focus more holistically on the growth pattern rather than looking at height in very limited windows during childhood or at adulthood. It is possible that the growth faltering that we observe in early years historically was simply the product of a growth pattern with a lower velocity and later pubertal growth spurt than that experienced by modern children. The only way to check this definitively is to

try to understand how the growth pattern from early childhood to adulthood has changed over time. This will involve returning to archives and searching for new sources of data that cover children's growth over a wider range of ages since most of the existing sources on child growth are too limited to capture the full growth pattern. However, understanding how the growth pattern of children shifted over time is fundamental to understanding the causes of the secular increase in height.

References

Archival Source References

Army Ministry. *Rikugunsho tokei nenpo* [Annual Statistical Report of the Army Ministry (volumes 41-49)]. Tokyo, 1931-1939.

Kayo, Nobufumi, ed., Todofuken nogyo kiso tokei [Basic statistics of agriculture in Japanese prefecture] (Tokyo, 1983)

Physical Education Bureau, Ministry of Education. (1931-1943). Statistics of school physical examination (1929-1939 editions). [in Japanese] Tokyo.

Physical Education Bureau, Ministry of Education. *Gakko shintaikensa tokei, Showa jyusannendo shogakko hen* [Statistics of school physical examination, primary school part (1938 edition)]. Tokyo, 1942a.

Physical Education Bureau, Ministry of Education. *Gakko shintaikensa tokei, Showa jyusannendo chugakko hen* [Statistics of school physical examination, secondary school part (1938 edition)]. Tokyo, 1942b.

Physical Education Bureau, Ministry of Education. *Gakko shintaikensa tokei, Showa jyuyonendo shogakko, yochien, mouagakko hen* [Statistics of school physical examination, primary school, kindergarden, and school for the blind or deaf children edition (1939 edition)]. Tokyo, 1943a.

Physical Education Bureau, Ministry of Education. *Gakko shintaikensa tokei, Showa jyuyonendo chugakko hen* [Statistics of school physical examination, secondary school part (1939 edition)]. Tokyo, 1943b.

Statistics Bureau of the Cabinet. (1924c–1936c). Nihonteikoku tokei nenkan [Statistical yearbook of the Japanese Empire, vol.43–55]. Tokyo.

Statistics Bureau of the Cabinet. (1937c–1941c). Dainihonteikoku tokei nenkan [Statistical yearbook of the greater Japanese Empire, vol.56–59]. Tokyo.

Statistics Bureau of the Cabinet. (various years). Kokusei chose hokoku, fukenhen [Population census of Japan, prefectural part, 1930 edition]. Tokyo.

Online Databases

Database of the Statistical Survey Department, Statistics Bureau, Ministry of Internal Affairs and Communications: http://www.stat.go.jp/data/chouki/zuhyou/02-05.xls (accessed on 23rd March 2017)

Database of the Japan Meteorological Agency: http: //www.data.jma.go.jp/gmd/risk/obsdl/index.php (accessed on 4th April 2017).

Secondary References

Bailey, R. E., Hatton, T. J., & Inwood, K. (2016a). Health, height, and the household at the turn of the twentieth century. *The Economic History Review*, *69*(1), 35–53. http://doi.org/10.1111/ehr.12099.

Bailey, R. E., Hatton, T. J., & Inwood, K. (2016b). Atmospheric Pollution and Child Health in Late Nineteenth Century Britain. *CEPR Discussion Paper*, 1–43.

Bassino, J. P. (2006). Inequality in Japan (1892–1941): Physical stature, income, and health. *Economics and Human Biology*, *4*(1), 62–88. http://doi.org/10.1016/j.ehb.2005.10.002.

Bassino, J. P., & Kato, N. (2010). Rich and slim, but relatively short Explaining the halt in the secular trend in Japan. *Center for Economic Institutions Working Paper Series*.

Baten, J., & Blum, M. (2012). Growing Tall but Unequal: New Findings and New Background Evidence on Anthropometric Welfare in 156 Countries, 1810–1989. *Economic History of Developing Regions*, *27*(sup1), S66–S85. http://doi.org/10.1080/20780389.2012.657489.

Beach, B., & Hanlon, W. W. (2017). Coal smoke and mortality in an early industrial economy. *Working Paper*.

Bleakley, H., & Lange, F. (2009). Chronic disease burden and the interaction of education, fertility, and growth. *The Review of Economics and Statistics*, *91*(1), 52–65.

Cameron, N. (1979). The growth of London schoolchildren 1904-1966: an analysis of secular trend and intra-county variation. *Annals of Human Biology*, *6*(6), 505–525.

Cole, T. J. (2016). Super Imposition by Translation and Rotation Growth Curve Analysis. Version: 1.0.8. Date: 2016-10-17. URL: https://github.com/statist7/sitar.

Cole, T. J., Donaldson, M. D. C., & Ben-Shlomo, Y. (2010). SITAR--a useful instrument for growth curve analysis. *International Journal of Epidemiology*, *39*(6), 1558–1566. http://doi.org/10.1093/ije/dyq115,

Cole, T.J., Cortina Borja, M., Sandhu, J., Kelly, F.P., Pan, H. (2008) Nonlinear growth generates age changes in the moments of the frequency distribution: the example of height in puberty. *Biostatistics*, 9, 159–71.

Deaton, A. (2007). Height, health, and development. *Proceedings of the National Academy of Sciences*, *104*(33), 13232-13237.

Deaton, A., & Cartwright, N. (2016). Understanding and Misunderstanding Randomized Controlled Trials. *NBER Working Paper*, no. 22595.

Drixler, F. F. (2016). Hidden in Plain Sight: Stillbirths and Infanticides in Imperial Japan. *Journal of Economic History*, *76*(3), 651–696. http://doi.org/10.1017/S0022050716000814.

Eveleth, P. B., & Tanner, J. M. (1976). *Worldwide Variation in Human Growth*. Cambridge: Cambridge University Press.

Eveleth, P. B., & Tanner, J. M. (1990). *Worldwide Variation in Human Growth*. Cambridge: Cambridge University Press.

Ferrie, J., & Troesken, W. (2008). Water and Chicago's mortality transition, 1850-1925. *Explorations in Economic History*, *45*(1), 1–16.

Gluckman, P., & Hanson, M. (2006). Evolution, development and timing of puberty. *Trends in Endocrinology & Metabolism*, *17*(1), 7–12.

Hatton, T. J. (2011). Infant mortality and the health of survivors: Britain, 1910–50. *The Economic History Review*, *64*(3), 951–972.

Hatton, T. J. (2014). How have Europeans grown so tall? *Oxford Economic Papers*, *66*(2), 349–372. http://doi.org/10.1093/oep/gpt030.

Ito, S., (1998). The decline in infant mortality in pre-war Japan [in Japanese]. *Shakaikeizaishigaku*, 63(6), 725–752.

Kawakami, T,. (1982). History of Sick People in Modern Japan [in Japanese]. Keisoshobo, Tokyo.

Meade, M. S. (1993). Beriberi. In K. F. Kiple (Ed.), *The Cambridge world history of human disease.* (pp. 1176–612). Cambridge: Cambridge University Press. http://doi.org/10.1017/CHOL9780521332866.077.

Mosk, C. (1996). Making Health Work: Human Growth in Modern Japan. University of California Press, Berkeley.

Nakagawa, S. (2001). Internal migration of Japan during the 1920s and 1930s [in Japanese]. *Kokumin keizaizasshi*, 184(4), 37-51.

Ogasawara, K. (2017a). Persistence of pandemic influenza on the development of children: Evidence from industrializing Japan. *Social Science & Medicine*, *181*, 43–53. http://doi.org/10.1016/j.socscimed.2017.03.023.

Ogaswara, K. (2017b). Child growth and the Great Kanto Earthquake. Working Paper.

Ogasawara, K., & Kobayashi, G. (2014). The impact of social workers on infant mortality in inter-war Tokyo: Bayesian dynamic panel quantile regression with endogenous variables. *Cliometrica*, *9*(1), 97–130. http://doi.org/10.1007/s11698-014-0110-1.

Ogasawara, K., Shirota, S., & Kobayashi, G. (2016). Public health improvements and mortality in interwar Tokyo: a Bayesian disease mapping approach. *Cliometrica*, *9*(4), 1–31. http://doi.org/10.1007/s11698-016-0148-3.

de Onis, M., & Blössner, M. (2003). The World Health Organization Global Database on Child Growth and Malnutrition: methodology and applications. *International Journal of Epidemiology*, *32*(4), 518–526. http://doi.org/10.1093/ije/dyg099.

de Onis, M., Onyango, A. W., Borghi, E., Siyam, A., Nishida, C., & Siekmann, J. (2007). Development of a WHO growth reference for school-aged children and adolescents. *Bulletin of the World Health Organization*, *85*(9), 660-667. http://doi.org/10.2471/BLT.07.043497.

Osamu Saito. (1989). Did Economic Development Bring About Mortality Decline? - Nutrition, Physique and Life Expectancy in the West and Japan. *Keizai Kenkyu*, 40(4), 339-356.

Oxley, D. (2006). "Pitted but not pitied" or, does smallpox make you small? *The Economic History Review*, *59*(3), 617–635. http://doi.org/10.1111/j.1468-0289.2006.00356.x.

Prentice, A. M., Ward, K. A., Goldberg, G. R., Jarjou, L. M., Moore, S. E., Fulford, A. J., & Prentice, A. (2013). Critical windows for nutritional interventions against stunting. *American Journal of Clinical Nutrition*, *97*(5), 911–918. http://doi.org/10.3945/ajcn.112.052332.

R Core Team. (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Reid, A. (2002). Infant feeding and post-neonatal mortality in Derbyshire, England, in the early twentieth century. *Population Studies*, *56*(2), 151–166. http://doi.org/10.1080/00324720215926.

Saito, O. (2003). Taii no seichoˆ to keizai hatten: Meijiki Yamanashi ken gakkô shinchô kiroku no bunseki [Human growth and economic development: an examination of school physical records, Yamanashi prefecture, Meiji Japan]. *Econ. Res. Keizai Kenkyu*, 54 (1), 19–32.

Saito, O. (2004). Senzen Nihon ni okeru taii no kettei yoin: Tôhoku nôson no shinchô kensa kiroku bunseki [The determinants of height, weight and weight-for-height in prewar Japan; An analysis of conscription records for a north-eastern village]. *Econ. Res. Keizai Kenkyu*, 55 (3), 193–203.

Saito, O. (2008a). Infant mortality and the aiiku village scheme in pre-war Japan [in Japanese]. *Shakai keizaishigaku*, 73(6), pp. 611-633.

Saito, O. (2008b). The role of the intermediary organisations in mother and child hygiene policy. In: Inoki T (ed) *Social groups and network in inter-war Japan* [in Japanese]. NTT Publishing, Tokyo, pp. 359-379.

Schneider, E. B. (2017a). 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. http://doi.org/10.1111/ehr.12484.

Schneider, E. B. (2017). Fetal health stagnation: Have health conditions in utero improved in the United States and Western and Northern Europe over the past 150 years? *Social Science & Medicine*, *179*, 18–26. http://doi.org/10.1016/j.socscimed.2017.02.018.

Stolz, Y., Baten, J., & Reis, J. (2013). Portuguese living standards, 1720–1980, in European comparison: heights, income, and human capital1. *The Economic History Review*, *66*(2), 545–578.

Takahashi, E. (1984). Secular trend in milk consumption and growth in Japan. *Human Biology*, *56*(3), 427-437.

Victora, C. G., de Onis, M., Hallal, P. C., Blossner, M., & Shrimpton, R. (2010). Worldwide Timing of Growth Faltering: Revisiting Implications for Interventions. *Pediatrics*, *125*(3), e473–e480. http://doi.org/10.1542/peds.2009-1519.

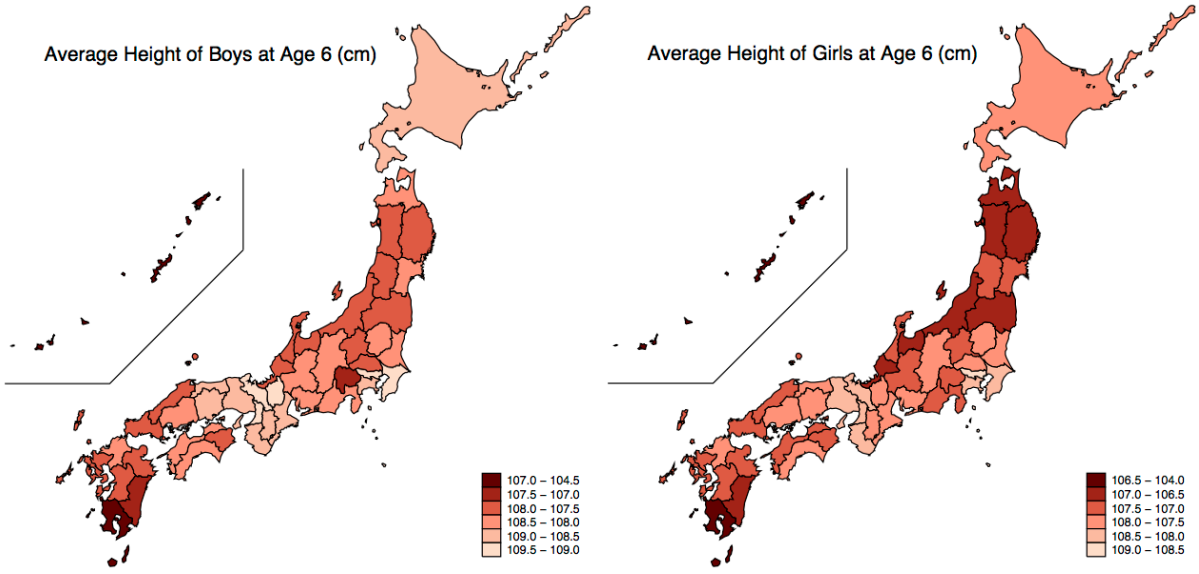
Woods, R. (1997). Infant Mortality in Britain: A Survey of Current Knowledge on Historical Trends and Variations. In Bideau, A., Desjardins, B., & Pérez Brignoli, H. (Ed.). *Infant and Child Mortality in the Past* (pp. 74-88). Oxford: Clarendon Press.

Woods, R. (2001). The Demography of Victorian England and Wales. Cambridge: Cambridge University Press.

Web Appendix

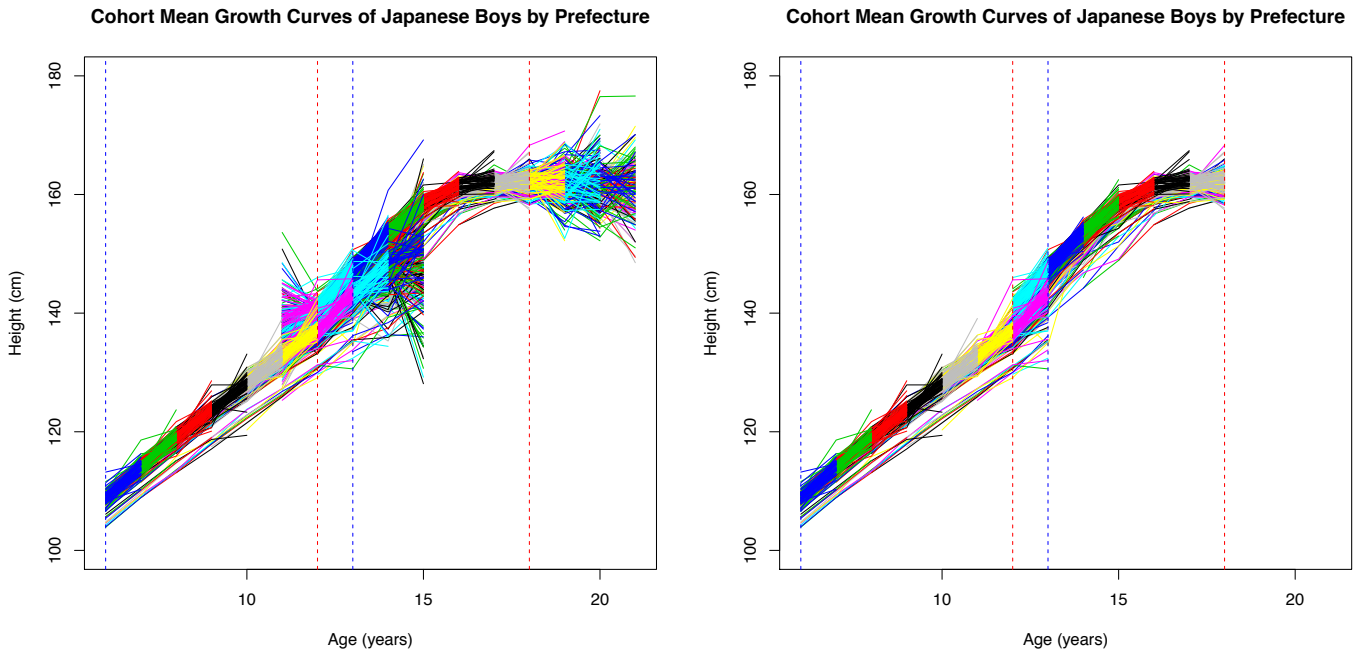
Appendix A: Supplementary Figures

Figure A1: Maps of average heights of six-year-old boys and girls in Japan, 1929-39.



Sources: See data appendix.

Figure A2: Individual cohort growth curves showing error that is eliminated by limiting the analysis to a narrower age range.



Notes: Left panel shows full data and right panel shows data in limited age range.

Sources: See data appendix.

Figure A3: Lexis diagrams of the prefecture dataset

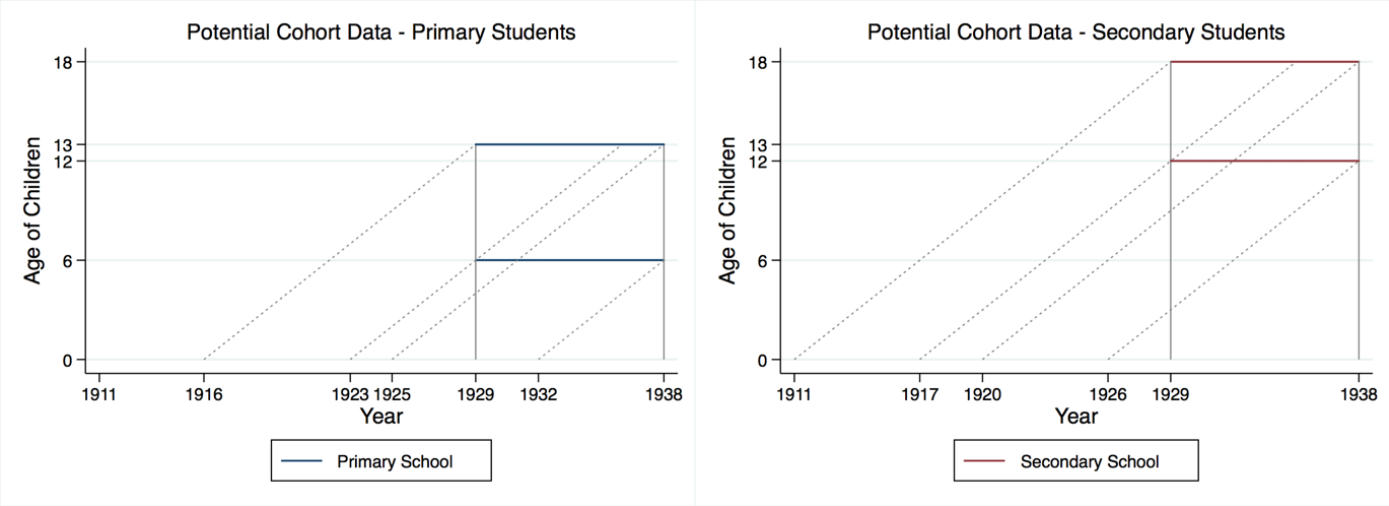
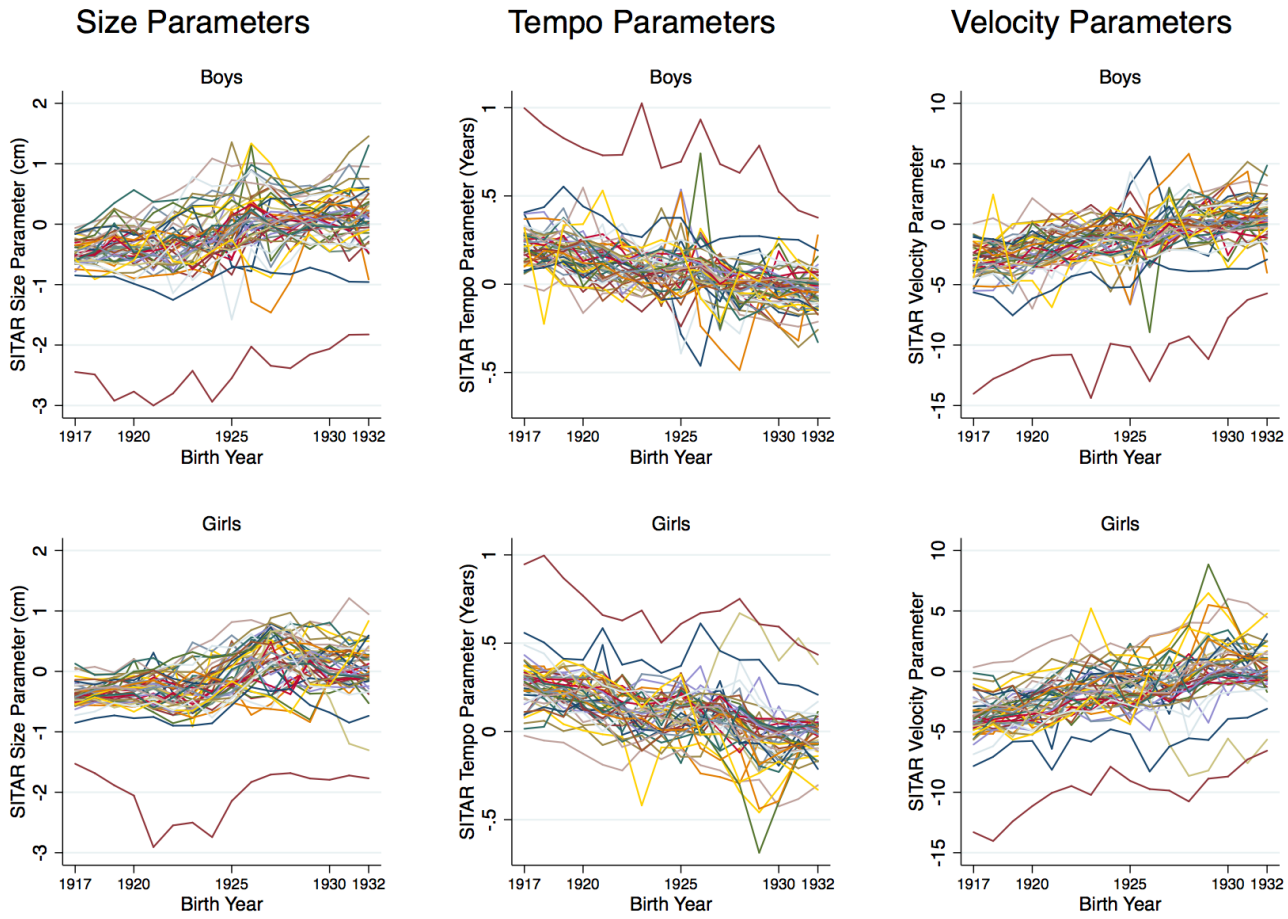


Figure A4: Predicted SITAR parameters for primary school boys and girls in each prefecture.



Notes: SITAR parameters for each birth cohort predicted using the model described in Table 1. Okinawa in brown is a clear outlier.

Sources: See data appendix.

Table A1: Descriptive statistics for growth pattern (SITAR parameter) regressions (Equation 2 and Table 2).

Variables	Definitions	Observations	Mean	Std. Dev.	Min	Max
Dependent Variables						
Size	SITAR size parameter	750	-0.17	0.54	-2.94	1.46
Tempo	SITAR tempo parameter	750	0.09	0.17	-0.49	1.02
Velocity	SITAR velocity parameter	750	-1.26	2.39	-14.37	5.82
Independent Variables						
Infant Mortality Rate	Number of infant deaths per 1000 live births	750	150.80	33.50	54.63	252.21
Average Temperature	Average annaul temperature, celsius	750	13.89	2.24	6.44	23.60
Population Density	Population per square kilometers	750	270.64	367.55	21.47	2706.87
Crude Birth Rate	Number of live births per 1000 population	750	33.99	3.89	22.21	46.29
% Peasant	Peasants per 100 workers in the agricultural sector	750	27.52	8.04	8.61	52.47
Rice Yield	Rice yields per hectare, hectoliter	750	33.62	6.31	2.45	49.48
Soy Yield	Soy yields per hectare, hectoliter	750	16.28	3.49	4.70	26.68
Milk production	Milk production per capita, liter	750	1.52	2.00	0.18	21.76
Doctors per 100 people	Number of doctors per 100 people	750	0.07	0.03	0.02	0.19
Midwives per 100 people	Number of midwives per 100 people	750	0.06	0.02	0.01	0.13
Hospitals per 100 people	Number of hospitals per 100 people	750	0.00	0.00	0.00	0.03

Sources: See data appendix.

Table A2: Descriptive statistics for instantaneous growth regressions (Equation 4 and Table 3).

Variables	Definitions	Observations	Mean	Std. Dev.	Min	Max
Dependent Variable						
Height (in cm)	Height of children	1645	120.29	8.13	103.90	136.30
Main Independent Variables						
Infant Mortality Rate	Number of infant deaths per 1000 live births	1645	121.63	23.90	45.98	212.92
Child Death Rate	Predicted child death rate	1645	20.82	3.75	12.42	35.18
Average Temperature	Average annual temperature, celsius	1645	13.91	2.33	6.44	22.53
Population Density	Population per square kilometers	1645	317.96	499.13	28.34	3225.42
Crude Birth Rate	Number of live births per 1000 population	1645	31.88	3.82	20.77	44.26
Milk Production	Milk production per capita, liter	1645	2.57	3.98	0.44	33.50
Coverage of Tap Water	Number of water taps per 100 people	1645	2.58	2.89	0.00	13.92
Doctors per 100 people	Number of doctors per 100 people	1645	0.07	0.04	0.02	0.23
Additional Controls						
% Peasant	Peasants per 100 workers at agricultural sector	1645	26.79	7.84	8.77	50.64
Coverage of Libraries	Number of public libraries per 10,000 people	1645	0.57	0.54	0.02	2.31
Rice yield	Rice yields per hectare, hectoliter	1645	35.54	7.20	3.80	51.68
Soy yield	Soy yields per hectare, hectoliter	1645	15.99	3.96	4.70	30.06
Midwives per 100 people	Number of midwives per 100 people	1645	0.08	0.02	0.02	0.12
Hospitals per 100 people	Number of hospitals per 100 people	1645	0.00	0.00	0.00	0.03
Additional Disease Controls						
Fetal Death Rate	Fetal deaths per 1000 births	1645	49.48	11.12	0.06	72.90
Congenital Infirmity Death Rate	Deaths from congenital infirmity per 1000 people	1410	1.09	0.34	0.02	1.96
Typhoid Death Rate	Typhoid deaths per 1000 people	1645	0.11	0.07	0.01	0.91
Tuberculosis Death Rate	TB deaths per 1000 people	1645	1.35	0.31	0.69	2.26
Beri beri Death Rate	Beri-beri deaths per 1000 people	1645	0.17	0.11	0.02	1.02
Measles Death Rate	Measles deaths per 1000 people	1645	0.12	0.09	0.00	0.72
Whooping Cough Death Rate	Deaths from whooping cough per 1000 people	1410	0.14	0.07	0.03	0.52

Sources: See data appendix.

Table A3: Regression of birth health conditions on the heights of children at age 6 and 10.

	Boys		Girls	
Dep. Var. Age	6	10	6	10
Birth Cohorts	1923-1933	1919-1929	1923-1933	1919-1929
Infant Mortality Rate	-0.002	-0.001	-0.001	-0.011**
	(0.005)	(0.004)	(0.003)	(0.005)
Average Temperature	0.270**	0.184	-0.075	-0.013
	(0.109)	(0.154)	(0.102)	(0.197)
Population Density	0.003	0.001	-0.003	-0.008**
	(0.003)	(0.001)	(0.003)	(0.003)
Crude Birth Rate	-0.004	-0.079*	0.014	-0.047
	(0.040)	(0.042)	(0.025)	(0.040)
% Peasants	-0.009	-0.073*	0.061*	-0.033
	(0.047)	(0.040)	(0.031)	(0.062)
Rice Yield	-0.002	-0.007	0.008	-0.012
	(0.008)	(0.009)	(0.006)	(0.011)
Soy Yield	0.024	0.015	0.020	0.007
	(0.041)	(0.026)	(0.018)	(0.027)
Milk Production per capita	0.009	0.021	0.054	-0.060
	(0.073)	(0.078)	(0.059)	(0.084)
Doctors per 100 people	3.509	2.406	-11.471	-4.018
	(7.236)	(6.066)	(8.094)	(5.355)
Midwives per 100 people	-0.434	-7.161	-4.939	5.717
	(7.421)	(6.686)	(4.700)	(14.786)
Hospitals per 100 people	-3.694	-149.430**	35.282	-19.091
	(23.213)	(59.985)	(54.187)	(55.867)
Prefecture Fixed Effects	yes	yes	yes	yes
Birth Year Fixed Effects	yes	yes	yes	yes
Prefecture-specific Time Trend	yes	yes	yes	yes
N	517	515	517	515
R-square	0.7707	0.7911	0.8229	0.7483

Notes: Standard errors in parentheses clustered at the prefecture level. * denotes significance at the 10% level. ** denotes significance at the 5% level. *** denotes significance at the 1% level.

Sources: See data appendix.

Table A4: Regressions of infant mortality in years before and after birth on the growth pattern (SITAR parameters) of Japanese children.

	Boys			Girls		
Dep. Var.	Size	Tempo	Velocity	Size	Tempo	Velocity
IMR -2	-0.001	-0.001	0.008	-0.000	-0.000	0.003
	(0.002)	(0.001)	(0.006)	(0.001)	(0.000)	(0.006)
IMR -1	-0.000	0.000	-0.000	-0.000	0.000	-0.002
	(0.001)	(0.000)	(0.006)	(0.001)	(0.000)	(0.005)
IMR	-0.002	-0.000	0.003	-0.000	0.000	-0.002
	(0.001)	(0.001)	(0.007)	(0.001)	(0.001)	(0.007)
IMR +1	-0.001	0.000	-0.006	-0.000	-0.000	0.004
	(0.001)	(0.001)	(0.007)	(0.001)	(0.000)	(0.006)
IMR +2	0.002	0.000	-0.004	0.000	0.000	-0.001
	(0.002)	(0.001)	(0.007)	(0.001)	(0.000)	(0.006)
IMR +3	-0.001	-0.000	0.004	0.001	-0.000	0.007
	(0.001)	(0.001)	(0.008)	(0.001)	(0.000)	(0.006)
IMR +4	0.000	0.000	-0.001	0.000	0.000	-0.001
	(0.001)	(0.000)	(0.006)	(0.001)	(0.000)	(0.006)

Notes: Coefficients on infant mortality rate (IMR) variables in various estimations of equation (3). Standard errors in parentheses clustered at the prefecture level. * denotes significance at the 10% level. ** denotes significance at the 5% level. *** denotes significance at the 1% level. All models include both prefecture and birth year fixed-effects as well as prefecture-specific time trends. The other control variables in Table 2 in the main text are also included in all specifications.

Sources: See data appendix.

Table A5: Regressions of infant mortality in years before and after birth on the heights of children at age 6 and 10.

	Boys		Girls	
Dep. Var. Age	6	10	6	10
Birth Cohorts	1923-1933	1919-1929	1923-1933	1919-1929
IMR -2	0.003	-0.002	-0.000	-0.007
	(0.003)	(0.004)	(0.002)	(0.006)
IMR -1	-0.002	0.000	0.005	0.011
	(0.004)	(0.005)	(0.003)	(0.008)
IMR	-0.002	-0.001	-0.001	-0.011**
	(0.005)	(0.004)	(0.003)	(0.005)
IMR +1	0.001	-0.000	0.001	-0.000
	(0.004)	(0.004)	(0.003)	(0.004)
IMR +2	-0.002	0.004	0.004	0.002
	(0.005)	(0.004)	(0.004)	(0.005)
IMR +3	0.002	0.002	-0.006	0.004
	(0.005)	(0.004)	(0.004)	(0.005)
IMR +4	0.003	-0.004	0.001	-0.012**
	(0.003)	(0.006)	(0.003)	(0.005)
N	517	515	517	515

Notes: Coefficients on infant mortality rate (IMR) variables in various estimations of equation (3) this time using the height of children at age 6 or 10 as the dependent variable. Standard errors in parentheses clustered at the prefecture level. * denotes significance at the 10% level. ** denotes significance at the 5% level. *** denotes significance at the 1% level. All models include both prefecture and birth year fixed-effects as well as prefecture-specific time trends. The other control variables in Table 2 in the main text are also included in all specifications.

Sources: See data appendix.

Appendix B: Data Appendix

The data were collected from a number of sources and constructed using the following methodology:

Anthropometric Data

Prefecture panel of anthropometric data

The panel dataset of heights of school children for the 47 prefectures was constructed from the reports entitled 'Statistics of School Physical Examination' (SSPE) published by the Physical Bureau, Ministry of Education (PEBME) between 1931 and 1943. The SSPE consisted of two types of reports: the *Koshiritsu shogakko chugakko kotojyogakko seitojido sintaikensatokei* (statistics of school physical examinations for public and private primary schools, junior high schools, and girls high schools in each prefecture and area) published by the Physical Education Bureau, Secretariat of Education in 1931, 1937, and 1938 (data for 1929—1936) and the *Gakko shintaikensatokei* (statistics of school physical examination) published by the PEBME in 1940, 1942, and 1943 (data for 1937—1939). For simplicity, we uniformly refer to these publications as the SSPE (1929—1939 editions).

Physical examinations were conducted in April of each year for all primary schools (*shogakko*), higher primary schools (*koto shogakko*), secondary schools (*chugakko*), and girls high schools (*koto jyogakko*). According to the estimates by Ogasawara (2017), the sample from the SSPE includes approximately 95% of children aged 6-11. The shares of boys aged 12-19 declined to 80.8, 69.3, 12.6, 9.2, 9.7, 2.5, 0.6, and 0.1 per cent, respectively, whereas the shares of girls aged 12-19 was 69.0, 57.1, 14.8, 12.0, 6.0, 0.7, 0.1, and 0.01 per cent, respectively. These shares declined so rapidly starting at age 12 because at that time compulsory education in Japan ended by age 12. Some children went up to higher primary schools or secondary schools or girls high schools after 12 years old.

Army height

Data on the army height is from the Annual Statistical Report of the Army Ministry (ASRAM). Almost all of the males ages 20 took physical examination for enlistment and thus this report includes the height of males aged 20. According to the Population Census of 1930, the number of males aged 20 was 607,136, while the number of examinees from the ASRAM was 592,161 (roughly 97.5 per cent of population).[1] Thus, these army heights provide very accurate and representative figures of the final height of males at that time. These data were used to measure selection in secondary schools. There were also used to estimate alternative SITAR parameters to test whether the inclusion of the selected data drawn from secondary schools had created substantial bias in the baseline SITAR parameters (described in Appendix C below).

[1] The number of examinees = (males who became 20-year old and took the examination) + (males who did not take the examination last year).

National time-series of anthropometric data

Time-series data on the average height, weight, and chest girth of children from 1900 to 1939 (except for 1921) are from *Gakko shintaikensa tokei* (Statistics of the physical examination of students) (1938 and 1939 editions) published by Physical Education Bureau, Ministry of Education (1942a; 1942b; 1943a; 1943b). These data were used to show long-run changes in the growth pattern of children in Figure 2 in the main text.

Control variables

Mortality and population

Data on live births and fetal deaths are from *Nihonteikoku jinkodotai tokei* (The Vital Statistics of the Empire of Japan; VSEJ) (1916–1931 editions) and *Nihon jinkodotai tokei* (The Vital Statistics of Japan; VSJ) (1932–1940 editions) published by the Statistics Bureau of the Cabinet between 1919 and 1942. Data on the number of infant deaths are from *Nihonteikoku tokei nenkan* (Statistical Yearbook of the Japanese Empire; SYJE) (vol.37–55) and *Dainihonteikoku tokei nenkan* (Statistical Yearbook of the Greater Japanese Empire; SYGJE) (vol.56–59) published by the Statistics Bureau of the Cabinet (1919–1941). The number of deaths from typhoid fever, diarrhea, measles, beriberi, tuberculosis, whooping caught, and congenital infirmity are from *Nihonteikoku shiin tokei* (Statistics of Causes of Death of the Empire of Japan; SCDEJ) (1916–1931 editions) published by the Statistics Bureau of the Cabinet between 1919 and 1939 and *Shiin tokei* (Statistics of Causes of Death; SCD) (1932–1938 editions) published by the Statistics Bureau of the Cabinet between 1925 and 1939.

Data on the population counts before 1935 is obtained from the SYJE (vol.34–55) and SYGJE (vol.56). Data on the number of population after 1936 is taken from the online database of the Statistical Survey Department, Statistics Bureau, Ministry of Internal Affairs and Communications (http://www.stat.go.jp/data/chouki/zuhyou/02-05.xls, accessed on 23rd March 2017). The information on the prefectural area in 1916 is obtained from the SYJE (vol.37).

The crude birth rate is the number of live births per 1,000 people. The infant mortality rate is the number of infant deaths per 1,000 live births. The fetal death rate is the number of fetal deaths per 1,000 births. The other cause-specific death rates are defined as the number of each death per 1,000 people.

Child death rate

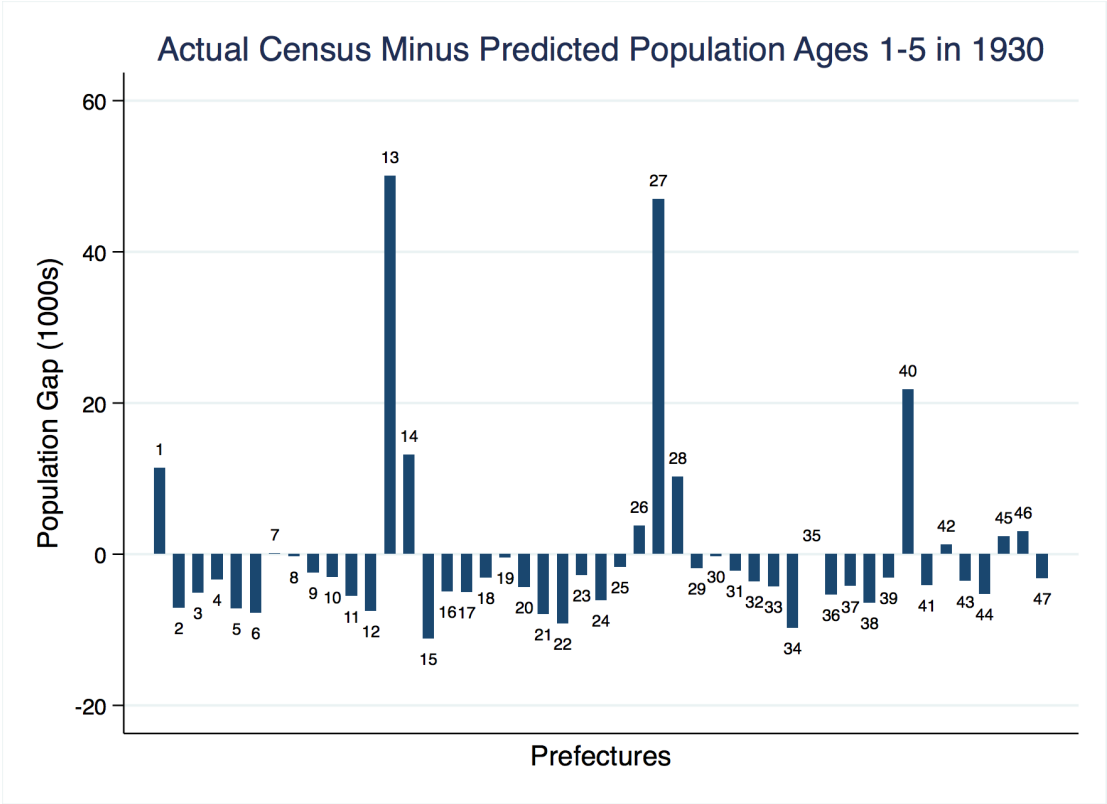
The information on the number of child deaths between ages 0 and 5 are from the VSEJ (1916–1931 editions) and VSJ (1932–1940 editions). Age-specific population data are compiled using *Kokusei chosa hokoku, fukenhen* (Population Census of Japan, Prefectural Part; PCJPP) (1930 edition) published by the Statistics Bureau of the Cabinet (various years). The PCJPP consists of 47 prefectural reports.

Unfortunately, the number of children aged 1-5 was not reported annually. Thus, in order to calculate a child death rate, we estimated the population of children at risk using perpetual inventory/life table methods. We observe the number of births and child deaths at

each age 0 to 5 in each prefecture and year. Thus, one can construct the population at risk by subtracting child deaths at later ages from the number of children born in a given cohort.

This method assumes that vital registration was perfect and migration was negligible. In actuality, when comparing the predicted population at risk from the vital statistics with the 1930 census, these assumptions do not hold perfectly. There appears to have been a very small amount of under-reporting of births across all prefectures, though this was much larger for Okinawa. Thus, the number of births were adjusted upward by 2.83 per cent across all prefectures with Okinawa's being adjusted by 28.1 per cent. This seems to have corrected for the imbalance of deaths and births so that the total estimated population at risk in 1930 matches the 1930 census. However, when looking at the difference between the actual population in 1930 and the estimated population at risk across prefectures (Figure B1), we do see our method underestimates the population at risk in urban prefectures because it does not take account of major cities. Prefectures 13, 27 and 40 include Tokyo, Osaka and Fukuoka respectively. These imbalances are somewhat problematic, but we do not adjust our life table method further. We believe that a dynamic estimate of the population at risk even with these limitations is better than assuming that the population of children aged 1 to 5 followed general trends in the population of Japan as a whole between censuses. In any case, the population gap was unlikely to change drastically from one year to another and we estimate all specifications with prefecture fixed effects, so the modest measurement error introduced by using the estimated population at risk is unlikely to influence our results.

Figure B1: Difference between actual population in 1930 and estimated population at risk.



Notes: Population gap is the actual population of children age 1-5 reported in the 1930 census minus the estimated population at risk from the perpetual inventory method. Clearly, migration to large cities is somewhat problematic.

Sources: See data appendix.

Health infrastructure and sanitation

Data on the number of doctors, midwives, and hospitals are from the SYJE (vol.37—55) and SYGJE (vol.56—59). Data on the number of water taps is from the SYJE (vol.42—55) and SYGJE (vol.56—59). The number of water taps in 1929 are linearly interpolated using both pre- and post-year data due to missing values. Coverage of doctors, midwives, and hospitals are defined as the number of doctors, midwives, and hospitals per 100 people, respectively. Coverage of tap water is the number of water taps per 100 people.

Peasants and agricultural production

Data on the number of peasants and agricultural households after 1924 are obtained from the SYJE (vol.45—55) and SYGJE (vol.56—59). The data before 1923 are from the *Todofuken nogyo kiso tokei* (Basic Statistics of Agriculture in Japanese Prefecture; BSAJP) edited by Nobufumi Kayo. The production of rice and rice acreage are taken from the SYJE (vol.37—55) and SYGJE (vol.56—59). The information on the milk and soy production as well as the soy planted area are also obtained from the BSAJP. Share of peasants is defined as the number of peasants per 100 workers at agricultural sector. The rice and soy yield are in hectoliter per hectare. The milk production is in liter per capita.

Meteorological data

Data on the meteorological variables, such as monthly average of the annual temperature and total amount of annual precipitation are downloaded from the database of the Japan Meteorological Agency. The meteorological observation stations were usually located in each city. We replicate some missing data on the observations at the nearest meteorological observing station. The data are publicly available and can be downloaded from http: //www.data.jma.go.jp/gmd/risk/obsdl/index.php (accessed on 4th April 2017).

Appendix C: Further Elaboration of the SITAR Analysis

We only present one potential SITAR model and configuration of data in the main text but we did estimate SITAR models for several configurations. This appendix discusses these briefly and justifies the choices we have made.

The first potential issue that we faced was outliers in cohort growth curves. Especially when looking at the female cohort growth curves, it became clear that there were a number of major outliers that were skewing the SITAR results. Visual outliers were first checked in the original record for transcription errors. However, in order to produce a more systematic way of excluding outliers, we converted the girls' heights into Z-scores of the modern WHO reference. We then excluded points where both the difference between the WHO Z-score in the current year and the previous year and the WHO Z-score in the current year the following year exceeded 0.5. Mean growth curves like the ones used in this paper should not jump up or down by a half standard deviation of modern height unless the underlying sample is very small and thus unreliable. This led to the exclusion of 47 outliers or 0.65 per cent of the sample. As expected, excluding these outliers reduces the standard deviation of the residuals in the SITAR estimation from 0.85 cm to 0.71 cm (see Table C1). Excluding the outliers also substantially reduced the BIC from 21,794.84 to 19,883.53. Thus, we feel that excluding these outliers is justified and we have used the model without outliers as our baseline specification in the paper. The male data did not suffer from similar problems so we did not exclude any data.

Table C1: Summaries of SITAR models calculated for female cohort growth curves excluding and including outliers.

	Girls No Outliers (Baseline)			Girls Including Outliers		
	Standard Deviations	Correlations		Standard Deviations	Correlations	
		Size	Tempo		Size	Tempo
Size (cm)	0.822			0.810		
Tempo (years)	0.318	-0.099		0.269	0.031	
Velocity (fractional)	0.035	0.244	-0.917	0.036	0.069	-0.994
Residual	0.713			0.849		
Degrees of Freedom	4			5		
BIC of Model	19,883.53			21,794.84		

Notes: See text for definition of outliers. SITAR models were estimated with varying degrees of freedom. We used the degrees of freedom that provided the lowest BIC score in line with Cole *et al.* (2010) as our baseline reported here.

Sources: See data appendix. Statistical citations: R Core Team (2016) and Cole (2017).

Another worry is that relying on the selected secondary data to draw the growth curve at later ages somehow biases the SITAR parameters predicted for the primary school children. The adjustment meant to normalise the primary and secondary curves is solely based on their overlap at ages 12 and 13. There are advantages and disadvantages about this overlap. It would clearly be beneficial if there were overlap over a wider range of ages to help SITAR determine the precise differences in size, tempo and velocity between the primary and secondary groups.

However, we are lucky in a way that the overlap occurs during the pubertal growth spurt since this is when differences in size, tempo and velocity are most easily measured. In order to test whether the narrow age overlap strongly influenced the SITAR parameters for primary school children, we tried a number of alternative estimation strategies that could mitigate the influence of the secondary data in our analysis. These are presented in the paragraphs below.

First, to entirely mitigate the secondary data, we tried estimating the cohort SITAR parameters only using the primary data for children aged 6-12. Unfortunately, this led to many errors warnings in the estimations because the SITAR model could not be fit properly. This is likely because SITAR struggles to distinguish vertical shifts in size from horizontal shifts in tempo when the growth curve is more or less linear as it is before the pubertal growth spurt.[2] Thus, it was not possible to compare the predicted SITAR parameters for primary school children only with the predicted parameters from our baseline model presented in the main text that included both primary and secondary students.

Second, we tried an additional robustness check incorporating the mean heights of men at age 20, measured when men registered for potential military service. These heights for each prefecture were included as part of the primary school data because both reflected over 95 per cent of the population and did not suffer from the selection bias present in the secondary data. This increases the age overlap of the primary and secondary data and provides another data point at later ages to help the SITAR model distinguish the primary and secondary data. However, there were a few key limitations with this robustness check. First, because it incorporates data drawn from military enlistment, we could only estimate the SITAR model for this robustness check for men. Second, the analysis could only be conducted using period rather than cohort growth curves as in the baseline analysis. We could not add the military data to cohort growth curves because the cohort growth curves included in the prefecture panel run from 1917 to 1932, which would correspond to men enlisting between 1937 and 1952. Data on the heights of military recruits is not available for many years within and shortly after World War II, and we worried that the loss of life during the war would bias our results if we connected the survivors at age 20 after the war with their birth conditions before the war. Thus, we estimated the SITAR parameters for the period growth curves 1929-38, i.e. children measured in a prefecture in 1929, etc. and compare this with the period SITAR parameters estimated without the army data.

Looking at the residuals and BIC of the SITAR estimation (see Table C2), it is clear that adding the army data improves the SITAR estimation reducing both the BIC and the standard deviation of the residuals, though the correlation between the parameters becomes worryingly high. On first appearance, this might lead one to believe that linking the primary data and army data was very important. However, when we compare the predicted SITAR parameters using the two separate models, the correlations between the size, tempo and velocity parameters was 0.93, 0.96 and 0.99 respectively. Despite the more efficient estimation, the addition of the army data barely influenced the predicted SITAR parameters. Thus, we have used the cohort growth curves for primary and secondary children only to estimate the SITAR parameters in the model. Although this is not perhaps the most optimal method, we believe that the error introduced by the differences in the estimation procedure is very small.

[2] Thanks to Tim Cole for making this point.

Table C2: Summaries of SITAR models calculated for period growth curves with and without linking the primary school data to the heights of army conscripts.

	Boys Period			Boys Period with Army		
	Standard Deviations	Correlations		Standard Deviations	Correlations	
		Size	Tempo		Size	Tempo
Size (cm)	0.961			1.006		
Tempo (years)	0.217	-0.295		0.206	-0.525	
Velocity (fractional)	0.035	0.406	-0.991	0.032	0.673	-0.982
Residual	0.730			0.646		
Degrees of Freedom	6			5		
BIC of Model	21,113.95			15,627.33		

Notes: SITAR models were estimated with varying degrees of freedom. We used the degrees of freedom that provided the lowest BIC score in line with Cole *et al.* (2010) as our baseline reported here.

Sources: See data appendix. Statistical citations: R Core Team (2016) and Cole (2017).