Health Shocks, Recovery and the First Thousand Days: The Effect of the Second World War on Height Growth in Japanese Children[1]

Eric B. Schneider,[a] Kota Ogasawara,[b] and Tim J. Cole[c] [a]London School of Economics and Political Science, CEPR, e.b.schneider@lse.ac.uk [b]Tokyo Institute of Technology, ogasawara.k.ab@m.titech.ac.jp [c]UCL Great Ormond Street Institute of Child Health, tim.cole@ucl.ac.uk

ARTICLE HISTORY

Compiled September 17, 2021

Forthcoming at *Population and Development Review*

ABSTRACT

This paper uses the health shock of the Second World War on Japanese civilians to understand the effects of health shocks at different developmental stages on children's long-run growth pattern and to test whether recovery is possible after an early life health shock. We construct a prefecture-level dataset of mean heights of boys and girls from ages six to nineteen from 1929 to 2015. Linking the heights recorded at different ages for the same birth cohort, we measure a counterfactual causal effect of the health shocks during the Second World War on the cohort growth pattern of children. We find that the war effect was greatest for cohorts exposed to the war in late childhood and adolescence: these cohorts were 1.7-3.0 cm shorter at adulthood and had delayed pubertal growth and slower maturation than they would have had if the war had never occurred. However, there were not persistent health penalties for children exposed to the war in early life, suggesting that catch-up growth was possible as health conditions improved after the war. These findings challenge the thousand-days consensus that children cannot recover from nutritional shocks in early life and indicate that adolescence is a sensitive period for health shocks.

KEYWORDS

Child growth; health shocks; catch-up growth; nutrition; Japan; Second World War

[1] We thank Janet Hunter, Deborah Oxley and the LSE Global Health Initiative Peer Review Group for comments on drafts of the paper. We also thank conference participants at the All-UC–APEBH Meeting in Pasadena, LSE Historical Economic Demography Workshop, the European Historical Economics Society Meeting in Paris, the Measuring Well-being in the Past Workshop in Utrecht and the NYUAD-CEPR Economic History Conference in Abu Dhabi and seminar participants at LSE, Oxford, Paris Sorbonne and Yale for useful comments. Minami Yumitori provided excellent research assistance. The usual disclaimer applies.

1. Introduction

Child malnutrition is a major source of ill health globally with large consequences for societies where malnutrition is prevalent. Child malnutrition, often proxied via the stunting rate,[2] matters because health conditions in the so-called first thousand days from conception to age two are critically important for later child growth, development and human capital formation (Almond and Currie 2011; Currie 2009; Victora et al. 2010; Wells 2017). Consequently, eradicating child stunting has been an important target in development policy with a prominent place in both the Millennium and Sustainable Development Goals. Child stunting rates have been declining since the 1980s (Black et al. 2013), yet as of 2019, there were still 149 million children globally who were stunted (UNICEF et al. 2019).

Despite this large number of stunted children, the focus of the global health literature has been on preventing stunting from occurring in the first place rather than intervening to improve the health status of already stunted children. This focus exists because there is wide consensus across a range of disciplines from development economics to human biology to nutrition that children experiencing health shocks in the first thousand days are unable to experience catch-up growth, faster than normal growth once health conditions improve at later ages, or at least the effectiveness of interventions decreases substantially outside the first thousand days (Alderman et al. 2006; Cunha and Heckman 2007; Hoddinott and Kinsey 2001; Proos and Gustafsson 2012; Wells 2017). However, there has been growing emphasis on testing whether health interventions outside the first thousand days can overcome early life health shocks. This research is still in its early stages, but there is evidence that at least for knowledgebased human capital, later interventions can overcome some of the deprivation of early-life shocks (Adhvaryu et al. 2018; Akresh et al. 2021; Almond et al. 2018; Arthi 2018). There is also evidence that catch-up growth may be possible in adolescence with or without health interventions (Prentice et al. 2013; Steckel 1987),

[2] The share of children who are too short for their age, i.e. less than 2 standard deviations below the mean in WHO Growth Standards.

though this has been challenged on methodological grounds (Leroy et al. 2015; Lundeen et al. 2014). Thus, it is worth considering again whether health interventions can ameliorate the consequences of early-life health shocks in order to help the millions of children already suffering from stunting.

In addition, recently a range of scholars have questioned whether the first thousand days is the only developmental stage where health shocks matter. A number of studies have begun to emphasise adolescence as another critical window for health shocks (Prentice et al. 2013) with several papers finding children more sensitive to health shocks in late childhood and adolescence than in early life (Akresh et al. 2021; Depauw and Oxley 2019; Schneider and Ogasawara 2018; van den Berg et al. 2014). Understanding the developmental periods when children are most sensitive to health shocks is important because it can guide policy makers about which children should be targeted with scarce resources in times of deprivation.

This paper uses food shortages and the deterioration in the health system in Japan from 1942 to 1947 associated with the Second World War to answer two questions: 1) what is the relative importance of health shocks at different developmental stages on the growth pattern of children; and 2) to what extent is catch-up growth possible for children experiencing poor health conditions in the first thousand days. Japan during the Second World War is an interesting case study to explore these two questions for several reasons.[3] First, although Japan was the industrial leader in Asia in the first half of the twentieth century and had a well-developed public health system, its GDP per capita and health outcomes were not dissimilar to outcomes in low and middle income

[3] A number of studies in development economics, population studies and economic history explore the influence of war on child health in other contexts. Akresh et al. (2011), Akresh et al. (2012b) and Akresh et al. (2012a) studied the effect of civil conflict in Rwanda in 1990-91, the Eritrean-Ethiopian conflict in the late 1990s and the Biafran War in Nigeria in the late 1960s respectively, finding that conflict led to worse child health and higher levels of child stunting. These findings are also supported by the historical literature on the effects of the Second World War on child growth. Mean heights of children declined in late childhood and adolescence in Germany, Italy, Belgium, Norway and to a lesser extent Finland during the Second World War, though Denmark and Sweden were left relatively untouched (Angell-Andersen et al. 2004; Brundtland et al. 1980; Daniele and Ghezzi 2018; Ellis 1945; Howe and Schiller 1952). Interestingly, heights actually increased during the Second World War in Britain in part because rationing improved the diets of British children (Harris 1995, p. 166; Magee 1946). However, despite considerable study of the European experience, very little has been written on the effect of the Second World War on Japanese children.

countries today where child malnutrition is a major problem. In 1936, before the beginning of the war in Asia, Japan had a GDP per capita of 3,986 2011 international dollars, similar to Bangladesh in 2018 (Fukao et al. 2015; Maddison Project et al. 2020). The infant mortality rate was falling but was still 116.7 infant deaths per 1000 live births (Japan Statistical Association 1987, p. 210), 44% higher than the country with the highest infant mortality rate in 2019, Central African Republic (UN Inter-agency Group for Child Mortality Estimation 2020). Finally, rough estimates of the stunting rate suggest that it was c. 45% before the war (Schneider 2018), similar to Guatemala in 2015 (UNICEF et al. 2019). Thus, although there are differences between mid-twentieth century Japan and low and middle income countries today, Japan's experience during the Second World War can yield insights into children's response to nutritional shocks and their recovery afterwards that are applicable to low and middle income countries today.

Japan is also a useful case study because the health shocks during the war were severe. As described in detail in Section 2, the decline in the availability of nutrition as well as the deterioration in health infrastructure during the war represents both an absolute and relative health shock affecting a wide range of birth cohorts at different ages. This allows us to test the sensitivity of children at different ages to health shocks. Finally, Japan's economic and health miracle in the decades after the war is the perfect opportunity to understand whether catch-up growth is possible following an early life health shock. In the post-war decades, Japan experienced rapid economic growth, large life expectancy gains and a more diversified and animal protein-rich diet (Cabinet Office 2020; Ikeda et al. 2011; Takahashi 1984). These conditions were very conducive to child growth, and mean child heights increased dramatically (Cole and Mori 2017; Headey et al. 2018). Thus, the post-war period brought with it a wide range of positive health interventions that we can use to understand the persistence of the war shock to health.

To understand at what ages children are sensitive to nutritional shocks and whether they are able to recover from them, we construct a prefecture-level dataset of population-representative mean heights of Japanese boys and girls from ages six to nineteen from 1929 to 2015.[4] We use the SuperImposition by Translation and Rotation (SITAR) growth model to predict cohort growth curves, the growth of the same group of individuals in a prefecture as they aged across time (Cole et al. 2010). We compare counterfactual growth curves that children would have experienced in the absence of the war with the growth curves children actually achieved after the war. This allows us to precisely quantify the effect of the war on children's growth and understand the influence of health shocks at different ages.

We also extend this literature by testing the effects of war on all three aspects of the growth pattern, the heights of children at a particular age, the timing of the pubertal growth spurt and the speed of maturation, rather than simply height at a particular age (see Figure 1). Analysing the growth pattern is interesting because there have been large changes in all three aspects of the growth pattern over time as populations became healthier. Not only has adult height increased, but the timing of the pubertal growth spurt has shifted to earlier ages and maturation has accelerated (Ali et al. 2000; Cole 2003; Cole and Mori 2017; Gao and Schneider 2021). Taking account of the growth pattern is also necessary because health shocks in early life may influence the speed of maturation and timing of puberty (Schneider 2017a). Thus, simply analysing changes in heights at a given age may overstate the consequences of a health shock on child height.

The paper unfolds as follows. Section 2 provides historical context on nutritional and health conditions during and after the Second World War. Section 3 presents the data and discusses their representativeness. Section 4 introduces the SITAR model and explains our identification strategy for estimating the causal effect of the war. Section 5 presents the results. Section 6 discusses the potential for selection by birth cohort and

[4] There have been a number of studies of Japanese children's growth using versions of the data we employ in this paper (Bassino and Kato 2010; Ogasawara 2017, 2018; Schneider and Ogasawara 2018; Yokoya and Higuchi 2016). Three use the national-level data to conduct studies on long run change in Japanese children's growth and briefly consider the effects of the Second World War on growth. These mention that the timing of the pubertal growth spurt increased after the war, but none developed a strong counterfactual, analysed the effects of the war for any other dimensions of the growth pattern or tested the effect of the health shock for cohorts exposed at different ages (Ali et al. 2000; Kagawa

measurement error and shows that neither explains our results. Finally, Section 7 concludes.

2. Historical Context

In order to understand the consequences of the war shock to health on Japanese children, we must first carefully describe and evaluate the deterioration in health during

et al. 2011; Mosk 1996).

the war and the subsequent improvements in health afterwards. We will describe nutritional conditions during the war first, then consider other aspects of the health environment before finishing with a discussion of the improvement in conditions following the war.

To begin, the nutrition available to the Japanese civilian population deteriorated substantially during the war. Although consumption declined somewhat from the beginning of the Sino-Japanese war in 1937 because of inflation and reductions in imports of food staples from the colonies (Hunter 2008), 1941 was the real turning point. Like many governments during the Second World War, the Japanese government instigated food rationing to ensure that there was enough food for the army and the civilian population. Rationing of rice began in the six principal cities in April 1941 and was expanded to the entire country by July 1942. Rationing for other non-staple foods such as vegetables, fruit, meat, fish and seafood began between 1940 and 1941. Rationing meant that civilians could no longer purchase food on the open market but instead were given tickets to collect a certain quantity of food from official dispensaries (Johnston 1953, pp. 165, 187).

Table T1 shows the allocation levels of staple foods, which varied by sex, age and work intensity. These ration levels served as an upper bound for government food allocation because the government often had to reduce temporarily the quantities given when poor weather affected agricultural production or when certain products were out of season. In addition, as the war progressed, additional staple crops such as wheat, barley, potatoes and sweet potatoes were included as a part of the staple ration calorie level, making it harder for people to purchase these other staple crops to supplement their rations. The ration levels were adjusted somewhat during the war, raising the staple ration for children aged 3-14 and providing more calories to workers directly in factories rather than in their homes (Johnston 1953, p. 202). However, these rations are very low compared to estimated calorie requirements by sex and age produced by the FAO with children receiving 35% to 50% of a typical child's energy expenditure (FAO 2004, pp. 26-27).

Of course, people did not only consume rice and other staple crops. Table 1 shows a supply-side calculation of calories and protein per capita available to civilians per day for each year as the war progressed. The amount of calories started at 2,105 kcal/capita/day in 1941 and declined across the war period reaching 1,793 kcal/capita/day by 1945. This decline was mainly driven by a sharp reduction in rice imports from Korea, Taiwan and Southeast Asia, which fell to 10% of the prewar level by 1945, but domestic production fell somewhat as well (Johnston 1953, p. 151,201). These population levels of calorie consumption are lower than estimates for Japan in the 1920s and 1930s, around 2,325 kcal/capita/day (Mosk 1978, p. 279; Hayami and Yamada 1970, p. 81), and of later estimates of consumption in 1955, 2,217 kcal/capita/day (Sanderson 1987).[5] It is also low by global standards today. Only six countries had lower calorie consumption per capita per day in 2013 according to the FAO's Food Balance Sheets (FAO 2020).

Interestingly, Table 1 shows that protein levels did not decline as sharply as calories during the war. In fact they were very similar to their pre-war (1934-38) values: 63.8 grams per capita per day. Most of the protein was derived from cereals and soy products with only 16% of protein coming from animal sources and fish making up 74% of animal protein (Johnston 1953, p. 277). Protein levels likely remained stable during the war because of increasing soy imports from Manchuria across the war (Cohen 1949,

[5] This pattern of the nutritional shock is very similar when controlling for changes in the age structure and sex composition of the population across the war (see Appendix C and Table T4 for details). The removal of millions of prime-age men from the civilian population did ameliorate the nutritional shock somewhat, but this effect was very small relative to the overall decline in food availability.

p. 370; Johnston 1953, p. 159). These high levels of protein may appear to be good for child growth, but the low levels of animal-based protein likely inhibited growth during the pre-war period and during the war (Headey et al. 2018; Takahashi 1984).

In addition to the temporal variation in the nutritional shock, the availability of calories varied across Japan depending on the extent to which people could produce their own food or purchase food on the black market (Scherer 1999).[6] Unfortunately, data on the black market are difficult to find and are often not comparable over space and time. The Institute for Science of Labour did conduct studies of black market purchases in March 1944 and September-October 1944 (see Table T2). They found that black market purchases of staple foods were relatively small (0.5-9.0%), but were much higher for vegetables, fish and meat (10-69%). In March 1944, the black market share was lower in Tokyo than in other prefectures, but conditions had worsened significantly in Tokyo by September 1944 (United States Strategic Bombing Survey 1947, pp. 28-29). These figures suggest that the black market was already important in the allocation of food even before the hardest months of the war in 1945. Table 2 presents additional data on non-ration shares of consumption for six cities in July 1945, reflecting the worst part of the nutritional crisis at the end of the war. These data are based on household budget surveys and show that 16-36% of calories and 21-49% of protein came from the black market by the end of the war, suggesting that the black market continued to be important in 1945.

Table 2 also highlights that nutritional conditions varied substantially across Japan. Average calorie intake from rations across the cities was very low and certainly could not have supported the population. However, total calorie intake, including non-ration food consumption, was much higher and varied substantially from 1,677 kcal/capita/day in Kyoto to 2,026 kcal/capita/day in Yamaguchi in Southwest Japan. The ability of people to procure food outside of official rations also varied across the

[6] While it is beyond the scope of this article to determine how much the food shortages were driven by food availability decline or exchange entitlements (Sen 1977), it seems clear that both factors were present and important. The fact that nearly all prefectures experienced negative change in the growth pattern across the war suggests that the shortages could not have been overcome simply by redistributing civilian food supplies within Japan. It is more tricky to determine whether the food shortages would have disappeared if civilians were given similar priority to soldiers in distribution.

cities depending on the strength of black markets and the proximity to agricultural areas. Urban dwellers often made trips to the countryside to purchase food at black market rates from rural dwellers (Scherer 1999, pp. 110-111). Farmers also had higher staple food allocations and were more likely able to produce food for their own consumption in gardens than urban residents. Thus, people in rural areas were shielded somewhat from the worst of the food crises (Johnston 1953, pp. 192-193).

Despite the grievous nutritional shortages for the civilian population, there were a few programmes that attempted to mitigate the worst effects of the war for pregnant women and children. These are described in detail in Appendix Section E.2. Pregnant mothers were given upgraded rations with milk, sugar and additional brown rice and the little milk production that existed was prioritised for infants whose mothers could not breastfeed them (Johnston 1953, p. 202, 208; Yoshida 1995, p. 267). In addition, 1.3 million urban children were evacuated to the countryside. Children evacuated to relatives may have experienced better health conditions (Partner 2007, p. 147) similar to children living in rural areas, but qualitative evidence suggests that children evacuated in school groups experienced hunger with conditions worsening toward the end of the war (Yamashita 2013). However, these mitigating factors were limited and were not able to counteract the deterioration in the health environment overall (Yoshida 1995, pp. 265-66).

Beyond nutrition, Koike (2019) analysed real outlays of households between 1940 and 1945. He found that by 1945 real outlays for urban and rural households fell to 30% and 60-70% of their 1940 level respectively, highlighting that people in rural areas may have been more protected since they could engage in self-provisioning. Koike then computes a real consumption (expenditure) per capita estimate and shows that real consumption per capita fell to the same level as 1875-80 by the end of the war. Thus, the war did not simply produce a nutritional shock. It limited household consumption very dramatically relative to the pre-war period.

Other aspects of the health system also deteriorated during the war. The number of doctors available to civilians declined from 67,612 in 1941 to 11,136 in 1944. Medicine was also preferentially distributed to the army. Many hospitals and other medical facilities were destroyed in the bombing of major cities, and the facilities surviving struggled to cope with the masses of injured and sick people. Matsuzawa Hospital in Tokyo saw its case fatality rate rise from 13.6% to 40.9% between 1943 and 1945. The share of deaths in the hospital due to undernutrition also increased during the war from 40.8% in 1943 to 50.0% in 1944 to 62.3% in 1945, showing how difficult the nutritional situation had become. Incidence rates of infectious diseases such as dysentery and typhoid fever also increased dramatically during the war (Japanese Statistical Association 1982, pp. 513-515, 522).

Unfortunately, nutritional conditions did not improve substantially in the aftermath of the war. Despite holding rations at the extremely low level set at the end of the war in the rest of 1945 and 1946, the Allied Occupation and Japanese government still struggled to meet this ration level in that year.[7] The domestic rice and sweet potato harvests in 1946 improved dramatically from the very low harvests of 1945 (Aldous 2011, pp. 242-45), but the growth of the population due to repatriation and unstable food imports meant that the early war level of rations (see Table T1) was not reached until 1948 and then only by supplementing the staple rations with sugar. The black market continued to be a major problem during the occupation period with only 55% of food consumed being allocated through the rationing system in December 1945 (Aldous 2011, p. 239) and other anecdotal evidence suggesting that this continued in 1946 (Johnston 1953, pp. 162-63).

After 1947, health conditions began improving dramatically. The Allied Occupation began promoting a better quality diet with the introduction of more animal protein, and consequently average consumption from animal protein increased. Protein from animal sources did not change much across the war, shifting from 10.4 grams per capita per day in 1934-38 to 10.5 grams in 1947 (Aldous 2011, p. 252; Johnston 1953, p. 277). However, after the war, animal protein consumption increased to 17.2 grams in 1950 and to 27.7 grams by 1963 (Aldous 2011, p. 252; Mosk 1978, p. 280). One way that the

[7] In the final months of the war, the government had to abandon its 330 grams of rice equivalent ration, reducing the ration to 301 grams of rice per typical adult (Johnston 1953, p. 214).

Allied Occupation achieved this goal was to reinitiate a school meals programme beginning in December 1946 that sought to provide children with protein-rich meals by introducing dried skimmed milk. In the first couple of years after the programme began, the Allied Occupation had trouble sourcing dried milk. However, by September 1951, the programme was providing four million children with regular meals of 600 calories and 25 grams of protein (Aldous 2011, pp. 253-54). This programme was very popular and was made permanent by the post-war Japanese government in 1954, leading to a substantial rise in the consumption of milk and eggs in the post-war period (Kagawa et al. 2011; Takahashi 1984). The 1950s also saw rapid economic growth, improvements in health infrastructure, sharp declines in infant and child mortality and large increases in life expectancy (Ikeda et al. 2011), all of which would create the conditions necessary for war-exposed cohorts to experience catch-up growth.

To summarise, the Second World War created a severe nutritional and health crisis for the Japanese civilian population. Food availability fell dramatically as the war progressed and other aspects of the health system also deteriorated as healthcare resources were focused on the military. Poor conditions continued for a few of years after the war until the end of 1947 when the Allied Occupation was able to begin improving the situation. These conditions certainly provided a shock to both nutrition and health infrastructure that we would expect to have a strong influence on child growth. However, the large improvements in nutritional and health conditions following the war, the most important of which might have been the introduction of school meals and more animal protein to the diet, suggest that catch-up growth may have been possible from the 1950s onward.

3. Data

To analyse how the food shortages and other health shocks of the Second World War affected Japanese children, we have constructed a panel dataset of the mean heights of boys and girls at various ages in each of Japan's 47 prefectures measured (with gaps) from 1929 to 2015.[8] In its final format the dataset includes over 80,000 observations of prefecture mean heights for over 6,000 prefecture-birth cohorts for each sex. We collected the data from various reports produced by the Japanese government (see Appendix A).

Figure 2 presents the structure of the data using boys as an example. Each graph shows the population weighted mean height for boys across all prefectures at a given age measured in a given year. These graphs serve as a summary, but we use the prefecture-level data in all estimations below (i.e. the 47 different versions of these figures, one for each prefecture). The period growth curves reflected in Figure 2a show the data as they were collected by the Japanese authorities, as cross-sections in a number of years. There are clearly periods where missing data prevent a more thorough study, especially during the Second World War, but overall the dataset is a very rich collection of repeated, prefecture mean height measurements.

Because the period data were collected in nearly every year, we can also use these data to construct growth curves for children born in the same year, birth cohorts. Figure 2b contains the same data points as Figure 2a, but shows that we can connect these points differently to reflect the experience of a birth cohort. The cohort growth curves are less complete than the period growth curves, but they capture the experience of children better since they reflect the growth of the same group of children over time. The cohort growth curves will be the focus of our analysis below.

For 1950 and earlier, the averages in our data are of all children enrolled in school, whereas from 1955 onward the mean heights were collected via a three-step stratified random sampling method (details in Appendix A.1). As Ali et al. (2000) and Mosk (1996) note, primary school enrolment was extraordinarily high in Japan from the early twentieth century with 94% of boys and 82% of girls of primary school age attending school. Thus, the data collected are representative of the entire population for children in primary school. However, a much smaller share of children continued on to secondary school when enrolment was no longer compulsory. Before 1970, there is clear evidence that children in secondary schools were positively selected on height relative to their

[8] Unfortunately, the underlying individual-level data do not survive for these sources.

primary school counterparts (see Appendix A.2 for full details).[9] Previous researchers have seriously downplayed the potential for selection bias across ages in the data.[10]

In order to overcome the selection bias in the data, we separate the clean, representative data for primary school children from the selected secondary school data. Thus, for most birth cohorts, we input two growth curves into the SITAR growth model, one representative and one selected, and treat the two as independent observations. All of the analysis of the influence of the war on children's growth is drawn from the representative data, but we need to include the selected data so that SITAR can draw the complete mean growth curve. Appendix A.2 explains how we define the representative and selected growth curves and how this changes over time.

4. Methods

4.1. SITAR Growth Model

We use the SITAR (SuperImposition by Translation and Rotation) growth model developed by Cole et al. (2010) to model the growth pattern of each prefecture-birth cohort: each cohort growth curve represented in Figure 2b. SITAR assumes three main characteristics of the growth pattern across individuals: differences in size, i.e. heights at various ages; differences in the age at peak velocity during the pubertal growth spurt; and differences in the speed of maturation, how quickly a child develops (see Figure 1). SITAR fits a mixed effects model to longitudinal growth curves in an attempt to simplify a vast array of individual-specific variation into a single mean growth curve. It is defined by the following equation:

$$y_{i,t,a} = \alpha_{i,t} + h\left(\frac{a-\beta_{i,t}}{exp(-\gamma_{i,t})}\right) \tag{1}$$

where $y_{i,t,a}$ is the mean height of children in prefecture i born in year t at age a, $h(a)$ is a natural cubic spline of height by age, and $\alpha_{i,t}$, $\beta_{i,t}$ and $\gamma_{i,t}$ are three random effects that

[9] See Schneider (2020) for a discussion of sample-selection biases in sources of historical children's growth and how sample-selection bias would influence inferences about the growth pattern.

[10] Mosk (1996, pp. 27-31) notes that selection bias may be present in the data but does not try to correct for it.

adjust the mean spline curve in three ways. $\alpha_{i,t}$ shifts the spline upward or downward, capturing differences in size or height between cohorts. $\beta_{i,t}$ shifts the spline left or right changing the timing of the pubertal growth spurt. Finally, $\gamma_{i,t}$ stretches or shrinks the age scale, which has the effect of capturing differences in the speed of maturation between cohorts. Figure 3 provides a graphical representation of how these parameters shift the growth curve. We will refer to $\alpha_{i,t}$, $\beta_{i,t}$ and $\gamma_{i,t}$ as the size, timing and intensity SITAR parameters respectively. SITAR estimates these parameters as random effects for each prefecture-birth cohort growth curve, and the parameters explain the difference between each prefecture-birth cohort growth curve and the mean growth curve predicted from within the model. Thus, we can use the parameters to track changes in children's growth pattern over time (Cole and Mori 2017; Schneider and Ogasawara 2018).

4.2. Identification Strategy

Our dataset is interesting because it presents a method for calculating the counterfactual influence of the Second World War on children's growth. This is because we observe some birth cohorts only before the war and others only after the war. For the 1933 cohort and earlier all of their height measurements were taken in 1939 or earlier, whereas for the 1934 cohort and later, all height measurements were taken in 1948 or later (Figure 2b). This allows us to counterfactually compare the growth pattern that the 1933 birth cohort would have achieved in the absence of the war to the actual growth pattern achieved by the 1934 birth cohort.

Figure 4 demonstrates the counterfactual graphically. The x-axis shows the data available for each birth cohort, and the y-axis shows the years that children in each birth cohort were measured. In this graph, horizontal lines by year of measurement reflect period growth curves (Figure 2a) and vertical lines by birth cohort reflect cohort growth curves (Figure 2b). We can then divide the cohorts into five groups. Group A consists of cohorts born before the war whose growth was unaffected by the war because they reached the end of their growing period (assumed to be age 20) before the war. Cohorts in group B were affected by the war, but the only growth observations for these cohorts come from before the war. Thus, we use the SITAR growth model to predict the counterfactual growth pattern that these cohorts would have experienced in the absence of the war. Cohorts in groups C and D were affected by the war, but we only observe growth for these cohorts after the war. The difference between groups C and D is that for group C, we only observe these cohorts' growth in the immediate aftermath of the war between 1948 and 1950, limiting the recovery possible for children in these cohorts. We observe group D on the other hand in the late 1950s and are able to capture catch-up growth that occurred for children as their health improved in the post-war period. Finally, group E were cohorts born after the war that were unaffected by the war. We draw the cutting point between cohorts D and E between 1948 and 1949 because children born in 1948 would have been *in utero* during the final year of the war shock in 1947.

The cleanest identification of the effect of the war comes from comparing cohorts around the 1933-34 split between groups B and C. This difference shows the effect of the war on children who were exposed in late childhood and adolescence since those are the ages at which group C was exposed to the war: see Figure 5 for a lexis diagram that shows at what ages the war affected each cohort. To increase the sample size and our confidence in the counterfactual, we analyse the birth cohorts 1931-33 for group B and 1934-36 for group C. There is little reason to believe that there are significant differences between these cohorts on pre-treatment (pre-war) characteristics: all cohorts were born before the war began in 1937, and there were not large differences in health conditions across these years. To prove this, in Section 5.2, we estimate inverse probability weighted average treatment effects balancing on pre-treatment cohort health characteristics to ensure that these treatment and control groups are truly comparable. To understand how the war affected young children and whether it was possible for these children to recover afterwards, we also make comparisons between groups C, D and E. However, here the counterfactual is potentially clouded by differential selection into fertility or other selection mechanisms by cohort. These issues are discussed briefly in Section 6 and in full detail in Appendix E, and while we cannot rule out all forms of selection, we do not think the selection would be large enough to explain away our main results.

5. Results

5.1. SITAR Results

We estimate the SITAR model using the SITAR package in R (Cole 2019; R Core Team 2019). A full discussion of the results is provided in Appendix D. In short, we analyse over 6,000 prefecture-birth cohorts for each sex, and the SITAR models provide a very good fit for our data. The residual standard deviations are low at 0.45 and 0.44 cm and the models explain 99.1% and 99.0% of the variance in height in our sample for boys and girls respectively. The standard deviation of the residuals is low relative to other samples analysed with SITAR (Cole et al. 2010), but this is in part because we are working with mean growth curves, which mitigates measurement error in our sample. The excellent fit of the models gives us confidence that we can use the predicted parameters to understand changes in the growth pattern over time and across the Second World War.

5.2. War Shock Results

Figure 6 presents the results from the SITAR estimations with boys in the left column and girls in the right column and each row representing one of the three SITAR parameters: size, timing and intensity. The solid line in each graph presents the mean and the shaded areas are 95% confidence intervals of the prefecture-cohort SITAR parameters both weighted by prefecture population in each year. The graphs show the long run changes in size, timing and intensity across the twentieth century. Size has increased as Japanese men and women are much taller than they were in the early twentieth century. Timing has decreased as the age at the peak velocity of the pubertal growth spurt has declined over time. Finally, intensity has increased over time. The mean intensity for the entire sample is zero, so the increase from negative to positive values suggests that children developed more quickly, growing at higher velocities and finishing their growing period at younger ages. These findings confirm the long run changes in the growth pattern of Japanese children discussed by Ali et al. (2000) using data free from selection bias and also concur with findings from the post-war period by

Cole and Mori (2017). These long run changes in the indicators provide context for the magnitude of the shock of the Second World War.

To understand the effects of the war on children exposed in late childhood and adolescence, we compare the counterfactual growth pattern of children in cohort B with the actual growth pattern of children treated by the war in cohort C (see Figure 4). To ensure that our war-treated cohorts (1934-36) and war-untreated cohorts (1931-33) are comparable, we estimate average treatment effects for the war shock using inverse probability weights to balance the treatment and control cohorts on pre-treatment health characteristics, i.e. health conditions in the year of birth for each cohort. We cluster our standard errors at the prefecture level to account for potential correlation of errors within prefectures. We balance the groups on eight health characteristics that would be most important for influencing the scarring or selective culling of the cohorts. We include the fetal death rate and infant mortality rate to capture selective culling since there is some evidence of selective culling through these mechanisms for past populations (Bozzoli et al. 2009; Bruckner and Catalano 2018; Bruckner et al. 2019; Schneider 2017b). The diarrhoeal death rate and water taps per 100 people control for the sanitary environment, which can be very important in causing child stunting (Lin et al. 2013; Spears et al. 2013). Rice, soy and milk output per capita capture the local nutritional environment (Headey et al. 2018), although Japanese markets were fairly well integrated in this period so we would not necessarily expect major food shortages before the war (Hunter and Ogasawara 2019; Ito et al. 2016). Finally, midwives per 1,000 births proxies the relative level of medical care in each prefecture and year, which again can dramatically influence birth outcomes, infant feeding practices and perinatal mortality (Saito 2008; Woods et al. 2006).

Table 3 presents the results. When balancing the war-treated and war-untreated cohorts on pre-treatment health characteristics, we find large and highly statistically significant shocks of the war. Japanese boys and girls treated by the war were 3.0 and 1.7 cm shorter than they would have been if the war had not occurred. The war also influenced the other characteristics of the growth pattern. Boys and girls had their pubertal growth spurt 0.51 and 0.46 years later respectively than they would have if the war had never occurred. They also experienced slower maturation as a result.[11] These counterfactual changes in the growth pattern are very large when compared against the secular changes in the growth pattern across the twentieth century (Figure 6): they tend to be a quarter to a third of the magnitude of the total change across the twentieth century. Again as Figure 5 shows, these large shifts in the growth pattern occurred in cohorts that were exposed to the war in late childhood and early adolescence well outside the critical first thousand days. Thus, health shocks to children in late childhood and adolescence had a very strong influence on the growth pattern, contrary to the typical thousand days consensus.

In addition, children in cohort D performed relatively well, despite being exposed to the war during their first thousand days (Figure 5). They did well in three respects. First, they were taller, had earlier growth spurts and matured more quickly than children in cohort C who experienced the war in late childhood and adolescence. Second, cohort D children were also often better off than counterfactual cohorts (B) reflecting how children would have grown in the absence of the war. Finally, the fact that there are no discontinuities in the trend across cohorts D and E suggests that the scarring effects of the war were relatively limited for children in cohort D. If the war scarring were persistent, we would expect to see a sudden change in the growth pattern between cohorts D and E, i.e. comparing the 1948 and 1949 cohorts, since cohort E was not exposed to the war. Given that the evidence for the war shock to health is overwhelming, the only interpretation for this pattern is that children exposed to health shocks in their first thousand days recovered from these shocks by experiencing catch-up growth.

The role of catch-up growth is also confirmed by the discontinuities in the intensity parameters at the border between cohorts C and D, i.e. comparing the 1940 and 1941 cohorts. There is no reason to expect the 1941 cohort to do so much better than the 1940

[11] Appendix I shows that the greater effect for boys is not driven by stronger selective culling of females across the war. We find excess male mortality across the war similar to the famine demography literature, which shows female advantage in mortality during most historical famines around the world (MacIntyre 2002, p. 243). It is possible that the harsher effects of the war shock for boys could be driven by more stringent physical work requirements for boys (Partner 2007, pp. 146-48; Cook and Cook 1992, pp. 235-38) and/or by a pro-girl shift in the allocation of household resources when men were away fighting or in factories and women had more autonomy over allocation decisions (Partner 2007, p. 150), but we leave definitive analysis of these questions for future work.

cohort when considering their exposure to the war: in fact, the 1941 cohort experienced the war shock at earlier ages. However, because of missing data, we only observe the 1940 cohort and the rest of cohort C immediately following the war (1948-50) whereas the 1941 cohort and the rest of cohort D is observed immediately after the war and again in 1955 and later (see Figure 4). Thus, the sudden increase in intensity for both boys and girls suggests that health conditions improved dramatically during the early 1950s, allowing the 1941 cohort to grow more rapidly than the 1940 cohort as we observe it. The range of health interventions in the 1950s including rapid economic growth, improvements in health infrastructure, the school meal programme and the introduction of more milk and eggs in the diet enabled children who were stunted early in life to recover and achieve better growth outcomes later in childhood and adolescence.

6. Threats to Inference: Selection Bias and Measurement Error

6.1. Scarring and Selection

As in all studies that analyse exposure to health shocks by cohort, our estimates could potentially be influenced by differential selection of those exposed and unexposed to the war. This may arise from three mechanisms (Doblhammer et al. 2013). First, there may be differential selection into fertility by cohort since couples giving birth during the war may have been different from those giving birth before or after the war. Second, the war may have led to selective culling of weak children. Thus, the survivors of the war may have been healthier than the average members of their initial cohort, biasing our parameter estimates for any give cohort. Finally, if there were selective culling of weak adults during the war, then the children born after the war shock to survivors of the shock would also have been positively selected relative to earlier cohorts (Gørgens et al. 2012).

Assessing these three possible selection mechanisms is difficult given that we do not have individual-level data that would allow us to analyse selection into fertility across the war, and we also do not observe most key health indicators during the war. However, in Appendix E we discuss these potential threats to inference and how they would influence our results and interpretation. Our main counterfactual comparing the 1931-33 and 1934-36 cohorts is the most robust, especially when balancing on pre-treatment health characteristics as in Table 3.

There is more potential for selection into fertility when we analyse the cohorts born during and after the war (discussed at length in Appendix E.2). The removal of many reproductive age men from society following the start of the Sino-Japanese war in 1937 and their return to Japan after the war, along with other Japanese people living in the colonies, could have influenced the mean latent health of children in different cohorts. There is some evidence that working-class men were more likely to be drafted than middle- or upper-class men (Watanabe 2014), suggesting that perhaps the latent health quality of children born during the war increased. The government also instituted pro-natal policies that counteracted the initial decline in births in the late 1930s (Mathias 1999, pp. 72-74; Partner 2007, pp. 150-51). At the end of the war, births slumped dramatically and there was a post-war baby boom in the late 1940s and early 1950s. Differential selection into fertility is possible across these cohorts. However, without more specific data, it is difficult to understand how important the change in the composition of births could be to the patterns that we find. The absence of sharp discontinuities in the predicted parameters year-to-year suggests that either selection into fertility did not matter or the selection effects were small relative to the war shock and subsequent catch-up.

In addition, the pattern that we observe across cohorts in Figure 6 does not match the expected selective culling or thousand-days scarring pattern from a health shock. If the thousand days scarring pattern were true in this case, we would not expect to see detrimental effects of the war until reaching cohorts who experienced the war during their first thousand days (Figure 7). Thus, we would not expect to see a difference between the 1933 and 1934 birth cohorts, and the worst-affected cohorts would be the 1944-1946 cohorts who experienced all of their first thousand days during the worst parts of the war. From the 1947 cohort onwards, we would expect to see improvements in height since the children would have only experienced part of their first thousand days in war conditions. Finally, we would expect to see a large gap between the last war-affected cohorts and the cohorts born after the war since they were not exposed to wartime conditions. This is clearly not the pattern observed in the data.

The data do not match a strong selective culling pattern either (Figure 7). Under a strong selective culling effect, we would also not expect to see a large difference between the 1933 and 1934 birth cohorts. The 1934 birth cohort would be slightly taller than the 1933 birth cohort because of the additional children dying during the war, but the effect would be small since the 1934 cohort experienced the war between ages 7 and 14, ages at which mortality rates are very low relative to infancy and early childhood. The positive selective culling effect would intensify for cohorts experiencing infancy and early childhood during the war since mortality rates are higher at these ages, driving mean height upward. Toward the end of the war, mean heights would fall somewhat again as the cohorts experienced only a few years with higher mortality.

Thus, although we cannot be as precise in estimating a causal effect of exposure to the war by comparing cohorts born during the war with those born before or after the war, the change in the SITAR parameters suggests that cohorts experiencing the war in late childhood and adolescence were more strongly affected by the war shock than cohorts born during the war.[12] The relatively small differences between those experiencing their first thousand days during the worst conditions of the war and those born after the war suggest that there was substantial scope for catch-up growth when conditions improved after the war.

6.2. Measurement Error from Post-war Repatriation and Internal Migration

In addition to the selection issues highlighted above, there are two sources of measurement error that could potentially threaten our inferences. The first is that following the war, 3.2 million Japanese civilians who had been living in Japan's colonial empire were repatriated to Japan (Watt 2009, p. 2). Some of these repatriates, though

[12] We also discuss the potential for positive selection of adults on health during the war in Appendix E.3 to account for Doblhammer et al. (2013)'s third selection mechanism, but this is unlikely to matter for our results.

we do not have precise figures, were children who would have joined Japan's school population after returning and are therefore present in the mean prefecture heights after the war. If these repatriates were systematically taller or shorter than Japanese children, then their inclusion in our post-war data will bias our results. Appendix G discusses the repatriates in detail. It is difficult to know whether repatriate children would have been taller or shorter than native Japanese. Those repatriated from active war zones of the Pacific War likely experienced very poor health conditions, but even children who were relatively removed from the war, such as those in Manchuria, could experience horrific conditions after the Japanese surrender (Itoh 2015, pp. 43-44; Watt 2009, pp. 102-104). However, we do not believe the repatriates are an important source of bias because they were likely a very small share of total children measured in Japan, perhaps around 2.8% of children.

Another threat to inference could arise from migration of individuals between prefectures over time. This could lead to biases in the cohort growth curves if migrants were systematically different from non-migrants. Appendix H deals with these concerns in considerable detail. Internal migration across the entire twentieth century was mainly characterised by rural to urban migration. Internal migration was at fairly low levels before the war, but it accelerated considerably in the post-war boom period with out-migration rates from many rural prefectures exceeding 1% per year and inmigration to Tokyo and Osaka and their surrounding prefectures and Aichi prefecture reaching 2% per year (Fukao et al. 2015, pp. 113-15). However, most migrants were in their late teens or early twenties (see Figure A7), limiting the effect of migration on the children in our data who were mostly under the age of 14. In addition, plausible back-of-the-envelope estimates of the potential bias created by this migration show that the potential for error is very small (see Figure A9). Thus, it is very unlikely that internal migration is substantially biasing our results.

7. Conclusion

In summary, we find that the deterioration in nutritional and health conditions in Japan during the Second World War led to substantial unhealthy changes in the growth pattern of Japanese children. Cohorts treated by the war in late childhood and adolescence were substantially shorter, had later pubertal growth spurts, and matured more slowly than counterfactual, war-untreated cohorts. The magnitude of the shock was a quarter to a third of the change in each parameter of the growth pattern over the twentieth century. We also show that our estimates are unlikely to be affected by differential selection into each cohort.

In addition, we find a peculiar age pattern of the effect of the war on the growth pattern. Children exposed to the war shock in late childhood and adolescence experienced much greater change in their growth pattern than children who were exposed to the war in early life. This age pattern does not accord with the thousand days consensus in development economics, nutrition and human biology in two respects (Alderman et al. 2006; Hoddinott and Kinsey 2001; Victora et al. 2010; Wells 2017). First, it suggests that shocks in late childhood and adolescence can have severe consequences for child growth, suggesting that late childhood and adolescence may be another critical window for growth (Akresh et al. 2021; Depauw and Oxley 2019; Prentice et al. 2013;
Schneider and Ogasawara 2018; van den Berg et al. 2014). This evidence suggests that policy makers should not only target scarce resources at young children during economic or health shocks because adolescents may need to be protected as well.

Second, the fact that cohorts born during the war often experienced healthier outcomes than the counterfactual, war-unaffected cohorts suggests that considerable catch-up growth was possible for these cohorts when conditions improved after the war. There is debate in the literature about whether children can achieve lasting catch-up growth after poor conditions in early life (Golden 1994; Hoddinott and Kinsey 2001; Prentice et al. 2013; Proos and Gustafsson 2012). Much of this debate centres around measuring catch-up growth relative to the WHO growth standard/reference. Lundeen et al. (2014) and Leroy et al. (2015) argue that children who catch up relative to the

WHO standard maintain the same height gap from the WHO mean, suggesting that the catch-up growth is simply a product of the increasing variance in height as children age. However, the SITAR parameters predicted in our study do not rely upon any growth reference, and therefore do not suffer from these potential methodological flaws. Thus, there is no doubt that the cohorts born during the war experienced substantial catch-up growth in the post-war period.

These findings relate to the “double shocks” literature in development economics, which finds that children experiencing poor health in early life can experience catchup in human capital following an intervention (Adhvaryu et al. 2018; Akresh et al. 2021; Almond et al. 2018). Perhaps it is unsurprising that children experienced catchup growth in post-war Japan. In the 1950s, real GDP per capita grew at 8.2% per year (Cabinet Office 2020); life expectancy increased by 7.3 and 8.7 years for men and women respectively, mainly driven by declines in mortality from communicable disease (Ikeda et al. 2011); the school lunch programme was expanded to feed all primary school children and many secondary school children as well; and the Japanese diet diversified with greater consumption of protein and nutrient-rich foods like milk and eggs (Takahashi 1984). This was an extraordinary intervention covering multiple dimensions of health and going far beyond what is possible in the typical experimental settings for current low and middle income countries. However, our results clearly demonstrate that catch-up growth is possible for children who suffered very poor health conditions in early life. While our results do not prove that any specific intervention was most effective, they highlight the imperative for renewed research on the effectiveness of health interventions following early life deprivation. These efforts should not replace the emphasis in the Sustainable Development Goals on preventing stunting and targeting children in their first thousand days. Instead, they should encourage researchers and policy makers to find effective ways to reduce the burden of stunting on the 149 million children already stunted in the world today (UNICEF et al. 2019).

References

Adhvaryu, A., Nyshadham, A., Molina, T., and Tamayo, J. (2018). Helping Children Catch Up: Early Life Shocks and the PROGRESA Experiment. *National Bureau of Economic Research Working Paper Series*.

Akresh, R., Bhalotra, S., Leone, M., and Osili, U. (2021). First and Second Generation Impacts of the Biafran War. *Journal of Human Resources*.

Akresh, R., Bhalotra, S., Leone, M., and Osili, U. O. (2012a). War and Stature: Growing Up during the Nigerian Civil War. *American Economic Review Papers and Proceedings*, 102(3):273–277.

Akresh, R., Lucchetti, L., and Thirumurthy, H. (2012b). Wars and child health: Evidence from the Eritrean–Ethiopian conflict. *Journal of Development Economics*, 99(2):330–340.

Akresh, R., Verwimp, P., and Bundervoet, T. (2011). Civil war, crop failure, and child stunting in Rwanda. *Economic Development and Cultural Change*, 59(4):777–810.

Alderman, H., Hoddinott, J., and Kinsey, B. (2006). Long term consequences of early childhood malnutrition. *Oxford Economic Papers*, 58(3):450–474.

Aldous, C. (2011). Contesting Famine: Hunger and Nutrition in Occupied Japan, 1945-1952. *Journal of American-East Asian Relations*, 17(3):230–256.

Ali, M. A., Lestrel, P. E., and Ohtsuki, F. (2000). Secular trends for takeoff and maximum adolescent growth for eight decades of Japanese cohort data. *American Journal of Human Biology*, 12(5):702–712.

Almond, D. and Currie, J. (2011). Killing Me Softly: The Fetal Origins Hypothesis. *The Journal of Economic Perspectives*, 25(3):153–172.

Almond, D., Currie, J., and Duque, V. (2018). Childhood Circumstances and Adult Outcomes: Act II. *Journal of Economic Literature*, 56(4):1360–1446.

Angell-Andersen, E., Tretli, S., Bjerknes, R., Fors´en, T., Sørensen, T. I. A., Eriksson, J. G., R¨as¨anen, L., and Grotmol, T. (2004). The association between nutritional conditions during World War II and childhood anthropometric variables in the Nordic countries. *Annals of Human Biology*, 31(3):342–355.

Arthi, V. (2018). "The Dust Was Long in Settling": Human Capital and the Lasting Impact of the American Dust Bowl. *Journal of Economic History*, 78(1):196–230.

Bassino, J.-P. and 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*.

Black, R. E., Victora, C. G., Walker, S. P., Bhutta, Z. A., Christian, P., de Onis MD, M., Ezzadi, M., Grantham-McGregor, S., Katz, J., Martorell, R., Uauy, R., and the Maternal and Child Nutrition Study Group (2013). Maternal and child undernutrition and overweight in low-income and middle-income countries. *The Lancet*, 382(9890):427–451.

Bozzoli, C., Deaton, A., and Quintana-Domeque, C. (2009). Adult height and childhood disease. *Demography*, 46(4):647–669.

Bruckner, T. A. and Catalano, R. (2018). Selection in utero and population health: Theory and typology of research. *SSM - Population Health*, 5:101–113.

Bruckner, T. A., Singh, P., Mortensen, L. H., and Løkke, A. (2019). The stillbirth sex ratio as a marker of population health among live-born males in Denmark, 1835-1923. *American Journal of Human Biology*, 94(6):e23241–8.

Brundtland, G. H., Liestøl, K., and Walløe, L. (1980). Height, weight and menarcheal age of Oslo schoolchildren during the last 60 years. *Annals of Human Biology*, 7(4):307–322.

Cabinet Office (2020). System of National Accounts (68SNA). Available at https://www5.cao.go.jp/j-j/wp/wp-je12/h10 data01.html. Data downloaded on 22 May 2020.

Cohen, J. B. (1949). *Japan's economy in war and reconstruction*. University of Minnesota Press, Minneapolis.

Cole, T. J. (2003). The secular trend in human physical growth: a biological view. *Economics and Human Biology*, 1(2):161–168.

Cole, T. J. (2019). Super imposition by translation and rotation growth curve analysis. https://github.com/statist7/sitar.

Cole, T. J., Donaldson, M. D. C., and Ben-Shlomo, Y. (2010). SITAR–a useful instrument for growth curve analysis. *International Journal of Epidemiology*, 39(6):1558–1566.

Cole, T. J. and Mori, H. (2017). Fifty years of child height and weight in Japan and South Korea: Contrasting secular trend patterns analyzed by SITAR. *American Journal of Human Biology*, 12:e23054.

Cook, H. T. and Cook, T. F. (1992). *Japan at war: an oral history*. New Press : Distributed by Norton, New York.

Cunha, F. and Heckman, J. (2007). The technology of skill formation. *American Economic Review*, 97(2):31–47.

Currie, J. (2009). Healthy, Wealthy, and Wise: Socioeconomic Status, Poor Health in Childhood, and Human Capital Development. *Journal of Economic Literature*, 47(1):87–122.

Daniele, V. and Ghezzi, R. (2018). The impact of World War II on nutrition and children's health in Italy. *Investigaciones de Historia Econ´omica - Economic History Research*, pages 1–13. de Onis, M., Onyango, A. W., Borghi, E., Siyam, A., Nishida, C., and 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.

Depauw, E. and Oxley, D. (2019). Toddlers, teenagers, and terminal heights: the importance of puberty for male adult stature, Flanders, 1800–76. *The Economic History Review*, 72(3):925–952.

Doblhammer, G., van den Berg, G. J., and Lumey, L. H. (2013). A re-analysis of the longterm effects on life expectancy of the Great Finnish Famine of 1866–68. *Population Studies*, 67(3):309–322.

Ellis, R. W. (1945). Growth and health of Belgian children: During and After the German Occupation (1940-1944). *Archives of disease in childhood*, 20(103):97–109.

FAO (2004). Human energy requirements. *Food and Nutrition Technical Report Series*, pages 1–103.

FAO (2020). New Food Balances. Available at http://www.fao.org/faostat/en/#data/FBS/metadata. Data downloaded on 22 May 2020.

Fukao, K., Bassino, J.-P., Makino, T., Paprzycki, R., Settsu, T., Takashima, M., and Tokui, J. (2015). *Regional Inequality and Industrial Structure in Japan: 1874-2008*. Maruzen Publishing Co., Ltd., Tokyo.

Gao, P. and Schneider, E. B. (2021). The Growth Pattern of British Children, 1850-1975. *Economic History Review*, 74(2):341–371.

Golden, M. H. (1994). Is complete catch-up possible for stunted malnourished children? *European Journal of Clinical Nutrition*, 48 Suppl 1:S58–70– discussion S71.

Gørgens, T., Meng, X., and Vaithianathan, R. (2012). Stunting and selection effects of famine: A case study of the Great Chinese Famine. *Journal of Development Economics*, 97(1):99– 111.

Harris, B. (1995). *The health of the schoolchild : a history of the school medical service in England and Wales*. Open University Press, Buckingham.

Hayami, Y. and Yamada, S. (1970). Kogyoka no shihatsuki niokeru nogyo no seisansei [Agricultural Productivity at the Beginning of Industrialisation]. In Shigeto, K. and Yuzuru, K., editors, *Nihonnogyo to keizaiseicho [Economic Development and Agriculture in Japan]*. The

Nippon Agricultural Research Institute, Tokyo.

Headey, D., Hirvonen, K., and Hoddinott, J. (2018). Animal Sourced Foods and Child Stunting. *American Journal of Agricultural Economics*, 100(5):1302–1319.

Hoddinott, J. and Kinsey, B. (2001). Child growth in the time of drought. *Oxford Bulletin of Economics and Statistics*, 63(4):409–436.

Howe, P. E. and Schiller, M. (1952). Growth responses of the school child to changes in diet and environmental factors. *Journal of Applied Physiology*, 5(2):51–61.

Human Mortality Database (2019). University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). Available at www.mortality.org. Data downloaded on 5 June 2019.

Hunter, J. (2008). Prices, Standards of Living and Material Incentives in Japan, 1937-1941. *Unpublished Manuscript*, pages 1–25.

Hunter, J. and Ogasawara, K. (2019). Price shocks in regional markets: Japan's Great Kant¯o Earthquake of 1923. *Economic History Review*, 72(4):1335–1362.

Ikeda, N., Saito, E., Kondo, N., Inoue, M., Ikeda, S., Satoh, T., Wada, K., Stickley, A., Katanoda, K., Mizoue, T., Noda, M., Iso, H., Fujino, Y., Sobue, T., Tsugane, S., Naghavi, M., Ezzati, M., and Shibuya, K. (2011). What has made the population of Japan healthy? *Lancet*, 378(9796):1094–1105.

Ito, M., Maeda, K., and Noda, A. (2016). Market Integration in the Prewar Japanese Rice Markets.

Itoh, M. (2015). *Japanese War Orphans in Manchuria: Forgotten Victims of World War II*. Palgrave Macmillan, New York.

Japan Statistical Association (1987). *Historical statistics of Japan. Vol.1.* Japan Statistical Association, Tokyo.

Japanese Statistical Association (1982). *History of patients in modern Japan*. Keisoshobyo, Tokyo.

Johnston, B. F. (1953). *Japanese Food Management in World War II*. Stanford University Press, Stanford.

Kagawa, M., Tahara, Y., Moji, K., Nakao, R., Aoyagi, K., and Hills, A. P. (2011). Secular changes in growth among Japanese children over 100 years (1900-2000). *Asia Pacific Journal of Clinical Nutrition*, 20(2):180–189.

Koike, R. (2019). Interpolation of Japan's Household Consumption during World War II. pages 1–59.

Leroy, J. L., Ruel, M., Habicht, J.-P., and Frongillo, E. A. (2015). Using height-for-age differences (HAD) instead of height-for-age z-scores (HAZ) for the meaningful measurement of population-level catch-up in linear growth in children less than 5 years of age. *BMC Pediatrics*, 15(1):1–11.

Lin, A., Arnold, B. F., Afreen, S., Goto, R., Huda, T. M. N., Haque, R., Raqib, R., Unicomb, L., Ahmed, T., Colford, J. M., and Luby, S. P. (2013). Household Environmental Conditions Are Associated with Enteropathy and Impaired Growth in Rural Bangladesh. *The American Journal of Tropical Medicine and Hygiene*, 89(1):130–137.

Lundeen, E. A., Stein, A. D., Adair, L. S., Behrman, J. R., Bhargava, S. K., Dearden, K. A., Gigante, D., Norris, S. A., Richter, L. M., Fall, C. H., Martorell, R., Sachdev, H. S., Victora, C. G., and on behalf of the COHORTS Investigators (2014). Height-for-age z scores increase despite increasing height deficits among children in 5 developing countries. *American Journal of Clinical Nutrition*, 100(3):821–825.

MacIntyre, K. (2002). Famine and the female mortality advantage. In Dyson, T. and O Gr´ada,´ C., editors, *Famine demography: perspectives from the past and present*, pages 240–259. Oxford University Press, Oxford.

Maddison Project, Bolt, J., and van Zanden, J. L. (2020). Maddison style estimates of the evolution of the world economy. A new 2020 update. *Maddison Project Working Paper*, pages 1–44.

Magee, H. E. (1946). Application of nutrition to public health; some lessons of the war. *British Medical Journal*, 1:475–482.

Mathias, R. (1999). Women and the War Economy in Japan. In Pauer, E., editor, *Japan's war economy*, pages 65–84. Routledge, London.

Mosk, C. (1978). Fecundity, infanticide, and food consumption in Japan. *Explorations in Economic History*, 15(3):269–289.

Mosk, C. (1996). *Making health work : human growth in modernJapan*. University of California Press, Berkeley.

Ogasawara, K. (2017). Persistence of pandemic influenza on the development of children: Evidence from industrializing Japan. *Social Science & Medicine*, 181:43–53.

Ogasawara, K. (2018). Persistence of Natural Disasters on Child Health: Evidence from the Great Kanto Earthquake of 1923. *Working Paper*, pages 1–50.

Ohara Institute for Social Research (1964). *Taiheiy¯osens¯oka no r¯od¯oshajy¯otai (The workers in the Pacific War). [in Japanese]*. T¯oy¯okeizaishinp¯osha, Tokyo.

Partner, S. (2007). Daily Life of Civilians in Wartime Japan, 1937-1945. In Lone, S., editor, *Daily Lives of Civilians in Wartime Asia: From the Taiping Rebellion to the Vietnam War*, pages 127–158. Greenwood Press, Westport, Connecticut.

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

Proos, L. and Gustafsson, J. (2012). Is Early Puberty Triggered by Catch-Up Growth Following Undernutrition? *International Journal of Environmental Research and Public Health*, 9(12):1791–1809.

R Core Team (2019). R: A language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Saito, O. (2008). Infant mortality and the Aiiku village schemes in pre-war Japan. *Shakai keizaishigaku*, 73(6):611–633.

Sanderson, F. H. (1987). *Japan's food prospects and policies*. University Microfilms International, Ann Arbor, Mich.

Scherer, A. (1999). Drawbacks to Controls on Food Distribution: Food Shortages, the Black Market and Economic Crime. In Pauer, E., editor, *Japan's war economy*, pages 106–123. Routledge, London.

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.

Schneider, E. B. (2017b). 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.

Schneider, E. B. (2018). Stunting: past, present, future. http://doi.org/10.21953/lse.9teyst78nhxh.

Schneider, E. B. (2020). Sample-Selection Biases and the Historical Growth Pattern of Children. *Social Science History*, 44(3):417–444.

Schneider, E. B. and Ogasawara, K. (2018). Disease and child growth in industrialising Japan: Critical windows and the growth pattern, 1917-39. *Explorations in Economic History*, 69:64–80.

Sen, A. (1977). Starvation and exchange entitlements: a general approach and its application to the great Bengal famine. *Cambridge Journal of Economics*, 1(1):33–59.

Spears, D., Ghosh, A., and Cumming, O. (2013). Open Defecation and Childhood Stunting in India: An Ecological Analysis of New Data from 112 Districts. *PLoS ONE*, 8(9):e73784–9.

Steckel, R. H. R. (1987). Growth depression and recovery: the remarkable case of American slaves. *Annals of Human Biology*, 14(2):111–132.

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

UN Inter-agency Group for Child Mortality Estimation (2020). Child Mortality Estimates. Available at http://childmortality.org. Data downloaded on 26 January 2021.

UNICEF, WHO, and World Bank Group (2019). Joint child malnutrition estimates - Levels and trends in child malnutrition: Key findings of the 2019 edition. Technical report.

United States Strategic Bombing Survey (1947). *The Japanese wartime standard of living and utilization of manpower.* Number xi, 146 p. U.S. Government Printing Office, [Washington].

van den Berg, G. J., Lundborg, P., Nystedt, P., and Rooth, D.-O. (2014). Critical Periods During Childhood and Adolescence. *Journal of the European Economic Association*, 12(6):1521–1557.

Victora, C. G., de Onis, M., Hallal, P. C., Blossner, M., and Shrimpton, R. (2010). Worldwide Timing of Growth Faltering: Revisiting Implications for Interventions. *Pediatrics*, 125(3):e473–e480.

Watanabe, T. (2014). Who became the soldiers? Part 2. *Working Paper*.

Watt, L. (2009). *When Empire Comes Home: Repatriation and Reintegration in Postwar Japan*. Harvard University Press, Cambridge, MA.

Wells, J. C. K. (2017). Worldwide variability in growth and its association with health: Incorporating body composition, developmental plasticity, and intergenerational effects. *American Journal of Human Biology*, 29(2):e22954–16.

Woods, R. I., Lokke, A., and van Poppel, F. (2006). Two hundred years of evidence-based perinatal care: late-fetal mortality in the past. *Archives of Disease in Childhood - Fetal and Neonatal Edition*, 91(6):F445–F447.

Yamashita, S. H. (2013). The 'Food Problem' of Evacuated Children in Wartime Japan, 1944-45. In Cwiertka, K. J., editor, *Food and War in Mid-Twentieth-Century East Asia*. Routledge, London.

Yokoya, M. and Higuchi, Y. (2016). Association between summer temperature and body weight in Japanese adolescents and children: An ecological analysis. *American Journal of Human Biology*, 28(6):789–795.

Yoshida, K. (1995). *Poverty in Japan* . Keisoshobyo, Tokyo.

Table 1. Calorie and protein intakes per capita per day between 1941 and 1945

Year	Calories (kcal)	Protein (grams)
1941	2105	64.7
1942	1971	60.2
1943	1961	60.6
1944	1927	61.2
1945	1793	65.3

Sources: Ohara Institute for Social Research (1964, Table 78 of Chapter 2, Section 5). The original figures are from the survey report by the Headquarters for Economic Stabilization.

Table 2. Calorie and protein intakes per capita per day in cities in July 1945

		[A] Intake from rationing		[B] Total intake		Non-ration share (B-A)/B	
City	Population 1940	Calories (kcal)	Protein (grams)	Calories (kcal)	Protein (grams)	% Calories	% Protein
Tokyo	6,778,804	1437	65	1798	82	20	21
Osaka	3,252,340	1277	51	1824	79	30	36
Kyoto	1,089,726	1413	51	1677	74	16	31
Maebashi	86,997	1221	45	1716	73	29	38
Morioka	79,478	1380	31	1745	53	21	42
Yamaguchi	34,579	1298	45	2026	88	36	49

Notes: Non-ration share could be purchases on the black market or home production in gardens. Home production would be a smaller component of the non-ration share for the larger cities.

Sources: Ohara Institute for Social Research (1964, Tables 79 and 81 of Chapter 2, Section 5). The original figures are from the survey report by the Ministry of Health and Welfare and from the survey report by the Osaka City.

Table 3. Inverse Probability Weighted Average Treatment Effects for the War Shock: 1931-33 vs. 1934-36

	Male			Female		
	Size	Timing	Intensity	Size	Timing	Intensity

	(cm)	(years)	(fractional)	(cm)	(years)	(fractional)
Potential Outcome Mean War-Untreated (Control)	164.0**	13.57**		152.4**	11.68**	
	(0.11)	(0.016)	(ref)	(0.10)	(0.021)	(ref)
Average Treatment Effect War-Treated	-3.0**	0.51**	-0.059**	-1.7**	0.46**	-0.043**
	(0.14)	(0.019)	(0.002)	(0.11)	(0.024)	(0.003)
N	276	276	276	276	276	276

Notes: Standard errors in parentheses clustered at the prefecture level. ** denotes statistical significance at the 1% level. The sample size of 276 reflects the six cohorts (1931-36) across 46 prefectures that we can observe before and after the war. We have not reported the potential outcome mean for intensity because intensity is a relative measure that is not meaningful on its own like the other two parameter means. The average treatment effects were estimated using inverse probability weights that balance the treatment (exposure to the war) with the control on pre-treatment characteristics. We included eight variables in the treatment model to capture the health environment for each prefecture-cohort: infant mortality rate, fetal death rate, diarrhoeal death rate, water taps per 100 people, rice output per capita, soy output per capita, milk output per capita, and midwives per 1,000 births. We also included the square terms for all variables except rice output to improve covariate balance. We compared standardized differences and variance ratios between the raw and weighted samples, and the mean differences were very close to zero and variance ratios close to one indicating covariate balance.

Sources: Japanese Prefecture Child Growth Dataset - see Appendix A.1 for sources.

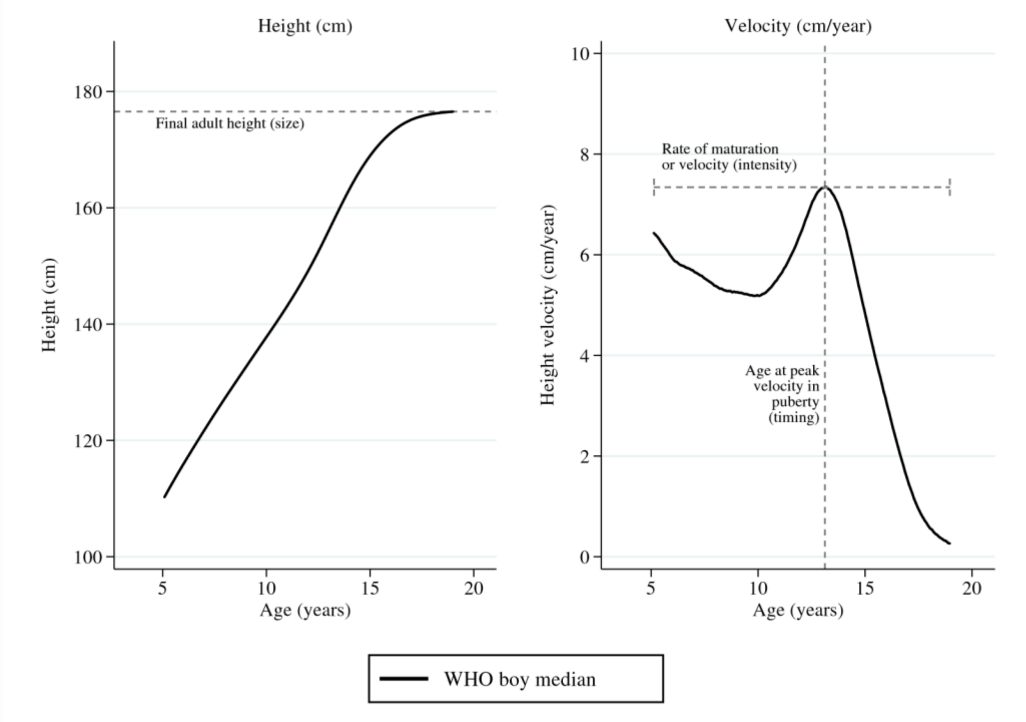


Figure 1. The growth pattern of boys illustrated using the WHO 2007 growth reference

Notes and Sources: WHO boy median is the median growth curve from the 2007 WHO schoolage child and adolescent growth reference (de Onis et al. 2007). Data from https://www.who.int/growthref/who2007_height_for_age/en/.

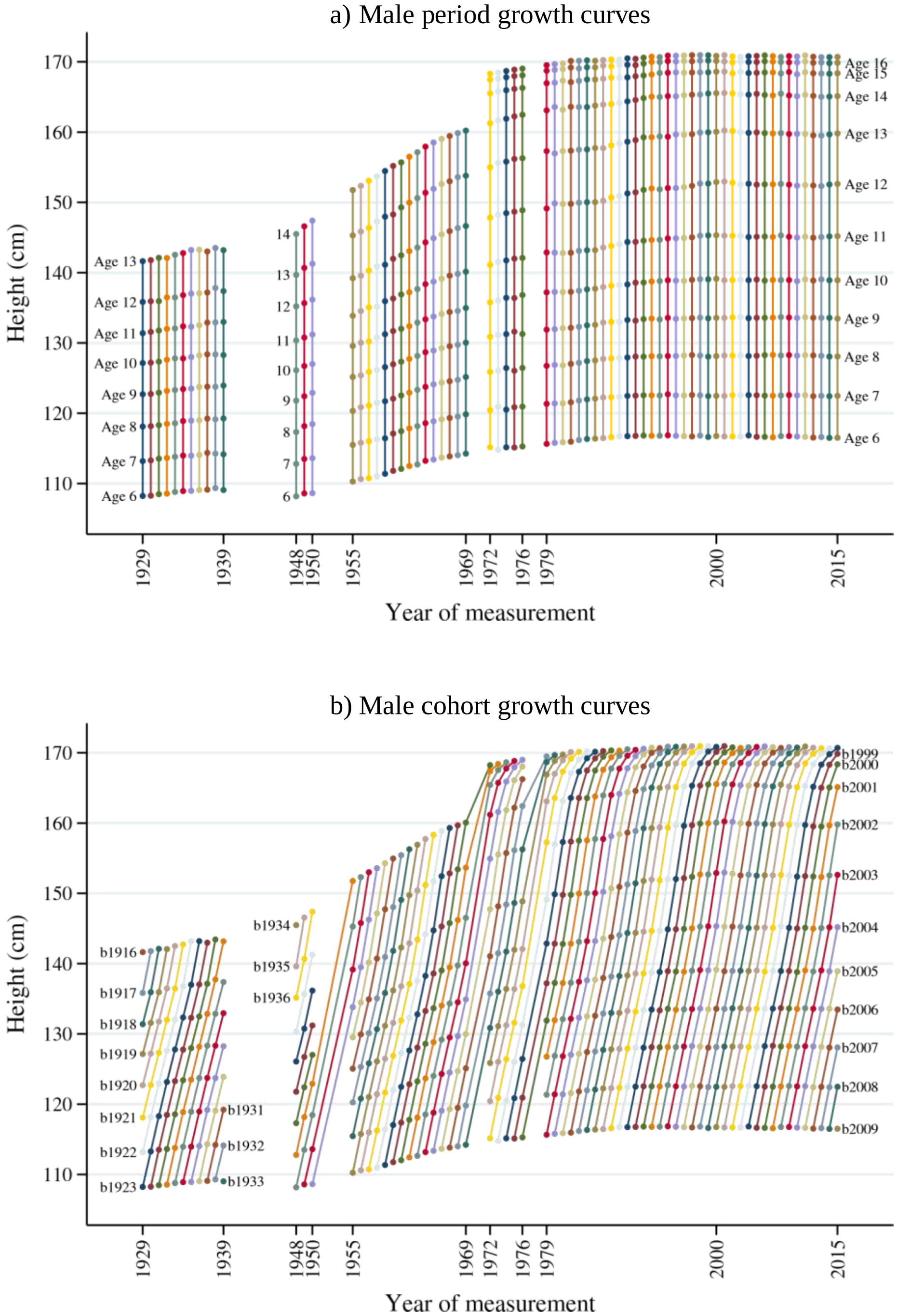


Figure 2. Graphical representation of period and cohort growth curves for boys

Notes: Each point reflects the population weighted mean height across all prefectures at each age in each year. Period growth curves are for children measured in the same year. Cohort growth curves relate to children born in the same year but measured in successive years as the cohort ages.

Sources: Japanese Prefecture Child Growth Dataset - see Appendix A.1 for sources.

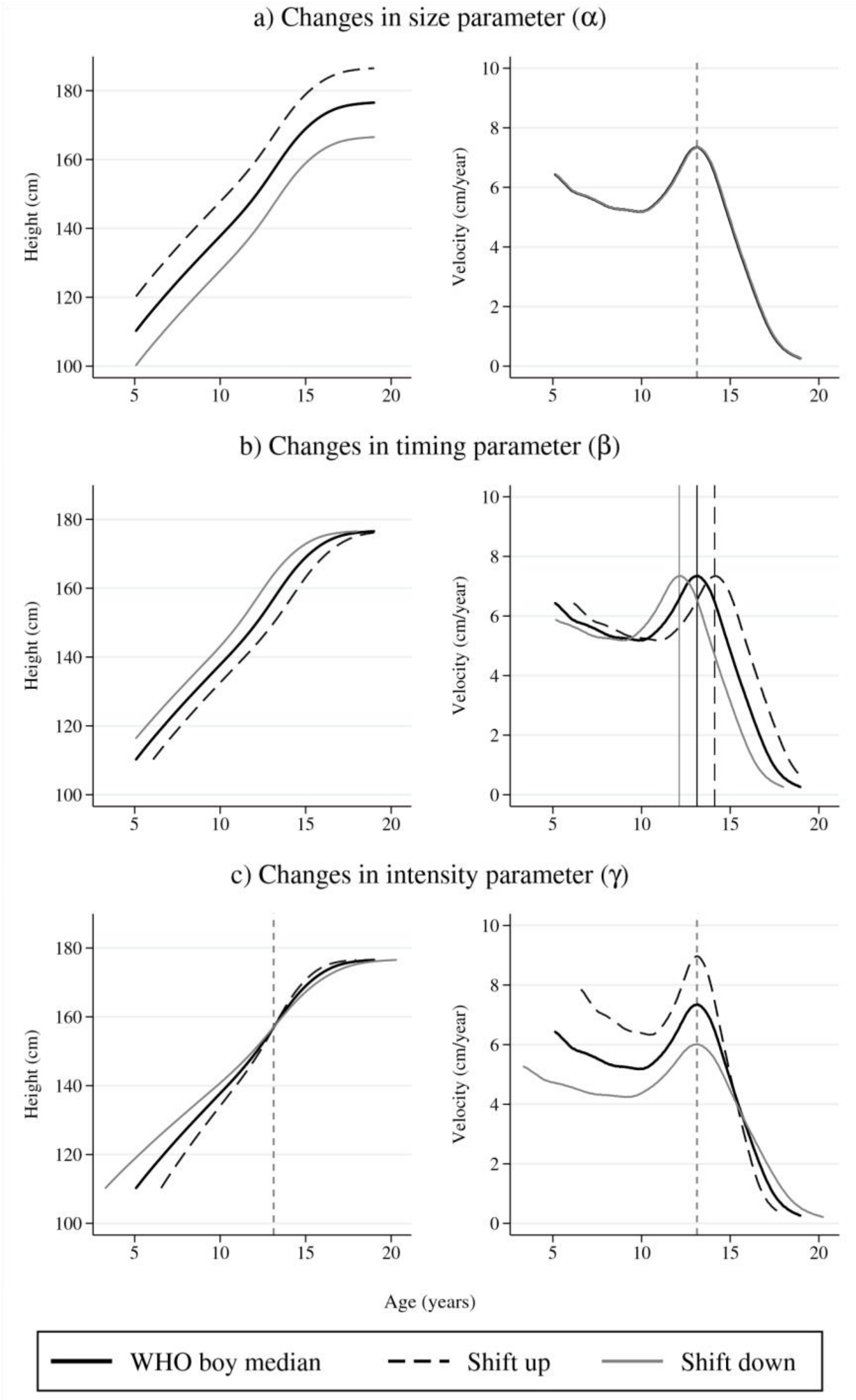


Figure 3. Demonstration of how each SITAR parameter adjusts the height and velocity curves

Notes and Sources: WHO Boy Median is the median growth curve from the 2007 WHO schoolage child and adolescent growth reference (de Onis et al. 2007). Data from https://www.who.int/growthref/who2007_height_for_age/en/. The median curve was adjusted for each parameter following Equation 1.

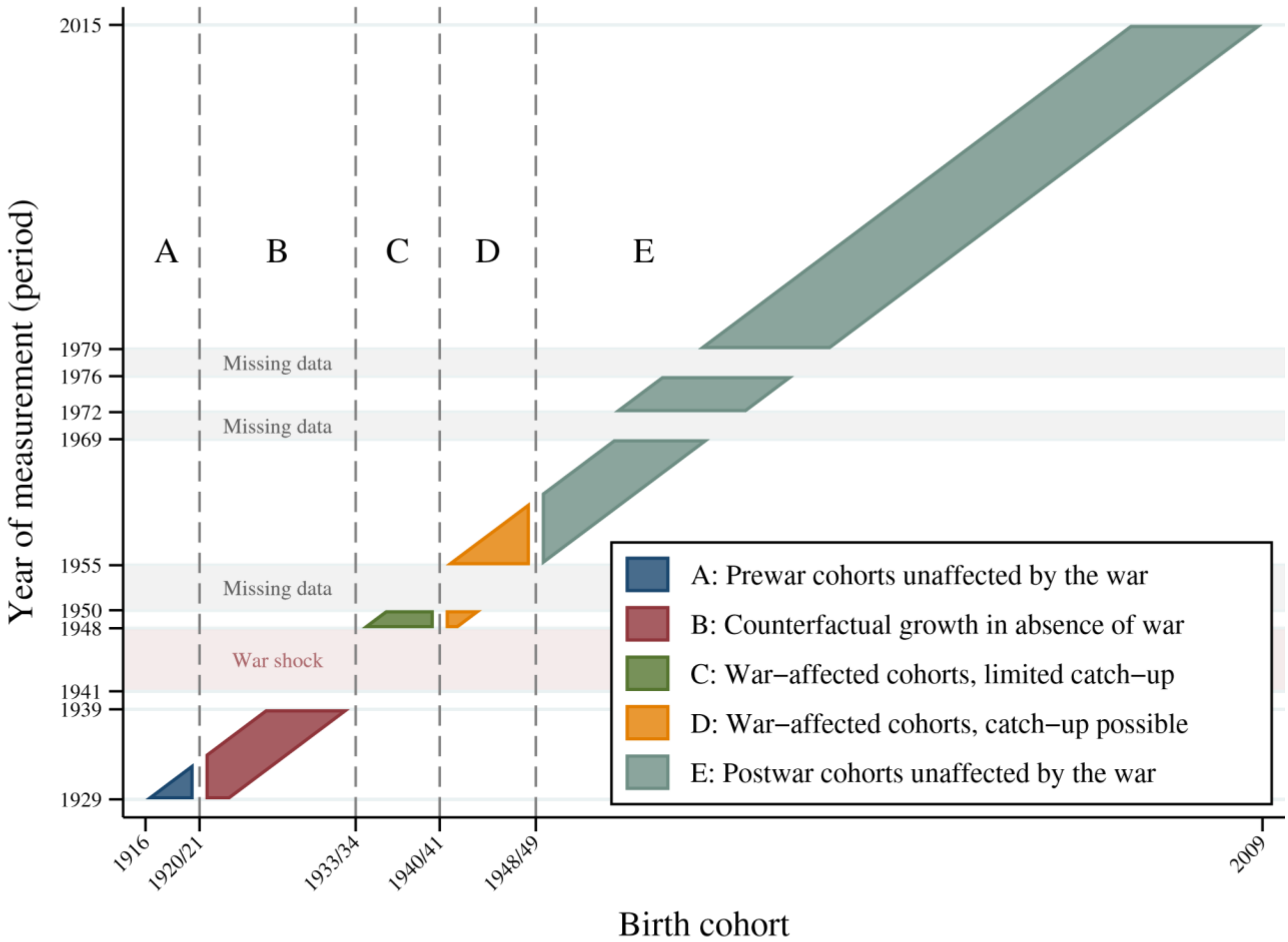


Figure 4. Graphical representation of data used to estimate growth curves

Notes: The x-axis reflects the birth year and the y-axis reflects the year in which a birth cohort was measured in our data. Thus, the height of the shaded areas is the number of age measurements available for each cohort. The width of the shaded areas changes when children across a wider range of ages are included in the measurements, e.g. when secondary school data becomes representative of children up to age 17 beginning in 1972. Dashed lines separate different sets of cohorts.

Sources: Japanese Prefecture Child Growth Dataset - see Appendix A.1 for sources.

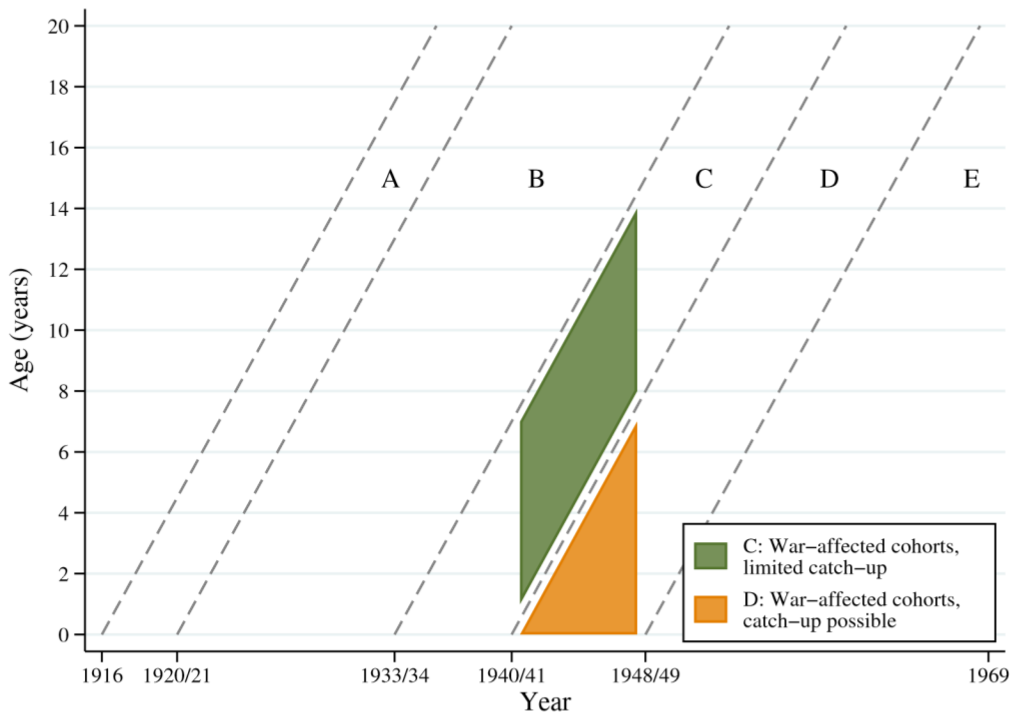


Figure 5. Lexis diagram showing the ages at which certain cohorts were exposed to the war

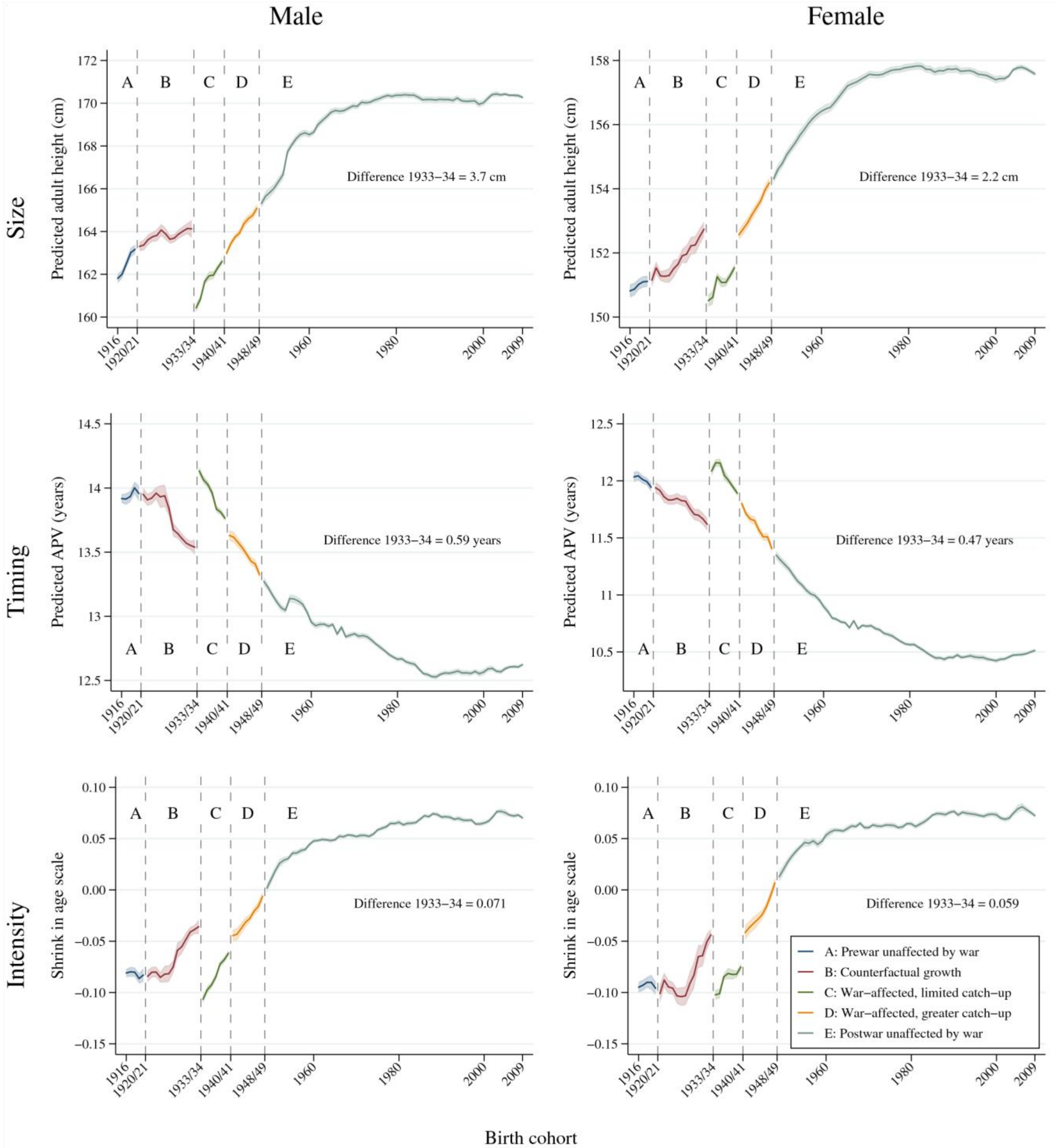


Figure 6. Mean SITAR parameters for males (left) and females (right)

Notes: The lines and shaded confidence intervals show the mean and 95% confidence interval of each SITAR parameter in each year across the prefectures. Both means and confidence intervals are weighted by prefecture population size in each year. APV is the age at peak velocity during the pubertal growth spurt.

Sources: Japanese Prefecture Child Growth Dataset - see Appendix A.1 for sources.

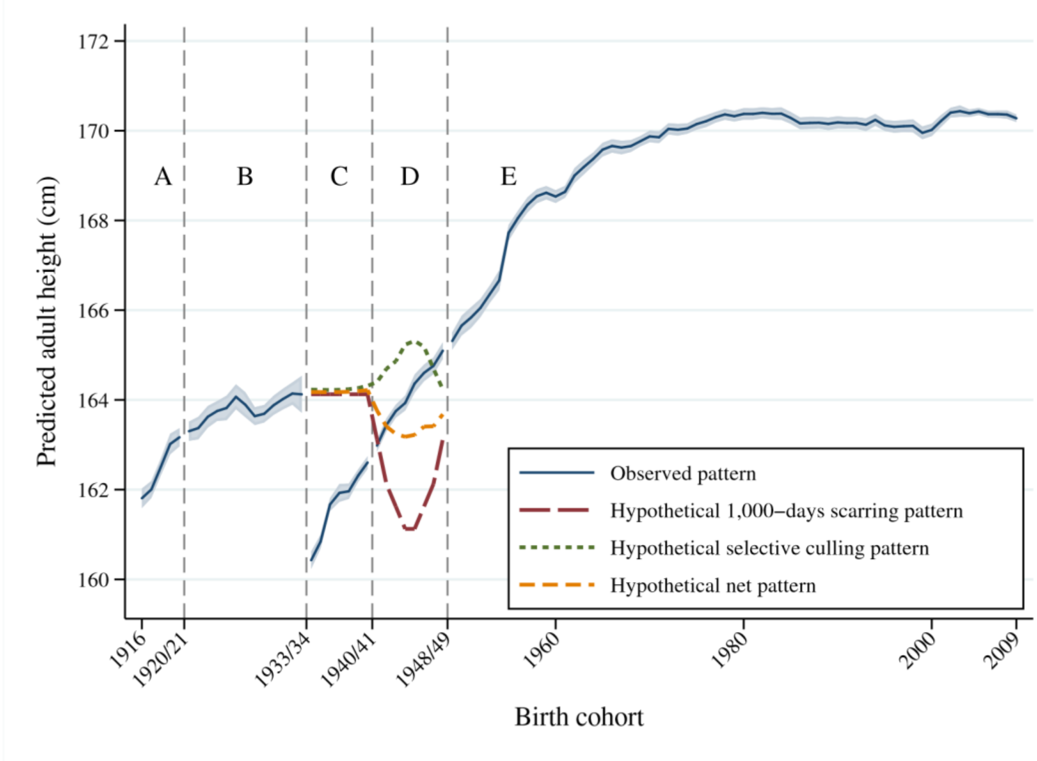


Figure 7. Patterns expected with pure scarring or culling effects

Notes: See Appendix F for full details about how the hypothetical scarring, selective culling and net patterns were generated.

Sources: Japanese Prefecture Child Growth Dataset - see Appendix A.1 for sources; Human Mortality Database (2019).

Appendix: Supplemental Materials

Health Shocks, Recovery and the First Thousand Days: The Effect of the Second World War on the Growth Pattern of Height in Japanese Children

Population and Development Review

Eric B. Schneider,[a] Kota Ogasawara,[b] and Tim J. Cole[c]

[a]London School of Economics and Political Science, e.b.schneider@lse.ac.uk

[b]Tokyo Institute of Technology, ogasawara.k.ab@m.titech.ac.jp [c]UCL Great Ormond Street Institute of Child Health, tim.cole@ucl.ac.uk

ARTICLE HISTORY

Compiled June 29, 2021

A. Data Appendix

A.1. Historical Sources of Height Data

For the pre-war period, 1929-39, the panel dataset of mean heights of primary school children for the 47 prefectures was constructed from the reports entitled 'Statistics of School Physical Examination' (SPPE) published by the Physical Bureau, Ministry of Education between 1931 and 1943. The SSPE consisted of two types of reports: the *K¯oshiritsu sh¯ogakk¯o chu¯gakk¯o k¯ot¯ojyogakk¯o seitojid¯o sintaikensat¯okei* (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 *Gakk¯o shintaikensat¯okei* (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). The SPPE sample in this period included all children in school in Japan, which was approximately 95% of children aged 6-11. More details on these sources is provided in Schneider and Ogasawara (2018).

For the post-war period, 1948-2015, the data were collected from the Annual Report of School Health Statistics Research (ARSHSR) (*Gakkō hokentōkei chōsa*) published by the Ministry of Education, Culture, Sports, Science and Technology. Between 1948 and 1950, the ARSHSR collected anthropometric data for all kindergarten, primary school, junior school, and high school students. However, after 1954, the ARSHSR began using a stratified random sampling method with three steps. In the first step, they stratified schools by type (i.e., kindergarten/primary/junior high/high) and number of students to ensure coverage of smaller schools. They then selected schools in each of the strata to survey. Finally, each selected school was stratified by age and sex to ensure large enough sample sizes in each bin. The Ministry of Education (1985– 2018) describes the details of the sampling in the post-war period. There were a few periods of missing data because prefecture-level reports were not published for these years (c.f. 1951-54, 1970-71 and 1977-78).

A.2. Selection Bias in the School Sources

As mentioned in Section 3 of the main text, we were concerned that although primary school enrolment was very high throughout the twentieth century, there might have been positive selection into secondary school when secondary school enrolment was low in our sample (Schneider 2020). This is because those remaining in secondary school may have been of higher socioeconomic status than the average person in the population. This Appendix builds from Schneider (2020) showing that there was positive selection into secondary school, and it describes how we manage this selection bias in our analysis.

Although it is not possible to view potential selection bias in prefecture average growth curves that mix different levels of schooling, in several short periods, the prefecture-level data are provided for different school types. In the pre-war period, Figure A1 presents the weighted mean height curve of boys in primary and secondary school measured in 1936 compared with the national height curve used by Ali et al. (2000) and Mosk (1996) for the same year. Looking first at the differences between primary and secondary schoolchildren, boys in secondary school were on average 3.4 cm taller than primary school boys at ages 12 and 13. In addition, at age 20 the secondary school boys were 1.3 cm taller than military conscripts, which again reflected 95% of the population. This positive selection is related to the share of boys in school also displayed in figure A1. The share of children in school fell from around 95% to 10% during the transition between primary and secondary school leading to the positive selection on height. This pattern is similar for girls as well. This selection is particularly problematic because it is not a simple shift in size. The larger height advantage of the secondary school boys in puberty relative to adulthood suggests that the secondary boys also experienced an earlier pubertal growth spurt and accelerated development relative to the primary boys.

Figure A1 also shows the potential bias created when using the national growth curves. The national curve traces the primary school data until there is overlap in the primary and secondary school data. Then, the national curve shifts to the secondary school data as children leave primary school and enter secondary school. The shift upward in the national growth curve from the representative primary school data

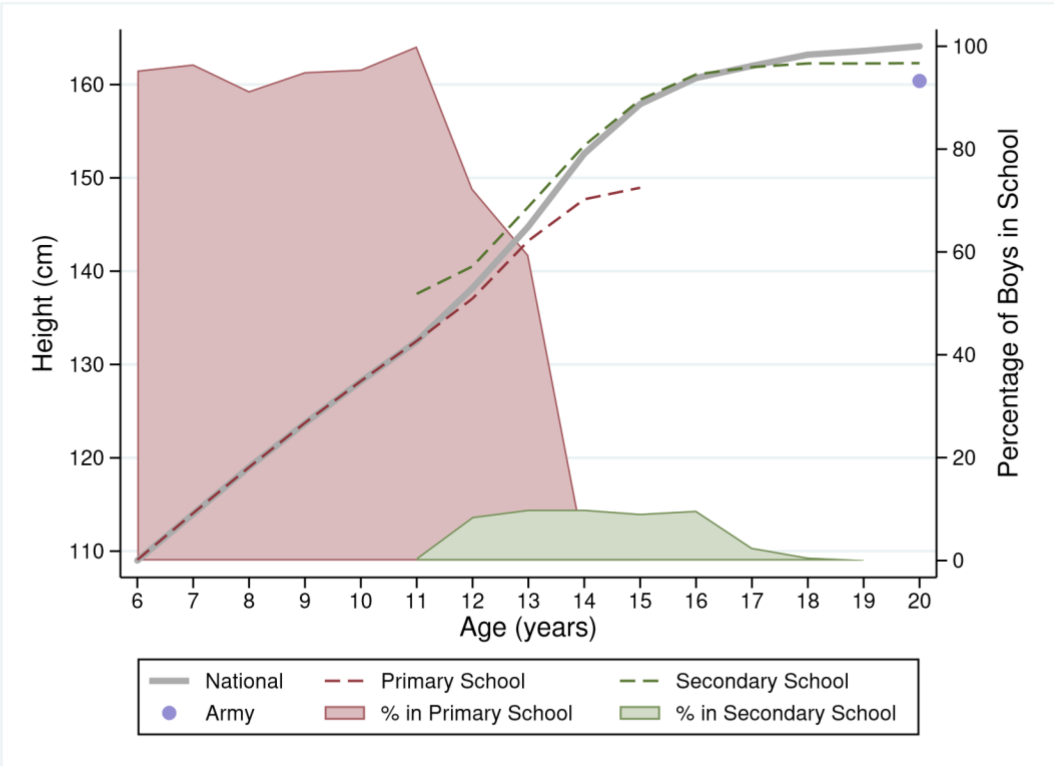


Figure A1. Positive selection of secondary school boys on height before the war (1936) leads to bias in the national growth curve after age 13.

Notes: National is the national mean period height curve measured in 1936, the same data that Ali et al. (2000) use. Army marks the population weighted mean height of men at age 20 upon their army medical inspection and covers nearly all men. Primary and Secondary are population weighted means of the mean prefecture heights of boys in each school type at each age measured in 1936. The area graphs show the percentage of boys in primary and secondary schools at each age and is based on data from 1938.

Sources: Japanese Prefecture Child Growth Dataset - see Appendix A.1 for sources; enrolment rates from the 1935 census (Statistics Bureau of the Cabinet 1939).

to the positively selected secondary school data artificially increases growth velocity during these years, accentuating the pubertal growth spurt and growth velocity during the growth spurt. The national growth curve is also positively selected in adulthood.

This pattern of selection is still present after the war. Figure A2 plots the weighted mean height for boys measured in 1948 at various school levels. Again the national cohort growth curve is plotted underneath. Positive selection does not appear to be a problem until children shift from junior school to high school after age 14. However, high school boys were substantially taller than their junior counterparts, and this selection again occurs at the moment when the vast majority of boys left school (see figure A2). The clear upward shift in the national data suggests that using the data up to age 17 as other authors have done would lead to bias. In addition, the small number of children in kindergarten raises questions about the representativeness of this sample, especially when looking at prefecture-level data where kindergarten policies may have differed. Therefore, we need a new strategy for analysing these datasets.

As mentioned in the main text, to overcome the selection bias problems, we enter the representative and selected cohort growth curves independently as separate individuals in the SITAR model. We need to include both growth curves because SITAR needs the non-linearities in growth in adolescence to distinguish between vertical shifts in the growth curve resulting from changes in size and horizontal shifts that affect the timing of the pubertal growth spurt. We define the representative and selected growth curves as follows. For the pre-war prefecture-level data, 1929-39, the representative data are the mean heights of primary school children from ages 6 to 13 when the proportion of children in primary school is still high and the bias seems minimal. The selected data are the mean heights of secondary school children from ages 12 to 19 or

18 for boys and girls respectively. Between 1948 and 1969, there is still evidence of positive selection into high school, but after 1950 we do not have data for separate school levels. Thus, we have taken the prefecture average height from ages 6 to 14 where the data appear to be minimally biased as the representative group and the mean heights from ages 15 to 17 as the selected group. After 1972, the share of students in secondary school was high enough that we feel confident in the prefecture mean heights from ages

6 to 17. Again, all analysis in the paper focuses on the representative data that are

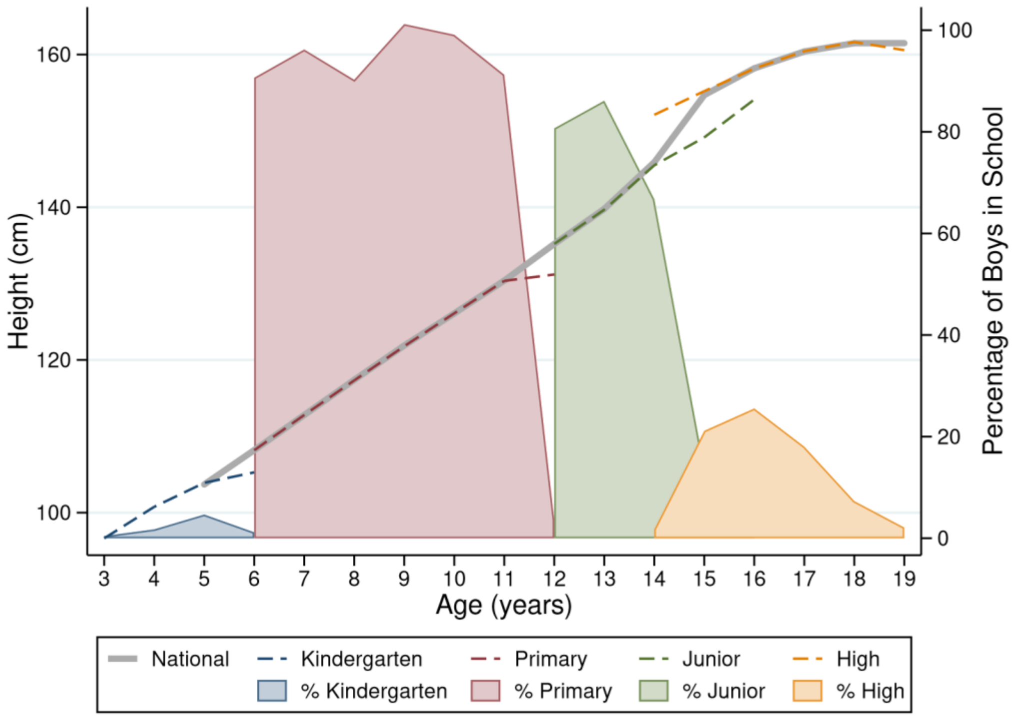


Figure A2. Positive selection of high school boys on height in the immediate aftermath of the war (1948) leads to bias in the national growth curve after age 14.

Notes: National is the national mean period height curve measured in 1948, the same data that Ali et al. (2000) use. Kindergarten, Primary, Junior and High are population weighted means of the mean prefecture heights of boys in each school type at each age measured in 1948. The area graphs show the percentage of boys enrolled in each school level at each age in 1948.

Sources: Japanese Prefecture Child Growth Dataset - see Appendix A.1 for sources; enrolment rates from the 1948 census (Statistics Bureau, Prime Minister's Office 1948).

independent of the selected data in the SITAR model. As a final point, note that there is no child growth data for Okinawa between 1940 and 1971, so it is excluded from the analysis when analysing the war shock.

A.3. Historical Sources for Figure A5

Sex Ratio

Statistics Japan (2007). *Population Estimates of Japan 1920-2000*. Table 6. Population by Sex and Sex ratio for Prefectures (as of October 1 of Each Year) - Total population (from 1920 to 2000). Available at https://www.e-stat.go.jp/en/ stat-search/files?page=1\&layout=datalist\&toukei=00200524\&tstat=000000090001\&cycle=0\&tclass1=000000090004\&tclass2=000000090005 (data downloaded on 9 April 2019).

Infant Mortality Rate and Total Births

Japanese Statistical Association (1987). *Historical Statistics of Japan*. Volume 1. Table 2 – 28, pp. 210-211.

Survivors in 1950

Japanese Statistical Association (1987). *Historical Statistics of Japan*. Volume 1. Table 2 – 9, p. 69.

B. Additional Figures and Tables

Table T1. Ration levels of staple crops per day in six principal cities in April 1941

	Rationlevelsformales		Rationlevelsforfemales	
Age and work intensity	Rice (grams)	Calories (kcal)	Rice (grams)	Calories (kcal)
0–4	120	421	120	421
5–9	200	702	200	702
10–59 (Regular work)	330	1158	330	1158
10–59 (Heavy work)	390	1369	350	1228
10–59 (Extremely heavy work)	570	2001	420	1474
60 + (Regular work)	300	1053	300	1053
60 + (Heavy work)	350	1228	320	1123
60 + (Extremely heavy work)	480	1685	380	1334

Sources: Ohara Institute for Social Research 1964, Table 74 of Chapter 2 of Section 5. The original figures are from the report by the Ministry of Agriculture and Forestry.

Table T2. Percentage of certain food products purchased on the black market

	Tokyo		Averageof24Prefectures
Food	March 1944	Sept/Oct 1944	March 1944

Rice and wheat	0.5	9.0†	4.0
Vegetables	10.0	68.5	36.0
Fish	12.5	37.7	39.0
Meat	18.5		22.5

Notes: Unfortunately, the USSBS publication does not provide details about which prefectures are included in the 24 prefecture average, and it was not possible to track down the original Institute for Science of Labor report during the pandemic. †Rice only.

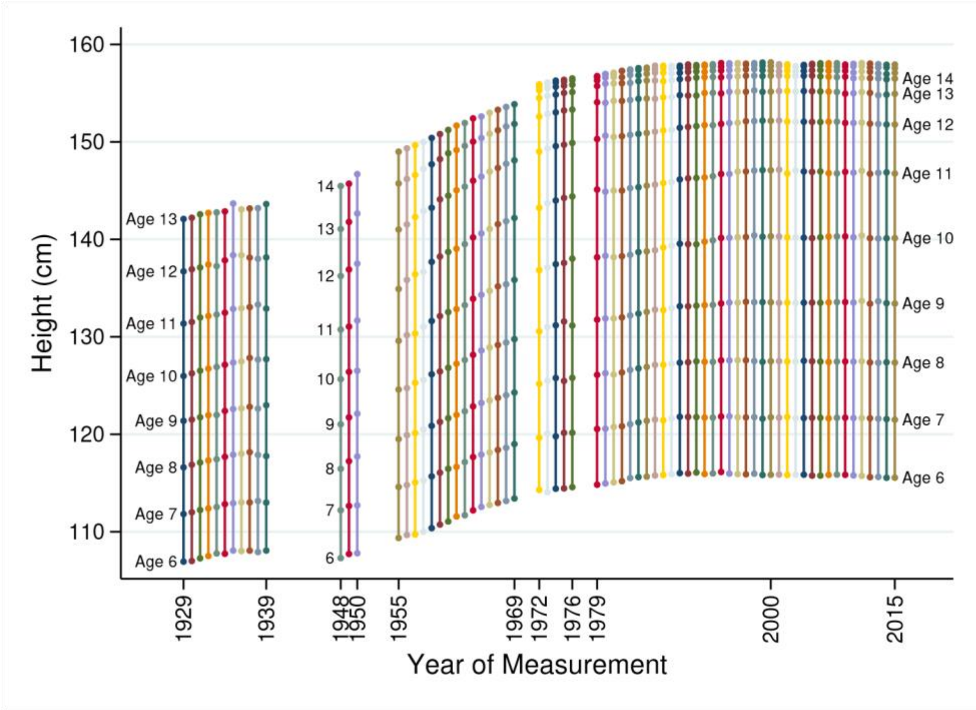
Sources: United States Strategic Bombing Survey (1947, p. 29). The original figures are from reports by the Institute for Science of Labor.

Table T3. Descriptive Statistics of the War Shock: 1931-33 vs. 1934-36

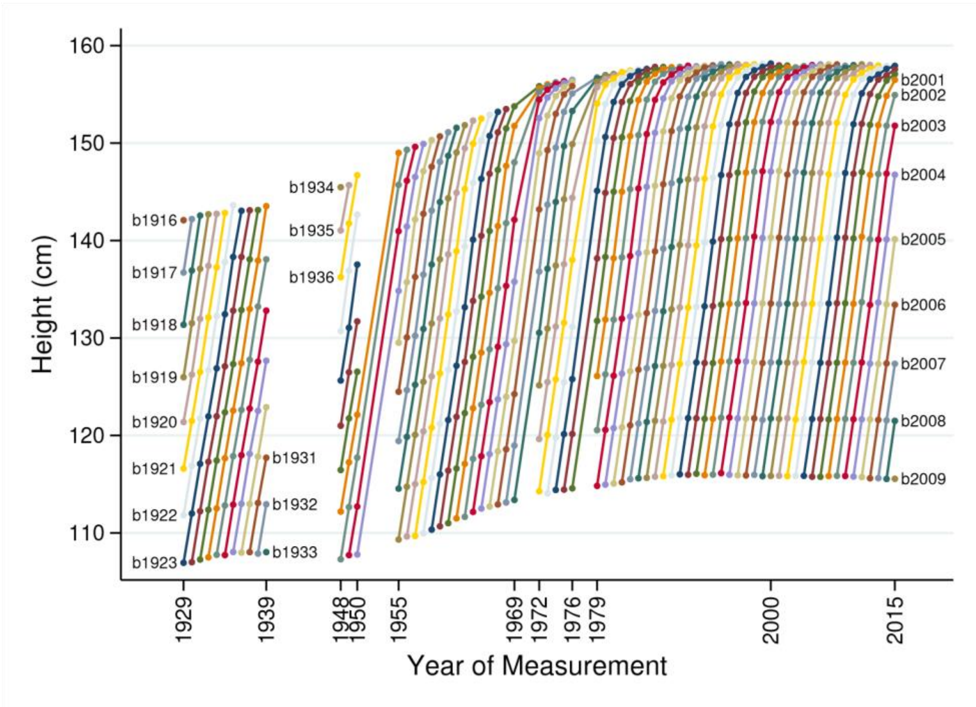
	Male			Female		
	Size (cm)	Timing (years)	Intensity (fractional)	Size (cm)	Timing (years)	Intensity (fractional)
Mean	2.9	0.50	0.058	1.7	0.46	0.043
Standard Deviation	0.9	0.12	0.014	0.8	0.16	0.025
Minimum	0.8	0.22	0.022	-0.6	0.08	-0.044
Maximum	4.9	0.77	0.089	4.4	0.91	0.14
N	46	46	46	46	46	46

Notes: The war shock is the mean of each SITAR parameter predicted for the 1934-36 birth cohorts minus the mean of each SITAR parameter predicted for the 1931-33 birth cohorts. We have multiplied the size and intensity parameters by −1 so that larger values of the war shock variable reflect a greater war shock.

Sources: Japanese Prefecture Child Growth Dataset - see Appendix A.1 for sources.



(a) Female period growth curves.



(b) Female cohort growth curves.

Figure A3. Graphical representation of period and cohort growth curves for girls

Notes: Each point reflects the population weighted mean height across all prefectures at each age in each year. Period growth curves are for children measured in the same year. Cohort growth curves relate to children born in the same year but measured in successive years as the cohort ages.

Sources: Japanese Prefecture Child Growth Dataset - see Appendix A.1 for sources.

C. Nutritional Availability Adjusted for Changing Age Structure and Sex Composition

One might reasonably wonder how the changing age structure and sex composition of the Japanese civilian population across the war would have affected the availability of nutrients. Prime-age adult males require more calories to survive than women and children, so the removal of many of these men from the population as they were sent across Asia to fight the war may have ameliorated the nutritional decline somewhat. This appendix shows how great this ameliorating effect may have been.

We start by pulling together the estimates of per capita calorie consumption per day in Japan from a variety of sources from 1925 to 1955. Unfortunately, this data is not available as a continuous series, but it does indicate trends in consumption before, during and after the war. To adjust these figures for the age structure and sex composition of the population, we collect data on population by age and sex from censuses matching our calorie estimates and also population estimates for the war period. Estimates for population by age and sex do not exist between 1941 and 1943, but estimates from population surveys are available for 1944-46. These surveys were taken on 22 February 1944, 1 November 1945 and 26 April 1946 respectively, so they are not able to capture the first half of 1945 when the most soldiers were overseas (Johnston 1953, p.245) and they do not perfectly match the timing of the nutritional availability estimates. However, these estimate can be useful in providing a sense of how important adjusting for age structure and sex composition might be.

To make the adjustment, we follow the method of Floud et al. (2011, pp. 165-68). Children and adults of each sex are assigned a consumption value as a share of the consumption of an adult male aged 20-39 (a consuming unit). For instance, boys age 0-4 have consumption equal to 0.4413 consuming units, whereas women over the age of 60 have consumption equal to 0.55 consuming units (the full weights are given in (Floud et al. 2011, p. 166)). We then multiply the share of the population in each age group by their consumption equivalent and sum these up to gain an adjustment factor for the population as a whole. This adjustment factor measures the number of consuming units present in a population per capita. Because populations always have large numbers of women and children, the adjustment factor is always less than one. To obtain an adjusted calorie figure per consuming unit, we simply divide the per capita calorie figures by the adjustment factor.

Table T4 displays the results. Obviously, making this adjustment will increase the calories available per consuming unit, but what is more important is the extent to which the pattern in calories available changes across the war. Column 4 shows that the basic patterns of nutrition described in the main text are replicated when looking at calories per consuming unit. In column 5, we compute calories per consuming unit holding the adjustment factor constant at the 1935 (pre-war) level to show how important adjusting for age structure and sex composition is during the war. A comparison between the two columns shows only minor differences in the series. Although the mass mobilization of young men did ameliorate the size of the nutritional shock by 21 kcal per consuming unit per day in 1945, this effect is very small relative to the 486 kcal per consuming unit per day decline in calorie availability between 1935 and 1945. Thus, adjusting for changing age structure and sex composition does not change the general story described in the main text, and we rely on the more straightforward figures there.

Table T4. Nutritional availability adjusted for the changing age structure and sex composition of Japan

Year	Calories (kcal/capita/day)	Age-Structure and Sex Adjustment Factor (consuming units/capita)	Adjusted Calories Current Year (kcal/consuming unit/day)	Adjusted Calories with Constant 1935 Age Structure (kcal/consuming unit/day)
1925	2325[a]	0.753	3089	3084
1935	2175[b]	0.754	2885	2885
1940		0.753		
1941	2105[c]			2792
1942	1971[c]			2614
1943	1961[c]			2601
1944	1927[c]	0.750	2569	2556

1945	1793[c]	0.747	2399	2378
1946		0.758		
1947		0.756		
1950		0.752		
1955	2217[d]	0.762	2908	2940
2015	2701[e]	0.746	3619	3582

Notes: See text of Appendix C for details of the calculations. Consuming units are adult male (aged 20-39) equivalents following the method of Floud et al. (2011, p. 166). Population figures from 1925, 1935, 1940, 1947, 1950, 1955 and 2015 are drawn from censuses. Population estimates between 1941 and 1943 were not available. Population estimates from 1944 to 1946 are based on population surveys that were unfortunately taken at different points during the year. The three surveys were taken on 22 February 1944, 1 November 1945 and 26 April 1946. This means that there is not a population estimate for the very end of the war when the largest number of men were overseas. We have confirmed that these estimates do not include men in the military overseas by comparing with general population tables reported in Johnston (1953, p. 245). The population figures for 1945 and 1946 also exclude Okinawa.

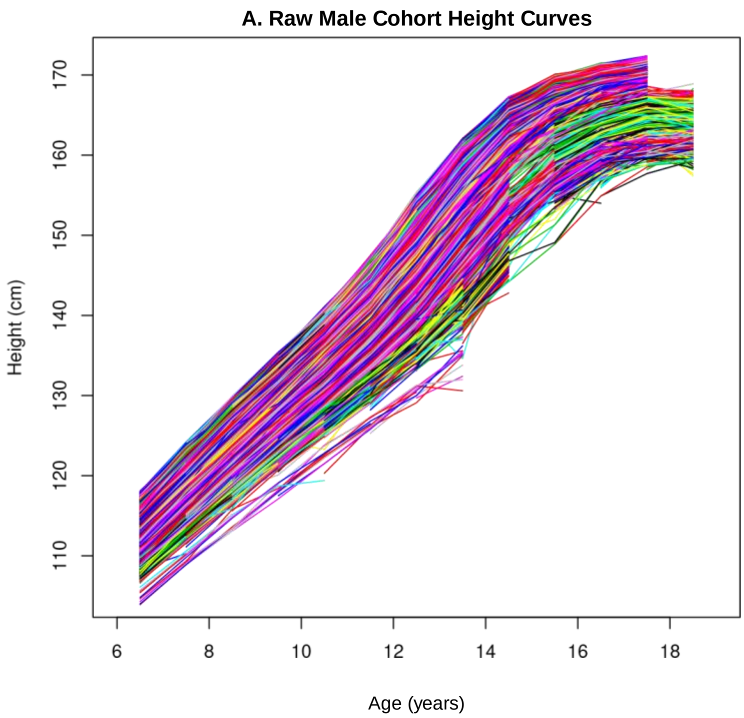
Sources: Population estimates by age and sex from 1925 to 1955 are from Statistics Japan (2007). *Population Estimates of Japan 1920-2000*. Table 3. Population by Age (5-Year Group and 3 Groups) and Sex (as of October 1 of Each Year) - Total population (from 1920 to 2000). Available at https://www.e-stat.go.jp/en/stat-search/files?page=1\&layout=datalist\&toukei=
00200524\&tstat=000000090001\&cycle=0\&tclass1=000000090004\&tclass2=000000090005
(data downloaded on 9 June 2021). Population estimates for 2015 drawn from the Human Mortality Database (2019). Sources for the calorie estimates are indicated in the table as follows: [a]Hayami and Yamada (1970, p. 81); [b]Johnston (1953, p. 277); [c]Ohara Institute for Social Research (1964, Table 78 of Chapter 2, Section 5); [d]Sanderson (1987); [e]FAO (2021).

D. Full Description of SITAR Model and Results

Building on the explanation provided in Section 4.1, Figure 3 shows how an increase or decrease in each SITAR parameter would affect the height and velocity curve using the WHO male median growth reference curve as a starting point. Increases or decreases in the size parameter shift the height curve vertically but do not affect the velocity curve. Increases or decreases in the timing parameter shift both the height and velocity curves horizontally, altering the timing of the pubertal growth spurt but not changing the shape of the velocity curve. Finally, increases or decreases in the intensity parameter shrink or stretch the age scale so that the same absolute increase in height is achieved in a shorter or longer time period. This shrinking/stretching of the age scale allows for differences between chronological age and developmental age since some children develop more quickly than others.

Thus, SITAR uses the three parameters to shift the mean curve that is predicted by the spline in the model to each cohort growth curve in our sample. Figure A4a presents the raw male cohort height curves from 1916 to 2010, and Figure A4b presents these curves after SITAR has adjusted them all for the three parameters. Clearly, the model is able to explain much of the variance in growth across the sample.

Table T5 presents the full results for the SITAR models, which were estimated with the SITAR package in R (Cole 2019; R Core Team 2019). We tested models with different degrees of freedom to see which provided a lower Bayesian Information Criterion (BIC) value. The best fit was for a spline with 5 degrees of freedom for males and 6 degrees of freedom for females. Comparing our models with others in the literature, the standard deviations of the parameters are lower than what has been found in other studies since our pooled cohort growth curves are averages of large numbers of children instead of individual-level observations. The residual standard deviations are also lower than individual-level studies as expected since grouped data will contain less measurement error (Cole et al. 2010). The pooled nature of our data also affects the correlations between the parameters which are very substantially higher than what is typically observed in individual-level data, although the positive correlation between size and intensity and the negative relationship between timing and



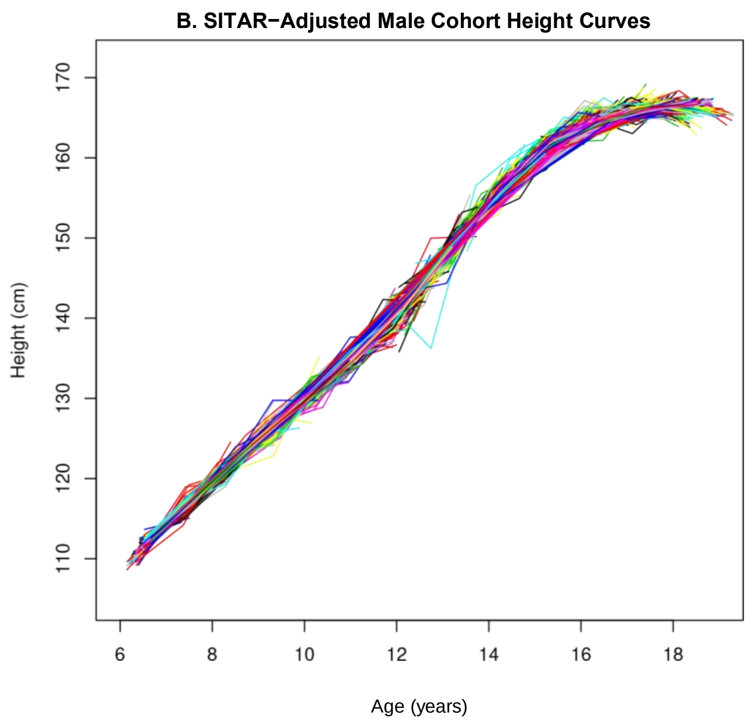


Figure A4. Raw and SITAR-adjusted height curves

Notes: Estimated using the SITAR package in R (Cole 2019; R Core Team 2019).

Sources: Japanese Prefecture Child Growth Dataset - see Appendix A.1 for sources.

Table T5. Summaries of SITAR models for boys and girls

	Boys			Girls		
	Standard Deviation	Correlations		Standard Deviation	Correlations	
		Timing	Intensity		Timing	Intensity
Size(cm)	3.37			2.68		
Timing (years)	0.52	-0.85		0.59	-0.91	
Intensity (fractional)	0.06	0.91	-0.95	0.07	0.94	-0.91
Residual (cm)	0.45			0.44		
Model Characteristics:						
Number of Height Observations	41,833			40,542		
Number of Cohort Growth Curves	6,251			6,114		
Degrees of Freedom	5			6		
% Variance Explained	99.1			99.0		

Notes: Estimated using the SITAR package in R (Cole 2019; R Core Team 2019).

Sources: Japanese Prefecture Child Growth Dataset - see Appendix A.1 for sources.

intensity is expected.

To gauge the effectiveness of the SITAR model in explaining the growth curves, we compute the variance explained by the model as follows:

$$\%\ \text{variance explained} = 100 \times \left(1 - \left(\frac{\sigma_2}{\sigma_1}\right)^2\right) \qquad \text{(E1)}$$

where σ_1 is the residual standard deviation from a model simply fitting a spline curve and σ_2 is the residual standard deviation from a model that adds random effects α, β and γ to allow each cohort growth curve to differ from the mean curve. Thus, this indicator explains the variation explained by adding random effects to the model. The variation explained by the model by this definition is 99.1% and 99.0% for males and females respectively, suggesting that the SITAR model is doing an excellent job of explaining the variation in the growth curves. Given the strength of fit of our data with SITAR, we feel confident in using the SITAR parameters to analyse change in the growth pattern of children across the war.

E. Selection by Cohort

E.1. Selection by Cohort: Main Counterfactual, Birth Cohorts 1931-33 vs. 1934-36

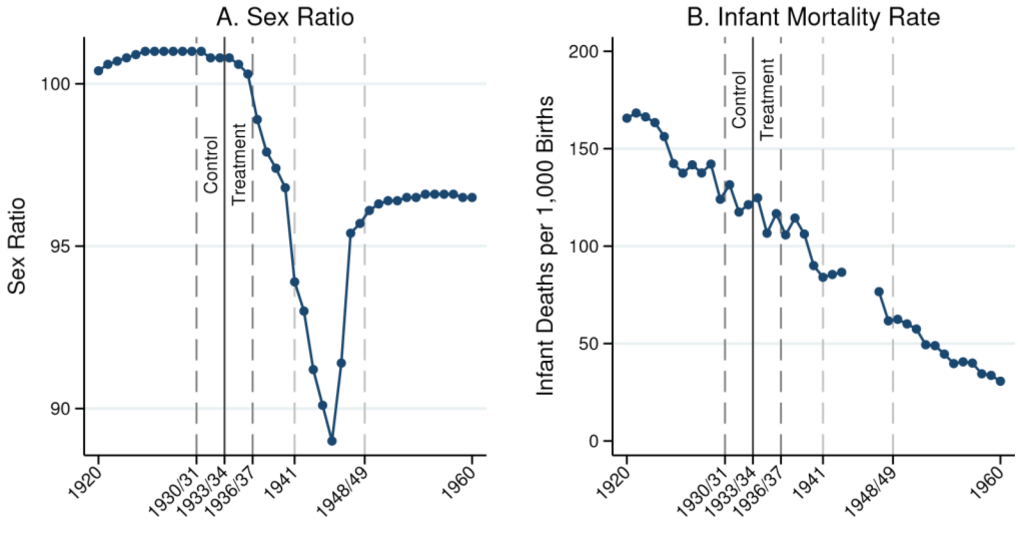
We start by discussing selection related to our main counterfactual comparison: birth cohorts 1931-33 vs. 1934-36. Beginning with selection into fertility, although there was year-to-year fluctuation in the total number of births, there is no reason to believe that there was differential selection into fertility by cohort between 1931 and 1936 (Figure A5c). Certainly, the typical mechanisms of fertility selection that are discussed in the famine literature would not be at play here since these years pre-dated the war.

Another potential source of selection bias by cohort would be the health conditions in early life for each cohort. Poor health conditions in early life may have had a detrimental effect on the latent health of a cohort, scarring the cohort relative to comparison cohorts. On the other hand, fluctuations in infant mortality year-to-year may have led to selective culling of the weak, which would leave the survivors with a higher level of latent health (Deaton 2007). In general, health conditions were improving across the 1930s before the war so if anything the 1934-36 cohorts should have experienced less scarring than the 1931-33 cohorts. Infant mortality rates were also falling (Figure A5b) suggesting that if there were an effect of selective culling of weak individuals, it would have become less pronounced over time.[13]

Our inverse probability weighting exercise described in Section 5, however, is the most convincing evidence that there is no real difference between the war-untreated and war-treated cohorts.

E.2. Selection by Cohort: Children Born during and after the War, 1937-50

It is more difficult to assume that children born during and after the war did not experience differential selection into fertility or selection and scarring across cohorts.



C. Registered Births or Survivors from Each Cohort in 1950

[13] Previous research has shown that there was not a strong (or statistically significant) relationship between infant mortality in the year of birth and any aspect of the growth pattern or the heights of children at ages 6 or 10 in interwar Japan (Schneider and Ogasawara 2018), suggesting that any scarring and selective culling effects may have been counteracting each other.

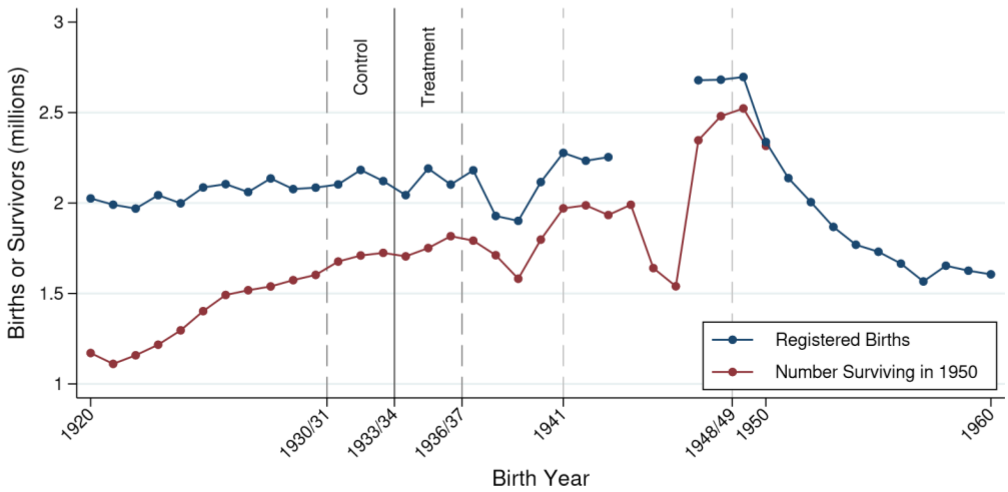


Figure A5. Sex ratio, infant mortality rate and total births of the Japanese population in Japan, 1920-60

Notes: The cohorts labelled treatment and control refer to our main counterfactual, which compares the 1931-33 and 1934-36 cohorts. Sex ratio data are estimates based on the standard censuses and population surveys in 1944, 1945 and 1946. Neither infant death nor live birth registration was reliable between 1944 and 1946, so these years have been excluded from the graphs. The number surviving in 1950 shows the number of people from each cohort enumerated in the 1950 census and provides an indication of what happened to the cohorts born during the war. It seems likely that there were substantially fewer births in 1945 and 1946, but the lower level of survivorship could also be explained by an increase in infant mortality.

Sources: See Appendix A.3.

Selection into fertility may have been important given the mass mobilisation of reproductive-aged soldiers during the war. Figure A5a shows the sex ratio for the Japanese population present in Japan between 1920 and 1960. Although stable before, beginning with the Second Sino-Japanese war in 1937, the sex ratio dropped from over 100.3 in 1936 to 89 in 1945 as more men entered the military and were sent to fight abroad. After the war, the sex ratio recovered to a stable level around 96.5 considerably below the pre-war level. Interestingly, the effect of this mass exodus of men did not have a straightforward effect on the number of births (Figure A5c). The number of births was more or less stable through 1937 with a substantial drop in 1938 and 1939. However, after a few years of lower births, the government instituted pronatal policies encouraging families to have more children and births rose above their pre-war level from 1941 to 1943 (Mathias 1999, pp. 72-74; Partner 2007, pp. 150-51).

Unfortunately, there were serious problems with vital registration between 1944 and 1946, so the statistics are considered unreliable.[14] This limits our ability to know how births were responding at the end of the war, though the low numbers of children surviving in 1950 for these cohorts suggests that births likely fell in 1945 and 1946. Finally, after the war there was the standard baby boom following a famine or other traumatic event.

These large changes in the number of births during the war suggest that there may have been substantial changes in the types of couples giving birth. For instance, Watanabe (2014) finds that working class men were more likely to be drafted than middle and upper class men, which would mean that children born during the war may have been positively selected relative to the pre-war and post-war periods. Without individual-level data, it is difficult to account for this compositional change. Thus, we must be somewhat circumspect about comparing cohorts after 1937 since selection into fertility may have influenced the distribution of latent health across cohorts.

In addition, there may be differences in scarring and selective culling across cohorts. Unfortunately, we are not able to estimate inverse probability weighted average treatment effects comparing cohorts born during the war to those born after the war because of missing health data. As mentioned in Section 2, the evidence of the income, nutritional and general health shock during the war is either aggregated at the national level or is based on incomplete, anecdotal evidence. Thus, we can only speculate about the balance of scarring and selection across these cohorts. The evidence for scarring is vast and emphasised strongly in Section 2 in the main text, so the real question is whether the intensity of selective culling increased for cohorts born and experiencing their early childhood years during the war. Figure A5b shows that infant mortality rates increased from 1941 to 1943 after which figures become unreliable. Although the infant mortality rate was still lower in 1943 than it was in 1940, mortality rates likely increased substantially for children during the war and could have led to

[14] The Japanese Historical Statistics Volume states that "vital statistics for 1944, 1945 and 1946 were extremely incomplete owing to war damages", so there was likely under-reporting and official statistics do not report values for these years (Japan Statistical Association 1987, p. 47).

increased selective pressure, especially from 1944 to 1946 (Nakashima 2011; Yoshida 1995, pp. 265-66).

However, there were also a number of policies and programmes targeted at promoting nutrition and health specifically among children and pregnant women that may have mitigated some of the worst effects of the war for these groups. For instance, beginning in July 1943, the government distributed maternal and child health handbooks to pregnant and nursing women. Mothers with these handbooks were given upgraded rations containing milk, sugar and 60 grams (211 kcal) of additional brown rice-equivalent staple foods (Johnston 1953, p. 202; Yoshida 1995, p. 267). In addition, although milk production was very low before the war and collapsed during the war, milk rations (as dried, evaporated or condensed milk) were prioritised for infants whose mothers were unable to breastfeed them (Johnston 1953, p. 208).

Finally, towards the end of the war in 1944 and 1945, 1.3 million urban children were evacuated to the countryside: 857 thousand were sent to live with relatives and 446 thousand third to sixth graders were evacuated as school groups (Yamashita 2013). Children living with relatives may have had access to a healthier diet than they would have been able to get in urban areas (Partner 2007, p. 147). However, children evacuated in groups suffered from prolonged hunger in the countryside with their diet deteriorating as the war progressed (Yamashita 2013). Evacuated children were also subjected to bullying and did not have local networks to help supplement their food consumption at the end of the war (Cook and Cook 1992, pp. 233-34). Thus, it is not clear to what extent evacuated children experienced better health conditions than their urban counterparts.

Determining the overall effect of these programmes along with other changes in health during the war is difficult. Mitigating factors may have reduced some of the worst effects of the war for young children, but these programmes were not completely effective at preventing an increase in infant mortality at the end of the war (Nakashima 2011; Yoshida 1995, pp. 265-66). There was likely space for selective culling to be an issue. Infant mortality rates were lower after the war from 1947 when the data becomes reliable again. Without reliable birth and infant death data between 1944 and 1946 it is very difficult to provide firm bounds on the selective culling effect.

However, despite this, the evidence presented in Figure 6 does not support a clear "thousand days" scarring or selective culling story for the cohorts born during and after the war. As mentioned in the main text, Figure 7 compares the observed pattern with hypothetical selection and scarring effects to make this clear. Thus, although we cannot be as precise in estimating a causal effect of exposure to the war by comparing cohorts born during the war with those born before or after the war, our finding that cohorts experiencing the war in late childhood or adolescence were more severely affected is unlikely to be explained by selection. Likewise, the relatively small differences between those experiencing their first thousand days during the worst conditions of the war and those born after the war suggest that there was substantial scope for catch-up growth when conditions improved after the war.

E.3. Selection from the Second World War

Aside from selection across cohorts, there may be a more general selection effect if people who survived the Second World War were healthier than those who died. We have seen how this might have influenced cohorts differently by adjusting age-specific mortality levels, but it is important to consider the question more broadly since it may bias our pre-war and post-war comparisons, especially if the healthier surviving adults influenced the health of children born after the war (Gørgens et al. 2012).

The scale of Japanese military and civilian deaths was very substantial. Gruhl (2007, p. 144) reports that Japan lost 2,315,878 military personnel and 672,000 civilians as a consequence of the war between 1942 and 1945. The military deaths were 6.7% of all men in Japan in 1935, and the civilian deaths were 1.0% of the total population in 1935. However, it is difficult to assess or understand the extent to which war deaths would have been selected on pre-war health characteristics in a way that would matter for our study.

To our knowledge there are no direct studies that have looked at whether Japanese battle deaths were related to soldiers' socioeconomic characteristics. However, evidence from other contexts suggests that the selection is not always straightforward. During World War I, mortality was higher among British officers than regular soldiers (Winter 1977). However, in France the picture was more mixed: sons of French urban white collar workers had lower mortality during World War I, but sons of wealthy rural inhabitants had higher mortality rates than their rural non-wealthy counterparts Kesztenbaum (2018). Returning to Japan, Watanabe (2014) has studied selection into military service during World War II and finds that blue collar workers were more likely to be drafted during the war but that there were no social class differences in the time spent in military service. Thus, even if surviving the war was not dependent on social class, there might have been some selective culling at the bottom of the social class distribution. Unfortunately, without a detailed study on Japanese military deaths during the Second World War, it is not possible to definitively understand whether selective culling during the war affected the health of survivors.

As for civilian war-related deaths, the majority of deaths on mainland Japan were caused by the Allied bombing campaigns. Intensive American bombing of Japanese cities began in June 1944. It was initially aimed at 'industrial, transportation, and communication targets' (Weber 2001, p. 113), but the aims were later expanded to target Japanese cities with incendiary bombs, with the US military arguing that industry was widespread in cities and that cities were vital to the Japanese war effort (Henderson 2011; Weber 2001). Incendiary attacks on cities accounted for 80% of bombings on Japan, and these attacks caused immense damage and civilian casualties: for instance, the bombing of Tokyo on 9 March 1945 burned 15 square miles of the city and killed at least 78,000 civilians (Henderson 2011, p. 313). The atomic bombs directly killed another 97,059 people according to Japanese accounts with at least as many deaths occurring later from radiation poisoning (Weber 2001, p. 114). However, given their indiscriminate nature, it is difficult to understand how these bombings would have influenced socioeconomic groups differentially. The bombs did target urban residents who were likely less healthy than their rural counterparts,[15] but the bombings would

[15] Infant mortality rates were higher in major cities until the late 1920s when they achieved lower mortality than other areas.

have contributed to the scarring of survivors as well since by one estimate the bombings injured 476,000 civilians and left 9.2 million people homeless (Henderson 2011, p. 314).

In any case, if the war did lead to generalised selective culling, then this would make our estimates biased toward finding no negative effect of the war. Since we find a very significant unhealthy change in the growth pattern between the 1933 and 1934 birth cohorts, this suggests that the effects of the war may have been even greater than our estimate.

F. Hypothetical Scarring and Selective Culling Effects in Figure 7

Figure 7 provides a hypothetical view of how a pure thousand days scarring or selective culling pattern would look. These are hypothetical curves drawn simply to illustrate the expected pattern in each of these cases. The thousand days scarring pattern is based on the intuition that cohorts who were exposed to the worst conditions of the war at early ages would be more strongly affected by the war than cohorts experiencing the war outside the first thousand days. Thus, we assume that cohorts who did not experience their first thousand days during the war were no different than the counterfactual cohorts and that the 1944 and 1945 cohorts that experienced the entirety of their first thousand days during the worst health conditions of the war would have been three centimetres shorter than the counterfactual cohorts. We do not have estimates for the potential size of the shock, so again the figure is simply an illustration of what a thousand days scarring pattern would look like across the cohorts.

In addition to the scarring effects, there may well have been some selective culling as well. To produce these estimates, we use the Japanese age-specific mortality schedule for the 1947 birth cohort from the Human Mortality Database (2019) to predict the number of deaths for each cohort relative to the 1933 cohort, allowing mortality rates to be 25% higher in the less difficult years of the war (1942-43, 1947-48) and to be 75% higher than normal in the most difficult years (1944-46). We then generate a normal distribution of heights for each cohort and remove the percentage of individuals predicted to die (relative to the 1933 cohort) from the lower tail of the height distribution. This is an upper-bound estimate of the selective culling effect since mortality is unlikely to be perfectly correlated with height. Selective culling does not have a strong effect on the early war-affected cohorts because the cohorts experienced the war mostly in late-childhood and adolescence when mortality rates are much lower than in infancy and early childhood (see Figure 5). However, once children began to experience the war in infancy, the potential for selective culling increased and was highest among those cohorts experiencing infancy and early childhood during the peak of the war. The selective culling effect then attenuated for children born late in the war since conditions improved shortly after they were born.

If we carry this thought experiment further and average the selective culling and scarring effects to get the net effect, we still cannot produce a pattern that matches the observed pattern. No combination of realistic patterns of thousand-days scarring and selective culling, despite what the magnitude of each effect might be, produces such a sharp difference between the 1933 and 1934 birth cohorts, nor does it explain the monotonic, linear improvement in the SITAR parameters across cohorts C and D (see Figure 6).

G. Post-war Civilian Repatriation

After the war 3.2 million civilians were repatriated to Japan, including a large number of children. These children potentially add measurement error to our comparison of the war-untreated 1931-33 cohorts and the war-treated 1934-36 cohorts since we only observe the war-treated cohorts after the war from 1948-50 when the repatriate children would be included in our prefecture mean height measurements. This appendix considers the post-war repatriates and shows that this measurement error is very unlikely to bias our results.

As Japan's empire had expanded in the late nineteenth and early twentieth centuries, many Japanese had moved abroad to serve as colonial administrators or in other semi-official positions, helping to manage the colonial state. Later, government programmes incentivised farming households to settle in Manchuria. Estimates of the number of Japanese civilians abroad by the end of the Second World War are not precise but were likely in the range of 3.2 to 3.5 million people. Table T6 gives a sense of the where Japanese civilians were located at the end of the war. Nearly half were settlers in Manchuria, but there were sizeable groups in the Korean Peninsula, China, Sakhalin and Taiwan (Nishizaki 2019; Watt 2009, p. 2).

Following Japan's surrender, the vast majority of these civilians along with 3.7 million demobilized soldiers were repatriated to Japan. The Allies were particularly focused on returning the demobilised soldiers quickly but soon realised that they would need to help facilitate the return of Japanese civilians as well since Japanese civilians faced violence and poverty in the former colonies (Watt 2009, pp. 39-41). Overall the repatriation process proceeded very quickly with 5.1 million civilian and military repatriates returning to Japan by the end of 1946 and an additional 744 thousand returning by the end of 1947, cumulatively 81% and 93% of the total returnees in each year respectively (Watt 2009, p. 77). Thus, the vast majority of repatriate children would appear in our war-treated cohorts, which are observed from 1948-50.

When reaching Japan, the repatriates were processed in 15 regional repatriation centres, most of which were located on Kyushu or southwest Honshu. From there the Ministry of Health and Welfare "provided transportation back to their previously **Table**

T6. Estimates of civilian Japanese nationals living abroad at the end of the Second World War

Location	Estimated Japanese Civilian Population 1945 (1000s)	Japanese Civilian Returnees by 1977 (1000s)
Manchuria	1,550	1,219
China, including Hong Kong	504	496
Korean Peninsula	720	713
Taiwan	350	322
Kuril Islands and Sakhalin	390	297
Southeast Asia and elsewhere	unavailable	137
Total	3,514	3,183

Source: Estimated civilian population in 1945 from Watt (2009, p. 39); Civilian Returnees from Nishizaki (2019, p. 4).

registered domicile in metropolitan Japan" (Watt 2009, pp. 70-76). Nishizaki (2019, p. 23) has studied the assimilation of the repatriates in Hiroshima prefecture and found that 65% of repatriates settled initially in the registered domicile. In a later survey taken in 1956 long after repatriation, 45% were still living in their hometowns, 27% were not living in their original registered domicile but were still living in Hiroshima prefecture, and 27% of repatriates in Hiroshima prefecture had moved there from other prefectures. Thus, it seems that migration between prefectures, at least in the short run, was not common for civilian repatriates, which allows us to account for repatriates by using their location in 1950.

We also need to need to consider whether children repatriated from the colonies would have experience better or worse health conditions than their counterparts who had remained in Japan. This is extremely difficult to establish definitively. One might expect that territories that experienced conflict during the war such as China would have had worse health conditions than territories that were relatively spared. However, conditions after the war also mattered. The experience of repatriation varied wildly across the various colonies. Repatriation from South Korea, Taiwan and China proceeded fairly peacefully and efficiently, but the experience of repatriation was very different in areas controlled by the Soviet Union: repatriation was delayed in these areas and repatriates suffered from violence, sexual assault, hunger and disease (Watt 2009, pp. 41-52).

The plight of Japanese civilians in Manchuria was particularly horrific. The Japanese government had conscripted most of the men in rural Manchuria by the end of the war leaving rural settlements filled with women, children and the elderly. This left these settlements unprotected, and the Japanese government estimates that around 90 thousand Japanese civilians in Manchuria died between the end of the war and repatriation. Rural settlers then had to walk hundreds of miles to cities and stay in camps where there was little food and the sanitary conditions were poor. Women were regularly sexually assaulted by Soviet troops (Itoh 2015, pp. 43-44; Nishizaki 2019, pp. 5-6; Watt 2009, pp. 102-104). Repatriation from Manchuria started in earnest in May 1946, and by October 1946, 1.01 million civilians had been repatriated (Watt 2009, p. 51). However, the health shocks experienced by civilians before being repatriated would have certainly matched the negative health shock of the war.

Thus, it is not entirely clear whether repatriate children were in better or poorer health than their counterparts who experienced the war in Japan. We should also highlight that repatriate children made up a small fraction of the total children in Japan. Unfortunately, there are no precise data showing the age distribution of civilian repatriates. In fact, there are no data in the censuses of 1950, 1955 or 1960 showing the numbers of civilian repatriates living in each prefecture. However, if we assume that one-third of repatriates were children, this would be around one million children. In 1950, there were 34.6 million children under the age of eighteen alive in Japan, suggesting that repatriates were 2.8% of children. Thus, it is highly unlikely that the repatriates would significantly bias our overall findings about the magnitude of the war shock.

However, if repatriates were not proportionally distributed across prefectures, then repatriate children may bias inferences about the shock of the war across prefectures. As mentioned above, there are no data available at the prefecture level that explicitly measures civilian repatriates let alone civilian children repatriates. Figure A6 shows military and civilian repatriates as a percentage of the population of each prefecture and shows that repatriates seemed to be clustered in Kyushu and southwestern Honshu. To ensure that bias from the repatriates is not driving our results, we regress prefecture variation in the war shock on post-war repatriates by including

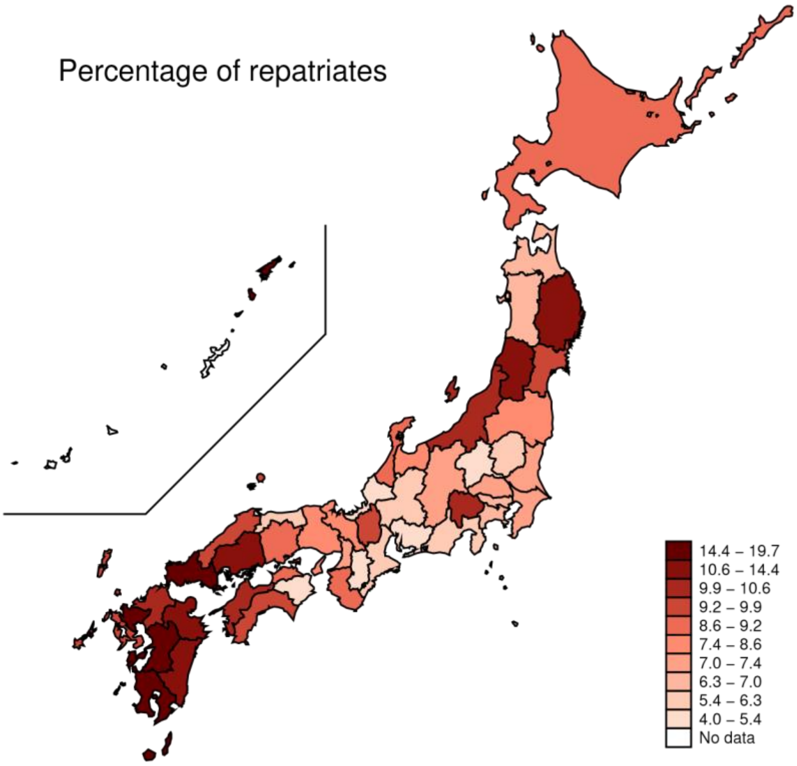


Figure A6. Military and civilian repatriates as a percentage of total population in 1950
Sources: Repatriation Relief Bureau (1950).

the foreign-born share of population in each prefecture (not reported), which would capture children born abroad. In separate regressions (not reported), we also include the percentage of the prefecture population that is both military and civilian repatriates. In all cases, these variables are statistically insignificant with small standardized coefficients.

We treat these results very sceptically for several reasons. First, measurement error in our indicators of repatriate children may be causing attenuation bias in the coefficient. Second, repatriates from different parts of the Japanese colonial empire may have had different underlying health status when arriving back in Japan, and we cannot test for that with our limited data. Finally, it may be that the repatriate children were such a small share of the population of children that they may not have mattered for our estimates. Unfortunately, we cannot distinguish between these three possible explanations.

H. Migration across Prefectures

As mentioned in the main text, another source of measurement error in our prefecturelevel estimates is internal migration. Internal migration was relatively small in the pre-war period, but it accelerated after the war and was generally characterised by migration from rural prefectures to the growing urban centres around Tokyo, Osaka and Aichi prefecture (Fukao et al. 2015, pp. 113-15). This rural to urban migration could cause bias in our estimates in three ways. First, any bias introduced from migrants would be magnified if the migration was particularly strong, so we need to understand the magnitude of internal migration in Japan. Second, if migrant sending prefectures had systematically taller or shorter children than migrant receiving prefectures, then inter-prefecture migration could bias mean heights. Finally, if migrants were positively selected on height relative to non-migrants in their home prefectures as observed with international migrants and internal migrants in England (EscamillaGuerrero and Lopez-Alonso 2019; Humphries and Leunig 2009; Spitzer and Zimran 2018), then migrants might actually increase the mean height in sending regions. We show evidence on each point, which suggests that migration is not a major source of bias in our results.

Starting with the strength of migratory flows across prefectures, we have performed a limited robustness check to try to understand how important the migration of children might have been in the post-war period. Rather than calculating net migration figures for all prefectures in Japan, we have focused on the four prefectures that make up the Tokyo Metropolitan Area (Tokyo, Saitama, Chiba and Kanagawa prefectures), which had the highest rates of net migration in the post-war period (Fukao et al. 2015, p. 114). We calculate net migration of five-year age groups as the survivaladjusted change in cohort size between two successive censuses (each five years apart).

We compute this using the following equation:

$$Mig_{i,a,t}^{a+5,t+5} = 100 \times \left(\left(\frac{P_{i,a+5,t+5}}{Surv_a} - P_{i,a,t} \right) / P_{i,a,t} \right) \quad \text{(E2)}$$

where $Mig_{a,t}^{a+5,t+5}$ is the net migration rate into a prefecture *i* over five years for a cohort between five-year age groups with starting age *a* and *a* + 5 and consequently between censuses in year *t* and *t* + 5. $P_{i,a,t}$ is the population in prefecture *i* in the five-year age group with starting age *a* at the first census in year *t*. $P_{i,a+5,t+5}$ is the population in prefecture *i* in the five-year age group five years later at starting age *a* + 5 at the second census five years later *t* + 5. $Surv_a$ is the mean of the five-year survival probabilities for each one-year age group as in the following equation:

$$Surv_{a,t} = \left(\sum_{a}^{a+4} {}_5L_{a,t} \right) / 5 \quad \text{(E3)}$$

where *a* is the starting age of a five-year age group and ${}_5L_{a,t}$ is the survival probability between exact age *a* and *a* + 5 estimated from the national complete period life table for year *t* in the Human Mortality Database (Human Mortality Database 2019). Of course, it would be better to have prefecture-specific life tables, but these were not readily available.

Figure A7 presents the five-year net migration rates into Tokyo and its surrounding prefectures between 1950 and 1970. The figures are means of the four prefectures weighted by the size of the age group in each prefecture at the first census. We interpret the first column as follows: adjusting for survival, the 5-9 age group across the four prefectures in 1955 was on average 8.0% larger than the 0-4 age group in 1950. Figure A7 shows that while there were some children moving with their families to Tokyo in the post-war period, the greatest net migration was of children in their late teens and early twenties who moved to Tokyo after finishing junior school after age 14. This leads us to two conclusions, suggesting that child migration is unlikely to be a large source of bias in our analysis. First, if child migration was small relative to migration in other age groups even in the wider Tokyo area where migration was greatest, then the scale of child migration was likely small across the entire country. Second, our analysis is principally focused on children in primary and junior school age 14 and under, so the strong migration among individuals above the age of 14 will not lead to measurement error in our prefecture height data.

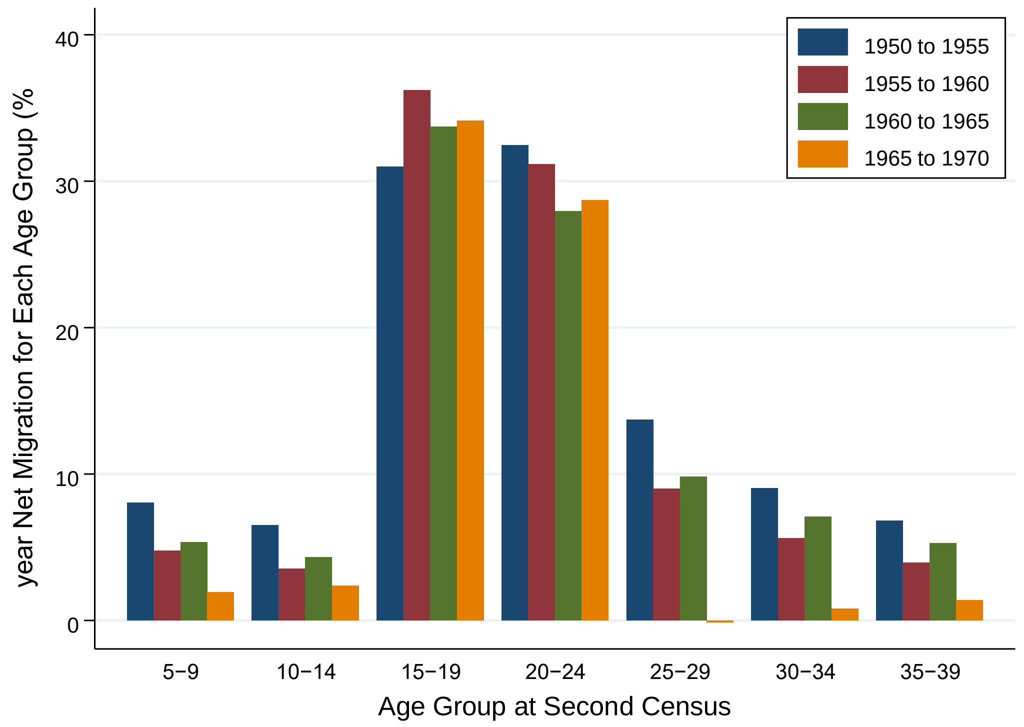


Figure A7. Mean net migration into Tokyo and its surrounding prefectures in the post-war period

Notes: The figure presents the mean net migration into each age group estimated using Equations E2 and E3 for four prefectures including and surrounding Tokyo: Tokyo, Saitama, Chiba and Kanagawa. See text for interpretation.

Sources: Japan Statistical Association (1987, Table 2–13, pp. 130-153).

Even though child migration was relatively small, it is important to check whether there were systematic differences in child height between migrant sending and receiving prefectures. These differences are particularly important at the beginning of the post-war period since migration accelerated after the war. To define migrant receiving prefectures, we rely on the analysis by Fukao et al. (2015, p. 114) showing net migration across the twentieth century. The main migrant receiving prefectures were Tokyo and its surrounding prefecture, Saitama, Chiba and Kanagawa; Osaka and some of its surrounding prefectures, Shiga, Hyogo and Nara; and finally Aichi prefecture. Most of the other prefectures had considerable out-migration in the post-war period. To understand whether there were differences in height between migrant sending and receiving prefectures, we compare the heights of boys and girls at age six at the beginning of the post-war period (1948-50) across the two groups. Figure A8 shows that six-year-olds for both boys and girls were 0.57 cm shorter in migrant sending regions than in receiving areas. Thus, it is possible that migration could have led to bias in both sending and receiving prefectures.

To test the scope of the potential bias, we focus again on the migrant receiving prefectures. The migrant receiving areas acquired larger shares of migrants than the sending areas lost, so they provide a plausible upper bound on the size of the bias created by migration. To measure the potential bias introduced by migrants, we perform a number of back-on-the-envelope calculations from the perspective of a single cohort. To simplify the exercise, we perform the calculations as Z-scores assuming a standard deviation of height of 5.1 cm, taken from the WHO growth reference for boys at age 6.5 de Onis et al. (2007). Keeping the mean of the migrant receiving prefectures at zero across all ages makes it easier to see the bias created by migration into the cohort over time. We assume that the cohort gains 2% of its initial size each year as the cohort ages from 0 to 14 so that by age 14, the population of the cohort is 28% larger than the initial cohort. This is an upper bound estimate of the annual net migration into migrant sending prefectures (see Figure A7). We then calculate a new mean height of children living in migrant receiving prefectures as a weighted average of the mean height of children in migrant receiving prefectures (0) and the height and number of migrants as follows:

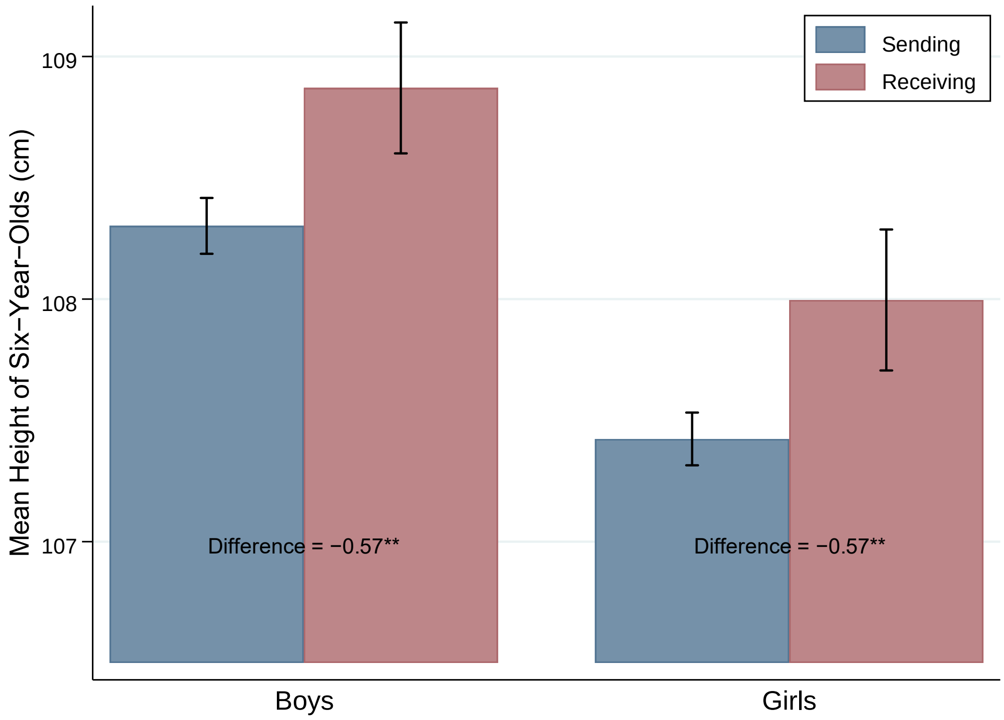


Figure A8. Mean height in migrant sending and receiving prefectures

Notes: Mean heights of six-year-olds measured in 1948-50, representing the initial height of cohorts in migrant sending and receiving prefectures at the beginning of the post-war period. Migrant receiving prefectures are taken from Fukao et al. (2015, p. 114) and are as follows: Saitama, Chiba, Tokyo, Kanagawa, Aichi, Shiga, Osaka, Hyogo and Nara. All other prefectures are considered migrant sending areas. ** indicates that the mean difference was statistically significant at the 1% level on a two-tailed t-test with unequal variances.

Sources: Japanese Prefecture Child Growth Dataset - see Appendix A.1 for sources.

$$height_a = \frac{height_{mig} \times a \times mig_{rate}}{1 + (a \times mig_{rate})} \tag{E4}$$

where $height_a$ is the height Z-score of all children at age a living in migrant receiving prefectures; $height_{mig}$ is the mean height Z-score of migrants, which does not change with age; and mig_{rate} is the annual net migration into migrant receiving prefectures, assumed to be 2%. We allow $height_{mig}$ to vary depending on our assumption about the extent to which migrant children were positively selected on height relative to the children in migrant sending prefectures.

Figure A9a presents the results. If we assume that migrant children were not positively selected, the blue line, we can see that as the receiving areas acquire more and more shorter migrants over time, the mean height of children falls below the original height of children born in migrant receiving areas (0). However, by age 14, this bias only amounts to -0.024 standard deviations less than the original mean, which is a very small difference. If we multiply this figure by the standard deviation of height at age 14.5 (7.8 cm) from the WHO reference, this provides a bias in the mean height of 14-year-olds of 0.19 cm. This bias would affect the SITAR size parameters, but we also need to consider how changing levels of bias across age would affect the intensity parameter. If mean heights are increasingly downwardly biased, then the growth velocity between ages will be biased downward as well because the height increment is smaller than it ought to be. Figure A9b shows the bias in the Z-score increment assuming no positive selection of migrants, and again the scale of the bias is very small. The Z-score increment at age 6 is -0.002, which is a difference in the height increment of 0.01 cm at that age. This bias is miniscule when compared to a typical height increment at age 6 which is above 5cm.

We also adjust our assumption of $height_{mig}$ to allow for positive selection of migrants relative to their counterparts in migrant sending areas. We take the mean height in the migrant sending prefectures to be 0.57 cm lower than migrant receiving areas as in Figure A8. We then generate correlated random variables for the height distribution and a variable capturing the propensity to migrate. We allow for four dif-

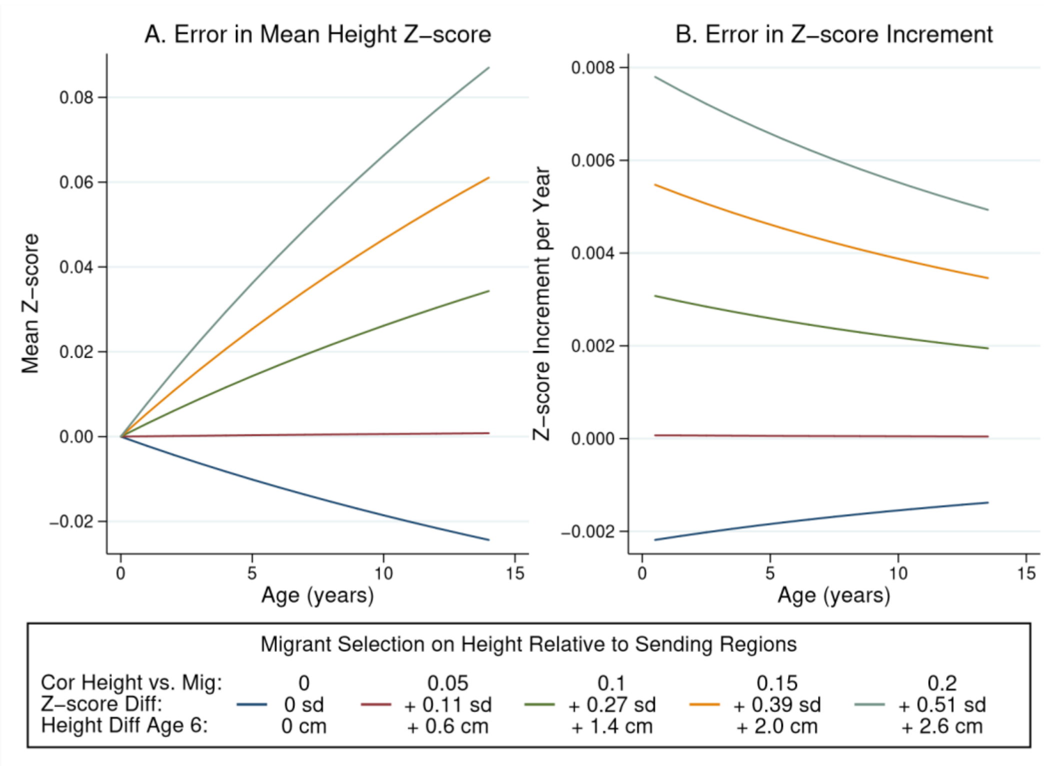


Figure A9. Potential for error in mean heights in migrant receiving prefectures caused by inter-prefecture migration

Notes: Mean Z-score of the cohort before migration is 0, so differences from zero show the bias from the original cohort mean Z-score (A) or original cohort Z-score increment (B). We assume in-migration rates of 2% of the original cohort every year so that by age 14 the cohort is 28% larger than at age 0. The lines are estimated using Equation E4. The different lines show different assumptions about the positive selection of migrants (see text for full details). These differences are very small, so internal migration does not present a major source of bias in our prefecture cohort growth curves.

Sources: Mean difference between migrant sending and receiving prefectures from Japanese Prefecture Child Growth Dataset - see Appendix A.1 for sources.

ferent levels of correlation between the migrant height distribution and the propensity to migrate variable starting with 0.05 and then increase the correlation in 0.05 increments until reaching 0.2. We assume a 1% out-migration rate from migrant sending prefectures. We then take the mean height Z-scores of the top 1% of the generated sample on the propensity to migrate variable. This gives us the mean height Z-score of migrants relative to non-migrants remaining in sending regions. As the correlation between propensity to migrate and height increases, the degree of positive selection of migrants increases to 0.51 standard deviations, which puts the mean height of migrants at age 6 2.6 cm above the height of non-migrants in migrant sending prefectures. This corresponds to a mean height difference of 3.7 cm at age 19. For reference, Spitzer and Zimran (2018) find positive selection at the local level of 0.037 standard deviations or 0.22 cm in adulthood for first-time Italian migrants to the United States. Other studies do find larger results. For instance, Mexican migrants to the United States in the early twentieth century could have been up to 4 cm taller in adulthood than non-migrants (Escamilla-Guerrero and Lopez-Alonso 2019; Kosack and Ward 2014). However, these studies do not have representative comparison groups for non-migrants like Spitzer and Zimran (2018), so they may overestimate migrant selection relative to the mean non-migrant. Thus, the levels of positive selection presented in Figure A8 represent a plausible range of selection of migrants.

Rather than interpret all of the lines in Figure A8, we will focus on the most positively selected migrants. By age 14 under this assumption, the mean height of the cohort in migrant receiving areas was 0.087 standard deviations taller than the initial cohort height. This reflects a bias of 0.68 cm at age 14, which is still very small relative to the height of boys at that age. In addition, the bias to the height increment was also small: 0.0065 standard deviations at age six, which is an upward bias in the height increment of 0.03 cm. Again, this is small compared to a typical height increment at that age above 5 cm. Obviously, this back-of-the-envelope procedure is not a perfect exercise: we have assumed constant migration at 2% for all age groups; we do not know the degree of migrant self-selection; we have not looked at variation within the migrant sending and receiving prefectures. However, given that our upper bound estimates of the effect of internal migration on prefecture mean heights are very small, it is clear that inter-prefecture migration, even when it was at its highest level in the post-war period, is unlikely to create substantial bias in the growth curves that we analyse in our study.

I. Excess Male Mortality during the War

One might be concerned that the lesser effect of the war on girls relative to boys is explained by greater selective culling among girls during the war. If this were the case, surviving girls would be taller than their original cohort, shrinking the war gap in the growth pattern. We test whether this was the case by computing implied mortality rates from the number of children born in each year or alive in 1935 to children in the same cohort alive in the 1940 and 1950 censuses. The implied death rate for each cohort of boys or girls is computed as follows:

$$ {}_nq_x = \frac{P_x - P_{x+n}}{P_x} \tag{E5} $$

where ${}_nq_x$ is the mortality rate for each cohort between exact age x and $x + n$, P_x is the population of the cohort at age x and P_{x+n} is the population of the cohort at age $x+n$. Figure A10 presents the male:female ratio of these implied mortality rates. The blue lines are the sex ratios of death rates calculated between exact age 0 and each cohort's age in the 1940 or 1950 censuses. The red lines are the sex ratios of death rates calculated between each cohort's age in the 1935 census and their age in the 1940 or 1950 censuses.

Figure A10 shows that there was always excess male mortality if we measure cohorts from birth to 1940 or 1950, suggesting that there was not excess female mortality during the war. However, these figures may be largely driven by deaths in the first few years of life. Thus, we can provide alternative figures by measuring mortality of cohorts between the 1935 census and the censuses of 1940 and 1950. Figure A10 shows that some cohorts experienced excess female mortality before the war. However, the mortality across the war shows a strong male bias, suggesting that differential selective culling by sex is unlikely to explain the smaller war shock experienced by girls in our data.

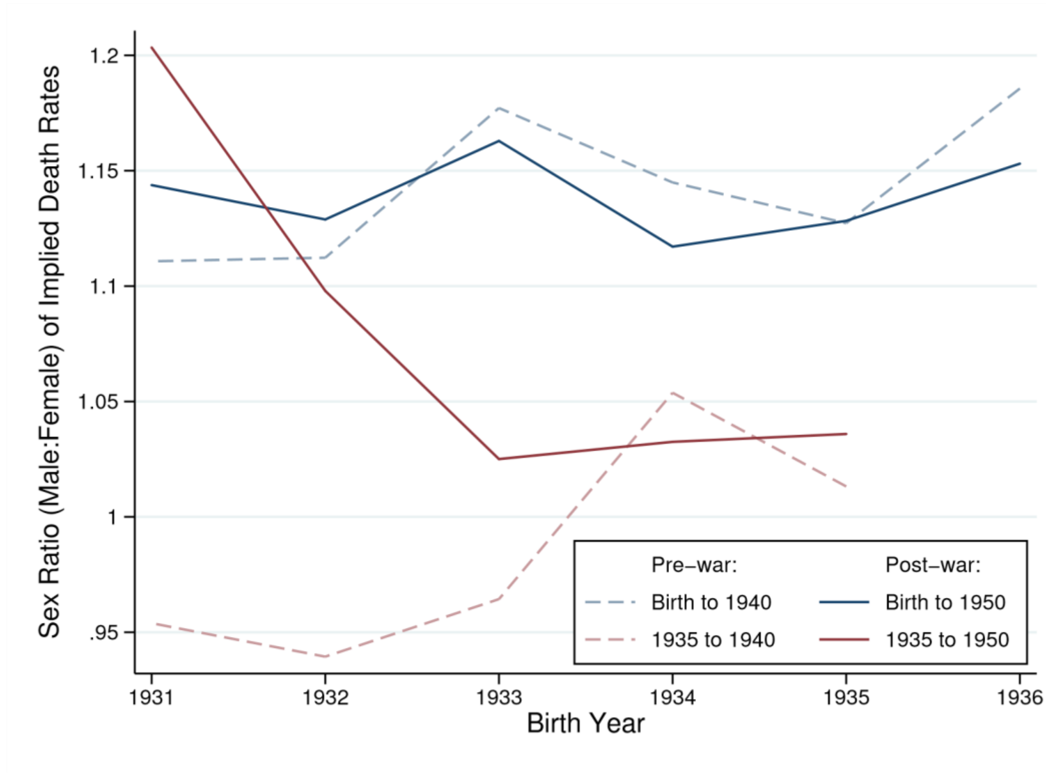


Figure A10. Sex ratio of implied death rates before and after the war

Notes: The implied death rate for each cohort of boys or girls is computed using Equation E5. The blue lines are the sex ratios of death rates calculated between exact age 0 and each cohort's age in the 1940 or 1950 censuses. The red lines are the sex ratios of death rates calculated between each cohort's age in the 1935 census and their age in the 1940 or 1950 censuses. We include these because infant mortality tends to be dominated by male deaths. Thus, we wanted to show that there was excess male mortality across the war even when excluding infant mortality.

Sources: Japan Statistical Association (1987, Tables 2-9, pp. 66-83).

References

Ali, M. A., Lestrel, P. E., and Ohtsuki, F. (2000). Secular trends for takeoff and maximum adolescent growth for eight decades of Japanese cohort data. *American Journal of Human Biology*, 12(5):702–712.

Cole, T. J. (2019). Super imposition by translation and rotation growth curve analysis. https://github.com/statist7/sitar.

Cole, T. J., Donaldson, M. D. C., and Ben-Shlomo, Y. (2010). SITAR–a useful instrument for growth curve analysis. *International Journal of Epidemiology*, 39(6):1558–1566.

Cook, H. T. and Cook, T. F. (1992). *Japan at war: an oral history*. New Press : Distributed by Norton, New York. de Onis, M., Onyango, A. W., Borghi, E., Siyam, A., Nishida, C., and 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.

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

Escamilla-Guerrero, D. and Lopez-Alonso, M. (2019). Self-selection of Mexican migrants in the presence of random shocks - Evidence from the Panic of 1907. *UNU WIDER Working Paper*, pages 1–31.

FAO (2021). New Food Balances. Available at http://www.fao.org/faostat/en/#data/FBS/metadata. Data downloaded on 9 June 2021.

Floud, R., Fogel, R. W., Harris, B., and Hong, S. C. (2011). *The Changing Body: Health, Nutrition, and Human Development in the Western World since 1700*. Cambridge University Press. Cambridge University Press.

Fukao, K., Bassino, J.-P., Makino, T., Paprzycki, R., Settsu, T., Takashima, M., and Tokui, J. (2015). *Regional Inequality and Industrial Structure in Japan: 1874-2008*. Maruzen Publishing Co., Ltd., Tokyo.

Gørgens, T., Meng, X., and Vaithianathan, R. (2012). Stunting and selection effects of famine: A case study of the Great Chinese Famine. *Journal of Development Economics*, 97(1):99– 111.

Gruhl, W. (2007). *Imperial Japan's World War Two, 1931-1945*. Transaction Publishers, New Brunswick, NJ.

Hayami, Y. and Yamada, S. (1970). Kogyoka no shihatsuki niokeru nogyo no seisansei [Agricultural Productivity at the Beginning of Industrialisation]. In Shigeto, K. and Yuzuru, K., editors, *Nihonnogyo to keizaiseicho [Economic Development and Agriculture in Japan]*. The Nippon Agricultural Research Institute, Tokyo.

Henderson, I. (2011). The Firebombing of Tokyo and Other Japanese Cities. In Tanaka, Y., McCormack, T., and Simpson, G., editors, *Beyond victor's justice?: the Tokyo War Crimes Trial revisited*, pages 311–321. Martinus Nijhoff Publishers, Leiden.

Human Mortality Database (2019). University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). Available at www.mortality.org. Data downloaded on 5 June 2019.

Humphries, J. and Leunig, T. (2009). Was Dick Whittington taller than those he left behind? Anthropometric measures, migration and the quality of life in early nineteenth century London. *Explorations in Economic History*, 46(1):120–131.

Itoh, M. (2015). *Japanese War Orphans in Manchuria: Forgotten Victims of World War II*. Palgrave Macmillan, New York.

Japan Statistical Association (1987). *Historical statistics of Japan. Vol.1.* Japan Statistical Association, Tokyo.

Johnston, B. F. (1953). *Japanese Food Management in World War II*. Stanford University Press, Stanford.

Kesztenbaum, L. (2018). Wealth in the trenches. Social class and survival during the Great War. *Working Paper*, pages 1–31.

Kosack, E. and Ward, Z. (2014). Who Crossed the Border? Self-Selection of Mexican Migrants in the Early Twentieth Century. *Journal of Economic History*, 74(4):1015–1044.

Mathias, R. (1999). Women and the War Economy in Japan. In Pauer, E., editor, *Japan's war economy*, pages 65–84. Routledge, London.

Mosk, C. (1996). *Making health work : human growth in modernJapan*. University of California Press, Berkeley.

Nakashima, M. (2011). Ninsanpu to nyuyoji no kenko wo shiensuru techoseido no hensen to koshueiseigyosei no iginitsuite. *Nihonkoeushi*, 58(7):515–525.

Nishizaki, S. (2019). Economic experiences of Japanese civilian repatriates in Hiroshima prefecture, 1945-1956. *LSE Economic History Working Paper*, pages 1–48.

Ohara Institute for Social Research (1964). *Taiheiy¯osens¯oka no r¯od¯oshajy¯otai (The workers in the Pacific War). [in Japanese]*. T¯oy¯okeizaishinp¯osha, Tokyo.

Partner, S. (2007). Daily Life of Civilians in Wartime Japan, 1937-1945. In Lone, S., editor, *Daily Lives of Civilians in Wartime Asia: From the Taiping Rebellion to the Vietnam War*, pages 127–158. Greenwood Press, Westport, Connecticut.

R Core Team (2019). R: A language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Repatriation Relief Bureau (1950). *Hikiageengo no kiroku*. Repatriation Relief Bureau, Tokyo.

Sanderson, F. H. (1987). *Japan's food prospects and policies*. University Microfilms International, Ann Arbor, Mich.

Schneider, E. B. (2020). Sample-Selection Biases and the Historical Growth Pattern of Children. *Social Science History*, 44(3):417–444.

Schneider, E. B. and Ogasawara, K. (2018). Disease and child growth in industrialising Japan: Critical windows and the growth pattern, 1917-39. *Explorations in Economic History*, 69:64–80.

Spitzer, Y. and Zimran, A. (2018). Migrant self-selection: Anthropometric evidence from the mass migration of Italians to the United States, 1907–1925. *Journal of Development Economics*, 134:226–247.

Statistics Bureau of the Cabinet (1939). *1935 Population Census of Japan, volume 1). [in Japanese]*. Statistics Bureau of the Cabinet, Tokyo.

Statistics Bureau, Prime Minister's Office (1948). *1948 Population Census of Japan). [in Japanese]*. Statistics Bureau, Prime Minister's Office, Tokyo.

United States Strategic Bombing Survey (1947). *The Japanese wartime standard of living and utilization of manpower.* Number xi, 146 p. U.S. Government Printing Office, [Washington].

Watanabe, T. (2014). Who became the soldiers? Part 2. *Working Paper*.

Watt, L. (2009). *When Empire Comes Home: Repatriation and Reintegration in Postwar Japan*. Harvard University Press, Cambridge, MA.

Weber, M. (2001). Bombing of Civilians. In Sandler, S., editor, *World War II in the Pacific: an encyclopedia*, pages 113–115. Garland, New York.

Winter, J. M. (1977). Britain's 'Lost Generation' of the First World War. *Population Studies: A Journal of Demography*, 31(3):449–466.

Yamashita, S. H. (2013). The 'Food Problem' of Evacuated Children in Wartime Japan, 1944-45. In Cwiertka, K. J., editor, *Food and War in Mid-Twentieth-Century East Asia*. Routledge, London.

Yoshida, K. (1995). *Poverty in Japan* . Keisoshobyo, Tokyo.