Health Effects of in Utero Exposure to the Dutch Hunger Winter*

Gabriella Conti[1,2], Stavros Poupakis[1], Peter Ekamper[3], Govert Bijwaard[3], Frans van Poppel[3], and L.H. Lumey[4]

[1]*Department of Economics, University College London*
[2]*Institute for Fiscal Studies*
[3]*Netherlands Interdisciplinary Demographic Institute*
[4]*Department of Epidemiology, Mailman School of Public Health, Columbia University*

November, 2018

Abstract

There is a vast literature on the health effects of in utero malnutrition, with the Dutch famine of 1944-1945 being among the most frequently studied adverse shocks. In this paper, we revisit the results of the highly influential 1970s studies of Stein et al. (1972) and Ravelli et al. (1976) who use male military recruits data to study the effects of prenatal famine exposure on mental development and obesity at age 18. Although the famine created a well-defined environment to study the effects of malnutrition, a binary indicator of exposure is mute on the mechanisms through which the famine affected these cohorts at the end of World War II. We enhance the analysis by linking the military recruits data with newly digitalised data on temperature, warfare, caloric and nutrients composition of the diet. While we find effects of in utero exposure on various health outcomes, these are concentrated on those exposed since early gestation and are driven by exposure to warfare and reduction in energy-adjusted protein intake. Moreover, we account for selection using a copula-based approach to relax the, rather restrictive, normality assumption and find evidence of both selection and scarring effects.

Keywords: Health, Fetal Origins Hypothesis, Famine, Prenatal Exposure.
JEL codes: I10, J13.

*We thank participants to the 23rd Annual Conference of the Italian Health Economics Association for their comments. This work was supported by the NIA/NIH grant "Prenatal Under Nutrition and Mortality Through Age 63" (R01AG028593-01A2, PI:LHL) and GC, PE, FvP, GB are co-Is. GC and SP have received funding from European Union's HORIZON 2020 research and innovation programme (grant no. 633595 DynaHEALTH). Authors' contributions: GC and SP: research design, data analysis and paper writing; PE and FvP: provided historical and demographic data and expertise and comments on the analysis and the draft; GB: provided comments on the analysis and the draft; LHL: research design, and provided the recruits and the birth data, famine and medical expertise and comments on the analysis and the draft. Corresponding author: Stavros Poupakis, s.poupakis@ucl.ac.uk, UCL, Drayton House, 30 Gordon St., London WC1H0AX, UK.

1 Introduction

The study of long-term effects of in utero conditions, such as malnutrition due to famine, has been the subject of a long established epidemiological literature (see Lumey et al. (2011) for a review of such studies using famines), with the literature in Economics catching up over the last 10 years (see relevant sections within reviews of Currie and Almond (2011), Almond et al. (2017) and Prinz et al. (2018)). According to Barker's fetal origins or 'fetal programming' hypothesis (Barker, 1990) the mechanisms of this relationship could be a biological pathway, although surprisingly empirical evidence on any specific mechanisms has been lacking until very recently (Tobi et al., 2018).

Prenatal malnutrition is often studied using a famine exposure, usually inflicted by a war or other unforeseen circumstances that affect widely entire cohorts in specific regions.[1] Such food shortcomings enable the study of short- and long-term effects on various outcomes for the members of these cohorts. One of the most well-studied such setting is the Dutch famine of 1944-1945, also known as the Hunger Winter. The famine occurred at the end of World War II (November 1944 - May 1945) and provided an unfortunate, but unique setting to study the effects of malnutrition as an exogenous event, as it was brief, clearly temporally and regionally defined and was not anticipated.

In this paper, we revisit the results of two highly influential studies of the 1970s on the Dutch famine by making several innovations over the previous analyses. These are the studies of Stein et al. (1972), who studied the effect of famine exposure on mental performance, and of Ravelli et al. (1976) who studied the same effect on the likelihood of obesity (defined by Ravelli as a weight/height ratio exceeding 120%), both at age 18 using the same military recruits data. First, we choose carefully the cities that constitute the control group based on common trends and a stable city population before the war. Second, we account for selective fertility by using only those who were already conceived at the start of the famine. Third,

[1] Note that wars and famines are not the only sources of prenatal malnutrition. For example, the Ramadan exposure for Muslim population has been studied in the literature with mixed evidence (Almond and Mazumder, 2011; Jürges, 2015; Majid, 2015)

we perform several robustness and placebo tests by testing the sensitivity of the results to each region inclusion, changing the exposure period at a time of no famine (two years after), and using a triple difference design. We also take good care of the inference by accounting for the moderate number of cities and for multiple hypothesis testing.

Moreover, we extend the analysis further in two important ways. First, we aim to understand possible mechanisms using information on temperature, warfare, changing conditions in the South, and caloric and nutrients composition in the diet. The famine occurred during a winter that was not unusually harsh overall except for a period at the end of January 1945[2] in an environment where the country was at the front of the war with some parts under occupation and some other parts (the South) being liberated. Thus, we use information on the warfare deaths and liberation dates as additional indicators of the impact of the occupation. In view of the timing of the famine during the winter period, we control for temperature – although the long-term effects of hot days in utero are consistent in the emerging literature (Bruckner et al., 2014; Isen et al., 2017; Wilde et al., 2017), results on the effects of cold days are still mixed (Karlsson and Ziebarth, 2017). Second, we account for selective survival using an encompassing definition which includes among the non-survivors stillbirths, neonatal and postneonatal deaths, and mortality until 18 years old (derived from historical records). We achieve this by employing sample selection models using copulas – to deviate from the, inappropriate in this setting, bivariate normality assumption of the standard model – and relying on exclusion restrictions obtained from historical data we linked on occupational and health care pre-war information.

Our findings suggest that exposure to the famine starting prenatally has significant effects compared to starting only postnatally. Most effects across the health outcomes are concentrated on those exposed since early gestation, while several outcomes are significant for exposure since middle and late gestation as well. These results are robust to account

[2]The winter of 1944-45 was assigned a Hellman cold index of 83.3 by the Dutch national meteorological institute (KNMI), which puts it at the 37th position in the ranking of coldest winters in the Netherlands since 1901. There was a period (from January 23-30, 1945) with seven ice days (maximum temperature below 0°C) and a lowest temperature of −13.3°C.

for seasonal trends, and placebo tests confirm the validity of our research design. Exploring the mechanisms of these effects, we find that these gestation effects are mostly driven by exposure to warfare and reduction of proteins in the diet. In addition, we observe evidence of both selection and scarring effects, with most (but not all) impacts reducing when we account for selection. While we find no effects in the main specifications for the ability outcomes, once we account for selection, we are able to detect a significant impact on the likelihood of impaired IQ for those exposed since early and since middle gestation.

The rest of the paper is organised as follows. Section 2 presents the Dutch famine of 1944-1945 and discusses the relevant literature on the relationship between *in utero* malnutrition and adult outcomes. Section 3 describes the data sources and Section 4 describes the econometric specifications used in this study. Section 5 presents the results and Section 6 concludes.

2 Background

2.1 The Dutch famine of 1944-1945

During the end of the 1940-1945 Nazi occupation of the Netherlands in the World War II, food – especially in the big cities – was distributed with rations (which included bread, potatoes, meat, butter and other fats). During the winter of 1944-1945, the occupied part of Netherlands experienced a severe famine as a result of the Nazi blockade, triggered by the Dutch national railways strike to facilitate the Allied liberation efforts. The situation became even worse due to the low temperatures in the winter period, the freezing of the canals, and the military stalemate of the Allied forces with regards to the Dutch front. While throughout the occupation the food rations were maintained around 1,800 calories per day per person, they dropped to below 1,000 calories per day by November 1944 and down to 500 calories per day by April 1945 (consisting mainly of bread and potatoes). The famine ended with the liberation of the occupied part in early May 1945.

This extreme shortage was experienced mainly in the western part of the country. In the North and East of the occupied country, while also experiencing a decline, the shortages were far more limited (see Lumey and van Poppel (1994) for further details), while the South was mostly already liberated. With 3.5 million people (of a total population of 9.3 million) living in the cities of the West – the most affected by the famine – the effects of this shortage were particularly severe. The estimated war related excess deaths vary between 15,000 and 25,000 (see Ekamper et al. (2017) for a discussion of various estimates). While the famine affected the entire population born (or conceived) during or before the famine started, more than 40,000 individuals were exposed *in utero*. Thus, it is possible to identify with some precision critical periods of human development before birth.

2.2 In utero malnutrition and adult health outcomes

The World War II had devastating consequences for the civilian populations across the world and, of course, Europe. The effects of the war can have several channels, physical and mental, such as experiencing hunger, dispossession, absence of the father, and war combats (Kesternich et al., 2014). In another study, looking at self-reported hunger episodes in the Netherlands, Germany and Greece, Van den Berg et al. (2016) find significant effects on height for men (but not for women), with the reduced form effect being a 0.7 cm reduction, and a causal estimate using the propensity to report hunger of a, rather substantial, 3.4 cm decline. However, all these studies used retrospective data on individuals, and while there is evidence that recall of some adverse childhood events is trustworthy (Havari and Mazzonna, 2015), the ability of someone to recall events in early infancy, or in utero, is undoubtedly diminished, and most of these results refer to childhood experiences.

In utero exposure to famine is examined in many studies focusing on the Dutch famine of 1944-1945. The study of Ravelli et al. (1976), using male military recruits at 18 years old, found an effect for those with prenatal exposure on the likelihood of obesity (defined as weight to height ratio being equal or greater than 120 percent). Furthermore, they found

that those exposed in the first half of pregnancy had higher obesity rates, whereas those exposed in the last trimester and first months of life had lower obesity rates. In two cohort studies of men and women followed from birth to late middle age, prenatal famine exposure was associated with increased BMI and waist circumference, specifically in women (Ravelli et al., 1999; Stein et al., 2007). Cognition was also evaluated in two studies of individuals followed from birth to late middle age. In one cohort, the men and women performed worse on a selective attention task (de Rooij et al., 2010) but this finding could not be confirmed in other birth cohorts with a more comprehensive evaluation of cognitive performance (de Groot et al., 2011). Scholte et al. (2015) use register data for a different cohort population in their 50s and find higher hospitalization rates in the years before retirement if exposure occurred in middle or late gestation stage. Besides, they find a significant decrease in the likelihood of being employed at age 55 for those exposed in early gestation, which they interpret as a proxy for cognitive ability later in life. However, Stein et al. (1972) failed to find any association between in utero malnutrition and mental performance at age 18, suggesting that the decline in cognitive ability appears in the period between 18 and 55 (Scholte et al., 2015). Portrait et al. (2017) look at height using a much broader group of cohorts from the Longitudinal Aging Study Amsterdam. Spanning from those born in 1930 to those born in 1945, they use place of birth as exposure (urban West as exposed; rural West and Nort/East as unexposed) and construct groups of exposure at certain age bands. They find a significant effect of 4 cm for the males exposed between gestation and age 2 compared to the unexposed, while no significant effect is found for any older exposure group.

Moving away from the Netherlands, Jürges (2013) finds effects of the German famine that occurred after the WWII on both education and labour market outcomes, which were stronger for early pregnancy exposure compared to late pregnancy exposure. Neelsen and Stratmann (2011) examine the effect of the Greek 1941-1942 famine using census data and compare those exposed in utero, in infancy and at their second year of life. They find the strongest effect on literacy for those exposed in infancy, whereas for education, the effects

for infancy and in utero exposure are both stronger than the older exposure. In addition, these effects are stronger for the urban-born cohorts compared to the rural-born cohorts, suggesting, like the other studies, that the famine was mostly experienced in the urban centres.[3]

In another study, Stanner et al. (1997) examine the effects of in utero malnutrition caused by prenatal exposure to the siege of Leningrad in 1941-1944. In contrast to the other studies, they find no association between in utero malnutrition and later in life health outcomes, but the study was small and underpowered. In the case of the Netherlands, the famine was followed by a period of prosperity, whereas in the USSR, economic growth was limited and food shortages continued for an extended period. This is consistent with the findings in Kesternich et al. (2015) for Germany, looking at the effect of those experienced hunger on food expenditures, suggesting higher consumption for the exposed cohorts. Moreover, they find that the effects were concentrated among lower-income adults, showing the behavioural pathway as a mechanism.

In the case of China, mixed results have been reported of the long term health impact of the Great Leap Forward famine of 1959-1963 on diabetes, blood pressure, heart disease, obesity, and other outcomes. Nearly all published studies report a negative impact of the famine, but these results could not be confirmed in a meta-analysis after appropriate age-adjustment of the study populations (Li and Lumey, 2017). This suggests that in most China studies the measures of famine severity may have been too imprecise to detect a long term impact and need significant improvement. Evidence from France shows that malnutrition caused by an income shock in the late 19th century did not result in worse adult health, although there was a significant effect on height (Banerjee et al., 2010). A possible

[3]Another famine, also widely studied, occurred in China and lasted for three years between 1959 and 1961 as a result of various economic and social reforms by the government in the preceding period, known as the Great Leap Forward. Chen and Zhou (2007) find an average effect of 3 cm on height for the cohorts born in 1959, 1960 and 1961 which experienced the 1959-1961 famine in China. However, contrary to the studies discussed above, the effects are concentrated in the rural populations, and for the urban-born cohorts there were no effects. This is due to the nature of the famine's origin and political and social structure of China at the time.

explanation suggested by the authors is that eventually the health of the affected cohorts was protected by other quality of life improvement, such as public health developments.

One problem in such famine studies, which might have contributed to these mixed results, is that these analyses rely on survivors only. However, the exposed cohort is subject to two, in the opposite direction, effects. Those who survive are 'scarred' and thus have worse health later in life, however the malnutrition raises the survival threshold which affects those in the bottom of the distribution and results in a population with better health. Bozzoli et al. (2009) argue that an environment with high infant mortality favours the selection effects to dominate, whereas in settings with better conditions and lower mortality scarring is more evident. Another limitation of such studies is that they measure exposure to famine areas and periods, without directly observing the individual experience of malnutrition. For example, family members might respond to such harsh conditions and try to give the pregnant woman a larger share of food rations. Moreover, most of the famine studies discussed here examine different windows of exposure, without looking at trimester-specific differential effects (apart from Scholte et al. (2015) as mentioned earlier)[4] and have not attempted to account for selection (see Lee (2009); Gørgens et al. (2012); Liu et al. (2015) for some examples that attempt to account for selection).

3 Data

3.1 Military Recruits Data

The primary data source in our analysis is the military recruits data also used in the studies of Stein et al. (1972); Ravelli et al. (1976) in the 1970s. This includes all men born in the Netherlands in 1944-1947 living in the Netherlands at age 18 and who were examined in the military centers. In addition to the medical examinations, the data contains date and

[4]For example, Almond and Mazumder (2011) examine trimester effects for malnutrition due to Ramadan exposure, finding stronger effects for those in the first and second trimester.

place of birth, and several demographic information, such as father's occupational status and family size.

To account for selective fertility, we only include individuals who were already conceived at the start of the famine. In other words, we exclude from the analysis everyone born after July 1945, unlike existing practice in the literature. Given that the famine lasted until the liberation on early May 1945, we define three treatment groups: exposure starting in the first trimester (born May-July 1945), exposure starting in the second trimester (born February-April 1945), and exposure starting in the third trimester (born November 1944-January 1945). The control group includes those exposed only postnatally in the first months of life (born May-October 1944).[5] Figure 1 shows the time periods corresponding to each of the three treatment groups and the control group. The famine period is also presented by the red vertical lines. As the extreme famine period lasted 6 months, individuals exposed in the first trimester were born after liberation and were not exposed in the third trimester.

We are well aware of the problems with defining exposure by counting backwards from the date of birth, as opposed to counting forwards from the date of conception (Currie and Rossin-Slater, 2013). The impact of the famine on gestation length was minimal (Stein and Susser, 1975) and given that we do not observe any significant effects on gestational age in our sample either (last column of Table 7 – and for girls, Table C4 in the Appendix) we do not consider this to be an issue. Moreover, we can rule out potential measurement error in the date of birth. In order to show that, we plot the distribution of the calendar day of birth for all the recruits in the years 1944-1947 (Figure D1), for those in our analytical sample (Figure D2), where for the latter we also plot it by city (Figure D3) and by month (Figure D4). It is clear from the graphs that there is no evidence indicating measurement error in the form of heaping.[6]

We took great care in selecting the cities included in the analysis, contrary to the literature

[5] See Lumey et al. (2007) for various alternative definitions of treatment and control cohorts in the context of the Dutch Famine.

[6] See for example Beach et al. (2018), where age heaping biases the estimates of the 1918 influenza exposure in the US by a factor of 2.

that usually places no criteria on the inclusion – and even if it does, this is done without any examination of the parallel trends assumption. Thus, we have selected our control group of cities through the following steps. Given that the famine has historically affected more the cities, we have first restricted our sample of interest to the 46 municipalities with a population greater than 25,000 inhabitants on January 1, 1940. We have further excluded from this group, using data from the Historical Ecological Databank of the Netherlands (Boonstra, 2016): (a) 4 municipalities where the majority of the population was not living in the largest place in the municipality (i.e. to be classified as rural not urban from their population dispersion pattern); (b) 13 municipalities where the population underwent major changes in size since 1930 (either increased more than 50% over the decade or had a decrease in the population after the onset of the war). Table 1 summarises the list of cities and their allocation into the Famine and the control areas.

We have finally tested and failed to reject that the remaining 29 cities follow the same trends before the start of WWII in the following outcomes: postnatal mortality rate, crude birth rate, crude death rate, crude marriage rate, infant mortality rate, mortality rate 1-14 years old, mortality rate 15-39 years old, mortality rate 40-59 years old, mortality rate 60+ years old, using data compiled from published monthly statistics by city from the Dutch Central Bureau of Statistics (CBS, 1935-1947). In Figures A1 to A9 in the Appendix, we present several graphs showing the trends for these outcomes. A formal Wald test is included at the right of each plot, testing whether the trends are parallel (by testing jointly if the slopes are equal)[7] for each group of cities. Hence, we are working with a consistently defined group of 7 treatment and 22 control cities (in the West, North-East, and South), for which we cannot reject the null hypothesis that they were on parallel trends in the years before WWII along several health-related dimensions. In contrast, in the case of the non-selected cities we reject the hypothesis of parallel trends.

[7]The estimated regression is $y = \beta_0 + \beta_1 year + \beta_{2,k} city_k + \beta_{3,k} city_k \times year$ where y is the corresponding outcome, $year$ takes values from 1935 to 1939, and $city_k$ is a dummy variables for each city $k = 1, ..., K$ for the K cities included in each group. Then, the Wald test is performed on whether all $\beta_{3,k}$ are equal to 0.

3.2 Additional Administrative Data

Using the date and city of birth of the military recruits data we were able to merge our data with newly digitalized information on calories and nutrients composition of the diet, temperature, warfare and the progressive liberation of the South to study the channels. First, we use the information on calories and nutrients composition of the diet from the official war information on the rations (Departement van Landbouw en Visserij, 1946). This information existed in weekly level for the West region and monthly level for the control regions (North-East and South). Thus, we match the examined cohorts at the appropriate level depending on the place of birth.

Figure 2 shows the caloric intakes and the shares of protein, fat and carbohydrates for each trimester by month of birth. The shares of protein and fat in the diet were calculated using the standard formulas $(Protein\ (grams) \times 4)/Calories\ (kcals)$ and $Fat\ share = 9 \times Fat(grams)/Calories(kcals)$, whereas for carbohydrates the two shares were subtracted from 100 $(Carbohydrates\ share = 100 - (Protein\ share + Fat\ share))$. However, the macronutrient we found to be the most significant is the protein and is the one included in the results throughout. The Figure shows how the drop of calories during the famine was accompanied by a drop in the protein share in the West.

Second, we use archived temperature information from the period (KNMI, 2018). At the time, only three meteorological stations were operating in the Netherlands. Thus, we use Inverse Distance Weighting (Pebesma, 2004) to interpolate the temperatures for each city in our sample. We do this for each month, for all the period covered in the analysis. For illustrative purposes, Figure 3 demonstrates an example of how this looks for two months, December 1944 and May 1945. The red squares are the meteorological stations, and the black dots are the cities in our study. The heatmap shows the predicted temperature across the Netherlands, as indicated in the legend, with each city receiving a temperature depending on its location.

Third, as indicators of the impact of the occupation on the population, we have available

estimates of civilian deaths from war related causes. Deaths due to warfare are all deaths classified with code "197 - Deaths of civilians due to operations of war" (within the main category "XVII Violent or Accidental Deaths") according to the International List of Causes of Death, Revision 5 (ICD-5) of 1938 (CBS, 1935-1947), to better separate direct and indirect mortality from the war (Jewell et al., 2018). Part of the number of deaths of civilians due to operations of war in the last years of the war were the result of incidental bombing by the Allied Forces. During the occupation, the Allied Forces carried out around 600 bomb attacks on Dutch territory (Korthals Altes, 1984; NIOD, 2018), aimed at strategic goals, such as ports, bridges, and railways. Most bombings caused no or relatively few deaths among the civilian population. However, in a few cases the bombings were errors or the intended goal was missed, resulting in unintentionally large numbers of civilian casualties (e.g. Nijmegen, February 22nd, 1944, nearly 800 victims, and The Hague, March 3rd, 1945, around 550 victims). Figure 4 shows the number of deaths of civilians due to operations of war per month for each city for the period May 1944 to July 1945. Each plot shows the civilian deaths for each trimester. The warfare affected cities in all three regions (West, North-East and South) within our study period for all three trimesters, thus allowing us to explore any trimester-specific effects.

Fourth, to account for the changing conditions in the South after liberation from the Nazi by the Allied forces, for each birth date, we use the weeks since liberation for the South region. The liberation dates for each city were: 14 September 1944 (Maastricht), 17 September 1944 (Heerlen), 18 September 1944 (Eindhoven), 23 September 1944 (Helmond), 5 October 1944 (Kerkrade), 27 October 1944 (Den Bosch and Tilburg), 29 October 1944 (Breda), 30 October 1944 (Bergen and Roosendaal), 1 March 1945 (Venlo).

Fifth, to account for selective survival and calculate survival rates of recruits up to age 18 we used additional information on monthly numbers of births per city for the period 1944-1947. Our primary source is the monthly city statistics published by Central Bureau of Statistics (1944-1948). Since monthly data for the period July 1944 - December 1945 are

missing from this source, we used additional published monthly municipal data for Amsterdam, Rotterdam, The Hague, and Utrecht (Gemeente Amsterdam, 1946; Gemeentelijk Bureau voor de Statistiek Rotterdam, 1946; Gemeente Den Haag , 1946; Gemeente Utrecht, 1946) and estimated the monthly patterns for all other cities by combining the regional monthly patterns published in Stein et al. (1975) and the annual city data from Central Bureau of Statistics (1944-1948).

Table 2 shows the descriptive statistics of the additional information included in the sample. In the analysis, we transform both calories and protein shares to their negative values so their coefficients represent the effects of a decrease in them. Additionally, we convert the calories in thousands for ease of the coefficients readability. For the warfare variables, we are using a log(x+1) transformation, due to the high skewness and to avoid loosing the observations with zero deaths.

Finally, in the models accounting for selection we use city-level pre-war information, using the following variables: medical staff per 1,000 of population (*Medical1930*), the share of inhabitants of the largest place in the municipality over the total number of inhabitants (*Largest1930*) and the share of people employed in agriculture (*Agriculture1930*), all measured at 1930 and taken from the Historical Ecological Databank (HED) of the Netherlands (Boonstra, 2016).

3.3 Birth data

The data used in the birth outcomes analysis are hospital records from five cities: Amsterdam, Rotterdam and Leiden (West), Groningen (North) and Heerlen (South). Every city has data from one hospital, while Amsterdam includes two hospitals (Stein and Susser, 1975). In contrast to the military recruits data, the birth data contains information on both boys and girls. Although at the time less than half of the births were taking place in hospitals, the admission procedures remained unchanged for the duration of the time period in our study (Stein and Susser, 1975). The total sample we used is all births that occurred between

May 1944 and July 1945. For each birth record, we have the gender of the newborn, weight, length, head circumference, along with placenta's weight, gestational age, and mother's age at the time of birth.

4 Econometric Framework

In order to assess the effect of the famine in the affected cities in the West we use a difference-in-differences design. This enables us to account for regional effects on estimating the effects on the various health outcomes at age 18. The main estimating equation is:

$$\begin{aligned} y_{ijm}^{K} =& \beta_1 WestF_{ijm} \times Late_{ijm} + \beta_2 WestF_{ijm} \times Middle_{ijm} \\ &+ \beta_3 WestF_{ijm} \times Early_{ijm} + \beta X_{ijm} + City_j + (Birth\ Month)_m + \epsilon_{ijm} \end{aligned} \tag{1}$$

where y_{ijm}^{K} is one of the following outcomes K for person i, born in city j in month m: *Height* (a frequently used indicator of adult health (Deaton, 2007)), *Weight*, *BMI* (kg/m^2), *overweight* ($BMI \geq 25$), *Obese Ravelli* (weight/height ratio $\geq 120\%$ – definition as in Ravelli et al. (1976) for comparison purposes), *Underweight* ($BMI < 18.5$), *Chest Height Ratio* (a measure of abdominal fat, used for example in Costa (2004)), *IQ Impaired* (ICD-6 325, primary or secondary diagnosis) and *Mental Deficiency Stein* (Stein et al. (1972) definition not including borderline and other/unspecified).

The $WestF_{ijm}$ is a dummy variable indicating those born in the West Famine region. The dummy variable $Late_{ijm}$ refers to those who would be exposed to the famine in the third trimester and postnatally (born between November 1944 and January 1945), the $Middle_{ijm}$ refers to those who would be exposed from the second trimester, and includes late (born between February 1945 and April 1945), and the $Early_{ijm}$ refers to those who would be exposed from the first trimester, and includes middle and some late (born between May 1945 and July 1945). The time periods are best explained in Figure 1. All these variables are defined in terms of month of birth, thus full exposure is defined by the interaction with

the $WestF_{ijm}$ variable. The control group consists of the selected cities in Non-Famine West, North-East and South regions (see Table 1). The examined population is that born between May 1944 and July 1945, thus the reference group is those born in the first six months of that period (born between May 1944 and October 1944). The choice of this control group was motivated by the fact that these cohorts are exposed to the famine period only postnatally. We do not include those conceived during or after the famine, in view of the significant reduction in conceptions and births related to the famine (Stein et al., 1975). Finally, $City_j$ and $(Month\ of\ birth)_m$ are entering the specification as fixed effects, along with a vector of controls X_{ijm} ($Father's\ Occupation\ Status$, $Older\ Brothers$, $Birth\ Order$, and $Religion$).[8]

The main parameters of interest are β_1, β_2 and β_3, the interaction terms of $WestF_{ijm}$ with the three exposure dummies. In all estimations we use clustered standard errors.[9] Given the relatively small number of cities, we follow the recommendation in Cameron et al. (2008) and compute the Wild cluster bootstrap standard errors and report the corresponding p-values (Roodman et al., 2018). To account for multiple hypothesis testing, we used the Romano and Wolf (2005, 2016) procedure and re-calculated the p-values for all the variables of interest. The results reported here are based on 5,000 replications for the Wild cluster bootstrap and on 1,000 replications for the Romano Wolf, although we experimented with different number of replications and the results are unchanged.

Moreover, we use in a placebo analysis the same estimating equation as in Equation 1 to test that the results are not driven by seasonality, using the cohorts born two years later (born between May 1946 and July 1947). This allows us to examine whether any

[8]Table C1 in the Appendix presents descriptive statistics of these variables. In our analytical sample, 28% of the cohort members had a father working as shop assistant, whereas 27% were in clerical occupation. 35% were first-born, 28% second-born and 17% third-born, with 45% having at least one older brother. Catholics constituted 40% of the sample and Protestants 28%.

[9]Note that we have the entire population of male births in the study period. Outcome differences between subpopulations defined by some attributes should simply be estimates which would be known with certainty (i.e., the standard errors should be zero). By reporting statistical significance nevertheless, we implicitly assume that there is a superpopulation from which the population is randomly sampled. As with samples drawn from the population, uncertainty in our case emerges from the unobservability of the superpopulation – this may be, for example, future populations, in which the uncertainty would emerge from year-to-year variation (for most recent discussion on the issue see Abadie et al. (2017)).

significant differences found in the previous specification are driven by other systematic differences between the defined groups, other than the famine exposure. Next, we perform a triple difference (difference-in-difference-in-differences) analysis by estimating Equation 2 to explicitly account for seasonality. The variables $WestF$ and $Late$, $Middle$ and $Early$ are defined as before. The variable War is a dummy variable taking value 1 for those born during the war period (May 1944 - July 1945) and 0 for those born two years after (May 1946 - July 1947). The coefficients of interest in Equation 2 are those of the triple interactions: β_9, β_{10} and β_{11}.

$$\begin{aligned} y_{ijm}^K = {} & \beta_1 WestF_{ijm} \times Late_{ijm} + \beta_2 WestF_{ijm} \times Middle_{ijm} + \beta_3 WestF_{ijm} \times Early_{ijm} \\ & + \beta_4 War_{ijm} + \beta_5 WestF_{ijm} \times War_{ijm} + \beta_6 Late_{ijm} \times War_{ijm} \\ & + \beta_7 Middle_{ijm} \times War_{ijm} + \beta_8 Early_{ijm} \times War_{ijm} \\ & + \beta_9 WestF_{ijm} \times Late_{ijm} \times War_{ijm} + \beta_{10} WestF_{ijm} \times Middle_{ijm} \times War_{ijm} \\ & + \beta_{11} WestF_{ijm} \times Early_{ijm} \times War_{ijm} \\ & + \beta X_{ijm} + City_j + (Birth\ Month)_m + \epsilon_{ijm} \end{aligned} \tag{2}$$

For the birth outcomes, we use a similar difference-in-differences specification on the cohorts born in the six hospitals in the birth data. The main estimating equation is:

$$\begin{aligned} y_{ijm}^K = {} & \beta_1 West_{ijm} \times Late_{ijm} + \beta_2 West_{ijm} \times Middle_{ijm} + \beta_3 West_{ijm} \times Early_{ijm} \\ & + \beta X_{ijm} + Hospital_j + (Birth\ Month)_m + \epsilon_{ijm} \end{aligned} \tag{3}$$

where y_{ij}^K is one of the following birth outcomes for child i born in city j: *Birth Weight*, *Low Birth Weight* (birth weight $< 2,500$ grams), *Birth Length*, *Head Circumference*, and *Placenta Weight* and being male (*Sex Ratio*). $Late_{ij}$ is exposure to the famine period starting on the third trimester (born between November 1944 and January 1945), $Middle_{ij}$ starting on the second trimester (born between February 1945 and April 1945), and $Early_{ij}$ is starting on the first trimester (born between May 1945 and July 1945). Those born from May 1944 until

October 1944 are defined as having postnatal exposure. Thus, the total period of births we include in the analysis is from May 1944 to July 1945. The $West_{ij}$ indicates those born in Amsterdam (two hospitals), Leiden and Rotterdam, whereas the controls are Groningen and Heerlen. The $Hospital_j$ and $(Birth\ Month)_m$ are the hospital and month of birth Fixed Effects and the vector X_{ij} includes as a control the mother's age.

Finally, we assess the robustness of the results after accounting for selection using the city-level variables on pre-war information (*Medical1930*, *Largest1930*, and *Agriculture1930*) in the selection equation. Regarding the estimation of the selection models, we depart from the bivariate normality assumption on the error terms by using copulas.[10] We implement this by performing a series of estimations using a variety of different copulas (Gaussian, FGM, Plackett, Clayton, AMH, Frank, Gumbel, and Joe) and choose the one that fits best based on the Bayesian Information Criterion (BIC). The estimating outcome equation is

$$\begin{aligned} y^K_{ijm} =& \beta_1 WestF_{ijm} \times Late_{ijm} + \beta_2 WestF_{ijm} \times Middle_{ijm} + \beta_3 WestF_{ijm} \times Early_{ijm} \\ &+ \beta X_{ijm} + City_j + (Birth\ Month)_m + \epsilon_{ijm} \end{aligned} \tag{4}$$

and the selection equation is

$$\begin{aligned} s^K_{ijm} =& \gamma_1 WestF_{ijm} \times Late_{ijm} + \gamma_2 WestF_{ijm} \times Middle_{ijm} + \gamma_3 WestF_{ijm} \times Early_{ijm} \\ &+ \gamma_4 Medical1930 + \gamma_5 Largest1930 + \gamma_6 Agriculture1930 \\ &+ City_j + (Birth\ Month)_m + u_{ijm} \end{aligned} \tag{5}$$

where s^K_{ijm} indicates whether the individual is alive at age 18, and the error terms ϵ_{ijm} and u_{ijm} have a joint distribution based on one of the proposed copulas. For example, in the case of a Gaussian copula, then the joint distribution is a bivariate Normal distribution, common to many applications of sample selection models.

In addition, we use Inverse Probability Weighting (IPW) as an alternative way to account

[10] While such implementation can be traced as back as in Lee (1983), the use of copulas in this context has become explicit since Smith (2003).

for selection, although more restrictive in the sense that it is assumed that selection is based only on observables. The weights are calculated for each city and each birth month, by taking the inverse of the survival rate at age 18, and the estimating equation is as in Equation 1 but estimated using weighted least squares.

5 Results

5.1 Age 18 Outcomes

Our baseline results, based on the difference-in-differences specification discussed previously, are presented in Table 3. All the coefficients reported are from linear regression models, so that they can be interpreted directly as marginal effects. The individuals exposed since early gestation (i.e. trimesters 1 and 2 only, as the famine period was limited to 6 months) have a significantly higher weight (652 grams w.r.t a control mean of 67.6 kg), BMI (0.2 w.r.t a control mean of 21.5), obesity (Ravelli definition, 0.6 p.p. w.r.t a control mean of 1.4%), underweight (reduction by 1.3 p.p. w.r.t a control mean of 5.6%) and abdominal obesity (0.010 increase in chest-height ratio w.r.t a control mean of 0.492); the individuals exposed since mid-gestation (i.e. trimesters 2 and 3) have only a significant increase in abdominal obesity (0.008 increase in chest-height ratio w.r.t a control mean of 0.492) and in mental deficiency (Stein definition, 1 p.p. increase w.r.t. a control mean of 3%). The individuals exposed since late gestation (i.e. trimester 3 only) have significantly *lower* weight, overweight and obesity (Ravelli definition) as compared to those with exclusive postnatal exposure. These effects are robust to controlling for multiple hypothesis testing by using the Romano and Wolf (2005) step-down method. No impacts for any exposure group are detected for height and IQ impaired.

To examine whether any significant differences found in Table 3 are driven by other systematic differences between the specified groups, other than the defined difference in famine exposure, we present in Table 4 the results from a placebo test analysis faking the

famine exposure. This is achieved by estimating the same equation as before using the cohorts born two years later (born between May 1946 and July 1947) and defining the exposure groups accordingly. As expected, we find no significant impacts in the placebo regressions for any of the outcomes. Since the placebo test confirms the identification strategy in our chosen analytical sample, we are extending the placebo test to the sample used by Ravelli et al. (1976), which doubles in size by including rural areas (presented in the Appendix Table C2). The placebo analysis fails, as we find significant differences, mostly in the opposite direction, for weight, BMI, underweight and IQ impaired. This further confirms the importance of carefully selecting the control group.

Then, we account for different seasonal trends by re-estimating our baseline model with a triple difference specification (also using the individuals born in 1946-1947 used in the placebo analysis). Table 5 presents the results of this estimation, where only the triple interaction terms $WestF \times Late \times War$, $WestF \times Middle \times War$, and $WestF \times Early \times War$ (as discussed in Equation 2) are shown. Our baseline results are confirmed, with the triple interaction for exposure since the first trimester increasing by almost 50% for weight, by 25% for BMI, and by 40% for being underweight and while for chest/height ratio were reduced as compared to the simple Difference-in-Differences specification, it remains significant at the 10% level. The exposure since the second trimester is again only significant for chest/height ratio, but only at the 10% level, whereas for exposure since the third trimester the triple interaction was not significant for any of the outcomes.

Lastly, we repeated the estimations in Tables 3 and 5 by excluding one control region at a time to test whether the results are sensitive to the choice of the comparison group. Results are presented in Figures B1 and B2 in the Appendix. The plots present the interaction term estimates as in the Tables 3 and 5, along with the 95% Confidence Interval. All plots include a vertical line at zero to ease examination of significance.

Then, we focus on understanding the mechanisms through which being in utero during the Dutch Hunger Winter led to adverse outcomes at 18. Since we do not find any effects on

height and mental performance in the main specifications, we exclude these outcomes from this part of analysis and focus on the rest. Table 6 shows the results from the specifications using the additional information on caloric and protein intake, temperature, warfare and liberation, merged to the military recruits data, with the Difference-in-Differences interaction terms.[11] Thus, we also present the coefficients for warfare experience by trimester (represented as dummies) and the coefficients for calories (in thousands) and protein share by trimester, both transformed into negative values so the effects can be interpreted as reductions in each. Temperature and liberation weeks are included as control variables and are omitted from the tables (as the other controls on father's occupation, older brothers, birth order and religion).

Our results show that controlling for the channels explains all the effect of being born during the famine (i.e. the coefficients become insignificant – apart from BMI since first trimester and Ravelli-obesity since second trimester, though they remain significant only at 10%). In particular, the impacts on weight, BMI, and overweight for those exposed since early gestation are driven by a combination of exposure to warfare in the third trimester and reduction of proteins in the diet (irrespective of the caloric value of the diet) in the third trimester; the impacts on underweight are also driven by exposure to warfare in the third trimester, but coupled with reduced caloric intake in the third trimester (i.e., irrespective of the composition of the diet).

5.2 Birth Outcomes

While the previous section is focused on the impacts of famine 18 years after the exposure, it is important to look at immediate effects because often the developmental origins of health and disease literature talks about latency effects. For example, Schwandt (2017) shows that while maternal influenza has damaging effects at birth, this can be through various different mechanisms that might appear only later – and not at birth.

[11] Table C3 in the Appendix presents the same specification without the Difference-in-Differences interaction terms.

Thus, we further examine the immediate effects of the famine exposure by investigating various anthropometrics at birth. For this, as described in Section 3.3, we use an additional dataset collected on selected birth clinics in three treated and two control cities. We present here the results using the male birth data, for consistency with the recruits. The results in Table 7 show that the individuals exposed since the first trimester of gestation have significantly lower placenta weight at birth (44 g w.r.t a control mean of 633 g).[12] Those exposed since the second trimester have significantly lower placenta weight and birth weight (40 g w.r.t. a control mean of 633 g and 169 g w.r.t. a control mean of 3,414 g, respectively). These findings differ from previous reporting on placental weights in this population (Lumey, 1998). Despite the significant negative effect on birth weight, there was no impact on low birth weight. An effect of 0.9cm is found on head circumference for those exposed since third trimester w.r.t. a control mean 38.9cm, however this was only significant at the 10% significance level. No effects are found on birth length. None of the effects are robust to controlling for multiple hypothesis testing. These results indicate that the adverse effects during utero might be latent at birth and appear only at later stages of life.

In addition, we find no effect on gestational age (in column (6) in Table 7) and on sex ratio (Table C5 in Appendix) for any of the exposure groups.[13] The fact that there are no effects on sex ratio and gestational age, together with the lack of detected measurement error in birth dates, validates our strategy of using the date of birth to identify the date of conception and so the exposure by trimester. Results are also presented in the Appendix for the girls of the birth sample, in Table C4.

5.3 Accounting for Selection

Our analysis is based on military recruits data collected when the respondents were 18 years old. This is conditional on being alive (and in the country), thus we observe only

[12] Since not all outcomes were measured for all individuals, sample sizes differ across the models. The results are robust to using a balanced sample (available upon request).

[13] The zero effect on sex ratio for the Dutch cohorts also shown and discussed in Cramer and Lumey (2010).

survivors up until that age. The cohorts born in the cities in the West faced much worse conditions than their counterparts in the rest of the country, hence we expect them to be less likely to survive. Using the additional information on births and deaths for that period, we are able to calculate survival rates for each city of birth. Figure 5 shows the proportion of survivals for the selected cities by cohort groups as defined previously. It is clear from the graphs that the cities in the West faced much higher mortality than the rest (including the West non-famine group), especially Amsterdam, The Hague, and Rotterdam, which were the biggest.

Table 8 shows the results from the difference-in-differences linear probability model on the likelihood of survival at age 18. The individuals exposed since early gestation have 4.8 p.p. lower probability of surviving until age 18 (w.r.t. a baseline mean of 85%), and those exposed since mid-gestation have 3.1 p.p. lower probability (but not significant with the Wild cluster bootstrap). For those exposed since late gestation, the effect was 1.5 p.p. although it was not significant at any significance level. Thus, we can see an almost linear increment on the effect on survival for every three months of exposure to the famine.

We experimented with a series of Monte Carlo simulations to explore alternative estimation solutions to account for the problem of selective survival (see Appendix E). The setup of the design included a cluster-level (region-level) equation with cluster-level variables (e.g., famine incident) and cluster-level error (u_j) and an individual-level equation with individual-level variables (e.g., height) in addition to the cluster-level and an idiosyncratic error term (ϵ_{ij}). Thus, the treatment effect is at regional-level, and the outcome of interest is at individual-level.

As we demonstrate, the best choice among estimation techniques depends on the crucial assumption of the correlation of the error terms among the outcome (indexed 1) and selection (indexed 2) equations in each level (between u_{1j} and u_{2j}, and ϵ_{1ij} and ϵ_{2ij}). If a correlation exists only at regional level (thus $Cov(\epsilon_{1ij}, \epsilon_{2ij}) = 0$), one can use Inverse Probability Weighting, or a GLS random-effects model to successfully account for the selective survival. On

the other hand, if a correlation also exists at the individual level (thus $Cov(u_{1j}, u_{2j}) = 0$), both of these two approaches will yield inconsistent estimates. However, a Heckman-type estimation will successfully produce consistent estimates. This, of course, relies on the availability of additional exclusions. Results from this Monte Carlo experiment are in line with previous research on selection bias in similar setups (see Grilli and Rampichini (2010)). We extend this work by relaxing the assumption of bivariate normality using copulas that preserve the dependence structure and allow various forms of excess joint asymmetry, skewness or kurtosis.[14]

We then use this information on the survival probability to estimate models which account for sample selection departing from the standard assumption of bivariate normality of the errors by using more flexible copulas functions as explained in Section 4. The results in Table 9 show evidence of both selection and scarring effects. On the one hand, looking at the sign and magnitude of Kendal's τ,[15] the survivors are negatively selected on height, and both underweight and overweight, but positively selected on weight, BMI, and the chest/height ratio: in other words, those shorter but bulkier appear to have been more likely to survive, and the BMI distribution of the survivors appears to have been truncated both on the left and the right. On the other hand, once we control for survival, we still detect a significant scarring effect, with those exposed since early gestation being 1 p.p. more likely to be overweight than the controls, and those exposed since middle gestation being 0.8 p.p. more likely to be underweight; the impacts on weight and BMI detected with the simple Difference-in-Differences results are instead driven to 0. Interestingly, once we account for selection, we are able to detect a significant impact of exposure on the probability of being IQ impaired, approximately 1 p.p. increase for those exposed since middle and early gestation, w.r.t. a control mean of 4.5% - hence, quite a sizeable increase. In sum, the impacts of exposure in middle and early gestation are reinforced once we account for selective survival. The

[14] See Winkelmann (2012); Gomes et al. (2018) for simulation studies on relative performance on the selection of copula structure.

[15] Given that the dependence parameter does not have the same interpretation across different copulas, it is transformed into a standard rank correlation, Kendal's τ coefficient.

results are robust to the choice of the variables used as exclusion restrictions (Table C7 in the Appendix shows the results from re-estimating the models using different combinations among the variables included).[16]

6 Conclusion

In this paper, we study the effects of prenatal malnutrition on health by examining various anthropometric and mental outcomes at age 18. We do that by using the military recruits data obtained for the whole population affected around birth at the time of the Dutch famine of 1944-1945. These data have been used in a series of influential studies in the 1970s but have not been re-examined with modern econometric methods. In this study, we use newly digitalised data on the various other circumstances at the time and link them to the recruits data. Thus, we are able to re-assess the results in the previous literature and enrich the analysis by using updated definitions of overweight and redefining the analytical strategy using the most recent novel econometric methods available.

Consistent with the previous studies, we find robust famine exposure effects for weight, BMI, and chest/height ratio for those exposed since early gestation. Moreover, using alternative specifications, we find that these effects are a result of a combination of warfare exposure and protein reduction. This is true after controlling for changing conditions in the South, seasonality, and weather conditions. Thus, we add to the literature by pointing out the importance of other exposures (although warfare can have also an income effect) and diet composition in an adverse prenatal environment of war and famine. Finally, we find evidence of both selection and scarring effects, which are affecting all the outcomes in the baseline estimations.

[16] As a comparison, Table C8 shows the results from the estimating models using Inverse Probability Weighting (IPW). The weights are calculated for each city and birth month, by taking the inverse of the survival rate at age 18 when we observe the outcomes. The control variables included are as in the previous models. The results are similar in magnitude and significance to the baseline difference-in-differences estimation in Table 3. This similarity might be an indication that there exists correlation among individual-level unobservables between the selection and equation equations, in addition to the city-level ones, and thus the IPW fails to account fully for selection, as demonstrated in our simulation study.

References

Abadie, A., Athey, S., Imbens, G. W., and Wooldridge, J. M. (2017). Sampling-based vs. design-based uncertainty in regression analysis. *arXiv preprint arXiv:1706.01778*.

Almond, D., Currie, J., and Duque, V. (2017). Childhood circumstances and adult outcomes: Act II. Working Paper 23017, National Bureau of Economic Research.

Almond, D. and Mazumder, B. (2011). Health capital and the prenatal environment: the effect of Ramadan observance during pregnancy. *American Economic Journal: Applied Economics*, 3(4):56–85.

Banerjee, A., Duflo, E., Postel-Vinay, G., and Watts, T. (2010). Long-run health impacts of income shocks: Wine and phylloxera in nineteenth-century France. *The Review of Economics and Statistics*, 92(4):714–728.

Barker, D. J. (1990). The fetal and infant origins of adult disease. *British Medical Journal*, 301(6761):1111.

Beach, B., Ferrie, J. P., and Saavedra, M. H. (2018). Fetal shock or selection? The 1918 influenza pandemic and human capital development. Working Paper 24725, National Bureau of Economic Research.

Boonstra, O. (2016). Historische Database Nederlandse Gemeenten. http://hdl.handle.net/10622/RPBVK4, IISH Dataverse, V1.

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

Bruckner, T. A., van den Berg, G. J., Smith, K. R., and Catalano, R. A. (2014). Ambient temperature during gestation and cold-related adult mortality in a Swedish cohort, 1915-2002. *Social Science & Medicine*, 119:191–197.

Cameron, A. C., Gelbach, J. B., and Miller, D. L. (2008). Bootstrap-based improvements for inference with clustered errors. *The Review of Economics and Statistics*, 90(3):414–427.

CBS Central Bureau of Statistics (1935-1947). Huwelijken, geboorte en sterfte naar den leeftijd en naar de oorzaken van den dood, onder de werkelijke bevolking van Nederland. *Maandschrift van het Centraal Bureau voor de Statistiek*. 's-Gravenhage: Gebr. Belinfante.

Chen, Y. and Zhou, L.-A. (2007). The long-term health and economic consequences of the 1959–1961 famine in China. *Journal of Health Economics*, 26(4):659–681.

Costa, D. L. (2004). The measure of man and older age mortality: Evidence from the Gould sample. *The Journal of Economic History*, 64(1):1–23.

Cramer, J. and Lumey, L. (2010). Maternal preconception diet and the sex ratio. *Human Biology*, 82(1):103–107.

Currie, J. and Almond, D. (2011). Human capital development before age five. In *Handbook of Labor Economics*, volume 4, pages 1315–1486. Elsevier.

Currie, J. and Rossin-Slater, M. (2013). Weathering the storm: Hurricanes and birth outcomes. *Journal of Health Economics*, 32(3):487–503.

de Groot, R. H., Stein, A. D., Jolles, J., van Boxtel, M. P., Blauw, G.-J., Van de Bor, M., and Lumey, L. H. (2011). Prenatal famine exposure and cognition at age 59 years. *International Journal of Epidemiology*, 40(2):327–337.

de Rooij, S. R., Wouters, H., Yonker, J. E., Painter, R. C., and Roseboom, T. J. (2010). Prenatal undernutrition and cognitive function in late adulthood. *Proceedings of the National Academy of Sciences*, 107(39):16881–16886.

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

Departement van Landbouw en Visserij, Rijksbureau voor de Voedselvoorziening in oorlogstijd, afdeling Voedingsvraagstukken, Onderafd. Statistiek, West-Nederland (1946). Overzicht van het verloop der rantsoenen in West-Nederland per product, per verbruikers- en arbeidersgroep. A. In grammen per week. B. in aantal calorieen en grammen koolhydraten, eiwitten en vetten, gemiddeld per dag. Over het tijdvak 1 Oct. 1944-5 Jan 1946. Onderafd Statistiek, vL/A1/H. 6 Mei 1946.

Ekamper, P., Bijwaard, G., van Poppel, F., and Lumey, L. (2017). War-related excess mortality in the Netherlands, 1944–45: New estimates of famine-and non-famine-related deaths from national death records. *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, 50(2):113–128.

Gemeente Amsterdam (1946). Maandbericht van het Bureau van de Statistiek der Gemeente Amsterdam. Amsterdam: Meulenhoff.

Gemeente Den Haag (1946). De haagsche bevolking een jaar na de bevrijding. Mededelingen 1946, nr. 4.

Gemeente Utrecht (1946). Statistische berichten der Gemeente Utrecht. Januari – December 1946.

Gemeentelijk Bureau voor de Statistiek Rotterdam (1946). Statistische mededelingen der gemeente Rotterdam. 1ste, 2de, 3de en 4de kwartaal met drie bijlagen.

Gomes, M., Radice, R., Camarena Brenes, J., and Marra, G. (2018). Copula selection models for non-gaussian outcomes that are missing not at random. *Statistics in Medicine*.

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.

Grilli, L. and Rampichini, C. (2010). Selection bias in linear mixed models. *Metron*, 68(3):309–329.

Havari, E. and Mazzonna, F. (2015). Can we trust older people's statements on their childhood circumstances? Evidence from SHARELIFE. *European Journal of Population*, 31(3):233–257.

Isen, A., Rossin-Slater, M., and Walker, R. (2017). Relationship between season of birth, temperature exposure, and later life wellbeing. *Proceedings of the National Academy of Sciences*, 114(51):13447–13452.

Jewell, N. P., Spagat, M., and Jewell, B. L. (2018). Accounting for civilian casualties: From the past to the future. *Social Science History*, 42(3):379–410.

Jürges, H. (2013). Collateral damage: The German food crisis, educational attainment and labor market outcomes of German post-war cohorts. *Journal of Health Economics*, 32(1):286–303.

Jürges, H. (2015). Ramadan fasting, sex-ratio at birth, and birth weight: No effects on Muslim infants born in Germany. *Economics Letters*, 137:13–16.

Karlsson, M. and Ziebarth, N. R. (2017). Population health effects and health-related costs of extreme temperatures: Comprehensive evidence from germany. Cornell, Mimeo.

Kesternich, I., Siflinger, B., Smith, J. P., and Winter, J. K. (2014). The effects of World War II on economic and health outcomes across Europe. *The Review of Economics and Statistics*, 96(1):103–118.

Kesternich, I., Siflinger, B., Smith, J. P., and Winter, J. K. (2015). Individual behaviour as a pathway between early-life shocks and adult health: Evidence from hunger episodes in post-war Germany. *The Economic Journal*, 125(588).

KNMI Royal Netherlands Meteorological Institute (2018). Daggegevens van het weer in Nederland. https://www.knmi.nl/nederland-nu/klimatologie/daggegevens.

Korthals Altes, A. (1984). *Luchtgevaar: Luchtaanvallen op Nederland 1940-1945*. Sijthoff, Amsterdam.

Lee, D. S. (2009). Training, wages, and sample selection: Estimating sharp bounds on treatment effects. *Review of Economic Studies*, 76(3):1071–1102.

Lee, L.-F. (1983). Generalized econometric models with selectivity. *Econometrica*, pages 507–512.

Li, C. and Lumey, L. (2017). Exposure to the Chinese famine of 1959-61 in early life and long-term health conditions: A systematic review and meta-analysis. *International Journal of Epidemiology*, 46(4):1157–1170.

Liu, E. M., Liu, J.-T., and Tseng, T.-Y. H. (2015). The impact of a natural disaster on the incidence of fetal losses and pregnancy outcomes. *Mimeo, University of Houston*.

Lumey, L. (1998). Compensatory placental growth after restricted maternal nutrition in early pregnancy. *Placenta*, 19(1):105–111.

Lumey, L., Stein, A. D., Kahn, H. S., Van der Pal-de Bruin, K. M., Blauw, G., Zybert, P. A., and Susser, E. S. (2007). Cohort profile: The Dutch Hunger Winter families study. *International Journal of Epidemiology*, 36(6):1196–1204.

Lumey, L. and van Poppel, F. W. (1994). The Dutch famine of 1944–45: Mortality and morbidity in past and present generations. *Social History of Medicine*, 7(2):229–246.

Lumey, L. H., Stein, A. D., and Susser, E. (2011). Prenatal famine and adult health. *Annual Review of Public Health*, 32:237–262.

Majid, M. F. (2015). The persistent effects of in utero nutrition shocks over the life cycle: Evidence from Ramadan fasting. *Journal of Development Economics*, 117:48–57.

Neelsen, S. and Stratmann, T. (2011). Effects of prenatal and early life malnutrition: Evidence from the Greek famine. *Journal of Health Economics*, 30(3):479–488.

NIOD Institute for War, Holocaust and Genocide Studies (2018). Slachtoffers van bombardementen. Anoniem overzicht bijgehouden in 1940-1945. NIOD dossier 249-0116A, A20.

Pebesma, E. J. (2004). Multivariable geostatistics in S: The gstat package. *Computers & Geosciences*, 30(7):683–691.

Portrait, F., van Wingerden, T., and Deeg, D. (2017). Early life undernutrition and adult height: The Dutch famine of 1944-45. *Economics & Human Biology*, 27:339–348.

Prinz, D., Chernew, M., Cutler, D., and Frakt, A. (2018). Health and economic activity over the lifecycle: Literature review. Working Paper 24865, National Bureau of Economic Research.

Ravelli, A. C., van der Meulen, J. H., Osmond, C., Barker, D. J., and Bleker, O. P. (1999). Obesity at the age of 50 y in men and women exposed to famine prenatally. *American Journal of Clinical Nutrition*, 70(5):811–816.

Ravelli, G.-P., Stein, Z. A., and Susser, M. W. (1976). Obesity in young men after famine exposure in utero and early infancy. *New England Journal of Medicine*, 295(7):349–353.

Romano, J. P. and Wolf, M. (2005). Stepwise multiple testing as formalized data snooping. *Econometrica*, 73(4):1237–1282.

Romano, J. P. and Wolf, M. (2016). Efficient computation of adjusted p-values for resampling-based stepdown multiple testing. *Statistics & Probability Letters*, 113:38–40.

Roodman, D., MacKinnon, J. G., Nielsen, M. O., and Webb, M. D. (2018). Fast and wild: Bootstrap inference in Stata using boottest. Working Papers 1406, Queen's University, Department of Economics.

Scholte, R. S., Van den Berg, G. J., and Lindeboom, M. (2015). Long-run effects of gestation during the Dutch hunger winter famine on labor market and hospitalization outcomes. *Journal of Health Economics*, 39:17–30.

Schwandt, H. (2017). The lasting legacy of seasonal influenza: In-utero exposure and labor market outcomes. Technical Report No. 10589, IZA Discussion Paper Series.

Smith, M. D. (2003). Modelling sample selection using Archimedean copulas. *The Econometrics Journal*, 6(1):99–123.

Stanner, S. A., Bulmer, K., Andres, C., Lantseva, O. E., Borodina, V., Poteen, V., and Yudkin, J. S. (1997). Does malnutrition in utero determine diabetes and coronary heart disease in adulthood? Results from the leningrad siege study, a cross sectional study. *British Medical Journal*, 315(7119):1342–1348.

Stein, A. D., Kahn, H. S., Rundle, A., Zybert, P. A., van der Pal-de Bruin, K., and Lumey, L. (2007). Anthropometric measures in middle age after exposure to famine during gestation: Evidence from the Dutch famine–. *American Journal of Clinical Nutrition*, 85(3):869–876.

Stein, Z. and Susser, M. (1975). The Dutch famine, 1944–1945, and the reproductive process. I. Effects on six indices at birth. *Pediatric Research*, 9(2):70.

Stein, Z., Susser, M., Saenger, G., and Marolla, F. (1972). Nutrition and mental performance. *Science*, 178(4062):708–713.

Stein, Z., Susser, M., Saenger, G., and Marolla, F. (1975). *Famine and human development: The Dutch hunger winter of 1944-1945.* Oxford University Press, New York.

Tobi, E. W., Slieker, R. C., Luijk, R., Dekkers, K. F., Stein, A. D., Xu, K. M., Slagboom, P. E., Van Zwet, E. W., Lumey, L., Heijmans, B. T., et al. (2018). DNA methylation as a mediator of the association between prenatal adversity and risk factors for metabolic disease in adulthood. *Science Advances*, 4(1):eaao4364.

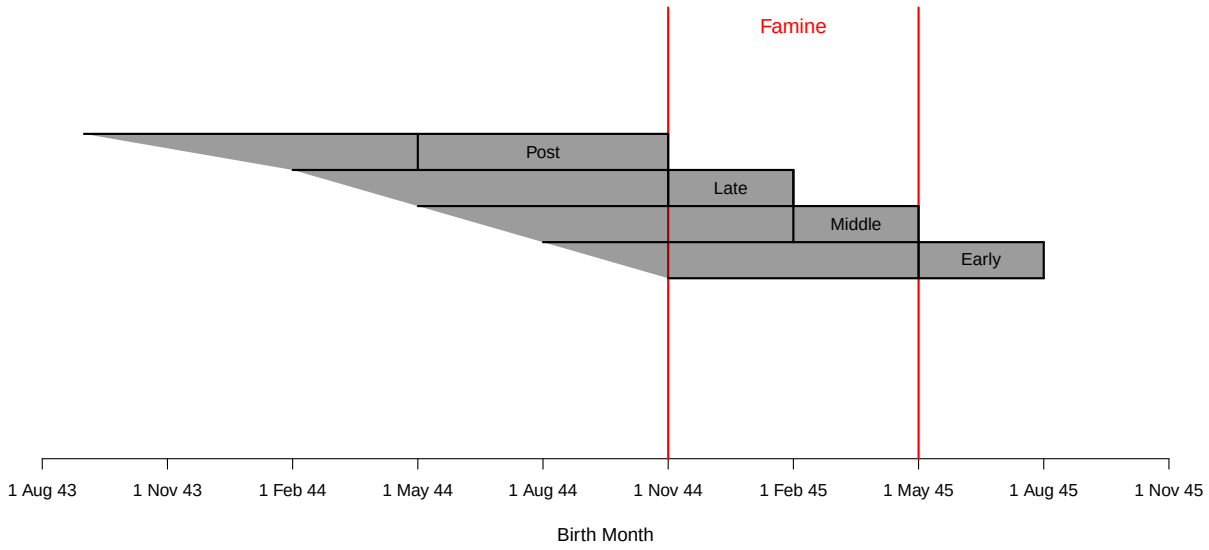
Van den Berg, G. J., Pinger, P. R., and Schoch, J. (2016). Instrumental variable estimation of the causal effect of hunger early in life on health later in life. *The Economic Journal*, 126(591):465–506.

Wilde, J., Apouey, B. H., and Jung, T. (2017). The effect of ambient temperature shocks during conception and early pregnancy on later life outcomes. *European Economic Review*, 97:87–107.

Winkelmann, R. (2012). Copula bivariate probit models: With an application to medical expenditures. *Health Economics*, 21(12):1444–1455.

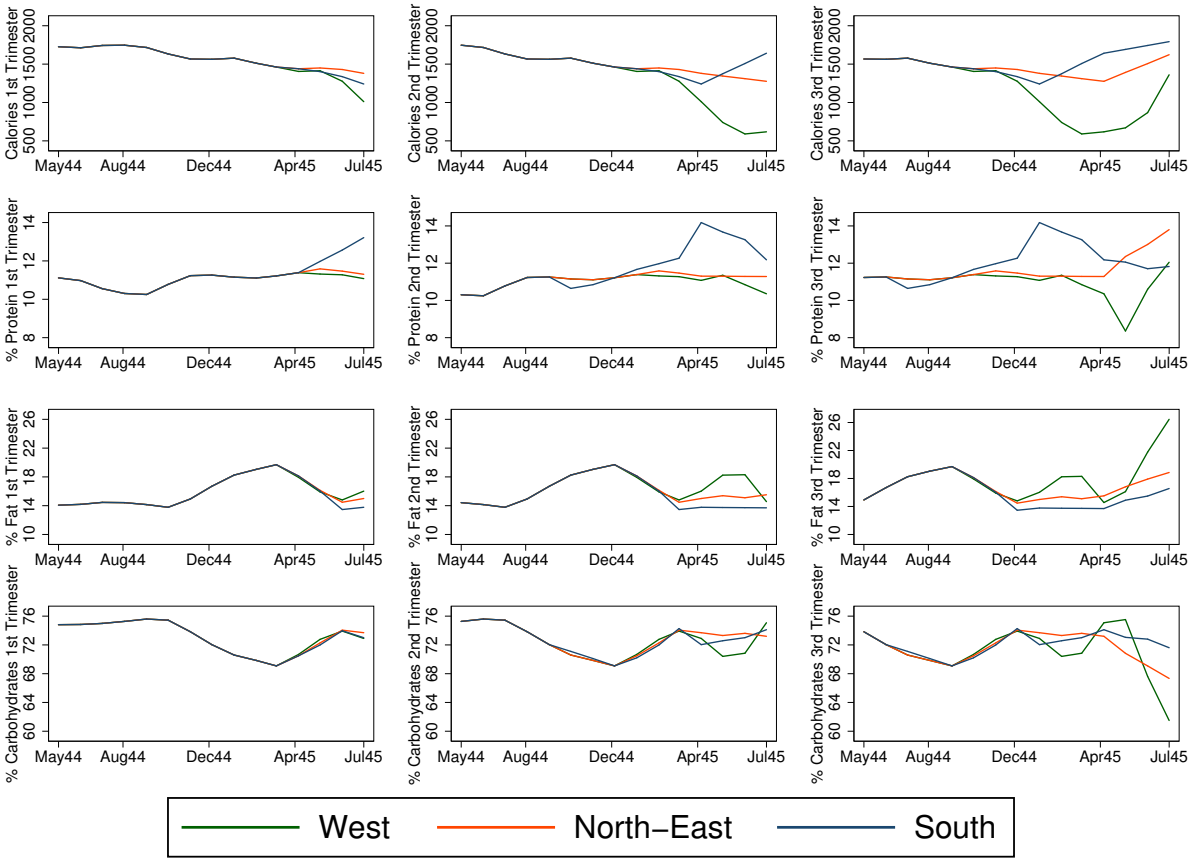
7 Figures and Tables

Figure 1: Treatment and control groups definition



Note: The time periods corresponding to each of the three treatment groups and the control group, with the birth months for each group enclosed in the respective boxes. The pregnancy period corresponding to each birth month is shaded in grey. The red vertical lines enclose the famine period.

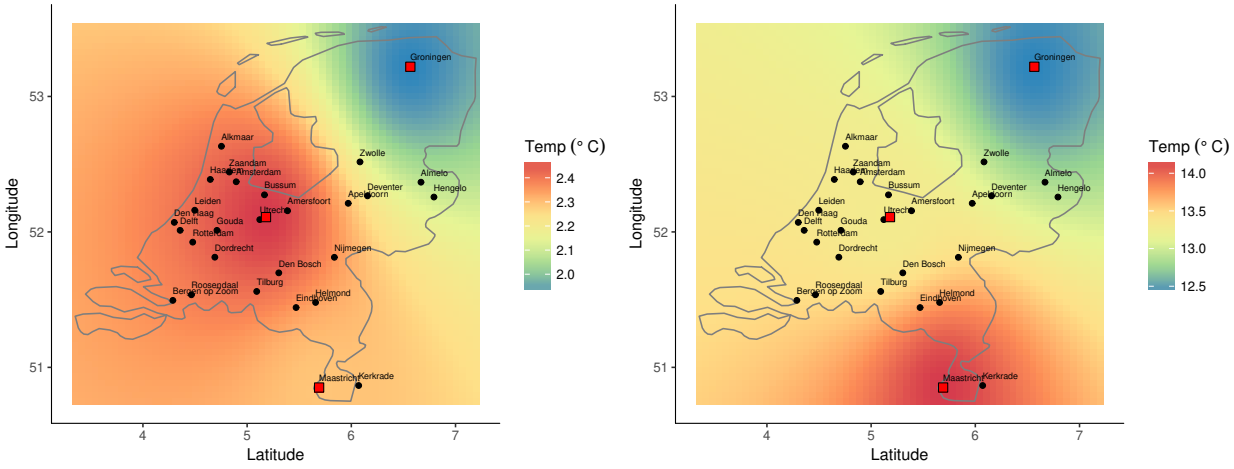
Figure 2: Calories and macronutrient share for each trimester by month of birth



Note: Data on caloric intakes and macronutrients on weekly level for the West and monthly level for North-East and South, using the official war information on the rations. Shares are calculated as $Protein\ share = 4 \times Protein(grams)/Calories(kcals)$, $Fat\ share = 9 \times Fat(grams)/Calories(kcals)$ and $Carbohydrates\ share = 100 - (Protein\ share + Fat\ share)$.

Figure 3: Example of Temperature Interpolation

(a) December 1944 (b) May 1945

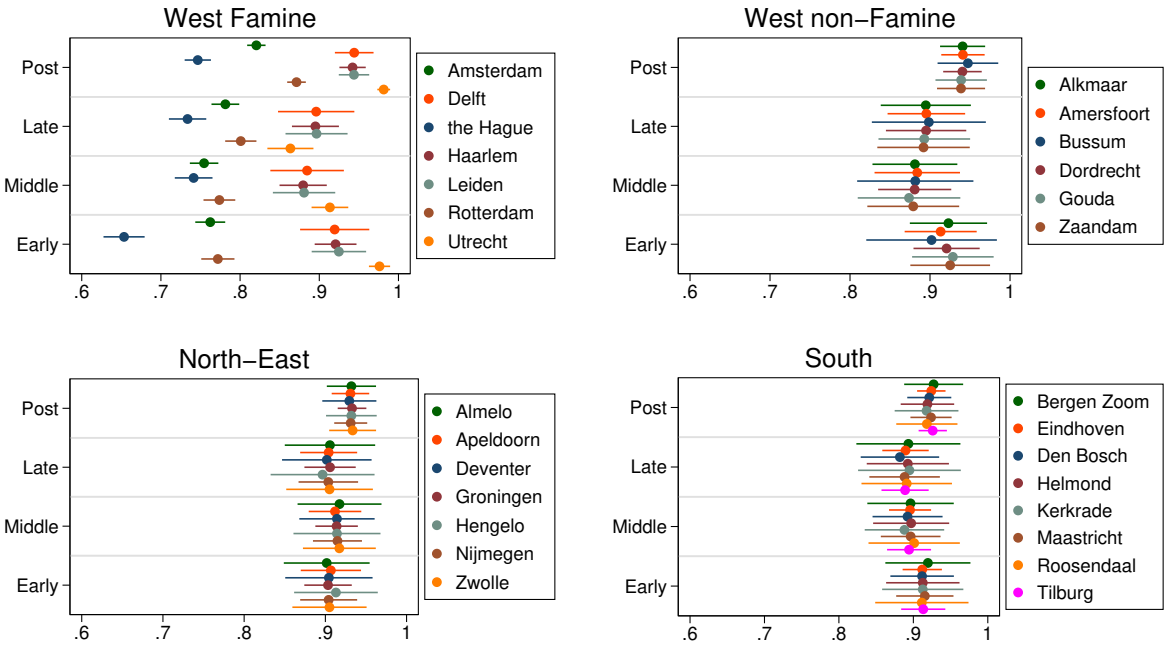


Note: The heatmap shows the predicted temperature across the Netherlands using the Inverse Distance Weighting method. The red squares are the meteorological stations, and the black dots are the cities in the study.

Figure 4: Deaths of civilians due to operations of war during pregnancy for each birth month by city

Note: Results for those born between May 1944 and July 1945. The following cities: Alkmaar, Almelo, Amersfoort, Bergen op Zoom, Bussum, Delft, Gouda, Groningen, Helmond, Kerkrade, Leiden, Zaandam, and Zwolle had the less variation and not shown here.

Figure 5: Alive at 18 by city and famine exposure



Note: Categories are defined as following: Post is born May1944-Oct1944; Late is born Nov1944-Jan1945; Middle is born Feb1945-Apr1945; Early is born May1945-Jul1945. Survival rates are calculated for each city and month of birth using information on births and deaths for that period.

Table 1: Selection of Cities Summary

Non-selected		Excl due pop dispersed	Excl due pop changing	Selected	
Region	Municipality			Region	Municipality
North-East	Arnhem		✓	West Famine	Amsterdam
	Ede	✓			Delft
	Emmen	✓			The Hague
	Enschede		✓		Haarlem
	Rheden	✓			Leiden
	Leeuwarden		✓		Rotterdam
South	Breda		✓		Utrecht
	Venlo		✓	North-East	Almelo
West	Haarlemmermeer	✓			Apeldoorn
	Den Helder		✓		Deventer
	Hilversum		✓		Hengelo
	Schiedam		✓		Nijmegen
	Velsen		✓		Zwolle
	Vlaardingen		✓		Groningen
	Vlissingen		✓	South	Bergen op Zoom
	Voorburg		✓		Eindhoven
	Zeist		✓		Helmond
					's Hertogenbosch
					Kerkrade
					Maastricht
					Roosendaal
					Tilburg
					Heerlen
				West	Alkmaar
				Non-Famine	Amersfoort
					Bussum
					Dordrecht
					Gouda
					Zaandam

Note: List of 46 cities with a population greater than 25,000 inhabitants on January 1, 1940. Furthermore, 4 municipalities were excluded because the majority of the population was not living in the largest place of the municipality and 13 municipalities were excluded because their population occurred major changes in size since 1930.

Table 2: Descriptive Statistics of Additional Controls

Variable	Level	Mean	Std Dev	Min	Max
Temperature (°C) 1st TR	Monthly	9.88	4.70	1.70	16.70
Temperature (°C) 2nd TR	Monthly	8.07	4.91	1.70	16.70
Temperature (°C) 3rd TR	Monthly	8.54	4.75	1.70	16.70
log(Warfare 1st TR + 1)	Monthly	2.17	1.56	0.00	6.55
log(Warfare 2nd TR + 1)	Monthly	2.57	1.64	0.00	6.55
log(Warfare 3rd TR + 1)	Monthly	2.92	1.71	0.00	6.39
-Calories 1st TR (1,000s)	Weekly	-1.490	0.26	-1.75	-0.59
-Calories 2nd TR (1,000s)	Weekly	-1.340	0.35	-1.75	-0.59
-Calories 3rd TR (1,000s)	Weekly	-1.314	0.37	-2.05	-0.59
-ProteinShare 1st TR	Weekly	-11.11	0.60	-14.18	-10.25
-ProteinShare 2nd TR	Weekly	-11.07	0.82	-14.18	-8.66
-ProteinShare 3rd TR	Weekly	-11.28	0.94	-14.18	-8.66
South Weeks Liberated	Weekly	1.70	8.57	-26.00	44.43

Note: For cohorts in analytical sample (May1944-July1945 in the selected cities). Number of observations for all the variables is 40,950.

Table 3: Difference-in-Differences Famine Effect on Age 18 Outcomes

	(1) Height	(2) Weight	(3) BMI (Weight/ Height^2)	(4) Overweight (BMI $\geq$25)	(5) Obese (Ravelli) (Weight/ Height>120%)	(6) Underweight (BMI< 18.5)	(7) Chest/ Height Ratio	(8) IQ Impaired (ICD-6 325)	(9) Mental Deficiency Stein (1972)
WestF×Late	-0.099	-0.350*	-0.092	-0.010*	-0.006**	0.001	0.000	0.001	0.002
	(0.150)	(0.193)	(0.058)	(0.005)	(0.003)	(0.006)	(0.001)	(0.006)	(0.006)
WestF×Middle	-0.384	-0.334	-0.012	-0.006	0.001	-0.003	0.008***	0.007	0.010*
	(0.305)	(0.445)	(0.106)	(0.006)	(0.004)	(0.006)	(0.002)	(0.006)	(0.006)
WestF×Early	-0.052	0.652**	0.218***	0.009	0.006**	-0.013*	0.010***	0.001	0.005
	(0.298)	(0.313)	(0.075)	(0.006)	(0.003)	(0.007)	(0.002)	(0.006)	(0.006)
Wild cluster bootstrap p-values:									
WestF×Late	0.530	0.102	0.140	0.078	0.062	0.824	0.729	0.825	0.742
WestF×Middle	0.303	0.586	0.914	0.410	0.837	0.687	0.001	0.289	0.094
WestF×Early	0.866	0.051	0.013	0.124	0.041	0.123	0.001	0.891	0.418
RW p-values:									
WestF×Late	0.822	0.093	0.154	0.082	0.058	0.935	0.935	0.935	0.935
WestF×Middle	0.277	0.558	0.893	0.461	0.893	0.824	0.000	0.389	0.065
WestF×Early	0.949	0.034	0.006	0.071	0.034	0.048	0.000	0.949	0.325
Observations	42,826	42,826	42,826	42,826	42,826	42,826	42,826	42,826	42,826
Controls and FE	✓	✓	✓	✓	✓	✓	✓	✓	✓
Control Mean	177.368	67.578	21.465	0.06	0.014	0.056	0.492	0.039	0.03

Note: Robust standard errors in parentheses clustered at the level of city. *** p<0.01, ** p<0.05, * p<0.1. *Late* refers to those born November 1944-January 1945, *Middle* to February 1945-April 1945, and *Early* to May 1945-July 1945. Controls include father's occupation, number of older brothers, birth order and religion. FE are included for city and for month of birth. Control Mean refers to the mean of the outcome for those born in the West Famine area with postnatal exposure only.

Table 4: Placebo Difference-in-Differences

	(1) Height	(2) Weight	(3) BMI	(4) Overweight	(5) Obese (Ravelli)	(6) Underweight	(7) Chest/Height Ratio	(8) IQ Impaired	(9) Mental Deficiency
WestF×PlaceboLate	-0.009	-0.280	-0.081	-0.006	0.002	-0.001	-0.002	-0.003	-0.001
	(0.132)	(0.197)	(0.061)	(0.005)	(0.004)	(0.007)	(0.002)	(0.007)	(0.007)
WestF×PlaceboMiddle	0.156	-0.268	-0.120	-0.000	0.000	0.001	-0.003	0.005	0.009
	(0.153)	(0.261)	(0.074)	(0.007)	(0.003)	(0.007)	(0.006)	(0.008)	(0.010)
WestF×PlaceboEarly	-0.122	-0.229	-0.044	0.008	0.003	0.005	-0.001	-0.004	-0.001
	(0.160)	(0.209)	(0.052)	(0.007)	(0.003)	(0.007)	(0.005)	(0.006)	(0.008)
Observations	59,040	59,040	59,040	59,040	59,040	59,040	59,040	59,040	59,040
Controls and FE	✓	✓	✓	✓	✓	✓	✓	✓	✓
Control Mean	177.884	68.537	21.644	0.074	0.019	0.055	0.499	0.039	0.03

Note: Robust standard errors in parentheses clustered at the level of city. *** p<0.01, ** p<0.05, * p<0.1. Controls include father's occupation, number of older brothers, birth order and religion. FE are included for city and for month of birth. Control Mean refers to the mean of the outcome for those born in the West Famine area with postnatal exposure only. Exposure cohorts are constructed using the same structure as in the main estimation, but two years after. Outcome and *Early, Middle, Late* definitions as in Table 3.

Table 5: Triple Difference Estimation

	(1) Height	(2) Weight	(3) BMI	(4) Overweight	(5) Obese (Ravelli)	(6) Underweight	(7) Chest/Height Ratio	(8) IQ Impaired	(9) Mental Deficiency
WestF×Late×War	-0.078	-0.095	-0.021	-0.004	-0.009	0.003	0.002	0.004	0.002
	(0.161)	(0.268)	(0.078)	(0.006)	(0.006)	(0.009)	(0.002)	(0.008)	(0.008)
WestF×Middle×War	-0.535	-0.064	0.108	-0.005	0.001	-0.003	0.012*	0.001	0.001
	(0.338)	(0.493)	(0.129)	(0.010)	(0.005)	(0.008)	(0.006)	(0.010)	(0.011)
WestF×Early×War	0.072	0.892**	0.265**	0.002	0.003	-0.018*	0.012*	0.004	0.005
	(0.296)	(0.392)	(0.101)	(0.009)	(0.005)	(0.010)	(0.007)	(0.008)	(0.009)
Wild cluster bootstrap p-values:									
WestF×Late	0.626	0.759	0.808	0.497	0.179	0.752	0.335	0.630	0.790
WestF×Middle	0.181	0.908	0.456	0.686	0.903	0.696	0.148	0.889	0.920
WestF×Early	0.827	0.029	0.025	0.804	0.601	0.129	0.156	0.599	0.589
RW p-values:									
WestF×Late	0.950	0.966	0.966	0.882	0.242	0.966	0.641	0.950	0.966
WestF×Middle	0.129	0.996	0.668	0.944	0.996	0.947	0.057	0.996	0.996
WestF×Early	0.920	0.027	0.002	0.920	0.888	0.102	0.079	0.888	0.874
Observations	101,866	101,866	101,866	101,866	101,866	101,866	101,866	101,866	101,866
Controls and FE	✓	✓	✓	✓	✓	✓	✓	✓	✓
Control Mean	177.368	67.578	21.465	0.06	0.014	0.056	0.492	0.039	0.03

Note: Robust standard errors in parentheses clustered at the level of city. *** p<0.01, ** p<0.05, * p<0.1. Controls include father's occupation, number of older brothers, birth order and religion. FE are included for city and for month of birth. Control Mean refers to the mean of the outcome for those born in the West Famine area with postnatal exposure only. Estimation is performed in the original and the placebo samples combined, using a triple interaction term for area, trimester and being in the original sample. Outcome and *Early, Middle, Late* definitions as in Table 3.

Table 6: Robustness of Difference-in-Differences on Age 18 Outcomes

	(1) Weight	(2) BMI	(3) Overweight	(4) Obese (Ravelli)	(5) Underweight	(6) Chest/Height Ratio
WestF×Late	-0.382	-0.037	-0.005	-0.004	0.001	-0.002
	(0.247)	(0.071)	(0.006)	(0.003)	(0.008)	(0.002)
WestF×Middle	-0.426	0.009	0.000	0.007*	0.001	0.003
	(0.362)	(0.101)	(0.009)	(0.004)	(0.010)	(0.004)
WestF×Early	0.482	0.141	0.010	0.007	-0.002	0.004
	(0.319)	(0.107)	(0.009)	(0.005)	(0.012)	(0.004)
log(Warfare1stTR+1)	0.040	-0.010	0.001	0.001	0.002	0.000
	(0.056)	(0.011)	(0.001)	(0.001)	(0.001)	(0.000)
log(Warfare2ndTR+1)	0.031	0.013	0.003**	-0.000	-0.001	-0.001
	(0.054)	(0.014)	(0.001)	(0.001)	(0.001)	(0.000)
log(Warfare3rdTR+1)	0.100*	0.030**	0.001	0.001	-0.001	0.000
	(0.053)	(0.013)	(0.001)	(0.001)	(0.001)	(0.000)
-Calories1stTR	0.575	-0.059	0.034	0.013	0.039	0.006
	(0.972)	(0.245)	(0.022)	(0.015)	(0.029)	(0.005)
-ProteinShare1stTR	-0.061	0.011	-0.001	0.006	0.008	0.001
	(0.352)	(0.079)	(0.008)	(0.005)	(0.008)	(0.001)
-Calories2ndTR	-0.773	-0.086	-0.031*	-0.021	-0.026	-0.003
	(0.898)	(0.208)	(0.016)	(0.014)	(0.016)	(0.003)
-ProteinShare2ndTR	0.171	0.041	-0.001	0.002	-0.006	0.001
	(0.145)	(0.040)	(0.004)	(0.003)	(0.005)	(0.001)
-Calories3rdTR	-0.839	-0.298	-0.014	-0.003	0.029**	0.005
	(0.602)	(0.201)	(0.015)	(0.006)	(0.011)	(0.004)
-ProteinShare3rdTR	0.428***	0.105***	0.008**	0.003*	-0.003	0.001*
	(0.115)	(0.023)	(0.003)	(0.002)	(0.003)	(0.000)
Wild cluster bootstrap *p*-values:						
WestF×Late	0.144	0.614	0.441	0.208	0.936	0.352
WestF×Middle	0.282	0.937	0.969	0.127	0.916	0.592
WestF×Early	0.165	0.266	0.302	0.243	0.906	0.395
log(Warfare1stTR+1)	0.530	0.340	0.565	0.463	0.215	0.814
log(Warfare2ndTR+1)	0.584	0.335	0.080	0.479	0.398	0.225
log(Warfare3rdTR+1)	0.170	0.111	0.713	0.408	0.320	0.812
-Calories1stTR	0.577	0.830	0.135	0.401	0.200	0.331
-ProteinShare1stTR	0.879	0.902	0.922	0.351	0.356	0.624
-Calories2ndTR	0.427	0.690	0.109	0.180	0.100	0.441
-ProteinShare2ndTR	0.260	0.341	0.838	0.482	0.274	0.330
-Calories3rdTR	0.214	0.206	0.393	0.667	0.017	0.225
-ProteinShare3rdTR	0.003	0.001	0.046	0.087	0.287	0.067
RW *p*-values:						
WestF×Late	0.203	0.702	0.551	0.294	0.915	0.410
WestF×Middle	0.339	0.997	0.997	0.109	0.997	0.719
WestF×Early	0.218	0.295	0.295	0.295	0.848	0.295
log(Warfare1stTR+1)	0.636	0.601	0.636	0.636	0.360	0.653
log(Warfare2ndTR+1)	0.633	0.585	0.065	0.633	0.623	0.291
log(Warfare3rdTR+1)	0.078	0.030	0.774	0.489	0.489	0.774
-Calories1stTR	0.561	0.721	0.170	0.487	0.230	0.359
-ProteinShare1stTR	0.984	0.984	0.984	0.421	0.532	0.822
-Calories2ndTR	0.373	0.573	0.082	0.098	0.098	0.373
-ProteinShare2ndTR	0.387	0.387	0.796	0.461	0.387	0.387
-Calories3rdTR	0.154	0.140	0.372	0.581	0.010	0.154
-ProteinShare3rdTR	0.000	0.000	0.012	0.052	0.174	0.052
Observations	40,950	40,950	40,950	40,950	40,950	40,950
Controls and FE	✓	✓	✓	✓	✓	✓

Note: Robust standard errors in parentheses clustered at the level of city. *** p<0.01, ** p<0.05, * p<0.1. Controls include father's occupation, number of older brothers, birth order, religion, temperature by trimester and weeks since liberation for the South. FE are included for city and for month of birth. Outcome and *Early, Middle, Late* definitions as in Table 3.

Table 7: Difference-in-Differences Results for the Birth Outcomes Males Sample

	(1) Birth Weight	(2) Low Birth Weight	(3) Birth Length	(4) Head Circumference	(5) Placenta Weight	(6) Gestational Age
West×Late	-76.681	-0.001	0.010	0.887*	-38.839	-0.079
	(87.265)	(0.034)	(0.372)	(0.508)	(26.568)	(0.286)
West×Middle	-168.993**	-0.007	-0.522	0.250	-39.760*	0.050
	(82.816)	(0.036)	(0.381)	(0.411)	(22.748)	(0.299)
West×Early	-76.137	-0.008	-0.320	0.181	-43.531*	-0.023
	(78.591)	(0.033)	(0.379)	(0.409)	(23.034)	(0.298)
RW p-values:						
West×Late	0.785	0.988	0.988	0.376	0.482	0.988
West×Middle	0.183	0.964	0.514	0.896	0.324	0.964
West×Early	0.809	0.957	0.865	0.957	0.248	0.957
$F_{West \times Middle = West \times Early}$	1.981	0.002	0.470	0.055	0.048	0.058
p-value	0.159	0.962	0.493	0.274	0.826	0.809
$F_{West \times Late = West \times Middle = West \times Early}$	1.150	0.031	1.504	1.297	0.033	0.098
p-value	0.317	0.970	0.223	0.815	0.968	0.907
Observations	1,931	1,931	1,869	1,275	1,295	1,259
Control and FE	✓	✓	✓	✓	✓	✓
Control Mean	3,414.378	.072	50.584	38.857	633.051	39.518

Note: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. All models include mother's age and FE for city and for month of birth. Control Mean refers to the mean of the outcome for those born in the West Famine area with postnatal exposure only. Results from hospital records in the birth data. *Early, Middle, Late* definitions as in Table 3.

Table 8: Difference-in-Differences Alive at 18

	(1) Alive at 18
WestF×Late	-0.015
	(0.012)
WestF×Middle	-0.031**
	(0.015)
WestF×Early	-0.048***
	(0.012)
Wild cluster bootstrap p-values:	
WestF×Late	0.256
WestF×Middle	0.185
WestF×Early	0.003
Observations	50,120
City FE & Birth month FE	✓
Control Mean	0.847

Note: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. All models include mother's age and city FE. Control Mean refers to the mean of the outcome for those born in the West Famine area with postnatal exposure only. Results from hospital records in the birth data. *Early, Middle, Late* definitions as in Table 3.

Table 9: Difference-in-Differences on Age 18 Outcomes with Sample Selection using Copulas

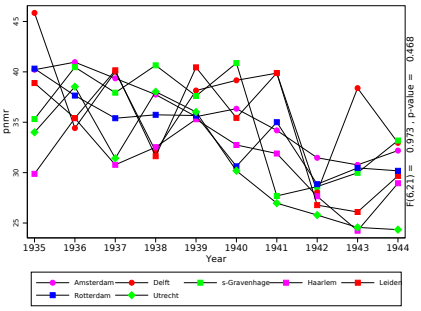
	(1) Height	(2) Weight	(3) BMI	(4) Overweight	(5) Obese (Ravelli)	(6) Underweight	(7) Chest/Height Ratio	(8) IQ Impaired	(9) Mental Deficiency
WestF×Late	-0.118	-0.364**	-0.064	0.000	-0.002	0.005	0.001	0.007	0.007
	(0.152)	(0.185)	(0.055)	(0.004)	(0.003)	(0.004)	(0.001)	(0.005)	(0.005)
WestF×Middle	-0.306	-0.617	-0.106	0.002	0.005	0.008**	0.008***	0.012**	0.014***
	(0.295)	(0.510)	(0.114)	(0.005)	(0.003)	(0.004)	(0.002)	(0.005)	(0.004)
WestF×Early	0.033	0.119	0.075	0.011**	0.011***	0.005	0.010***	0.009*	0.012**
	(0.282)	(0.299)	(0.056)	(0.004)	(0.003)	(0.004)	(0.003)	(0.005)	(0.005)
Observations	49,211	49,211	49,211	49,211	49,211	49,211	49,211	49,211	49,211
Controls and FE	✓	✓	✓	✓	✓	✓	✓	✓	✓
Copula	FGM	Clayton	Clayton	Joe	Gumbel	Joe	Clayton	Frank	Frank
Kendal's τ	-0.214	0.436	0.474	-0.697	-0.667	-0.712	0.275	-0.647	-0.673
Wald Test	29.63	570.42	567.93	2,788.58	246.10	1,643.62	11.08	585.81	125.18
p-value	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001

Note: Robust standard errors in parentheses clustered at the level of city. *** p<0.01, ** p<0.05, * p<0.1. Controls include father's occupation, number of older brothers, birth order, and religion (outcome equation only). FE are included for city and for month of birth (both outcome and selection equations). Wald Test for independence of the error terms in selection and outcome equations. First stage results are presented in Table C6. These results correspond to the results in Table 3 without selection. Outcome and *Early, Middle, Late* definitions as in Table 3.

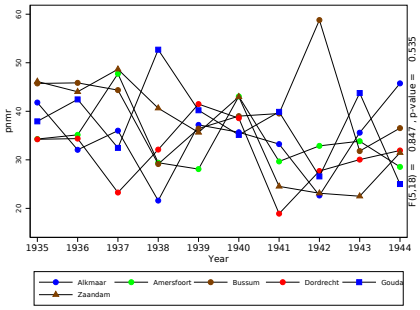
Appendix A Pretrends for selection of cities

Figure A1: Pretrends for Postnatal Mortality Rate

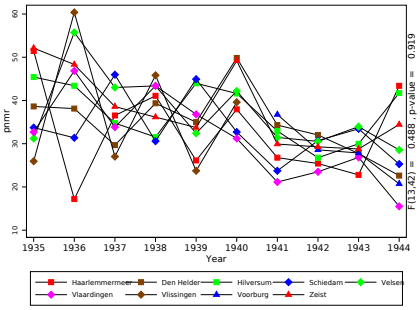
(a) West Famine



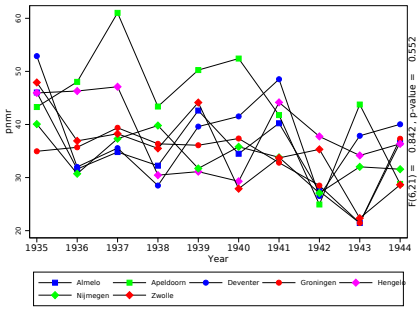
(b) West non-Famine Selected



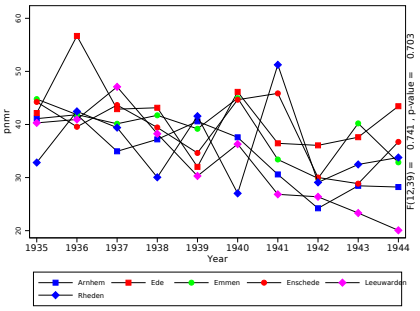
(c) West non-Famine non-Selected



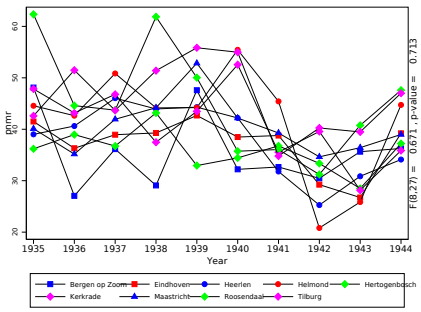
(d) North-East Selected



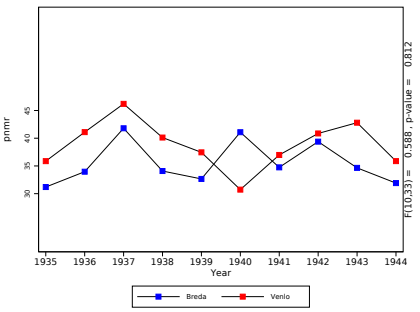
(e) North-East non-Selected



(f) South Selected



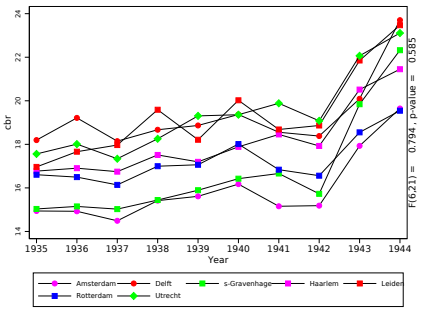
(g) South non-Selected



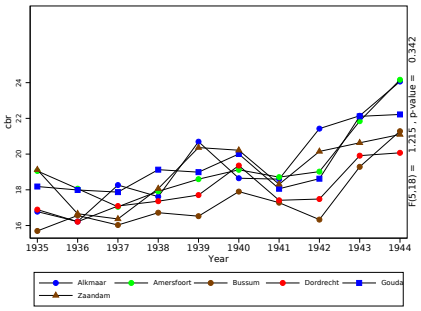
Note: Pretrends shown for groups of selected, and non-selected, in the analysis cities. A Wald test is performed in each group, by fitting a linear regression with city-specific slopes and testing jointly whether the slopes are equal.

Figure A2: Pretrends for Crude Birth Rate

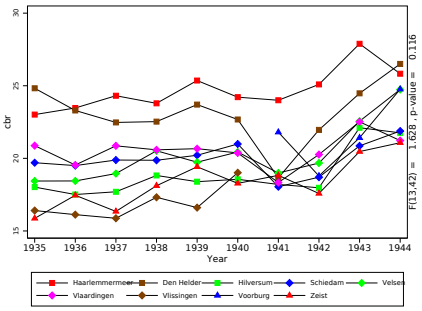
(a) West Famine



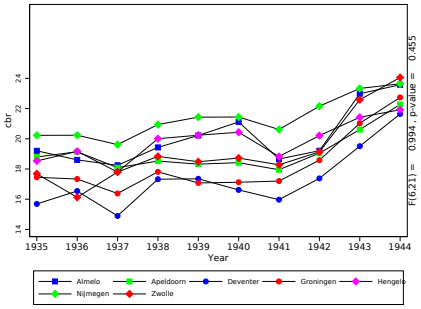
(b) West non-Famine Selected



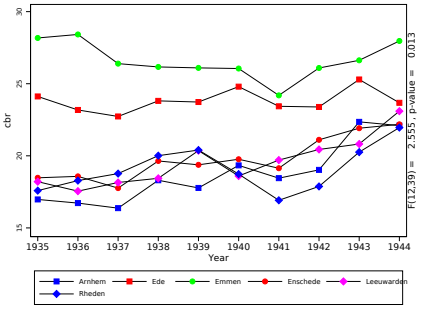
(c) West non-Famine non-Selected



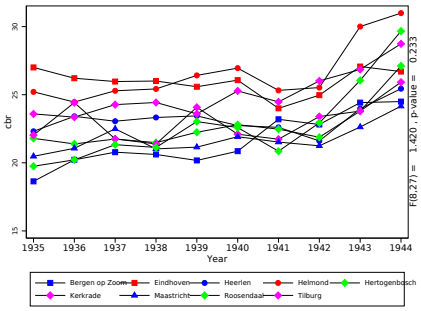
(d) North-East Selected



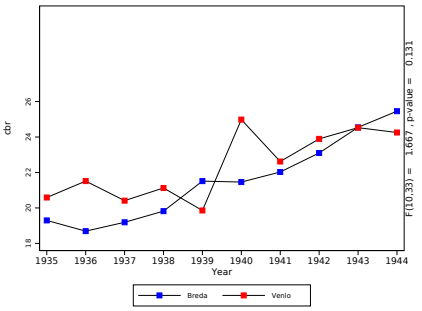
(e) North-East non-Selected



(f) South Selected



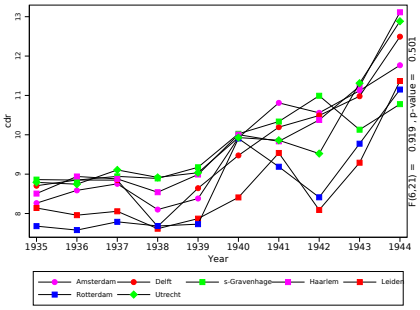
(g) South non-Selected



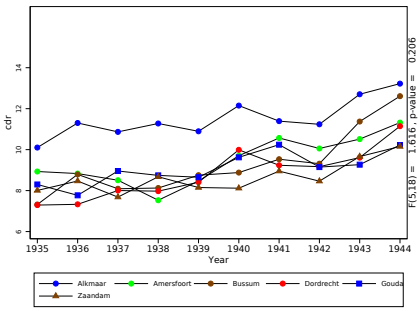
Note: Pretrends shown for groups of selected, and non-selected, in the analysis cities. A Wald test is performed in each group, by fitting a linear regression with city-specific slopes and testing jointly whether the slopes are equal.

Figure A3: Pretrends for Crude Death Rate

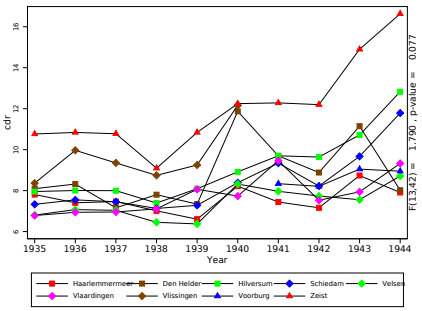
(a) West Famine



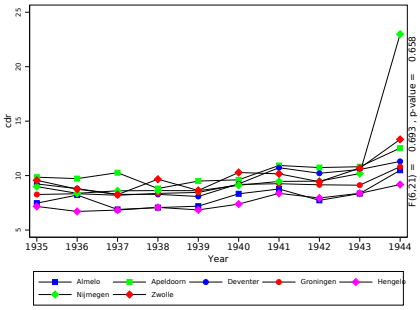
(b) West non-Famine Selected



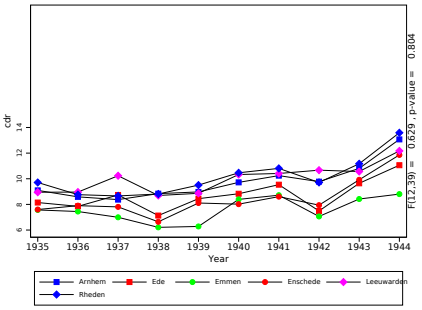
(c) West non-Famine non-Selected



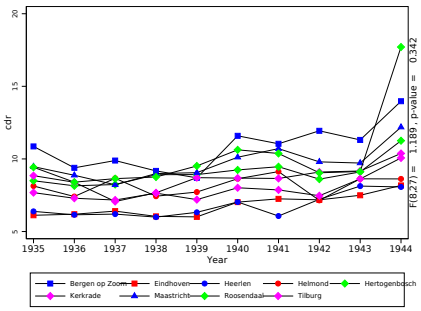
(d) North-East Selected



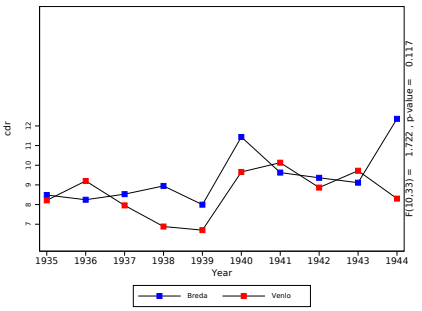
(e) North-East non-Selected



(f) South Selected



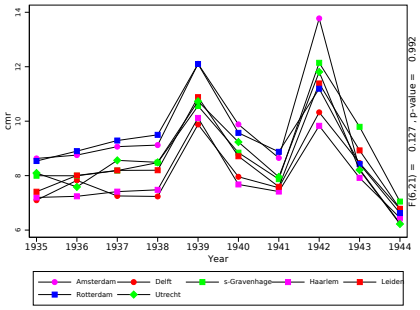
(g) South non-Selected



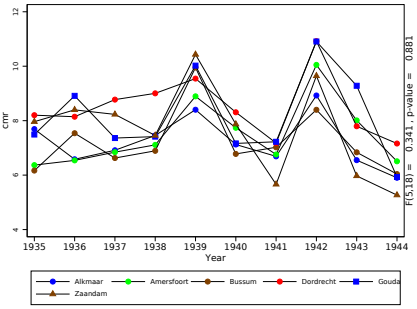
Note: Pretrends shown for groups of selected, and non-selected, in the analysis cities. A Wald test is performed in each group, by fitting a linear regression with city-specific slopes and testing jointly whether the slopes are equal.

Figure A4: Pretrends for Crude Marriage Rate

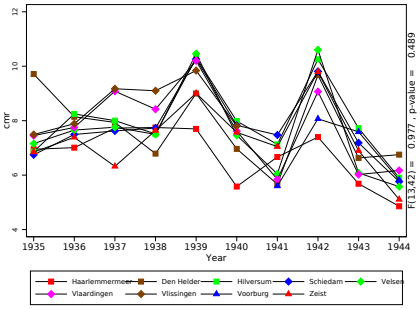
(a) West Famine



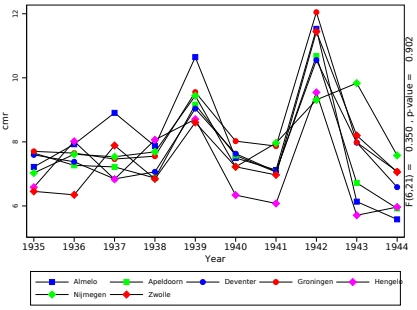
(b) West non-Famine Selected



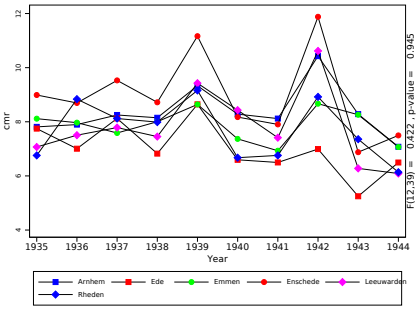
(c) West non-Famine non-Selected



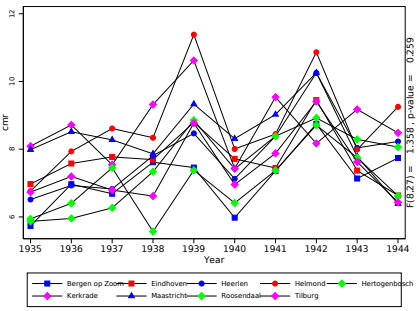
(d) North-East Selected



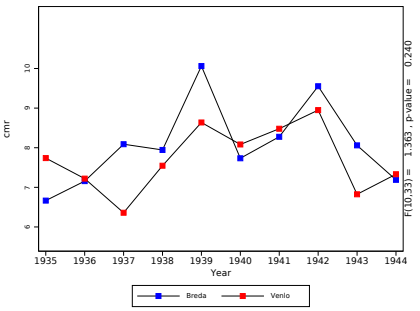
(e) North-East non-Selected



(f) South Selected



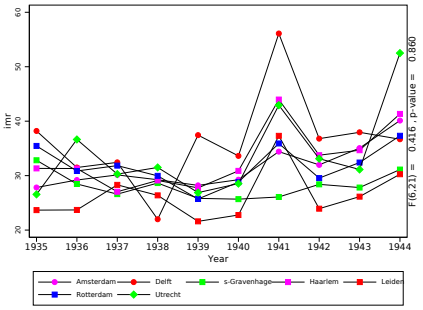
(g) South non-Selected



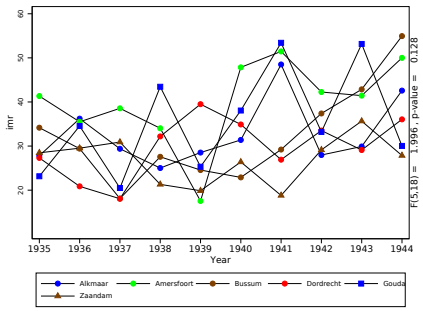
Note: Pretrends shown for groups of selected, and non-selected, in the analysis cities. A Wald test is performed in each group, by fitting a linear regression with city-specific slopes and testing jointly whether the slopes are equal.

Figure A5: Pretrends for Infant Mortality Rate

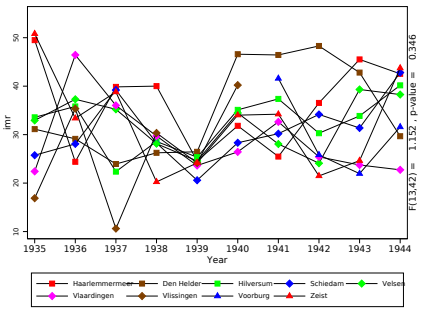
(a) West Famine



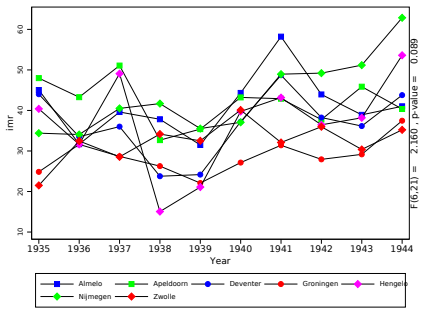
(b) West non-Famine Selected



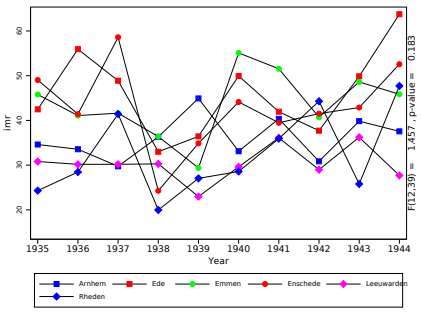
(c) West non-Famine non-Selected



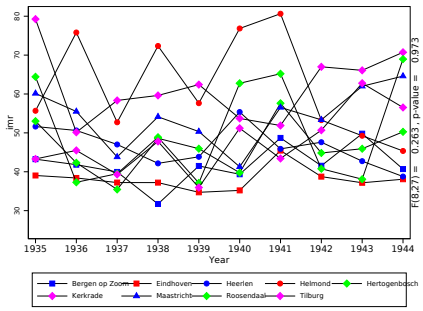
(d) North-East Selected



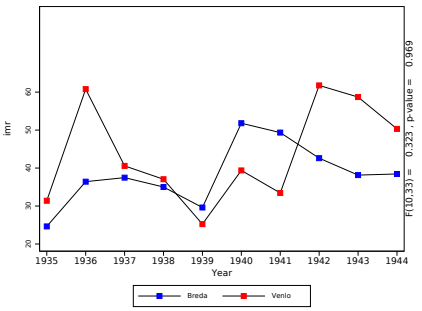
(e) North-East non-Selected



(f) South Selected



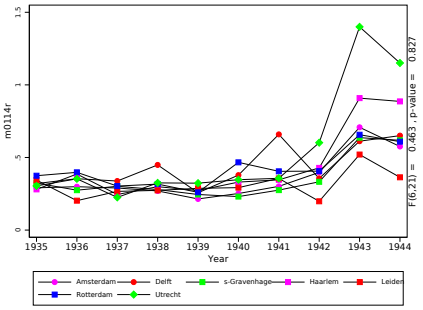
(g) South non-Selected



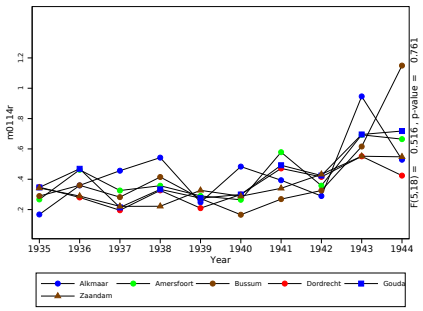
Note: Pretrends shown for groups of selected, and non-selected, in the analysis cities. A Wald test is performed in each group, by fitting a linear regression with city-specific slopes and testing jointly whether the slopes are equal.

Figure A6: Pretrends for Crude Death Rate: Age 1-14 years

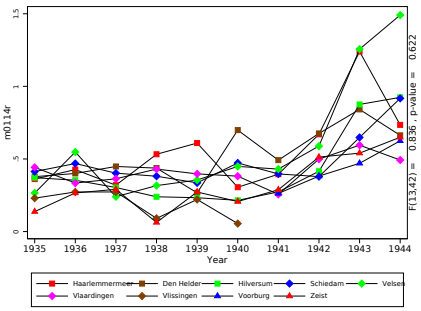
(a) West Famine



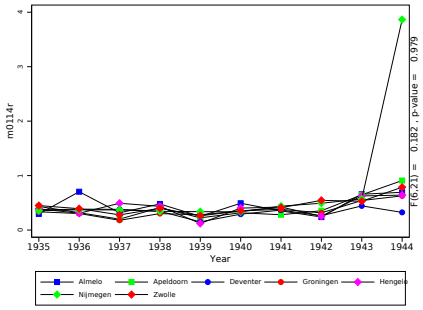
(b) West non-Famine Selected



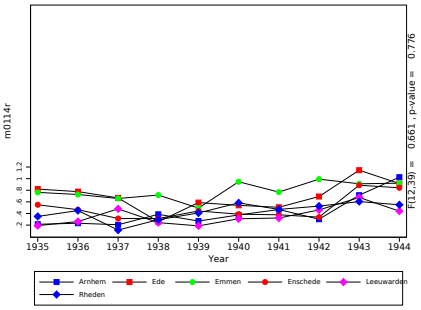
(c) West non-Famine non-Selected



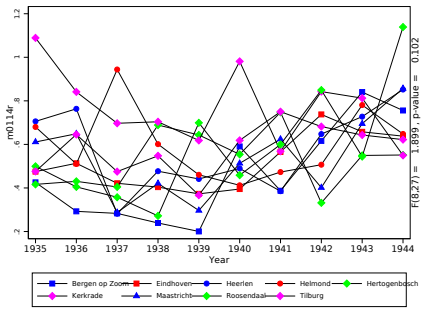
(d) North-East Selected



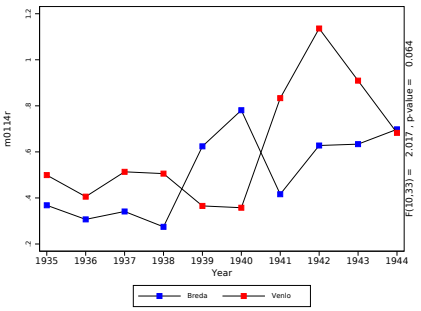
(e) North-East non-Selected



(f) South Selected



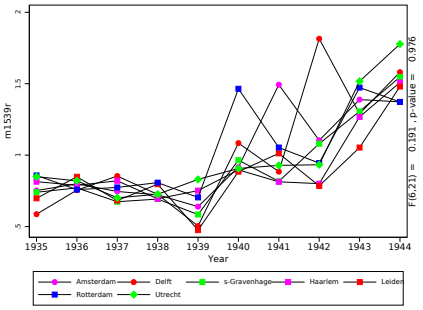
(g) South non-Selected



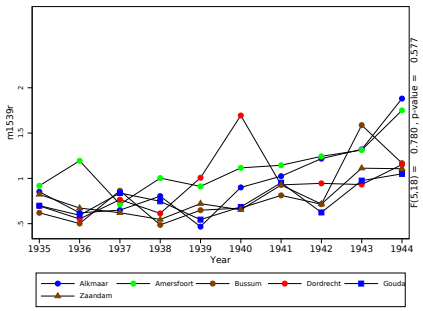
Note: Pretrends shown for groups of selected, and non-selected, in the analysis cities. A Wald test is performed in each group, by fitting a linear regression with city-specific slopes and testing jointly whether the slopes are equal.

Figure A7: Pretrends for Crude Death Rate: Age 15-39 years

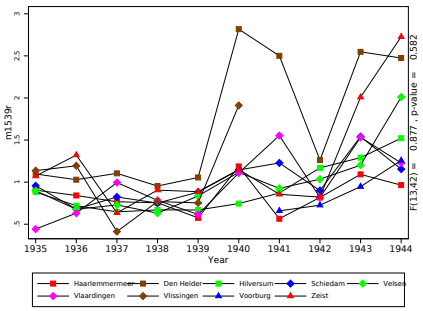
(a) West Famine



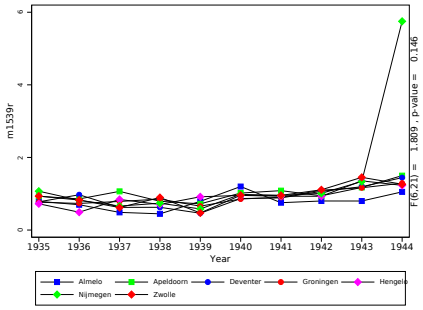
(b) West non-Famine Selected



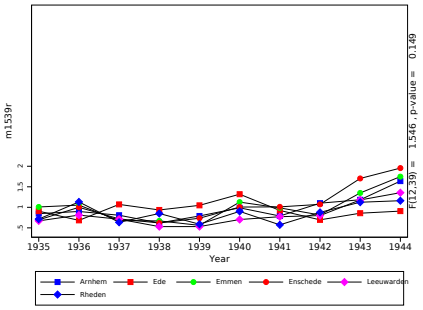
(c) West non-Famine non-Selected



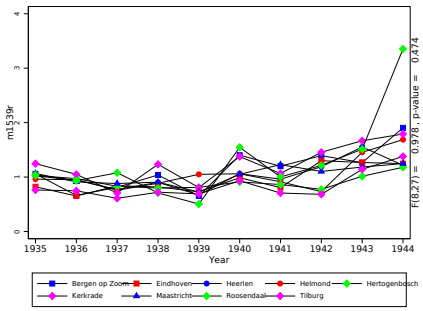
(d) North-East Selected



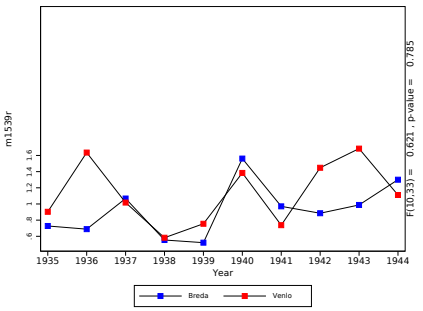
(e) North-East non-Selected



(f) South Selected



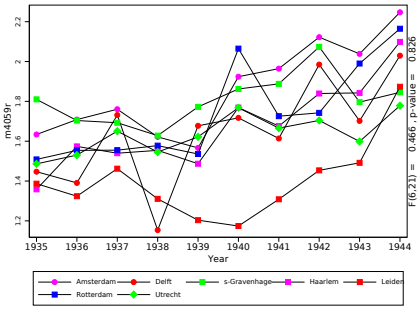
(g) South non-Selected



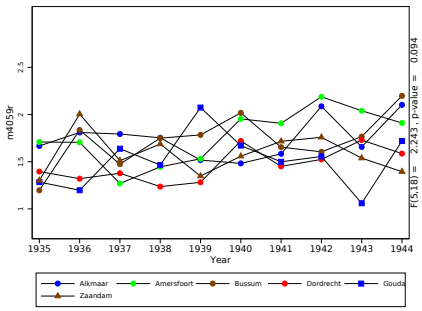
Note: Pretrends shown for groups of selected, and non-selected, in the analysis cities. A Wald test is performed in each group, by fitting a linear regression with city-specific slopes and testing jointly whether the slopes are equal.

Figure A8: Pretrends for Crude Death Rate: Age 40-59 years

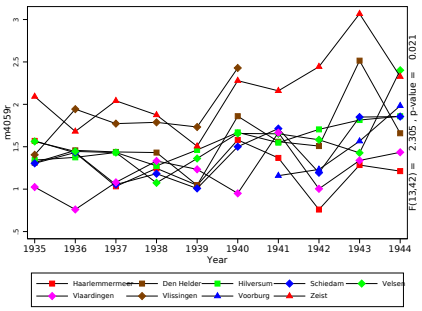
(a) West Famine



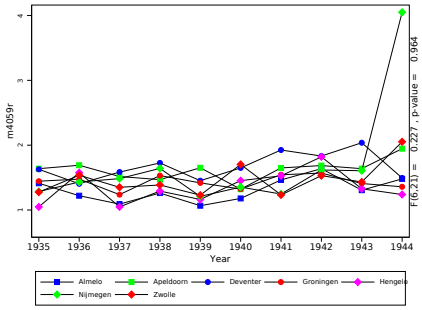
(b) West non-Famine Selected



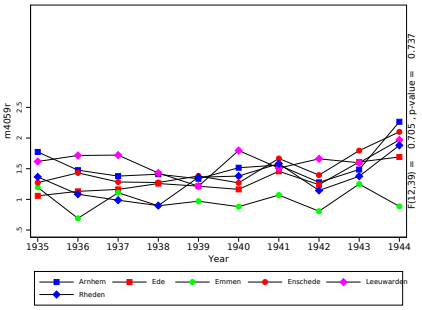
(c) West non-Famine non-Selected



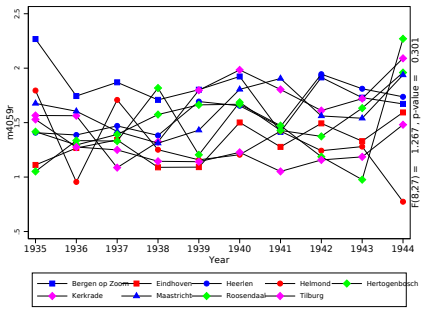
(d) North-East Selected



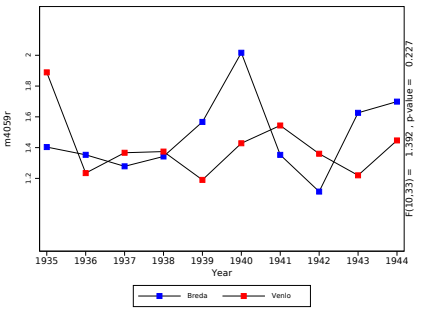
(e) North-East non-Selected



(f) South Selected



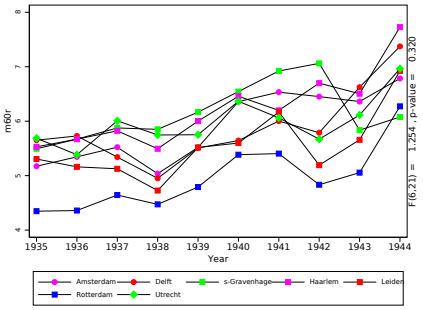
(g) South non-Selected



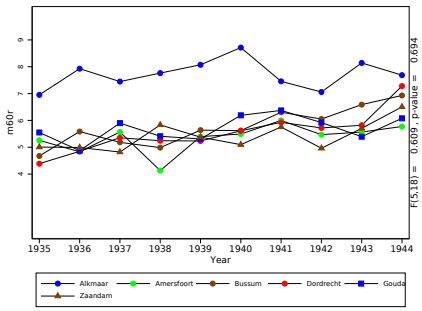
Note: Pretrends shown for groups of selected, and non-selected, in the analysis cities. A Wald test is performed in each group, by fitting a linear regression with city-specific slopes and testing jointly whether the slopes are equal.

Figure A9: Pretrends for Crude Death Rate: Age 60+ years

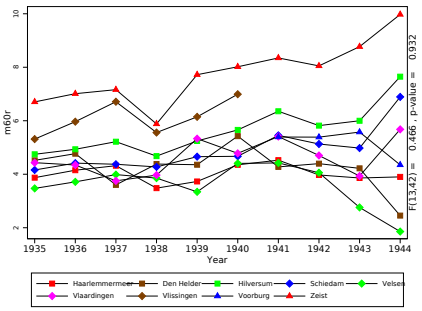
(a) West Famine



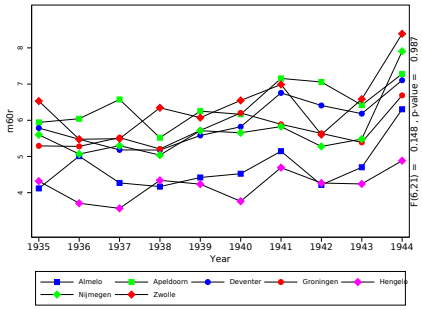
(b) West non-Famine Selected



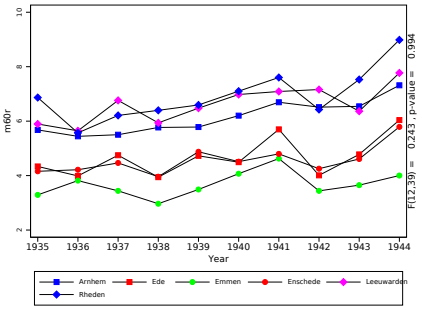
(c) West non-Famine non-Selected



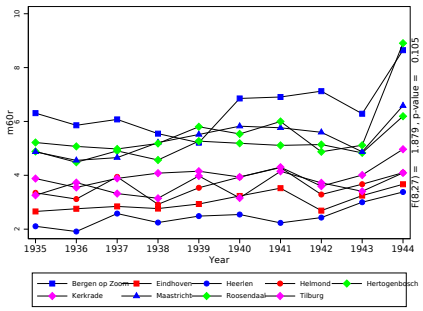
(d) North-East Selected



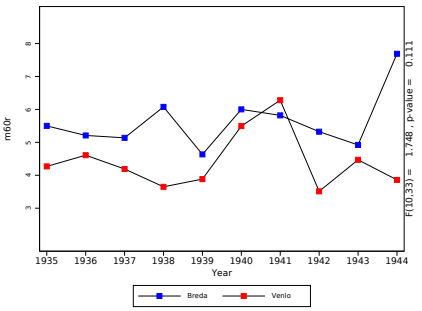
(e) North-East non-Selected



(f) South Selected



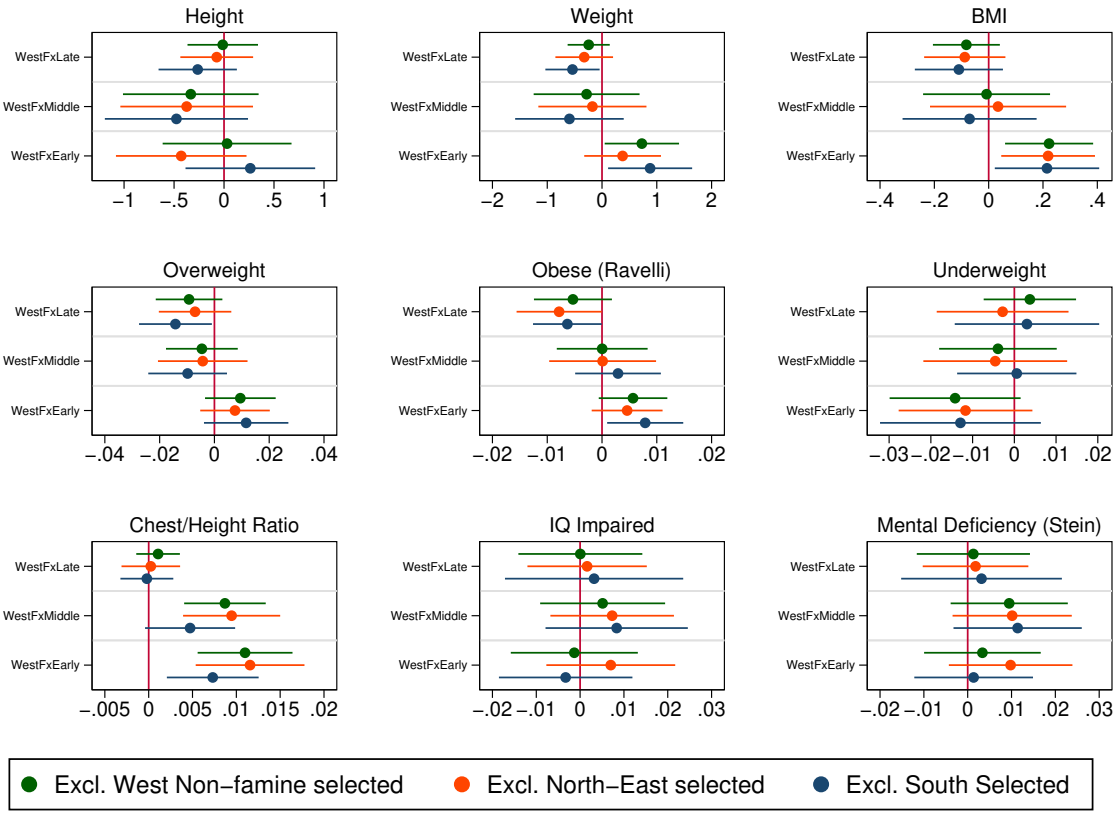
(g) South non-Selected



Note: Pretrends shown for groups of selected, and non-selected, in the analysis cities. A Wald test is performed in each group, by fitting a linear regression with city-specific slopes and testing jointly whether the slopes are equal.

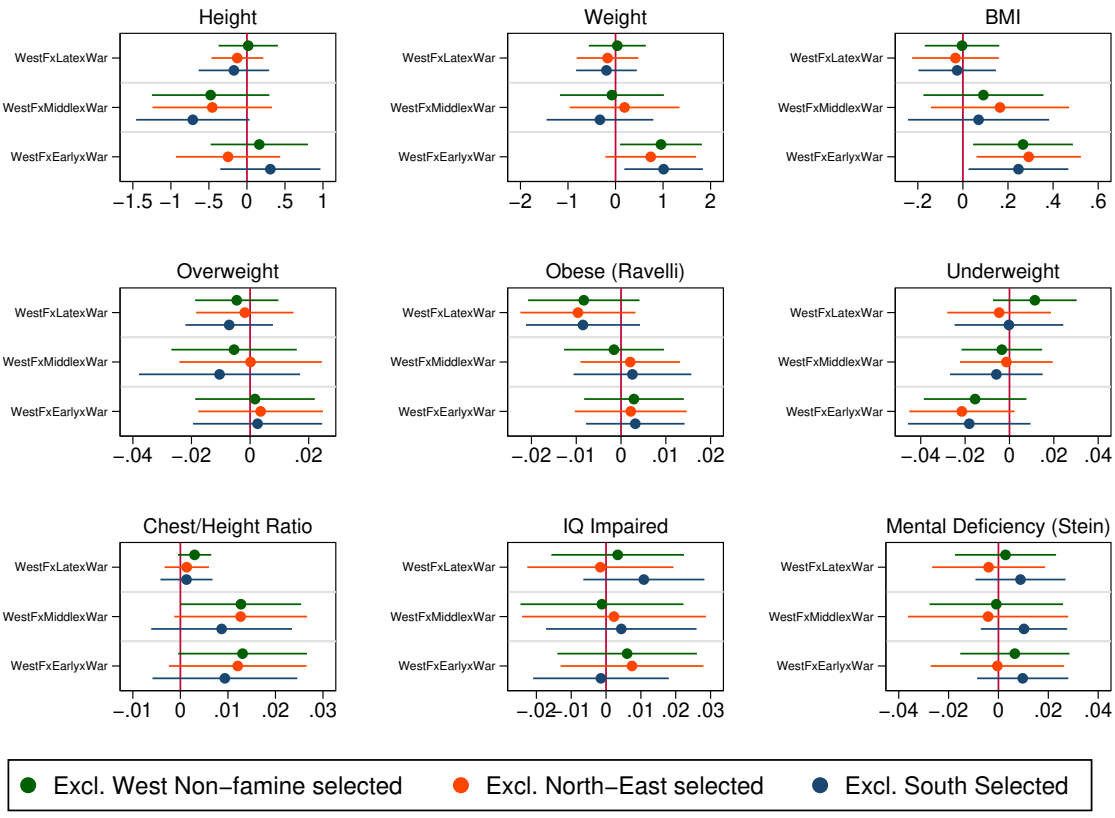
Appendix B Sensitivity Analysis: Inclusion of Regions

Figure B1: Difference-in-Differences model excluding one region at a time



Note: For each model, the three interaction term estimates are presented along with the 95% Confidence Interval. A vertical line at zero is added to ease examination of significance. Results correspond to models in Table 3.

Figure B2: Triple Difference model excluding one region at a time



Note: For each model, the three triple interaction term estimates are presented along with the 95% Confidence Interval. A vertical line at zero is added to ease examination of significance. Results correspond to models in Table 5.

Appendix C Further Tables

Table C1: Descriptive Statistics of Control Variables

Variable		Mean	Std.Dev.	Min	Max
Father's Occupation Status	Low professional	0.031	0.173	0	1
	Manager	0.107	0.309	0	1
	Clerical	0.274	0.446	0	1
	Self-employed	0.069	0.254	0	1
	Farm owner	0.017	0.128	0	1
	Shop assistant	0.285	0.451	0	1
	Personal assistant	0.044	0.204	0	1
	Miner	0.006	0.075	0	1
	Farmer	0.006	0.079	0	1
	Laborer	0.077	0.267	0	1
	Other	0.060	0.238	0	1
Older Brothers	1	0.282	0.450	0	1
	2	0.115	0.319	0	1
	3	0.034	0.180	0	1
	4	0.011	0.106	0	1
	5	0.004	0.065	0	1
	6	0.002	0.039	0	1
	7	0.001	0.024	0	1
	8	0.000	0.016	0	1
	9+	0.000	0.015	0	1
Birth Order	2	0.287	0.452	0	1
	3	0.172	0.377	0	1
	4	0.087	0.281	0	1
	5	0.044	0.206	0	1
	6	0.024	0.153	0	1
	7	0.014	0.116	0	1
	8	0.007	0.085	0	1
	9+	0.011	0.102	0	1
Religion	Catholic	0.407	0.491	0	1
	Protestant	0.275	0.447	0	1
	Jewish	0.062	0.242	0	1
	Unknown	0.007	0.081	0	1

Note: Descriptive statistics for dummy variables entering the specifications as controls, presented here for the cohorts in the analytical sample (May 1944-July 1945 in the selected cities). Number of observations for all the variables is 40,950.

Table C2: Placebo Difference-in-Differences – Ravelli et al. (1976) sample

	(1) Height	(2) Weight	(3) BMI	(4) Overweight	(5) Obese (Ravelli)	(6) Underweight	(7) Chest/Height Ratio	(8) IQ Impaired	(9) Mental Deficiency
WestF×PlaceboLate	-0.132	-0.232*	-0.040	-0.003	-0.001	-0.002	-0.000	0.001	0.001
	(0.092)	(0.134)	(0.047)	(0.004)	(0.003)	(0.005)	(0.002)	(0.005)	(0.005)
WestF×PlaceboMiddle	0.013	-0.325**	-0.106**	-0.003	0.002	0.005	-0.003	0.001	0.005
	(0.124)	(0.122)	(0.040)	(0.005)	(0.002)	(0.005)	(0.005)	(0.006)	(0.007)
WestF×PlaceboEarly	-0.046	-0.251*	-0.071*	0.004	0.001	0.007*	-0.001	-0.004	-0.001
	(0.113)	(0.148)	(0.041)	(0.006)	(0.002)	(0.004)	(0.005)	(0.004)	(0.006)
Observations	107,549	107,549	107,549	107,549	107,549	107,549	107,549	107,549	107,549
Controls and FE	✓	✓	✓	✓	✓	✓	✓	✓	✓
Control Mean	177.884	68.537	21.644	0.074	0.019	0.055	0.499	0.039	0.030

Note: Robust standard errors in parentheses clustered at the level of city. *** p<0.01, ** p<0.05, * p<0.1. Controls include father's occupation, number of older brothers, birth order and religion. FE are included for city and for month of birth. Control Mean refers to the mean of the outcome for those born in the West Famine area with postnatal exposure only. Exposure cohorts are constructed using the same structure as in the main estimation, but two years after. Outcome and *Early, Middle, Late* definitions as in Table 3.

Table C3: Mechanisms on Age 18 Outcomes

	(1) Weight	(2) BMI	(3) Overweight	(4) Obese (Ravelli)	(5) Underweight	(6) Chest/Height Ratio
log(Warfare1stTR+1)	0.028	-0.014	0.001	0.000	0.002	-0.000
	(0.066)	(0.011)	(0.001)	(0.001)	(0.001)	(0.000)
log(Warfare2ndTR+1)	0.026	0.010	0.002**	-0.000	-0.001	-0.001
	(0.049)	(0.012)	(0.001)	(0.001)	(0.001)	(0.000)
log(Warfare3rdTR+1)	0.097*	0.029**	0.000	0.001	-0.001	0.000
	(0.051)	(0.013)	(0.001)	(0.001)	(0.001)	(0.000)
-Calories1stTR	0.623	0.003	0.037	0.016	0.038	0.008**
	(0.946)	(0.236)	(0.022)	(0.014)	(0.030)	(0.004)
-ProteinShare1stTR	0.005	0.025	0.000	0.006	0.007	0.001
	(0.347)	(0.073)	(0.008)	(0.005)	(0.008)	(0.001)
-Calories2ndTR	-0.414	-0.011	-0.024	-0.016	-0.027	-0.000
	(0.813)	(0.198)	(0.017)	(0.014)	(0.016)	(0.003)
-ProteinShare2ndTR	0.185	0.044	-0.001	0.002	-0.006	0.001
	(0.151)	(0.042)	(0.004)	(0.003)	(0.005)	(0.001)
-Calories3rdTR	-1.334**	-0.326*	-0.017	0.001	0.030***	0.006**
	(0.573)	(0.188)	(0.012)	(0.006)	(0.010)	(0.003)
-ProteinShare3rdTR	0.432***	0.107***	0.008**	0.003*	-0.003	0.001**
	(0.113)	(0.021)	(0.003)	(0.002)	(0.003)	(0.000)
Wild cluster bootstrap p-values:						
log(Warfare1stTR+1)	0.717	0.187	0.732	0.720	0.178	0.802
log(Warfare2ndTR+1)	0.610	0.369	0.068	0.463	0.365	0.202
log(Warfare3rdTR+1)	0.157	0.099	0.758	0.478	0.321	0.869
-Calories1stTR	0.547	0.992	0.103	0.269	0.218	0.048
-ProteinShare1stTR	0.989	0.740	0.982	0.311	0.348	0.420
-Calories2ndTR	0.615	0.958	0.216	0.319	0.096	0.943
-ProteinShare2ndTR	0.257	0.327	0.897	0.392	0.264	0.193
-Calories3rdTR	0.072	0.144	0.185	0.880	0.003	0.044
-ProteinShare3rdTR	0.005	0.000	0.041	0.074	0.273	0.032
RW p-values:						
log(Warfare1stTR+1)	0.934	0.299	0.934	0.934	0.299	0.934
log(Warfare2ndTR+1)	0.632	0.632	0.049	0.632	0.632	0.253
log(Warfare3rdTR+1)	0.077	0.023	0.825	0.549	0.503	0.825
-Calories1stTR	0.525	0.981	0.112	0.292	0.264	0.039
-ProteinShare1stTR	0.984	0.916	0.916	0.350	0.529	0.529
-Calories2ndTR	0.715	0.988	0.189	0.236	0.153	0.988
-ProteinShare2ndTR	0.295	0.314	0.865	0.338	0.314	0.190
-Calories3rdTR	0.013	0.064	0.119	0.853	0.001	0.017
-ProteinShare3rdTR	0.003	0.000	0.013	0.032	0.144	0.032
Observations	40,950	40,950	40,950	40,950	40,950	40,950
Controls and FE	✓	✓	✓	✓	✓	✓

Note: Robust standard errors in parentheses clustered at the level of city. *** p<0.01, ** p<0.05, * p<0.1. Controls include father's occupation, number of older brothers, birth order, religion, temperature by trimester and weeks since liberation for the South. FE are included for city and for month of birth. Outcome and *Early, Middle, Late* definitions as in Table 3.

Table C4: Difference-in-Differences Results for the Birth Outcomes Females Sample

	(1) Birth Weight	(2) Low Birth Weight	(3) Birth Length	(4) Head Circumference	(5) Placenta Weight	(6) Gestational Age
West×Late	-143.154	0.074	-0.277	-0.341	11.843	-0.306
	(94.673)	(0.045)	(0.480)	(0.424)	(29.616)	(0.370)
West×Middle	-196.852**	0.061	-1.094***	-0.153	1.576	-0.691**
	(83.020)	(0.039)	(0.404)	(0.359)	(24.746)	(0.345)
West×Early	-98.478	0.030	-0.426	0.203	8.035	-0.274
	(84.575)	(0.039)	(0.397)	(0.363)	(25.062)	(0.343)
RW *p*-values:						
West×Late	0.456	0.411	0.860	0.860	0.860	0.860
West×Middle	0.073	0.298	0.037	0.871	0.955	0.170
West×Early	0.732	0.866	0.755	0.866	0.866	0.866
$F_{West \times Middle = West \times Early}$	2.433	0.828	4.711	1.785	0.126	1.645
p-value	0.119	0.363	0.030	0.182	0.895	0.200
$F_{West \times Late = West \times Middle = West \times Early}$	1.221	0.657	2.998	1.437	0.111	0.965
p-value	0.295	0.519	0.050	0.238	0.722	0.381
Observations	1,766	1,766	1,700	1,203	1,230	1,162
Control and FE	✓	✓	✓	✓	✓	✓
Control Mean	3,271.69	0.075	50.026	38.46	574.107	39.519

Note: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. All models include mother's age and FE for city and for month of birth. Control Mean refers to the mean of the outcome for those born in the West Famine area with postnatal exposure only. Results from hospital records in the birth data. *Early, Middle, Late* definitions as in Table 3.

Table C5: Difference-in-Differences Birth for Sex ratio

	(1) Male
West×Late	0.030
	(0.057)
West×Middle	0.052
	(0.052)
West×Early	0.004
	(0.053)
Observations	3,697
Control Mean	0.531

Note: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. All models include mother's age and FE for city and for month of birth. Control Mean refers to the mean of the outcome for those born in the West Famine area with postnatal exposure only. Results from hospital records in the birth data. *Early, Middle, Late* definitions as in Table 3.

Table C6: Selection equation AMEs of copula selection models

	(1) Height	(2) Weight	(3) BMI	(4) Overweight	(5) Obese (Ravelli)	(6) Underweight	(7) Chest/Height Ratio	(8) IQ Impaired	(9) Mental Deficiency
WestF×Late	0.003	-0.001	-0.004	0.003	0.007	0.011	-0.002	0.022	0.026
	(0.014)	(0.013)	(0.013)	(0.020)	(0.017)	(0.020)	(0.014)	(0.020)	(0.021)
WestF×Middle	-0.011	-0.017	-0.021	-0.017	-0.012	-0.007	-0.018	0.010	0.018
	(0.017)	(0.015)	(0.016)	(0.023)	(0.019)	(0.022)	(0.016)	(0.022)	(0.021)
WestF×Early	-0.029***	-0.034***	-0.036***	-0.025**	-0.036***	-0.047***	-0.034***	-0.017	-0.010
	(0.010)	(0.010)	(0.010)	(0.012)	(0.011)	(0.014)	(0.011)	(0.013)	(0.017)
Medical1930	-0.073***	-0.057***	-0.072***	-0.139***	-0.098***	-0.071***	-0.033**	-0.096***	-0.150***
	(0.010)	(0.008)	(0.008)	(0.014)	(0.010)	(0.013)	(0.015)	(0.015)	(0.019)
Largest1930	0.049***	0.039***	0.047***	0.090***	0.066***	0.053***	0.026***	0.066***	0.096***
	(0.007)	(0.006)	(0.006)	(0.010)	(0.007)	(0.009)	(0.009)	(0.010)	(0.012)
Agriculture1930	0.144***	0.110***	0.135***	0.269***	0.194***	0.140***	0.063**	0.183***	0.281***
	(0.020)	(0.016)	(0.016)	(0.028)	(0.020)	(0.027)	(0.030)	(0.030)	(0.035)
Observations	41,645	41,645	41,645	41,645	41,645	41,645	41,639	41,645	41,645
FE	✓	✓	✓	✓	✓	✓	✓	✓	✓

Note: Robust standard errors in parentheses clustered at the level of city. *** p<0.01, ** p<0.05, * p<0.1. Average Marginal Effects calculated for first-stage (selection) equation estimated jointly with the results in Table 9. FE are included for city and for month of birth. Outcome and *Early, Middle, Late* definitions as in Table 3.

Table C7: Selection copula models with different variables used as exclusion restrictions

	(1) Height	(2) Weight	(3) BMI	(4) Overweight	(5) Obese (Ravelli)	(6) Underweight	(7) Chest/Height Ratio	(8) IQ Impaired	(9) Mental Deficiency
Incl: Medical1930 & Largest1930									
WestF×Late	-0.089	-0.351*	-0.068	0.000	-0.002	0.004	0.001	0.004	0.007
	(0.157)	(0.197)	(0.057)	(0.004)	(0.003)	(0.004)	(0.001)	(0.006)	(0.005)
WestF×Middle	-0.287	-0.594	-0.11	0.000	0.005	0.008**	0.008***	0.011**	0.014***
	(0.310)	(0.529)	(0.116)	(0.005)	(0.003)	(0.004)	(0.002)	(0.005)	(0.004)
WestF×Early	0.047	0.065	0.052	0.010**	0.010***	0.004	0.010***	0.010*	0.012**
	(0.323)	(0.311)	(0.059)	(0.004)	(0.003)	(0.004)	(0.003)	(0.006)	(0.005)
Kendall's τ	-0.132	0.441	0.478	-0.697	-0.672	-0.712	0.268	-0.613	-0.665
Wald test	6.29	836.90	481.74	3592.70	284.61	2077.66	14.26	475.61	129.173
p-value	0.012	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Incl: Medical1930 & Agriculture1930									
WestF×Late	-0.089	-0.346*	-0.065	0.000	-0.002	0.004	0.001	0.004	0.007
	(0.159)	(0.192)	(0.057)	(0.004)	(0.003)	(0.004)	(0.001)	(0.006)	(0.005)
WestF×Middle	-0.286	-0.587	-0.107	0.000	0.004	0.007*	0.008***	0.011**	0.013***
	(0.310)	(0.533)	(0.118)	(0.005)	(0.003)	(0.004)	(0.002)	(0.005)	(0.004)
WestF×Early	0.055	0.062	0.051	0.010**	0.010***	0.004	0.010***	0.010*	0.012**
	(0.327)	(0.308)	(0.059)	(0.004)	(0.003)	(0.004)	(0.003)	(0.006)	(0.005)
Kendall's τ	-0.143	0.442	0.481	-0.697	-0.672	-0.713	0.267	-0.614	-0.666
Wald test	6.50	877.41	531.99	3427.14	273.02	2008.43	13.94	495.56	131.498
p-value	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Incl: Largest1930 & Agriculture1930									
WestF×Late	-0.095	-0.324*	-0.062	0.000	-0.002	0.003	0.001	0.004	0.007
	(0.156)	(0.196)	(0.057)	(0.004)	(0.003)	(0.004)	(0.001)	(0.006)	(0.005)
WestF×Middle	-0.291	-0.606	-0.113	0.000	0.005	0.008*	0.008***	0.011**	0.014***
	(0.310)	(0.525)	(0.115)	(0.005)	(0.003)	(0.004)	(0.002)	(0.005)	(0.004)
WestF×Early	0.040	0.053	0.049	0.010**	0.010***	0.004	0.010***	0.011*	0.012**
	(0.321)	(0.306)	(0.058)	(0.004)	(0.003)	(0.004)	(0.003)	(0.006)	(0.005)
Kendall's τ	-0.121	0.441	0.479	-0.697	-0.672	-0.712	0.260	-0.614	-0.666
Wald test	7.65	703.80	486.92	3212.89	292.80	2158.05	14.41	484.49	135.126
p-value	0.006	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Incl: Medical1930									
WestF×Late	-0.087	-0.354*	-0.067	0.000	-0.002	0.004	0.001	0.004	0.007
	(0.158)	(0.196)	(0.057)	(0.004)	(0.003)	(0.004)	(0.001)	(0.006)	(0.005)
WestF×Middle	-0.286	-0.587	-0.108	0.000	0.005	0.007*	0.008***	0.011**	0.013***
	(0.310)	(0.534)	(0.118)	(0.005)	(0.003)	(0.004)	(0.002)	(0.005)	(0.004)
WestF×Early	0.055	0.064	0.052	0.010**	0.010***	0.004	0.010***	0.010*	0.012**
	(0.327)	(0.308)	(0.059)	(0.004)	(0.003)	(0.004)	(0.003)	(0.006)	(0.005)
Kendall's τ	-0.142	0.442	0.481	-0.697	-0.672	-0.713	0.268	-0.614	-0.666
Wald test	6.47	876.41	536.84	3559.28	275.70	2016.49	14.24	487.52	131.563
p-value	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Incl: Largest1930									
WestF×Late	-0.092	-0.343*	-0.066	0.000	-0.002	0.004	0.001	0.004	0.007
	(0.156)	(0.197)	(0.058)	(0.004)	(0.003)	(0.004)	(0.001)	(0.006)	(0.005)
WestF×Middle	-0.290	-0.602	-0.112	0.000	0.005	0.008*	0.008***	0.011**	0.014***
	(0.310)	(0.527)	(0.116)	(0.005)	(0.003)	(0.004)	(0.002)	(0.005)	(0.004)
WestF×Early	0.042	0.057	0.050	0.010**	0.010***	0.004	0.010***	0.011*	0.012**
	(0.323)	(0.307)	(0.058)	(0.004)	(0.003)	(0.004)	(0.003)	(0.006)	(0.005)
Kendall's τ	-0.124	0.442	0.479	-0.697	-0.672	-0.712	0.261	-0.613	-0.666
Wald test	7.39	690.12	454.65	3392.83	294.68	2148.51	15.03	480.60	134.823
p-value	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Incl: Agriculture1930									
WestF×Late	-0.091	-0.338*	-0.064	0.000	-0.002	0.003	0.001	0.004	0.007
	(0.157)	(0.193)	(0.057)	(0.004)	(0.003)	(0.004)	(0.001)	(0.006)	(0.005)
WestF×Middle	-0.290	-0.595	-0.110	0.000	0.004	0.007*	0.008***	0.011**	0.013***
	(0.310)	(0.531)	(0.117)	(0.005)	(0.003)	(0.004)	(0.002)	(0.005)	(0.004)
WestF×Early	0.049	0.053	0.049	0.010**	0.010***	0.004	0.010***	0.010*	0.012**
	(0.328)	(0.302)	(0.058)	(0.004)	(0.003)	(0.004)	(0.003)	(0.006)	(0.005)
Kendall's τ	-0.133	0.444	0.482	-0.697	-0.673	-0.713	0.260	-0.615	-0.668
Wald test	7.79	726.13	525.98	3270.40	279.71	2094.98	14.59	500.44	136.539
p-value	0.005	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Note: Robust standard errors in parentheses clustered at the level of city. *** p<0.01, ** p<0.05, * p<0.1. These results correspond to the results in Table 9 with different variables used as exclusion restrictions in the selection equation. Outcome and *Early, Middle, Late* definitions as in Table 3.

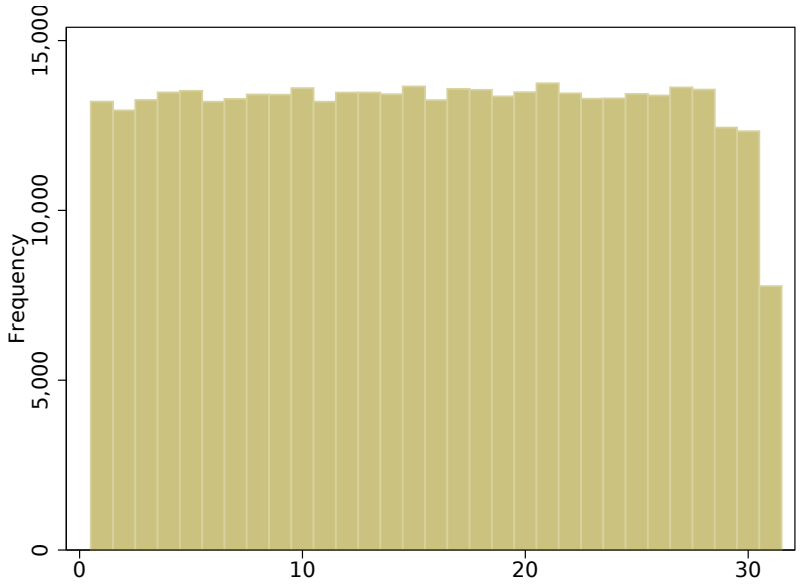
Table C8: Difference-in-Differences on Age 18 Outcomes with IPW

	(1) Height	(2) Weight	(3) BMI	(4) Overweight	(5) Obese (Ravelli)	(6) Underweight	(7) Chest/Height Ratio	(8) IQ Impaired	(9) Mental Deficiency
WestF×Late	-0.164	-0.348*	-0.076	-0.009	-0.006*	0.000	0.001	0.003	0.003
	(0.143)	(0.203)	(0.058)	(0.005)	(0.003)	(0.006)	(0.001)	(0.007)	(0.006)
WestF×Middle	-0.422	-0.425	-0.032	-0.007	0.001	-0.002	0.008***	0.007	0.011*
	(0.307)	(0.457)	(0.109)	(0.006)	(0.004)	(0.007)	(0.002)	(0.006)	(0.006)
WestF×Early	-0.120	0.612*	0.220***	0.008	0.006*	-0.015**	0.010***	-0.000	0.004
	(0.315)	(0.334)	(0.076)	(0.006)	(0.003)	(0.007)	(0.003)	(0.007)	(0.006)
Observations	41,639	41,639	41,639	41,639	41,639	41,639	41,639	41,639	41,639
Controls and FE	✓	✓	✓	✓	✓	✓	✓	✓	✓

Note: Robust standard errors in parentheses clustered at the level of city. *** p<0.01, ** p<0.05, * p<0.1. Controls include father's occupation, number of older brothers, birth order, and religion. FE are included for city and for month of birth. Weights are constructed as $1/Pr(Alive\ at\ 18)$ defined for each city and birth month. These results correspond to the results in Table 3 without selection. Outcome and *Early, Middle, Late* definitions as in Table 3.

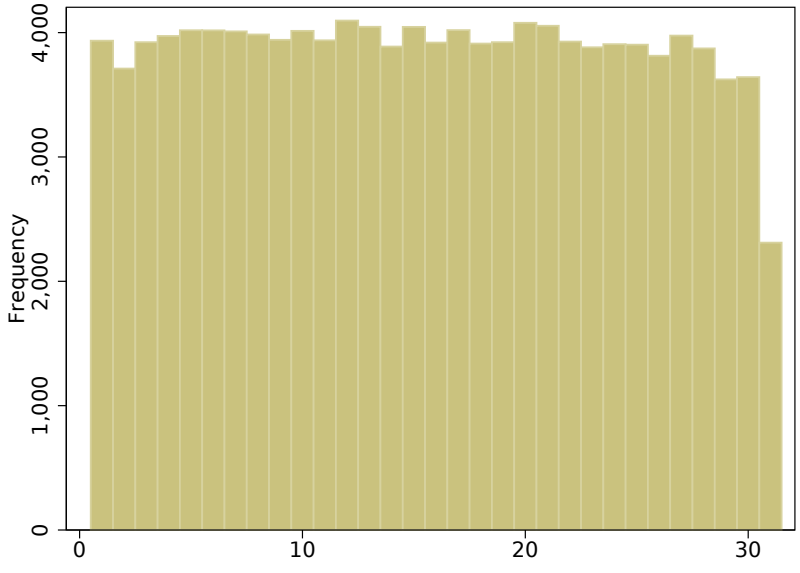
Appendix D Measurement Error

Figure D1: Births by day of month (Jan1944-Dec1947) All cities



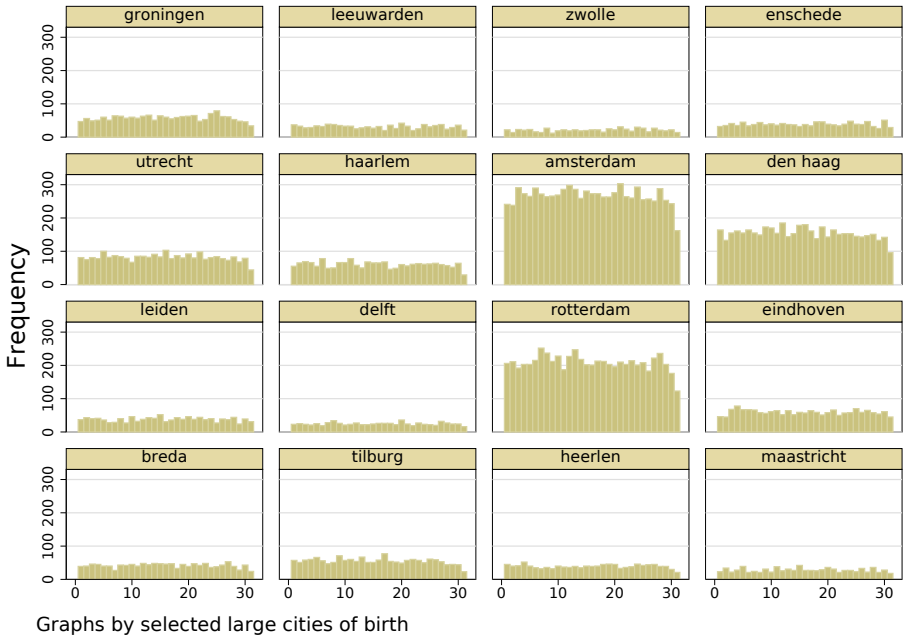
Data source: Military recruits data.

Figure D2: Births by day of month (May1944-July1945) All cities



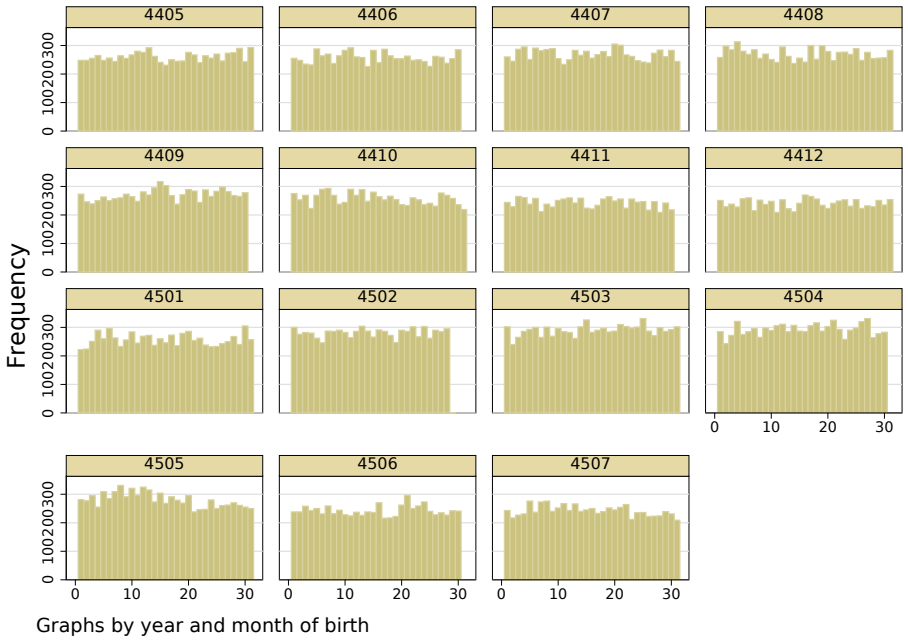
Data source: Military recruits data.

Figure D3: Births by day of month and by city (May1944-July1945) for the 16 largest cities



Data source: Military recruits data.

Figure D4: Births by day of month - Recruits by Month (All cities)



Data source: Military recruits data.

Appendix E Simulation Study

We conduct the following simulation study. For $j = 1, ..., n_j$ and $i = 1, ..., n_i$, we have a sample size of $n_j \times n_i$ observations clustered in n_j regions. The variable that we are interested to estimate its effect is constructed as $f_j \sim Bernoulli(0.5)$. We construct the error terms to have a Clayton copula structure (see Winkelmann (2012) for a construction method) with dependence parameters θ_u and θ_ϵ, for regional level and individual level dependence, respectively. Depending on whether we impose independence or not, the parameters are taking values of either 2 or 0. The survival is constructed as:

$$D_{ij} = I(1 - 0.8f_j + 0.3x_{ij} + \gamma z_j + u_{1j} + \epsilon_{1ij} > 0)$$

where f_j is a binary variable for famine (1 if famine, 0 otherwise), x_{ij} is some individual-level exogenous variable, and z_j is a region-level exogenous variables. We estimate the probability of survival at regional level as $P_j = \frac{\sum_i D_{ij}}{n_j}$ (in all simulations, n_j is constant across regions). The individual-level variable ($x_{ij} \sim N(0, 2)$) is included in both equations, and the region-level ($z_j \sim N(0, 1)$) is included only in the selection equation. Finally, the outcome equation of interest is:

$$y_{ij} = 1x_{ij} - 1f_j + u_{2j} + \epsilon_{2ij}$$

Table E1 shows the results for 100 regions with 100 individuals in each region from six different DGPs. Three of them are using an exclusion restriction that has a weak relationship with survival ($\gamma = 0.05$), and three with a strong relationship ($\gamma = 0.3$). The first panel shows the case with error correlation only at individual level, the middle panel only at region level, and the bottom panel at both levels. For each scenario, we estimate the outcome equation with OLS before setting the outcome to missing when $D_{ij} = 0$ (FULL) and after (OLS). We also perform an IPW estimation using the calculated P_j described above (IPW) and a GLS estimation (GLS). Finally, we estimate the traditional sample selection model using full maximum likelihood assuming normality in the error terms (HECK) and the sample

selection model using the Clayton copula for the dependence of the error terms (COP).

Looking at the results in Table E1 it is evident that the naive OLS estimator is always biased. In the case where the error-term correlation exists only at region level, the IPW and GLS estimators are unbiased – as expected – but not in the case where there is correlation at the individual level. In fact, IPW and GLS are performing almost identically across all scenarios. In the case of correlation only at region level they are also efficient. Finally, in the case of a non-zero correlation in the individual error terms, the best behaved estimators are the HECK and COP. The superiority of COP can be seen in the variance of the estimators, as it is the efficient one. The Standard Deviation is reduced by around 20% across all the scenarios presented in the Table. Moreover, the magnitude of the effect of z is affecting the means of both estimators, since both are closer to the true value 1 when γ increases.

Table E1: Monte Carlo Results with Clayton copula as DGP

Estimator	Weak: $\gamma = 0.05$		Strong: $\gamma = 0.3$	
	Mean	Std.Dev.	Mean	Std.Dev.
$\sigma_u > 0$ and $\sigma_e = 0$				
FULL	-1.000	0.198	-1.002	0.205
OLS	-0.879	0.179	-0.887	0.192
IPW	-0.971	0.192	-0.967	0.201
GLS	-0.978	0.192	-0.975	0.201
HECK	-0.874	0.264	-0.972	0.259
COP	-0.922	0.192	-0.942	0.205
$\sigma_u = 0$ and $\sigma_e > 0$				
FULL	-0.999	0.205	-0.999	0.205
OLS	-0.877	0.223	-0.881	0.224
IPW	-0.834	0.211	-0.837	0.210
GLS	-0.830	0.211	-0.833	0.210
HECK	-0.952	0.295	-0.974	0.292
COP	-0.951	0.228	-0.959	0.231
$\sigma_u > 0$ and $\sigma_e > 0$				
FULL	-1.000	0.199	-0.998	0.202
OLS	-0.754	0.164	-0.761	0.172
IPW	-0.802	0.166	-0.803	0.173
GLS	-0.808	0.166	-0.809	0.173
HECK	-0.909	0.287	-1.010	0.244
COP	-0.935	0.186	-0.943	0.191

Note:Results based on sample size of 10,000 and 2,000 replications.