Journal Pre-proof

Severe prenatal shocks and adolescent health: Evidence from the Dutch Hunger Winter

Gabriella Conti, Stavros Poupakis, Peter Ekamper, Govert E. Bijwaard, L.H. Lumey

PII: S1570-677X(24)00024-8
DOI: https://doi.org/10.1016/j.ehb.2024.101372
Reference: EHB 101372

To appear in: *Economics and Human Biology*

Received date : 7 June 2023
Revised date : 23 February 2024
Accepted date : 1 March 2024

Please cite this article as: G. Conti, S. Poupakis, P. Ekamper et al., Severe prenatal shocks and adolescent health: Evidence from the Dutch Hunger Winter. *Economics and Human Biology* (2024), doi: https://doi.org/10.1016/j.ehb.2024.101372.

Severe Prenatal Shocks and Adolescent Health: Evidence from the Dutch Hunger Winter

Gabriella Conti, Stavros Poupakis, Peter Ekamper, Govert E. Bijwaard and L.H. Lumey

December 3, 2023

Abstract

This paper investigates health impacts at the end of adolescence of prenatal exposure to multiple shocks, by exploiting the unique natural experiment of the Dutch Hunger Winter. At the end of World War II, a famine occurred abruptly in the Western Netherlands (November 1944 - May 1945), pushing the previously and subsequently well-nourished Dutch population to the brink of starvation. We link high-quality military recruits data with objective health measurements for the cohorts born in the years surrounding WWII with newly digitised historical records on calories and nutrient composition of the war rations, daily temperature, and warfare deaths. Using difference-in-differences and triple differences research designs, we first show that the cohorts exposed to the Dutch Hunger Winter since early gestation have a higher Body Mass Index and an increased probability of being obese at age 18. We then find that this effect is partly moderated by warfare exposure and a reduction in energy-adjusted protein intake. Lastly, we account for selective mortality using a copula-based approach and newly-digitised data on survival rates, and find evidence of both selection and scarring effects. These results emphasise the complexity of the mechanisms at play in studying the consequences of early conditions.

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

*G. Conti: Department of Economics and Social Research Institute, University College London; Institute for Fiscal Studies, CEPR and IZA. S. Poupakis: Institute for Global Health, University College London. P. Ekamper and G. Bijwaard: Netherlands Interdisciplinary Demographic Institute. L.H. Lumey: Department of Epidemiology, Mailman School of Public Health, Columbia University. We thank participants to the NBER Cohort Studies Meeting Spring 2019, to the 23rd Annual Conference of the Italian Health Economics Association and to the University of Strathclyde for their comments. This work was supported by the NIA/NIH grant "Prenatal Under Nutrition and Mortality Through Age 63" (R01AG028593-01A2, PI:LHL) where GC, PE, GB are co-Is. GC and SP have received funding from European Union's HORIZON 2020 research and innovation programme (grant no. 633595 DynaHEALTH). GC has also received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme grant agreement No 819752 - DEVORHBIOSHIP - ERC-2018COG. Authors' contributions: GC and SP: research design, data analysis and paper writing; PE: 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: Gabriella Conti, University College London, Department of Economics, Drayton House, 30 Gordon Street, London WC1H 0AX, United Kingdom. Tel.:+44(0)2076794696. Email: gabriella.conti@ucl.ac.uk.

1 Introduction

The study of the long-term health effects of early life conditions has been at the core of life course epidemiology since its origins in the 1990s (Ben-Shlomo and Kuh, 2002). In the last two decades, the economic literature has substantially contributed to the field, going beyond associational evidence to show causal impacts of early circumstances not only on health, but also on a wide variety of socioeconomic outcomes (see the reviews of Currie and Almond (2011), Almond et al. (2018), Prinz et al. (2018) and Conti et al. (2019)). While the field has been burgeoning, knowledge of the mechanisms through which early conditions affect life course outcomes is still scarce. Among the shocks most widely studied, prenatal malnutrition has often been proxied by exposure to a famine, caused by a war or other circumstances that affect entire cohorts in specific regions for limited periods of time.[1] One of the most studied is the Dutch famine (also known as the "Hunger Winter"), which occurred at the end of World War II (November 1944 - May 1945) and was brief, unanticipated, and temporally and regionally defined.[2]

Previous work in both epidemiology and economics has convincingly shown lifelong health and socio-economic impacts of prenatal exposure to famines, starting with the influential paper by Ravelli et al. (1976) on the Dutch Hunger Winter; however, few studies have tackled two key issues we focus on, mostly because of data limitations. First, existing studies have mostly used a difference-in-differences methodology, comparing cohorts born in different periods in regions with varying exposure to the famine; hence, the relative importance of the multiple, co-occurring shocks which affected these cohorts – low-calorie and imbalanced diet, harsh weather conditions, warfare, to mention a few – has not been ascertained. As Van den Berg and Lindeboom (2018) note 'If this point is ignored, this affects the interpretation of

[1] Famine exposure is not the only design used to study the impact of prenatal malnutrition. For example, the fasting observed during the Ramadan by the Muslim population has been also studied, with mixed results (Almond and Mazumder, 2011; Jürges, 2015; Majid, 2015).

[2] Van den Berg and Lindeboom (2018), in their critical review of the literature on famines and health, describe it as the 'textbook example'. Havari and Peracchi (2017) show that there was a sharp increase in self-reported experiences of hunger in the Netherlands at the end of 1944.

the famine as an effect through undernutrition'. Second, most previous work has studied the impacts of the famine on the survivors, with few studies investigating (albeit all acknowledging) the role played by selection effects. Third, we add to the small but growing group of studies analyzing the "missing middle", i.e. the period between the early shock and the adult outcomes.

In this paper, we re-examine the original military recruits data used in the first, influential epidemiological papers on the Dutch Hunger Winter (Ravelli et al. (1976) and Stein et al. (1972)), and we combine them with newly collected and digitised data from historical sources to provide a more comprehensive analysis of mid-term impacts of multiple, co-occurring severe prenatal shocks. The outcomes we study were objectively measured during the routine military visits at age 18 and include: height and weight, which we use to compute indicators of underweight and overweight/obesity; chest circumference, which we use to construct a measure of body build (chest-height ratio); and intellectual disability (ICD-6 325).[3] The novel data we have collected for this paper include: original information from the war rations on calories and nutrient composition of the diet; average daily temperature; number of deaths due to warfare; city-level demographic and socioeconomic characteristics; number of stillbirths, neonatal and post-neonatal deaths, and mortality until age 18.

Using these data, we contribute to the literature in the following ways. First, we use newly digitised data on the pre-war years to carefully select the cities in the control group, on the basis of common trends along a variety of health indicators, rather than solely on the basis of geographical location.[4] Second, we account for selective fertility, by focusing on the sample of women who were already pregnant at the start of the famine. Third, in addition to our baseline difference-in-differences specification, we perform several robustness and placebo tests, including a triple difference design. Fourth, we use the newly collected information

[3] Another study using objective health measurements in studying the long-term health impacts of early-life food deprivation (measured with self-reported hunger episodes) during WWII is Mink et al. (2020) for the case of France.

[4] As we explain in a later section, we are unable to test for parallel trends in the outcomes themselves, due to the unavailability of our main datasource in pre-war periods.

on caloric content and nutrient composition of the rations, average daily temperature, and number of deaths due to warfare, to understand the impacts of the different severe shocks by which the cohorts in utero during the 1944-45 winter in West Netherlands were hit.[5] Fifth, we account for selective survival, combining information on the number of stillbirths, neonatal and post-neonatal deaths, and mortality until 18 years of age (derived from newly digitised records), with historical pre-war data on local occupational structure and availability of health care resources, within a flexible copula selection model. Throughout, we take good care of the inference by accounting for the moderate number of cities and for multiple hypothesis testing.

We provide a variety of results. We first confirm previous epidemiological findings that being exposed to the famine since the prenatal period is more harmful than being exposed only postnatally; most of the negative health impacts are concentrated among those exposed since early gestation.[6] Affected cohorts have, on average, a significantly higher weight and a larger chest circumference, but not a taller height: hence, they have a higher prevalence of obesity, by 0.6 p.p. in the basic specification, which amounts to a 43% increase with respect to the control group. The choice of the control group matters: we show that the results obtained based on a control group of cities more broadly defined (as done in previous work) fail a simple placebo test. We then show that the obesity results are partly moderated by a change in the

[5] As noted in Van den Berg and Lindeboom (2018), "famine often accompany other types of societal disruption. If this point is ignored, this affects the interpretation of the famine as an effect through undernutrition".

[6] Like the majority of the studies in this area, we don't have objective information on the nutritional status of the households in which the recruits were born, hence our estimates are intention-to-treat (ITT). Immediately after Liberation, data on nutritional status at the household level was collected through nutritional surveys in selected communities by Allied medical teams (Burger et al., 1948). These surveys document the extreme nutritional deprivation suffered by many households. Official weekly food rations per household depended on the family composition (family size and age structure), with additional rations for men with hard physical labour and for expecting or lactating women: see Dols and Van Arcken (1946) who provide a detailed description of the food supply and nutrition patterns for different ages and types of workers during and after WWII in the Netherlands. Additionally, even in cases where such information on retrospective self-reported hunger episodes is available, exposure to a famine has been used as instrument for nutritional status – a case in which the exclusion restriction might be violated. In some very recent work, Deng and Lindeboom (2022) propose a new TSIV method (first developed in Van den Berg et al. (2016)) and a bounding exercise for the case of the Chinese famine, and show that the causal treatment effects (obtained by instrumenting self-reported hunger episodes with exposure to the famine) are much larger than the ITT effects.

nutritional content of the rations – a reduction in the proportion of proteins – and partly by a dramatic fall in the calories consumed, coupled with the stress and destruction caused by exposure to warfare (proxied by civilian deaths). Lastly, we confirm that conditioning on the sample of survivors leads to a downward bias of the estimated famine effects, and provide evidence of two-sided selection and scarring: on the one hand, those underweight and overweight were both less likely to survive than those with a normal weight; on the other hand, even after accounting for selection, the famine caused significant scarring, with a 11.67% and a 35.71% higher prevalence of overweight and obesity, respectively, among the survivors. Additionally, once we account for selection, we are able to detect a significant increase among those exposed since middle gestation in the likelihood of having an intellectual disability by 46.67%; this negative impact was only marginally significant in the absence of selection correction, since those more likely to be mentally impaired were less likely to survive.

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

2 Background

2.1 The Dutch Hunger Winter of 1944-1945

Towards the end of the 1940-1945 Nazi occupation of the Netherlands during 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 Western part of the Netherlands (which was still occupied) experienced a severe famine, as result of a Nazi blockade which had been 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

on the Dutch front. While throughout the occupation the food rations were maintained at around 1,800 calories per day per person, they dropped to below 1,000 calories per day by November 1944 and 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 Netherlands; in the North and East the rationing of the food was far more limited (see Lumey and van Poppel (1994) for details), while the South was mostly already liberated. For the 3.5 million people (of a total population of 9.3 million) living in the cities of the West, the effects of this shortage were particularly severe, with estimated war-related excess deaths ranging between 15,000 and 25,000 (see Ekamper et al. (2017) for a discussion of various estimates). While the famine affected the entire population living in the Western Netherlands, more than 40,000 individuals were exposed *in utero*, making it a suitable "natural" experiment to study the consequences of severe prenatal shocks.

2.2 In Utero Shocks and Adult Health Outcomes

As found in many studies using retrospective data on childhood exposures,[7] World War II had devastating consequences for the civilian populations across Europe. War can affect physical and mental health via multiple channels, such as experience of hunger, dispossession, absence of the father, and stress from combats (Kesternich et al., 2014). Examining 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 the instrumental variable estimate (using the propensity to report hunger) a, rather substantial, 3.4 cm decline.

In the context of World War II, many studies have examined the Dutch famine of 1944-1945. The landmark study of Ravelli et al. (1976) used data on military recruits examined at 18 years old, and found an effect for those with prenatal exposure on the likelihood of obesity

[7]There is evidence that recall of some adverse childhood events is trustworthy (Havari and Mazzonna, 2015).

(defined as the weight to height ratio being equal to or greater than 120 percent, as per the WHO Jelliffe (1966) growth standards). More specifically, the authors found that those exposed in the first half of pregnancy had higher obesity rates, while those exposed in the last trimester of pregnancy and first months of life had lower obesity rates.[8] Using the same data on military recruits, Stein et al. (1972) failed to find any significant association between in utero malnutrition and mental performance at age 18. In other 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 in women (Ravelli et al., 1999; Stein et al., 2007). Two further studies evaluating the effects on cognition, instead, found contradictory results.[9] Two recent papers in economics have also studied the effects of the Dutch famine. Scholte et al. (2015) use high-quality register data and find higher hospitalisation rates in the years before retirement if exposure occurred in middle or late gestation, and 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 decline in cognitive ability. Portrait et al. (2017) use the Longitudinal Ageing Study Amsterdam (which includes cohorts born in 1930–1945), and find a 4 cm significant reduction in height for both males and females exposed between gestation and age 2.

Other papers have studied episodes of famine experienced in other countries at the time of WWII. Jürges (2013) finds negative effects of the German famine (proxied by birth cohort-month exposure) on both education and labour market outcomes, stronger for early pregnancy than for late pregnancy exposure. Also for Germany, Kesternich et al. (2015) find higher food expenditures among lower-income adults who experienced hunger in childhood (self-reported episodes validated with comparisons to official food rations) – suggesting a possible behavioural pathway through which the health effects could manifest. Akbulut-

[8] In Lumey et al. (2021) we evaluate the Ravelli et al. (1976) results using the modern definition of obesity (i.e. BMI>30); no other outcomes apart from overweight and obesity are analyzed, and no econometric techniques are used.

[9] de Rooij et al. (2010) found that men and women exposed prenatally performed worse on a selective attention task, but this finding was not replicated in another birth cohort with a more comprehensive evaluation of cognitive performance (de Groot et al., 2011a).

Yuksel (2017), instead, studies the effect of the intensity of WWII destruction in Germany and finds that individuals who were exposed during the prenatal and early postnatal period have higher BMI and are more likely to be obese as adults – an effect that she attributes to early life malnutrition. Neelsen and Stratmann (2011) examine the effects of the Greek famine 1941-1942 using census data and find negative impacts on education and literacy for those exposed in infancy; additionally, they find stronger effects for the urban-born cohorts than for the rural-born ones, suggesting, like other studies, that the famine was mostly experienced in the urban centres.[10] Atella et al. (2020) exploit Nazi raids on municipalities in Italy during WWII and find that workers exposed to raids in utero have worse labor market outcomes. Lastly, Allais et al. (2021) show that both the intensity of WWII exposure in France (measured as the number of military casualties) and self-reported episodes of hunger are associated with worse health outcomes for a sample of women over 20s.

Hence, while multiple studies have examined long-term impacts of exposure to famine, most of them have focused on samples of survivors. In this case, the exposed cohort is subject to two effects working in opposite directions: on the one hand, those who survive are 'scarred' and thus have worse health; on the other hand, those at the bottom of the health distribution are less likely to survive, resulting in left truncation and 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, i.e. the survivors are generally shorter and less healthy. Among the few papers which have accounted for selection there are Gørgens et al. (2012) and Lindeboom et al. (2010), who use the Historical Sample of the Netherlands with continuous life histories and find that most of the selection effects take place at early ages, and that their impacts can be substantial.

[10] 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 implemented by the government, known as the Great Leap Forward. Li and Lumey (2017) provide a systematic review and meta-analysis of the studies on the long-term health effects of the Chinese famine.

3 Data and Sample Definition

3.1 Military Recruits Data

The primary data source for our analysis is the original military recruits data used in the well-known studies of Stein et al. (1972) and Ravelli et al. (1976) in the 1970s. This includes all men born in the Netherlands in the period 1944-1947 who were examined in the military centres at age 18.[11] In addition to the results of the medical examinations, the data contains the exact date and place of birth, and basic demographic information (father's occupational status and family size).

As mentioned in section 1, to account for selective fertility, we only include in our analytical sample those individuals who were already conceived at the start of the famine, i.e. those born up to July 1945. We then define three treatment groups: early exposure starting in the first trimester (born May-July 1945), middle exposure starting in the second trimester (born February-April 1945), and late 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).[12] Figure 1 shows the time periods corresponding to each of the three treatment groups and the control group; the famine period is indicated by the red vertical lines. It is worth noting that, given the 6-month duration of the famine, nobody was exposed for the full duration of the pregnancy: those exposed since mid- and late-gestation had also some postnatal exposure, and those exposed since the first trimester were born after liberation and not exposed in the third trimester.[13] However, we do not consider this to be an issue in our application, given that the famine did not affect

[11] At the time, 100% of Dutch males were called up for military examinations, to determine their fitness for the military service. Hence, data on medical examinations includes all Dutch male citizens aged 18 years in the national population registers, expect those living in psychiatric hospitals or in special institutions for the blind or for the deaf-mute (0.6%). For those who did not undergo physical examinations, available health information from the hospitals or institutions was added to the comprehensive military recruits data. The data also includes all examination results on individuals who were later exempted from service.

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

[13] Defining exposure by counting backwards from the date of birth, as opposed to counting forwards from the date of conception, has well-known problems (Currie and Rossin-Slater, 2013).

gestation length (Stein and Susser, 1975), as we confirm in our birth sample, which is based on high-quality hospital records (column (7) of Table 7).

Additionally, as also mentioned in section 1, we take great care in selecting the cities included in the analytical sample. We proceed according to the following steps. First, given that the famine historically affected more the cities (Banning, 1946; De Jong, 1981), we restrict our sample of interest to the 46 administrative units (municipalities) with a population greater than 25,000 inhabitants on January 1, 1940; in this way we include in the sample a predominantly urban population. Second, using data from the Historical Ecological Databank of the Netherlands (Boonstra, 2016), we exclude: (a) 4 municipalities classified as rural (because of a population highly dispersed across the municipal land area mainly comprising farming and woodlands); (b) 13 municipalities where the population underwent major changes in size since 1930 (i.e. either increased by more than 50% over a decade or decreased after the onset of the war).[14] Table 1 lists the excluded and retained cities, and their allocation into Famine or control areas. Third, we test and fail to reject that, in the remaining 29 cities, control cities follow the same trends as the treated cities before the start of WWII (for the years 1935-1939) in several health outcomes (postnatal mortality rate, crude birth rate, crude death 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).[15] The results are presented in Figures B1 to B8 in the Appendix: here we also see that, in the case of the non-selected cities, we reject the null hypothesis that they were on the same trends for many outcomes.[16] Hence, we work with a

[14]The rationale for doing this is that both rurality and population size are historically major determinants of susceptibility to famine. For example, Breda, Leeuwarden, Venlo and Vlaardingen underwent municipal merges (in 1942, 1944, 1941 and 1942, respectively), which increased their population size; den Helder, Velsen and Vlissingen were evacuated in 1941, 1942 and 1940 respectively, because the use of their harbours as German naval bases implied a heightened risk of the bombing by the allied for the civilians.

[15]Ideally we would test directly for parallel trends in the outcomes of interest; however, military recruits data for the pre-war years is not available to us.

[16]A Wald test is performed for each control group, by fitting a linear regression with treatment group and year dummies interacted, and testing jointly whether the interactions are zero.

consistently defined group of 7 treatment (in the West) and 22 control cities (6 in the West, 7 in the North-East, and 9 in the South).

3.2 Newly Digitised Historical Data

To investigate the nature of the shocks, we enhance the military recruits data by merging them (using the date and city of birth) with newly digitalised information on caloric content and nutrients composition of the rations, temperature, civilian deaths from warfare and the dates in which the cities in the South were liberated by the Allied forces. First, we extract the data on caloric content and nutrients composition from the official war information on the rations (Departement van Landbouw en Visserij, 1946) (sample document is in Figure A1 in the Appendix); this information is available at weekly level for the West and at monthly level for the North-East and South of the Netherlands.[17] Figure 2 shows the distribution of calories and shares of protein, fat and carbohydrates for each trimester by month of birth, separately for the West, North-East and South.[18] The Figure clearly shows that the drop in calories during the famine was accompanied by a drop in the protein share at the end of gestation among those with early exposure in the West, and that the liberated South had rations with higher caloric intake and greater nutritional value.

Second, we extract from the CBS registries information on the number of civilian deaths from war-related causes.[19] 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-

[17] Rations data reflect distributed food coupons based on the number of people in the household, Actual food intake will depend on restribution within the family and additional food sources – information which is not available to us.

[18] The shares were calculated using the following standard formulae: $Protein\ share = (Protein\ (grams) \times 4)/Calories\ (kcals)$, $Fat\ share = 9 \times Fat(grams)/Calories(kcals)$, and $(Carbohydrates\ share = 100 - (Protein\ share + Fat\ share))$.

[19] As noted in Van den Berg and Lindeboom (2018), 'one of the reasons why the Dutch famine has been widely used in the literature is that it was so short that the quality of the population registers for the exposed cohorts was not heavily affected'. Indeed, we can rule out potential measurement error in the date of birth: when we plot the distribution of the calendar day of birth for all the recruits in the years 1944-1947, we see no sign of heaping, both overall, by month, and by city (graphs not reported here for data confidentiality reasons but available upon request).

5) of 1938 (CBS, 1935-1947); this classification allows to better separate direct and indirect mortality from the war (Jewell et al., 2018). These civilian deaths were in part the result of bombings by the Allied Forces (Ekamper, 2020). During the occupation, the Allied Forces carried out around 600 bomb attacks on Dutch territory (Korthals Altes, 1984; NIOD, 2018), aimed at strategic targets, such as ports, bridges, and railways. Most bombings caused no or relatively few deaths among the civilian population; however, in few cases the attacks caused large numbers of civilian casualties (because of errors or missed targets, for example in Nijmegen there were nearly 800 victims on February 22nd, 1944, and in The Hague there were around 550 victims on March 3rd, 1945). Figure 3 shows the number of civilian deaths due to war operations for each birth month by city for the period May 1944 to July 1945; each plot has three lines, showing the civilian deaths for each trimester. The warfare did not affect the three regions in the same way (Ekamper, 2020). Warfare casualties in the cities in the North-East and South were mainly caused by operations of war and combat activity during the liberation battle in the Netherlands from September 1944 to December 1944 in the South and from February 1945 to May 1945 in the North-East. In the meantime, the west of the country was still occupied until the German surrender in May 1945 and not affected by heavy combat activity or periods of heavy bombing. Warfare casualties in the cities in the West were mainly caused by incidental Allied bombings aimed at strategic targets of military importance. These bombings did not interfere with ration card and food distribution in the cities in the West. This reinforces our interpretation of the warfare as stress effects.

Third, we extract temperature information for the period of interest, from the archives of the Royal Netherlands Meteorological Institute (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 average daily temperatures for each city in our sample, separately for each month. For illustrative purposes, Figure D1 presents an example of the interpolated data for two months, December 1944 and May 1945, based

on the data from the meteorological stations (the red squares). The heatmap shows the predicted temperatures across the Netherlands, as indicated in the legend, with each city (the black dots) receiving a value depending on its location. The famine occurred during a winter that, from a purely climatological perspective, was not unusually harsh overall except for a period at the end of January 1945,[20] when parts of the country were still under occupation and some others (the South) were being liberated.

Fourth, we extract from historical sources information on the dates in which each city in the South was liberated from the Nazi by the Allied forces: 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). We merge this information by date of birth and control for the variable 'number of days since liberation' to account for the fact that conditions in the South changed during the study period. Table 2 shows the descriptive statistics of these additional data merged with the recruits sample.[21]

Fifth, we calculate survival rates up to age 18 (proportion of those born who are alive at age 18) by combining the information in the recruits data with data on live births and deaths (still births[22], and deaths <6, 7-29, 30-89, 90-364 days and 1-18 years) by sex, region and month from Stein et al. (1975) (Table 1), whose reliability has been ascertained on the

[20]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 (Ekamper et al., 2017).

[21]In the analysis, we transform both calories and protein shares to their negative values, so that their coefficients can be interpreted as the effects of a one-unit decrease. Additionally, we convert the calories in thousands to ease readability of the estimated coefficients; and we use a log(x+1) transformation for the warfare variables, due to the high skewness and to avoid loosing the observations with zero deaths.

[22]Reported stillbirth rates did not change much over time between 1935-1947 (De Zwarte, 2023); nonetheless, they might be subject to reporting errors. Unfortunately, we don't have information on miscarriages. However, in Table 7 we show that there is no effect of in utero famine exposure on sex ratios: and given that males are more vulnerable to maternal stress in pregnancy (Kraemer, 2000), we can rule out famine impacts on fetal deaths (see Sanders and Stoecker (2015) for a related use of sex ratio of live births as indicator of fetal deaths).

basis of extensive follow-up of predetermined cohorts from birth to age 18.[23] Lastly, we extract from the Historical Ecological Databank (HED) of the Netherlands (Boonstra, 2016) the following city-level pre-war information (measured in 1930), which we include in the survival selection model (equation (4) in Section 4): medical staff per 1,000 of population, the share of inhabitants of the largest place in the municipality (over the total number of inhabitants) and the number of religious groups with ≥25 inhabitants in 1930. The choice of these exclusions has been guided by the following considerations. First, given that the Nazi embargo closed the imports also of medications (and of other vital supplies in addition to food) and that the disease environment also worsened (Van den Berg and Lindeboom, 2018; Banning, 1946), it is plausible that individuals born in cities with greater availability of medical infrastructures were more likely to survive. Second, mortality in towns with less population density was lower (Banning, 1946). Third, it has been documented that religious associations (in particular the Inter-Church Council, I.K.O.) were very active in helping the vulnerable population at risk of starvation (Banning, 1946).

3.3 Birth Hospital Data

We complement the analysis of age 18 outcomes based on the military recruits data with unique birth data from the hospital records of five cities: Amsterdam, Rotterdam and Leiden (West), Groningen (North) and Heerlen (South) (Stein and Susser, 1975).[24] Although at the time less than half of the births were taking place in hospitals, the admission procedures remained unchanged during the war (Stein and Susser, 1975). We use data on all births that occurred between May 1944 and July 1945. For each birth record, we have information on the sex of the newborn, weight, length, head circumference, along with placenta's weight, gestational age, and mother's age at birth.

[23] For a subset of these cohorts in Amsterdam and Tilburg, Stein et al. (1975) also demonstrated that deaths and out-migrations were low and no births were missing from the military register.

[24] We have data from one hospital per city, with the exception of Amsterdam for which we have data for two hospitals.

4 Econometric Framework

We use a difference-in-differences design to estimate the impact of the famine on various health outcomes at age 18. Our main estimating equation is:

$$\begin{aligned} y_{ijm}^{K} =& \beta_0 + \beta_1 WestF_{ijm} \times Late_{ijm} + \beta_2 WestF_{ijm} \times Middle_{ijm} \\ &+ \beta_3 WestF_{ijm} \times Early_{ijm} + \beta X_{ijm} + City_j + (Birth\ Month)_m + \epsilon_{ijm} \end{aligned} \tag{1}$$

where y_{ijm}^{K} is one of the following outcomes K for person i, born in city j in month m: *Height*, *Weight*, *BMI* (kg/m^2), *Overweight* ($BMI \geq 25$), *Obesity* (weight/height ratio $\geq 120\%$ – we use the definition as in Ravelli et al. (1976) for comparison purposes, which uses the 1966 WHO reference standards in Jelliffe (1966)), *Underweight* ($BMI < 18.5$), *Chest-Height Ratio* (a measure of body size, used among others in Costa (2004)), *Intellectual Disability* (ICD-6 325, primary or secondary diagnosis, as defined in Stein et al. (1972)). $WestF_{ijm}$ is a binary indicator which takes value 1 for those born in the West Famine region. $Late_{ijm}$, $Middle_{ijm}$ and $Early_{ijm}$ are binary indicators which take value 1 for those born between November 1944 and January 1945, between February 1945 and April 1945, and between May 1945 and July 1945, respectively (see Figure 1) – i.e. with exposure starting from the third, the second and the first trimester.

The reference group includes those born in the selected cities in Non-Famine West, North-East and South regions (see Table 1), between May and October 1944 – hence with only postnatal exposure to the famine. We do not include those conceived during or after the famine, in view of the significant changes in conceptions and births (first reductions, and then increases, respectively), particularly among those with low socioeconomic status (Stein et al., 1975). Finally, $City_j$ and $(Birth\ Month)_m$ are fixed effects for city and month of birth, and X_{ijm} is a vector of controls (binary indicators for father's occupational status, number of older brothers, birth order, and religion).[25]

[25] Table D1 in the Appendix presents descriptive statistics of these variables. In our analytical sample, 28% of the cohort members had fathers who were semi-skilled workers, whereas 27% had fathers who were

The main parameters of interest are β_1, β_2 and β_3, the interaction terms of $WestF_{ijm}$ with the three exposure dummies, which can be interpreted as an Intention-to-Treat (ITT) effect. In all estimations we use clustered standard errors;[26] given the relatively small number of cities, we follow the recommendation in Cameron et al. (2008) and compute wild cluster bootstrap standard errors, reporting the corresponding p-values (Roodman et al., 2019). To account for the multiplicity of hypotheses tested, we use the Romano and Wolf (2005, 2016) procedure.[27]

After having estimated our baseline specification, we perform two additional analyses to allow for the fact that seasonal effects might differ between famine and non-famine regions. First, we estimate the same model as in Equation 1, but on a sample of cohorts born two years later (i.e. between May 1946 and July 1947), as a placebo regression. Second, we estimate the following triple difference (difference-in-difference-in-differences) specification:

$$\begin{aligned} y^K_{ijm} = {} & \beta_0 + \beta_1 WestF_{ijm} \times Late_{ijm} + \beta_2 WestF_{ijm} \times Middle_{ijm} + \beta_3 WestF_{ijm} \times Early_{ijm} \\ & + \beta_4 War_{ijm} + \beta_5 WestF_{ijm} \times War_{ijm} + \beta_6 Late_{ijm} \times War_{ijm} \\ & + \beta_7 Middle_{ijm} \times War_{ijm} + \beta_8 Early_{ijm} \times War_{ijm} \\ & + \beta_9 WestF_{ijm} \times Late_{ijm} \times War_{ijm} + \beta_{10} WestF_{ijm} \times Middle_{ijm} \times War_{ijm} \\ & + \beta_{11} WestF_{ijm} \times Early_{ijm} \times War_{ijm} + \beta X_{ijm} + City_j + (Birth\ Month)_m + \epsilon_{ijm} \end{aligned} \tag{2}$$

where the variables $WestF$ and $Late$, $Middle$ and $Early$ are defined as before. The variable

white collar workers. 35% were first-born, 29% second-born and 17% third-born, with 45% having at least one older brother. Catholics constituted 40% of the sample and Protestants 28%.

[26] 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. (2020)).

[27] The results reported here are based on 5,000 replications for the wild cluster bootstrap and on 1,000 replications for the Romano and Wolf procedure; the results are not sensitive to different numbers of replications.

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}.

Then, to investigate any short-term effects, we complement the analysis of the health outcomes at age 18 with that of the birth outcomes, using the data on the six hospitals. The main estimating equation is:

$$y^K_{ijm} = \beta_0 + \beta_1 West_{ijm} \times Late_{ijm} + \beta_2 West_{ijm} \times Middle_{ijm} + \beta_3 West_{ijm} \times Early_{ijm} + \beta X_{ijm} + Hospital_j + (Birth\ Month)_m + \epsilon_{ijm} \quad (3)$$

where y^K_{ijm} is one of the following birth outcomes for child i born in city j in month m: *Birth Weight*, *Low Birth Weight* (birth weight $< 2,500$ grams), *Birth Length*, *Head Circumference*, *Placenta Weight*, *Gestation Age*, and *Sex Ratio*. The analytical sample includes only male births (for comparison with the main results), except for *Sex Ratio* which includes both males and females. $Late_{ij}$, $Middle_{ij}$, $Early_{ij}$ are defined as before. $West_{ij}$ is a dummy variable taking value 1 for those born in one of the hospitals in Amsterdam, Leiden and Rotterdam, and 0 for those born in Groningen and Heerlen. $Hospital_j$ and $(Birth\ Month)_m$ are the hospital and month of birth fixed effects and X_{ij} includes mother's age as control.

Finally, we assess the robustness of the results to accounting for selection into survival up to age 18.[28] We use a flexible copula selection model, which allows us to depart from the bivariate normality assumption on the error terms;[29] in our implementation we choose among different copulas (Gaussian, FGM, Plackett, Clayton, AMH, Frank, Gumbel, and Joe) using the Bayesian Information Criterion (BIC). The estimating outcome equation is

[28] Mobility is unlikely to be an issue in our setup. First, as mentioned in Section 3.1, we exclude from our control group the cities which have been evacuated; second, as mentioned in Section 3.2, Stein et al. (1975) showed that out-migration was not an issue.

[29] While a general implementation can be traced back to Lee (1983), the use of copulas in this context has become explicit since Smith (2003).

equation (1), and the selection equation is:

$$\begin{aligned} s^K_{ijm} =& \gamma_0 + \gamma_1 WestF_{ijm} \times Late_{ijm} + \gamma_2 WestF_{ijm} \times Middle_{ijm} + \gamma_3 WestF_{ijm} \times Early_{ijm} \\ &+ \gamma_4 WestF_{ijm} + \gamma_5 Late_{ijm} + \gamma_6 Middle_{ijm} + \gamma_7 Early_{ijm} \\ &+ \gamma_8 Med1930_j + \gamma_9 Larg1930_j + \gamma_{10} Relig1930_j + u_{ijm} \end{aligned} \tag{4}$$

where s^K_{ijm} indicates whether the individual is alive at age 18 (constructed as explained in section 3.2), and the error terms ϵ_{ijm} (equation 1) and u_{ijm} (equation 4) have a joint distribution based on one of the proposed copulas. For example, in the case of a Gaussian copula, the joint distribution is a bivariate Normal distribution, commonly used in many applications of sample selection models (Hasebe, 2013). The rest of the variables are defined as before, and $Med1930$, $Larg1930$ and $Relig1930$ are the 1930 city-level variables described in section 3.2: medical staff per 1,000 population, the share of inhabitants of the largest place in the municipality, and the number of religious groups with ≥25 inhabitants, all likely to affect survival for the reasons documented in section 3.2.

5 Results

Baseline results Our baseline results, based on the difference-in-differences specification (Equation 1), are presented in Table 3. All coefficients are from linear regression models, so that they can be interpreted directly as marginal effects. The individuals exposed since early gestation (i.e. exposed in trimesters 1 and 2 only, as the famine period was limited to 6 months) have a significantly higher weight (by 652 grams w.r.t a control mean of 67.6 kg), BMI (by 0.2 w.r.t a control mean of 21.5), obesity (using the Ravelli definition,[30] by 0.6 p.p. w.r.t a control mean of 1.4%, quite a sizeable increase), and chest-height ratio (a 0.010

[30]Note that here we define obesity as weight/height>120, as used by Ravelli, and not with the current definition of BMI ≥30.

increase w.r.t a control mean of 0.492); they are also significantly less likely to be underweight (by 1.3 p.p. w.r.t a control mean of 5.6%). The individuals exposed since mid-gestation (i.e. exposed in trimesters 2 and 3) have only a significant increase in chest-height ratio (a 0.008 increase w.r.t a control mean of 0.492) and in intellectual disability (by 1 p.p. w.r.t. a control mean of 3%). The individuals exposed since late gestation (i.e. exposed in trimester 3 and in the first three months of life) have significantly *lower* weight, overweight and obesity (using the Ravelli definition), as compared to those with exclusive postnatal exposure. All these effects are robust to accounting for the small number of cities by computing the Wild cluster bootstrap standard errors, and to controlling for multiple hypothesis testing by using the Romano and Wolf (2005) step-down method. We are unable to detect any impacts for any exposure group for height.

What might be driving the differential impacts on weight-related outcomes in the early and late exposure groups? Ravelli et al. (1976) (who first reported this finding) propose two hypotheses to link critical periods of growth for specific tissues or organ adipose tissue to permanent effects later in life: the first concerned the number of adipose cells laid down during sensitive periods and the second the development of hypothalamic center regulating appetite control. The obesity effects linked to early famine exposure can be easily explained within the 'mismatch hypothesis', i.e the fact that experiencing an adverse fetal environment, followed by normal food availability postnatally (and consequent 'catch-up' growth), is conducive to obesity and other chronic conditions (Gluckman and Hanson, 2006). The programming of adult obesity by intrauterine food deprivation has been shown experimentally in rodent models: the offspring of pregnant rats underfed in the first two weeks of gestation developed hypertrophied adipocytes (Jones and Friedman, 1982). Similar findings have been obtained experimentally in sheep models: maternal nutrient restriction, targeted in the period in which placental growth occurs, resulted in lower maternal plasma cortisol, leptin (the 'satiety' hormone), T4, and IGF-I; these adaptations were accompanied by an increase in fetal adipose tissue deposition in the offspring (Bispham et al., 2003). As possible

mechanism, Crescenzo et al. (2003) have studied the role of suppressed thermogenesis favoring catch-up fat in the pathophysiology of catch-up growth. In humans, famine exposure in early gestation (but not in mid or late gestation) has also been linked to selective fetal survival and (in adults) to DNA methylation changes at the imprinted IgF2 gene, increased T2DM in middle age, and increased mortality (Ekamper et al., 2015). In case of exposure to famine in late gestation, instead, there is no such mismatch, since the environment of scarcity continues postnatally: there is just a lower nutritional intake, which naturally leads to lower weight and obesity.

To examine whether any significant effects found in Table 3 might be driven by systematic differences between the specified groups other than by famine exposure, we then present in Table 4 the results from a placebo test, where we estimate the same equation as before but on the cohorts born two years later (i.e. between May 1946 and July 1947). As expected, we find no significant impacts in the placebo regressions for any of the outcomes: all significant coefficients in Table 4 are either of much reduced magnitude and/or have the opposite sign as in Table 3.[31] We then perform the placebo test on the sample used by Ravelli et al. (1976), which is twice as large as ours, since it includes rural areas. The results, presented in the Appendix Table D2, show that the placebo test fails: we find significant differences for weight, BMI, and underweight, likely due to differences across groups which cannot be attributed to the famine. This further confirms the importance of carefully selecting the cities belonging to the control group, and casts doubts on the comparisons made in Ravelli et al. (1976).

Next, we combine our main wartime sample (i.e. births May 1944 - July 1945) and the placebo sample (births May 1946 - July 1947) in a triple difference specification. The results are reported in Table 5, where we display the triple interaction terms $WestF \times Late \times War$, $WestF \times Middle \times War$, and $WestF \times Early \times War$ (from Equation 2). Most of our main results for exposure since the first trimester are confirmed, with the coefficients on the triple

[31] For example, the coefficient on the interaction $WestF \times Early$ is 0.218 (significant at 1%) in Table 3, and -0.044 (not statistically significant) in Table 4.

interaction terms increasing by almost 50% for weight, by 25% for BMI, and by 40% for being underweight; the coefficient on obesity, instead, is halved in size and no longer significant. Additionally, the other coefficients for exposure since middle and late gestation are smaller in magnitude and no longer statistically significant.[32]

Lastly, we repeat 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 C1 and C2 in the Appendix, where the plots show the interaction term estimates as in the Tables 3 and 5, along with their 90% confidence intervals. Across both models, the BMI and Chest-Height ratio results are robust to the choice of the region used as comparison group.

Understanding the nature of the shocks We now exploit our newly digitised data to investigate the relative importance of the different shocks occurring during the Dutch Hunger Winter; we focus on the outcomes for which we have found a significant effect in the main specification (equation 1, Table 3). Table 6 shows the results from our basic difference-in-differences specification (equation (1)), with the additional inclusion of variables on calories, protein share, warfare deaths, temperature and weeks since liberation, separately by trimester.[33] While this exercise provides purely correlational evidence, it is nonetheless informative about the relative importance of different war-related shocks experienced by these cohorts born during the Dutch Hunger Winter.

Our results show that both the poor nutritional quality of the diet and the warfare-related stress and destruction play a significant role in moderating the effects of being born during the famine on health at age 18. More specifically, the impacts on weight, BMI (whose

[32]The only exception is the coefficient on the chest-height ratio for mid-pregnancy exposure, which is only significant at 10% level.

[33]Calories and protein shares have been transformed into negative values, so the effect can be interpreted as reductions in each. Temperature and weeks since liberation in the South are also controlled for in every specification (they are not shown in the table for readability, since their associated coefficients fail to achieve statistical significance). Table D3 in the Appendix presents the same specification without the Difference-in-Differences interaction terms.

coefficients are reduced by 26% and 35%, respectively[34]) and overweight for those exposed since early gestation appear to be moderated by a combination of increased exposure to warfare *and* reduction of proteins in the rations (at a low caloric content), which became particularly severe at the end of the famine, when those with early exposure (born between May and July 1945) were in the third trimester. Instead, the impacts on chest-height ratio (whose coefficient is reduced by 60%[35]) and obesity appear driven *entirely* by a reduction in protein share at the end of gestation (again, at a low caloric content[36]).[37] These results are robust to controlling for multiple hypothesis testing and the choice of cities included in our sample (see Figure C3 in the Appendix).

These findings are consistent with an established nutrition literature which has linked, in animal models, a protein-restricted diet in mothers with impaired metabolic function in the offspring (Stocker et al., 2005); the effects of unbalanced protein intake in pregnancy have also been found in humans (see e.g. Imdad and Bhutta (2011) for a review of studies on birth outcomes). More specifically, Alejandro et al. (2020) have found that maternal low-protein diet at the end of pregnancy contributes to increased obesity, insulin resistance and β-cell dysfunction in the mouse offspring; and perturbations in miRNA expression seem to be among the biological mechanisms involved in causing the impacts of a low-protein diet during gestation on offspring obesity (Zheng et al., 2017). As mentioned, the cohorts exposed since earlier in gestation have also experienced a 'mismatch' between a harsher famine environment before birth and a more prosperous post-liberation environment after birth, which could have exacerbated the effect of the prenatal shock (Gluckman and Hanson, 2004).[38] However, while we are able to disentangle in part the nature of the shocks which

[34]We are comparing the coefficients in Table 6 and in Table 3.

[35]Again, we are comparing the coefficients in Table 6 and in Table 3.

[36]It is useful to reiterate that, although statistically we control for both caloric intake and protein share, in practice we have little separate identifying variation, since the reduction in calories was partly achieved via a substantial fall in the protein content of the diet.

[37]The famine effects on underweight appear to be moderated only by the reduction in calories in the third trimester, highlighting how changes in different parts of the BMI distribution might be moderated by different determinants.

[38]More specifically, the famine could have caused a 'developmental mismatch' (Gluckman et al., 2019), which occurs when the phenotype induced during development encounters a different environment post-

hit those exposed in utero to the Dutch Hunger Winter, we are unable to fully account for other shocks, such as income losses, and for behavioural responses: hence, our reduced-form coefficients might be bigger or smaller than a 'purely biological effect' (Yi et al., 2015).[39]

Birth Outcomes As described in Section 3.3, to explore impacts of famine exposure on birth outcomes, we then use the additional dataset with information collected in selected birth hospital clinics in three treated and two control cities. For consistency with the recruits data, we focus on the sample of male newborns. We first show in Table D4 that we are able to replicate the main results in the subsample of cities for which we have the hospital birth data. We then present the birth results in Table 7. They show a reduction in birth weight and placenta weight for those exposed since middle and early gestation, respectively; however, the trimester-specific effects are neither significantly different from each other, nor robust to controlling for multiple hypothesis testing.[40] In addition, we find no effect on gestational age (column 7) and on sex ratio (column 8, where we use all births, i.e. males and females) for any of the exposure groups.[41] The fact that there are no effects on sex ratio and gestational age further validates our strategy of using the date of birth to identify the date of conception (and so the exposure by trimester), and rules out any concerns related to the fact that we don't observe miscarriages in our mortality data.

Accounting for Selection In this last section we account for selective mortality using a copula-based approach, which is less commonly used, but more flexible, than a normal

development: such induced alterations are advantageous for short-term survival, but detrimental in the long term.

[39]For example, in some families pregnant women might have been protected, and given a higher share of resources than the one implied by the rations; in others, parents might have allocated a higher share of food consumption postnatally to those children prenatally undernourished, possibly creating unhealthy eating habits (Kesternich et al., 2015).

[40]Stein et al. (1995) finds that women who delivered at the height of the famine no longer had third trimester pregnancy weight gains as before and that the birth weights of their offspring declined by 300g; however, this study has limitations since it only includes birth at the University of Amsterdam Teaching Hospital in a before-after comparison.

[41]The zero effect on sex ratio for the Dutch cohorts is shown and discussed in Cramer and Lumey (2010). Other studies, instead, have found significant impacts of early shocks on sex ratios, see for example Catalano et al. (2006), Almond et al. (2007) and Sanders and Stoecker (2015).

selection model. Recall that our main analysis is based on military recruits data, which was collected when the respondents were 18 years old, conditional on being alive and resident in the Netherlands; additionally, since the cohorts born in the Western cities faced much worse conditions than those in the rest of the country, we expect them to be less likely to survive. Figure D2 shows that there is indeed substantial variation in survival rates (proportion of those born in the period May 1944-July 1945 in one of the selected cities who are alive at age 18) across cities. Table D5 shows the results for the selection equation: the individuals exposed during gestation have indeed a lower probability of surviving until age 18 as compared to those with only postnatal exposure, with the greater reduction in survival experienced by those exposed since mid-gestation. Reassuringly, the excluded variables have the expected sign: a greater availability of medical staff in the city and a larger number of religious groups is associated with greater survival, while living in a larger municipality is associated with lower survival (both because of reduced availability of food and also because of easier spread of infections).

As described in Section 4, we then jointly estimate equation (1) and (4), and we model the errors with more flexible copulas functions. The results are presented in Table 8, and show evidence of *both* selection *and* scarring effects. First of all, as expected, the survivors appear to be positively selected: the sign and magnitude of Kendal's τ[42] show that the survivors are less likely to be either underweight or overweight, and have on average a taller height and larger weight, and higher BMI; in other words, those taller and more robust (without being overweight) appear to have been more likely to survive, and so the BMI distribution of the survivors appears truncated both on the left and the right.[43] Nonetheless, once we account for survival, we still detect a significant scarring effect, with those exposed since early gestation having a higher BMI and CHR (by 0.188 and 0.011) and being 0.7 p.p. and 0.5 p.p. more likely to be overweight and obese and 0.8 p.p. less likely to be underweight

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

[43] Gørgens et al. (2012) also found that taller children were more likely to survive the Chinese famine.

than the controls, respectively – all impacts in line with the baseline specification (Table 3), as expected given small selection effects among those with early exposure (Table D5).

Second, accounting for survival seems particularly important to understand the famine impacts on intellectual disability (column 8 of Table 8): we detect not only a negative selection effect,[44] but also a significant scarring effect of famine exposure in middle and late gestation on intellectual disability, by 1.4 p.p. and 0.7 p.p., respectively (w.r.t. a control mean of 3%, and estimates of 1p.p. and 0.2p.p. in the baseline specification). The middle gestation exposure effect, in particular, is significantly different from the early and the late ones at the 1% level (while the early and the late one are not statistically significantly different from each other), and it implies a 47% increase in intellectual disability. While such finding had not been uncovered in Stein et al. (1972), both de Groot et al. (2011b) and Wiegersma et al. (2022) are in line with our results.[45] Additionally, a body of biological evidence confirms the importance of mid-pregnancy exposure to stressors for infant neurodevelopment, partly via hippocampal myelination and synaptophysin expression (see for example Xu et al. (2013) and Ellman et al. (2008)): indeed, it is in the second trimester that the fetal hippocampus undergoes several critical changes (Ge et al., 2015). In sum, accounting for survivorship bias strengthens our main results: prenatal exposure to the famine had both selection and scarring effects, and caused increases in both physical health conditions (overweight and obesity) and intellectual disabilities at age 18.

[44]The sign of Kendal's τ points to a negative correlation between survival status at 18 and having an intellectual disability: hence, with very few intellectually disabled surviving until 18, it is difficult to uncover a significant impact of the famine on this dimension when using a survivors-only sample. Indeed, previous research has only been able to detect impacts on cognitive performance in older samples, when cognitive impairments had started to be visible also among survivors.

[45]More specifically, these studies have found that, in line with our results, earlier gestation exposure for men relates to worse cognitive performance at age 59 (de Groot et al., 2011b) and more self-perceived cognitive problems and need for medical attention at age 72 (Wiegersma et al., 2022).

6 Conclusions

In this paper, we have investigated health impacts at the end of adolescence of prenatal exposure to multiple shocks, by exploiting the unique natural experiment of the Dutch Hunger Winter, which hit abruptly the Western Netherlands at the end of World War II. We have linked military recruits data with objective health measurements on the cohorts born in the years surrounding World War II, used in landmark epidemiological papers, with newly digitised historical records on calories and nutrient composition of the war rations, daily temperature, and warfare deaths. We have also collected a variety of health indicators for all the cities in the Netherlands since the mid 1930s, to carefully select the cities in the control group on the basis of common trends. Armed with this new rich resource, we have provided the following contributions to the interdisciplinary literature on the developmental origins of health.

First, we have confirmed previous epidemiological findings that the cohorts exposed to the Dutch Hunger Winter since early gestation have an increased probability of being obese at age 18: by 0.6 p.p. in the basic DiD specification, which is a 43% increase with respect to the control group, and it amounts to 1.6 times the obesity differential between recruits of high and low socioeconomic status (as proxied by the father having a manual or non-manual job).[46] While the magnitude of these effects appears smaller than that found for adults by previous work (Akbulut-Yuksel, 2017), it has been demonstrated that obesity at age 18 significantly increases the risk of being obese in adulthood, and consequently of developing type 2 diabetes and hypertension (Goldhaber-Fiebert et al., 2013), as previous work on the Dutch famine has shown (Portrait et al., 2011). Additionally, exposure to famine in early gestation has been linked with an increase in mortality by age 63 (Ekamper et al., 2015): hence, our results contribute to understanding how shocks of different nature (nutritional versus stress-related) affect health – which is crucial to design effective remediation strategies.

[46] We have socioeconomic status in the recruits data and we do not find any heterogeneity in the impacts of the famine by socioeconomic status (results available upon request).

Second, by providing different placebo and robustness tests, we have shown that the careful selection of cities to include in the control group matters. Third, we have found that the adverse health effects of early in utero exposure to the Dutch Hunger Winter are moderated, in part, by a combination of warfare exposure and energy-adjusted protein reduction in the rations at the end of gestation, while the winter temperature does not seem to have played a role. Fourth, we have carried out a flexible selection correction procedure by means of copula models, and we have shown evidence of *both* selection ('survival of the fittest') *and* scarring effects; crucially, even after accounting for selective survival, early in utero exposure to the Dutch Hunger Winter leads to long-term scarring in physical health and mental capability.

Although our study is based on a historical context, wars and famines are still a sad reality in the modern world (like the ongoing Russian-Ukraine and Hamas-Israel conflicts, or recent experiences in Africa),[47] hence lessons learnt from the Dutch Hunger Winter can still be applied today. For example, Ampaabeng and Tan (2013) find long-term cognitive consequences of infancy exposure to the 1983 Ghana famine; Dercon and Porter (2014) find long-term impacts on height for survivors of the 1984 Ethiopia famine. Hence, although cities might be rebuilt and return to their previous patterns, poor nutrition and war-related destruction place substantial burden on children's development, which might last a lifetime. Protecting vulnerable subjects in times of conflict is both a moral and a public health imperative.

[47] In the latest *Food Security Update* from November 2023, the World Bank listed Burkina Faso, Mali, Palestine, South Sudan, and Sudan as hunger hotspots of the highest concern, with significant numbers of people at risk of famine.

References

Abadie, A., S. Athey, G. W. Imbens, and J. M. Wooldridge (2020). Sampling-based versus design-based uncertainty in regression analysis. *Econometrica 88*(1), 265–296.

Akbulut-Yuksel, M. (2017). War during childhood: The long run effects of warfare on health. *Journal of health economics 53*, 117–130.

Alejandro, E. U., S. Jo, B. Akhaphong, P. R. Llacer, M. Gianchandani, B. Gregg, S. D. Parlee, O. A. MacDougald, and E. Bernal-Mizrachi (2020). Maternal low-protein diet on the last week of pregnancy contributes to insulin resistance and β-cell dysfunction in the mouse offspring. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology 319*(4), R485–R496.

Allais, O., G. Fagherazzi, and J. Mink (2021). The long-run effects of war on health: Evidence from world war ii in france. *Social Science & Medicine 276*, 113812.

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

Almond, D., L. Edlund, H. Li, and J. Zhang (2007). Long-term effects of the 1959-1961 china famine: Mainland china and hong kong.

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

Ampaabeng, S. K. and C. M. Tan (2013). The long-term cognitive consequences of early childhood malnutrition: the case of famine in ghana. *Journal of health economics 32*(6), 1013–1027.

Atella, V., E. Di Porto, J. Kopinska, and M. Lindeboom (2020). Maternal stress and offspring lifelong labor market outcomes. *IZA DP 13744*.

Banning, C. (1946). Food shortage and public health, first half of 1945. *The ANNALS of the American Academy of Political and Social Science 245*(1), 93–110.

Ben-Shlomo, Y. and D. Kuh (2002). A life course approach to chronic disease epidemiology: conceptual models, empirical challenges and interdisciplinary perspectives.

Bispham, J., G. Gopalakrishnan, J. Dandrea, V. Wilson, H. Budge, D. Keisler, F. Broughton Pipkin, T. Stephenson, and M. Symonds (2003). Maternal endocrine adaptation throughout pregnancy to nutritional manipulation: consequences for maternal plasma leptin and cortisol and the programming of fetal adipose tissue development. *Endocrinology 144*(8), 3575–3585.

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

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

Burger, G., J. Drummond, and H. Sandstead (1948). *Malnutrition and starvation in Western Netherlands*. Part I. The Hague, Netherlands: General State Printing Office.

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

Catalano, R., T. Bruckner, A. R. Marks, and B. Eskenazi (2006). Exogenous shocks to the human sex ratio: the case of september 11, 2001 in new york city. *Human reproduction 21*(12), 3127–3131.

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.

Conti, G., G. Mason, and S. Poupakis (2019). Developmental origins of health inequality. In *Oxford Research Encyclopedia of Economics and Finance*. Oxford University Press.

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

Crescenzo, R., S. Samec, V. Antic, F. Rohner-Jeanrenaud, J. Seydoux, J.-P. Montani, and A. G. Dulloo (2003). A role for suppressed thermogenesis favoring catch-up fat in the pathophysiology of catch-up growth. *Diabetes 52*(5), 1090–1097.

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

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

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

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

De Jong, L. (1981). *Het Koninkrijk der Nederlanden in de Tweede Wereldoorlog. Het laatste jaar II. Eerste helft, Volume 10b.* 's-Gravenhage: Staatsuitgeverij.

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

De Zwarte, Ingrid, e. a. (2023). Stillbirths. *Population Studies*.

Deng, Z. and M. Lindeboom (2022). Early-life famine exposure, hunger recall, and later-life health. *Journal of Applied Econometrics*.

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.

Dercon, S. and C. Porter (2014). Live aid revisited: long-term impacts of the 1984 ethiopian famine on children. *Journal of the European Economic Association 12*(4), 927–948.

Dols, M. and D. Van Arcken (1946). Food supply and nutrition in the netherlands during and immediately after world war ii. *The Milbank Memorial Fund Quarterly*, 319–358.

Ekamper, P. (2020). World War II municipal liberation dates in the Netherlands, 1944-1945. https://doi.org/10.17026/dans-zcq-7grx, deposited dataset DANS-EASY.

Ekamper, P., G. Bijwaard, F. van Poppel, and L. Lumey (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.

Ekamper, P., F. van Poppel, A. D. Stein, G. E. Bijwaard, and L. Lumey (2015). Prenatal famine exposure and adult mortality from cancer, cardiovascular disease, and other causes through age 63 years. *American journal of epidemiology 181*(4), 271–279.

Ellman, L. M., C. D. Schetter, C. J. Hobel, A. Chicz-DeMet, L. M. Glynn, and C. A. Sandman (2008). Timing of fetal exposure to stress hormones: effects on newborn phys-

ical and neuromuscular maturation. *Developmental Psychobiology: The Journal of the International Society for Developmental Psychobiology 50*(3), 232–241.

Ge, X., Y. Shi, J. Li, Z. Zhang, X. Lin, J. Zhan, H. Ge, J. Xu, Q. Yu, Y. Leng, et al. (2015). Development of the human fetal hippocampal formation during early second trimester. *Neuroimage 119*, 33–43.

Gluckman, P. D. and M. A. Hanson (2004). Developmental origins of disease paradigm: a mechanistic and evolutionary perspective. *Pediatric research 56*(3), 311–317.

Gluckman, P. D. and M. A. Hanson (2006). The developmental origins of health and disease. *Early life origins of health and disease*, 1–7.

Gluckman, P. D., M. A. Hanson, and F. M. Low (2019). Evolutionary and developmental mismatches are consequences of adaptive developmental plasticity in humans and have implications for later disease risk. *Philosophical Transactions of the Royal Society B 374*(1770).

Goldhaber-Fiebert, J. D., R. E. Rubinfeld, J. Bhattacharya, T. N. Robinson, and P. H. Wise (2013). The utility of childhood and adolescent obesity assessment in relation to adult health. *Medical Decision Making 33*(2), 163–175.

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

Hasebe, T. (2013). Copula-based maximum-likelihood estimation of sample-selection models. *The Stata Journal 13*(3), 547–573.

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

Havari, E. and F. Peracchi (2017). Growing up in wartime: Evidence from the era of two world wars. *Economics & Human Biology 25*(9–32).

Imdad, A. and Z. A. Bhutta (2011). Effect of balanced protein energy supplementation during pregnancy on birth outcomes. *BMC public health 11*(3), 1–9.

Jelliffe, D. B. (1966). The assessment of the nutritional status of the community. *World Health Organization Monograph Series* (53).

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

Jones, A. P. and M. I. Friedman (1982). Obesity and adipocyte abnormalities in offspring of rats undernourished during pregnancy. *Science 215*(4539), 1518–1519.

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.

Kesternich, I., B. Siflinger, J. P. Smith, and J. K. Winter (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., B. Siflinger, J. P. Smith, and J. K. Winter (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), F372–F393.

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*. Amsterdam: Sijthoff.

Kraemer, S. (2000). *The fragile male*. BMJ Publishing Group.

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

Li, C. and L. Lumey (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.

Lindeboom, M., F. Portrait, and G. J. Van den Berg (2010). Long-run effects on longevity of a nutritional shock early in life: the dutch potato famine of 1846–1847. *Journal of health economics 29*(5), 617–629.

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

Lumey, L. and F. W. van Poppel (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., P. Ekamper, G. Bijwaard, G. Conti, and F. van Poppel (2021). Overweight and obesity at age 19 after pre-natal famine exposure. *International Journal of Obesity 45*(8), 1668–1676.

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.

Mink, J., M.-C. Boutron-Ruault, M.-A. Charles, O. Allais, and G. Fagherazzi (2020). Associations between early-life food deprivation during world war ii and risk of hypertension and type 2 diabetes at adulthood. *Scientific reports 10*(1), 1–9.

Neelsen, S. and T. Stratmann (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., E. Teeuwiszen, and D. Deeg (2011). Early life undernutrition and chronic diseases at older ages: the effects of the dutch famine on cardiovascular diseases and diabetes. *Social science & medicine 73*(5), 711–718.

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

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

Ravelli, A. C., J. H. van der Meulen, C. Osmond, D. J. Barker, and O. P. Bleker (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., Z. A. Stein, and M. W. Susser (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 M. Wolf (2005). Stepwise multiple testing as formalized data snooping. *Econometrica 73*(4), 1237–1282.

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

Roodman, D., M. Ø. Nielsen, J. G. MacKinnon, and M. D. Webb (2019). Fast and wild: Bootstrap inference in stata using boottest. *The Stata Journal 19*(1), 4–60.

Sanders, N. J. and C. Stoecker (2015). Where have all the young men gone? using sex ratios to measure fetal death rates. *Journal of health economics 41*, 30–45.

Scholte, R. S., G. J. Van den Berg, and M. Lindeboom (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.

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

Stein, A., A. Ravelli, and L. Lumey (1995). Famine, third-trimester pregnancy weight gain, and intrauterine growth: the dutch famine birth cohort study. *Human Biology 67*(1), 135–50.

Stein, A. D., H. S. Kahn, A. Rundle, P. A. Zybert, K. van der Pal-de Bruin, and L. Lumey (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 M. Susser (1975). The Dutch famine, 1944–1945, and the reproductive process. I. Effects on six indices at birth. *Pediatric Research 9*(2), 70.

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

Stein, Z., M. Susser, G. Saenger, and F. Marolla (1975). *Famine and human development: The Dutch hunger winter of 1944/45.* New York: Oxford University Press.

Stocker, C. J., J. R. Arch, and M. A. Cawthorne (2005). Fetal origins of insulin resistance and obesity. *Proceedings of the Nutrition Society 64*(2), 143–151.

Van den Berg, G. J. and M. Lindeboom (2018). Famines, hunger, and later-life health. In *Oxford Research Encyclopedia of Economics and Finance*.

Van den Berg, G. J., P. R. Pinger, and J. Schoch (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.

Wiegersma, A. M., A. Boots, T. J. Roseboom, and S. R. de Rooij (2022). Prenatal exposure to the dutch famine is associated with more self-perceived cognitive problems at 72 years of age. *BMC geriatrics 22*(1), 1–8.

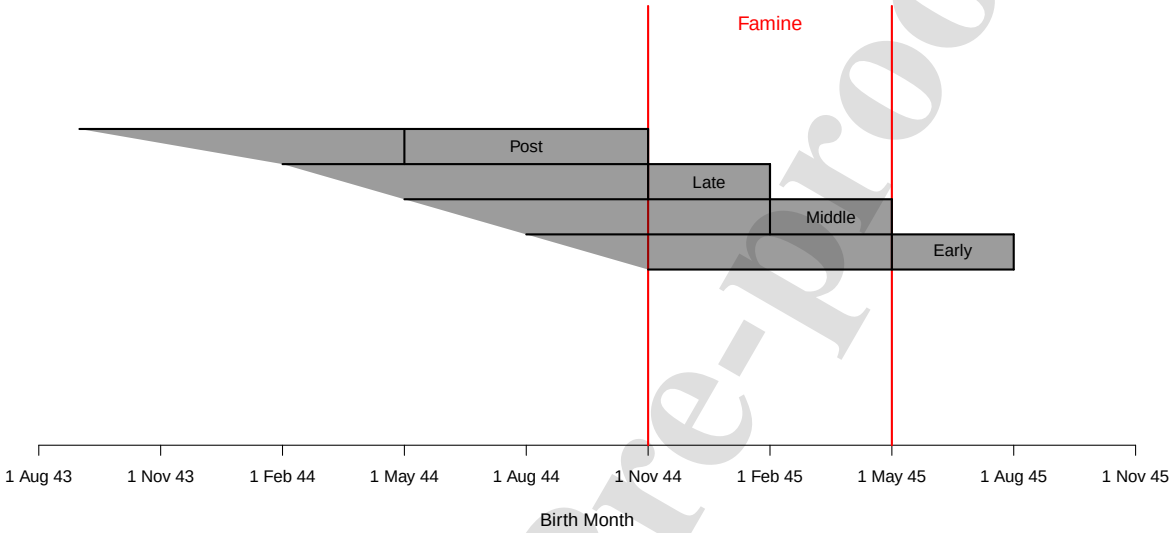
Xu, J., B. Yang, C. Yan, H. Hu, S. Cai, J. Liu, M. Wu, F. Ouyang, and X. Shen (2013). Effects of duration and timing of prenatal stress on hippocampal myelination and synaptophysin expression. *Brain research 1527*, 57–66.

Yi, J., J. J. Heckman, J. Zhang, and G. Conti (2015). Early health shocks, intra-household resource allocation and child outcomes. *The Economic Journal 125*(588), F347–F371.

Zheng, J., X. Xiao, Q. Zhang, T. Wang, M. Yu, and J. Xu (2017). Maternal low-protein diet modulates glucose metabolism and hepatic micrornas expression in the early life of offspring. *Nutrients 9*(3), 205.

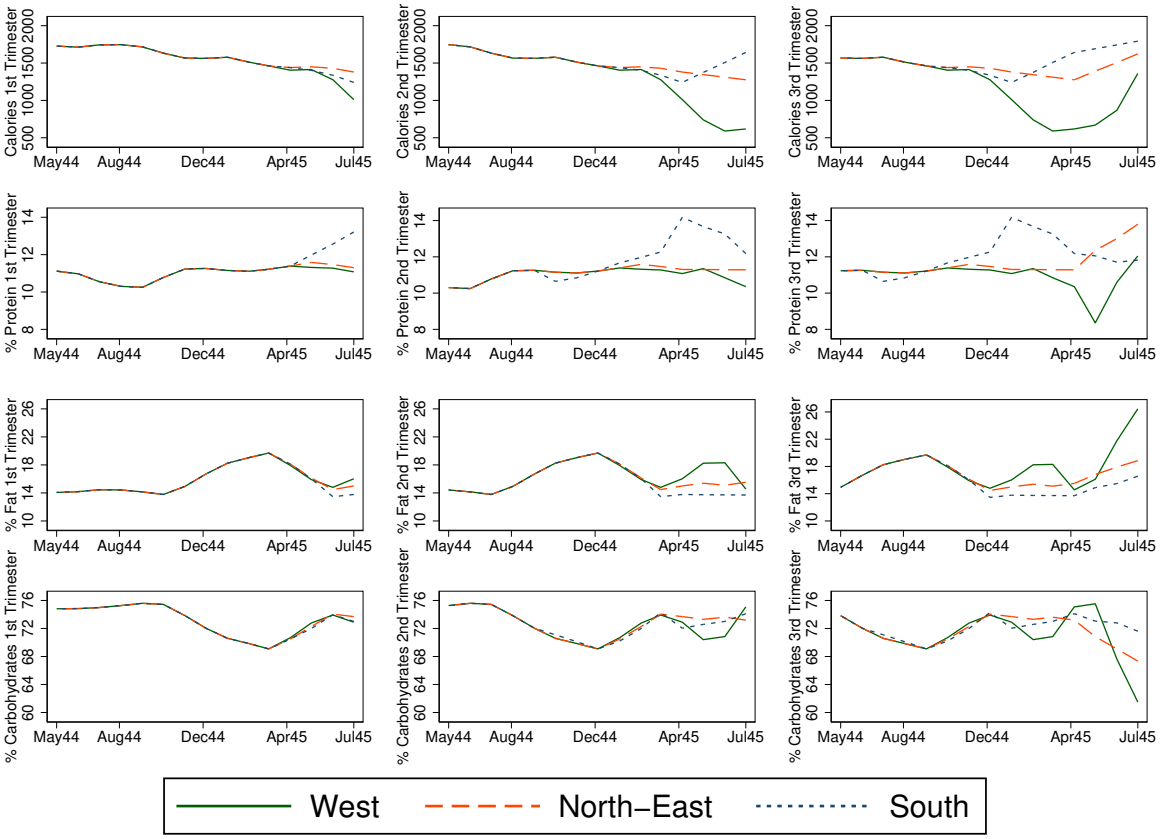
Figures and Tables

Figure 1: Treatment and control groups definition



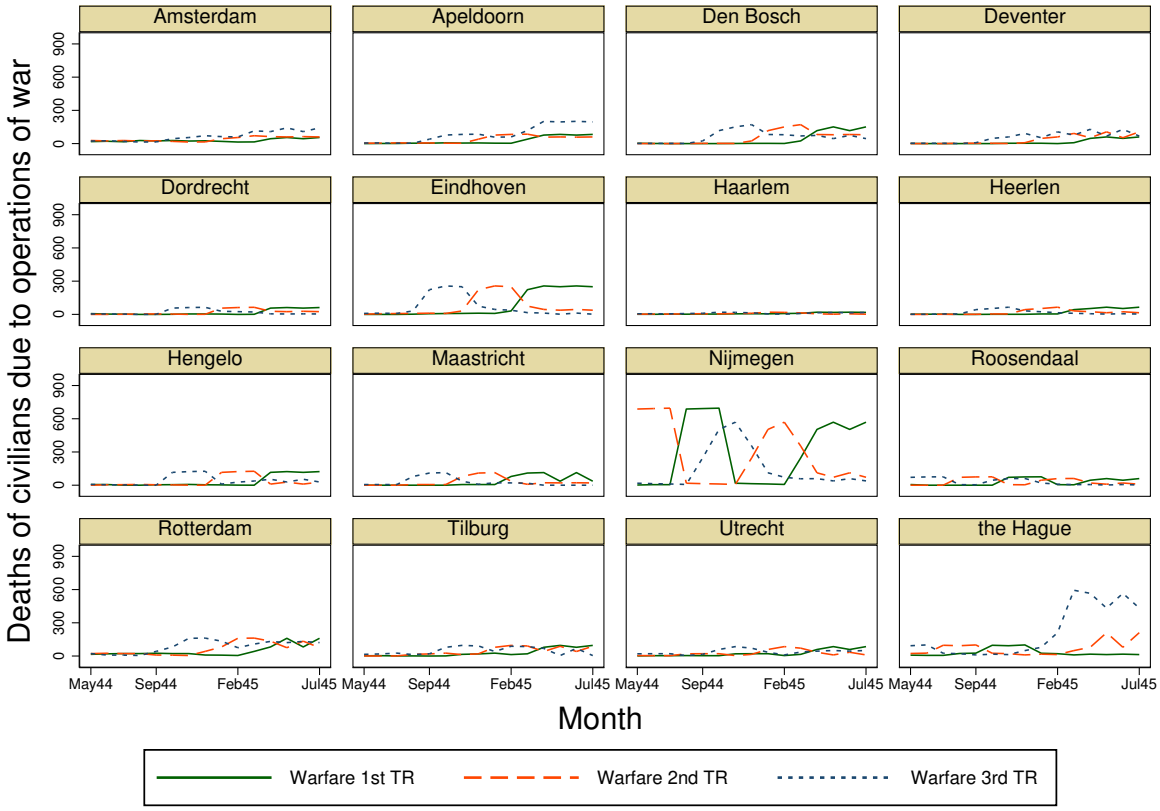
Note: The figure shows the time periods corresponding to each of the three treatment groups and the control group, with the birth months for each group enclosed in the respective grey boxes labelled "Post" (May-October 1944), "Late" (November 1944-January 1945), "Middle" (February-April 1945), "Early" (May-July 1945). The pregnancy period corresponding to each birth month is shaded in grey. The red vertical lines enclose the famine period.

Figure 2: Rations for each trimester by month of birth and region



Note: Data on calories (first row) and macronutrients shares (% protein, fat and carbohydrates in second, third and fourth row, respectively) at weekly level for the West and at monthly level for the North-East and South, by month of birth (shown on the horizontal axis). Source: official war information on the rations. Note that West here includes both the treated and the control cities. 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: Civilian deaths due to war operations during pregnancy for each trimester by month of birth and city



Note: Results for those born between May 1944 and July 1945 (month of birth displayed on the horizontal axis). The cities of Alkmaar, Almelo, Amersfoort, Bergen op Zoom, Bussum, Delft, Gouda, Groningen, Helmond, Kerkrade, Leiden, Zaandam, and Zwolle had little variation and are not shown here. Source: CBS registries, deaths with code "197 – Deaths of civilians due to operations of war".

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		✓		Groningen
	Schiedam		✓		Hengelo
	Velsen		✓		Nijmegen
	Vlaardingen		✓		Zwolle
	Vlissingen		✓	South	Bergen op Zoom
	Voorburg		✓		Eindhoven
	Zeist		✓		Heerlen
					Helmond
					's Hertogenbosch
					Kerkrade
					Maastricht
					Roosendaal
					Tilburg
				West	Alkmaar
				Non-Famine	Amersfoort
					Bussum
					Dordrecht
					Gouda
					Zaandam

Note: We start from a list of 46 cities with a population greater than 25,000 inhabitants on January 1, 1940. We exclude 4 municipalities which were classified as rural on the basis of the population dispersion patterns (i.e. the majority of the population was not living in the largest place of the municipality) and additional 13 municipalities because their population underwent major changes in size since 1930 (i.e. an increase or decrease by more than 50%). Source: the Historical Ecological Databank of the Netherlands (Boonstra, 2016).

Table 2: Newly Digitized Historical Data, Descriptive Statistics

Variable	Level	Mean	Std. Dev.	Min	Max
Temperature (°C) 1st TR	Monthly	8.54	4.75	1.70	16.70
Temperature (°C) 2nd TR	Monthly	8.07	4.91	1.70	16.70
Temperature (°C) 3rd TR	Monthly	9.88	4.70	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	1.49	0.26	0.59	1.75
Calories 2nd TR (1,000s)	Weekly	1.34	0.35	0.59	1.75
Calories 3rd TR (1,000s)	Weekly	1.31	0.37	0.59	2.05
Protein Share 1st TR	Weekly	11.11	0.60	10.25	14.18
Protein Share 2nd TR	Weekly	11.07	0.82	8.66	14.18
Protein Share 3rd TR	Weekly	11.28	0.94	8.66	14.18
South Weeks Liberated	Weekly	1.70	8.57	26.00	44.43

Note: The statistics refer to the cohorts in the analytical sample (May1944-July1945 in the selected cities). The number of observations for all the variables is 40,950. The calories and protein shares are transformed into negative values for the estimation.

Table 3: Difference-in-Differences Estimates of the Dutch Hunger Winter Effects on Age 18 Outcomes

	(1) Height	(2) Weight	(3) BMI (Weight/ Height^2)	(4) Overweight (BMI ≥25)	(5) Obese (Weight/ Height>120%)	(6) Underweight (BMI< 18.5)	(7) Chest/ Height Ratio	(8) Intellectual Disability (ICD-6 325)
WestF×Late	-0.099	-0.350*	-0.092	-0.010*	-0.006**	0.001	0.000	0.002
	(0.150)	(0.193)	(0.058)	(0.005)	(0.003)	(0.006)	(0.001)	(0.006)
WestF×Middle	-0.384	-0.334	-0.012	-0.006	0.001	-0.003	0.008***	0.010*
	(0.305)	(0.445)	(0.106)	(0.006)	(0.004)	(0.006)	(0.002)	(0.006)
WestF×Early	-0.052	0.652**	0.218***	0.009	0.006**	-0.013*	0.010***	0.005
	(0.298)	(0.313)	(0.075)	(0.006)	(0.003)	(0.007)	(0.002)	(0.006)
$F_{Early=Late}$	0.021	10.417	15.527	9.759	18.552	3.499	13.326	0.261
p-value	0.885	0.003	0.000	0.004	0.000	0.072	0.001	0.614
$F_{Early=Middle}$	1.331	7.733	7.049	6.428	1.878	2.222	2.638	1.324
p-value	0.258	0.010	0.013	0.017	0.181	0.147	0.116	0.260
$F_{Middle=Late}$	1.090	0.002	0.598	0.328	3.322	0.366	13.448	4.801
p-value	0.305	0.969	0.446	0.571	0.079	0.550	0.001	0.037
Wild cluster bootstrap p-values:								
WestF×Late	0.530	0.102	0.140	0.078	0.062	0.824	0.729	0.742
WestF×Middle	0.303	0.586	0.914	0.410	0.837	0.687	0.001	0.094
WestF×Early	0.866	0.051	0.013	0.124	0.041	0.123	0.001	0.418
Romano & Wolf p-values:								
WestF×Late	0.822	0.093	0.154	0.082	0.058	0.935	0.935	0.935
WestF×Middle	0.277	0.558	0.893	0.461	0.893	0.824	0.000	0.065
WestF×Early	0.949	0.034	0.006	0.071	0.034	0.048	0.000	0.325
Observations	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	0.03

Note: Robust standard errors in parentheses clustered at the level of city. *** p<0.01, ** p<0.05, * p<0.1. *Late* refers to births November 1944-January 1945, *Middle* to February 1945-April 1945, and *Early* to May 1945-July 1945. Additional controls include father's occupation, number of older brothers, birth order and religion. Fixed effects 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 cities with postnatal exposure only (May-October 1944).

Table 4: Placebo Difference-in-Differences Estimates of the Dutch Hunger Winter Effects on Age 18 Outcomes

	(1) Height	(2) Weight	(3) BMI (Weight/ Height^2)	(4) Overweight (BMI $\geq$25)	(5) Obese (Weight/ Height>120%)	(6) Underweight (BMI< 18.5)	(7) Chest/ Height Ratio	(8) Intellectual Disability (ICD-6 325)
WestF×PlaceboLate	-0.009	-0.280	-0.081	-0.006	0.002	-0.001	-0.002	-0.001
	(0.132)	(0.197)	(0.061)	(0.005)	(0.004)	(0.007)	(0.002)	(0.007)
WestF×PlaceboMiddle	0.156	-0.268	-0.120	-0.000	0.000	0.001	-0.003	0.009
	(0.153)	(0.261)	(0.074)	(0.007)	(0.003)	(0.007)	(0.006)	(0.010)
WestF×PlaceboEarly	-0.122	-0.229	-0.044	0.008	0.003	0.005	-0.001	-0.001
	(0.160)	(0.209)	(0.052)	(0.007)	(0.003)	(0.007)	(0.005)	(0.008)
$F_{Early=Late}$	0.653	0.050	0.414	5.825	0.078	1.054	0.000	0.001
p-value	0.426	0.824	0.525	0.023	0.782	0.313	0.989	0.974
$F_{Early=Middle}$	2.887	0.031	1.515	1.416	0.771	0.360	3.717	2.807
p-value	0.100	0.861	0.229	0.244	0.387	0.553	0.064	0.105
$F_{Middle=Late}$	1.506	0.002	0.259	0.453	0.383	0.057	0.236	1.910
p-value	0.230	0.962	0.615	0.506	0.541	0.814	0.631	0.178
Observations	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	0.030

Note: Robust standard errors in parentheses clustered at the city level. *** p<0.01, ** p<0.05, * p<0.1. Exposure cohorts are constructed using the same windows as in the main estimation, but two years after: "PlaceboLate" refers to births in November 1946-January 1947, "PlaceboMiddle" to births in February-April 1947, and "PlaceboEarly" in May-July 1947. Control mean refers to the mean of the outcome for those born in the West Famine cities between May-October 1946. Outcomes, exposures and controls are defined as in Table 3.

Table 5: Triple Difference Estimates of the Dutch Hunger Winter Effects on Age 18 Outcomes

	(1) Height	(2) Weight	(3) BMI (Weight/ Height^2)	(4) Overweight (BMI ≥25)	(5) Obese (Weight/ Height>120%)	(6) Underweight (BMI< 18.5)	(7) Chest/ Height Ratio	(8) Intellectual Disability (ICD-6 325)
WestF×Late×War	-0.078	-0.095	-0.021	-0.004	-0.009	0.003	0.002	0.002
	(0.161)	(0.268)	(0.078)	(0.006)	(0.006)	(0.009)	(0.002)	(0.008)
WestF×Middle×War	-0.535	-0.064	0.108	-0.005	0.001	-0.003	0.012*	0.001
	(0.338)	(0.493)	(0.129)	(0.010)	(0.005)	(0.008)	(0.006)	(0.011)
WestF×Early×War	0.072	0.892**	0.265**	0.002	0.003	-0.018*	0.012*	0.005
	(0.296)	(0.392)	(0.101)	(0.009)	(0.005)	(0.010)	(0.007)	(0.009)
$F_{Early=Late}$	0.196	6.040	11.86	0.644	6.989	4.143	3.166	0.134
p-value	0.661	0.020	0.002	0.429	0.013	0.051	0.086	0.717
$F_{Early=Middle}$	3.495	5.992	2.733	0.603	0.163	3.938	0.013	0.383
p-value	0.072	0.021	0.109	0.444	0.690	0.057	0.910	0.541
$F_{Middle=Late}$	2.225	0.005	1.608	0.004	4.655	0.518	3.445	0.014
p-value	0.147	0.947	0.215	0.949	0.040	0.477	0.074	0.906
Wild cluster bootstrap p-values:								
WestF×Late	0.626	0.759	0.808	0.497	0.179	0.752	0.335	0.790
WestF×Middle	0.181	0.908	0.456	0.686	0.903	0.696	0.148	0.920
WestF×Early	0.827	0.029	0.025	0.804	0.601	0.129	0.156	0.589
RW p-values:								
WestF×Late	0.950	0.966	0.966	0.882	0.242	0.966	0.641	0.966
WestF×Middle	0.129	0.996	0.668	0.944	0.996	0.947	0.057	0.996
WestF×Early	0.920	0.027	0.002	0.920	0.888	0.102	0.079	0.874
Observations	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	0.03

Note: Robust standard errors in parentheses clustered at the city level. *** p<0.01, ** p<0.05, * p<0.1. The estimation is performed on the original (i.e. births May 1944-July 1945) and the placebo (i.e. births May 1946-July 1947) samples combined, using a triple interaction term for area, trimester and being in the original (war) sample.Outcomes, exposures and controls are defined as in Table 3.

Table 6: Difference-in-Differences Estimates of the Dutch Hunger Winter Effects on Age 18 Outcomes – With Nutrition and Warfare Shocks

	(1) Weight	(2) BMI (Weight/ Height^2)	(3) Overweight (BMI ≥25)	(4) Obese (Weight/ Height>120%)	(5) Underweight (BMI< 18.5)	(6) Chest/ Height Ratio
West×Late	-0.387	-0.046	-0.008	-0.005	0.001	-0.002
	(0.236)	(0.070)	(0.006)	(0.003)	(0.008)	(0.002)
West×Middle	-0.459	0.007	-0.001	0.006	0.001	0.003
	(0.361)	(0.098)	(0.008)	(0.004)	(0.010)	(0.004)
West×Early	0.462	0.146	0.008	0.005	-0.003	0.004
	(0.347)	(0.105)	(0.009)	(0.005)	(0.011)	(0.004)
log(Warfare1stTR+1)	0.033	-0.003	0.001	0.000	0.001	0.000
	(0.050)	(0.009)	(0.001)	(0.001)	(0.001)	(0.000)
log(Warfare2ndTR+1)	0.032	0.020	0.003**	-0.000	-0.001	-0.000
	(0.047)	(0.014)	(0.001)	(0.001)	(0.001)	(0.000)
log(Warfare3rdTR+1)	0.062	0.025*	0.000	0.000	-0.001	0.000
	(0.064)	(0.015)	(0.001)	(0.001)	(0.001)	(0.000)
-Calories1stTR	0.820	-0.034	0.045*	0.019	0.042	0.005
	(0.856)	(0.231)	(0.023)	(0.015)	(0.026)	(0.005)
-ProteinShare1stTR	0.048	0.036	-0.005	0.007	0.005	0.001
	(0.325)	(0.074)	(0.007)	(0.004)	(0.008)	(0.001)
-Calories2ndTR	-1.107	-0.209	-0.030*	-0.026*	-0.015	-0.005
	(0.824)	(0.201)	(0.015)	(0.013)	(0.017)	(0.003)
-ProteinShare2ndTR	0.206	0.067	-0.002	0.002	-0.010**	0.001*
	(0.142)	(0.041)	(0.004)	(0.002)	(0.005)	(0.001)
-Calories3rdTR	-0.670	-0.257	-0.011	-0.002	0.028**	0.005
	(0.580)	(0.201)	(0.016)	(0.006)	(0.011)	(0.004)
-ProteinShare3rdTR	0.431***	0.110***	0.007**	0.004**	-0.005	0.001**
	(0.103)	(0.022)	(0.003)	(0.001)	(0.003)	(0.000)
Wild cluster bootstrap p-values:						
West×Late	0.128	0.519	0.258	0.117	0.951	0.360
West×Middle	0.265	0.947	0.900	0.188	0.949	0.603
West×Early	0.203	0.245	0.391	0.331	0.810	0.382
log(Warfare1stTR+1)	0.568	0.701	0.543	0.595	0.426	0.732
log(Warfare2ndTR+1)	0.509	0.152	0.049	0.398	0.315	0.238
log(Warfare3rdTR+1)	0.497	0.224	0.855	0.521	0.362	0.661
-Calories1stTR	0.371	0.889	0.067	0.240	0.115	0.350
-ProteinShare1stTR	0.891	0.631	0.575	0.124	0.588	0.334
-Calories2ndTR	0.211	0.331	0.066	0.066	0.387	0.131
-ProteinShare2ndTR	0.172	0.125	0.615	0.417	0.033	0.083
-Calories3rdTR	0.282	0.269	0.507	0.754	0.013	0.231
-ProteinShare3rdTR	0.002	0.000	0.034	0.019	0.166	0.027
RW p-values:						
West×Late	0.140	0.561	0.257	0.140	0.936	0.328
West×Middle	0.275	0.994	0.994	0.226	0.994	0.753
West×Early	0.332	0.318	0.424	0.424	0.746	0.424
log(Warfare1stTR+1)	0.822	0.822	0.822	0.822	0.752	0.822
log(Warfare2ndTR+1)	0.487	0.225	0.035	0.487	0.487	0.295
log(Warfare3rdTR+1)	0.567	0.128	0.829	0.626	0.567	0.678
-Calories1stTR	0.309	0.866	0.053	0.210	0.109	0.283
-ProteinShare1stTR	0.842	0.735	0.735	0.138	0.735	0.516
-Calories2ndTR	0.156	0.295	0.073	0.071	0.295	0.104
-ProteinShare2ndTR	0.105	0.066	0.426	0.364	0.020	0.041
-Calories3rdTR	0.254	0.201	0.514	0.713	0.010	0.161
-ProteinShare3rdTR	0.000	0.000	0.018	0.012	0.052	0.018
Observations	42,826	42,826	42,826	42,826	42,826	42,826
Controls and FE	✓	✓	✓	✓	✓	✓

Note: Robust standard errors in parentheses clustered at the level of city. *** p<0.01, ** p<0.05, * p<0.1. Additional 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. Calories are in thousands, see Table 2 for summary statistics.

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

	(1) Birth Weight	(2) Log Birth Weight	(3) Low Birth Weight	(4) Birth Length	(5) Head Circumference	(6) Placenta Weight	(7) Gestational Age	(8) Male
West×Late	-76.681	-0.021	-0.001	0.010	0.887*	-38.839	-0.079	0.030
	(87.265)	(0.028)	(0.034)	(0.372)	(0.508)	(26.568)	(0.286)	(0.057)
West×Middle	-168.993**	-0.050*	-0.007	-0.522	0.250	-39.760*	0.050	0.052
	(82.816)	(0.028)	(0.036)	(0.381)	(0.411)	(22.748)	(0.299)	(0.052)
West×Early	-76.137	-0.025	-0.008	-0.320	0.181	-43.531*	-0.023	0.004
	(78.591)	(0.026)	(0.033)	(0.379)	(0.409)	(23.034)	(0.298)	(0.053)
$F_{Early=Late}$	0.000	0.030	0.058	1.137	2.507	0.043	0.036	0.243
p-value	0.994	0.863	0.962	0.493	0.155	0.967	0.809	0.668
$F_{Early=Middle}$	1.981	1.174	0.002	0.470	0.055	0.048	0.058	1.117
p-value	0.159	0.279	0.862	0.084	0.815	0.836	0.849	0.622
$F_{Middle=Late}$	1.377	1.187	0.030	2.983	2.025	0.002	0.195	0.184
p-value	0.241	0.276	0.809	0.286	0.114	0.826	0.658	0.291
RW p-values								
West×Late	0.790	0.853	0.998	0.998	0.361	0.483	0.987	
West×Middle	0.177	0.320	0.970	0.466	0.889	0.320	0.970	
West×Early	0.794	0.788	0.950	0.848	0.950	0.320	0.950	
Observations	1,931	1,931	1,931	1,869	1,275	1,295	1,259	3,697
Controls and FE	✓	✓	✓	✓	✓	✓	✓	✓
Control Mean	3,347.86	8.098	0.073	50.323	38.657	604.348	39.519	0.531

Note: Columns (1)-(6) are based on males only sample. Column (7) sample includes all births (males and females). 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 Estimates of the Dutch Hunger Winter Effects on Age 18 Outcomes with Sample Selection Correction using Copulas

	(1) Height	(2) Weight	(3) BMI (Weight/ Height^2)	(4) Overweight (BMI ≥25)	(5) Obese (Weight/ Height>120%)	(6) Underweight (BMI< 18.5)	(7) Chest/ Height Ratio	(8) Intellectual Disability (ICD-6 325)
WestF×Late	-0.224	-0.445**	-0.081	-0.003	-0.003	0.003	0.001	0.007*
	(0.143)	(0.180)	(0.055)	(0.003)	(0.002)	(0.004)	(0.001)	(0.004)
WestF×Middle	-0.522*	-0.765*	-0.143	0.002	0.004	0.006	0.007***	0.014***
	(0.303)	(0.438)	(0.097)	(0.004)	(0.003)	(0.004)	(0.002)	(0.004)
WestF×Early	-0.125	0.492	0.188**	0.007*	0.005**	-0.008*	0.011***	0.004
	(0.300)	(0.314)	(0.074)	(0.004)	(0.002)	(0.005)	(0.003)	(0.004)
$F_{Early=Late}$	0.084	8.511	11.831	7.023	12.802	4.830	13.310	0.640
p-value	0.772	0.004	0.001	0.008	0.000	0.028	0.000	0.424
$F_{Early=Middle}$	1.662	10.858	17.166	1.759	0.056	10.255	7.042	10.946
p-value	0.197	0.001	0.000	0.185	0.812	0.001	0.008	0.001
$F_{Middle=Late}$	1.216	0.679	0.431	1.413	5.794	0.443	8.384	9.242
p-value	0.270	0.41	0.512	0.235	0.016	0.506	0.004	0.002
Observations	45,614	45,614	45,614	45,614	45,614	45,614	45,614	45,614
Controls and FE	✓	✓	✓	✓	✓	✓	✓	✓
Copula	Clayton	Clayton	Clayton	Joe	Plackett	Joe	Gumbel	Joe
Kendal's τ	0.170	0.490	0.535	-0.702	0.002	-0.713	0.712	-0.756
Wald Test	61.58	680.46	905.92	3333.71	4900	2118.9	51.18	271.86
p-value	<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). Fixed effects are included for city and for month of birth in the outcome equation. The selection equation includes dummies for West and Late, Middle, Early. Wald Test for independence of the error terms in selection and outcome equations. The results for the selection equation are presented in Table D5. Results without selection correction are reported in Table 3. Outcome and *Early, Middle, Late* definitions as in Table 3.

- The Dutch Hunger Winter was an abrupt famine occurring at the end of World War II in the well-nourished Western Netherlands (November 1944 - May 1945).
- We link high-quality military recruits data with objective health measurements for the affected cohorts with newly digitised historical records on calories and nutrient composition of the war rations, and warfare deaths.
- The cohorts exposed to the Dutch Hunger Winter since early gestation have an increased probability of being obese at age 18.
- This effect is partly moderated by warfare exposure and a reduction in energy-adjusted protein intake.
- We also finds evidence of both selection and scarring effects in the affected cohorts.

Authors' contributions: GC and SP: research design, data analysis and paper writing; PE: 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.