TI 2024-027/V
Tinbergen Institute Discussion Paper

Estimating the Lifecycle Fertility Consequences of WWII Using Bunching

Esmée Zwiers[1]

1 University of Amsterdam and Tinbergen Institute

Estimating the Lifecycle Fertility Consequences of WWII Using Bunching*

Esmée Zwiers†

April 15, 2024

Abstract

In the Netherlands, an immediate baby boom followed the end of WWII and the baby bust of the 1930s. I propose a novel application of the bunching methodology to examine whether the war shifted the timing of fertility or changed women's completed fertility. I disaggregate the number of births by age for cohorts of mothers, and estimate counterfactual distributions of births by exploiting that women experienced the war at different ages. I show that the rise in fertility after the liberation did not make up for the "missed" births that did not occur prior to the war, as fertility would have been 9.4% higher in absence of WWII.

Keywords: Lifecycle fertility, bunching, World War II, The Netherlands
JEL codes: J11, J13, J18, N34, N44

*I am grateful to Alícia Adsera, Rachel Anderson, Martha Bailey, Leah Boustan, Adriana Corredor Waldron, Janet Currie, Anne Gielen, Kyra Hanemaaijer, Frank Kleibergen, Olivier Marie, Karsten Muller, Lira Mota, Jonathan Payne, Sebastian Toft Roelsgaard, Andrei Tuiu, Yinuo Zhang, seminar participants at Princeton University, the 2023 Economic Demography Workshop, and 2023 EALE meetings for their helpful comments. I thank Peter Ekamper for sharing his data, and Statistics Netherlands for assisting me in accessing and using their data. Generous financial support from Princeton University's Center for Health and Wellbeing (CHW) is gratefully acknowledged. All errors are my own.

†Department of Quantitative Economics, Amsterdam School of Economics, University of Amsterdam, IZA and Tinbergen Institute. Email: e.s.zwiers@uva.nl

1 Introduction

Many developed countries experienced relatively low fertility during the 1930s and a rise in fertility after the Second World War (WWII). The post-WWII fertility response is widely known as the baby boom, with varying lengths and magnitudes across countries (see Van Bavel and Reher, 2013). The economics literature studying explanations for the baby boom has focused mainly on the United States. One strand of literature argues that the baby boom occurred because economic aspirations in adulthood exceeded those in childhood for affected cohorts (Easterlin, 1966, 1971, 1987). Another strand of literature focuses on the role of relative prices and absolute incomes for fertility decisions (Becker, 1960, 1965), with technological progress in the household (Greenwood et al., 2005), changed labor market opportunities for women (Doepke et al., 2015; Brodeur and Kattan, 2022), and improved maternal health (Albanesi and Olivetti, 2014, 2016) as possible channels.

I instead make a connection between the post-war baby boom and the "baby bust" occurring in earlier years. I ask whether observed peaks in period-fertility rates can be interpreted as a boom when taking into account lower fertility rates in the prior period. This is important because two effects could be at play. The post-war fertility response could be explained by a *tempo* effect if births were postponed to after the war. However, if births do (not) occur because of the prevailing conditions there could be a *quantum* effect on cohort-fertility, warranting a different interpretation of a fertility increase after the war. Understanding these mechanisms is important as changes in the age structure of populations can have profound consequences for healthcare systems, pensions, economic growth, and the environment. Specifically, in the context of the Netherlands, I show that the post-WWII peak in fertility did not compensate for births that did not occur during WWII and the interwar period.

In this paper, I propose a novel application of the bunching methodology to estimate and visualize how cohort fertility is affected by external factors. Previous applications of the bunching methodology in the public finance literature as pioneered by Saez (2010), Chetty et al. (2011), and Kleven and Waseem (2013) examine gaps and notches in earnings distributions, and use the unaffected parts of these distributions to estimate what the affected parts would have looked like in absence of manipulation. Instead of using the information of one distribution to estimate a counterfactual, my approach leverages information on a number of distributions for subsequent cohorts to estimate a counterfactual. This approach can be applied more broadly to virtually any outcomes that show similar (age-)profiles for subsequent cohorts.

This empirical approach offers several advantages for studying fertility. First, as opposed to earlier papers that evaluate short- and long-run fertility effects at certain ages (e.g., at the end of childbearing ages) or at certain points in time, I propose a disaggregation of fertility by using the number of births by age for cohorts of mothers. The bunching method can then show fertility responses across the full lifecycle of cohorts, which is particularly relevant when women are exposed to a series of adverse conditions during childbearing ages. In these cases it is hard to disentangle which factors affected women's fertility decisions at what ages, and proxies for completed fertility may not be informative. Second, the bunching method can be applied to these complex fertility questions without making assumptions on the type of exposure and the age of exposure. Hence, it can be considered nonparametric. Third, the estimation of a counterfactual — that is, what fertility would have looked like in absence of the war — allows for a conceptualization of the baby bust and baby boom in terms of "extra" or "missed" births.

This paper is set in the Netherlands. The Netherlands experienced poor economic conditions after the first World War (WWI) and during the Great Depression, and was occupied by Nazi Germany during WWII. I show that birth rates were low over the course of the 1920s and 1930s, and a large increase in fertility is observed in 1946 after the end of WWII as the general fertility rate increased by 43.5% compared to pre-war levels. Interestingly, when examining completed fertility for birth cohorts of mothers, the post-war fertility response is not driven by any cohorts in particular. This motivates the use of a different approach — the bunching method — to study the lifecycle fertility effects of the war.

The setting of this paper in the Netherlands has a few advantages. First, the availability of registry data allows me to construct disaggregated cohort fertility (births by age) for different cohorts of mothers. Second, the experiences of the war varied considerably by location, which I exploit in my empirical strategy as well as when examining effect heterogeneity. This aspect also allows me to zoom in on the effects of the war on fertility decisions while abstracting from the mechanical effects of famine and a (temporary) absence of men on fertility. Third, most of the literature studying the baby boom has focused on the United States — one of the few countries where the fertility rise lasted until 1960 (Van Bavel and Reher, 2013). The Dutch setting is more representative of the post-war fertility response observed in other countries.

To explore the lifecycle effects of WWII, I exploit that women in different cohorts experienced the war at different ages. I propose an application of the bunching methodology to estimate a counterfactual distribution of births for each cohort of mothers. Bunching methodologies estimate what "manipulated" distributions would have looked like in absence of "manipulation" by exploiting information on the shape of "unmanipulated" parts of the distribution. In this paper, this implies that I use information on the distributions of births for women whose fertility was not affected by the war, or affected at different ages, to estimate counterfactual distributions of births; that is, what the distributions would have looked like in the absence of the war.

More specifically, I show that the distributions of births by age for subsequent birth cohorts of mothers differ by exactly one year. I then exploit that a fertility rise after the war can only be observed for women who were between the ages of 21 and 40 in 1945 — and is strongest for women in prime fertility ages — whereas this fertility response cannot be observed for women who were aged 15 to 20 in 1945. The latter group was too young for their fertility to be instantaneously affected by the war and serves as a control group. As fertility can change across cohorts over time, which implies that fertility of these younger cohorts is different from older cohorts, I introduce an additional control group by making a distinction between births in areas with and without air raids during WWII. I show that trends in fertility develop similarly in areas with and without air raids prior to the war, and that fertility differences across areas are unrelated with determinants that matter for the fertility rise, akin to a parallel pre-trends test.

The estimated cohort-specific counterfactual distributions of births exceed the actual distributions of births for most cohorts, which implies that the peaks in fertility that occurred right after WWII did not make up for the "missed births" in earlier years. I estimate that $255,472$ fewer children are born compared to the counterfactual for cohorts 1905–1924, which is 9.4% of the actual number of births for these cohorts. These effects are particularly salient for women in the oldest birth cohorts, as fertility is already much lower than predicted for them even before the start of the war due to the poor economic conditions in the interwar period. For women in cohorts 1905–1911, I estimate 21.8% fewer births compared to the counterfactual. Sensitivity checks reveal that these numbers are robust to different identification choices, estimating similar magnitudes of "missed births", with a lower bound of 4.9% for all cohorts over time.

Finally, I am the first to use the bunching method to shed light on the two hotly debated opposing schools of thought explaining the baby boom in the United States — i.e., Becker (1960, 1965) versus Easterlin (1966, 1971, 1987). I use historical data on unemployment rates to distinguish between areas that suffered below- or above-median unemployment rates during the Great Depression (1932–1938) in the Netherlands. Using my bunching approach, I estimate and plot how the number of "missed births" develops by area over time and by cohort. This exercise shows that fertility does not develop according to Easterlin's hypothesis, who would predict higher birth rates especially for those cohorts exposed to poor economic conditions in childhood. Instead, I find that poor economic conditions in childhood and adolescence have long-term negative effects on fertility. This can most likely be explained by a "scarring" of marriage and family formation opportunities due to poor economic conditions earlier in life, which is consistent with Becker's theory.

This paper shows that the post-war baby boom did not make up for the pre-war baby bust in the Netherlands, and is the first to comprehensively study the link between the baby bust of the 1930s and the post-war baby boom. Most related are two earlier papers that linked the baby bust and baby boom in the United States. Jones and Schoonbroodt (2016) build a theoretical framework that is used to link the baby bust with the baby boom. Chabé-Ferret and Gobbi (2018) argue that the reduced economic uncertainty experienced by cohorts entering childbearing ages after the war explains more than 60% of the rise in completed fertility during the baby boom in the United States. More broadly, this paper also adds to the literature studying the origins of the baby boom (Greenwood et al., 2005; Doepke et al., 2015; Albanesi and Olivetti, 2014, 2016; Brodeur and Kattan, 2022), and the 20th century demographic transition more generally (e.g., Bailey et al., 2014).

Studying the link between the baby bust and baby boom is not only interesting from a historical perspective but also because it allows for a reconciliation of the literature studying the effects of economic conditions on fertility (e.g., Dehejia and Lleras-Muney, 2004; Del Bono et al., 2012; Currie and Schwandt, 2014; Currie et al., 2015; Huttunen and Kellokumpu, 2016; Chevalier and Marie, 2017; Dettling and Kearney, 2023) and the literature studying the effect of others shocks (such as natural disasters and wars) on fertility (e.g., Lee, 1997; Lindstrom and Berhanu, 1999; Agadjanian and Prata, 2002; Heuveline and Poch, 2007; Nobles et al., 2015). This is particularly relevant considering the recent Covid-19 pandemic which combined a global pandemic, uncertainty, and adverse economic conditions. Drops in fertility were predicted in the United States during the pandemic, but instead Bailey et al. (2023) show that fertility increased for U.S. born mothers, and particularly for first births and for mothers with college education.

Finally, this paper adds to the bunching literature as introduced by Saez (2010); Chetty et al. (2011); Kleven and Waseem (2013). The new application put forward in this paper is most closely related to the application by Persson (2020) studying the marriage effects of social insurance in Sweden, which uses unaffected and affected cohorts to construct a counterfactual distribution of marriages in absence of a social insurance reform. I extend this idea by leveraging the similarity of distributions of births by age for many subsequent cohorts of mothers. This approach can be applied to study different fertility questions, but also more generally to virtually any outcome exhibiting similar (age-)profiles across cohorts.

2 Historical setting

This section outlines the historical background and describes how the birth rate and cohort fertility measures developed over the course of the twentieth century in the Netherlands.

2.1 Historical background

WWI and the interwar period Figure 1 shows the crude birth rate (CBR), the general fertility rate (GFR), and two cohort-based measures of "children ever born" to (married) women ages 41 and older.[1] The Figure shows that fertility rates sloped downward from the beginning of the 20th century until the start of WWII. The decline was interrupted by the onset of the first World War in 1914. The Netherlands was not involved in WWI and hence its army was not in combat. However, the Netherlands did mobilize 100,000 soldiers to maintain the neutrality of the country, which consequently affected fertility during the war, and after the war as men returned home (De Jong and De Graaf, 1999; De Graaf, 2008). Indeed, fertility rates declined during WWI and the influenza pandemic of 1918–1919, and show an uptick in 1920.

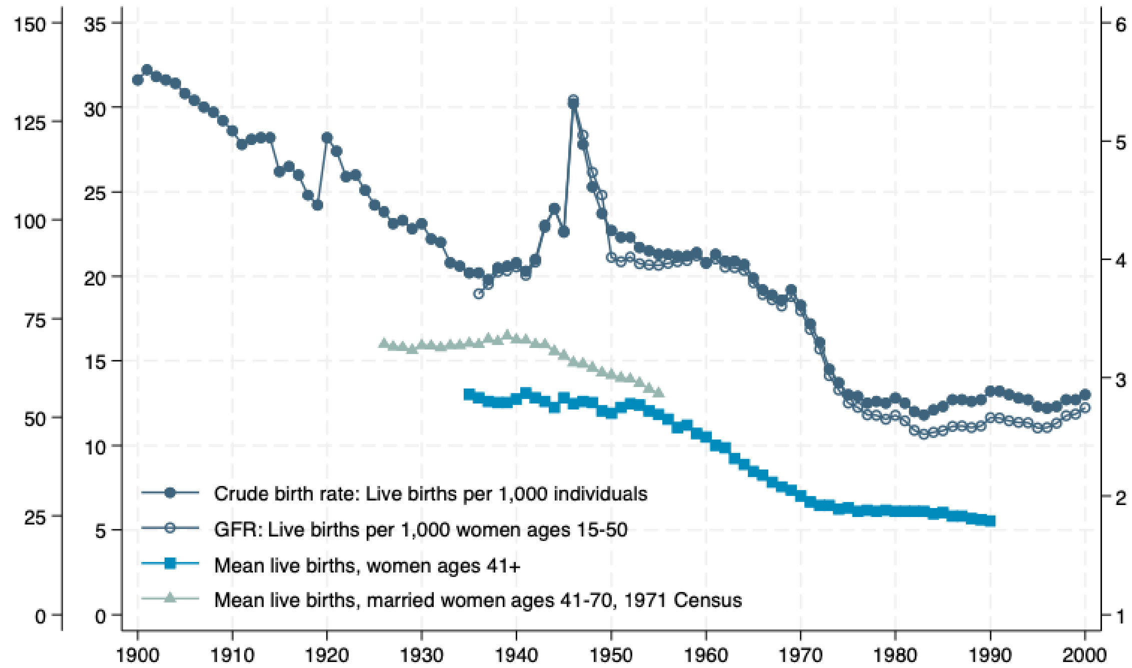
Note that until the 1960s in the Netherlands, whether and when individuals got married was mainly determined by whether they had the financial means to do so (De Jong and De Graaf, 1999). Fertility heavily depended on marriage, as only 1.6% of births were born out of wedlock between 1930–1960.[2] Unemployment rates were often high over the course of the 1920s, and even though the economy recovered in 1929, economic conditions became even worse with the onset of the Great Depression in 1931/1932. This can explain why the fertility rate declined through the 1920s, and even more through the 1930s with the start of the Great Depression. Unemployment rates in the Netherlands remained high up until the start of WWII (Kloosterman, 1985), and consequently fertility rates remained low.

WWII World War II started in September 1939 with the German invasion of Poland, and it took until May 1940 for Germany to invade the Netherlands. The occupation had a large effect on the Dutch population, and one of the direct consequences was the constant potential hazard from the air. Interestingly, Figure 1 shows that the fertility rate increased during the first years of the war. The increase can be explained by improved economic conditions in the first years after the start of the war, as the unemployment that prevailed in the Great Depression prior the

[1] The GFR is the desired measure to examine fertility patterns because it divides the number of live births by the number of women in childbearing ages. Unfortunately the GFR is only available from 1935, and therefore I use to CBR (the number of live births divided by population size) before 1935. Despite differences in the scale for the CBR and the GFR due to a different denominator, the trends in both measures look very similar post-1935, which suggests that the CBR should provide a reasonable approximation of the GFR before 1935.

[2] Calculated using CBS Statline table: *Bevolking, huishoudens en bevolkingsontwikkeling; vanaf 1899.*

Figure 1: Crude birth rate, general fertility rate (GFR), and completed childbearing, the Netherlands, 1900–2000



Notes: The crude birth rate (second left vertical axis) is the number of live births per 1,000 individuals in the population. The general fertility rate (GFR) (first left vertical axis) is the number of live births per 1,000 women ages 15-50 in the population (note that the rates for 1936–1949 are based on the number of women aged 15–45 in the population). Mean live births (right vertical axis) is the mean number of children ever born for each birth cohort (indexed to year by adding 25 years to a woman's year of birth; e.g., mean children ever born for birth cohort of 1910 corresponds to the year 1935 on the graph's horizontal axis). Data is reported for cohorts 1910–1965 who were 41 and older in 2006 — the year the data was reported. Mean live births for ever married women ages 41–70 in cohorts 1901–1930 are also reported.
Sources: Fertility rates are retrieved from Statistics Netherlands Statline (statline.cbs.nl), table: *Bevolking, huishoudens en bevolkingsontwikkeling; vanaf 1899.* Information on mean live births comes from (De Graaf, 2008) who reports data compiled by Statistics Netherlands up until 2006. The data for ever married women is compiled by the author using the 1971 census.

the war disappeared by 1941. Moreover, even after the collapse of the economy at the end of 1941, pro-family policies (i.e., tax benefits and child subsidies) were introduced by the occupying authorities during the war (Klemann, 2002; Te Slaa, 2016) and had a positive effect on fertility.

The increase in fertility during the war because of improved economic conditions could occur because of other war-related conditions specific to the Netherlands. First, the Dutch had access to sufficient food supplies in the first years of the occupation (Stein et al., 1975). Even though all food was rationed by 1941, the number of calories for the average adult declined only from 3,000 to 2,700 in the first years of the occupation (Trienekens, 1985). Second, healthcare did not necessarily deteriorate during the war (although access may have been more restricted for some than for others), as the numbers of doctors, dentists and midwives increased during the war (Klemann, 2002). Panel (a) of Figure A1 shows that the rates of stillbirths declined during war years. Moreover, panel (a) and panel (b) of Figure A1 show that infant mortality and mortality for individuals aged 1–65 was similar to pre-war levels in the early war years.[3] Third, the number of working women did not increase during the WWII in the Netherlands as in other countries, if anything, the number of married working women even decreased (Blok, 1988).[4]

Finally, and unlike in other countries, Dutch men were not away from home fighting during WWII. However, in 1942 Nazi Germany started to induce Dutch laborers to work in German factories. The number of Dutch men working in Germany was low in the early war years, but methods became more forceful and led to large raids in the cities of Rotterdam and the Hague in late 1944 (Sijes, 1990). A temporary absence of men could have affected fertility in later war years, but this was not the case in earlier years. The potential absence of men during war years also induced marriages, which could have affected fertility positively during the war. Peaks in the marriage rate (see Figure A3) can be observed in 1939 because married soldiers received higher remunerations, and in 1942 because married men faced lower probabilities for forced labor in Germany (De Graaf, 2008).

The end of WWII The Allied Forces started their attempt to liberate Western Europe in the spring of 1944. This led to the liberation of the southern part of the Netherlands in September 1944, but the Allied Forces were unable to proceed further to the north and west of the Netherlands. The Dutch "government in exile" called for a railroad strike in the still occupied parts of the country to support the Allied Forces. The occupying forces responded by enforcing an embargo on food transport to the urban west of the country. This, combined with an unusually early and harsh winter, led to a famine in the winter of 1944–1945 that is also known as the Hunger Winter (Scholte et al., 2015).

The drop in the birth rate in 1945 as shown in Figure 1 can be explained by the famine of 1944–1945. The famine mostly affected individuals living in the larger cities and towns in the west of the country, as individuals in rural areas were able to produce food and support themselves (Stein et al., 1975). Estimates suggest that the famine led to 15,000–25,000 deaths, which can explain the higher mortality rates in the late war years (Figure A1). The famine affected fertility as 36% of women exposed to the famine experienced irregular menses, and about 50% of women in the urban west did not menstruate at all (De Zwarte, 2019). The famine ended with the liberation of the Netherlands in May 1945.

It took until the end of March 1945 before the Allied Forces could resume their attempt to liberate the remaining parts of the Netherlands. The country was officially liberated on May 5, 1945. Figure 1 shows that a large peak in the fertility rate can be observed a year after the liberation. In 1946, 130.4 children per 1,000 women in childbearing ages were born, which is a 34.2% increase compared to the 1945 GFR (97.2 births per 1,000 women), and a 43.5% increase compared to the pre-war 1940 GFR (88.1 births per 1,000 women).

After WWII Even though the post-war peak in fertility was relatively short-lived, the birth rate remained relatively high after the war until the late 1960s. The 1950s and 1960s were

[3] Note that mortality increased in later later war-years due to the onset of poorer living conditions.

[4] The Netherlands is different from other countries for three reasons. First, it is important to note that the Netherlands was occupied and did not have an army that took away (working) men. Second, when men entered forced labor in Germany in later war years, it was seen as betrayal if women took over these jobs. Third, the Netherlands did not have a big war industry because it lacked the natural resources (Blok, 1988).

characterized by economic growth and good employment prospects for young individuals, which led to more marriages, and marriages at younger ages (De Jong and De Graaf, 1999). The drop in the birth rate that can be observed after 1970 can be explained by the liberalization of oral contraceptives for young women in the Netherlands (Marie and Zwiers, 2023). In the 1980s, the general fertility rate stabilized at around 48 births per 1,000 women in childbearing ages.

2.2 The impact of WWII on *cohort*-fertility

The *period* fertility rates show large effects of WWII but cannot explain the impact of the war on women in different birth cohorts. This is important because the period fertility rates do not show whether the war led women to delay births until later ages, or whether some of these delayed births never took place. The existing literature examines these *cohort*-specific effects by studying a cohort's mean number of live births at the end of their fertile ages (in this case, ages 41 and up).[5]

Figure 1 reports mean live births for women ages 41 and older for birth cohorts 1910–1965 from De Graaf (2008). I also compiled mean live births for ever married women ages 41-70 for birth cohorts 1901-1930 using the 1971 Census. The mean number of children born is a cohort-based measure, and for reporting purposes I advanced the series 25 years (which is approximately the period in which the birth cohort would start childbearing). Note that age at first birth was somewhat over 26 in 1950 (De Graaf, 2008). This implies that women born in 1910 (linked to 1935 in Figure 1) had on average 2.9 children in their life. This number remained about 2.7 and higher for birth cohorts 1910–1930 (linked to 1935–1955 in the figure) and dropped below 2 for cohorts 1946–1965 (linked to 1971–1990 in the figure), who were born after the war.

The cohort-based measure follows the period-based closely after 1950 but not for the period during WWII. The post-war birth peak cannot be observed in cohort-fertility rates, which implies that children born in 1946 — whose birth may have been delayed during the war — were born from mothers in different birth cohorts. This implies that, in this case, cohort-based measures do not show an effect of the war on lifecycle fertility. To study the effects of WWII on fertility I introduce a new application of the bunching method, which is able to distinguish between changes in fertility timing as well as completed fertility.

3 Data

3.1 Population-level administrative data

For the bunching analyses detailed information about all births that occurred over a cohort's lifecycle is needed. I use administrative population-level data from Statistics Netherlands (a detailed explanation of the data can be found in Appendix B1).[6] For individuals registered in a

[5]It is important to note that completed fertility measures suffer from biases. First, not all women may have completed fertility by age 41, and hence mean live births may be underestimated for younger birth cohorts. Second, if lower-income women are less likely to survive to later ages (and hence we are less likely to observe completed fertility in the 1971 census), and if lower-income women have more children on average this could bias completed fertility measures. Similarly, women who have more children are more likely to die during childbirth, and therefore may be less likely to be observed in censuses (Bailey et al., 2014).

[6]These data are available at a remote-access facility after signing a confidentiality agreement.

Dutch municipality by 1995 I can observe their month and year of birth and the year and month of birth of their parents.[7] Since I am interested in births to mothers in birth cohorts that were in childbearing ages during WWII, information on the child is used to infer the birth cohort of their mothers. I focus on individuals who were born in the Netherlands, and I merge in the information on the individual's municipality of birth to later match births to local circumstances around the time of conception/birth.[8] I take into account changes in municipal boundaries over time (the process is described in Appendix B2). The child's place of birth is likely a good representation of the mother's place of residence because by 1955 — the earliest year for which data is available — 76.1% of births were delivered at home in the Netherlands (De Vries, 2004).

I aim to study how fertility decisions respond to the depression and war. Especially during the war, there were frictions that prevented conceptions beyond choice. To illustrate, the 1944–1945 famine affected women's ability to conceive, and in some municipalities an absence of men might have made it difficult to conceive. For that reason I exclude births in areas that were most severely affected by the famine to make sure that this biological channel is not driving the results. Following Stein et al. (1975) and Scholte et al. (2015), I exclude the seven cities with a population over 40,000 in 1944 that were most severely affected by the famine (i.e., Amsterdam, Delft, Den Haag, Haarlem, Leiden, Rotterdam and Utrecht). The municipalities of Rotterdam and The Hague also suffered the largest absence of men in the fall of 1944.[9]

The final sample consists of $4,688,942$ individuals who were born in the Netherlands to mothers in birth cohorts 1905–1930, and for whom I can identify place of birth (99.8%). The mothers of these individuals were aged 15–40 at the time of the liberation in 1945 and hence in childbearing ages. I drop $1,075,384$ (22.9%) individuals who were born in the seven municipalities most severely affected by the famine. The final sample consists of $3,613,558$ individuals in 895 municipalities. This sample allows me to track births by age for cohorts of mothers born 1905–1930.

3.2 1971 Census

A limitation of the registry data is that reliable parent-child linkages cannot be made because the mothers of the births in the sample would have to still be alive in 1995, which is when some cohorts were already at relatively advanced ages (65–90). This implies that I cannot identify completed fertility using this data, nor do I have information on the mothers' marital state. For this reason, the registry data is supplemented with information from the 1971 Census.

I start with the 1971 full count population census (*VT1971*), which contains information on $13,133,333$ individuals. I harmonize municipal boundaries so that I can match individuals to war circumstances (the process is described in Appendix B2). I drop $126,439$ individuals for whom place of residence in 1971 and gender is unknown and am left with $13,006,894$ observations. For the analysis, I restrict the sample to women in birth cohorts 1901–1930 (who were ages 41–70 in 1971, and for whom a reliable measure of completed fertility can be determined), which leaves a sample of 1,984,667 individuals. After dropping individuals who were born in the seven

[7] I calculate that 82% of births that occurred between 1930–1960 are observed in my data as I do not observe individuals who died or moved abroad before 1995. Appendix B3 further discusses the quality of the data.

[8] The Dutch concept of a municipality is similar to the U.S. concept of a county.

[9] Section 5.3 shows that the estimated counterfactual densities look similar when including these municipalities.

municipalities most affected by the famine, a final sample of 1,487,168 individuals remains.

I create a variable detailing whether women were ever married in 1971. For individuals who were ever married, the census captures information on the number of live births a woman had up until 1971. For ever married women, I define a variable for remaining childless, which is equal to one if they never had a live birth. Fertility outcomes for women who were never married are unfortunately not available in the census. I also create a dummy variable for whether women were born in the Netherlands, and indicators for religious denomination (a distinction is made between Catholics, Orthodox Protestants, Liberal Protestants, other religions, and individuals without religious denomination). Finally, I create a indicator for individuals living in more urban areas, defined as municipalities with $20,000$ inhabitants or more.

3.3 Data on air raids during WWII

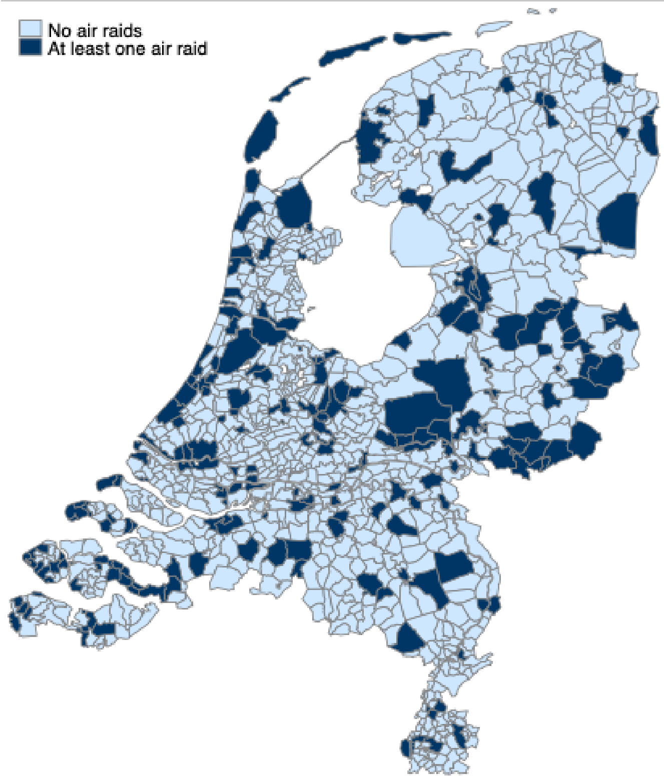
To estimate counterfactual distributions of births, a distinction is made between areas based on their exposure to more-severe war circumstances. To proxy for more-severe war circumstances, data on air raids across Dutch municipalities between 1940–1945 is used. This data comes from a report commissioned by the Dutch State Service for Monument Maintenance aimed at bundling information on damages in the Netherlands during WWII (Van Blankenstein, 2006). After the German occupation in May 1940, the Netherlands was targeted by about 600 air raids from the Allied Forces. Between 1940–1945, various air raids were aimed at military, maritime, and industrial targets (e.g., harbors, airports, military depots, industries aiding the occupation). Most air raids did not lead to civilian deaths, but about 500,000 civilians lost their home during the war. I digitized a chronological overview of air raids on pages 219–306 in Van Blankenstein (2006). The first air raids occurred in May 1940 with the start of the German occupation, and the last air raids occurred on April 26th of 1945 nearing the end of the war.[10]

After digitizing the data, I linked every air raid target to a Dutch municipality in 1946. I harmonized the municipalities so that they match the municipality changes as accounted for in the microdata and the 1971 census. Given that one air raid can have multiple targets (that is, municipalities), the observation unit is at the air raid-municipality level. I can identify 1,220 air raids that took place between 1940–1945, that targeted 20% of municipalities in the Netherlands (47% of these municipalities experienced more than one air raid). Table A1 shows that most air raids were executed in 1940 (25%) and 1944 (21.4%), and Table A2 shows that about 75% of air raids were executed by the Royal Air Force (RAF).

Figure 2 shows that air raids were geographically dispersed across the country. Panel (a) of Figure A2 shows that particularly the municipalities of Rotterdam (114 air raids), Vlissingen (66 air raids), Haarlemmermeer (65 air raids), Den Helder (63 air raids), and Gilze en Rijen (62 air raids) were affected. Not surprisingly, these municipalities were known during WWII for their harbors (Rotterdam and Vlissingen), their airports (Haarlemmermeer and Gilze en Rijen) or a military airport, naval base and a large dockyard (Den Helder). Panels (b), (c), and (d) of Figures A2 show the percentage of homes that got destroyed, heavily damaged and lightly damaged during WWII by province. The percentage of destroyed homes appears particularly

[10]The overview excludes exploration flights, air raids targeted at (rail)roads and hostile transport through the water. The placing of landmines by Allied Forces are also excluded from this overview.

Figure 2: Dutch municipalities that experienced at least one air raid, 1940–1945



Notes: Municipalities that experienced at least one air raid between 1940–1945. The data comes from Van Blankenstein (2006), pages 219–306.

high in the coastal provinces, whereas the distribution of damaged houses is more dispersed over the country.

4 Empirical Strategy

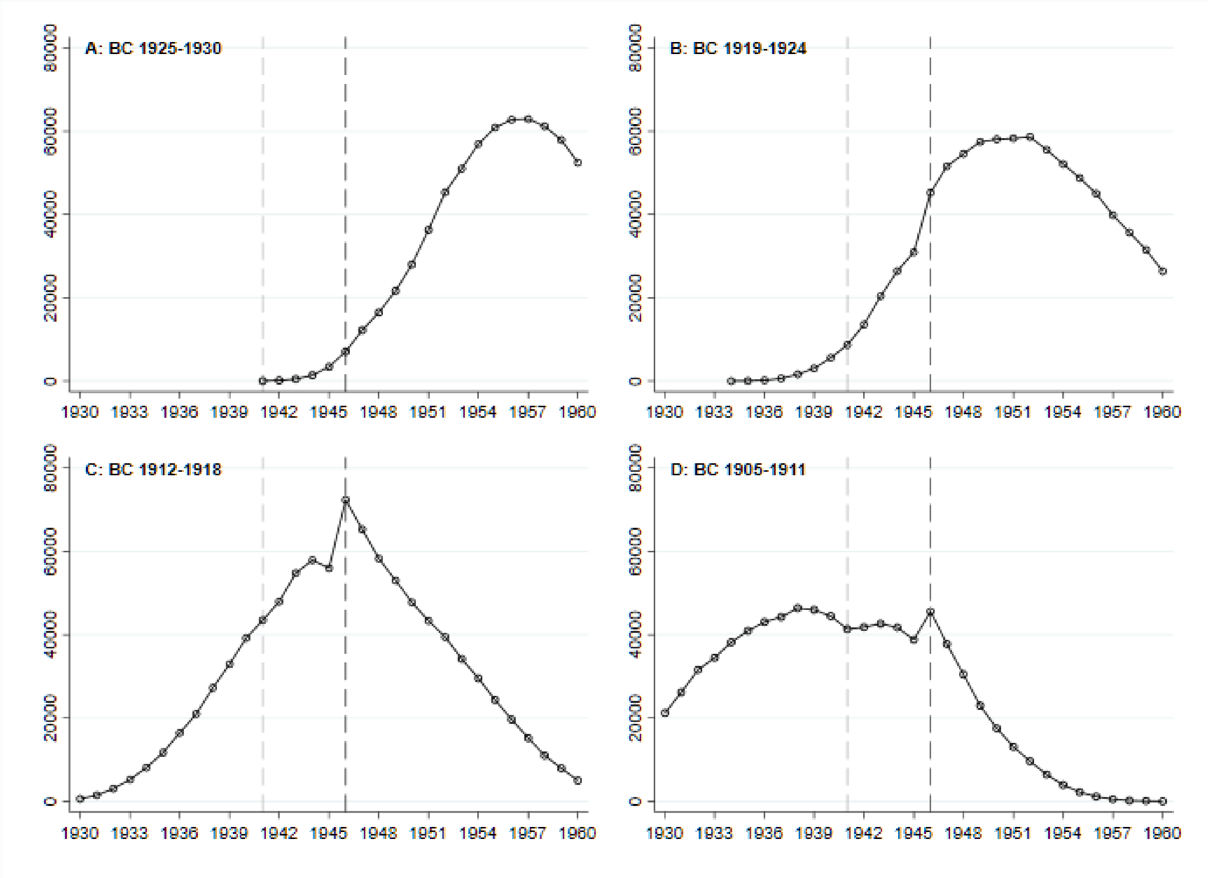
Fertility increased in the Netherlands after the war. It is unclear whether this fertility rise represents a catching up of births that did not occur due to the war or whether some of these postponed births never took place. This section proposes the bunching methodology as a new framework to analyze the lifecycle effects of fertility shocks.

4.1 Heterogeneity across maternal birth cohorts

Women were impacted differently by the Great Depression and war due to their age at the time they experienced these events. I identify four age groups of mothers to illustrate this: (1) women who were too young for their fertility to be contemporaneously affected by the war (birth cohorts 1925–1930, aged 15–20 in 1945), (2) women who were affected before peak fertility age (birth cohorts 1919–1924, aged 21–26 in 1945), (3) women who were affected during peak fertility age (birth cohorts 1912–1918, aged 27–33 in 1945), and finally, (4) women who were affected after peak fertility age (birth cohorts 1905–1911, aged 34–40 in 1945).[11]

[11]These groups of cohorts are formed based on observed patterns in fertility and mean age at birth. Mean age at birth in the administrative data was 28.0 in 1939 (prior to the start of WWII) for cohorts 1905–1930.

Figure 3: Empirical distributions of births by maternal birth cohort



Notes: Figure shows the empirical distributions of yearly births from 1930–1960 by birth cohort of the mother. Group A consists of births to women from birth cohorts 1925–1930 (ages 15–20 in 1945), group B consists of births to women in birth cohorts 1919–1924 (ages 21–26 in 1945), group C consists of births to women in birth cohorts 1912–1918 (ages 27–33 in 1945), and group D consists of births to women in birth cohorts 1905–1911 (ages 34–40 in 1945).

Figure 3 plots the number of births per year for the four groups defined above. The dashed lines reflect the year after the start of WWII and the year after the liberation to indicate when a response in the number of births would be expected. One can clearly observe the bell-shaped fertility curves for all four groups. For the group of women who were too young to be affected by the war (panel (a), birth cohorts 1925–1930, aged 15–20 in 1945), no effects of the liberation on fertility are visible. For the group that was affected at ages younger than peak fertility age (panel (b), birth cohorts 1919–1924, aged 21–26 in 1945), a fertility response is visible before the top of the bell curve is reached. For the women in peak fertility age (panel (c), birth cohorts 1912–1918, aged 27–33 in 1945), we see a spike exactly at the top of the bell curve. Finally the group that was affected beyond peak fertility age (panel (d), birth cohorts 1905–1911, aged 34–40 in 1945) experiences a peak beyond the top of the bell curve.

Table 1 shows differences across cohorts in the shapes of the distribution of births by maternal age as well as characteristics of mothers using the 1971 census. Panel A shows the characteristics of the distributions of births by age. Mean age at birth is higher for women in older birth

This informed the labeling of groups affected before/during/after peak fertility ages. Cohorts 1925–1930 were too young for their fertility to be instantly affected by WWII because they were minors (under 21) during the war.

Table 1: Summary statistics by mother's birth cohort

	All cohorts (1)	1925–1930 (2)	1919–1924 (3)	1912–1918 (4)	1905–1911 (5)
A: Microdata - distribution of births by age					
Maternal age at birth	30.88	29.83	30.56	31.33	31.82
Birth < age 20	0.015	0.018	0.014	0.015	0.014
Birth < age 25	0.158	0.182	0.164	0.141	0.146
Birth < age 30	0.460	0.531	0.485	0.420	0.405
Birth < age 35	0.750	0.826	0.774	0.729	0.670
Birth ≥ age 35	0.250	0.174	0.226	0.271	0.330
N	3,613,558	890,694	906,746	960,173	855,945
B: 1971 census - cohort characteristics					
Live births*	3.36	3.06	3.29	3.53	3.52
Childless*	0.093	0.078	0.087	0.090	0.111
Ever married	0.910	0.928	0.921	0.909	0.890
Dutch	0.939	0.948	0.944	0.940	0.926
Catholic	0.415	0.446	0.421	0.408	0.395
Orthodox Protestant	0.097	0.094	0.095	0.097	0.101
Liberal Protestant	0.290	0.253	0.277	0.298	0.317
Other religion	0.036	0.030	0.034	0.037	0.041
No religion	0.162	0.177	0.173	0.160	0.146
Urban	0.579	0.572	0.582	0.581	0.580
N*	1,352,645	330,297	314,551	306,368	271,060
N	1,487,168	355,912	341,490	337,108	304,632

Notes: Panel (a) shows characteristics of the distributions of the number of births by maternal age using the microdata for different groups of birth cohorts. Panel (b) show characteristics of birth cohorts using the 1971 census. Note that the number of observations is lower for "live births" and "childless" because this variable is only observed for women who were ever married.

cohorts.[12] The other measures indicate what proportion of births occurred before or after a particular age. The two youngest birth cohorts are more likely to have given birth before age thirty, which suggests that the distribution of births by age shifted to the left for younger cohorts of mothers. Panel B shows cohort characteristics of the 1971 census and shows that women in the oldest birth cohorts of 1905-1918 have highest completed fertility (i.e., number of live births by 1971). Note that the number of live births is lowest for the youngest cohorts, but also that their fertility outcomes are measured at ages 41–46, and not all of these women may have completed their fertility. Interestingly, the youngest cohorts are more likely to have been ever married by 1971. Women in the youngest cohorts are more likely to have been born in the Netherlands (although differences are minor). Differences in religious denomination and urbanicity are small.

4.2 Estimating counterfactual densities of births

To estimate counterfactual densities, that is, what the densities of births would have looked like in the absence of the Great Depression and the war, I exploit the timing of these events across mothers of different ages in a bunching methodology. Traditional applications of bunching methodologies in the public finance literature exploit discontinuities in the choice sets of individuals and firms, and show how behavior changes as a result of these different choice sets.

[12] Based on this data it is unclear if a higher mean age at birth is caused by a higher age at first birth, or rather that these older cohorts have more children. The latter can mechanically lead to a higher maternal age at birth.

Most early examples of these applications consider changes in tax rates and transfers (Saez, 2010; Chetty et al., 2011; Kleven and Waseem, 2013). These papers exploit that changed choice sets can generate gaps and notches in earnings distributions, and then use the unaffected parts of the earnings distribution to show what the affected part of the earnings distribution would have looked like in absence of manipulation.

I propose a new approach to using the bunching methodology. Bunching methodologies predict what manipulated distributions would have looked like in the absence of manipulation by exploiting information on the shape of the unmanipulated parts of the distribution. Instead of using information on the shape of the same distribution, I propose using a number of distributions for different cohorts to model the manipulated parts of the distribution.[13] The key assumption of bunching methodologies is that in the absence of the manipulation (in this case, WWII), the distribution of births over time can be estimated from the polynomial that is derived from the unmanipulated parts of the distribution. Intuitively, my identification strategy can be thought of as re-centering the distribution of births across maternal age at birth and then exploiting the similarities of these distributions for subsequent birth cohorts of mothers.

This idea is illustrated in Figure 4, where the number of births are now re-centered by maternal age at birth (in years) for all maternal birth cohorts from 1905 to 1930. Panel (a) shows these distributions for birth cohorts 1925–1930, who were aged 15–20 at the end of the war and too young to be affected by the depression and war. The distributions of births by age are very similar for subsequent cohorts of births. Panel (b) shows these distributions for birth cohorts 1919–1924 and who were aged 21–26 in 1945. Again the distributions of births for subsequent cohorts look very similar. Compared to Panel (a) these distributions look less smooth as the mothers in these birth cohorts were more strongly affected by the war and the depression. A fertility response after the liberation is particularly visible for the oldest cohort (i.e., 1919) in this group.

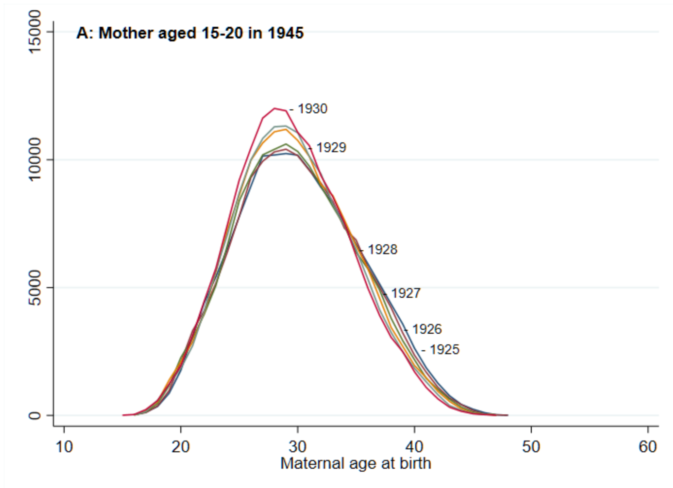
Panel (c) shows the distributions for birth cohorts 1912–1918. These mothers are between the ages of 27 and 33 at the end of the war, and hence are affected at peak fertility age. One can clearly see that the distributions of births for subsequent birth cohorts of mothers look similar, apart from experiencing a peak in the number of births exactly one year apart. Panel (d) shows the distributions of births by age for cohorts 1905–1911, who were aged of 34–40 in 1945. Again this figure shows that the distributions of births by age look very similar across subsequent birth cohorts of mothers, apart from experiencing the impact of war exactly one year apart.

Overall, Figure 4 shows that the distributions of births by age for subsequent maternal cohorts look very similar, despite the women's experience of the fertility consequences of war at different ages (depending on the mother's age at the time of the liberation). Mothers in the 1925–1930 birth cohorts appear to have been too young for their fertility to be directly affected by war. I use the properties of the distributions of these unaffected women (the first control group) as well as the densities of mothers who were affected by war — but at different ages — to estimate counterfactual distributions of births.

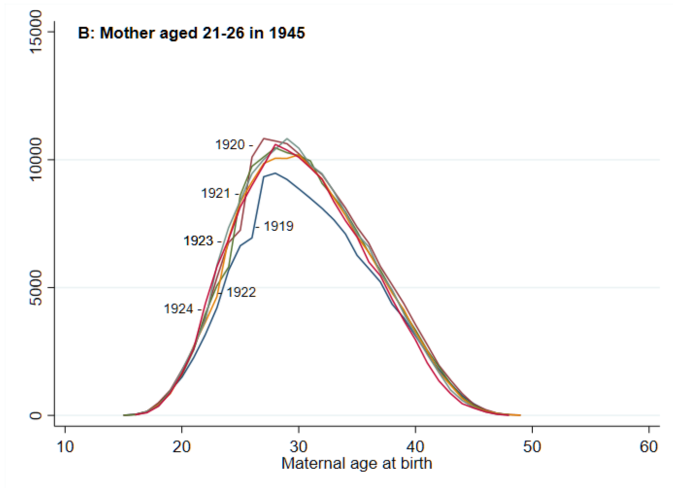
[13]This approach is most closely related to the bunching application by Persson (2020) who examines how a change in the financial benefits for married couples affected selection into marriage in Sweden. She exploits that different cohorts faced different incentives to marry earlier or later, and uses both distributions of marriages over time to estimate what would have happened in absence of the policy change in benefits for married couples.

Figure 4: Empirical distributions of births by age

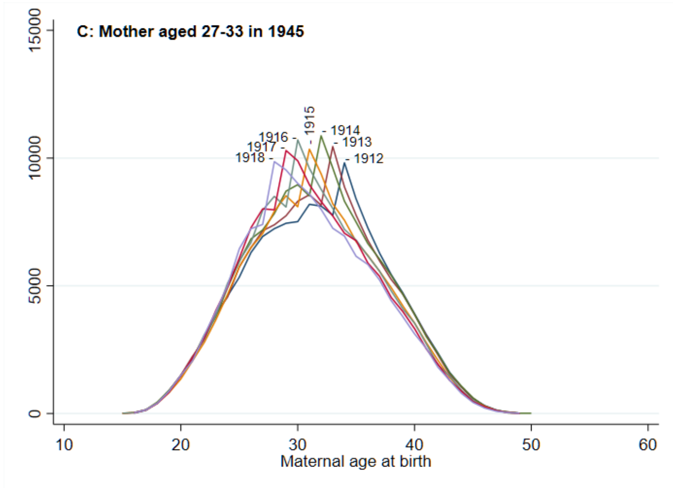
(a) Group 1: Cohorts 1925–1930



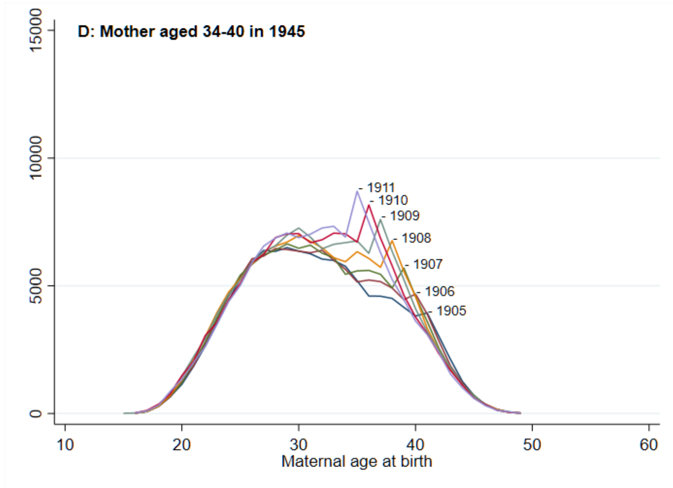
(b) Group 2: Cohorts 1919–1924



(c) Group 3: Cohorts 1912–1918



(d) Group 4: Cohorts 1905–1911



Notes: The figure shows the empirical distributions of births by maternal age at giving birth, by maternal birth cohort. Group A consists of births to women from birth cohorts 1925–1930, group B consists of births to women in birth cohorts 1919–1924, group C consists of births to women in birth cohorts 1912–1918, and group D consists of births to women in birth cohorts 1905–1911.

The key assumption of this approach is that the shapes of the distributions of younger unaffected cohorts (i.e., 1925–1930), and other cohorts who were affected by the war at different ages, are a good comparison group to estimate a polynomial for fertility in the absence of war. Table 1 showed that measures of completed fertility and marriage are different across cohorts, suggesting that younger cohorts may not be the ideal group to estimate a common polynomial. To construct a better counterfactual for fertility, while taking into account these differences across cohorts, I split the sample by whether the municipality was targeted by an air raid between 1940–1945. In Section 4.4 I show that fertility indicators develop similarly across cohorts in municipalities with and without an air raid before the start of the war (akin a test for common pre-trends), showing that introducing this distinction as an additional control group should allow for the estimation of a more reliable counterfactual.

4.3 Specification

A distinction is made between the treated sample, that is, women whose fertility was affected by WWII and the untreated sample, that is, women who were too young for their fertility to be directly affected by the war. The treated sample (T) is divided into 20 maternal birth cohorts. Each cohort $c \in \{1, ..., 20\}$ consists of women born from 1905 to 1924, and who are between the ages of 21 and 40 at the end of the war. The untreated sample (U) is divided into six maternal birth cohorts, where each cohort $c \in \{21, ..., 26\}$ consists of women born from 1925 to 1930 and thus were between the ages of 15 and 20 at the end of the war. A second control group is set up by exploiting regional variation in the incidence of air raids during the war, and specifically by making a distinction between areas with and without air raids during the war. This implies that for each cohort $c \in \{1, ..., 26\}$, I observe two distributions of births by maternal age, one for areas with at least one air raid and one for areas without air raids. Note that 20% of municipalities were targeted by air raids that house about 52% of births in the sample.

I collapse the data in the cohort sample at the cohort, area, and maternal age at birth (in years) level. In the spirit of Saez (2010), Chetty et al. (2011), Kleven and Waseem (2013), and Persson (2020), I estimate the following equation:

$$
\begin{aligned}
n_{cmt} &= \alpha + g(a_{ct}) + \mu_{cm} \\
&+ \sum_{c^*=1}^{20} \mathbb{1}[c^* = c]\Big(\mathbb{1}[t = 1946](\beta_{c^*} + \gamma_{c^*} * AnyAirRaid_m) \\
&+ \mathbb{1}[1941 \leq t \leq 1944](\phi_{c^*} + \theta_{c^*} * AnyAirRaid_m) \\
&+ \mathbb{1}[t = 1945](\delta_{c^*} + \psi_{c^*} * AnyAirRaid_m)\Big) + \varepsilon_{cmt},
\end{aligned}
\tag{1}
$$

where n_{cmt} denotes the natural logarithm[14] of the number of births to women in birth cohort c, in area m (targeted or not targeted by any air raid during the war), at age t; $AnyAirRaid_m$ is an indicator variable that takes a value of one for births occurring in areas with at least one air raid. Polynomial $g(a_{ct})$ is a higher-order polynomial in age (years), and μ_{cm} are cohort-area

[14]The cohort-specific distributions of births by age have nonlinear properties, which is why I use the natural logarithm: $n_{cmt} = \ln(N_{cmt})$.

fixed effects that allow for cohort- and area-specific shifts in the distribution of births by age $g(a_{ct})$.[15]

To control for nonlinearities in the densities of births by age as shown in Figure 4, I add controls for the cohort's age while experiencing the war. The magnitude of these responses to war circumstances are allowed to be different across birth cohorts and to be different within birth cohorts but across areas with and without air raids. Bunching at the notch, and hence the magnitude of the birth peak, is captured by β_{c^*} for births in areas without air raid and $\beta_{c^*} + \gamma_{c^*}$ for births in areas with at least one air raid. I also control for the cohort's age from 1941 to 1944, which are the years in which the children who were conceived during the war (1940–1943) would have been born. ϕ_{c^*} captures the cohort-specific impact of the war years on fertility for areas without air raids, and $\phi_{c^*} + \theta_{c^*}$ captures the cohort-specific impact of the war on fertility for areas with air raids. I include a separate indicator for births in 1945 (conceptions in 1944) to capture the cohort-specific impact of the famine for areas without air raid (δ_{c^*}) and with air raid ($\delta_{c^*} + \psi_{c^*}$).

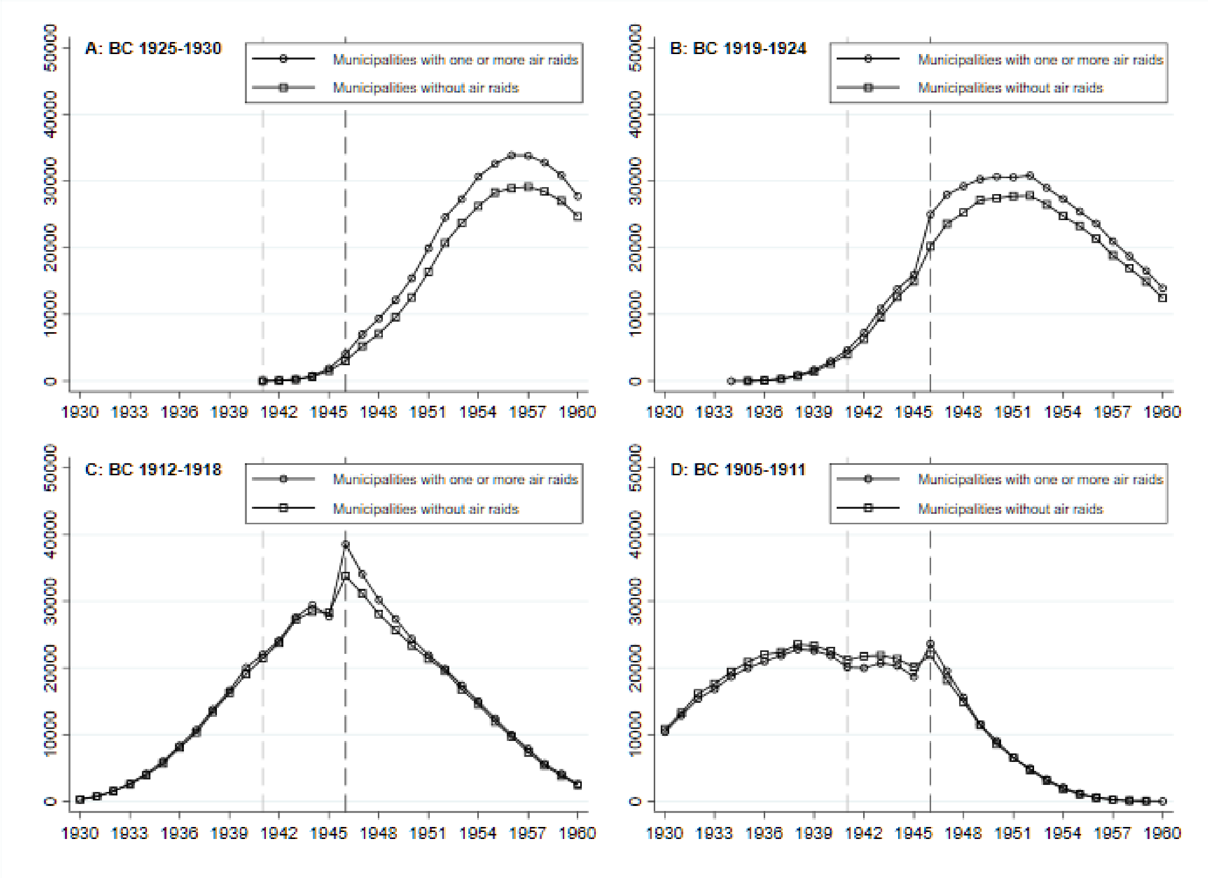
The intuition behind this empirical strategy is that I use information on the shapes of the distributions of both untreated cohorts ($c \in \{21, ..., 26\}$) as well as treated cohorts ($c \in \{1, ..., 20\}$) to estimate a common shape of distribution $g(a_{ct})$ that reflects what fertility of each cohort would have looked like in the absence of the war. Note that the distribution of births by age is allowed to be different across birth cohorts, and that shifts in these distributions are accounted for by the cohort-area fixed effects μ_{cm}. These fixed effects also account for area-level differences in time-invariant characteristics that could affect fertility.

Recovering the counterfactual distributions To recover the counterfactual distribution for each maternal birth cohort, I use the coefficients that are estimated in Equation (1), but set indicators $\mathbb{1}[t = 1946]$, $\mathbb{1}[1941 \leq t \leq 1944]$, and $\mathbb{1}[t = 1945]$ to zero. I then predict cohort- and area-specific frequencies to construct counterfactual densities that portray a world in which the war did not take place (hence for all birth cohorts in areas (not) targeted by air raids). I add the area-specific predicted frequencies (for areas targeted and not targeted by air raids) within each cohort to create counterfactual distributions of births for each maternal birth cohort.

In what follows I will refer to the concept of "missing" or "extra" births. At the places of the distribution where the counterfactual distribution of births exceeds the actual distribution of births, the actual number of births is lower than predicted by my model, which indicates "missing" births. The reverse is true when the distribution of the actual number of births exceeds the counterfactual distribution of births. In this case there were more births than predicted by my model, and this would indicate "extra" births.

[15] I use the Akaike information criterion (AIC) to determine the degree of polynomial best suitable for $g(a_{ct})$. I choose the functional form with up to fourth-order polynomials because these have the lowest AIC at 4298.477. This compares to 4687.781 for a model with up to a third-degree polynomial, and 4839.465 for a model with up to a second-degree polynomial. The good fit of the quartic polynomial is not surprising given the bimodal shapes of the empirical distributions of births.

Figure 5: Empirical distributions of births by birth cohort — by exposure to air raids



Notes: Figure shows the empirical distributions of yearly births from 1930–1960 by birth cohort of the mother. Group A consists of births to women from birth cohorts 1925–1930 (ages 15–20 in 1945), group B consists of births to women in birth cohorts 1919–1924 (ages 21–26 in 1945), group C consists of births to women in birth cohorts 1912–1918 (ages 27–33 in 1945), and group D consists of births to women in birth cohorts 1905–1911 (ages 34–40 in 1945).

4.4 Validity checks

Common pre-trends The main identifying assumption is that, in the absence of the war, the manipulated parts of the distribution can be estimated from the unmanipulated parts of the distribution. Hence, that fertility of women in different birth cohorts would have behaved similarly in the absence of war. Table 1 showed that fertility indicators are different for younger and older cohorts, and that younger cohorts may therefore not be a suitable control group to estimate a polynomial for older cohorts. For this reason, I make an additional distinction between municipalities with and without air raids during WWII.

Figure 5 shows the empirical distributions of births by maternal birth cohort while distinguishing between areas with and without air raids. The figure shows that the trends in number of births deviate during and after WWII, which could be explained by the prevalence of air raids across areas. Most importantly, this figure shows that before the onset of WWII (and hence before areas are exposed to air raids), fertility is similar across areas. This figure essentially functions as a test for common pre-trends for areas with and without air raids.

It is important to note that areas with air raids have different characteristics than areas without air raids. Table A3 shows that municipalities with air raids are larger (average population size of 30,924 compared to 7,739 for municipalities without air raids), and have a lower share of Catholics (36.1% compared to 43.2% for municipalities without air raids). This does not imply that areas with air raids are clustered together as Figure 2 shows that air raids were geographically dispersed across the Netherlands. Differences across urbanicity and religion may be important because fertility is different across religious groups for cohorts over time, and different between urban and rural areas.[16] Therefore it is important to assert if any differences in fertility and marriage outcomes for different birth cohorts across areas with and without air raids can be explained by factors like religion and urbanicity. That is, to assert that any similarity of trends is not driven by changing compositions of cohorts.

Panel (a) of Figure 6 plots the mean number of live births while making a distinction between areas with and without air raids using the 1971 census. Although the mean number of live births is higher in areas without air raids, this difference remains similar up until birth cohort 1915 (who were aged 25 at the start of WWII). The difference in mean live births across areas is converging for later cohorts, which may be explained by exposure to air raids during WWII for these cohorts of whom most did not yet begin family formation in 1940.[17]

To check if these differences across areas can be explained by individual-level characteristics that are relevant for fertility, I regress the individual-level number of live births on age in 1971 (linear and quadratic), an indicator for whether the individual was born in the Netherlands, indicators for religious denomination, a municipality-level indicator for population size, whether the municipality is considered an urban area (more than 20,000 inhabitants). I also control for the municipal liberation date compared to the official liberation date of the Netherlands (days from May 5th 1945, and an indicator for the municipality being liberated after May 5th 1945), to account for the fact that some areas where exposed to war longer than others.[18] I then save the residuals of this regression and collapse the data by birth cohort and at the area-level (with and without air raids). Panel (b) shows that after residualizing — i.e., controlling for individual- and area-level characteristics that could affect fertility — there are still differences in fertility between areas with and without air raids, which indicates that individual- and area-level characteristics cannot fully explain the fertility differences across areas. Most importantly, the mean number of live births still looks similar across areas with and without air raids for birth cohorts 1915 and earlier (that is, parallel pre-trends).

Panel (c) of Figure 6 shows the same exercise for the proportion of women who were ever married by 1971. The marriage rate looks very similar in areas with and without air raids over time, and in particular for cohorts before 1924. When residualizing this variable like described above, the trends still look similar over time as shown in panel (d). These exercises show that even though there may be differences in the levels of family formation measures across areas, the trends for cohorts that most likely started their fertile lives before WWII (i.e., cohorts 1915 and

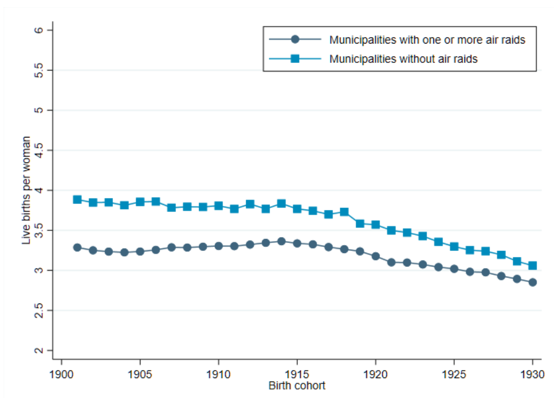
[16] Figure A4 shows that the mean number of live births is relatively high for Catholics and Orthodox Protestants compared those with other religious denomination and those without religious denomination. For marriage, especially those without religious affiliation are most likely to be ever married across all birth cohorts. Differences can also be observed across urban and rural areas. Panels (c) and (d) show that the mean number of live births is higher in rural areas, although trends seem to converge slightly for younger cohorts, the pattern in marriage is very similar in urban and rural areas across birth cohorts.

[17] Mean age at first birth was somewhat over 26 in 1950 (De Graaf, 2008).

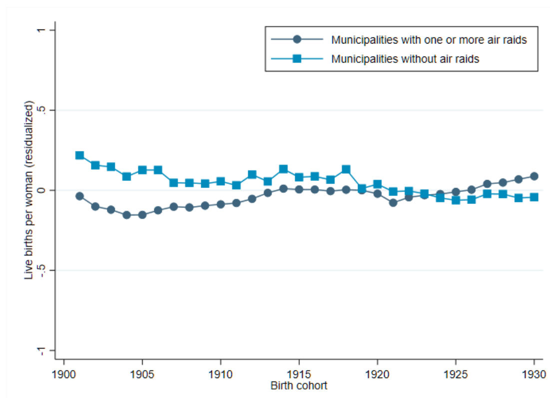
[18] Data on municipal liberation dates comes from Ekamper (2020).

Figure 6: Fertility and marriage across cohorts and areas with and without air raid

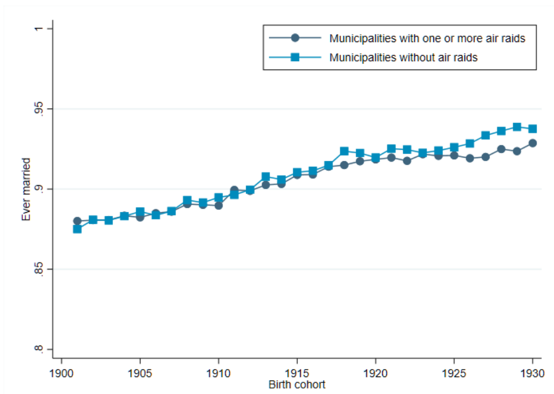
(a) Mean live births



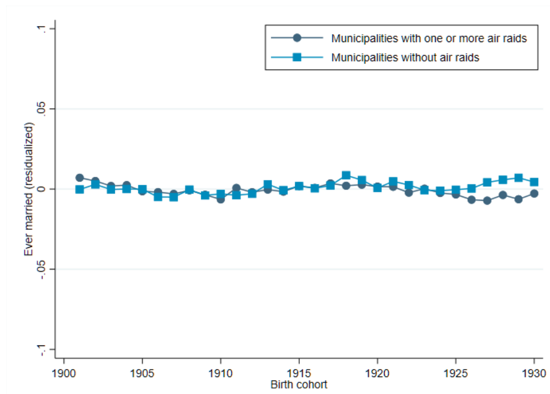
(b) Mean live births (residualized)



(c) Proportion ever married



(d) Proportion ever married (residualized)



Notes: This figure is compiled using data from the 1971 census. Panels (a) and (c) show the mean number of live births and the proportion of women who were ever married for every birth cohort while making a distinction between areas with and without air raid during WWII. Panels (b) and (d) show the mean residualized measures after controlling for individual-level (age and age-squared in 1971, born in the Netherlands, religious group) and municipality-level (population size, urbanicity, and municipal liberation date) characteristics.

earlier) look very similar in areas with and without air raids. This is also true after accounting for individual- and municipality-level characteristics that could explain fertility differentials. These tests provide evidence for the common pre-trend assumption when making a distinction between areas with and without air raids in the identification strategy.

Sensitivity to identification assumptions To estimate counterfactual distributions of births, various identification assumptions are made. I relax some of these assumptions to examine the sensitivity of these choices. First, the counterfactuals are estimated while not making a distinction between areas with and without air raids. This exercise examines the sensitivity of results if the counterfactual is only estimated using the distributions of younger and older cohorts, which are shown to be different over time. Second, I estimate counterfactual distributions of births by excluding the youngest cohorts (birth cohorts 1925–1930), whose fertility outcomes differ most from the older cohorts (birth cohorts 1905–1924). Finally, I estimate counterfactual distributions while including the seven cities most severely affected by the famine of 1944–45. These individuals are left out of the main analysis because the famine could have affected women's ability to conceive, and because some of these municipalities also experienced the biggest absence of men during the war, implying that fertility could have been affected by factors beyond choice in these areas.

5 Results

This section discusses the estimated counterfactual distributions of births, how they compare to the actual distributions of births, and what this eventually means in terms of thinking about the rise in births after the liberation in the sense of "extra" or "missing" births. I also check the sensitivity of results depending on the identification choices made.

5.1 Counterfactual distributions

Panels (a)–(d) of Figure 7 show the cohort-specific counterfactual distributions (in red) and distributions of births as observed in the data (in green) for the different age groups of mothers as specified before. For every group of birth cohorts (1925–1930, 1919–1924, 1912–1918 and 1905–1911), I show the actual and counterfactual distributions of births for four birth cohorts. The red dashed lines represent the ages of the women in the cohort in the year after the start of the war (1941, left line) and their age in the year after the war (1946, right line).[19]

Panel (a) shows the results for the birth cohorts 1925–1930: the mothers who were aged 15–20 in 1945. These women were likely too young to have their fertility directly affected by the war. This is also shown by the estimated counterfactual distributions that follow the distribution of actual births. It does appear that the distributions of actual births are somewhat to the left of the counterfactual distributions of births, which suggests that these cohorts started with childbearing at somewhat earlier ages as predicted by the counterfactual.

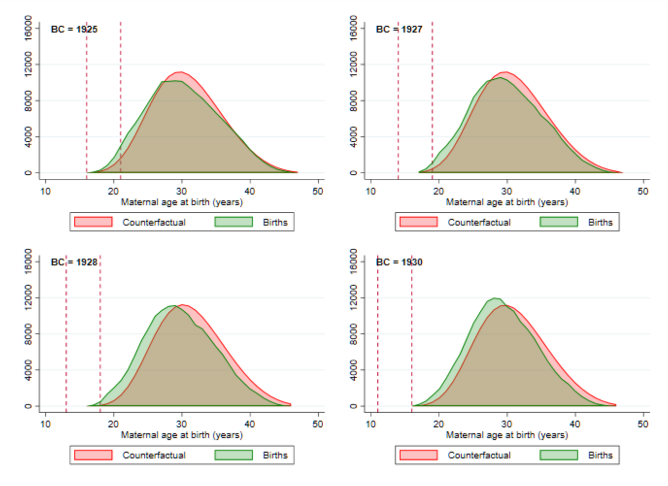
The results for birth cohorts of 1919–1924 are shown in Panel (b). These cohorts represent women who were affected by the war before peak fertility age and were aged 21–26 in 1945. Especially for the slightly older cohorts of 1919 and 1921, one can see that the war directly affected fertility and that a fertility rise is visible in the year after the war. The counterfactual distributions of births follow the actual distributions very closely. However, this is not true for the oldest birth cohort in this group. The counterfactual distribution of births (in red) exceeds the actual distribution of births (in green), which implies that the peak does not compensate for "missed" births that occurred during the war for the 1919 birth cohort.

The pattern observed for the 1919 birth cohort is more salient for birth cohorts 1912–1918 in Panel (c). These women were aged 27–33 in 1945 and were affected by the war around peak fertility age. The differences between the distributions of counterfactual and actual births is even more stark for these cohorts. These women had fewer births during the war and after the war than predicted, again implying that the post-war peak did not make up for missing births during the war. Notably, the discrepancy between both distributions started prior to the war, which could be the result of the poor economic conditions in the interwar period.

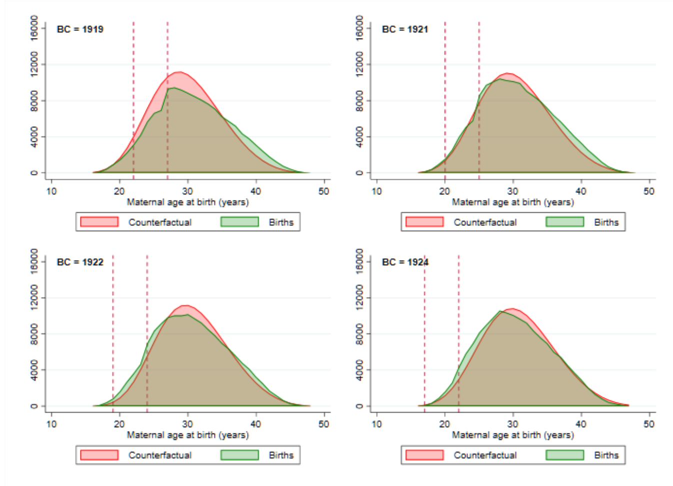
[19] The results for the birth cohorts that are not shown look similar to the ones that are shown and are available on request.

Figure 7: Counterfactual distributions

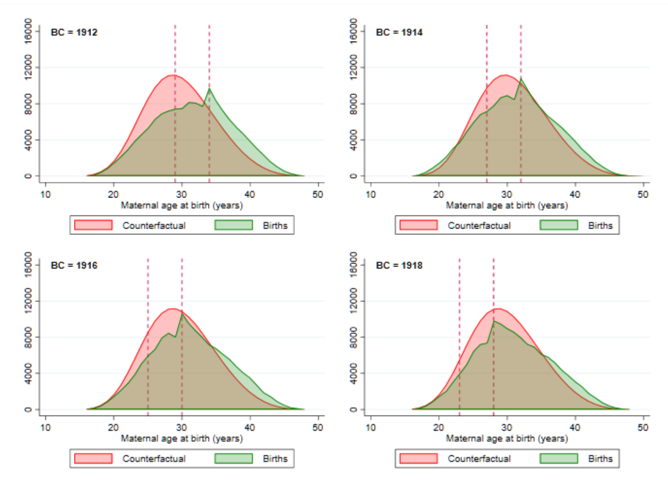
(a) Group 1: Cohorts 1925–1930



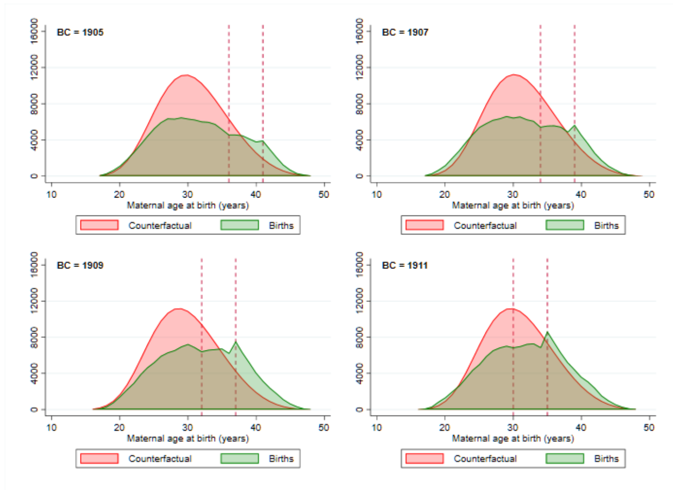
(b) Group 2: Cohorts 1919–1924



(c) Group 3: Cohorts 1912–1918



(d) Group 4: Cohorts 1905–1911



Notes: The figures show the results of the bunching counterfactual distribution estimation. The procedure to create these densities is discussed in Section 4.2. The four panels represent the four broad age groups that were introduced in Figure 3 that faced different incentives in their fertility response to the liberation. The counterfactual distributions of births are plotted in red and the actual distribution of births are plotted in green. For every group, I show four maternal birth cohorts, but figures for the remaining birth cohorts that are not plotted show similar patterns.

The latter effects are even larger for the oldest cohorts in my sample: birth cohorts 1905–1911 who were aged 34–40 in 1945. These women were affected by the war after peak fertility ages, and the resulting distributions of counterfactual and actual births are shown in Panel (d). Again, as for the mothers in the third group, the poor circumstances during the interwar period already had a large effect on fertility as the number of births was much lower than predicted at early ages. These effects are only strengthened by the influence of the war. The peak that occurs quite late in the lifecycle for these women does not make up for the missed births they experienced earlier in life.

5.2 Decomposing "missing births"

These results are summarized in Table 2, where "missed births" — the difference between the counterfactual number of births and the actual number of births — are plotted for different time periods. When looking at Figure 7 one can think of the "missed births" measure as the difference between the counterfactual distribution (in red) and the actual distribution of births (in green). The "missed births" measure is positive for years in which the red density exceeds the green density (hence, fewer births than predicted by the counterfactual), and negative in years where the green density exceeds the red density (more births than predicted by the counterfactual).

Using the data underlying Figure 7, I calculate the number of "missed births" over different time periods for different groups of cohorts. I focus on birth cohort groups 1905–1911, 1912–1918, and 1919–1924, because their fertility was contemporaneously affected by WWII. A distinction is made between conceptions during the Great Depression and before the war (births during 1931–1940), conceptions during the war (births during 1941–1945), and conceptions during the post-war time period (births during 1946–1960). To provide a magnitude for the number of "missed births" I divide this number by the total number of actual births for the same birth cohorts/time period, this is shown in the even columns.

Table 2: Number of "missed births" per cohort and period

	Total missed births	%	Years 1931–1940	%	Years 1941–1945	%	Years 1946–1960	%
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
All cohorts	255,471.8	9.4	256,051.9	44.7	136,247.3	24.1	-120,545.6	-8.4
1905–1911	186,549.6	21.8	214,074.3	54.1	46,115.6	22.4	-73,053.3	-38.2
1912–1918	82,361.6	8.6	42,177.8	25.4	90,895.2	34.9	-47,799.7	-9.1
1919–1924	-13,439.5	-1.5	-200.2	-1.8	-763.6	-0.76	307.4	0.0

Notes: Table shows the number of "missed births" for cohort groups over different time periods. The number of "missed births" is calculated by subtracting the number of actual births from the number of births in the counterfactual. The measure is positive when the counterfactual distributions exceed the actual distribution of births (and the other way around). The columns with percentages divide the number of "missed births" by the total number of actual births for the same group and period.

Table 2 shows that for birth cohorts 1905–1924 I estimate that 255,472 fewer births are born compared to the counterfactual. This number amounts up to 9.4% of all births that actually occurred, which implies that there would have been 9.4% more births in the absence of the events surrounding WWII. This number masks heterogeneity across birth cohorts depending on their exposure to the interwar period and WWII. The younger cohorts of 1919–1924 had 1.5% more births compared to the counterfactual, which is consistent with Figure 7 that shows that, if anything, these cohorts started childbearing at earlier ages. For birth cohorts 1912–1918 the number of births would have been 8.6% higher without WWII. The number is largest for birth cohorts 1905–1911 who were exposed to WWII and the Great Depression (aged 19-25 at the start of the Great Depression in 1930) during their fertile lives. For these cohorts, the number of births would have been 21.8% higher in the absence of these events.

This exercise also allows for distinguishing between the impact of different periods, which is particularly interesting because the Netherlands was first exposed to poor economic conditions during the Great Depression (1930–1939) after which WWII started and the Netherlands was occupied by Nazi Germany (1940–1945). For birth cohorts 1905–1911, who were exposed to poor economic conditions and war throughout almost all of their childbearing ages, the number of births was 54.1% lower during 1931–1940, and 22.4% lower during WWII. The post-war years (1946–1960) show a catching up of fertility for these birth cohorts, but it is not compensating for the "missed births" in earlier years. For birth cohorts 1912–1918 — who were ages 22–28 in 1940 — the number of births would have been 25.4% higher during 1931–1940, and 34.9% higher during WWII. Also for these groups there is a catching up after the war (1946–1960), but again this number does not compensate for "missed births" in earlier years. Birth cohorts 1919–1924 experience a lower impact of WWII on their fertility, which is not surprising given their ages at the start of the war (i.e., ages 16–21 in 1940).

Overall, these numbers clearly show that the recovery of fertility after WWII did not make up for missed births in earlier years. The effect of the poor economic conditions preceding WWII was largest for mothers in birth cohorts 1905–1911, who were aged 25–31 at the height of the Great Depression in 1936. The impact of the war is largest for mothers in birth cohorts 1912–1918 who were aged 27–33 at the time of the liberation. Hence, fertility of women in prime fertile ages was hit hardest by the poor circumstances of the depression and war.

5.3 Sensitivity of results

The key assumption of the identification strategy is that the polynomial of the manipulated distributions can be estimated from the unmanipulated distributions. In Section 4 I outline the choices made to estimate a valid counterfactual that can be used to benchmark the actual number of births that occurred over a cohort's lifecycle. In this section, I check the sensitivity of these results by relaxing three identification choices: (1) not making a distinction between areas with and without air raids, (2) excluding the youngest "control" cohorts, and (3) including famine-affected areas.

First, I estimate counterfactual distributions while not making a distinction between areas with or without air raids. Note that without making this distinction, the estimation of the common polynomial is only based on the shape of the distributions for older and younger cohorts. The counterfactual distributions compared to the actual distributions of birth resulting from this exercise are shown in Figure A5, and the aggregated numbers by birth cohort are shown in the second row of Table 3. The estimated counterfactuals look similar to the baseline counterfactuals, and the aggregated numbers of "missed births" are also comparable to the baseline estimates. If anything, the estimated number of "missed births" is somewhat lower when not making the distinction between areas with and without air raids.

Table 3: Sensitivity checks: number of "missed births" per cohort and period

	Total missed births (1)	% (2)	Cohorts 1905–1911 (3)	% (4)	Cohorts 1912–1918 (5)	% % (6)	Cohorts 1919–1924 (7)	% % (8)
Baseline	255,471.8	9.4	186,549.6	21.8	82,361.6	8.6	-13,439.5	-1.5
No treatment/control	238,608.3	8.8	180,584.6	21.1	76,320.4	7.9	-18,296.9	-2.0
Excl. 1924–1930 cohorts	132,958.0	4.9	143,701.3	16.8	39,523.4	4.1	-50,266.7	-5.5
Incl. famine-affected areas	321,002.0	9.1	244,359.6	22.1	100,080.3	8.0	-23,437.8	-2.0

Notes: Table shows the number of "missed births" for cohort groups over different time periods. The number of "missed births" is calculated by subtracting the number of actual births from the number of births in the counterfactual. The measure is positive when the counterfactual distributions exceed the actual distribution of births (and the other way around). The columns with percentages divide the number of "missed births" by the total number of actual births for the same group and period.

Second, I re-estimate the models while excluding the 1925–1930 birth cohorts of "control" mothers. I do this to check how much the estimated counterfactual distributions rely on the control cohorts of women who too young to be contemporaneously affected by the depression and the war. Figure A6 shows that, as expected, the number of "missed births" for the different cohorts is lower when the distributions of births for these unaffected women are not used, which is not surprising given that only "manipulated" distributions are used to estimate counterfactual distributions of births. However, the patterns in counterfactual distributions are very similar to the baseline specification. The third row of Table 3 shows that, when excluding the youngest cohorts as a control group, the total number of "missed births" would be lower compared to the number estimated using the baseline specification (4.9% versus 9.4%).

Third, I check how the results change when including the seven large cities that were affected by the famine. These cities are excluded because fertility may be affected beyond choice, given that frictions like food shortages and the absence of men may have affected women's ability to conceive in the last year before the end of the war. Figure A7 shows the results of leaving in births that occurred in these cities. The figure shows that the observed patterns in counterfactual distributions are similar to the baseline specification, apart from differences in the levels of the number of births. Hence, even though in these cases "missed births" could have been caused by other frictions and not just choice, the patterns in "missed births" are similar. Moreover, Table 3 confirms that results are very similar when leaving in famine-affected areas.

These sensitivity checks show that regardless of the identification choices made, the fertility rise after the war did not make up for births in earlier years. My preferred specification is very close to the results when excluding an extra control group (areas with and without air raids), and when including famine-affected areas. The numbers are less stark when not including the untreated cohorts of 1925–1930, those who were too young for their fertility to be contemporaneously affected by the war, although the number of "missed births" is still substantial.

6 Becker versus Easterlin

The previous section showed that the post-war birth peak did not make up for the "missed births" that did not occur earlier in time. In this section, I test whether a higher exposure to the Great Depression led to differential fertility responses. This exercise sheds light on the two schools of thought that have dominated the discussion on the factors underlying the U.S. baby boom, and examines which of the two views is responsible for the fertility response after the liberation in the Netherlands.

The existing literature studying explanations for the U.S. baby boom is based on two main strands of literature (Bailey et al., 2014). First, Gary Becker's neoclassical theory suggests that prices and absolute incomes matter for fertility decisions (Becker, 1960, 1965). Becker's theory argues that a negative association between childbearing and income can be explained by differences in the opportunity costs of childbearing (i.e., higher wages lead to a lower demand for children). Several papers have examined how changes in the "price" of having children could have led to the U.S. baby boom. These shocks to the opportunity costs of childbearing vary from technological progress in the household (Greenwood et al., 2005), changed labor market opportunities for women (Doepke et al., 2015; Brodeur and Kattan, 2022), and medical progress leading to declines in maternal morality (Albanesi and Olivetti, 2014, 2016).

On the other hand, the "relative income hypothesis" by Richard Easterlin (Easterlin, 1966, 1971, 1987) argues that a cohort's earnings potential relative to their "material aspirations" in childhood is a key determinant of fertility. To illustrate, "material aspirations" of children who grew up during the Great Depression were low, and circumstances became better when these cohorts entered adulthood after WWII. When circumstances improved compared to those experienced in childhood, these cohorts had relatively more children, which is mentioned as an explanation for the baby boom in the United States.

6.1 Empirical strategy and data

To test for these two hypotheses, I distinguish between areas that were more or less affected by the Great Depression. We have seen that for the oldest cohorts, fertility was already lower before the start of the war because of poor economic conditions in the interwar period. In this section, the effects of poor economic conditions are separated by area and cohort.

Regional unemployment rates Data on the unemployment rate for several municipalities between 1931–1939 come from pages 335–338 of Kloosterman (1985). Data on unemployment rates are only reported in all years for 37 municipalities with more than 20,000 inhabitants (excluding the seven large cities in the west). I focus the analysis on the 36 municipalities that also experienced at least one air raid between 1940–1945. These remaining municipalities cover 1,160,274 births, which is about 32% of the sample used in the baseline analysis.

To distinguish between areas that suffered from low and high unemployment, I calculate the mean unemployment rate over the 1932–1938 period, which is the period in which the Great Depression affected the Netherlands. Mean unemployment in the sample over this period is 26.4%, with a median of 27%. I then split up the sample by whether municipalities experienced higher than median or lower than median unemployment over the 1932–1938 period.

Table A4 shows characteristics are very similar in areas with above or below median unemployment between 1932–1938. These areas have similar rates of marriage, childlessness, and live births as observed in the 1971 census. At the municipality-level, these areas have a similar religious composition, similar proportion of individuals who are born in the Netherlands, and experienced a very similar number of air raids during WWII. If anything, areas with below median unemployment have larger populations compared to those with above median unemployment, but given that this sample only includes urban areas (i.e., with populations of 20,000 or more) this should not have an effect on fertility.

Estimating counterfactual distributions of births I estimate Equation (1) but instead of making a distinction between areas with or without air raids, I substitute this with an indicator variable that is equal to one if an area experienced above median unemployment between 1932–1938. Note that all municipalities that remain in the final sample experienced at least one air raid, so that this extra control group has become redundant. Using these results, I then add up the cohort-specific number of actual and counterfactual births in each year and at the area-level (i.e., areas with above or below median unemployment). I calculate the number of "missing births" by subtracting the yearly number of actual births from the number of counterfactual births. Plotting these allows me to observe how the measure of "missing births" evolves over time and in particular across areas with above and below median unemployment during the Great Depression.

6.2 Results

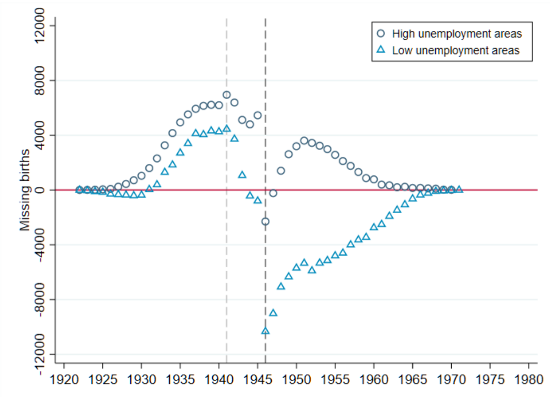
The results of estimating the bunching methodology when making a distinction between areas with high and low unemployment and plotting the number of "missed births" over time are shown in Figure 8. The number of "missed births" is positive when there are fewer births than predicted, and negative when there are more births that predicted.

Panel (a) shows that for all birth cohorts (i.e., 1905–1924) the number of "missed births" is similar in high- and low-unemployment areas in the 1920s, but also that this measure starts to diverge during the early 1930s, which is when the Great Depression hit. The number of "missed births" is higher in high unemployment areas starting the early 1930s, and this pattern remains until the end of WWII. The catch-up of fertility after the war (a negative shock in "missed births") appears similar in both areas, although the number of "missed births" is much higher in areas characterized by high unemployment between 1932–1938. This figure suggests that the Great Depression had an instantaneous and long-lasting impact on fertility for those exposed to more-severe economic conditions in the interwar period. I calculate that high unemployment areas would have had 30.9% more births in the absence of the Great Depression and war.

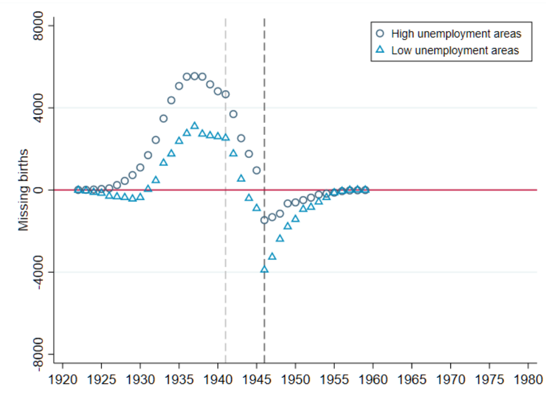
For women in the oldest birth cohorts 1905–1911, and who were closest to peak fertility ages at the start of the Great Depression in 1932 (ages 21–27), panel (b) shows that women in high unemployment areas have more missed births from the beginning of the 1930s, which persists until the end of the war. For women in birth cohorts 1912–1918, the picture looks different as can be seen in panel (c). These women were ages 14–20 in 1932, and hence exposed to more-severe economic conditions before they started family formation. Interestingly, the number of "missed births" is similar for these women across areas before the start of the war, but starts to deviate during the war. After the start of WWII, areas that were characterized by high unemployment rates before the war, have higher levels of "missed births" compared to areas that experienced lower unemployment rates.

Figure 8: "Missed births" over time, areas with above/below unemployment between 1932–1938

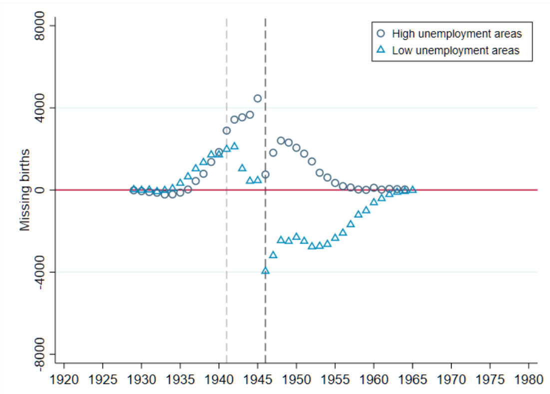
(a) Birth cohorts 1905–1924



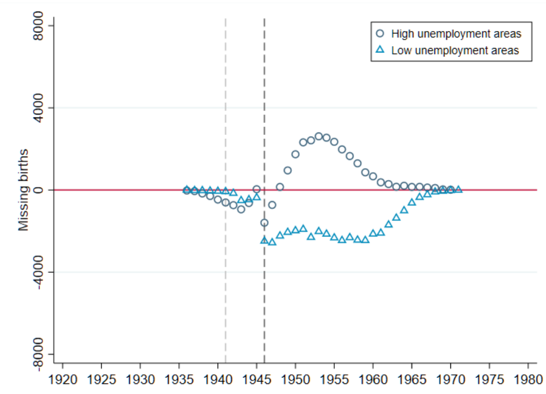
(b) Birth cohorts 1905–1911



(c) Birth cohorts 1912–1918



(d) Birth cohorts 1919–1924



Notes: The figures show the results of the bunching counterfactual distribution estimation. This analysis is based on 36 municipalities that all experienced at least one air raid between 1940–1945. A distinction is made between municipalities that experienced above and below median unemployment between 1932–1938. For every group of birth cohorts the number of "missed births" is plotted over time. This measure is positive if the counterfactual number of births is higher than the actual number of births, and negative when the actual number of births exceeds the counterfactual number of births.

Panel (d) shows that the picture for women in the youngest birth cohorts 1919–1924 looks even more different. These women were ages 9–13 in 1932 when the Great Depression started, and were exposed to poor economic conditions throughout childhood. Fertility looks very similar before and during the war, most likely because the majority of these women were too young to start family formation. After the war, a differential fertility response can be observed as there are more "missed births" for women exposed to more-severe economic conditions in childhood during the interwar period.

Easterlin predicts that if children are exposed to adverse economic conditions in childhood, childbearing increases when economic conditions improve relative to circumstances in childhood. Figure 8 shows instantaneous effects of the Great Depression on fertility for women who were in prime childbearing ages while being exposed to these conditions. For cohorts that were exposed earlier in life (adolescence for cohorts 1912–1918 and childhood for cohorts 1919–1924) fertility remains lower throughout later periods, even after economic prospects improve after the end of the war. Hence, I find no evidence that the Easterlin hypothesis can explain the birth peak after

WWII in the Netherlands.

On the other hand, Becker's theory about relative prices and absolute incomes affecting the opportunity cost of childbearing is consistent with the patterns observed in Figure 8. It was difficult to start family in the studied time period when economic conditions were bad, as women started childbearing after getting married, and marriage required financial means. Hence the Great Depression made it difficult to start family formation, which had an instantaneous effect on fertility for older cohorts 1905–1911, and had long-lasting effects on fertility for younger cohorts 1912–1924.

7 Conclusion

This paper shows that the fertility rise that occurred after the end of WWII in the Netherlands did not make up for the "missed" births that did not take place during the war and the interwar period. Hence, experiencing these adverse economic or living circumstances during fertile ages did not only lead to fertility delay (a *tempo* effect) but also to lower completed fertility as some of these postponed births never occurred (a *quantum* effect). I estimate that 255,472 fewer children were born to mothers in birth cohorts 1905–1924, which amounts to 9.4% of the total actual number of births for these cohorts. The magnitude of "missed" births is even higher for cohorts exposed to the interwar period and war in prime fertile ages. I show that the mechanism behind these fertility effects is consistent with Becker's neoclassical theory, but not with Easterlin's relative income hypothesis.

A new application of the bunching method is introduced in this paper that exploits similarities across multiple distributions of affected and unaffected cohorts to estimate counterfactual distributions of births. Using the bunching method to study fertility has many advantages. It allows one to study changes in fertility over the full lifecycles of cohorts, instead of differences at pre-specified ages or points in time. This is especially relevant when a series of adverse conditions affect women at different ages, which makes it difficult to disentangle the fertility impact of changes in living conditions. Also, the bunching method can be applied without making assumptions on the age or time of exposure to these adverse conditions. Finally, the estimation of counterfactual densities of births allows for an interpretation of whether external shocks lead to "extra" or "missed" births. The bunching method could be an opportunity to study the consequences of other shocks to fertility or to basically any other outcome that exhibits an age profile. This is particularly relevant given the rise in the availability of administrative datasets that allow for disaggregations of data in several dimensions.

These results call for a different interpretation of the baby boom as can be observed in period-fertility rates in the Netherlands. Given that measures of cohort-fertility (i.e., mean live births for married women aged 41 and older) did not show cohort-specific responses to the war, these results also stress the importance of examining the full lifecycles of cohorts when examining fertility questions. This is relevant from a demographic perspective, but also because age-structures of populations can have long-lasting impacts on public welfare systems (e.g., pensions, healthcare), economic growth, and the environment. The results of this paper also underscore that issues experienced by many developed countries due to ageing populations may have been larger in the absence of the events of the 1930s–1950s. This emphasizes the importance of taking into account demographic change when designing (long-term) public policies.

A number of questions remain. It is unclear how these results extend to different institutional settings, such as the baby bust and boom in the United States and in other countries, as well as different times (e.g., the Covid-19 pandemic). Moreover, this paper does not address how child outcomes are affected by the war. Fertility fluctuations resulting from the war could affect child outcomes through the size of the birth cohort, family size, and who chooses (not) to become a parent. These questions are left for future research.

References

Agadjanian, V. and Prata, N. (2002). War, peace, and fertility in Angola. *Demography*, 39(2):215–231.

Albanesi, S. and Olivetti, C. (2014). Maternal health and the baby boom. *Quantitative Economics*, 5(2):225–269.

Albanesi, S. and Olivetti, C. (2016). Gender roles and medical progress. *Journal of Political Economy*, 124(3):650–695.

Bailey, M. J., Currie, J., and Schwandt, H. (2023). The covid-19 baby bump in the united states. *Proceedings of the National Academy of Sciences*, 120(34):e2222075120.

Bailey, M. J., Guldi, M., and Hershbein, B. J. (2014). Is there a case for a "second demographic transition"? Three distinctive features of the post-1960 U.S. fertility decline. In *Human Capital and History: The American Record*, page 273. University of Chicago Press, Chicago, IL. Boustan, Frydman, and Margo (eds.).

Becker, G. S. (1960). An economic analysis of fertility. In *Demographic and Economic Change in Developed Countries*, pages 209–240. Columbia University Press.

Becker, G. S. (1965). A theory of the allocation of time. *The Economic Journal*, 75(299):493–517.

Blok, E. (1988). De maatschappelijke positie van de vrouw in nederland. *M. Grever, C. Wijers (ed.) Vrouwen in de twintigste eeuw. De positie van de vrouw in Nederland en de Verenigde Staten van Amerika 1929–1965. IJsselstein: Vereniging van Docenten in Geschiedenis en Staatsinrichting in Nederland (VGN)*, pages 15–24.

Brodeur, A. and Kattan, L. (2022). World War II, the baby boom, and employment: County-level evidence. *Journal of Labor Economics*, 40(2):437–471.

Chabé-Ferret, B. and Gobbi, P. (2018). Economic uncertainty and fertility cycles: The case of the post-WWII baby boom. *CEPR Discussion Paper No. DP13374*.

Chetty, R., Friedman, J. N., Olsen, T., and Pistaferri, L. (2011). Adjustment costs, firm responses, and micro vs. macro labor supply elasticities: Evidence from Danish tax records. *The Quarterly Journal of Economics*, 126(2):749–804.

Chevalier, A. and Marie, O. (2017). Economic uncertainty, parental selection, and children's educational outcomes. *Journal of Political Economy*, 125(2):393–430.

Currie, J., Duque, V., and Garfinkel, I. (2015). The Great Recession and mothers' health. *The Economic Journal*, 125(588):311–346.

Currie, J. and Schwandt, H. (2014). Short-and long-term effects of unemployment on fertility. *Proceedings of the National Academy of Sciences*, 111(41):14734–14739.

De Graaf, A. (2008). Vruchtbaarheid in de twintigste eeuw. *Bevolkingstrends*, 1(2008):14–22.

De Jong, A. and De Graaf, A. (1999). Marriage: From cornerstone to outdated institution? *Vital Events. Past, Present and Future of the Dutch Population. Statistics Netherlands, Voorburg*, pages 37–50.

De Vries, R. (2004). *A Pleasing Birth: Midwives and Maternity Care*. Philadelphia: Temple University Press.

De Zwarte, I. (2019). *De Hongerwinter*. Amsterdam: Prometheus.

Dehejia, R. and Lleras-Muney, A. (2004). Booms, busts, and babies' health. *The Quarterly Journal of Economics*, 119(3):1091–1130.

Del Bono, E., Weber, A., and Winter-Ebmer, R. (2012). Clash of career and family: Fertility decisions after job displacement. *Journal of the European Economic Association*, 10(4):659–683.

Dettling, L. J. and Kearney, M. S. (2023). The cyclicality of births and babies' health, revisited: Evidence from unemployment insurance. *National Bureau of Economic Research Working Paper No. w30937*.

Doepke, M., Hazan, M., and Maoz, Y. D. (2015). The baby boom and World War II: A macroeconomic analysis. *The Review of Economic Studies*, 82(3):1031–1073.

Easterlin, R. A. (1966). On the relation of economic factors to recent and projected fertility changes. *Demography*, 3(1):131–153.

Easterlin, R. A. (1971). Does human fertility adjust to the environment? *The American Economic Review*, 61(2):399–407.

Easterlin, R. A. (1987). *Birth and Fortune: The Impact of Numbers on Personal Welfare*. University of Chicago Press.

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

Greenwood, J., Seshadri, A., and Vandenbroucke, G. (2005). The baby boom and baby bust. *American Economic Review*, 95(1):183–207.

Heuveline, P. and Poch, B. (2007). The phoenix population: Demographic crisis and rebound in Cambodia. *Demography*, 44(2):405–426.

Huttunen, K. and Kellokumpu, J. (2016). The effect of job displacement on couples' fertility decisions. *Journal of Labor Economics*, 34(2):403–442.

Jones, L. E. and Schoonbroodt, A. (2016). Baby busts and baby booms: The fertility response to shocks in dynastic models. *Review of Economic Dynamics*, 22:157–178.

Klemann, H. A. (2002). *Nederland 1938-1948: Economie en Samenleving in Jaren van Oorlog en Bezetting*. Boom Koninklijke Uitgevers.

Kleven, H. J. and Waseem, M. (2013). Using notches to uncover optimization frictions and structural elasticities: Theory and evidence from Pakistan. *The Quarterly Journal of Economics*, 128(2):669–723.

Kloosterman, R. (1985). *Werkloosheid in Nederland, 1920-1939.* Hes Uitgevers, Utrecht.

Lee, R. D. (1997). Population dynamics: Equilibrium, disequilibrium, and consequences of fluctuations. *Handbook of Population and Family Economics*, 1:1063–1115.

Lindstrom, D. P. and Berhanu, B. (1999). The impact of war, famine, and economic decline on marital fertility in Ethiopia. *Demography*, 36(2):247–261.

Marie, O. and Zwiers, E. S. (2023). Religious barriers to birth control access. *IZA Discussion Paper No. 16051.*

Nobles, J., Frankenberg, E., and Thomas, D. (2015). The effects of mortality on fertility: Population dynamics after a natural disaster. *Demography*, 52(1):15–38.

Persson, P. (2020). Social insurance and the marriage market. *Journal of Political Economy*, 128(1):252–300.

Saez, E. (2010). Do taxpayers bunch at kink points? *American Economic Journal: Economic Policy*, 2(3):180–212.

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

Sijes, B. A. (1990). *De Arbeidsinzet: De Gedwongen Arbeid van Nederlanders in Duitsland, 1940-1945.* Number Nr. 11. Netherlands State Institute for War Documentation, Amsterdam.

Stein, Z., Susser, M., Saenger, G., and Marolla, F. (1975). *Famine and Human Development: The Dutch Hunger Winter of 1944–1945.* Oxford University Press.

Te Slaa, R. (2016). *1941: Het Masker Valt.* Spectrum/NIOD.

Trienekens, G. M. (1985). *Tussen Ons Volk en de Honger: De Voedselvoorziening, 1940-1945.* Doctoral Thesis, Landbouwhogeschool te Wageningen.

Van Bavel, J. and Reher, D. S. (2013). The baby boom and its causes: What we know and what we need to know. *Population and Development Review*, 39(2):257–288.

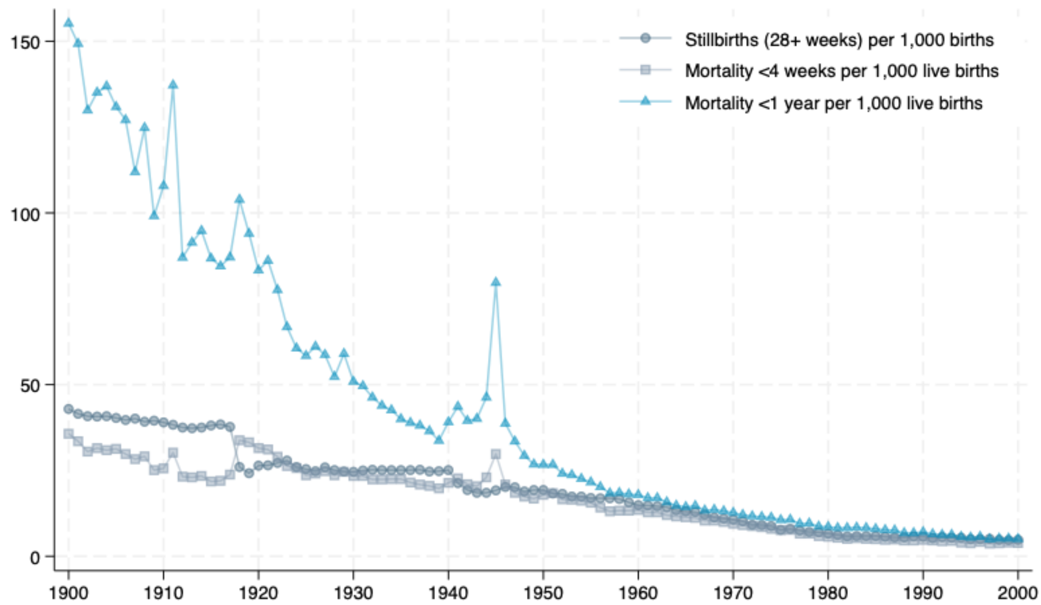
Van Blankenstein, E. (2006). *Defensie-en Oorlogsschade in Kaart Gebracht (1939-1945).* Rijksdienst voor de Monumentenzorg.

Appendix

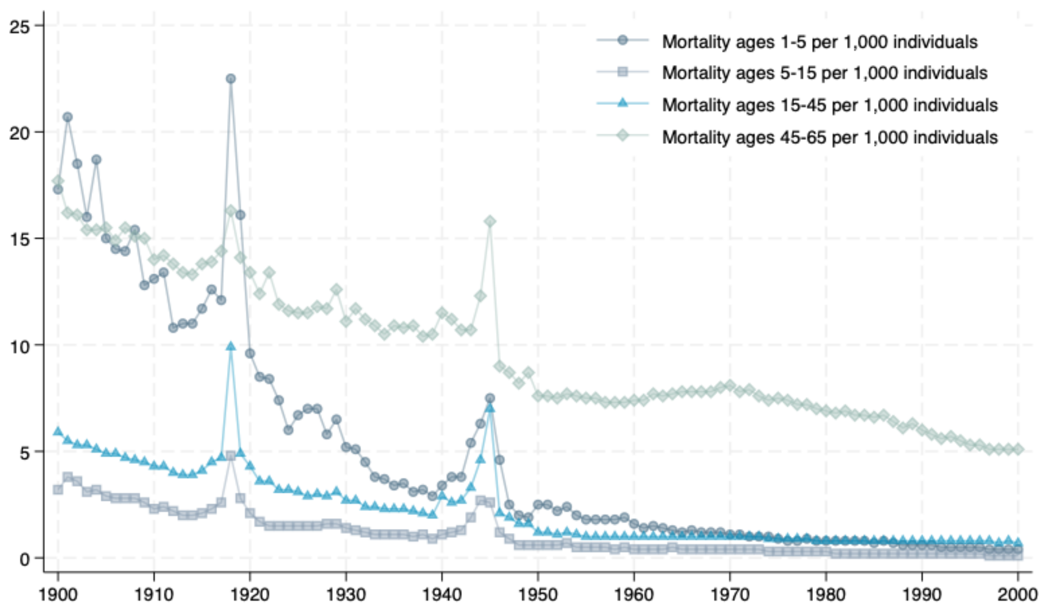
A: Supplemental figures and tables

Figure A1: Mortality rate by age group, the Netherlands, 1900–2000

(a) Infant mortality rate (stillbirths, within four weeks and one year after birth)



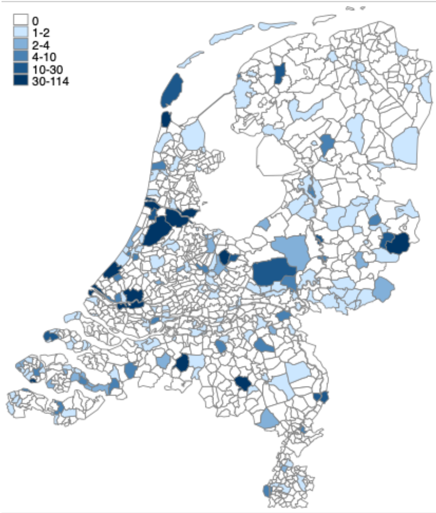
(b) Mortality rate for ages 1–65, by group



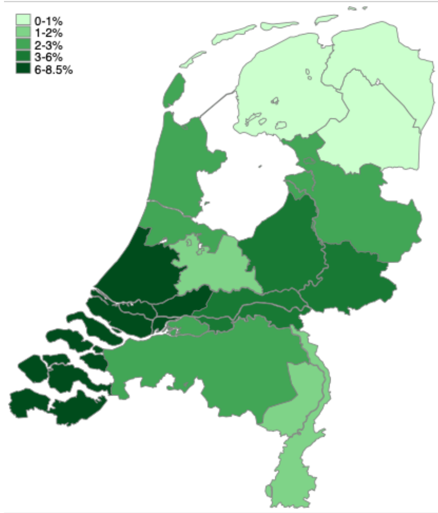
Notes: Panel (a) shows the the number of stillbirths after a pregnancy duration of 28 weeks or more relative to 1,000 births, and the number of deaths within four weeks and one year after birth per 1,000 live births. Panel (b) shows the number of deaths by age-group per 1,000 population in that same age-group. Source: statline.cbs.nl; Bevolking, huishoudens en bevolkingsontwikkeling; vanaf 1899

Figure A2: Air raids and damages by area, the Netherlands, 1940–1945

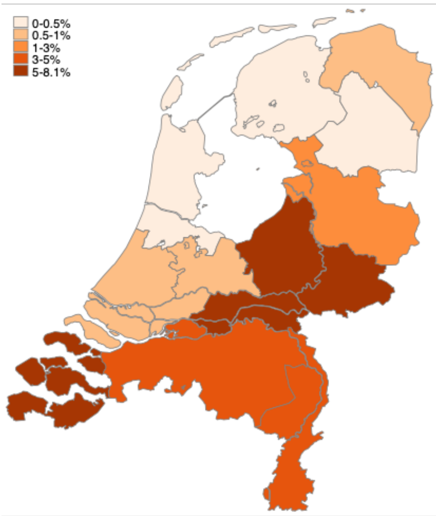
(a) Number of air raids across Dutch municipalities, 1940–1945



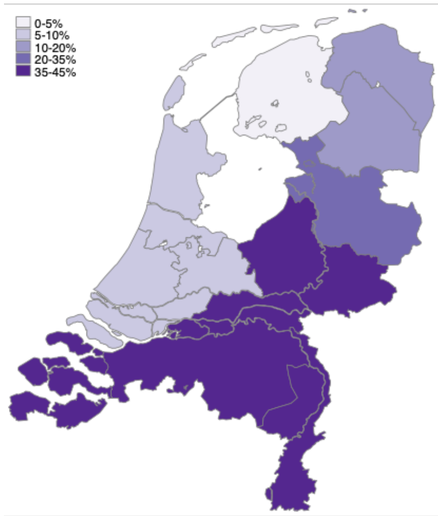
(b) Percentage of destroyed homes in 1945 by province



(c) Percentage of heavily damaged homes in 1945 by province

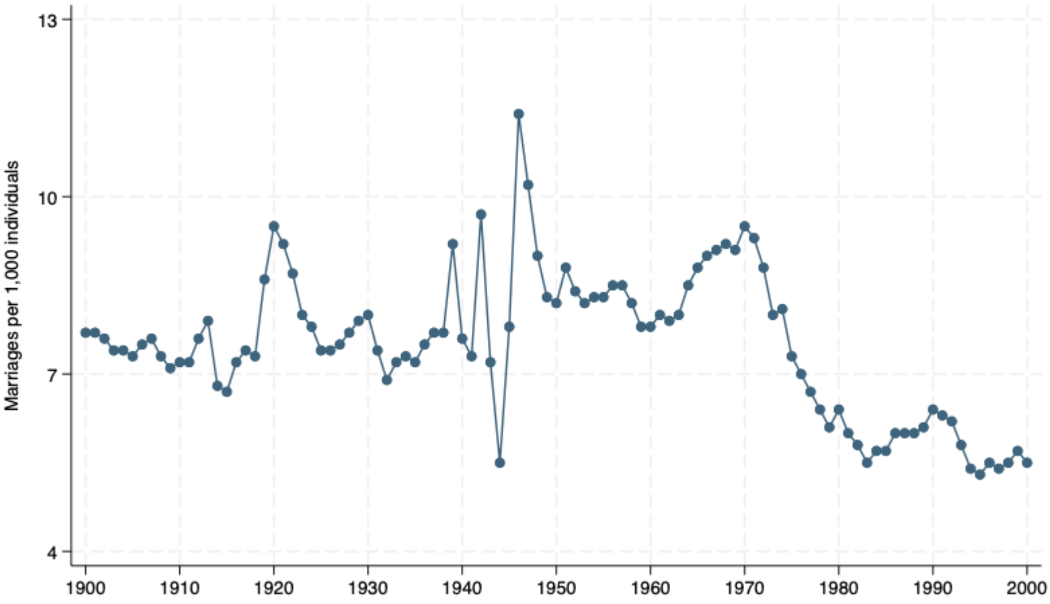


(d) Percentage of lightly damaged homes in 1945 by province



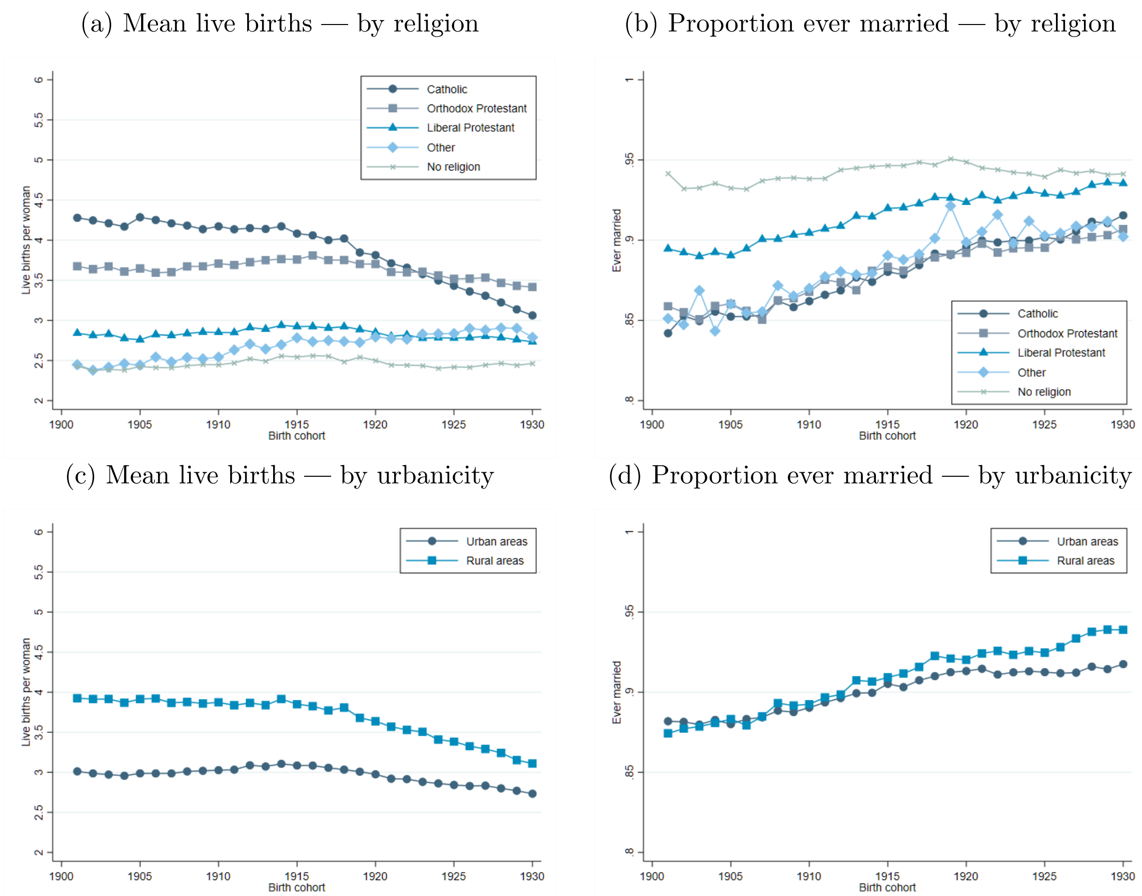
Notes: Panel (a) shows the number of air raids by municipality in the Netherlands between 1940–1945 and is constructed using author's calculations based on Van Blankenstein (2006), pages 219–306. The data in panels (b)–(d) come from Van Blankenstein (2006), Table 1.3, page 15. The figures show the percentages of homes per province that were either destroyed (b), heavily damaged (c) or lightly damaged (d) by May 1945.

Figure A3: Marriage rate, the Netherlands, 1900–2000



Notes: The figure shows the number of marriages per 1,000 individuals. Source: statline.cbs.nl; Bevolking, huishoudens en bevolkingsontwikkeling; vanaf 1899.

Figure A4: Fertility and marriage by cohort and group (religion, urban/rural)



(a) Mean live births — by religion

(b) Proportion ever married — by religion

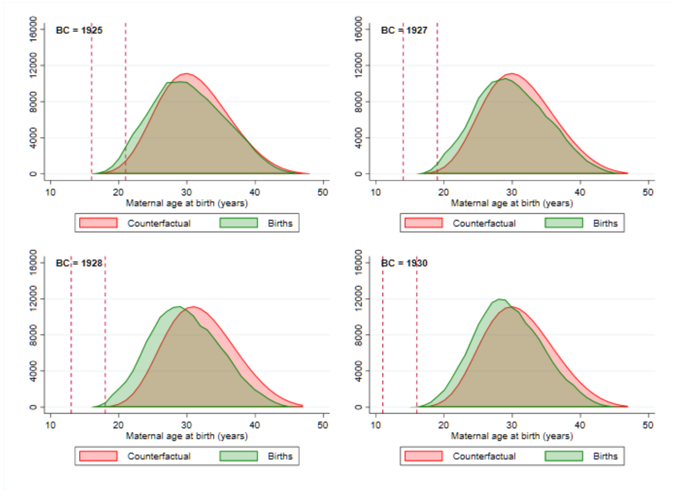
(c) Mean live births — by urbanicity

(d) Proportion ever married — by urbanicity

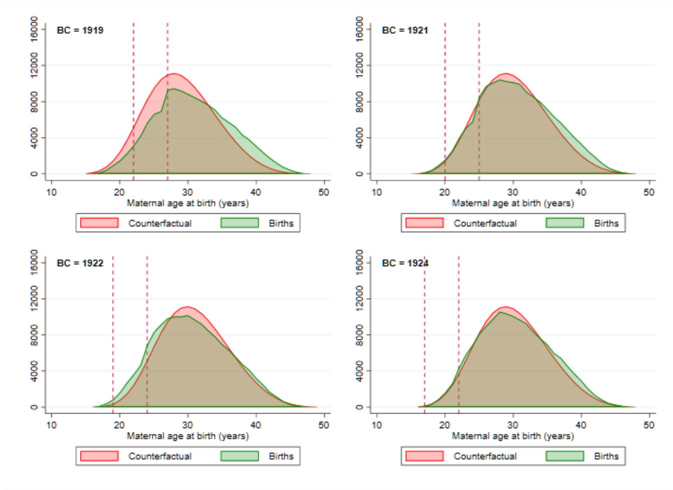
Notes: This figure is compiled using data from the 1971 census. Panels (a) and (b) show the mean number of live births and the proportion of women who were ever married by birth while making a distinction between five religious groups. Panels (c) and (d) show the mean number of live births and the proportion of women who were ever married by birth while making a distinction between urban (municipalities with more than 20,000 inhabitants) and rural areas.

Figure A5: Sensitivity check: Not making an additional distinction between areas with and without air raid between 1940–1945

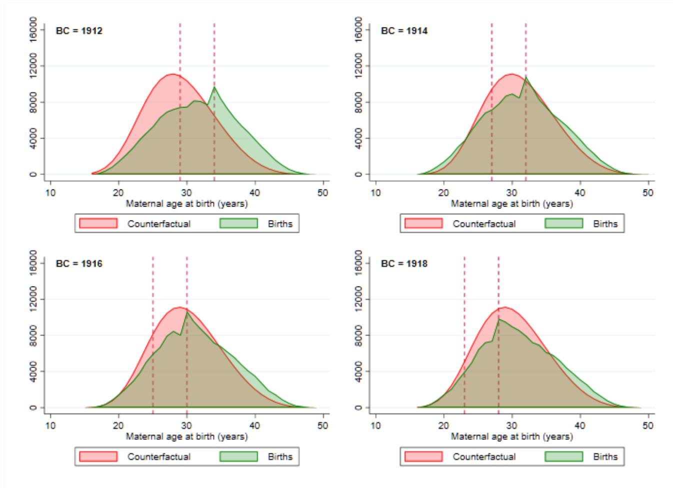
(a) Group 1: 1925–1930 cohorts



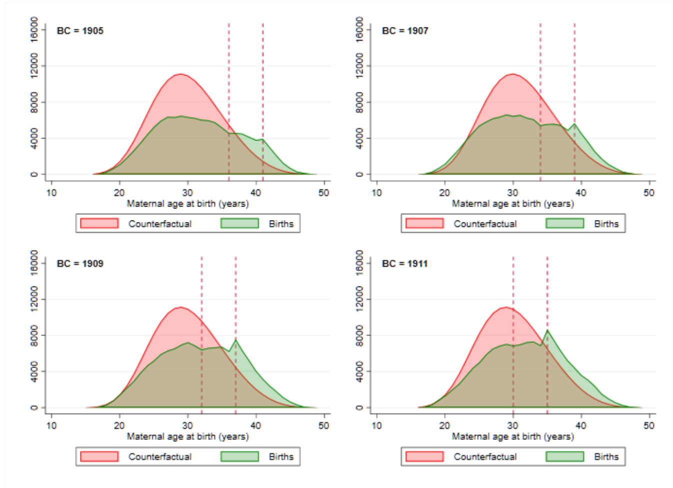
(b) Group 2: 1919–1924 cohorts



(c) Group 3: 1912–1918 cohorts



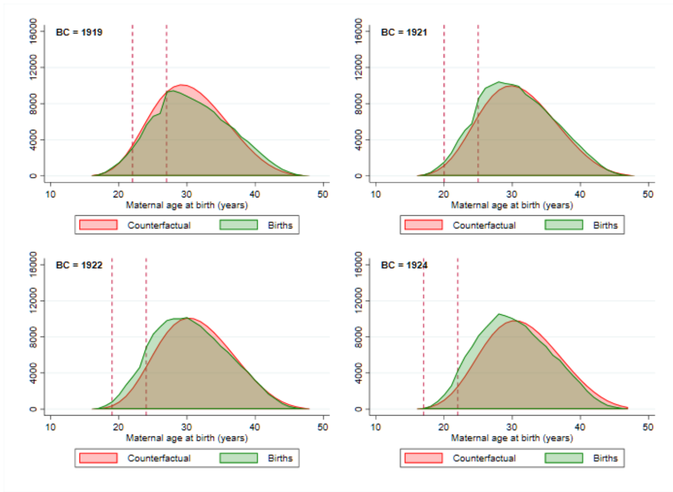
(d) Group 4: 1905–1911 cohorts



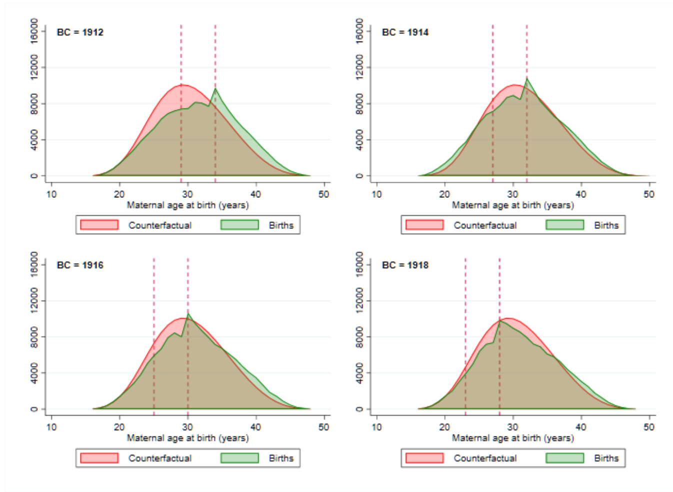
Notes: The figures show the results of the bunching counterfactual distribution estimation. For this robustness check, I do not make an extra distinction between areas with and without air raid (as an additional control group). The procedure to create these densities is discussed in Section 4.2. The counterfactual distributions of births are plotted in red, and the actual distribution of births is plotted in green.

Figure A6: Sensitivity check: Counterfactual distributions when leaving out control cohorts of 1925–1930

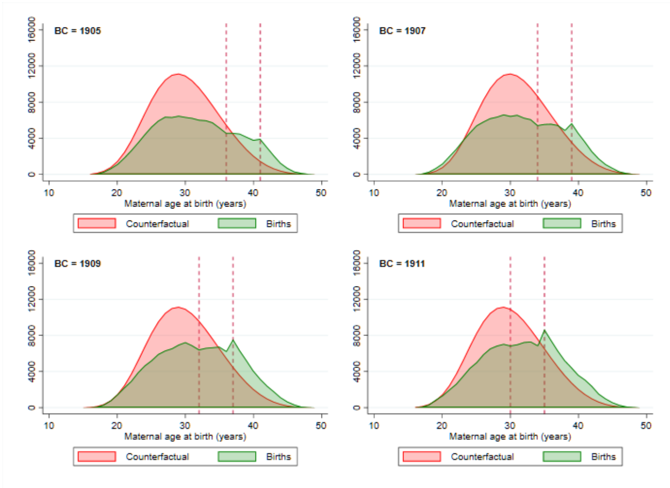
(a) Group 2: 1919–1924 cohorts



(b) Group 3: 1912–1918 cohorts



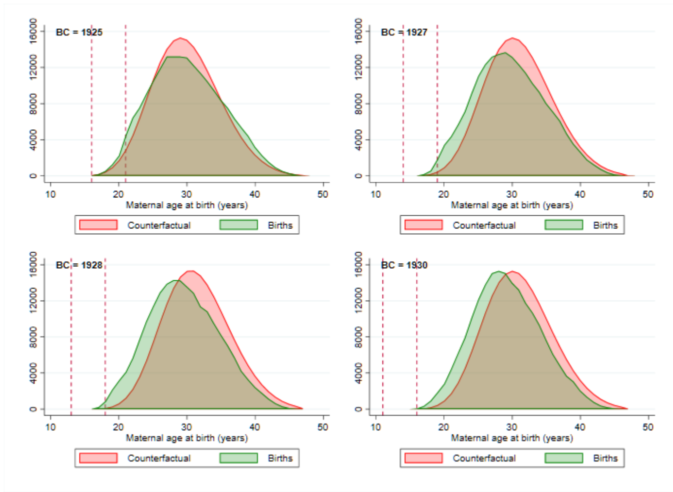
(c) Group 4: 1905–1911 cohorts



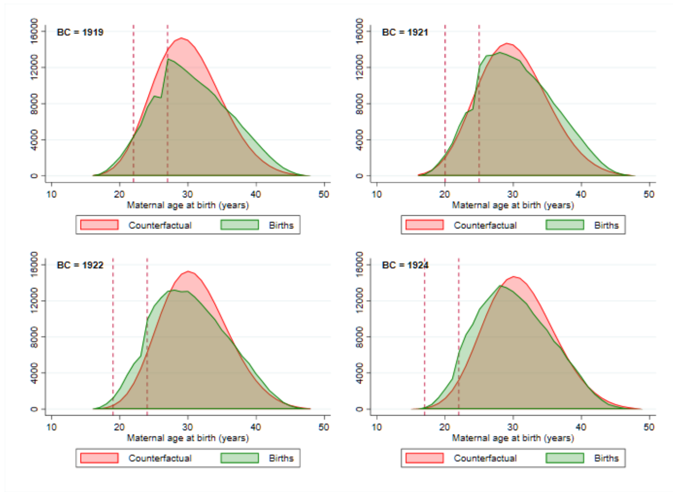
Notes: The figures show the results of the bunching counterfactual distribution estimation. For this robustness check, births to mothers in the 1925–1930 birth cohorts are excluded. The procedure to create these densities is discussed in Section 4.2. The counterfactual distributions of births are plotted in red, and the actual distribution of births is plotted in green.

Figure A7: Sensitivity check: Counterfactual distributions when leaving in municipalities affected by the famine

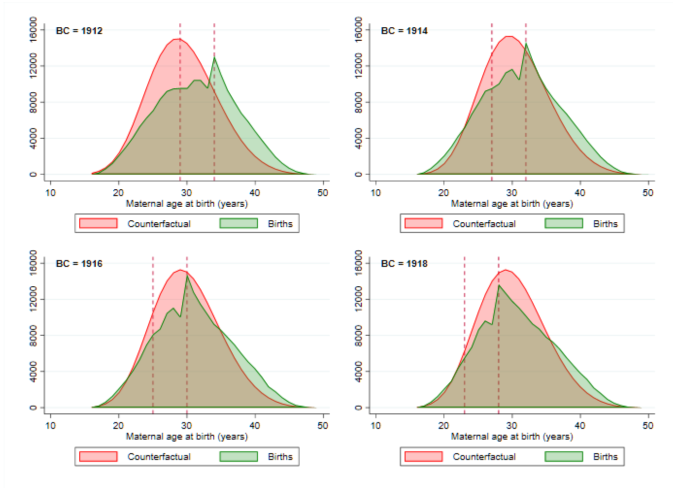
(a) Group 1: 1925–1930 cohorts



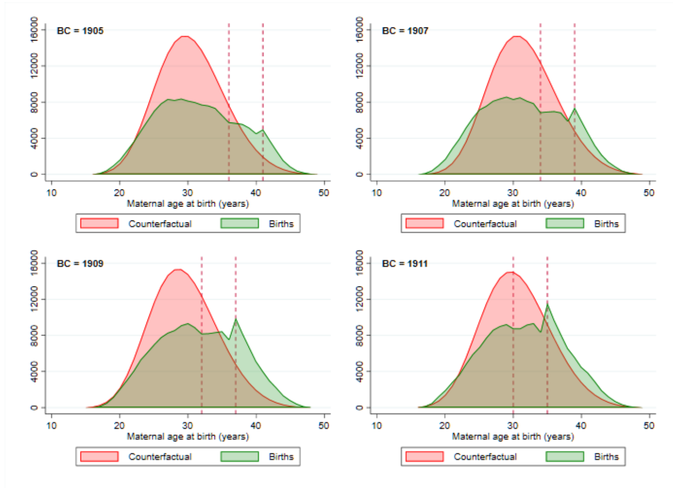
(b) Group 2: 1919–1924 cohorts



(c) Group 3: 1912–1918 cohorts



(d) Group 4: 1905–1911 cohorts



Notes: The figures show the results of the bunching counterfactual distribution estimation. For this robustness check, I include the seven municipalities that were most severely affected by the famine. The procedure to create these densities is discussed in Section 4.2. The counterfactual distributions of births are plotted in red, and the actual distribution of births is plotted in green.

Table A1: Air raids by year, the Netherlands, 1940–1945

Year	Number of air raids	Percentage of total
1940	305	25.0
1941	197	16.2
1942	171	14.0
1943	127	10.4
1944	261	21.4
1945	159	13.0
Total	1,220	100.0

Notes: Number of air raids by year. Author's calculations using data from Van Blankenstein (2006).

Table A2: Air raids by Air Force, the Netherlands, 1940–1945

Air Force	Number of air raids	Percentage of total
RAF	913	74.8
USAAF	91	7.5
Luftwaffe (Germany)	65	5.3
RAF + USAAF	34	2.8
AEAF	26	2.1
Unknown/other	91	7.5
Total	1,220	100.0

Notes: The RAF is the British Royal Air Force, the USAAF is the United States Army Air Force, the Luftwaffe is the German Air Force, and the AEAF is the Allied Expeditionary Air Force. Air raids that cannot be linked to an Air Force, and 2 air raids that were executed by other (combinations) of Air Forces, are grouped under unknown/other. Author's calculations using data from Van Blankenstein (2006).

Table A3: Summary statistics by prevalence of air raids in the municipality 1940–1945

	Any air raid (1)	No air raid (2)
A: Census — Individual-level characterstics		
Live births*	3.17	3.57
Childless*	0.096	0.091
Ever married	0.907	0.912
N*	726,028	626,617
N	800,160	687,008
B: Census — Municipality-level characterstics		
Total population	30,924.8	7,738.7
Urban	0.433	0.068
Proportion Dutch	0.954	0.967
Proportion Catholic	0.361	0.432
Proportion Orthodox Protestant	0.105	0.102
Proportion Liberal Protestant	0.342	0.337
Proportion other religion	0.033	0.019
Proportion no religion	0.159	0.110
N	171	649

Notes: Panel (a) shows characteristics of women in birth cohorts 1905–1930 depending on whether they live in municipalities that were or were not targeted by an air raids. The seven cities most severely affected by the famine of 1944–45 are left out. Note that the number of observations is lower for "live births" and "childless" because these measures are only observed for married women. Panel (b) shows municipality-level characteristics by the same split.

Table A4: Summary statistics by median unemployment 1932–1938

	>Median unemployment (1)	≤Median unemployment (2)
A: Census — Individual-level characterstics		
Live births*	2.99	3.08
Childless*	0.097	0.097
Ever married	0.913	0.901
N*	194,672	258,195
N	213,296	286,711
B: Census — Municipality-level characterstics		
Total population	74,654.8	101,302.7
Proportion Dutch	0.939	0.933
Proportion Catholic	0.423	0.419
Proportion Orthodox Protestant	0.069	0.078
Proportion Liberal Protestant	0.219	0.227
Proportion other religion	0.042	0.044
Proportion no religion	0.248	0.232
Total number air raids	14.44	14.78
N	18	18

Notes: Panel (a) shows characteristics of women in birth cohorts 1905–1930 depending on whether they live in municipalities with above or below median unemployment between 1932–1938. Statistics are reported for the 36 municipalities with more than 20,000 inhabitants for which information on unemployment rates are observed, and that experienced at least one air raid during WWII. The seven cities most severely affected by the famine of 1944–45 are left out. Note that the number of observations is lower for "live births" and "childless" because these measures are only observed for married women. Panel (b) shows municipality-level characteristics by the same split.

B: Data Appendix

I use administrative data from Statistics Netherlands to set up the sample of individuals born to mothers in birth cohorts 1905–1930. These data are available at a remote-access facility after signing a confidentiality agreement. An explanation is given below.

B1 Population-level administrative data

The period sample contains all individuals who were born between 1930 and 1960. To set up this sample, I start with the registry of persons (*GBAPERSOONTAB*). This registry contains information on all individuals who are registered in a Dutch municipality by 1995, which implies that I do not observe individuals who either moved abroad or died prior to 1995. I select all individuals born to mothers in birth cohorts 1905–1930. I start with the registry of persons and select all individuals who were born in the Netherlands and whose mothers were born between 1905 and 1930 (N = 4,699,028). I merge in information on the individual's place of birth (*VRLGBAGEBOORTEGEMEENTE*), and I drop 0.2% of individuals for whom the place of birth is missing. After this, the sample contains 4,688,942 individuals who were born to mothers in birth cohorts 1905–1930, and 4,215,264 of those were born between 1930–1960. Finally, I drop 1,075,384 individuals who were born in cities most-severely affected by the famine before running the analyses (22.9%). This leaves a final sample of 3,613,558 individuals born to mothers in birth cohorts 1905–1930, of which 3,236,340 were born between 1930–1960.

B2 Changes in municipal boundaries

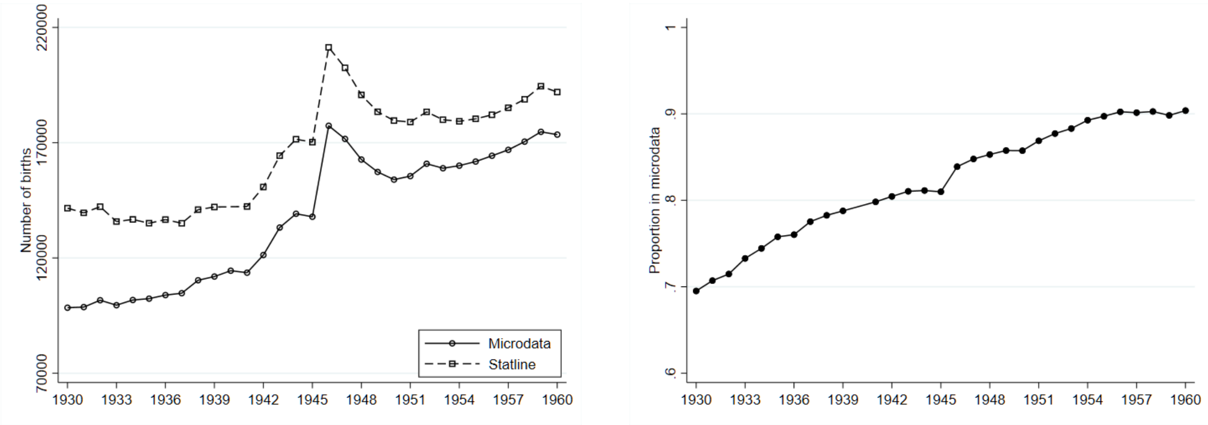
I use information on an individual's place of birth to match births to local circumstances around the time of conception and/or birth. The place of birth file matches an individual to their municipality code in the year of birth. The Netherlands went through many changes in municipal boundaries over time, and decreased from 1,078 municipalities in 1930 to 994 municipalities in 1960, so it is important to harmonize the boundaries of these municipalities so that a woman born in municipality X in 1930 can be compared to a woman born in municipality X in 1960. The general tendency of the Dutch government was to decrease the number of municipalities and hence increase the size of municipalities.

In case a new municipality was developed out of already existing municipalities, I assign a new municipality code to individuals who were born in the municipalities with the old municipality codes (and hence before the change in municipal boundaries). To illustrate, the municipalities of Ambt-Hardenberg (municipality code 1011) and Stad-Hardenberg (municipality code 1191) were merged into the municipality of Hardenberg (municipality code 160) in May 1941. In this case I assign the municipality code of the newly developed municipality of Hardenberg (160) to individuals born in all three municipalities.

I apply similar procedures to the outcomes and variables at the municipality level. Hence, to obtain the number of bombings by municipality, I aggregate the number of bombings for municipalities that merged. To obtain the liberation date for municipalities that merged, I take I take the latest municipal liberation date for municipalities that merged. After taking into account municipality changes and dropping the seven cities that were most severely affected by

Figure B1: Quality of microdata against aggregate population statistics

(a) Births in microdata vs aggregate data (b) Proportion observed in microdata



Notes: Panel (a) shows the number of births as observed in the microdata between 1930–1960 compared to the number of births as observed in aggregate data between 1930–1960 (i.e., Statline). Panel (b) shows the proportion of births observed in the microdata as a proportion of all aggregated births as in the population registries (from CBS Statline). Both panels exclude births in cities most severely affected by the Hunger Winter (Amsterdam, Delft, Den Haag, Haarlem, Leiden, Rotterdam, and Utrecht).

the famine, I am left with 895 municipalities.

B3 Data quality

Individuals are only observed in the microdata if they are registered in an Dutch municipality in 1995. That implies that I may not observe all individuals who were born between 1930 and 1960 as individuals could have moved abroad or died before 1995 and therefore are not observed in the microdata.

To check the coverage of the administrative data I start with the registry of persons (*GBAPERSOONTAB*) and restrict my sample to individuals born in the Netherlands between 1930–1960 and for whom I can observe the birth year of their mother (N = 5,548,452). I then merge in information on the individual's place of birth (*VRLGBAGEBOORTEGEMEENTE*), which I can identify for 99.8% of the individuals in my sample. Observations for whom place of birth cannot be identified are dropped from the sample. The sample contains 5,538,834 individuals who were born between 1930–1960. To match the analysis sample, I drop 1,275,315 individuals born in the cities most-severely affected by the famine, which leaves a final sample of 4,263,519 individuals who were born between 1930 and 1960.

To check what proportion of observations are actually observed in the microdata, I collect data on the number of births for the Netherlands from 1930 to 1960 from Statline (Statistics Netherlands' public data portal, statline.cbs.nl). The main analyses in this paper exclude children born in cities that were most severely affected by the Hunger Winter (Amsterdam, Delft, Den Haag, Haarlem, Leiden, Rotterdam, and Utrecht), and hence I collect extra data on the number of births in those cities for 1930–1960 from the Historical Collection of Statistics Netherlands (historisch.cbs.nl). Using this information, I calculate the proportion of births that are observed in the microdata, excluding the births in famine-affected areas.

On average, I observe 82.3% of births over the entire period, and Panel (a) of Figure B1 shows that this proportion is the lowest in 1930 (69.5%) and increases to 90.4% in 1960. This follows naturally from the fact that individuals born in 1930 are older and thus more likely to have died or moved abroad before 1995 than individuals who were born in 1960. The line is quite smooth, which suggests that there is no non random process going on that makes me more likely to observe births in some years. There does appear to be a dip in 1945, which can likely be explained by the onset of the Hunger Winter in 1944–1945. However, overall it seems that even though I do not observe all individuals in the data, the process seems to be similar over time.