Using data on biomarkers and siblings to study early-life economic determinants of type-2 diabetes

Alessie, Rob J.M.; Angelini, Viola; van den Berg, Gerard J.; Mierau, Jochen O.; Niccodemi, Gianmaria

Published in:
Health Economics

DOI:
10.1002/hec.4815

Received: 23 March 2023 | Revised: 9 January 2024 | Accepted: 6 February 2024

DOI: 10.1002/hec.4815

RESEARCH ARTICLE

Health Economics WILEY

Using data on biomarkers and siblings to study early-life economic determinants of type-2 diabetes

Rob J. M. Alessie[1,2] | **Viola Angelini**[1,2] | **Gerard J. van den Berg**[1,3,4,5] | **Jochen O. Mierau**[1,3,6] | **Gianmaria Niccodemi**[7]

[1]University of Groningen, Groningen, The Netherlands

[2]Netspar, Tilburg, The Netherlands

[3]University Medical Centre Groningen, Groningen, The Netherlands

[4]IZA, Bonn, Germany

[5]ZEW, Mannheim, Germany

[6]Lifelines, Roden, The Netherlands

[7]University of Luxembourg, Esch-sur-Alzette, Luxembourg

Correspondence
Jochen O. Mierau.
Email: j.o.mierau@rug.nl

Abstract

We study the effect of economic conditions early in life on the occurrence of type-2 diabetes in adulthood using contextual economic indicators and within-sibling pair variation. We use data from Lifelines: a longitudinal cohort study and biobank including 51,270 siblings born in the Netherlands from 1950 onward. Sibling fixed-effects account for selective fertility. To identify type-2 diabetes we use biomarkers on the hemoglobin A1c concentration and fasting glucose in the blood. We find that adverse economic conditions around birth increase the probability of type-2 diabetes later in life both in males and in females. Inference based on self-reported diabetes leads to biased results, incorrectly suggesting the absence of an effect. The same applies to inference that does not account for selective fertility.

KEYWORDS
business cycle, developmental origins, early-life conditions, unemployment

JEL CLASSIFICATION
I15, J11

1 | INTRODUCTION

It is well-established that adverse conditions in utero and shortly after birth negatively affect individual health later in life (Almond & Currie, 2011; Almond et al., 2018; Barker, 1995, 1998). Fetuses and newborns are sensitive to stressful environments and are "programmed" to deal with survival in similar environments after birth. A major discrepancy between the early-life environment and the environment later in life (e.g., birth in poverty and adulthood in an affluent society) may then negatively impact health outcomes over the life-cycle.

Early-life conditions can be measured by indicators of the parental living circumstances at birth. However, an association between such indicators and health in adulthood does not automatically imply a causal effect as unobserved factors may affect both. For example, parental genetic predisposition to some diseases may decrease the parental socioeconomic status at birth and worsen the health outcomes in adulthood. To overcome this endogeneity issue, a growing literature exploits natural experiments, using exogenous variation in contextual (macro)

circumstances at birth that are quasi-randomly assigned to individuals as proxies for economic conditions early in life. Landmark examples of this approach exploit exposure to famines or recessions (see the overviews by Almond and Currie (2011) and Almond et al. (2018)). In the famine-based literature, the late-life outcome that seems most responsive to early-life malnutrition is type-2 diabetes.[1] The comprehensive overview by Lumey et al. (2011) shows that this adverse effect of early-life malnutrition has been found across a range of different famines.[2]

Famines are catastrophic events, typically dating back a long time and/or having occurred in low and middle income countries. In this paper we expand on the existing evidence by exploiting milder fluctuations in early-life conditions in more recent eras. From a policy perspective, knowledge on determinants of type-2 diabetes is of high interest, as the economic burden of this disease is high and is expected to further increase in the next decades worldwide (Lin et al., 2020). In the Netherlands, the total economic burden of type-2 diabetes amounts to roughly 1% of the Gross Domestic Product (see Peters et al., 2017).

We exploit spatial and temporal variation in unemployment rates at birth as contextual variation in early-life economic conditions. Effectively, we compare type-2 diabetes outcomes of those born in recessions to those born in adjacent boom years, in the Netherlands in post-1950 birth cohorts. The paper makes a number of innovative methodological contributions. The first of these concerns the observability and measurement of type-2 diabetes. Type-2 diabetes is known to be severely under-diagnosed in the population. Many of those who are affected are unaware of having this condition, leading to under-reporting in survey data. Those with lower socio-economic status may consult doctors less systematically, and the disease may be asymptomatic in younger and prime-aged individuals. For example, Whicher et al. (2020) report that in the UK in 2019 six out of ten individuals did not have any symptoms at the time of the diagnosis and that nearly 1 million people had undiagnosed type-2 diabetes. We solve this by identifying type-2 diabetes through biomarkers. In accordance to the diagnosis criteria of the American Diabetes Association 2020, we consider the individuals to be affected by type-2 diabetes if the hemoglobin A1c (HbA1c) concentration in the blood is equal to or higher than 6.5% or if the fasting glucose level is equal to or higher than 7 mmol/L. The results are robust with respect to small deviations of the thresholds used. We examine to which extent the biomarker outcome variable is related to self-reported type-2 diabetes, in the same sample of individuals. More importantly, we compare the empirical findings to those that only rely on self-reported diabetes.

A second contribution concerns the econometric correction for potentially selective fertility. Recent studies on long-run health effects have found evidence of such selective fertility depending on the unemployment rate at birth (see van den Berg et al., 2020, for an overview). For example, Dehejia and Lleras-Muney (2004) find that the socio-economic status affects how fertility changes when the unemployment rate is high. As we shall see, our own data reveal substantial selective fertility in the sense that low-educated parents target fertility to periods of low unemployment. As parental education is observed, we can control for this. To account for selective fertility driven by unobservables across families we adopt a sibling study design, effectively comparing outcomes of siblings born at different stages of the business cycle. This approach follows other studies, such as van den Berg et al. (2020) for outcomes in infancy and van den Berg et al. (2017) for high-age outcomes. Note that the sibling design also controls for potential associations between parental genetic predisposition to type-2 diabetes and fertility. It has been estimated that the heritability of type-2 diabetes is of at least 20% (Ali, 2013).

We use the Lifelines data which is a cohort study of over 167,000 individuals, consisting of longitudinal surveys and a biobank, based on a multi-generational sampling design. Most individuals reside in the Northern part of the Netherlands (see Scholtens et al., 2015, for an overview). The data include over 51,000 siblings born between 1950 and 1995. Sibling-sibling relationships are self-reported or identified through questionnaire information such as coded surname and date of birth. These relationships were all checked and, if necessary, corrected by using linked administrative register data. Interestingly, a sub-sample of the Lifelines data, called Lifelines-UGLI, contains genetic records that enable us to identify sibling-sibling relationships in a completely different way, namely by using genetic similarity. Exploiting both types of sibling identifier opens up exciting possibilities for the study of health determinants. As it happens, the sub-sample of biological but non-self-reported siblings is currently too small for quantitative inference beyond simple categorizations of differences in terms of covariates. The lack of sample size also prevents separate analysis with non-biological sibling-sibling relationships due to adoption.

The outline of the paper is as follows. In Section 2 we discuss the data. In Section 3 we discuss the sibling fixed-effects econometric model that we estimate; in Section 4 we present and discuss the main results as well as several sensitivity analysis; in Section 5 we discuss some policy suggestions and conclude the paper.

2 | DATA

2.1 | The Lifelines data

Lifelines is a multi-disciplinary prospective population-based cohort study examining in a unique three-generation design the health and health-related behaviors of 167,729 persons living in the North of the Netherlands. It employs a broad range of investigative procedures in assessing the biomedical, socio-demographic, behavioral, physical and psychological factors which contribute to the health and disease of the general population, with a special focus on multi-morbidity and complex genetics.

The participants, usually recruited by the general practitioner in the three northern provinces (Groningen, Friesland and Drenthe), were often asked to invite their family members. Consequently the number of siblings in the cohort study is relatively high. Initially, siblings could only be identified if at least one of their parents also participated in Lifelines. Recently, however, through the so-called Pedigree file, all siblings could be identified. This file links individuals to a unique identifier for both the mother and the father regardless of the presence of the mother and the father in the database. It was constructed primarily by invoking the parent-child and sibling-sibling relationships that were declared by Lifelines participants. Other variables were used to complete the file, as it could not be ruled out that siblings who participated in the study without their parents did not declare their family ties. In particular, the Lifelines questionnaires contain questions on number of siblings, year of birth of the siblings and coded surnames. All the sibling-sibling relationships were verified and, if necessary, corrected through administrative data.

Having a unique identifier for both the mother and the father enables us to define as siblings those who have the same mother and father. This is an advantage compared to data that only allow them to be defined by a common mother. Lifelines-UGLI, a sub-sample of the original Lifelines data that includes genetic records, was used to identify genetically defined sibling-sibling relationships as well as non-biological (due to adoption) sibling-sibling relationships.[3] From the 19,005 genetically defined siblings we uncovered 110 additional siblings whereas 13 were identified as non-biological adopted siblings. The latter were dropped. The total number of siblings aged 18 or above is 51,270.

Some of these siblings could not be used in the analyses as they could not be linked to provincial unemployment data. These include individuals born abroad (376) or in homonymous Dutch cities that belong to different provinces (387). Moreover, because provincial unemployment data is not available from before 1950 we drop individuals born before that (3,484). Flevoland, which became an independent province in 1986, only has unemployment data since 1973, leading to the additional exclusion of 199 individuals. We further exclude from the data 1107 individuals for whom neither the hemoglobin A1c (HbA1c) concentration in the blood nor the fasting glucose level in the blood used for the diagnosis of type-2 diabetes was collected. Finally, we exclude those individuals who declare to have type-1 diabetes (88), other forms of diabetes (109) or who do not know which type of diabetes they have (47). Obviously, excluding individuals entails that for some families only one sibling remains in the sample. Such individuals do not contribute to fixed-effects econometric analyses so we exclude those as well (2,342). The resulting number of siblings used in our study is 43,131. Of these, 25,895 are females and 17,236 are males. After these exclusions the number of siblings for whom the genetic records are available is 16,375, of which 68 are biological siblings linked to their parents only through the genetic data and 12 are the adopted siblings excluded from the data.

Following the standard diagnosis criteria defined by the American Diabetes Association (2020), we define the binary dependent variable *T2D* to be equal to one if the HbA1c concentration in the blood is equal to or higher than 6.5% or if the fasting glucose level in the blood is equal to or higher than 7 mmol/L, and to be equal to 0 otherwise. Both the HbA1c concentration and the fasting glucose level in the blood are commonly used for the diagnosis of type-2 diabetes. The HbA1c concentration in the blood is a measure of *chronic* glucose exposure as it reflects the average glucose concentration in the previous 3 months and is relatively stable over time. Consequently, measuring the HbA1c concentration in the blood does not require the individuals to fast before the test, while the actual glucose level in the blood varies with food digestion and so requires the individual to fast for 8 h beforehand.[4] It is important to note that our sample does not include pregnant women as their measurement was rescheduled to 6 months after pregnancy or 3 months after breastfeeding.

Descriptive statistics are reported in Table 1. The prevalence of type-2 diabetes in our sample is 1.47% (1.27% for females and 1.76% for males). This is lower than in the Dutch population, where for 2013, the average is estimated to be 5.13% (Kleefstra et al., 2016). This is explained by the age distribution in our sample which does not include individuals over 63. The data also show that the between-siblings standard deviation of *T2D* is 8.42% while the within siblings standard deviation is 8.71%. Age is measured in years at baseline. The provincial unemployment rate is measured in

TABLE 1 Descriptive statistics.

Variable	Obs.	Mean	Std. Dev.	Min	Max
Full sample					
T2D in females (%)	25,895	1.27	11.18	0	100
T2D in males (%)	17,236	1.76	13.16	0	100
Unemployment rate (%)	43,131	5.22	3.17	0.5	13.8
Age	43,131	41.18	10.51	18	63
First-born	43,131	0.44	0.50	0	1
Low education	42,394	0.61	0.49	0	1
Genetically-confirmed biological siblings					
T2D in females (%)	9795	1.03	10.10	0	100
T2D in males (%)	6580	1.49	12.11	0	100
Unemployment rate (%)	16,375	5.77	3.38	0.5	13.8
Age	16,375	39.26	11.26	18	63
First-born	16,375	0.41	0.49	0	1
Low education	16,117	0.57	0.49	0	1

Note: *T2D* is a dummy variable equal to one if the HbA1c concentration in the blood is equal to or higher than 6.5% or if the fasting glucose level is equal to or higher than 7 mmol/L, and equal to 0 otherwise. "Low education" refers to the education of the individual. It is not controlled for in the analysis and it is not observed for all the individuals in our sample but it is included for completeness.
Genetically-confirmed biological siblings include those siblings who were biologically related to their mother and father in the data.

annual percentages. The binary variable *First-born* is defined to equal 1 for the first-born sibling and to equal 0 otherwise as well as for twins.[5] The medical literature has found evidence that first-born children are more likely to develop diabetes (Ayyavoo et al., 2013; Lammi et al., 2007). Although the medical explanation for this association is not entirely clear, it is thought to be related to changes in the uterus occurring between the first and subsequent pregnancies.

2.2 | Selective fertility by socio-economic status

A relationship between early-life socio-economic conditions and health outcomes later in life may be driven by selective fertility whereby certain groups are more likely or less likely to have children during economic downturns. We use the education level of the parents to gauge selective fertility along this dimension over different unemployment levels in the birth year (see Subsection 2.3 below for an explanation of how we use unemployment to capture variation in economic early-life conditions). Note that at this stage we do not deal with selection on unobservables yet. The education level of the parents is not reported by the siblings, so we have this data only for the parents who participated in the study. Specifically, we have information on the completed education level of the mothers of 13,685 siblings and the fathers of 9949 siblings. For 8102 siblings we have information on the completed education level of both the mother and the father. We define as high education a level of senior general secondary education[6] or above and as low education all the other levels.

It turns out that 2595 siblings have mothers with high education, 11,090 siblings have mothers with low education, 2757 siblings have fathers with high education, and 7192 siblings have fathers with low education. Finally, among 1068 siblings both parents have high education, among 5134 siblings both have low education, and among 1900 siblings one parent has a high education and one a low education.

Figure 1 depicts kernel estimates of the probability density functions of unemployment rates at birth, by education level of the parents (with bandwidths of 0.8% points throughout). Specifically, in Figure 1a we report the birth distribution for the cases where (i) both parents have low education, (ii) both parents have high education, and (iii) one

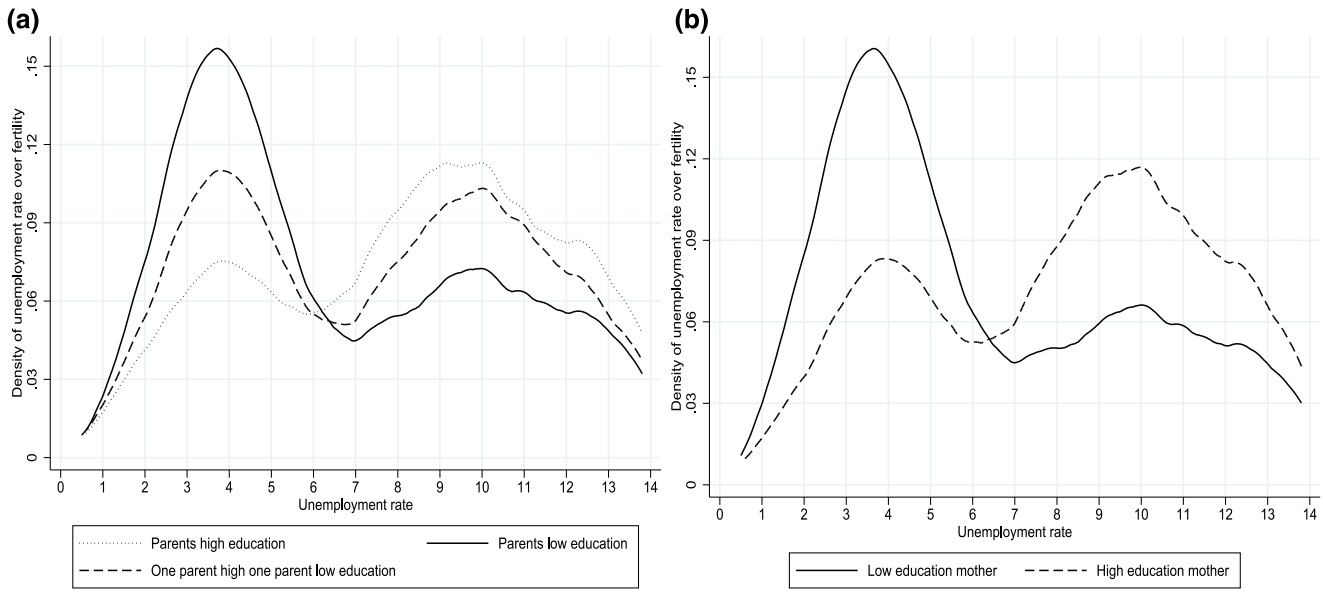

FIGURE 1 Selective fertility by socio-economic status over unemployment. (a) Selection by education of the parents. (b) Selection by education of the mother.

parent has high education while the other has low education. It can be seen that parents with lower education tend to have children when unemployment is low. We produce the same graph based on education of the mother only (Figure 1b) and we observe a similar pattern. Although we only observe individuals who became parents and so we are unable to compute the fertility rates by socio-economic status, these graphs suggest that unemployment rates are not randomly assigned to newborns in the sense that they are somewhat related to the parental socio-economic status. This in turn suggests that there may also be systematic fertility differences in unobserved parental characteristics, strengthening the case for a sibling fixed-effects approach when estimating long-run effects of conditions around birth.[7]

2.3 | Unemployment

We may distinguish between two phases in the economic development of the Netherlands in the era after World War II. In the so-called Golden Age (1950–1973) the Dutch economy experienced high growth, led by the liberalization of trade with Western European countries and the growth of the petro-chemical and agricultural sector (Van Zanden, 2005). After 1973 economic growth slowed down, leading to an economic crisis in the early 1980s due to the collapse of the Bretton Woods system (1971), the two oil crisis (1973 and 1979) and the expansion of the public sector (Smits et al., 1999; Van Zanden, 2005). The manufacturing sector declined and unemployment increased substantially. In many respects, these developments were mirrored in other West-European countries.

The two phases are clearly visible in Figure 2, with the unemployment rates decreasing until the early 1970s and then increasing during the various economic crises leading up to the early 1980s. The three provinces in the Northern Netherlands experienced higher unemployment rates than most other provinces in the Netherlands. To gauge the relevance of the years depicted on the horizontal axis of Figure 2 it should be considered that over 91% of the individuals in our data with type-2 diabetes were born before 1973. This reflects the fact that prevalence increases with current age. Importantly, Figure A1 in the Appendix shows that there is within family variation in unemployment across siblings.[8]

In the absence of reliable data on regional GDP, we use the provincial unemployment rate as a proxy for transitory contextual economic conditions. In the birth years in the data, some sort of unemployment benefits and welfare was typically available for unemployed workers. However, unemployment did lead to income loss and uncertainty about future work opportunities. Also, in recessions, families that did not encounter unemployment were nevertheless affected by the risk of future unemployment. In all this it is worth noting that female labor force participation in the period 1950–1973 was low by international standards and most household had at most one breadwinner.[9]

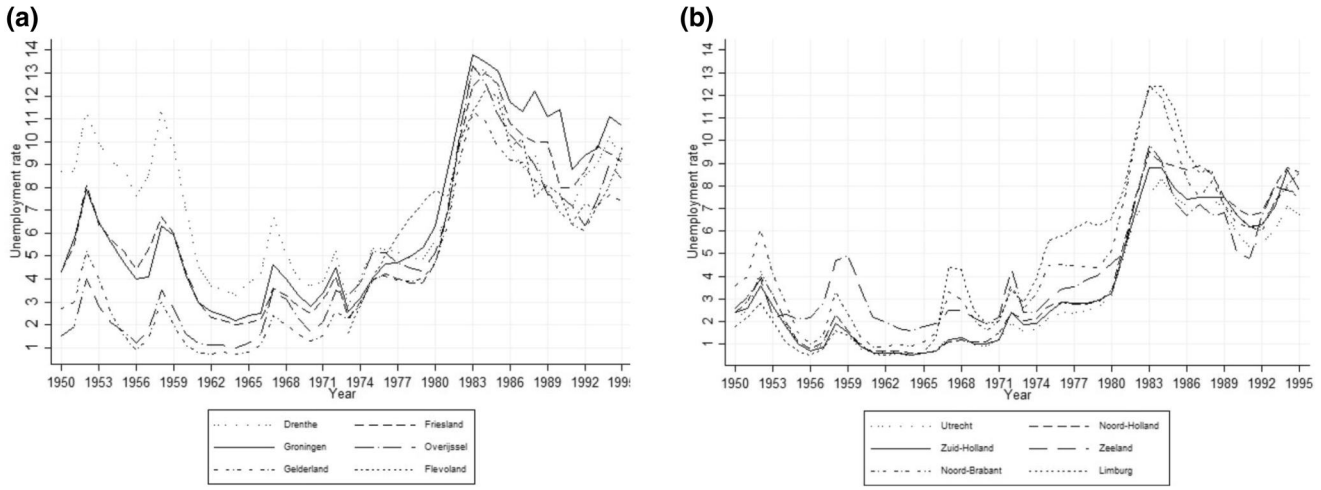

FIGURE 2 Provincial unemployment rate series. (a) North and East Netherlands. (b) West and South Netherlands.

3 | EMPIRICAL IMPLEMENTATION

We define the following model:

$$T2D_i = \beta_1 g_i \cdot u_{p(i)} + \beta_2 (1 - g_i) \cdot u_{p(i)} + x_i \beta_3 + \alpha_{j(i)} + \varepsilon_i \quad (1)$$

where the indices i, j, p denote individual, family and province, respectively. $T2D_i$ is a binary variable equal to one if the individual is affected by type-2 diabetes in the year that this was measured, and equal to 0 otherwise. To avoid misunderstandings, note that we only observe $T2D$ once per individual. In the notation we suppress that $T2D_i$ and covariates x_i are measured at the Lifelines baseline collection date of individual i. It should also be pointed out that the baseline collection date is not identical across individuals, so $T2D_i$ may be measured at different points in time for different individuals i. The binary gender indicator $g_i = 1$ if the individual is female. Next, $u_{p(i)}$ is the unemployment rate in the birth province and birth year. The coefficients β_1 and β_2 are the parameters of interest. The residual term in Equation (1) is the sum of a component $\alpha_{j(i)}$ shared among siblings and an idiosyncratic component ε_i. To account for selective fertility and other family-shared determinants of diabetes, we control for dependence between $\alpha_{j(i)}$ on the one hand and $u_{p(i)}$, g_i and x_i on the other hand, by estimating sibling fixed-effects regressions.

The vector x_i contains covariates that may differ between siblings, such as gender, being first-born, current age at baseline, the birth province (via "fixed effects" γ_p) and the birth year (via "fixed effects" δ_t for birth years t). The latter warrants some discussion because it relates to our source of identifying variation. In a cross-sectional sample collected in a given year, current age and birth year (as parameterized by δ_t) are fully collinear. In our sampling design, their effects are separately identified because siblings may have had their Lifelines baseline date at different points in time, somewhere in a window of 8 years covering 2006–2013. Clearly, this is a fragile way of identifying these effects. Moreover, if we do not impose any functional form on the δ_t sequence then the estimated sequence will capture the national business cycle effect so that the effects of $u_{p(i)}$ are only identified by variation across provinces within birth years. Clearly, variation in the national business cycle over time quantitatively dominates cross-sectional variation across provincial business cycles.[10] We therefore also estimate more robust model specifications where the δ_t parameters are absent and where current age is included by way of a flexible polynomial.[11]

In addition, we estimate model 1 with sibling random effects in order to subsequently test it against the fixed-effects specification, using the so-called Mundlak approach (see Wooldridge, 2010). As unemployment at birth is exogenous from the individual perspective, rejecting random effects in favor of fixed effects in the Mundlak approach suggests that parental selective fertility should be accounted for. Put differently, if unemployment at birth is randomly assigned to newborns and so there is no significant difference in the coefficient estimates between the random-effects and fixed-effects specifications, the Mundlak test leads to adoption of the random-effects specification with a more efficient estimator than the fixed-effects estimator.

4 | RESULTS

The sibling fixed-effects estimates based on Equation (1) are reported in Table 2. We find a significant effect of the unemployment rate at birth on type-2 diabetes: 1% point higher unemployment at birth increases by 0.3% points the probability of having type-2 diabetes in both females and males. As the prevalence of diabetes in our sample is 1.47%, this can be considered a large effect. We conclude from this that there is a causal effect of adverse economic conditions around birth on the development of type-2 diabetes later in life.

The random-effects estimates are also reported in Table 2. Here, the effect of the unemployment rate at birth is not significant. However, the Mundlak test rejects the random-effects specification in favor of the fixed-effects specification at the 10% level.[12]

To ascertain the robustness of our results we perform a series of sensitivity analyses.[13] These are discussed by topic in the remainder of this section:

4.1 | Sibling status

First, we consider the impact of the way in which siblings are identified. The estimates reported in Table 2 are virtually identical to those based on data that are not corrected through the genetic records, namely including adopted siblings and excluding siblings whose blood-relationship is not reported or identified through questionnaire information. The same applies if we define as siblings those who have the mother in common but not necessarily the father (43,247 in total).

As an alternative to the sibling design one may sample cousin-cousin relationships in order to run cousin fixed-effects. The advantage of using cousins instead of siblings is that they can be born at the same point in time and in different provinces, so more variation of unemployment rate at birth can be exploited. Unfortunately, we can only identify these relationships if the sibling-sibling relationships of the respective parents are identified in Lifelines, so most of the cousin-cousin relationships remain unidentified. This severely reduces the sample size compared to our

TABLE 2 Sibling FE and RE estimates on *T2D*.

	FE	RE
Unem. × female	0.0030*	0.0009
	(0.0016)	(0.0009)
Unem. × male	0.0032**	0.0008
	(0.0016)	(0.0009)
Female	−0.0044	−0.0058**
	(0.0030)	(0.0022)
Age	0.0007	0.0009**
	(0.0008)	(0.0004)
First born	0.0049***	0.0036***
	(0.0018)	(0.0011)
Province FE	Yes	Yes
Year of birth FE	Yes	Yes
Mundlak test *p*-value		0.0620
$\hat{\sigma}^2_\alpha$		0.0006
$\hat{\sigma}^2_\epsilon$		0.0136
Obs.	43,131	43,131

Note: *T2D* is a dummy variable equal to 1 if the HbA1c concentration in the blood is equal to or higher than 6.5% or if the fasting glucose level is equal to or higher than 7 mmol/L, and equal to 0 otherwise. Standard errors (in parentheses) are clustered at the family level. *$p < 0.1$ **$p < 0.05$, ***$p < 0.01$. α and ϵ are the family-specific component and the idiosyncratic component of the error term, respectively.

sibling sample. Moreover, the cousins thus identified are on average younger than the siblings in our data, implying that type-2 diabetes is rarer. Additional challenges are (*i*) that cousins can give rise to chains of relationships where two individuals that are not cousins of each other are each cousins of a third individual, and (*ii*) that an individual can have cousins that are siblings of each other. Any simple econometric approach would therefore require the cousin sample size to be reduced even further. In conclusion, a cousin study design is infeasible and we do not pursue this further. Finally, our results are robust to the exclusion of the top 5% of individuals with the largest spacing.

4.2 | Unemployment, province of birth, and age in the model specification

To check whether the results are driven by inter-province mobility due to differences in unemployment rates, we estimate the model omitting those 3881 siblings that, within a family, were born in different provinces. Obviously, we now exclude the province fixed-effects γ_p. We obtain very similar results to the estimates reported in Table 2, although the effect among females is now only marginally significant.

Next, we turn to the role of age. It is well known that age is a risk factor for the occurrence of type-2 diabetes. The American Diabetes Association 2020 states that testing individuals for high diabetes risk should begin at age 45. The estimated age effect may be conflated with the effects of parity and, in particular, birth-year fixed-effects, as discussed in the previous section. We proceed in a number of ways. First, we interact the unemployment effects with age. The estimates are reported in Table A2 in Appendix A while the marginal effects at different values of age are reported in Table 3. It can be seen that the effect of unemployment manifests itself with age with a significant impact among males above the age of 40. Second, we estimate the gender-specific average effect of unemployment at birth for the population of siblings over the age of 45. The estimates are reported in Table 4. As expected, the average effect of unemployment at birth is much higher for this sub-population of elderly siblings. Finally, and importantly, we estimate the main model excluding the birth-year fixed-effects δ_t while including age as well as age squared. We compute the marginal effect of age and we find that this is significant and increases with age, as expected (see Table A3 in Appendix A).[14]

It should also be pointed out that the results are robust with respect to the choice of the interval of birth years in the sample. In particular, larger intervals lead to virtually identical results.

4.3 | Operationalization of type-2 diabetes

In Table A4 in Appendix A we also report the estimates based on self-reported type-2 diabetes as dependent variable. It can be seen that the estimates of the parameters of interest are not significant. This is likely due to underdiagnosis of the disease, suggesting that self-reported data on type-2 diabetes is not reliable. In principle, measurement error in the dependent variable should not bias the estimates as long as this measurement error is randomly distributed. To investigate whether this is the case, we divide individuals who have type-2 diabetes based on the biomarkers in two groups. Individuals in the first group report having the disease while individuals in the second group report not having the disease. In Table A5 of Appendix A we display descriptive statistics of the two groups regarding age, gender and education and we run a *t*-test on the mean-comparisons between the two samples. A number of things stand out. First, more than 50% of the individuals who, based on the biomarkers, have type-2 diabetes do not report the disease. Second, the two groups differ by gender (although this difference is not significant), education and age, with younger individuals more likely to have undiagnosed type-2 diabetes. The latter may be because the disease is asymptomatic at early stages and because, as already mentioned, recommendations point toward testing after the age of 45. The difference in the education level might be explained by the age difference between the two samples, with younger individuals more likely to have higher education.

4.4 | Placebo regressions

To lend additional credibility to our results, we conduct a series of placebo regressions. First, macroeconomic conditions before conception should not have any impact on the probability of contracting type-2 diabetes, otherwise our main results might be spurious rather than true causal effects. Therefore we test the robustness of our main results by

regressing type-2 diabetes on unemployment rate 2 years before birth (so at least 1 year before conception). For this regression we exclude those who were born before 1952 or in the Flevoland province before 1975 from the data, as we do not have unemployment data for those years. Results are reported in Table A6 in Appendix A. Clearly, there is no significant effect of the unemployment rate on the occurrence of type-2 diabetes in adulthood. Second, in Table A7, we randomly reassign each province of birth to another one, ensuring that all individuals born in a specific province are assigned the same 'new' province. Also in this case, we find no significant effect of the unemployment rate in the randomly assigned province of birth. Finally, we conduct a similar exercise by randomly assigning the year of birth. In Table A8, we randomly assign each possible year of birth to another one, ensuring that all individuals born in a specific

TABLE 3 Marginal effect of unemployment by age and gender on *T2D*.

Age	Female	Male
20	0.0011	0.0009
	(0.0017)	(0.0017)
30	0.0016	0.0020
	(0.0013)	(0.0014)
40	0.0021	0.0031**
	(0.0014)	(0.0014)
50	0.0026	0.0042**
	(0.0018)	(0.0019)
60	0.0031	0.0054**
	(0.0025)	(0.0025)

*$p < 0.1$ **$p < 0.05$, ***$p < 0.01$.

TABLE 4 Sibling FE and RE estimates on *T2D*: siblings over age 45.

	FE	RE
Unem. × female	0.0059*	0.0018
	(0.0033)	(0.0023)
Unem. × male	0.0084**	0.0037
	(0.0034)	(0.0024)
Female	0.0023	−0.0014
	(0.0074)	(0.0057)
Age	0.0009	0.0007
	(0.0017)	(0.0010)
First born	0.0077**	0.0073***
	(0.0039)	(0.0027)
Province FE	Yes	Yes
Year of birth FE	Yes	Yes
Mundlak test *p*-value		0.0948
$\hat{\sigma}^2_\alpha$		0.0018
$\hat{\sigma}^2_\epsilon$		0.0247
Obs.	15,617	15,617

Note: *T2D* is a dummy variable equal to 1 if the HbA1c concentration in the blood is equal to or higher than 6.5% or if the fasting glucose level is equal to or higher than 7 mmol/L, and equal to 0 otherwise. Standard errors (in parentheses) are clustered at the family level. α and ϵ are the family-specific component and the idiosyncratic component of the error term, respectively.

*$p < 0.1$ **$p < 0.05$, ***$p < 0.01$.

year are assigned the same 'new' year. In Table A9, we randomly assigned the year of birth at the individual level. In both cases, we observe no significant effect of the unemployment rate in a randomly assigned year of birth on diabetes.

5 | CONCLUSION

Among males, a one percentage point higher provincial unemployment rate at birth increases the probability of developing type-2 diabetes in adulthood by 0.3% points. This increases up to 0.8% points for the sub-population of individuals older than 45. In each case we reject the null hypothesis of no effect. Among females, the estimated effect sizes are similar, but the coefficients are only marginally significant. In our explanatory framework, the results implies that at least among men, adverse economic conditions in utero and shortly after birth increase the probability of type-2 diabetes later in life.

The fact that diabetes in modern societies responds to modest changes in economic conditions around birth means that the landmark findings on long-run effects of famines on diabetes can be generalized from extreme historical settings to common contemporaneous settings. Our study therefore extends the relevance of the findings in the famine studies, toward modern economies.

We find that analyses that solely rely on the self-reported prevalence of type-2 diabetes produce systematically different results.[15] This suggests that self-reported data are not reliable. Individuals who do not report diabetes but have diabetes according to the official diagnostic guidelines may be expected to suffer from long-term health damage if no medical help is requested and/or if the diabetes goes undetected for many years.

Our study shows that not accounting for selective fertility leads to insignificant and hence erroneous estimates for the effects of economic conditions at birth on type-2 diabetes. One may argue that in studies like ours, where individuals have a high degree of control over fertility, the potential for endogenous fertility selection is larger than in historical studies. The fact that it matters a great deal if selection is properly dealt with is therefore of importance for the wider literature using relatively recent birth cohorts.

The data enable a comparison between self-reported sibship and sibship status based on genetic identifiers. The sub-sample of biological but unreported siblings is currently too small for quantitative inference, but we do observe systematic differences in terms of covariates. In the future, the availability of both types of sibling identifier may open up exciting research opportunities for the study of health determinants.

The results of the paper have some obvious policy implications. First of all, note that screening of individuals on diabetes is important because, as we have seen, self-reported diabetes underestimates clinically diagnosed diabetes. When targeting individuals for screening it would be efficient to take into account if individuals experienced adverse conditions around birth (e.g., if they are known to have grown up in a poor area or society) as their susceptibility is higher. Similarly, it would be efficient to focus on such groups when attempting to raise awareness of type-2 diabetes. All this is particularly relevant for type-2 diabetes as a health hazard because, if diagnosed early, it may be fully reversed for at least some groups of individuals, by way of specific weight-loss programs (see Lean et al., 2018; Taylor, 2019). A more down-to-earth policy recommendation would be to inform and support pregnant women and parents of newborns living in poor economic conditions. In this sense, prevention of type-2 diabetes may start before birth.

ACKNOWLEDGMENTS

We thank the services and participants of the Lifelines Cohort Study and the contributing research centers delivering data to Lifelines. The Lifelines initiative has been made possible by subsidy from the Dutch Ministry of Health, Welfare and Sport, the Dutch Ministry of Economic Affairs, the University Medical Center Groningen, University of Groningen and the Provinces in the North of the Netherlands (Drenthe, Friesland, Groningen). We are grateful to Govert Bijwaard, Aga Postepska, participants of the Econometric Society World Congress, the conference "New Data for the New Challenges of Population and Society", the American-European Health Economics Workshop 2022 and of seminars at the University of Rome La Sapienza and the University of Groningen for comments.

CONFLICT OF INTEREST STATEMENT

None.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available on request from the Lifelines Biobank and Cohort Study. The data are not publicly available due to privacy or ethical restrictions.

ORCID

Viola Angelini https://orcid.org/0000-0002-5177-3447
Jochen O. Mierau https://orcid.org/0000-0002-9968-8807

ENDNOTES

[1] For studies focusing on type-1 diabetes see, for instance, Eriksen et al. (2021).

[2] These findings are confirmed in animal experiments (Fernandez-Twinn & Ozanne, 2006; Portha et al., 2011) as well as in other natural experiments (Sotomayor, 2013) (see also Calkins & Devaskar, 2011, for another review).

[3] See Lopera Maya et al. (2020) for a description of the genotyping data.

[4] In our sample, the correlation between HbA1c and the fasting glucose level is 0.60. This confirms that the two underlying biomarkers are somewhat different. The correlation between the binary versions based on their diabetes threshold values is 0.71.

[5] Twins are only included if they have further siblings, otherwise they are excluded due to lack of between-sibling variation in early-life conditions.

[6] Senior general secondary education includes pre-univerity education such as HAVO, VWO, Atheneum, Gymnasium, HBS and MMS.

[7] As we shall see below, unemployment rates were highest in the most recent birth years. To the extent that parental education increases in the year of birth, some of the effects visualized in this subsection may be attributed to this. However, the relation between the unemployment rate and birth year in our sample is actually U-shaped. Moreover, the correlations between paternal and maternal education on the one hand and birth year on the other hand are only 0.15 and 0.25, respectively. These are low numbers, confirming that the patterns are not driven by secular increases in the attained level of education over time.

[8] The zeros in the figure are due to boundary bias in kernel density estimation. In practice, there are only a couple of families with no within variation (those with siblings born in the same calendar year)

[9] It should be pointed out that the Lifelines data do not provide detailed information on individual labor market outcomes over the years 1950–1995. This is relevant as it means that instrumental variable analyses with parental unemployment as treatment variable are not feasible.

[10] Nevertheless, Figure A2 in the Appendix shows that there is variation even when we demean provincial unemployment by subtracting the national average for that year.

[11] Similarly, the γ_p parameters are identified because some siblings who belong to the same family were born in different provinces.

[12] We also estimate a probit model with correlated sibling random effects, controlling for the family mean of the covariates (see Wooldridge, 2010). This approach fits well with skewed outcome distributions, as in our case. The estimates are reported in Table A1 in Appendix A. Simple probit estimates and marginal effects are also reported for comparison. Reassuringly, the estimated marginal effect of unemployment in the probit model with correlated random effects is very close to that in the fixed-effects linear probability model.

[13] When not reported, results of these analyses are available upon request.

[14] If year fixed-effects are also included in here, the point estimates of the marginal effect of age do not change a lot but the standard errors increase dramatically, again as expected. See again Table A3 in Appendix A.

[15] In particular, they give insignificant estimates of the long-run effects, presumably because self-reported prevalence is much lower than diagnosed.

REFERENCES

Ali, O. (2013). Genetics of type 2 diabetes. *World Journal of Diabetes*, *4*(4), 114–123. https://doi.org/10.4239/wjd.v4.i4.114

Almond, D., & Currie, J. (2011). Killing me softly: The fetal origins hypothesis. *The Journal of Economic Perspectives*, *25*(3), 153–172. https://doi.org/10.1257/jep.25.3.153

Almond, D., Currie, J., & Duque, V. (2018). Childhood circumstances and adult outcomes: Act II. *Journal of Economic Literature*, *56*(4), 1360–1446. https://doi.org/10.1257/jel.20171164

American Diabetes Association. (2020). Classification and diagnosis of diabetes: Standards of medical care in diabetes in 2020. *Diabetes Care*, *43*(Supplement 1), S14–S31. https://doi.org/10.2337/dc20-s002

Ayyavoo, A., Savage, T., Derraik, J. G., Hofman, P. L., & Cutfield, W. S. (2013). First-born children have reduced insulin sensitivity and higher daytime blood pressure compared to later-born children. *The Journal of Clinical Endocrinology & Metabolism*, *98*(3), 1248–1253. https://doi.org/10.1210/jc.2012-3531

Barker, D. J. (1995). Fetal origins of coronary heart disease. *BMJ*, *311*(6998), 171–174. https://doi.org/10.1136/bmj.311.6998.171

Barker, D. J. (1998). In utero programming of chronic disease. *Clinical Science*, *95*(2), 115–128. https://doi.org/10.1042/cs0950115

Calkins, K., & Devaskar, S. U. (2011). Fetal origins of adult disease. *Current Problems in Pediatric and Adolescent Health Care*, *41*(6), 158–176. https://doi.org/10.1016/j.cppeds.2011.01.001

Dehejia, R., & Lleras-Muney, A. (2004). Booms, busts, and babies' health. *Quarterly Journal of Economics*, *119*(3), 1091–1130. https://doi.org/10.1162/0033553041502216

Eriksen, T. L. M., Gaulke, A., Skipper, N., & Svensson, J. (2021). The impact of childhood health shocks on parental labor supply. *Journal of Health Economics*, *78*, 102486. https://doi.org/10.1016/j.jhealeco.2021.102486

Fernandez-Twinn, D., & Ozanne, S. (2006). Mechanisms by which poor early growth programs type-2 diabetes, obesity and the metabolic syndrome. *Physiology & Behavior*, *88*(3), 234–243. https://doi.org/10.1016/j.physbeh.2006.05.039

Kleefstra, N., Landman, G. W., Van Hateren, K. J., Meulepas, M., Romeijnders, A., Rutten, G. E., Klomp, M., Houweling, S. T., & Bilo, H. J. (2016). Dutch diabetes prevalence estimates (dude-1). *Journal of Diabetes*, *8*(6), 863–865. https://doi.org/10.1111/1753-0407.12370

Lammi, N., Moltchanova, E., Blomstedt, P., Eriksson, J., Taskinen, O., Sarti, C., Tuomilehto, J., & Karvonen, M. (2007). The effect of birth order and parental age on the risk of type 1 and 2 diabetes among young adults. *Diabetologia*, *50*(12), 2433–2438. https://doi.org/10.1007/s00125-007-0843-5

Lean, M. E., Leslie, W. S., Barnes, A. C., Brosnahan, N., Thom, G., McCombie, L., Peters, C., Zhyzhneuskaya, S., Al-Mrabeh, A., Hollingsworth, K. G., Rodrigues, A. M., Rehackova, L., Adamson, A. J., Sniehotta, F. F., Mathers, J. C., Ross, H. M., McIlvenna, Y., Stefanetti, R., Trenell, M., & Taylor, R. (2018). Primary care-led weight management for remission of type 2 diabetes (DiRECT): An open-label, cluster-randomised trial. *Lancet*, *391*(10120), 541–551. https://doi.org/10.1016/s0140-6736(17)33102-1

Lin, X., Xu, Y., Pan, X., Xu, J., Ding, Y., Sun, X., Song, X., Ren, Y., & Shan, P.-F. (2020). Global, regional, and national burden and trend of diabetes in 195 countries and territories: An analysis from 1990 to 2025. *Scientific Reports*, *10*(1), 14790. https://doi.org/10.1038/s41598-020-71908-9

Lopera Maya, E. A., van der Graaf, A., Lanting, P., van der Geest, M., Fu, J., Swertz, M., Franke, L., Wijmenga, C., Deelen, P., Zhernakova, A., & Sanna, S., & Lifelines Cohort Study. (2020). Lack of association between genetic variants at ACE2 and TMPRSS2 genes involved in SARS-CoV-2 infection and human quantitative phenotypes. *Frontiers in Genetics*, *11*, 613. https://doi.org/10.3389/fgene.2020.00613

Lumey, L. H., Stein, A. D., & Susser, E. (2011). Prenatal famine and adult health. *Annual Review of Public Health*, *32*(1), 237–262. https://doi.org/10.1146/annurev-publhealth-031210-101230

Peters, M., Huisman, E., Schoonen, M., & Wolffenbuttel, B. (2017). The current total economic burden of diabetes mellitus in The Netherlands. *The Netherlands Journal of Medicine*, *75*(7), 281–297.

Portha, B., Chavey, A., & Movassat, J. (2011). Early-life origins of type 2 diabetes: Fetal programming of the beta-cell mass. *Experimental Diabetes Research*, *2011*. https://doi.org/10.1155/2011/105076

Scholtens, S., Smidt, N., Swertz, M. A., Bakker, S. J., Dotinga, A., Vonk, J. M., Van Dijk, F., van Zon, S. K., Wijmenga, C., Wolffenbuttel, B. H., & Stolk, R. P. (2015). Cohort profile: Lifelines, a three-generation cohort study and biobank. *International Journal of Epidemiology*, *44*(4), 1172–1180. https://doi.org/10.1093/ije/dyu229

Smits, J. P., De Jong, H., & Van Ark, B. (1999). *Three phases of Dutch economic growth and technological change, 1815-1997*. GGDC Research Memorandum 199942.

Sotomayor, O. (2013). Fetal and infant origins of diabetes and ill health: Evidence from Puerto Rico's 1928 and 1932 hurricanes. *Economics and Human Biology*, *11*(3), 281–293. https://doi.org/10.1016/j.ehb.2012.02.009

Taylor, R. (2019). Calorie restriction for long-term remission of type 2 diabetes. *Clinical Medicine*, *19*(1), 37–42. https://doi.org/10.7861/clinmedicine.19-1-37

van den Berg, G. J., Lindeboom, M., Poplawska, M., & Portrait, F. R. (2017). *Bad potatoes: A sibling analysis of early-life famine exposure and effects on longevity*. Working paper. VU University.

van den Berg, G. J., Paul, A., & Reinhold, S. (2020). Economic conditions and the health of newborns: Evidence from comprehensive register data. *Labour Economics*, *63*, 101795. https://doi.org/10.1016/j.labeco.2020.101795

Van Zanden, J. L. (2005). *The economic history of The Netherlands 1914-1995: A small open economy in the "long" twentieth century* (Vol. 1). Routledge.

Whicher, C., O'Neill, S., & Holt, R. G. (2020). Diabetes in the UK: 2019. *Diabetic Medicine*, *37*(2), 242–247. https://doi.org/10.1111/dme.14225

Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data*. MIT press.

How to cite this article: Alessie, R. J. M., Angelini, V., van den Berg, G. J., Mierau, J. O., & Niccodemi, G. (2024). Using data on biomarkers and siblings to study early-life economic determinants of type-2 diabetes. *Health Economics*, 1–18. https://doi.org/10.1002/hec.4815

APPENDIX A

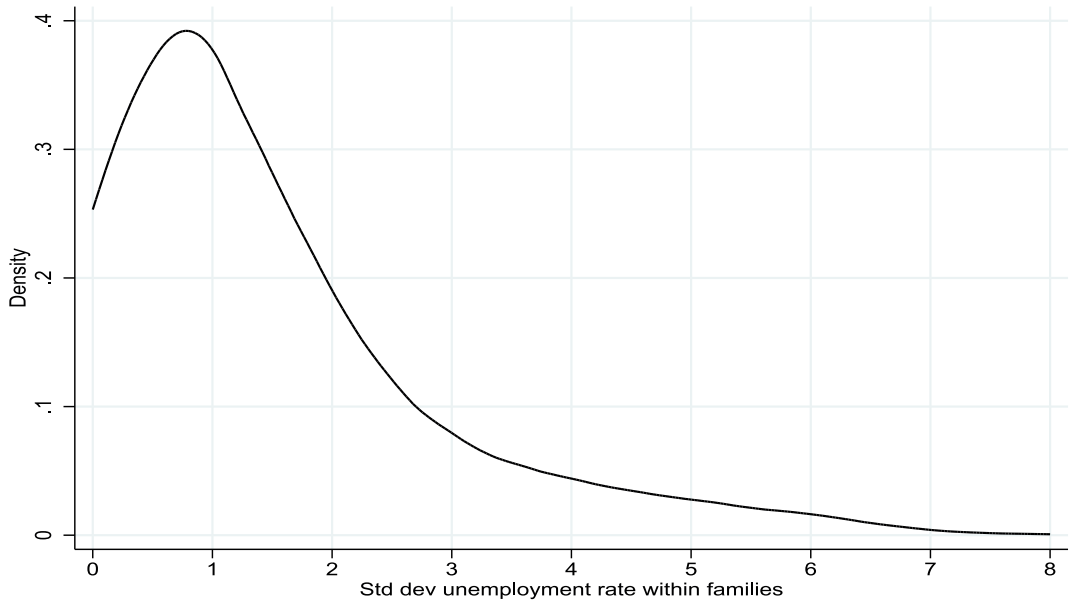

FIGURE A1 Within-family standard deviation of unemployment.

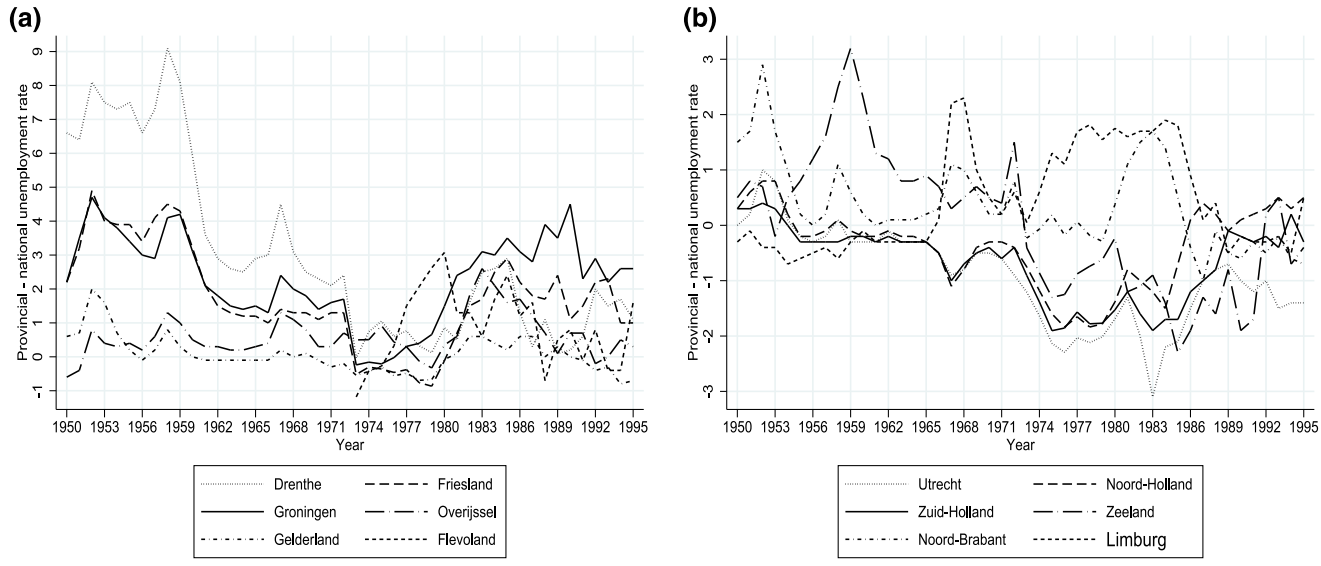

FIGURE A2 Demeaned provincial unemployment rate series. (a) North and East Netherlands. (b) West and South Netherlands.

TABLE A1 Probit estimates on *T2D*.

	Probit	Probit (Chamberlain)
Unem. × female	0.0033	0.0651
	(0.0222)	(0.0401)
Unem. × male	0.0130	0.0878**
	(0.0221)	(0.0408)
Female	−0.1083	−0.0604
	(0.0672)	(0.0940)
Age	0.0277**	0.0204
	(0.0113)	(0.0221)
First born	0.1045***	0.1153**
	(0.0337)	(0.0537)
Province FE	Yes	Yes
Year of birth FE	Yes	Yes
Obs.	43,131	43,131
Average marginal effect Unem. × female	0.0001	0.0023
	(0.0008)	(0.0014)
Average marginal effect Unem. × male	0.0005	0.0031**
	(0.0008)	(0.0015)

Note: *T2D* is a dummy variable equal to 1 if the HbA1c concentration in the blood is equal to or higher than 6.5% or if the fasting glucose level is equal to or higher than 7 mmol/L, and equal to 0 otherwise. Probit (Chamberlain) includes correlated sibling random effects as well as the family mean of the covariates (estimates not reported). Standard errors (in parentheses) are clustered at the family level.
*$p < 0.1$ **$p < 0.05$, ***$p < 0.01$.

TABLE A2 Sibling FE estimates on *T2D*: unemployment age interaction.

	FE	RE
Unem. × female	0.0000	−0.0001
	(0.0030)	(0.0017)
Unem. × male	−0.0013	−0.0017
	(0.0031)	(0.0018)
Unem. × female × age	0.0001	0.0000
	(0.0001)	(0.0000)
Unem. × male × age	0.0001	0.0001
	(0.0001)	(0.0000)
Female	−0.0009	−0.0023
	(0.0037)	(0.0029)
Age	0.0004	0.0008
	(0.0009)	(0.0005)
First born	0.0049***	0.0036***
	(0.0018)	(0.0011)

TABLE A2 (Continued)

	FE	RE
Province FE	Yes	Yes
Year of birth FE	Yes	Yes
Mundlak test *p*-value		0.0494
$\hat{\sigma}^2_\alpha$		0.0006
$\hat{\sigma}^2_\epsilon$		0.0136
Obs.	43,131	43,131

Note: *T2D* is a dummy variable equal to 1 if the HbA1c concentration in the blood is equal to or higher than 6.5% or if the fasting glucose level is equal to or higher than 7 mmol/L, and equal to 0 otherwise. The unemployment variable is the unemployment rate of the province of birth 2 years before birth. Standard errors (in parentheses) are clustered at the family level. α and ϵ are the family-specific component and the idiosyncratic component of the error term, respectively.

*$p < 0.1$ **$p < 0.05$, ***$p < 0.01$.

TABLE A3 Sibling FE estimates on *T2D*: alternative specifications.

	FE	RE
Unem. × female	−0.0003	0.0029*
	(0.0004)	(0.0016)
Unem. × male	−0.0001	0.0032*
	(0.0005)	(0.0016)
Female	−0.0045	−0.0044
	(0.0030)	(0.0030)
Age	−0.0049***	−0.0037
	(0.0010)	(0.0033)
Age^2	0.0001***	0.0001
	(0.0000)	(0.0000)
First born	0.0052***	0.0049***
	(0.0017)	(0.0018)
Province FE	Yes	Yes
Year of birth FE	No	Yes
Obs.	43,131	43,131
Marginal effect of age at age:		
age = 40	0.0006**	0.0004
	(0.0003)	(0.0007)
age = 60	0.0033***	0.0024
	(0.0005)	(0.0020)

Note: *T2D* is a dummy variable equal to 1 if the HbA1c concentration in the blood is equal to or higher than 6.5% or if the fasting glucose level is equal to or higher than 7 mmol/L, and equal to 0 otherwise. Standard errors (in parentheses) are clustered at the family level. Reported marginal effects are for females (effects are slightly higher for males).

*$p < 0.1$ **$p < 0.05$, ***$p < 0.01$.

TABLE A4 Sibling FE and RE estimates on self-reported *T2D*.

	FE	RE
Unem. × female	0.0006	0.0007
	(0.0014)	(0.0007)
Unem. × male	0.0004	0.0007
	(0.0014)	(0.0007)
Female	−0.0025	−0.0018
	(0.0025)	(0.0017)
Age	0.0027***	0.0014***
	(0.0007)	(0.0003)
First born	0.0005	0.0012
	(0.0015)	(0.0009)
Province FE	Yes	Yes
Year of birth FE	Yes	Yes
Mundlak test *p*-value		0.0002
$\hat{\sigma}^2_\alpha$		0.0003
$\hat{\sigma}^2_\epsilon$		0.0088
Obs.	43,131	43,131

Note: Self-reported *T2D* is a dummy variable equal to 1 if the individual self-reports to have type-2 diabetes, and equal to 0 otherwise. Standard errors (in parentheses) are clustered at the family level. α and ϵ are the family-specific component and the idiosyncratic component of the error term, respectively. $^*p < 0.1$ $^{**}p < 0.05$, $^{***}p < 0.01$.

TABLE A5 Means-comparison *t*-test: individuals with unreported and reported type-2 diabetes.

	Unreported *T2D*			Reported *T2D*			
	Obs.	Mean	Std. dev.	Obs.	Mean	Std. dev.	*t*-Test (*p*-value)
Female	338	0.49	0.50	294	0.55	0.50	0.180
Age	338	47.69	9.05	294	51.10	6.64	0.000
Low education	333	0.75	0.43	285	0.82	0.38	0.033

Note: We test the null hypothesis $\mu_x = \mu_z$ against the alternative $\mu_x \neq \mu_z$. The *t*-Test is given by $t = \frac{\bar{x} - \bar{z}}{\left(s_x^2/n_x + s_z^2/n_z\right)^{1/2}}$ and is distributed as a *t*-Student with $t = \frac{\left(s_x^2/n_x + s_z^2/n_z\right)^2}{\frac{\left(s_x^2/n_x\right)^2}{n_x - 1} + \frac{\left(s_z^2/n_z\right)^2}{n_z - 1}}$ degrees of freedom.

TABLE A6 Sibling FE estimates on *T2D*: placebo regressions. Unemployment rate 2 years before birth.

	FE	RE
Unem. × female	−0.0002	0.0001
	(0.0016)	(0.0008)
Unem. × male	0.0002	0.0001
	(0.0015)	(0.0008)
Female	−0.0039	−0.0050**
	(0.0029)	(0.0021)
Age	0.0006	0.0009**
	(0.0008)	(0.0004)

TABLE A6 (Continued)

	FE	RE
First born	0.0040**	0.0031***
	(0.0018)	(0.0011)
Province FE	Yes	Yes
Year of birth FE	Yes	Yes
Mundlak test *p*-value		0.1041
$\hat{\sigma}^2_\alpha$		0.0006
$\hat{\sigma}^2_\epsilon$		0.0128
Obs.	41,756	41,756

Note: *T2D* is a dummy variable equal to 1 if the HbA1c concentration in the blood is equal to or higher than 6.5% or if the fasting glucose level is equal to or higher than 7 mmol/L, and equal to 0 otherwise. The unemployment variable is the unemployment rate of the province of birth 2 years before birth. Standard errors (in parentheses) are clustered at the family level. α and ϵ are the family-specific component and the idiosyncratic component of the error term, respectively.

*$p < 0.1$ **$p < 0.05$, ***$p < 0.01$.

TABLE A7 Sibling FE estimates on *T2D*: placebo regressions. Random assignment of the province of birth to another province.

	FE	RE
Unem. × female	−0.0013	0.0006
	(0.0013)	(0.0007)
Unem. × male	−0.0007	0.0008
	(0.0014)	(0.0008)
Female	−0.0025	−0.0044*
	(0.0031)	(0.0023)
Age	0.0007	0.0009**
	(0.0008)	(0.0004)
First born	0.0049***	0.0036***
	(0.0018)	(0.0011)
Province FE	Yes	Yes
Year of birth FE	Yes	Yes
Mundlak test *p*-value		0.0412
$\hat{\sigma}^2_\alpha$		0.0006
$\hat{\sigma}^2_\epsilon$		0.0136
Obs.	43,095	43,095

Note: *T2D* is a dummy variable equal to 1 if the HbA1c concentration in the blood is equal to or higher than 6.5% or if the fasting glucose level is equal to or higher than 7 mmol/L, and equal to 0 otherwise. The unemployment variable is the unemployment rate of a randomly reassigned province of birth, where all individuals born in a specific province were assigned the same 'new' province. Standard errors (in parentheses) are clustered at the family level. α and ϵ are the family-specific component and the idiosyncratic component of the error term, respectively.

*$p < 0.1$ **$p < 0.05$, ***$p < 0.01$.

TABLE A8 Sibling FE estimates on *T2D*: placebo regressions. Random assignment (at the year level) of the year of birth to another year.

	FE	RE
Unem. × female	0.0011	0.0008
	(0.0008)	(0.0006)
Unem. × male	0.0002	0.0000
	(0.0008)	(0.0006)
Female	−0.0112***	−0.0104***
	(0.0036)	(0.0028)
Age	0.0008	0.0009**
	(0.0008)	(0.0004)
First born	0.0048**	0.0036***
	(0.0018)	(0.0011)
Province FE	Yes	Yes
Year of birth FE	Yes	Yes
Mundlak test *p*-value		0.0600
$\hat{\sigma}^2_\alpha$		0.0006
$\hat{\sigma}^2_\epsilon$		0.0137
Obs.	43,074	43,074

Note: *T2D* is a dummy variable equal to 1 if the HbA1c concentration in the blood is equal to or higher than 6.5% or if the fasting glucose level is equal to or higher than 7 mmol/L, and equal to 0 otherwise. The unemployment variable is the unemployment rate in a randomly assigned year of birth, where all individuals born in a specific year were assigned the same 'new' year. Standard errors (in parentheses) are clustered at the family level. α and ϵ are the family-specific component and the idiosyncratic component of the error term, respectively.

*$p < 0.1$ **$p < 0.05$, ***$p < 0.01$.

TABLE A9 Sibling FE estimates on *T2D*: placebo regressions. Random assignment of the year of birth (at the individual level).

	FE	RE
Unem. × female	0.0011	0.0002
	(0.0009)	(0.0007)
Unem. × male	0.0009	0.0002
	(0.0004)	(0.0003)
Female	−0.0066*	−0.0053**
	(0.0036)	(0.0027)
Age	0.0013***	0.0010***
	(0.0003)	(0.0001)
First born	0.0038**	0.0049***
	(0.0016)	(0.0011)
Province FE	Yes	Yes
Year of birth FE	Yes	Yes
Mundlak test *p*-value		0.3098
$\hat{\sigma}^2_\alpha$		0.0006
$\hat{\sigma}^2_\epsilon$		0.0137
Obs.	43,063	43,063

Note: *T2D* is a dummy variable equal to 1 if the HbA1c concentration in the blood is equal to or higher than 6.5% or if the fasting glucose level is equal to or higher than 7 mmol/L, and equal to 0 otherwise. The unemployment variable is the unemployment rate in a randomly assigned year of birth, where the random assignment is at the individual level. Standard errors (in parentheses) are clustered at the family level. α and ϵ are the family-specific component and the idiosyncratic component of the error term, respectively.

*$p < 0.1$ **$p < 0.05$, ***$p < 0.01$.