BUDAPEST WORKING PAPERS
ON THE LABOUR MARKET
BWP – 2017/12

Health Differences at Birth between Roma and Non-Roma Children in Hungary

Long-Run Trends and Decompositions

TAMÁS HAJDU – GÁBOR KERTESI – GÁBOR KÉZDI

BWP 2017/12

INSTITUTE OF ECONOMICS, CENTRE FOR ECONOMIC AND REGIONAL STUDIES
HUNGARIAN ACADEMY OF SCIENCES
BUDAPEST, 2017

Budapest Working Papers on the Labour Market

BWP – 2017/12

Institute of Economics, Centre for Economic and Regional Studies,
Hungarian Academy of Sciences

Health Differences at Birth between Roma and Non-Roma Children in Hungary
Long-Run Trends and Decompositions

Authors:

Tamás Hajdu
research fellow
Institute of Economics, Center for Economic and Regional Studies
of the Hungarian Academy of Sciences
hajdu.tamas@krtk.mta.hu

Gábor Kertesi
senior research fellow
Institute of Economics, Center for Economic and Regional Studies
of the Hungarian Academy of Sciences
kertesi.gabor@krtk.mta.hu

Gábor Kézdi
Survey Research Center, University of Michigan
and
senior research fellow
Institute of Economics, Center for Economic and Regional Studies
of the Hungarian Academy of Sciences
kezdi@umich.edu

November 2017

ISBN 978-615-5754-24-1
ISSN 1785 3788

Health Differences at Birth between Roma and Non-Roma Children in Hungary

Long-Run Trends and Decompositions

Tamás Hajdu – Gábor Kertesi – Gábor Kézdi

Abstract

This paper uses birth records linked to census data to document health differences at birth between Roma and non-Roma children in Hungary between 1981 and 2010. It focuses on differences in average birth weight and average gestational age, as well as the likelihood of low birth weight and the likelihood of preterm birth. The paper shows large gaps in all indicators over the 30 years, with a small narrowing of the gap in absolute terms but not in relative terms. Roma mothers are twice as likely to give birth to babies with low birth weight and before the 37^{th} week. Standard decompositions show that around 80% of the gap is explained by socioeconomic factors, and education alone explains more than half. Despite significant changes in society, the explanatory power of education and other factors remains constant. Narrowing the gap in educational attainment, especially at higher levels, may have the highest potential to improve the relative health of Roma births.

Keywords: health at birth, birth weight, preterm birth, minorities

JEL classification: I14, J15

Acknowledgement:

We thank Zsolt Németh (KSH) and dr. Miklós Szabó (SOTE Department of Pediatrics I) for their thoughtful comments. Bence Szabó provided excellent research assistantship.
The authors received the following funding to carry out the research summarized in this article: Nemzeti Kutatási, Fejlesztési és Innovációs Hivatal (National Research, Development and Innovation Office of Hungary) research grant NKFI-116354. This work was supported by the Horizon 2020 Twinning project EdEN (grant no. 691676) as well.

Roma és nem roma újszülöttek születéskori fejlődési különbségei Magyarországon, 1981-2010

Hajdu Tamás – Kertesi Gábor – Kézdi Gábor

Összefoglaló

Tanulmányunk a kelet-közép-európai régióban hozzáférhető legrészletesebb roma etnikai adatokra (a 2011. évi népszámlálás nemzetiségi adataira), illetve népszámlálással egyéni szinten összekapcsolt adminisztratív adatokra (a KSH élveszületési regiszterére) támaszkodva, korábban nem tapasztalható részletességgel és mélységben vizsgálja a roma népesség születéskori egészségi helyzetét. Négy születéskori fejlődési mutatót vizsgálunk meg részletesebben: az átlagos születési testsúlyt, a születés átlagos hetét, valamint az alacsony (2500 gram alatti) születési testsúly, illetve a koraszülés (a 37. gesztációs hét előtti születés) gyakoriságát és ezen indikátorok hosszú távú alakulását. Ez a beszámoló az első kísérlet arra, hogy egy ország roma népességének születéskori egészségi állapotáról és annak hosszú távú trendjeiről teljes körű adminisztratív adatok alapján alkothassunk képet. A mért etnikai különbségek valamennyi évben igen nagyok. A különbségek abszolút értelemben kis mértékben csökkentek ugyan 1981 és 2010 között, relatív értelemben azonban szinte semmit sem változtak: a roma anyák, a nem roma anyákhoz viszonyítva, kétszer akkora eséllyel adtak életet koraszülötteknek, illetve kis súlyú újszülötteknek a vizsgált harminc éves időszak egészében. E különbségek 80 százalékát ismert társadalmi-gazdasági tényezők magyarázzák, közülük egyedül a szülők iskolázottsága több mint 50 százalékot. Noha e három évtized alatt óriási társadalmi változások mentek végbe, a születéskori etnikai különbségeket meghatározó összefüggések azonban, súlyukat és jelentőségüket tekintve, időben meglepően állandónak bizonyultak. A legnagyobb lehetőség a változásra az új roma generációk iskolai végzettségének növelésében, különösen az érettségizett és diplomás roma fiatalok számának növelésében rejlik.

Tárgyszavak: születéskori egészség, születési testsúly, koraszülés, roma újszülöttek

JEL kódok: I14, J15

1. INTRODUCTION

Health at birth predicts important outcomes in later in life (Aizer & Currie, 2014; Almond & Currie, 2011; Currie, 2009, 2011). Birth weight, the most widely used measure, is a strong predictor of infant mortality (Elder, Goddeeris, & Haider, 2011, 2016), childhood, adolescent and adult health (Case, Fertig, & Paxson, 2005; Case, Lubotsky, & Paxson, 2002; Hack, Klein, & Taylor, 1995; Palloni, Milesi, White, & Turner, 2009), as well as educational and labor market outcomes (J. R. Behrman & Rosenzweig, 2004; Bharadwaj, Lundborg, & Rooth, 2017; Black, Devereux, & Salvanes, 2007; Breslau et al., 1994; Currie & Hyson, 1999; Figlio, Guryan, Karbownik, & Roth, 2014; Oreopoulos, Stabile, Walld, & Roos, 2008; Reichman, 2005). Infant health is often strongly associated with early delivery. As a result, preterm birth, defined as delivery before the 37th week, is another important indicator of health at birth (R. E. Behrman & Butler, 2007; Howson, Kinney, & Lawn, 2012; Lawn, Cousens, & Zupan, 2005).

It is important to quantify and understand the disadvantages of ethnic and racial minorities in infant health, both in their own right and because they may predict disadvantages in adulthood. Particularly interesting questions include the extent to which such differences may be due to poverty, education, or geographic isolation. We know more about these questions for some minorities than for others. The gap in health at birth between Whites and African Americans and Hispanics in the U.S. is well documented and analyzed (Costa, 2004; Lhila & Long, 2012; Martin, Hamilton, Osterman, Driscoll, & Mathews, 2017; Morisaki, Kawachi, Oken, & Fujiwara, 2017). Substantially less is known about the Roma of Central and Eastern Europe, even though the Roma are one of the largest and poorest ethnic minorities in Europe (FRA & UNDP, 2012). Most studies on the Roma analyze small samples, in selected regions within countries, and as single cross sections (Balázs, Rákóczi, Grenczer, & Foley, 2013; see, for example Balázs, Rákôczi, Grenczer, & Foley, 2014; Bobak, Dejmek, Solansky, & Sram, 2005; Joubert, 1991; Stanković et al., 2016). The likely reason is lack of appropriate data. Birth records in most European countries do not have ethnic markers, making comprehensive analysis difficult.

In this paper we use comprehensive data from birth records in Hungary that we linked to census data to obtain ethnic markers. Hungary has a sizeable Roma minority and a rather homogeneous non-Roma majority. We consider four indicators: birth weight, the incidence of low birth weight (< 2500 g), gestational age (length of pregnancy), and the incidence of preterm birth (< 37 weeks). First, we show trends in differences in these indicators between Roma and non-Roma Hungarians from 1981 to 2010. Then we decompose the ethnic differences into a part explained by parental socioeconomic characteristics and a residual part. We carry out the decomposition in a fully comparable fashion over the 30 years spanned by our data.

Our contribution is threefold. This is the first paper to measure the health gap at birth between the children of the Roma minority and the non-Roma majority for an entire country. Second, we present trends spanning 30 years of large social changes, from communism through to membership in the European Union. Third, based on decomposing the gap into a part attributable to observed parental characteristics and a residual part over the entire 30-year period, we present trends in the two parts.

We find that the gap narrows for most indicators in absolute terms, e.g., from 10 to 8 percentage points for low birth weight and from 7 to 5 percentage points for preterm birth. However, in relative terms the gap remains very similar: Roma births are shown to be more than twice as likely to be of low weight and slightly less than twice as likely to be preterm over the entire 30-year period.

Socio-economic differences explain the majority of this gap, between 77% and 98% across indicators. The bulk of the differences explained are due to differences in education, the second most important factor being the presence of the father. Remarkably, differences in geography play a negligible role, even though the Roma are substantially more likely to live in rural areas and the poorest counties. Our decomposition results suggest that the narrowing of the ethnic gap is due to the narrowing of the gap in educational attainment at the lowest end of the educational distribution.

2. DATA

We linked two administrative datasets for this analysis: birth records and the census of 2011. The birth records of Hungary contain all live births, with content harmonized since 1981. The records contain information on the date of birth, birth weight, and gestational age of the newborn babies, as well as the date of birth, level of education, employment, and residence of both the mother and the father. Birth records do not contain ethnic markers. The census of 2011 identifies the ethnic identity of all adult respondents; we discuss the quality of ethnic markers below. Birth records and census data are maintained by the Kozponti Statisztikai Hivatal (KSH, Central Statistical Office of Hungary). We accessed and linked the datasets in the secure data environment of KSH. In line with the literature, we excluded all twin births and birth records with missing key information (less than 4% of all births in total).

We linked the records of single births to the census of 2011 to identify the ethnicity of the mother. This linkage was feasible for children and mothers alive in September 2011, the month of the census count. As a result, our time series ends in 2010, the last complete year covered. The first year of our time series is 1981. Appendix 1 in the Supporting Information contains more detail on the linking procedure and the quality of the linking. The proportion of linked records is high: 90% of live births after 1995 are successfully linked, and the success rate is still around 60% in 1981. Systematic differences between linked and not linked births are not very great. Linked births appear somewhat healthier over the time period. Remarkably, however, the differences are quite stable across time, despite the trend in the fraction of successful links. This suggests that the additional births not linked in earlier years are selected by factors that are largely unrelated to outcomes.

Ethnicity in the Hungarian census of 2011 was measured by two questions of every adult respondent, thus allowing for the reporting of multiple identities. The total number of Roma marked this way in the census of 2011 is 315,583, or 3.2% of the population. This is around 60% as large as other estimates of the size of the Roma minority in Hungary (Kemény & Janky, 2006)). Using data from a survey that represents the entire Roma population we find that the Roma populations represented by the two samples are very close. Since the proportion of unidentified Roma in the census is low relative to the entire non-Roma population their presence does not introduce significant bias in the non-Roma results. Thus, we conclude that while there is substantial selection into successful linkage, as well as Roma identification in the census, they are unlikely to induce substantial biases in our estimates of the ethnic gaps in birth outcomes.

3. THIRTY-YEAR TRENDS

Figures 1–4 show the 30-year trends in average birth weight, proportion with low birth weight, average gestation age, and proportion of preterm births. The values are calculated for calendar years of the births. Solid lines show values for the non-Roma and dashed lines show values for the Roma. We draw 95% confidence intervals around each line, reflecting the notion that linked births are a sample of all births, and actual births may also be considered a sample of potential births. The confidence intervals are narrow for the non-Roma figures, so they are hard to detect on the graphs; the confidence intervals for the Roma figures are wider.

As shown in Figure 1, average birth weight among non-Roma children increases from 3200 g in 1981 to over 3300 g by 2008, and levels off to 2010. Roma children are born with substantially lower birth weight, on average, throughout the entire time period, starting at 2900 g in 1981 and reaching slightly over 3000 g by 2005. The gap remains relatively stable, at 300 g, through the entire period. This gap is comparable to the approximately 250 g difference currently measured between African American and non-Hispanic White children in the U.S. (Morisaki et al., 2017). While the gap is similar in magnitude, the levels are lower by 100 g.

Figure 1

Trends in average birth weight

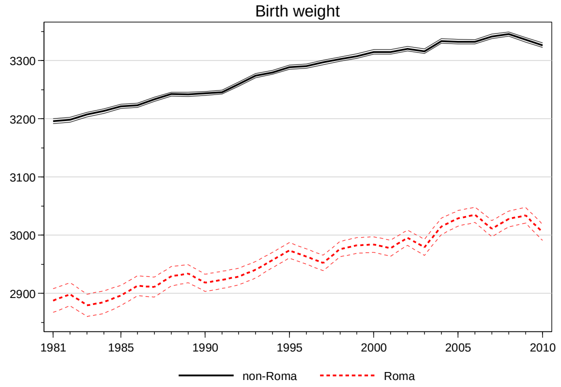


Figure 2

Trends in the fraction of low birth weights

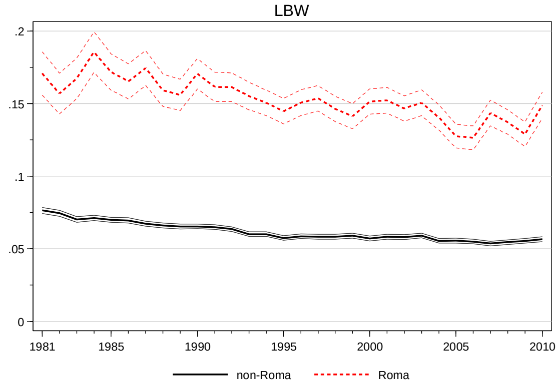


Figure 2 shows that the proportion of newborn infants with low birth weight (less than 2500 g) decreases from 7.5% to 5.5% among the non-Roma, and from around 17% to 14% among the Roma. For both groups the decrease is more pronounced before 1995. In absolute terms, the gap decreases from 10 percentage points to 8 percentage points. In relative terms the gap does not decrease: the Roma incidence remains slightly more than twice as high.

Our estimate of the gap in the likelihood of low birth weight between Roma and non-Roma births is in the middle of the wide range of estimates in non-representative samples for the respective time periods. Bobak et al. (2005), and Pongracz and S. Molnar (1994) find larger gaps in two Czech districts and in a non-representative Hungarian sample; Balazs et al. (2013) find a substantially smaller gap in two Hungarian counties; Koval et al. (2012) find a very similar gap and similar levels in a Slovak region.

Our estimate of the gap in low birth weight (14% versus 5.5%) is somewhat wider than the corresponding difference between African American and White children in the U.S. in 2010, (13.5% versus 7%) (Martin et al., 2017). The incidence of low birth weight is similar among Roma and African American children, but it is higher among White Americans than non-Roma Hungarians (7% versus 5.5%).

Average gestational age is constant for non-Roma Hungarians at 39 weeks until around 2000 and declines to 38.8 weeks by 2010. The average is shorter, at around 38.3 weeks, among Roma births through the entire period, with no clear trends.

Figure 3

Trends in average gestation age

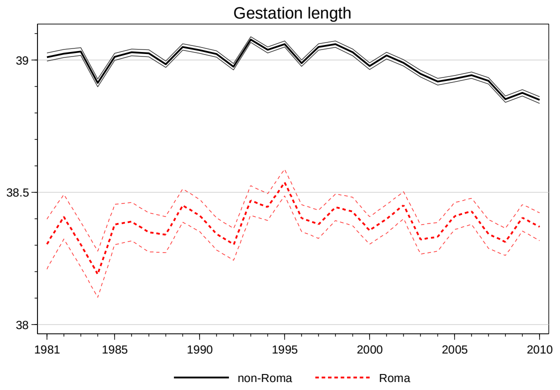
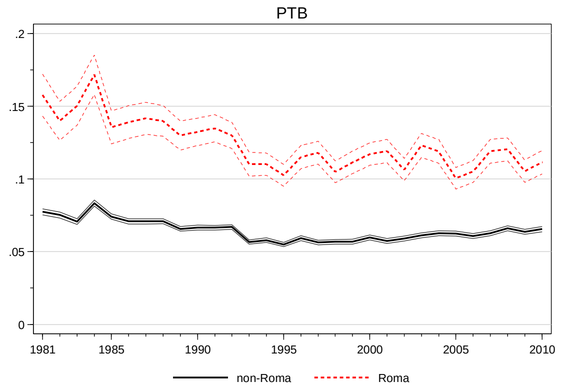


Figure 4

Trends in the fraction of preterm births



The incidence of preterm birth, defined as birth before the 37th week, decreases until around 1995 for both Roma and non-Roma, but levels off afterwards, with a slight increase among non-Roma births. The Roma figures start with 15% of births being preterm in the early 1980s, compared to 7.5% for the non-Roma. By 1995, the Roma incidence of preterm births decreases to 11%, while the non-Roma incidence decreases to 5.5%. The Roma figures do not change considerably following this, while the non-Roma figures increase to 6.5%. As a result, the ethnic gap in preterm births decreases from over 7 percentage points in the 1980s to below 5 percentage points by 2010. Again, the relative gap does not decrease near as much: Roma mothers are around twice as likely to have a preterm birth as non-Roma mothers throughout the entire time period.

Our estimate of the gap in preterm births between Roma and non-Roma newborns is again in the middle of the estimates in the literature. Similar to their low birth weight gap estimates, Bobak et al. (2005) and Pongracz and S. Molnar (1994) find larger gaps than us, but Balazs et al. (2013) find a gap very similar to our estimate. The gap in preterm births between Roma and non-Roma in Hungary is similar to the Black–White gap in the U.S., but the Hungarian levels are lower. In the U.S., 13% of African American births are preterm, compared to 9% of White births (Martin et al., 2017), compared with the 11% and 6.5% rates in Hungary.

Figures A3–A6 in the Supporting Information show the corresponding results using inverse probability weights to account for selection on observable characteristics into successful linkage of birth records to census records. The results suggest that our estimates of the gaps and their time trends are robust to this kind of selection.

Summarizing the results, Roma children have substantially worse health at birth in Hungary than their non-Roma peers. They are more than twice as likely to have low birth weight (13% versus 5.5%), and somewhat less than twice as likely to be born preterm (11% versus 6.5%). These differences are comparable to the differences between African American and White babies in the U.S. The 30 years see an improvement in the health indicators both among the Roma and non-Roma, and the gap narrows in absolute terms but not in relative terms.

4. GAP DECOMPOSITION

We decomposed the gap between the health indicators of Roma and non-Roma newborn babies using observable characteristics of the parents in the birth records. Besides the ethnicity of the parents, which is quite a stable personal characteristic, we did not use any information from the census. The other variables used in our analysis are level of education of the mother (five categories), her employment status before delivery (five categories), her age (categories for five-year age groups), a generated indicator variable for marital status (whether information is available on the father, and if yes, whether they are married), education and employment of the father if known, the mother's history of pregnancies (number of previous abortions, fetal losses, and live births), as well as geography of residence (interaction of county and whether in Budapest, other large city, small town, or rural area).

In this section, we present the results from the most recent years, combining data from 2008, 2009, and 2010. As the variables in the birth records data are fully harmonized, we also carried out analogous decompositions for earlier years. Those results will be presented in the next section.

Table A3 in the Supporting Information presents summary statistics for the Roma and non-Roma subsamples, and the differences in mean values. The differences are substantial. Roma mothers give birth at a younger age: the modal age group is 20–24 versus 30–34 among non-Roma mothers; 27% of the Roma mothers are 19 years old or younger, compared to 4% among non-Roma mothers. As a mirror image, 19% of Roma mothers are over 30 years of age, compared to 53% among the non-Roma. Roma mothers are less educated: 85% of Roma mothers had 8 years of education or less at the time of the birth of their child, compared to 14% among non-Roma mothers. The corresponding differences among fathers with known education are similar, 70% versus 10%. The prevalence of employment among Roma mothers and fathers is 12% and 46% respectively; the corresponding figures for non-Roma are 70% and 91%. Information on the father is missing for 40% of Roma births, but only 7% of non-Roma births. Only 26% of Roma mothers are reported to be married, compared to 64% of non-Roma mothers. Roma mothers have previously had more live births and more abortions. The Roma are more rural (55% versus 29%), and they are more likely to live in counties with higher levels of unemployment (Heves, Szabolcs, Borsod).

We carried out decompositions in multiple ways. In the main text we present results from the simplest approach: pooled linear regressions. We pooled the Roma and non-Roma subsamples and estimated two kinds of linear regression models: one with the Roma mother indicator variable on the right-hand side only, and one adding all the covariates. The coefficient on the Roma mother indicator variable in the first type of regressions shows the *raw gap*: the difference between average Roma and non-Roma births. The coefficient on the Roma indicator variable when all covariates enter approximates the *residual gap*: the estimated difference between Roma and non-Roma births with parents sharing the same observable characteristics.

Table 1 presents our estimates of the residual gaps from the pooled linear regressions, together with the corresponding raw gaps.

Table 1.

Raw gaps (A) and residual gaps (B) in the health indicators of births to Roma mothers versus non-Roma mothers

		(1) Birth weight	(2) Low birth weight	(3) Gestation age	(4) Preterm birth
(A)	Roma mother	−313***	0.083***	−0.498***	0.047***
		(4.2)	(0.003)	(0.016)	(0.002)
	Controls	No	No	No	No
	Adj. R-squared	0.02	0.01	0.01	0.00
	N	249,200	249,200	249,159	249,159
(B)	Roma mother	−72***	0.015***	−0.083***	0.001
		(4.9)	(0.003)	(0.018)	(0.003)
	Controls	Yes	Yes	Yes	Yes
	Adj. R-squared	0.09	0.04	0.04	0.02
	N	249,200	249,200	249,159	249,159
	Non-Roma average	3336	0.056	38.86	0.065

Notes. Coefficient estimates on the Roma indicator variable from OLS regressions. Live births in Hungary in 2008–2010 matched to the 2011 census to obtain ethnic markers. Controls: gender of the newborn child, month of delivery, whether information on father is missing; marital status of the mother if father is known; age, education, labor force status of mother and father; number of previous abortions, miscarriages, and live births, county of residence interacted with settlement type of residence (Budapest, large city, small town, rural).
Robust standard errors are in parentheses. * p<.10, ** p<.05, *** p<.01

In 2008–2010, Roma mothers gave birth to babies 300 g lighter than non-Roma mothers, and they were 8.3 percentage points more likely to give birth to babies with low birth weight. Our results suggest that around 80% of these differences can be explained by the few demographic and socioeconomic variables observed in our data. Children born to Roma mothers are only 72 g lighter on average, and they are only 1.5 percentage points more likely to have low birth weight than children born to non-Roma mothers with the same demographic and socioeconomic characteristics. This is all the more remarkable as the R-squared of the regression with all of the variables is only 4%. The incidence of low birth weight is primarily determined by factors other than the observed demographic and socioeconomic variables, but the ethnicity of the mother is not strongly correlated with those other, unobservable factors.

The results are even more striking for average gestational age and the incidence of preterm birth. In 2008–2010 Roma mothers gave birth half a week earlier than non-Roma mothers, and they were 4.7 percentage points more likely to have a preterm birth. More than 80% of the average difference is explained by the few observable characteristics in our data, and 98% of the difference in the incidence of preterm birth. As a result, Roma mothers with the same demographic and socio-economic characteristics as their non-Roma peers have the same propensity to give preterm birth. The control variables explain, in the regression sense, the entire ethnic gap, even though their overall explanatory power is also small here (R-squared value of 4% and 2%).

We carried out extensive robustness checks using more sophisticated models (the results are summarized in Table A4 in the Supporting Information). First, we re-estimated the regressions with inverse probability weights that account for selection on observable characteristics into successful linking of birth records to the census. The results are very similar to the unweighted estimates. Second, we replaced the linear regression models with logit models for the binary outcome variables. The marginal effect estimates from these models are virtually identical to the linear regression models. Third, we carried out Oaxaca–Blinder decompositions, using the non-Roma coefficients on the covariates to remove composition effects. Fourth, we carried out Oaxaca–Blinder decompositions of the binary outcome variables using logit models (Fairlie, 2005), similar to the Black–White gap analysis of Lhila and Long (2012). Fifth, we estimated the gap between Roma and non-Roma births after propensity score matching on the observed characteristics. All of these models give results very similar to our simple estimates.

Using the results of the Oaxaca–Blinder decompositions, we can characterize the explanatory power of subsets of the control variables. Table 2 shows the fraction of the raw gap each group of control variables explains, in percentage terms, the raw gaps being set to 100%. Table A5 in the Supporting Information shows that the corresponding results from logit-based decompositions for the two binary outcomes are very similar to the linear estimates.

Table 2.

The share of parental characteristics in explaining the raw gap

	Birth weight	Low birth weight	Gestation age	Preterm birth
Education	53%	52%	49%	60%
Labor force status	10%	13%	12%	15%
Pregnancy history	−3%	2%	9%	8%
Geography	4%	3%	6%	4%
Marital status	18%	30%	29%	60%
Age	−3%	−17%	−18%	−46%
Other	0%	0%	0%	0%
Residual gap	21%	17%	13%	−1%
Total	100%	100%	100%	100%

Differences in education explain by far the largest part of the ethnic gap in health at birth. Indeed, they explain 50% or more of the entire gaps, with 60% for preterm birth. Another 10–15% is explained by differences in labor force status, mostly driven by differences in employment. In addition, 20–60% is explained by marital status, the main difference being due to whether information about the father is available. Roma mothers give birth at a younger age, on average, which is associated with lower incidence of low birth weight and especially preterm birth. Without the age differences, the ethnic gap would be even larger. With the exception of teen pregnancy, younger age at delivery is associated with better birth outcomes in general, and in particular leads to a substantially lower incidence of preterm birth (R. E. Behrman & Butler, 2007, pp. 125–127), in line with our results. Other characteristics, including county and rural/urban nature of residence, do not matter at all for the gaps. The fact that geographic differences do not affect the ethnic gap in health at birth is remarkable, given the large ethnic differences in geography.

5. TRENDS IN THE COMPONENTS OF THE GAPS

We estimated the residual gap in low birth weights and preterm births for five-year periods between 1981 and 2010. Figures 5–8 show the estimates of the residual gaps (dashed lines), together with the raw gap estimates (continuous lines). The figures also show the corresponding 95% confidence intervals.

Figures 5–8.

Trends in ethnic gaps: raw gaps (continuous lines) and residual gaps (dashed lines)

Figure 5

Average birth weight

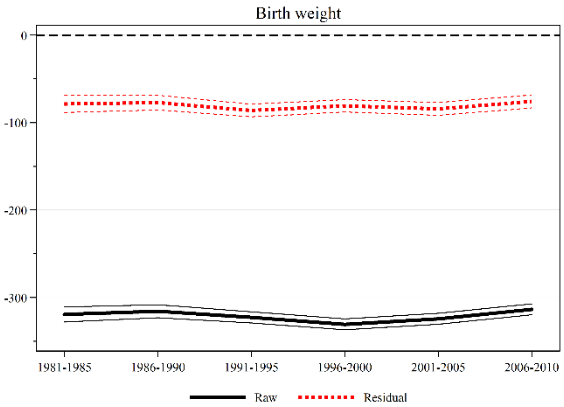


Figure 6

Likelihood of low birth weight

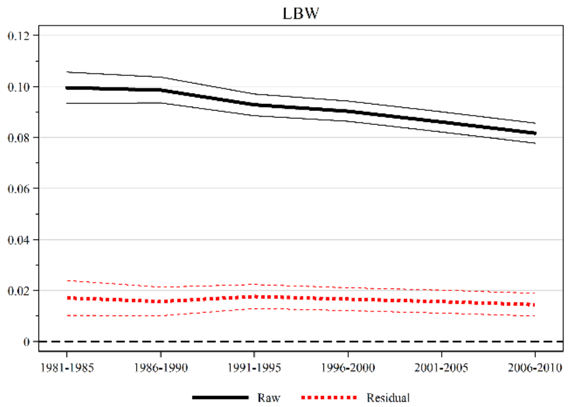


Figure 7

Average gestation age

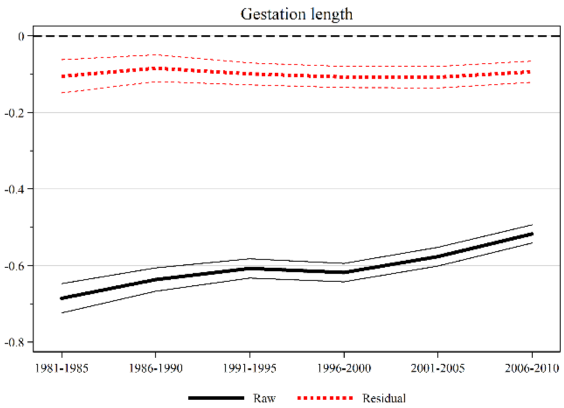
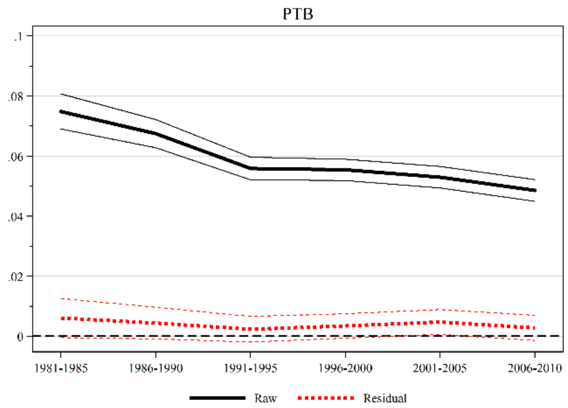


Figure 8

Likelihood of preterm births



As we have seen when interpreting the raw trends (Figures 1–4), the ethnic gaps shrink somewhat in terms of the incidence of low birth weight and preterm birth, as well as average gestation age. At the same time, our results show that the residual gaps remain constant for all indicators. Moreover, the explanatory power of specific observable characteristics remains the same through the entire 30-year period (Table A6 in the Supporting Information shows the results for five-year time periods starting in 1981).

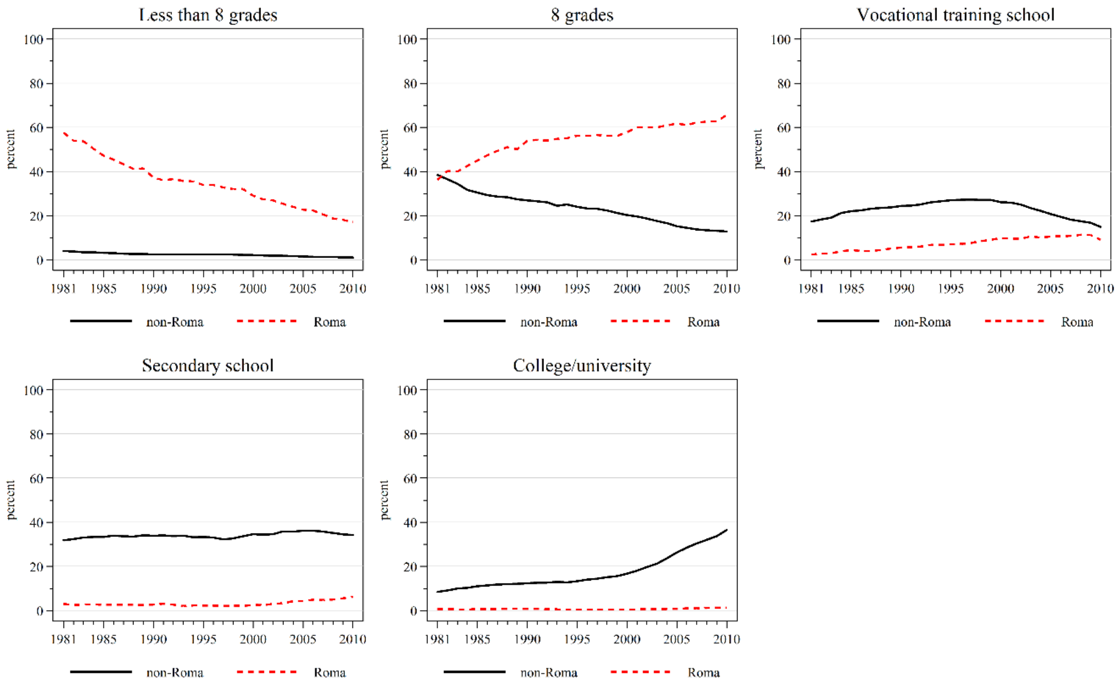
Taken together, the results of our analysis suggest that the decreases observed in the absolute gaps in the health indicators are due to decreases in the differences in observed parental characteristics between Roma and non-Roma parents. Parental education plays by far the larger role in explaining the gaps, and Hungary experienced major changes in educational attainment in the 30-year time period we examined. Thus, it is likely that trends in education may explain the trends in the observed gaps.

Figure 9 shows the trends in educational attainment of Roma and non-Roma mothers in our sample. The fraction of Roma mothers with 7 years of education or less decreases dramatically, from 60% in 1981 to 20% in 2010, while the corresponding figures for non-Roma are 4% and 1%. This decrease of 40 percentage points in the fraction of uneducated Roma mothers is mirrored by an increase of 20 percentage points among mothers with 8 grades of education, and another 20% increase for vocational and secondary education combined. The fraction of Roma mothers with a college education remains negligible, while the fraction of non-Roma mothers with college education increases substantially, from 5% to almost 40%. To some extent, these trends are due to the increasing average age of the mothers at the birth of the child, but to a large extent they are due to genuine expansions in educational attainment.

To sum up the conclusions from these graphs, the ethnic gap in educational attainment narrows substantially at the lowest end of the educational distribution, but increases significantly at the highest end. It appears, therefore, that the modest decrease in the ethnic gap in health at birth is due to the large decrease in the educational gap at the lower end of the distribution.

Figure 9.

Trends in educational attainment of Roma (dashed line) and non-Roma (continuous line) mothers



6. DISCUSSION AND CONCLUSIONS

Our study is the first to use national data going back 30 years to examine the gap in health indicators between Roma and non-Roma births. We focused on Hungary and used birth records linked to census data to obtain ethnic markers. The indicators we examined were birth weight, incidence of low birth weight, age of gestation, and incidence of preterm birth. Our results suggest large differences in Hungary. These differences decreased somewhat in the 30 years between 1981 and 2010 in absolute terms, but not in relative terms. Roma children were more than twice as likely to have low birth weight, and almost twice as likely to be preterm births in 2010 just like in 1981. We showed that over 80% of the gap in low birth weight, and the entire gap in preterm birth, is explained by differences in the level of demographic and socio-economic characteristics of the parents. Of these, differences in education and the presence of the father are the most important. Ethnic differences in geography, while substantial in themselves, explain nothing of the ethnic gap in health at birth. The role of these factors in explaining the gaps in health at birth remained remarkably constant between 1981 and 2010. Trends in educational attainment by ethnicity suggest that narrowing of the gap at the lowest levels of education may be responsible for the observed narrowing of the gaps in health at birth.

The findings that observable parental characteristics explain most of the gaps in health at birth, and that education plays the most important role, are in line with corresponding estimates of gaps between Roma and non-Roma in Hungarian regions (Balázs et al., 2013) and Czech regions (Bobak et al., 2005). These results are at odds with what the literature found in the U.S. between African Americans and Whites, where the residual gaps are still around half the size of raw gaps (Lhila & Long, 2012; Morisaki et al., 2017). This contrast turns out to be analogous to what we see in terms of gaps in test scores of students. While the raw gaps are at least as large in Hungary as in the U.S., a substantially larger part of this is explained by parental characteristics (Kertesi & Kézdi, 2011). We can only speculate on the reasons. It may be related to the fact that according to many indicators, such as education, employment, and unemployment, the relative socioeconomic disadvantage of the Hungarian Roma is far larger than that of African Americans in the U.S.

The relative stability of ethnic differences in health at birth and the stability of the factors explaining those differences are remarkable. The 30 years analyzed saw transition from communism to capitalism, accompanied by large recessions and rapid economic growth, and large changes to the welfare system in Hungary. They also experienced a substantial expansion in education, with the largest increase of completion of secondary school and college among the non-Roma, and the largest increase of completion of 8 grades among the Roma. Our results suggest that despite all the large shifts in society, the drivers of the relative position of the Roma minority changed little.

In particular, the large ethnic gap in educational attainment and its slight improvement may explain why the gap in health at birth remained large. The results of our regressions suggest that closing the educational gap may lead to a 50–60% drop in the gaps in the health indicators, holding all other factors constant. Further calculations using the same results suggest a 25% drop if, instead of a full closing of the gap, only the fraction of Roma with 0–8 grades of education were to be reduced to the corresponding non-Roma fraction (and the previously less-educated Roma all completed vocational training instead). Of course, many of those other factors, such as employment, age at birth, and marital status, are likely to be affected by changes in educational attainment in complex ways. Many of those additional effects may lead to further reductions in the gaps, while others may work in the other direction. Nevertheless, improved education is likely to hold high potential for narrowing the ethnic gap in health at birth.

The Roma made substantial progress in closing the educational gap at the lower end over the period, leaving little room to make further progress there in the future. At the same time, their disadvantage in terms of secondary and higher education increased considerably as they missed much of the massive educational expansion experienced by the non-Roma. Further progress is possible only through increased participation in secondary and tertiary education. The continued exclusion of the Roma minority from the expansion of education at higher levels would likely conserve many of their huge disadvantages. Among these, the Roma would continue to give birth to less healthy babies, who thus start their lives already disadvantaged.

REFERENCES

Aizer, A., & Currie, J. (2014). The intergenerational transmission of inequality: Maternal disadvantage and health at birth. *Science*, *344*(6186), 856–861. https://doi.org/10.1126/science.1251872

Almond, D., & Currie, J. (2011). Killing Me Softly: The Fetal Origins Hypothesis. *Journal of Economic Perspectives*, *25*(3), 153–172. https://doi.org/10.1257/jep.25.3.153

Balázs, P., Rákóczi, I., Grenczer, A., & Foley, K. L. (2013). Risk factors of preterm birth and low birth weight babies among Roma and non-Roma mothers: a population-based study. *European Journal of Public Health*, *23*(3), 480–485. https://doi.org/10.1093/eurpub/cks089

Balázs, P., Rákôczi, I., Grenczer, A., & Foley, K. L. (2014). Birth-Weight Differences of Roma and Non-Roma Neonates - Public Health Implications from a Population-Based Study in Hungary. *Central European Journal of Public Health*, *22*(1), 24.

Behrman, J. R., & Rosenzweig, M. R. (2004). Returns to Birthweight. *The Review of Economics and Statistics*, *86*(2), 586–601. https://doi.org/10.1162/003465304323031139

Behrman, R. E., & Butler, A. S. (Eds.). (2007). *Preterm Birth: Causes, Consequences, and Prevention*. Washington (DC): National Academies Press (US). Retrieved from http://www.ncbi.nlm.nih.gov/books/NBK11362/

Bharadwaj, P., Lundborg, P., & Rooth, D.-O. (2017). Birth Weight in the Long Run. *Journal of Human Resources*, 0715–7235R. https://doi.org/10.3368/jhr.53.1.0715-7235R

Black, S. E., Devereux, P. J., & Salvanes, K. G. (2007). From the Cradle to the Labor Market? The Effect of Birth Weight on Adult Outcomes. *The Quarterly Journal of Economics*, *122*(1), 409–439. https://doi.org/10.1162/qjec.122.1.409

Bobak, M., Dejmek, J., Solansky, I., & Sram, R. J. (2005). Unfavourable birth outcomes of the Roma women in the Czech Republic and the potential explanations: a population-based study. *BMC Public Health*, *5*, 106. https://doi.org/10.1186/1471-2458-5-106

Breslau, N., DelDotto, J. E., Brown, G. G., Kumar, S., Ezhuthachan, S., Hufnagle, K. G., & Peterson, E. L. (1994). A Gradient Relationship Between Low Birth Weight and IQ at Age 6 Years. *Archives of Pediatrics & Adolescent Medicine*, *148*(4), 377–383. https://doi.org/10.1001/archpedi.1994.02170040043007

Case, A., Fertig, A., & Paxson, C. (2005). The lasting impact of childhood health and circumstance. *Journal of Health Economics*, *24*(2), 365–389. https://doi.org/10.1016/j.jhealeco.2004.09.008

Case, A., Lubotsky, D., & Paxson, C. (2002). Economic Status and Health in Childhood: The Origins of the Gradient. *American Economic Review*, *92*(5), 1308–1334. https://doi.org/10.1257/000282802762024520

Costa, D. L. (2004). Race and Pregnancy Outcomes in the Twentieth Century: A Long-Term Comparison. *The Journal of Economic History*, *64*(4), 1056–1086. https://doi.org/10.1017/S0022050704043086

Currie, J. (2009). Healthy, Wealthy, and Wise: Socioeconomic Status, Poor Health in Childhood, and Human Capital Development. *Journal of Economic Literature*, *47*(1), 87–122. https://doi.org/10.1257/jel.47.1.87

Currie, J. (2011). Inequality at Birth: Some Causes and Consequences. *American Economic Review*, *101*, 1–22. https://doi.org/10.1257/aer.101.3.1

Currie, J., & Hyson, R. (1999). Is the Impact of Health Shocks Cushioned by Socioeconomic Status? The Case of Low Birthweight. *American Economic Review*, *89*(2), 245–250. https://doi.org/10.1257/aer.89.2.245

Elder, T. E., Goddeeris, J. H., & Haider, S. J. (2011). A Deadly Disparity: A Unified Assessment of the Black-White Infant Mortality Gap. *The B.E. Journal of Economic Analysis & Policy*, *11*(1). https://doi.org/10.2202/1935-1682.2821

Elder, T. E., Goddeeris, J. H., & Haider, S. J. (2016). Racial and ethnic infant mortality gaps and the role of socio-economic status. *Labour Economics*, *43*, 42–54. https://doi.org/10.1016/j.labeco.2016.04.001

Fairlie, R. W. (2005). An extension of the Blinder-Oaxaca decomposition technique to logit and probit models. *Journal of Economic and Social Measurement*, *30*(4), 305–316.

Figlio, D., Guryan, J., Karbownik, K., & Roth, J. (2014). The Effects of Poor Neonatal Health on Children's Cognitive Development. *American Economic Review*, *104*(12), 3921–3955. https://doi.org/10.1257/aer.104.12.3921

FRA, & UNDP. (2012). *The situation of Roma in 11 EU Member States*. Luxembourg: Publications Office of the European Union.

Hack, M., Klein, N. K., & Taylor, H. G. (1995). Long-Term Developmental Outcomes of Low Birth Weight Infants. *The Future of Children*, *5*(1), 176–196. https://doi.org/10.2307/1602514

Howson, C., Kinney, M., & Lawn, J. (Eds.). (2012). *Born Too Soon: The Global Action Report on Preterm Birth*. Geneva: World Health Organization.

Joubert, K. (1991). Size at birth and some sociodemographic factors in gypsies in Hungary. *Journal of Biosocial Science*, *23*(1), 39–47.
https://doi.org/10.1017/S0021932000019052

Kemény, I., & Janky, B. (2006). Roma Population of Hungary 1971–2003. In I. Kemény (Ed.) (pp. 70–225). New York: Columbia University Press.

Kertesi, G., & Kézdi, G. (2011). The Roma/Non-Roma Test Score Gap in Hungary. *The American Economic Review*, *101*(3), 519–525. https://doi.org/10.1257/aer.101.3.519

Koval', J., Mrosková, S., & Schlosserová, A. (2012). Natality and infant mortality in Roma children in the Prešov region. *Medycyna Środowiskowa - Environmental Medicine*, *15*(2), 92–101.

Lawn, J. E., Cousens, S., & Zupan, J. (2005). 4 million neonatal deaths: When? Where? Why? *The Lancet*, *365*(9462), 891–900. https://doi.org/10.1016/S0140-6736(05)71048-5

Lhila, A., & Long, S. (2012). What is driving the black–white difference in low birthweight in the US? *Health Economics*, *21*(3), 301–315. https://doi.org/10.1002/hec.1715

Martin, J. A., Hamilton, B. E., Osterman, M. J. K., Driscoll, A. K., & Mathews, T. J. (2017). Births: Final Data for 2015. *National Vital Statistics Reports: From the Centers for Disease Control and Prevention, National Center for Health Statistics, National Vital Statistics System*, *66*(1), 1.

Morisaki, N., Kawachi, I., Oken, E., & Fujiwara, T. (2017). Social and anthropometric factors explaining racial/ethnical differences in birth weight in the United States. *Scientific Reports*, *7*, srep46657. https://doi.org/10.1038/srep46657

Oreopoulos, P., Stabile, M., Walld, R., & Roos, L. L. (2008). Short-, Medium-, and Long-Term Consequences of Poor Infant Health An Analysis Using Siblings and Twins. *Journal of Human Resources*, *43*(1), 88–138. https://doi.org/10.3368/jhr.43.1.88

Palloni, A., Milesi, C., White, R. G., & Turner, A. (2009). Early childhood health, reproduction of economic inequalities and the persistence of health and mortality differentials. *Social Science & Medicine*, *68*(9), 1574–1582. https://doi.org/10.1016/j.socscimed.2009.02.009

Pongrácz, T., & S. Molnár, E. (1994). A serdülőkori szülések néhány egészségügyi vonatkozása. *Demográfia*, *37*(2), 178–190.

Reichman, N. E. (2005). Low Birth Weight and School Readiness. *The Future of Children*, *15*(1), 91–116.

Stanković, S., Živić, S., Ignjatović, A., Stojanović, M., Bogdanović, D., Novak, S., … Vorgučin, I. (2016). Comparison of weight and length at birth of non-Roma and Roma newborn in Serbia. *International Journal of Public Health*, *61*(1), 69–73. https://doi.org/10.1007/s00038-015-0736-1

Appendix

Appendix 1: Details of the data

We linked two administrative datasets for this analysis: birth records and the census of 2011. The birth records of Hungary contain all live births starting in 1970. The content of the data is completely harmonized across time starting in 1981. The records contain information on the date of birth, birth weight, and age of gestation of the newborn babies, as well as the date of birth, level of education, employment, and residence of both the mother and the father. Birth records do not contain ethnic markers. The census of 2011 identifies the ethnic identity of all adult respondents; we discuss the quality of ethnic markers below. Birth records and census data are maintained by the Kozponti Statisztikai Hivatal (KSH, Central Statistical Office of Hungary). We accessed and linked the datasets in the secure data environment of KSH. In line with the literature, we excluded all twin births and birth records with missing key information (the proportion of twins in all births presents a small but growing rate, from around 2.1% in 1981 to 3.4% in 2010; the proportion of records with missing key information was around 1% throughout).

We linked the records of single births to the census of 2011 to identify the ethnicity of the mother. This linkage was feasible for children and mothers alive in September 2011, the month of the census count. As a result, our time series ends in 2010, the last complete year covered. The first year of our time series is 1981. Neither birth records nor census records have personal identifiers in Hungary, such as social security numbers, that would help in linking them. Names are permanently erased from the census records and are not recorded for birth records. At the same time, both datasets contain the gender and the exact date of birth of the newborn child and the mother, as well as the city, town or village of residence at birth, and other variables on the mother as well as the father. The most important variables we used for the linkage were the exact date of birth of the child and the mother, the gender of the child, and the residence of the mother at the time of the birth of the child. We found some additional matches when we narrowed the set of multiple matches by including other variables (parity of the child meaning whether first born, second born, etc.; gender and birthdates of siblings, date of birth of the father, levels of education of the parents), and we linked some of the originally unmatched records using fewer variables. Dates of previous live births to the mother were available in the birth records, which helped link siblings. We linked many births to census records via their siblings.

The proportion of linked records is high, and systematic differences between linked and not linked births are not very great. Figure A1 shows the linkage rates. Panel

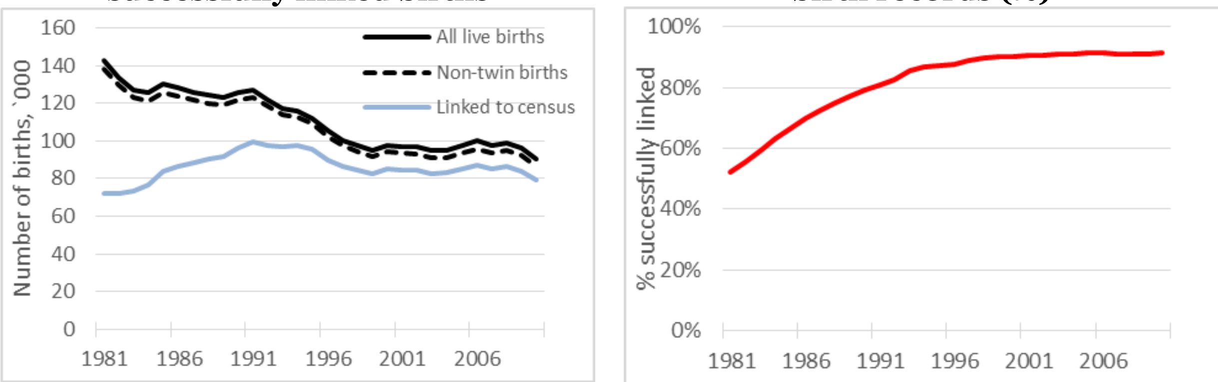
A shows the number of all live births, the number of births that we included in the search (excluding twin births and mothers born outside Hungary), and the number of successfully linked birth records. Panel B shows the percentage of birth records successfully linked: 90% of live births after 1995 are successfully linked, but the success rate declines continuously as we consider earlier births, to below 60% in 1981. The main reason for fewer successful links in earlier years is that links are created by using information both on the children and their mothers, and thus successful links are more likely for children living with their parents during the census of 2011. Children born in 1994 were 16 years old in 2011; children born earlier were older and thus had a smaller chance of living with their parents.

Figure A1.

Linkage rates

(A) Number of births and number of successfully linked births

(B) Fraction of successfully linked non-twin birth records (%)

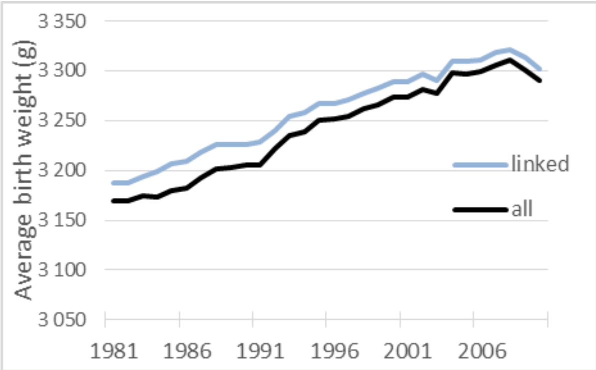


Linked births appear somewhat healthier over the time period. Figure A2 shows statistics of the main outcome variables for all (non-twin) births and successfully linked births. Linked births are of higher average weight, lower incidence of low birth weight, and lower incidence of preterm births. As a result, our analysis of the linked data misses a disproportionately large fraction of unhealthy newborns. Remarkably, however, the differences are quite stable across time, despite the strong trend in the fraction of successful links. This suggests that the additional births not linked in earlier years are selected by factors that are largely unrelated to outcomes.

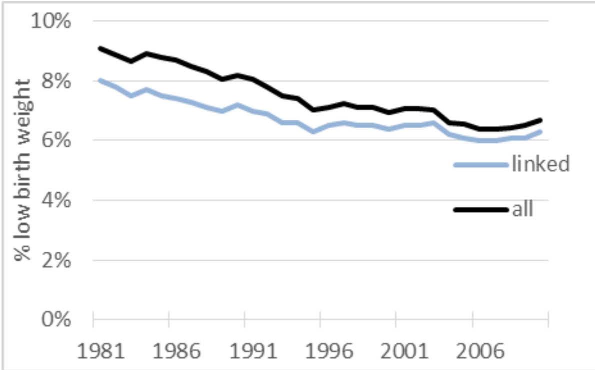
Figure A2.

Systematic differences in outcomes by linkage

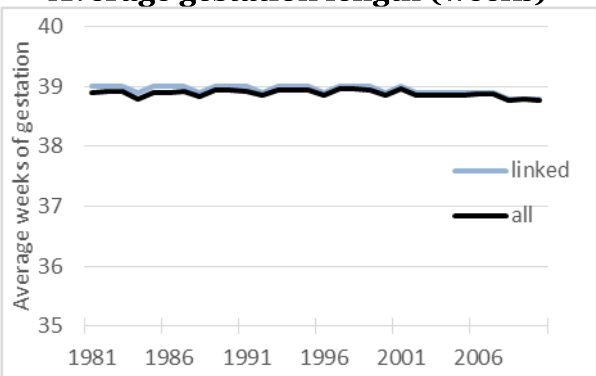
(A)
Average birth weight (g)



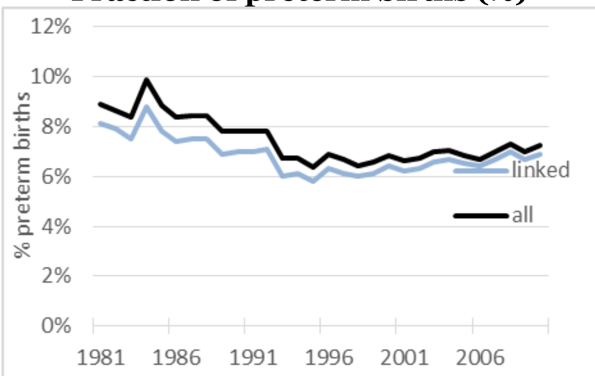
(B)
Fraction with low birth weight (%)



(C)
Average gestation length (weeks)



(D)
Fraction of preterm births (%)



To investigate selection to linkage further, we estimated probability models to predict successful linkage with all information available in the birth records. Table A1 shows results for two pooled time periods, 1981–1995 and 1996–2010. The results are rather similar in the two time periods. They show that linkage is more likely to be successful for mothers with higher education, if there is information on the father, if the father is employed, and for rural households. These results point to at least two opposing forces of selection. Rural households are easier to link because fewer births take place in a village than in a large city. Rural households are also of lower social status on average. At the same time, conditional on residence, births to mothers with higher social status are more likely to be linked, partly perhaps their children study longer and are more likely to reside with their parents after the age of 16, and partly because the information they report on the birth records and the census may be more complete and more accurate. The small differences in observed outcomes between linked and non-linked births are thus due to the presence of these opposing forces, as well as the probably large role of random factors in the earlier years. Selection into successful linking is unlikely to lead to large biases in our analysis. Nevertheless, we re-estimated all of our main results with inverse probability weighting among our robustness checks.

Table A1:

Selection into successful linkage. Regression results with covariates from the birth records

	1981-1995 OLS	1996-2010 OLS
Girl	-0.036**	0.002**
	(0.001)	(0.000)
Month of birth Feb	-0.001	0.001
	(0.002)	(0.001)
Month of birth Mar	0.004*	0.001
	(0.002)	(0.001)
Month of birth Apr	0.003+	0.000
	(0.002)	(0.001)
Month of birth May	0.004**	0.001
	(0.002)	(0.001)
Month of birth Jun	0.008**	0.002
	(0.002)	(0.001)
Month of birth Jul	0.009**	0.000
	(0.002)	(0.001)
Month of birth Aug	0.015**	0.002+
	(0.002)	(0.001)
Month of birth Sep	0.016**	0.002
	(0.002)	(0.001)
Month of birth Oct	0.021**	0.004**
	(0.002)	(0.001)
Month of birth Nov	0.021**	-0.000
	(0.002)	(0.001)
Month of birth Dec	0.022**	0.003*
	(0.002)	(0.001)
Mother's ed. 8 grade	0.033**	0.014**
	(0.002)	(0.002)
Mother's ed. Voc. training school	0.091**	0.035**
	(0.002)	(0.002)
Mother's ed. Matura	0.074**	0.044**
	(0.002)	(0.002)
Mother's ed. College/university	0.083**	0.055**
	(0.002)	(0.002)
Mother's ed. Missing	-0.070**	-0.037**
	(0.020)	(0.008)
Mother's age 20-24	-0.013**	0.006**
	(0.001)	(0.001)
Mother's age 25-29	-0.045**	0.014**
	(0.001)	(0.001)
Mother's age 30-34	-0.074**	0.014**
	(0.002)	(0.002)
Mother's age 35-39	-0.093**	0.005**
	(0.002)	(0.002)

Mother's age 40-44	-0.123** (0.004)	-0.005+ (0.003)
Mother's age 45+	-0.195** (0.020)	-0.010 (0.012)
non-missing father + married	-0.017+ (0.010)	0.024** (0.008)
non-missing father + not married	-0.032** (0.009)	0.005 (0.008)
Mother on parental leave	0.027** (0.001)	0.002** (0.001)
Mother unemployed	0.083** (0.002)	-0.024** (0.001)
Mother student	0.012** (0.004)	-0.037** (0.003)
Mother on other nonworking	-0.015** (0.001)	-0.026** (0.001)
Mother employment missing	-0.018** (0.005)	-0.067** (0.006)
# of abortions=1	-0.027** (0.001)	-0.004** (0.001)
# of abortions=2	-0.042** (0.002)	-0.008** (0.002)
# of abortions=3	-0.060** (0.004)	-0.009** (0.003)
# of abortions=4	-0.064** (0.007)	-0.006 (0.006)
# of abortions=5	-0.094** (0.009)	-0.001 (0.007)
# of abortions=missing	-0.118 (0.142)	
# of misscarriages=1	-0.008** (0.001)	-0.004** (0.001)
# of misscarriages=2	-0.017** (0.002)	-0.010** (0.002)
# of misscarriages=3	-0.031** (0.005)	-0.017** (0.003)
# of misscarriages=4	-0.033** (0.008)	-0.021** (0.006)
# of misscarriages=5	-0.078** (0.011)	-0.025** (0.008)
# of misscarriages= missing	0.072 (0.100)	
# of previous live births=1	-0.013** (0.001)	0.035** (0.001)
# of previous live births=2	0.015** (0.001)	0.045** (0.001)
# of previous live births=3	0.016** (0.002)	0.040** (0.001)
# of previous live births=4	-0.003 (0.003)	0.029** (0.002)
# of previous live births=5	-0.025** (0.005)	-0.005 (0.003)
# of previous live births=6	-0.058** (0.006)	-0.045** (0.005)
# of previous live births=7	-0.095** (0.009)	-0.070** (0.007)

# of previous live births=8	-0.137**	-0.095**
	(0.012)	(0.010)
# of previous live births=9	-0.161**	-0.126**
	(0.010)	(0.010)
Father's ed. 8 grades	0.043**	0.005+
	(0.003)	(0.003)
Father's ed. voc. training school	0.058**	0.011**
	(0.003)	(0.003)
Father's ed. Matura	0.037**	0.009**
	(0.003)	(0.003)
Father's ed. College/university	0.011**	0.010**
	(0.003)	(0.003)
Father's ed. missing	0.005	0.011+
	(0.008)	(0.006)
Father on parental leave	0.078**	-0.022+
	(0.020)	(0.012)
Father unemployed	0.122**	-0.010**
	(0.002)	(0.001)
Father student	-0.035**	-0.018**
	(0.005)	(0.005)
Father other nonworking	0.081**	-0.008**
	(0.002)	(0.001)
Father employment missing	0.069**	-0.007
	(0.006)	(0.005)
Father's age 25-29	-0.004	0.009**
	(0.002)	(0.003)
Father's age 30-34	-0.004	0.007*
	(0.003)	(0.003)
Father's age 35-39	-0.001	0.003
	(0.003)	(0.003)
Father's age 40-44	-0.009**	-0.005
	(0.003)	(0.003)
Father's age 45+	-0.036**	-0.018**
	(0.004)	(0.004)
Missing	-0.077**	-0.039**
	(0.008)	(0.009)
County Baranya	-0.006**	0.038**
	(0.002)	(0.002)
County Bács-Kiskun	-0.008**	0.043**
	(0.002)	(0.001)
County Békés	-0.034**	0.037**
	(0.002)	(0.002)
County Borsod-Abaúj-Zemplén	-0.003*	0.041**
	(0.002)	(0.001)
County Csongrád	0.022**	0.049**
	(0.002)	(0.001)
County Fejér	0.030**	0.047**
	(0.002)	(0.001)
County Győr-Moson-Sopron	0.040**	0.051**
	(0.002)	(0.001)
County Hajdú-Bihar	0.030**	0.054**
	(0.002)	(0.001)
County Heves	0.001	0.038**
	(0.002)	(0.002)
County Komárom-Esztergom	-0.001	0.039**
	(0.002)	(0.002)

County Nógrád	-0.000	0.048**
	(0.002)	(0.002)
County Pest	0.051**	0.039**
	(0.001)	(0.001)
County Somogy	-0.018**	0.036**
	(0.002)	(0.002)
County Szabolcs-Szatmár-Bereg	0.001	0.054**
	(0.002)	(0.001)
County Jász-Nagykun-Szolnok	-0.013**	0.050**
	(0.002)	(0.001)
County Tolna	-0.026**	0.035**
	(0.002)	(0.002)
County Vas	0.024**	0.045**
	(0.002)	(0.002)
County Veszprém	0.005*	0.044**
	(0.002)	(0.002)
County Zala	-0.006*	0.041**
	(0.002)	(0.002)
Big city (other than Budapest)	-0.029**	-0.030**
	(0.001)	(0.001)
Smaller town	-0.021**	-0.017**
	(0.001)	(0.001)
Rural	0.000	0.000
	(.)	(.)
Constant	0.692**	0.796**
	(0.010)	(0.009)
Adjusted R-squared	0.02	0.03
N	1820546	1403755

Reference categories: Month of birth January, education category 0-7 grades, age category 15-19 years, employment category working, information on father is missing, number of miscarriages, abortions, previous live births all zero, residence Budapest.
Robust standard errors are in parentheses + p<.10, * p<.05, ** p<.01

The Hungarian census of 2011 asked two questions of every adult respondent concerning their ethnic or national identity, thus allowing for the reporting of multiple identities. It also contained a question on mother tongue, and the language used in the family and with friends. We marked all mothers we found in the census as Roma if they identified themselves as Roma or indicated Roma as their mother tongue or a language spoken in the family or with friends. The total number of Roma marked this way in the census of 2011 is 315,583, or 3.2% of the population. This is around 60% as large as other estimates of the size of the Roma minority in Hungary (Kemény and Janky, 2006). In principle, selection into reporting on the census may bias our estimates of the Roma figures. However, the Roma populations represented by the two samples are very close. The last available data representative of the entire Roma population of Hungary is from the Roma survey of 2003 (representing all households whose members were classified as Roma by outside observers, as well as self-identified as Roma). We compared mothers who gave birth between 2000 and 2002 in the census of 2011 and the Roma survey of 2003. Table A2 shows that the two samples are very close in terms of education and age at delivery. Mothers in the census are more likely to be divorced, which may be due to the fact that marital status is measured eight years later in the census for the same cohort of mothers.

Table A2:
Selection into Roma identity in the census. Comparing the characteristics of Roma mothers that gave birth in 2000-2002, in the census of 2011 and the representative Roma survey of 2003

	Census 2011 (N=17503)	Roma survey 2003 (N=349)
Mother's education [a]		
0-7 classes	0.25	0.30
8 classes	0.61	0.53
vocational training	0.10	0.13
Secondary	0.03	0.03
Tertiary	0.01	0.00
Missing	0.00	0.01
Average age of the mother [b]		
All deliveries	24.1	24.1
First deliveries	20.1	19.7
Mother's marital status [a]		
Unmarried	0.42	0.44
Married	0.44	0.50
Widowed	0.02	0.03
Divorced	0.12	0.01
Missing	0.00	0.02
Average parity	2.55	2.73

Notes.
[a] Measured at the year of the interview: 2011 for the Census, 2003 for the Roma survey.
[b] Measured at the year of the delivery: 2000 to 2002 for both the Census and the Roma survey.
Data: Roma mothers that gave birth between 2000 and 2002, measured in the Census of 2011 and the Roma survey of 2003. Mother's age at the delivery is calculated using the date of the delivery and the birth date of the mother in the Census, and the year of the delivery and the birth year of the mother in the Roma survey. Parity is calculated only for the deliveries between 2000-2002.

Even if the subpopulation of Roma mothers in the sample represents the population of all Roma mothers, the fact that not all of them are identified may bias the non-Roma estimates. This is because mothers not marked as Roma in the census who would otherwise identify as Roma are counted among the non-Roma group in our analysis. However, the share of unidentified Roma is small, at around 3% of the entire population, so this downward bias is likely to be very small. Thus, we conclude that while there is substantial selection into successful linkage, as well as Roma identification in the census, they are unlikely to induce substantial biases in our estimates.

Appendix 2: Additional results

Figures

Figure A3

Trends in average birth weight (using inverse probability weights to account for selection)

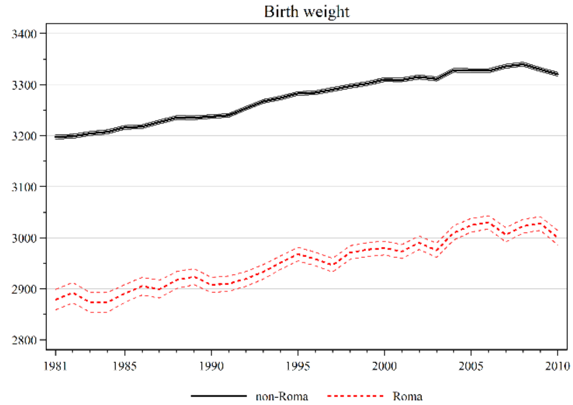


Figure A4

Trends in the fraction of low birth weights (using inverse probability weights to account for selection)

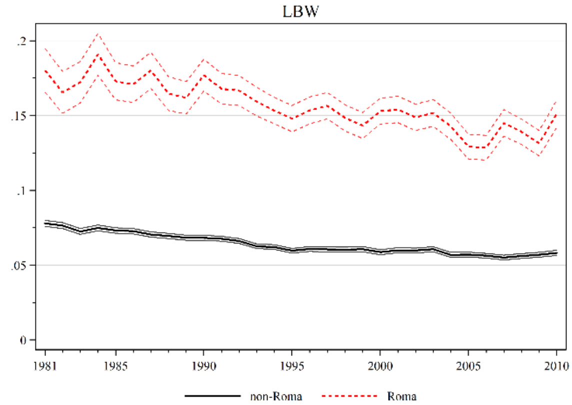


Figure A5

Trends in average gestation length (using inverse probability weights to account for selection)

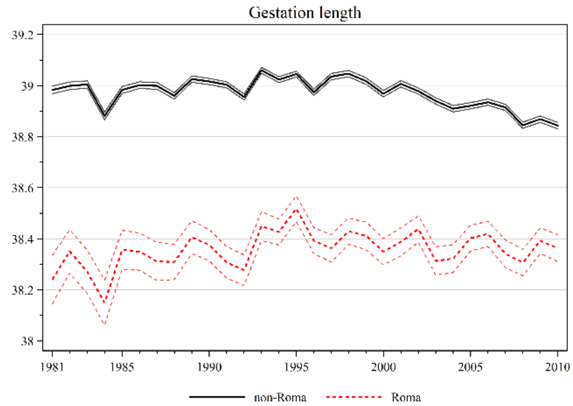


Figure A6

Trends in the fraction of preterm births (using inverse probability weights to account for selection)

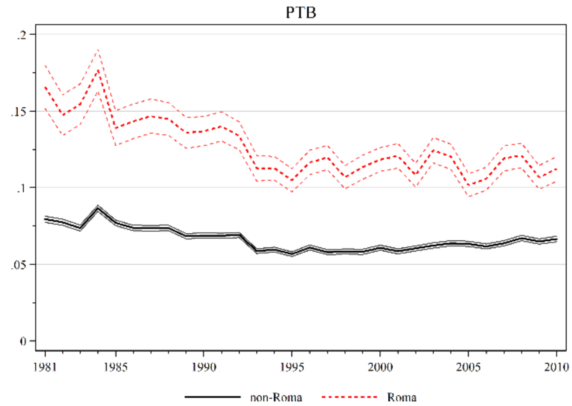


Table A3:

Descriptive statistics, 2008-2010

	Roma (N=18,417)		non-Roma (N=230,853)		Difference: Roma – non-Roma	
	Mean	SD	Mean	SD	Mean diff	p-value
Birth weight	3023	546	3336	543	-313.416	0.000
LBW (<2500g)	0.138	0.345	0.056	0.229	0.083	0.000
Gestation length	38.4	2.1	38.9	1.7	-0.5	0.000
PTB (<37 weeks)	0.113	0.316	0.065	0.247	0.047	0.000
Girl	0.482	0.500	0.487	0.500	-0.005	0.196
Birth month						
january	0.087	0.282	0.087	0.282	-0.000	0.863
february	0.075	0.263	0.079	0.270	-0.004	0.030
march	0.082	0.274	0.084	0.277	-0.002	0.293
april	0.074	0.262	0.081	0.272	-0.006	0.002
may	0.082	0.274	0.079	0.269	0.003	0.140
june	0.080	0.271	0.082	0.274	-0.002	0.298
july	0.093	0.290	0.091	0.287	0.002	0.397
august	0.089	0.284	0.086	0.280	0.003	0.173
september	0.088	0.283	0.088	0.283	-0.000	0.909
october	0.085	0.279	0.084	0.277	0.001	0.558
november	0.078	0.268	0.078	0.269	-0.001	0.761
december	0.089	0.284	0.082	0.274	0.007	0.001
Mother's age						
-19	0.268	0.443	0.038	0.191	0.230	0.000
20-24	0.337	0.473	0.125	0.330	0.213	0.000
25-29	0.204	0.403	0.306	0.461	-0.102	0.000
30-34	0.127	0.332	0.371	0.483	-0.244	0.000
35-39	0.054	0.227	0.137	0.344	-0.083	0.000
40-44	0.010	0.098	0.023	0.151	-0.013	0.000
45-	0.001	0.023	0.001	0.029	-0.000	0.102
Mother's education						
0-7	0.180	0.384	0.011	0.105	0.168	0.000
8	0.636	0.481	0.131	0.338	0.505	0.000
vocational training	0.106	0.307	0.164	0.370	-0.058	0.000
secondary	0.056	0.229	0.347	0.476	-0.291	0.000
tertiary	0.012	0.110	0.340	0.474	-0.328	0.000
missing	0.011	0.103	0.007	0.083	0.004	0.000
Mother's labor force status						
employed	0.115	0.319	0.700	0.458	-0.585	0.000
parental leave	0.418	0.493	0.154	0.361	0.265	0.000
unemployed	0.149	0.356	0.056	0.230	0.093	0.000
student	0.052	0.221	0.019	0.138	0.032	0.000
other	0.255	0.436	0.063	0.243	0.192	0.000
missing	0.011	0.104	0.008	0.091	0.002	0.002
Mother's marital status						
missing father	0.399	0.490	0.069	0.254	0.329	0.000
non-missing father + married	0.260	0.438	0.635	0.481	-0.375	0.000
non-missing father + not married	0.342	0.474	0.296	0.456	0.046	0.000

N of abortions						
0	0.731	0.443	0.840	0.366	-0.109	0.000
1	0.151	0.358	0.118	0.323	0.033	0.000
2	0.066	0.249	0.030	0.171	0.036	0.000
3	0.027	0.163	0.008	0.087	0.020	0.000
4	0.012	0.108	0.002	0.049	0.010	0.000
5+	0.013	0.111	0.002	0.040	0.011	0.000
N of fetal losses						
0	0.841	0.366	0.858	0.350	-0.017	0.000
1	0.119	0.324	0.112	0.316	0.007	0.005
2	0.029	0.169	0.023	0.150	0.006	0.000
3	0.007	0.083	0.005	0.072	0.002	0.006
4	0.002	0.048	0.001	0.038	0.001	0.016
5+	0.002	0.039	0.001	0.025	0.001	0.002
N of previous live births						
0	0.308	0.462	0.468	0.499	-0.160	0.000
1	0.257	0.437	0.344	0.475	-0.087	0.000
2	0.179	0.383	0.123	0.328	0.056	0.000
3	0.112	0.315	0.038	0.191	0.074	0.000
4	0.064	0.244	0.015	0.120	0.049	0.000
5	0.036	0.187	0.006	0.080	0.030	0.000
6	0.021	0.142	0.003	0.055	0.018	0.000
7	0.011	0.106	0.002	0.040	0.010	0.000
8	0.006	0.077	0.001	0.030	0.005	0.000
9+	0.006	0.076	0.001	0.029	0.005	0.000
Father's age						
-19	0.047	0.212	0.006	0.079	0.041	0.000
20-24	0.165	0.371	0.048	0.214	0.117	0.000
25-29	0.152	0.359	0.188	0.390	-0.036	0.000
30-34	0.117	0.322	0.374	0.484	-0.257	0.000
35-39	0.066	0.249	0.207	0.405	-0.141	0.000
40-44	0.028	0.165	0.072	0.259	-0.044	0.000
45-	0.016	0.127	0.032	0.176	-0.016	0.000
missing	0.408	0.491	0.072	0.259	0.336	0.000
Father's education						
0-7	0.049	0.216	0.004	0.061	0.045	0.000
8	0.346	0.476	0.092	0.289	0.254	0.000
vocational training	0.130	0.336	0.272	0.445	-0.142	0.000
secondary	0.037	0.188	0.291	0.454	-0.254	0.000
tertiary	0.010	0.099	0.258	0.437	-0.248	0.000
missing	0.428	0.495	0.083	0.276	0.345	0.000
Father's labor force status						
employed	0.260	0.439	0.837	0.369	-0.577	0.000
parental leave	0.000	0.013	0.000	0.017	-0.000	0.221
unemployed	0.182	0.386	0.050	0.219	0.132	0.000
Student	0.008	0.091	0.003	0.057	0.005	0.000
other	0.118	0.323	0.023	0.150	0.095	0.000
missing	0.430	0.495	0.086	0.280	0.344	0.000
Settlement type						
Budapest	0.046	0.209	0.188	0.390	-0.142	0.000
big city	0.084	0.278	0.205	0.404	-0.120	0.000
small town	0.322	0.467	0.318	0.466	0.004	0.264
rural	0.548	0.498	0.290	0.454	0.258	0.000
County						
Budapest	0.046	0.209	0.188	0.390	-0.142	0.000
Heves	0.062	0.240	0.027	0.162	0.034	0.000

Komárom-Esztergom	0.016	0.126	0.033	0.177	-0.016	0.000
Nógrád	0.044	0.206	0.017	0.128	0.028	0.000
Pest	0.064	0.245	0.144	0.352	-0.080	0.000
Somogy	0.048	0.215	0.027	0.161	0.022	0.000
Szabolcs-Szatmár-Bereg	0.155	0.362	0.054	0.227	0.101	0.000
Jász-Nagykun-Szolnok	0.065	0.246	0.036	0.188	0.028	0.000
Hajdú-Bihar	0.065	0.247	0.055	0.228	0.010	0.000
Tolna	0.024	0.155	0.021	0.143	0.004	0.002
Vas	0.006	0.080	0.024	0.152	-0.017	0.000
Veszprém	0.017	0.128	0.035	0.183	-0.018	0.000
Baranya	0.048	0.213	0.037	0.188	0.011	0.000
Zala	0.025	0.155	0.025	0.156	-0.000	0.924
Bács-Kiskun	0.038	0.191	0.052	0.222	-0.014	0.000
Békés	0.029	0.169	0.030	0.172	-0.001	0.384
Borsod-Abaúj-Zemplén	0.208	0.406	0.063	0.243	0.145	0.000
Csongrád	0.012	0.110	0.041	0.198	-0.029	0.000
Fejér	0.018	0.133	0.045	0.206	-0.027	0.000
Győr-Moson-Sopron	0.010	0.100	0.048	0.213	-0.038	0.000

Table A4.

Residual gap estimates using alternative models

		(1) Birth weight	(2) Low birth weight	(3) Gestation length	(4) Preterm birth
(A)	Weighted least squares	-70.8** (4.9)	0.015** (0.003)	-0.082** (0.018)	0.001 (0.003)
(B)	Logit regression (marginal effects)	NA	0.008** (0.002)	NA	0.002 (0.002)
(C)	Oaxaca-Blinder decomposition (linear)	-68.5** (5.0)	0.014** (0.003)	-0.070** (0.019)	-0.000 (0.003)
(D)	Oaxaca-Blinder decomposition (logit)	NA	0.014** (0.003)	NA	0.001 (0.003)
(E)	Propensity score matching (ATET)	-52.2** (6.9)	0.014** (0.004)	-0.053* (0.026)	-0.002 (0.004)

(C) and (D): non-Roma coefficients used in estimating composition effects.
Robust standard errors are in parentheses. + p<.10, * p<.05, ** p<.01

Table A5.

The share of parental characteristics in explaining the raw gap. Results from logit-based decompositions of binary outcomes

	Low birth weight	Preterm birth
Education	63%	78%
Labor force status	16%	19%
Pregnancy history	-1%	3%
Geography	4%	6%
Marital status	21%	56%
Age	-19%	-64%
Other	0%	0%
Residual gap	16%	2%
Total	100%	100%

Table A6.

The share of parental characteristics in explaining the raw gap in five-year intervals

	1981-1985	1986-1990	1991-1995	1996-2000	2001-2005	2006-2010
Birth weight						
Education	51%	54%	60%	54%	49%	53%
Labor force status	11%	11%	12%	11%	13%	10%
Pregnancy history	-1%	0%	0%	0%	0%	-3%
Geography	4%	5%	4%	5%	5%	4%
Marital status	7%	-2%	-4%	2%	3%	13%
Age	5%	10%	3%	6%	8%	-1%
Other	1%	0%	0%	0%	0%	0%
Residual gap	22%	22%	25%	22%	22%	24%
Total	100%	100%	100%	100%	100%	100%
Low birth weight						
	1981-1985	1986-1990	1991-1995	1996-2000	2001-2005	2006-2010
Education	55%	59%	68%	50%	48%	55%
Labor force status	13%	8%	14%	18%	16%	11%
Pregnancy history	4%	6%	7%	8%	7%	3%
Geography	0%	2%	1%	2%	4%	3%
Marital status	14%	-4%	-14%	-2%	9%	24%
Age	0%	15%	6%	8%	-1%	-13%
Other	0%	0%	0%	0%	0%	0%
Residual gap	14%	14%	18%	16%	17%	17%
Total	100%	100%	100%	100%	100%	100%
Gestation length						
	1981-1985	1986-1990	1991-1995	1996-2000	2001-2005	2006-2010
Education	37%	45%	66%	38%	34%	46%
Labor force status	17%	10%	11%	20%	20%	13%
Pregnancy history	11%	13%	13%	15%	14%	10%
Geography	2%	4%	5%	8%	10%	7%
Marital status	13%	-9%	-26%	-11%	-12%	22%
Age	9%	27%	17%	16%	19%	-14%
Other	0%	0%	0%	0%	0%	0%
Residual gap	11%	10%	14%	14%	15%	16%
Total	100%	100%	100%	100%	100%	100%
Preterm birth						
	1981-1985	1986-1990	1991-1995	1996-2000	2001-2005	2006-2010
Education	41%	55%	82%	51%	41%	58%
Labor force status	18%	9%	18%	21%	26%	15%
Pregnancy history	12%	15%	16%	17%	14%	8%
Geography	-1%	-1%	-1%	3%	6%	5%
Marital status	14%	-11%	-29%	-16%	-13%	44%
Age	11%	30%	13%	21%	21%	-34%
Other	0%	0%	0%	0%	0%	0%
Residual gap	5%	3%	1%	3%	5%	4%
Total	100%	100%	100%	100%	100%	100%