Universitat Pompeu Fabra
Departament d'Economia i Empresa

DOCTORAT EN ECONOMIA, FINANCES I EMPRESA

ESSAYS IN HEALTH AND GENDER ECONOMICS

Tesi Doctoral de:
Ana Rodríguez González

Directora de la tesi:
Libertad González

TESI DOCTORAL UPF/ANY 2020

Aos meus pais e a Miquel

ACKNOWLEDGEMENTS

I am extremely grateful to my advisor, Libertad González, for her excellent guidance and her support in the completion of this thesis. I have learned a great deal from her about how to be a better researcher.

I am also very thankful to Christian Fons-Rosen, Alessandro Tarozzi and Ana Tur-Prats for their immense help during these years, as well as to Cristina Bellés, Rosa Ferrer, Albrecht Glitz, Gianmarco León, Vicente Ortún, Maria Petrova, Dijana Zejcirovic and the rest of participants in the LPD Breakfast/Applied Lunch seminar for their comments and feedback. A big thanks also to Marta Araque and Laura Agustí for all their support and incredible efficiency. The department and all PhD students are so very lucky to have them.

This thesis would not have been possible without my great coauthors: Ana, Miquel, Carlos, Lauri, Mika, and Libertad. I have learned a lot working with all of you. A special thank you to Ana, "la meua companya infatigable d'aventures"—the best PhD mate and friend I could have had during these years. I am so happy to have shared it with you.

I am also indebted to those who have made these years in Barcelona so much better in so many ways. Thank you to Patricia, Matt, Dani, Ana, Miquel S., Óscar and Korie, who have been like family to me. To Christian and Christoph, for always brightening my lunchtimes and days in the office. I a la meva altra família—Carme, Carles, Aleix, Vicky—per fer-me sentir sempre a casa.

Por último, grazas aos meus pais polo seu apoio incondicional e por animarme sempre a seguir medrando. Ao meu irmán, polo seu agarimo. I a en Miquel, per fer-me costat, sempre.

Abstract

This thesis consists of four independent articles. The first two study the causal effect of being born by cesarean section on child health. In the first paper, we use data from Spain and show that avoidable unplanned C-sections have a negative impact on neonatal health, which is however small compared to the associations reported by previous literature. The second paper uses administrative data from Finland to study the impact of C-sections on children's longer-term outcomes until age 15. Our results show that unplanned C-sections increase the risk of childhood asthma, but do not seem to affect the probability of other immune-related diagnoses previously associated with C-sections. In the third paper, I study the effects of the increasing female-male gap in education in the marriage market on marriage and fertility, exploiting the gradual implementation of a school reform in Finland that increased women's relative level of education. My results show decreases in marriage and fertility in marriage markets with a larger female advantage in education. Finally, the last paper analyzes the evolution of inequality in mortality in Spain during 1990-2014, focusing on age-specific mortality and considering inequality across small geographical areas, ranked by average socioeconomic status. We find that mortality decreased substantially during this period, with little change in inequality in most age groups.

Resum

Aquesta tesi es composa de quatre articles independents. Els dos primers estudien l'efecte de néixer per cesària en la salut infantil. En el primer article mostrem, amb dades d'Espanya, que les cesàries no programades evitables tenen un impacte negatiu en la salut neonatal. Aquest impacte, però, és petit en comparació amb les associacions trobades per estudis previs. El segon article fa ús de dades administratives de Finlàndia per estudiar l'efecte de les cesàries en salut infantil a més llarg termini, fins als 15 anys d'edat. Els resultats mostren que las cesàries no programades augmenten el risc d'asma infantil, però no semblen afectar la probabilitat de patir altres malalties relacionades amb el sistema immunitari que havien estat associades prèviament amb les cesàries. El tercer treball estudia l'efecte d'un augment en la bretxa de gènere en nivell educatiu a favor de les dones al mercat matrimonial, fent ús d'una reforma escolar a Finlàndia que va augmentar el nivell educatiu relatiu femení. Els resultats mostren que en mercats amb un avantatge educatiu femení més gran els matrimonis i la fertilitat van decréixer. Finalment, el quart article analitza l'evolució de la desigualtat en mortalitat a Espanya entre 1990 i 2014, centrant-se en la mortalitat específica per edat i considerant desigualtat entre àrees geogràfiques petites, ordenades per nivell socioeconòmic mitjà. Trobem baixades substancials en mortalitat durant aquests anys, amb poc canvi en desigualtat a la majoria de grups d'edat.

Resumen

Esta tesis se compone de cuatro artículos independientes. Los dos primeros estudian el efecto causal de nacer por cesárea en salud infantil. En el primer artículo mostramos, usando datos de España, que las cesáreas no programadas evitables tienen un impacto negativo en salud neonatal. Este, sin embargo, es pequeño en comparación con las asociaciones que estudios previos habían encontrado. En el segundo trabajo se usan datos administrativos de Finlandia para estudiar el efecto de las cesáreas en salud infantil a más largo plazo, hasta los 15 años de edad. Nuestros resultados muestran que las cesáreas no programadas aumentan el riesgo de asma infantil, pero no parecen afectar a la probabilidad de padecer otras enfermedades relacionadas con el sistema inmunitario que se habían asociado previamente con las cesáreas. En el tercer artículo se estudia el efecto de una mayor brecha de género en nivel educativo a favor de las mujeres en el mercado matrimonial, usando la implementación gradual de una reforma escolar en Finlandia que incrementó el nivel educativo relativo de las mujeres. Los resultados muestran que en mercados con una mayor ventaja educativa femenina disminuyeron los matrimonios y la fertilidad. Por último, el cuarto artículo analiza la evolución de la desigualdad en mortalidad en España entre 1990 y 2014, centrándose en mortalidad específica por edades y considerando desigualdad entre pequeñas áreas geográficas, ordenadas por nivel socioeconómico medio. Encontramos que la mortalidad disminuyó sustancialmente durante estos años, con pocos cambios en desigualdad en la mayoría de grupos de edad.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

1

INTRODUCTION

This thesis consists of four independent articles within the fields of health and gender economics.

The first two papers study the causal effect of cesarean sections on child health. The literature has recognized the importance of early-life circumstances for a wide range of long-term outcomes. At the same time, a large number of studies have reported associations between cesarean sections and worse infant health. However, we know relatively little about the causal nature of this relationship. These papers contribute to filling this gap by providing credible causal estimates of the impact of C-sections.

In the first paper, joint with Ana Costa-Ramón, Miquel Serra-Burriel and Carlos Campillo-Artero, we use data from public hospitals in Spain and exploit exogenous variation in C-section rates by time of birth to study the impact of avoidable unplanned C-sections on newborn health. Our findings show that C-sections have a negative impact on neonatal health indicators, which is however small compared to previous associations reported in the medical literature.

In the second paper, coauthored with Ana Costa-Ramón, Mika Kortelainen and Lauri Sääksvuori, we use high-quality administrative data from Finland to study the impact of C-sections on children's longer-term outcomes, following them until age 15. For identification of the causal effect, we combine an instrumental variable strategy that exploits the increase in C-sections on days preceding a weekend or public holiday, with a difference-in-difference analysis that exploits variation within and across sibling pairs. Our results show that unplanned C-sections increase the risk of asthma from early childhood, but we do not find evidence that they affect the probability of other immune-related diagnoses previously associated with C-sections. Hence, while our findings highlight the long-term costs of potentially avoidable interventions at birth, they also paint a more nuanced picture of the effects of cesarean delivery.

The third paper studies the effects of the increasing female-male gap in education in the marriage market on marriage and fertility. Recent years have seen a reversal of the traditional gender gap in education in favor of men in many countries. This emerging phenomenon could have profound implications for the family, challenging traditional patterns of union formation, which were characterized by educational hypergamy (that is, an educational advantage in favor of the husband). My empirical strategy exploits the gradual implementation of a large school reform in Finland that increased women's relative level of education. I study the reduced-form relationship between marriage market exposure to the reform and marriage and fertility outcomes. The results show that in marriage markets with a larger female advantage in education, men had fewer children and a lower probability of being in a couple by age 40. I provide suggestive evidence that these results are mostly driven by the mismatch between the distributions of educational attainment of men and women, which present "excess" numbers of high-educated women and low-educated men. My findings also point at potential negative consequences for low-educated men's health behaviors and mental health.

Finally, the fourth article, joint with Libertad González, analyzes the evolution of inequality in mortality in Spain during recent decades, from 1990 to 2014. We follow the recent literature by focusing on age-specific mortality and considering inequality across small geographical areas, ranked by average socioeconomic status. Our results show substantial

declines in mortality for most age groups, which were particularly pronounced for men. We find low levels of inequality during the whole period, except for the elderly, and no evidence of an increase after the recent recession.

2

IT'S ABOUT TIME: CESAREAN SECTIONS AND NEONATAL HEALTH

Joint with Ana María Costa-Ramón (UPF), Miquel Serra-Burriel (Universitat de Barcelona) and Carlos Campillo-Artero (Servei de Salut de les Illes Balears)

Published as: Costa-Ramón, A. M., Rodríguez-González, A., Serra-Burriel, M., and Campillo-Artero, C. (2018). It's about time: Cesarean sections and neonatal health. *Journal of Health Economics*, 59:46 – 59. http://www.sciencedirect.com/science/article/pii/S0167629617307609. DOI: https://doi.org/10.1016/j.jhealeco.2018.03.004.

2.1 INTRODUCTION

Recent years have seen increasing concern over the rise in cesarean section births. Among OECD countries in 2013, on average more than 1 out of 4 births involved a c-section, compared to 1 out of 5 in 2000 (OECD, 2013). This rise has been largely debated because c-sections are associated with greater complications and higher maternal and infant mortality and morbidity compared to vaginal births. However, the available studies may suffer from omitted variable bias, as mothers who give birth by c-sections may be different from those who have vaginal births in terms of characteristics that can affect the health outcomes of the child and the mother after birth. Along these lines, the WHO has recently pointed out the need for more research in order to better understand the health effects of cesarean sections on immediate and future outcomes, remarking that "the effects of cesarean section rates on other outcomes, such as maternal and neonatal morbidity, pediatric outcomes and psychological or social well-being, are still unclear" (WHO, 2015).

This paper aims to help fill this research gap by providing new evidence of a causal link between unplanned cesarean sections and newborn health outcomes. Understanding the impact of c-sections on neonatal health is of relevance, as fetal and neonatal outcomes have been shown to be determinants not only of future health, but also of other later life outcomes, such as test scores, educational attainment, and income (Almond and Currie, 2011). In particular, we look at the impact of c-sections on Apgar scores, a widely used measure of newborn well-being. Apgar scores have been found to be predictive of health, cognitive ability, and behavioral problems of children at age three (Almond et al., 2005), of reading and math test scores in grades 3-8 (Figlio et al., 2014), and of school attainment and social assistance receipt after age 18 (Oreopoulos et al., 2008). We also analyze the effect of c-sections on other indicators of newborn wellbeing, such as needing reanimation or being admitted to the intensive care unit (ICU).

In order to show the existence of a causal relationship between unscheduled c-sections and health, we use exogenous variation in the probability of having a c-section at different times of day. Indeed, although nature distributes births and associated problems uniformly, some studies have demonstrated that time-dependent variables related to physicians' demand for leisure are significant predictors of unplanned c-sections (Brown, 1996). Using a sample of birth registries in public hospitals in Spain, we first document that, in this context, unplanned c-sections are more likely to be performed in the early hours of the night (from 11 pm to 4 am). We discuss how the structure of medical shifts and the higher opportunity cost in terms of time that vaginal deliveries imply might explain physicians' incentives to perform more c-sections during this time of day. We then show that mothers giving birth at different times of day are observationally similar, also in terms of pregnancy and labor characteristics that might predict a medically-indicated c-section. The results thus suggest that the excess number of c-sections observed at the early night are due to non-medical reasons. We consequently adopt an instrumental variable approach, using time of birth as an instrument for the mode of delivery. In other words, we estimate the local average treatment effect of c-sections on neonatal health for mothers whose mode of delivery is affected by time of birth. This allows us to interpret our estimates as causal and to focus on avoidable c-sections, as medically-indicated cesareans will be performed independently of the time of birth. Our results suggest that these non-medically indicated c-sections lead to a significant worsening of Apgar scores of approximately one standard deviation, but we do not find effects on more extreme outcomes such as needing reanimation, being admitted to the ICU or on neonatal death.

In order for our instrument to be valid, it must satisfy two conditions: first, that there is no selection of mothers with different characteristics giving birth at different times of day and, second, that giving birth during the early hours of the night only affects infant health through the increased probability of having a c-section. The comparison of maternal and pregnancy characteristics across times of day provides reassuring evidence regarding the first assumption. In order to support the validity of the exclusion restriction and, in particular, to show that variation in quality of care across time is not driving our results, we perform a robustness check restricting the analysis to births that take place during the night. Moreover, section 5 includes further supplementary tests that support our interpretation of the findings.

This paper contributes to two different strands of the literature. First, we contribute to studies on the effects of c-sections on newborn health outcomes. A large number of papers have documented a robust association between c-sections and respiratory morbidity, both at birth (Zanardo et al., 2004; Hansen et al., 2008) and in the longer-term in the form of asthma (Davidson et al., 2010; Sevelsted et al., 2015).

To the best of our knowledge, the only paper that endeavors to identify the causal impact of cesareans on later infant health is Jachetta (2015)[1]. The author uses variation in medical malpractice premia at the Metropolitan Statistical Area (MSA) level in the US as an instrument for the rate of risk-adjusted cesarean sections and finds that higher rates lead to an increase in the rate of total hospitalizations and of hospitalizations that present asthma. Although the author identifies several potential threats to the validity of the instrument, the paper is a first step towards providing evidence of the causal link between c-sections and health outcomes. We advance the existing knowledge by using a new instrument that allows us to credibly isolate the causal impact of non-medically indicated c-sections on newborn health. In particular, our setting allows us to focus on mothers that give birth in the same hospital and have similar observable characteristics, differing only in the time of delivery. Moreover, because we measure the impact on health at birth, we are able to establish a direct connection between c-sections and health outcomes.

Second, our work is also related to the literature that documents or uses time variation in the probability of having a c-section. Brown (1996) was one of the first to show that the probability of unplanned c-sections is non-uniformly distributed across time. Using data from military hospitals in the US, the author finds that cesarean sections were less likely to occur during the weekend and more likely from 6 pm to 12 am. He interprets these results as evidence that non-clinical

[1]Recent work by Jensen and Wüst (2015) and Mühlrad (2017) examines the impact of medically necessary c-sections on health for a particular group of at-risk babies: those in breech position at term. Their findings suggest positive short and long-run effects of medically indicated cesareans for this group.

variables, in particular physicians' demand for leisure, also play a role in doctors' decision-making. In our setting, we find that the probability of unplanned c-sections is higher during the early hours of the night. It is during this time that doctors appear to have a higher incentive to perform a c-section when facing ambiguous cases, as the opportunity cost in terms of time for a vaginal delivery is higher.

There is one paper that uses time variation in the probability of having a c-section to study maternal outcomes. Halla et al. (2016) use administrative data from Austria to show that the probability of a c-section birth is lower on weekends and public holidays. They use this as an instrument for mode of delivery, and find that c-sections reduce subsequent fertility and that this translates into an increase in maternal labor supply over a period of about six years. Our paper also makes use of time variation but our data allow us to use finer variation and rule out potential exogeneity problems: we study mothers in the same hospital, on the same day, but giving birth at different times. Moreover, we are also able to precisely identify and restrict our sample to non-scheduled c-sections.

The structure of the rest of the paper is as follows. In the next section we provide background information on the choice of mode of delivery, on the institutional setting and physicians' shifts, and on why we would expect to find an adverse effect of c-sections on health outcomes. The third section introduces the data, describes the variation in the c-section rate across a 24-hour cycle and presents the empirical strategy. In section 4 we show and discuss our results. Section 5 presents some robustness checks and supplementary analysis and, finally, section 6 concludes.

2.2 BACKGROUND

2.2.1 *Choice of the mode of delivery*

Cesarean sections can be performed for several reasons and at different lengths of pregnancy. First, c-sections can be scheduled in advance – also known as planned c-sections – when there are medical indications that make a vaginal delivery inadvisable. Examples of such indications include multiple pregnancies with non-cephalic presentation of the first twin or placenta previa (NICE, 2016). In principle, c-sections can also be scheduled if they are demand-determined; that is, if the mother requests to deliver via a c-section. However, in the context of public hospitals in Spain, these elective c-sections are very uncommon and are not, in fact, included in the portfolio of services offered by the public system (Marcos, 2008). In any case, we exclude scheduled c-sections from our sample as these women are likely to be different from those delivering vaginally.

If there is no scheduled c-section, an attempt of vaginal delivery begins with the onset of labor or medical induction. If an immediate threat to the life of the woman or fetus emerges, a c-section should be performed as quickly as possible (NICE, 2011). However, some indications such as dystocia (failure to progress or cephalopelvic disproportion) have a more imprecise diagnosis which leaves the door open to a more discretionary interpretation and present large variability among clinicians (Fraser et al., 1987; Barber et al., 2011). Therefore, in some cases, whether or not a c-section is needed is not obvious, and the choice between a vaginal delivery or a c-section will depend on the subjective assessment of the doctor. Unfortunately, our data does not contain the specific indication registered by the medical team to justify the c-section. However, given that emergencies should be uniformly distributed across time, we expect any observed time variation in the c-section rate to be due to indications falling in this gray area.

As Shurtz (2013) points out, a c-section is a common procedure known to be sensitive to physician incentives. Several papers have found, for example, that financial fees can influence doctors' behavior (Grant, 2009). When fees are higher for a c-section than for a vaginal delivery, physicians have a greater incentive to perform a c-section. Other studies suggest that physicians perform more c-sections as a defensive strategy reflecting a fear of malpractice lawsuits (Baicker et al., 2006; Currie and MacLeod, 2008; Jachetta, 2015). Finally, physicians have more incentives to perform c-sections when the opportunity cost of time is higher, as vaginal deliveries take longer than c-sections and thus the latter can be seen as a time-saving device (Lefèvre, 2014). We focus here on this last type of incentive given that, by performing our analysis within hospital and exploiting variation across time of day, we abstract from variations in malpractice premia and financial fees.

In particular, the average duration of vaginal deliveries among first-time mothers is around 11 hours (NICE, 2014). The first stage of established labor[2] usually lasts about 8 hours and is rarely longer than 18 hours. After that, birth is expected to take place within 3 hours of the start of the active second stage[3]. In contrast, a c-section takes much shorter; in general the average duration of this procedure is between 30 and 75 minutes (NICE, 2014). The baby is usually delivered in the first 5-15 minutes, with the remaining time being used for closing the incision (APA, 2017). Moreover, complications during this procedure are very uncommon. According to NICE (2011), c-sections increase the risk of hysterectomy (14 more per 100,000) and of cardiac arrest (15 more per 10,000). Therefore, given the low risk in terms of complications and the expected time gain, doctors may have larger incentives to perform a cesarean section when the opportunity cost of time is higher.

[2]Mothers are considered to be in the first stage of established labor when the cervix has dilated to about 4 cm (NICE, 2014).

[3]The mother is considered to be in active second stage of labor when either the baby is visible, or the full dilatation of the cervix has been accomplished and one of the following conditions is satisfied: either the mother has expulsive contractions or there is active maternal effort.

2.2.2 *Mechanisms: the impact of c-sections on newborn health*

Cesarean sections have been associated with several adverse health outcomes for newborns. Hyde et al. (2012) provide an extensive review of such findings, concluding that although further research is needed, the available evidence suggests that "normal vaginal delivery is an important programming event with life-long health consequences." More specifically, the absence or modification of a vaginal delivery has been linked to several health alterations, which they classify as either short- or long-term. In what follows we summarize some of these findings, in particular those that are more relevant to understand how c-sections might affect our outcome variables. Before doing so, however, it should be noted that any negative health effect of c-sections is outweighed by its benefits when there is a clear medical necessity. For instance, in the case of breech babies, Jensen and Wüst (2015) find that c-sections decrease the probability of having low Apgar scores and the number of doctor visits in the first year of life. More generally, cesareans save lives when severe complications arise during birth.

The adverse short-term outcomes with which c-sections have been associated include the increased risk of impaired lung functioning and altered behavioral responses to stress. With regard to the former, one of the most common causes of respiratory distress among newborns is transient tachypnea or the presence of retained lung fluid. While in the amniotic sac, a baby's lungs are filled with amniotic fluid, but during labor the baby releases chemicals which, together with the pressure of the birth canal on the baby's chest, help expel the amniotic fluid from their lungs. This process does not occur when babies are born by cesarean section, such that the presence of fluid in their lungs after birth is more common. Moreover, catecholamines, one of the chemicals released by the fetus during labor, are also correlated with muscle tone and excitability. Otamiri et al. (1991) find that babies born by cesarean section responded worse to neurological tests a few days after birth. In our setting, we can proxy the impact of c-sections on these outcomes by looking at Apgar scores at minute 1 and 5 after birth, which capture, among other aspects, respiration, reflexes and muscle tone. Severe effects, in particular serious respiratory morbidity, could also be reflected in increased need for assisted ventilation or ICU admission (Grivell and Dodd, 2011).

In the longer-term, cesarean births have also been associated with a higher risk of asthma (Sevelsted et al., 2015). While one possible mechanism is change in infant microbiome as a result of not passing through the birth canal, Hyde et al. (2012) also highlight that altered lung functioning at birth may lead to the development of future respiratory problems. Finally, there is evidence that the reduction in excitability among cesarean newborns may be a sympton of further alterations in the programming of the central nervous system, as affected by the catecholamine surge at birth (Boksa and Zhang, 2008). These findings generally suggest that any health worsening at birth we detect may have long-lasting consequences.

2.2.3 *Institutional setting*

2.2.3.1 Childbirth in Spanish public hospitals

In Spain, maternity care coverage is universal under the provision of the Spanish National Health Service. Antenatal and postnatal care for women are mainly provided at local health centers by midwives, while deliveries are supervised in hospitals by teams of both midwives and obstetricians. Expectant women do not have a pre-assigned doctor or midwife for the delivery. Rather, they are assigned to the professional available at the time of admission to the hospital. During labor, women are assisted by midwives who monitor the baby, check how labor is progressing, and call a doctor if they notice any issues. If no complications arise, midwives might manage the whole delivery. However, the obstetrician is in charge of any instrumented assistance and makes decisions regarding the mode of delivery.

Women may opt for private care, but most deliveries – 8 out of 10 births – take place under the public health system (Ministerio de Sanidad, Servicios Sociales e Igualdad, 2015). Pregnant women are in general assigned to give birth at the hospital that is closest to their residence. In big cities where there are several public hospitals, mothers can request a change in the assigned hospital through an administrative procedure. However, hospitals in our sample are located in medium-size towns in which there are no other public hospitals.

In the year 2014, the c-section rate in the public health system was 22.1%, lower than the 25.4% rate of the whole sector, combining both public and private hospitals (*ibid.*). It is important to note that within the public system, obstetricians' wages are independent of the method of delivery used or the number of c-sections performed.

2.2.3.2 Physicians' shifts

In our setting, the typical work shift for a doctor is from 8 am to 3 pm; night shifts are covered by doctors that are on duty and must stay in the hospital for 24 hours (from 8 am to 8 am next morning). All doctors younger than 55 are required by law to work these longer shifts (Ministerio de Trabajo y Asuntos Sociales, 1997). When doctors are on duty, they provide assistance in (relatively uncommon) gynecological emergencies, occasionally monitor mothers' health after birth, and are present in the labor room when decisions regarding a delivery are made, or if complications arise. Midwives, on the other hand, work 12-hour shifts (from 8 am to 8 pm).

For all of the hospitals in our sample, there are at least two obstetricians and two midwives on duty during the night, and each doctor assists on average between 1 and 2 deliveries per night. During these times, each delivery thus accounts for a major part of a doctor's duties. Although in our setting doctors cannot leave the hospital while they are on duty, beds are available to rest when there is no emergency or complication that requires their presence (Ministerio de Sanidad y Política Social, 2009).

2.3 DATA AND METHODS

2.3.1 *Description of the data*

Our data consists of all 6,163 birth records from four public hospitals in different Autonomous Regions in Spain during the years 2014-2016[4]. The characteristics of the hospitals in our sample are comparable to that of the majority of public hospitals in Spain, in particular with regard to the volume of births attended per year (between 300 and 1500). In terms of c-section rates, three of the four hospitals are in the left tail of the distribution, while one is just at the mode, with a c-section rate around 21%. This comparison can be found in figure 2.A.1 in the appendix.

Each birth registry contains information on the mother's characteristics (age, nationality, education, marital status, etc.), on the pregnancy, on the type of birth (planned cesarean, unscheduled cesarean, eutocic delivery, etc.), on medical interventions during labor, on a series of medical indicators collected before, during, and after the delivery, on the newborn (birth weight, Apgar scores, etc.), and on the date and time of birth. Table 3.A.1 shows some summary statistics of the variables of interest[5]. In our data, 5% of women delivered via a planned c-section, more than 11% via an unplanned c-section, and 68% had an eutocic

[4] Data collection was approved and financed by the Spanish Ministry of Health under the Strategy for Assistance at Normal Childbirth in the National Health System (PI/01445).

[5] For comparison, in table 2.A.2 we show descriptive statistics of the coincident variables reported in the Spanish National Statistics Institute birth registries for all births that took place in hospitals in Spain in the years 2014-2015. We see a slightly higher proportion of non-Spanish women in our data and also less multiple pregnancies, but similar characteristics in terms of age, gestational length or birth weight.

delivery, that is, a vaginal delivery without other interventions (i.e. spatula, forceps, or vacuum). Vaginal deliveries with such interventions represent around 15% of the sample. We eliminate non-single births, planned c-sections and breech vaginal babies[6]: our final sample consists of 5,783 observations.

Our main outcome variables are Apgar scores at minutes 1 and 5 after birth. These result from the examination of the health status of the newborn performed by the midwife or the pediatrician one and five minutes after birth, respectively (AEPED, 2014)[7]. In particular, they assess and grade between 0 and 2 points each of the following aspects: appearance (skin color), pulse (heart rate), grimace (reflex irritability), activity (muscle tone), and respiration. These variables thus take values between 0 and 10. We study both the levels of these scores and also the probability of the scores being below different thresholds. We also look at whether the newborn needed reanimation (assisted ventilation), whether they were admitted to the intensive care unit, and at the event of neonatal death.

Some other medical variables included in our analysis need further clarification. Besides the outcome variables presented above, another one of interest is the umbilical cord pH, which is an indicator of fetal distress. A sample of blood from the umbilical cord artery is collected after cord clamping, and the levels of pH are measured. There is some variation in the literature in what is considered the range of normal values for this outcome, with thresholds for acidemia (low pH) spanning from 7 to 7.20 (Malin et al., 2010). In our analysis we consider thresholds of 7.20, 7.15,

[6]Breech vaginal babies – that is, babies that were in breech position and were born vaginally – are a rare case: we only have 8 of those in our sample. This is because attending such type of birth requires special caution and expertise (American College of Obstetricians and Gynecologists, 2006) – most fetus in breech position are delivered by planned c-section. Therefore, these kind of births are not a plausible counterfactual for unplanned cesareans.

[7] In general, Apgar scores can be determined by a pediatrician, a midwife or a nurse present in the labor room – this depends mainly on the routines of each hospital. In the hospitals in our sample, this task is normally assigned to midwives.

and 7.10. A related variable is the fetal scalp pH or intrapartum pH, which is a measure of fetal distress during labor, before birth. In this case, the pH is measured from a sample collected from the baby's head when it becomes visible. Too low values of this variable – in particular, pH lower than 7.20 – suggest that the baby is not getting enough oxygen, and thus a cesarean section might be necessary (SEGO, 2005). Finally, one relevant control we include in our preferred specifications is obstetric risk. This is recorded by the medical professionals who prepared our data, and defined as a dummy variable that takes value one if, during pregnancy, some risk factors were detected that could lead to an adverse pregnancy outcome[8].

2.3.2 *Variation in the c-section rate by time of day*

Figure 3.1a shows the c-section rate at different times of day for our sample of public hospitals in Spain. We can observe that the distribution of unscheduled c-sections by time of birth is not uniform. The proportion of women that deliver via an unplanned c-section is higher in the early hours of the night (from 11 pm to 4 am), and much lower during the remaining hours of the night and the rest of the day. This pattern is not matched by either the total number of births or the number of vaginal births (see figure 2.A.2 in the appendix). More importantly, this variation is not driven by differences in maternal or pregnancy characteristics of the deliveries that take place at different times of day. In the next section, Table 2.1 confirms the balance of a very large set of mother and pregnancy characteristics between women delivering in

[8] More specifically, obstetric risk was defined as the presence during pregnancy of one or more of the following factors that increase the chance of an adverse pregnancy outcome: cholestasis, chorioamnionitis, 486 diabetes insulin and non-insulin dependent, chronologically prolonged pregnancy, multiple pregnancy, hellp syndrome, hypertension, isoimmunization in pregnancy, stained amniotic fluid, fetal malformation, uterine malformation, fetal malposition, myomectomy, oligoamnios, previous preterm labor, placenta praevia, plyhydramnios, preeclampsia, premature rupture of membranes, siphylis, toxoplasmosis, previous c-section, repeated abortions, previous miscarriages, anteparturm alteration of fetal wellbeing.

the early hours of the night and during the rest of the day. As we will discuss in further detail, this allows us to use this exogenous variation as an instrument for mode of delivery.

We are not the first to document this early night spike in unscheduled c-section deliveries. For example, Fraser et al. (1987), Brown (1996), and Spetz et al. (2001) show an increase in the probability of a c-section at the end of the day up until midnight, and Hueston et al. (1996) documents a peak in the unplanned c-section rate between 9 pm and 3 am. These authors have interpreted these evening or night peaks as evidence that convenience and doctors' demand for leisure influence the timing and mode of delivery. Similarly, several studies find that the probability of a c-section increases when doctors can go to sleep or return home after the birth, likely linked to the fact that cesarean sections require on average less total time devoted to the patient (Klasko et al., 1995; Spong et al., 2012).

This explanation is consistent with the time pattern that we observe in our data. Given the medical shift structure and the larger time-cost of surveillance implied by vaginal deliveries, doctors' incentives to perform c-sections in ambiguous cases may vary by time of day. In particular, we expect doctors to have a larger incentive to perform c-sections in the early hours of the night. By this time, on-duty doctors have already been working for more than 12 straight hours (see Figure 2.A.3 in the appendix[9]). If they perform a c-section and do not have other mothers to care for, they can expect to rest for the remainder of their shift. Alternatively, if they do not perform a c-section, they will need to occasionally monitor the vaginal delivery throughout the night. Moreover, ongoing deliveries in the early hours of the night have a high probability of falling under the responsibility of the doctor on

[9] Figure 2.A.3 shows the proportion of unplanned c-sections as a function of the number of hours worked by physicians: 0 hours corresponds to 8 am. As can be seen, the proportion of c-sections starts to increase when doctors have been working for already 12 hours, and reaches its maximum when hours worked are between 15 and 20. The proportion of unplanned c-section decreases in the last hours of their shift.

duty[10], as opposed to deliveries which begin later and are more likely to finish past the doctor's shift. These conditions would suggest that a higher share of deliveries with ambiguous indications end up as cesarean sections during the early hours of the night, as compared to the rest of the day. Consistent with this interpretation, we find that the probability of doctors performing a c-section at these times increases when there is only one ongoing delivery at the beginning of the night, that is, when the expected marginal gain of a c-section is larger[11].

Other alternative explanations are not compatible with this variation. For example, if either patient's or physician's fatigue increased the probability of c-sections, we would expect to see a higher unplanned c-section rate during the late hours rather than the early hours of the night. We can also rule out that this is driven by an accumulation of births during these hours, as we do not observe the same time pattern for the number of births (see figure 2.A.2 in the appendix). Finally, the early night spike in c-sections cannot be explained by selection of highly interventionist doctors at different times of day, as deliveries are not pre-assigned to a given obstetrician. We also provide evidence that this is not the case in Figure 2.A.4 in the appendix[12], where we show that there are no systematic differences among doctors in the probability of attending births during the early hours of the night.

2.3.3 *Identification strategy*

Our objective is to identify the causal impact of non-medically indicated c-sections on infants' health at birth. The simple comparison of women

[10]Average duration for the first stage of labor in vaginal deliveries among first-time mothers is around 8 hours (NICE, 2014), and for the second stage around 3 hours. See section 2.2.1 for more detail.

[11]Table 2.A.3 in the appendix shows that the increase in the probability of cesarean birth at the early hours of the night (from 11 pm to 4 am) is larger in days when there is only one birth at night compared to days with more than one birth.

[12]Figure 2.A.4 plots, for a small sample of births for which we know the doctor who attended the delivery, the probability of attending births during the early hours of the night across different doctors.

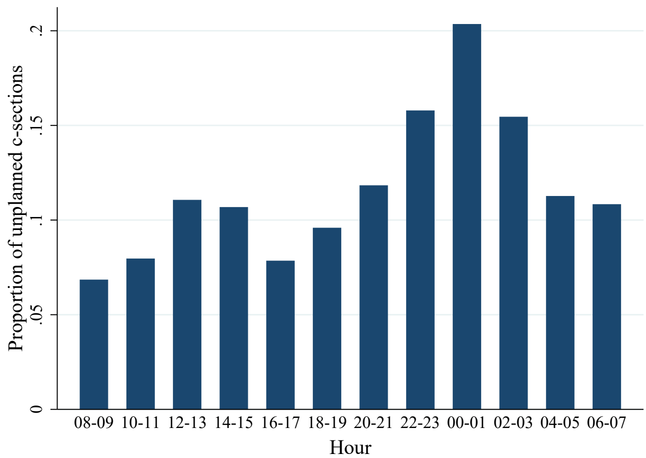


Figure 2.1: Proportion of Unplanned C-sections by Time of Day

Notes: The figure represents the proportion of unplanned c-sections by time of day over the sample of unplanned c-sections and vaginal births. Sample is restricted to single births, unscheduled c-sections and vaginal births (excluding breech vaginal babies).

who had a c-section and those who delivered vaginally is likely to suffer from omitted variable bias, as these groups likely differ in characteristics that influence the outcome variables. Table 3.A.1 in the appendix compares observable characteristics of these two types of mothers. We observe, in fact, that these mothers are significantly different in terms of several relevant aspects such as age, gestational length, obstetric risk, or educational achievement, all potentially related to the health of the newborn. There are thus reasons to be concerned that they might also differ in other characteristics we cannot observe. Moreover, a comparison of vaginal deliveries and births by c-section does not allow to identify which kind of c-section is causing whatever health effects are found, since we observe the outcomes of both medically and non-medically indicated interventions. In order to overcome these issues, we use variation in the probability of having a c-section by time of day. The purpose of the instrument is thus twofold: to compare similar women, and to precisely identify the impact of non-medically indicated cesareans.

We define a binary variable CS_i equal to one if the mode of delivery is an unplanned c-section and zero if it is a vaginal delivery (eutocic or operative). Infant health H_i refers to either Apgar scores or other measures of neonatal health. We would thus like to estimate the following equation:

$$H_i = \beta_0 + \beta_1 CS_i + \beta_2 X_i + \varepsilon_i \tag{2.1}$$

where X_i is a set of covariates that include information on mothers' personal and pregnancy characteristics. As discussed earlier, the estimation of equation (3.1) is, however, likely to provide biased estimates of β_1. To overcome this potential endogeneity, we use an IV approach, instrumenting the type of birth with an indicator for the time of day the infant is born. Therefore, our first stage is as follows:

$$CS_i = \gamma_0 + \gamma_1 earlynight_i + \gamma_2 X_i + \upsilon_i \tag{2.2}$$

where $earlynight_i$ is an indicator variable equal to 1 if woman i gives birth during the beginning of the night (from 11 pm to 4 am). We expect a positive $\widehat{\gamma}_1$ since obstetricians are more likely to initiate a c-section during these hours of the night in order to gain time for rest or leisure.

The identifying assumption is that $earlynight_i$ is not correlated with ε_i, but this assumption entails two conditions. The first is that the instrument is as good as randomly assigned. We provide suggestive evidence that this is the case by comparing personal and pregnancy characteristics of mothers who give birth between 11 pm and 4 am and those during the rest of the day in Table 2.1. Mothers are similar with respect to their age, educational level, weight and height, alcohol and tobacco consumption habits during pregnancy, gestational length, obstetric risk, weight of the newborn, or previous c-sections. The level of intrapartum pH, a measure of fetal distress during labor – a major cause of emergency c-sections – is also equivalent. Mothers are also comparable in terms of the average time that they have been in the hospital, that is, time between admission and time of birth. We find some slight differences between mothers across time of day with respect to nationality (there are slightly more non-Spanish women during the day shift) and marital status (more unmarried women during the day). However, these differences are very small in magnitude. We also find that the proportion of women whose labor was induced is higher during the early hours of the night (28.5%) compared to the rest of the day (22.6%). This is something one might expect from our institutional setting, since in the hospitals in our sample most inductions are performed in the morning and, given the average duration of labor, these women are more likely to give birth during the early hours of the night. We control in our main specification for all of these differences and perform a robustness check excluding inductions in Section 2.5.2, where we find that our conclusions still hold. Overall, we thus feel confident with the assumption that there is no selection of women into the different times that could threaten our identification.

Additionally, identification requires the exclusion restriction to hold; that is, the instrument should affect infant health only through the increased probability of having a c-section. One potential concern is that the quality of medical care could change depending on the time/shift. Although we do not have a direct measure of hospital service quality, we have some information about the doctors attending the birth for a subsample of births. In table 2.A.5 we show that the number of doctors and the proportion of male doctors is balanced across different times of day. Additionally, we provide more systematic evidence in favor of our exclusion restriction by performing the analysis using variation in the probability of having a c-section only during the night, thus holding the quality of medical care constant (see section 2.5.1).

Table 2.1: Maternal Characteristics by Time at Delivery

	Means		p-value
	Rest of the day	Early night	for difference
A. Personal characteristics			
Mother's age	31.729	31.888	0.349
Level of education			
No school	0.033	0.025	0.146
Primary school	0.254	0.262	0.563
Secondary school	0.525	0.523	0.906
University education	0.187	0.189	0.876
Non-Spanish	0.256	0.223	0.015
Single	0.019	0.009	0.017
Mother's weight	65.561	65.779	0.630
Mother's height	1.650	1.607	0.534
B. Pregnancy characteristics			
Tobacco during pregnancy	0.120	0.126	0.606
Alcohol during pregnancy	0.004	0.004	0.891
Gestation weeks	39.263	39.274	0.853
Previous c-section	0.090	0.103	0.173
Obstetric Risk	0.388	0.409	0.161
Intrapartum pH*	7.271	7.278	0.402
Birth weight	3277.356	3270.303	0.662
Induction	0.226	0.285	0.000
Time in hospital (in hours)*	9.891	10.156	0.450
Observations	4478	1305	5783

Notes: The table shows means for a set of maternal and pregnancy characteristics by time of day and the p-value for the difference between the means of the two groups. Sample is restricted to single births, unscheduled c-sections and vaginal births (excluding breech vaginal babies). Variables marked with an asterisk (*) are not available for the whole sample. Intrapartum pH is only available for a sample of births (425 observations), and time in hospital is only available for one hospital (2289 observations).

2.4 RESULTS

Tables 2.2 and 2.3 present the results for the OLS estimation of equation (3.1) for the different measures of neonatal health. In table 2.2, the first column for each outcome presents the results without controls, the second column incorporates controls for maternal characteristics, and finally the third column adds information about the pregnancy. All specifications include hospital and weekday fixed effects, the sample is restricted to single births, unplanned c-sections and vaginal deliveries, and we cluster standard errors at the hospital-shift level[13]. The results show that delivering via a c-section is associated with a significant decline of Apgar scores 1 and 5. Table 2.3 presents the results for other outcomes of neonatal health. As it can be seen, babies born by cesarean section are more likely to need reanimation and to go to the intensive care unit, but they are no more likely to die.

As explained above, these estimates are likely to be biased because mothers giving birth by c-section and vaginally are not comparable, and because we cannot identify which kind of c-section is driving the results. The results for the IV estimation of the effects of non-medically indicated c-sections on Apgar scores 1 and 5 are shown in Table 2.4[14]. The first stage F-statistics are larger than 34 for the different specifications, so following Stock and Yogo (2005) critical values with one endogenous variable and one IV (16.38), we can reject the null hypothesis that our instrument is weak. In line with our descriptive analysis, Panel B shows that births that take place between 11 pm and 4 am are around 6 percentage points more likely to be by cesarean[15].

[13] All estimations hereafter use clustered standard errors at the hospital-shift level. We show in Table 2.A.6 in the appendix that our IV results are robust to alternative standard error estimations.

[14] The full regression output for both the first and second stage can be found in tables 2.B.1 and 2.B.2 in the appendix.

[15] We have also considered alternative specifications of the IV, using dummies for single hours in the window from 11 pm to 4 am. Our second stage results are similar but the first stage is weaker, thus harming precision and raising concerns about bias

Table 2.2: OLS Results – Apgar Scores

	Apgar Score 1			Apgar Score 5		
	(1)	(2)	(3)	(1)	(2)	(3)
Unplanned CS	-0.528***	-0.524***	-0.419***	-0.219***	-0.219***	-0.142***
	(0.057)	(0.057)	(0.061)	(0.038)	(0.037)	(0.043)
Mean of Y		8.895			9.798	
Observations		5783			5781	
Maternal controls		✓	✓		✓	✓
Pregnancy controls			✓			✓

Notes: The table shows the results of OLS regressions of Apgar scores 1 and 5, respectively, on an indicator for an unplanned cesarean birth. The first column for each outcome shows the results of this regression controlling only for weekday and hospital fixed effects; in the second column maternal controls are added, and in the third column pregnancy controls are also included. Maternal controls comprise: level of education, nationality, maternal weight, height, age, and marital status. Pregnancy controls include: an indicator for previous c-section, the trimester in which prenatal care began, an indicator for obstetric risk, an indicator for preterm birth, and an indicator for induced labor. Mean of Y refers to the average of the outcome variable in the sample. The sample is restricted to single births, unscheduled c-sections, and vaginal deliveries (excluding breech vaginal babies). Standard errors (in parentheses) are clustered at the hospital-shift level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 2.3: OLS Results – Other Outcomes

	Intensive Care Unit		Reanimation		Neonatal death	
	(1)	(2)	(3)	(4)	(5)	(6)
Unplanned CS	0.137***	0.102***	0.081***	0.062***	-0.001	-0.005
	(0.016)	(0.014)	(0.014)	(0.014)	(0.002)	(0.003)
Mean of Y	0.060		0.082		0.004	
Observations	5783		5782		5783	
Maternal controls	✓	✓	✓	✓	✓	✓
Pregnancy controls		✓		✓		✓

Notes: The table shows the results of OLS regressions of different indicators of neonatal health on an indicator for an unplanned cesarean birth. The outcome variable in columns (1)-(2) is a dummy variable equal to one if the newborn was admitted to the intensive care unit; in columns (3)-(4), an indicator for whether the newborn needed reanimation (assisted ventilation), and in columns (5)-(6) an indicator of neonatal death. The first column for each outcome shows the results of this regression controlling for maternal characteristics, weekday and hospital fixed effects; in the second column pregnancy controls are also added. Maternal controls comprise: level of education, nationality, maternal weight, height, age, and marital status. Pregnancy controls include: an indicator for previous c-section, the trimester in which prenatal care began, an indicator for obstetric risk, an indicator for preterm birth, and an indicator for induced labor. Mean of Y refers to the average of the outcome variable in the sample. The sample is restricted to single births, unscheduled c-sections, and vaginal deliveries (excluding breech vaginal babies). Standard errors (in parentheses) are clustered at the hospital-shift level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

In the first row of the table below (Panel A), we observe that a c-section has a negative impact on both Apgar score 1 and Apgar score 5. The estimated effects are large and significant. In the specification with the full set of controls (column 3), an unscheduled c-section reduces Apgar score 1 by 0.992 points. This effect is around 0.9 standard deviations (1.117) and is significant at the 10% significance level. A c-section also has a negative impact on Apgar score 5. In this case the coefficient is -0.936, larger than one standard deviation (0.818) and significant at the 5% significance level.

Most of the newborns in our sample have an Apgar score 1 equal to 9 and an Apgar score 5 equal to 10 (see figure 2.A.5). We thus perform a similar analysis but using as dependent variable an indicator for having Apgar scores 1 and 5, respectively, lower than 10 (table 2.A.7), and both scores lower than 9 (table 2.A.8). Our qualitative conclusions hold, as we find that a non-medically justified c-section, as compared to a vaginal delivery, increases the probability of having Apgar scores 1 and 5, respectively, below 10 by around 25 and 40 percentage points, and the probability of having Apgar scores 1 and 5 below 9 by 36 and 19 percentage points. Finally, Figure 2.A.6 in the appendix provides an overview of the size of the coefficients for different thresholds of Apgar 1 and 5, respectively, as dependent variables. This is relevant, since decreases in Apgar scores are non-linearly related to the health of the newborn. We see a clearer pattern for Apgar scores 5: there seems to be an effect of these non-medically justified interventions on the probability of having Apgar scores lower than 10, 9 and 8, but not lower than 7 or inferior levels. Therefore, these marginal c-sections increase the probability of deviating from the perfect scores, which are the mode in our sample, but we do not see significant effects in the left tail of the distribution.

We also perform the same analysis for other infant health outcomes. Results can be found in Table 2.5. Although we might expect an effect

of the 2SLS. Results are available upon request.

Table 2.4: IV Estimation – Apgar Scores

	Apgar Score 1			Apgar Score 5		
	(1)	(2)	(3)	(1)	(2)	(3)
Panel A. 2SLS						
Unplanned CS	-1.122**	-1.147**	-0.992*	-0.956**	-0.987**	-0.936**
	(0.497)	(0.501)	(0.572)	(0.404)	(0.408)	(0.464)
Mean of Y		8.895			9.798	
Panel B. First stage						
Early night	0.073***	0.073***	0.063***	0.073***	0.073***	0.063***
	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
Observations	5783	5783	5783	5781	5781	5781
First-stage F	41.661	41.591	34.234	41.570	41.487	34.159
Maternal controls		✓	✓		✓	✓
Pregnancy controls			✓			✓

Notes: The table shows the instrumental variables estimates of the effect of an unplanned c-section on Apgar scores 1 and 5, respectively. The endogenous variable, an indicator for an unplanned cesarean birth, is instrumented with a dummy variable equal to one for births between 11 pm to 4 am (early night). Panel A shows the second stage coefficients, while Panel B displays the corresponding first stage results. First-stage F statistics are reported at the bottom of the table. The first column for each outcome shows the results of this regression controlling only for weekday and hospital fixed effects; in the second column maternal controls are added, and in the third column pregnancy controls are also included. Maternal controls comprise: level of education, nationality, maternal weight, height, age, and marital status. Pregnancy controls include: an indicator for previous c-section, the trimester in which prenatal care began, an indicator for obstetric risk, an indicator for preterm birth, and an indicator for induced labor. Mean of Y refers to the average of the outcome variable in the sample. The sample is restricted to single births, unscheduled c-sections, and vaginal deliveries (excluding breech vaginal babies). Standard errors (in parentheses) are clustered at the hospital-shift level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

on needing intensive care, reanimation, or neonatal mortality, we do not observe any significant impact.

Our IV identifies the local average treatment effect for the "marginal" women, that is, for the deliveries that are sensitive to the subjective assessment of the doctor. More specifically, we capture cases in which the time of birth affects the decision of the doctor to perform a cesarean section. We therefore focus on c-sections that are not strictly necessary in the medical sense and that are potentially avoidable surgeries. These are, in fact, arguably the most relevant from a policy point of view.

Table 2.5: IV Estimation – Other Outcomes

	Intensive Care Unit		Reanimation		Neonatal death	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. 2SLS						
Unplanned CS	0.154	0.092	0.101	0.057	0.030	0.026
	(0.103)	(0.114)	(0.114)	(0.133)	(0.031)	(0.035)
Mean of Y	0.060		0.082		0.004	
Panel B. First stage						
Early night	0.073***	0.063***	0.073***	0.063***	0.073***	0.063***
	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
Observations	5783	5783	5782	5782	5783	5783
First-stage F	41.591	34.234	41.576	34.149	41.591	34.234
Maternal controls	✓	✓	✓	✓	✓	✓
Pregnancy controls		✓		✓		✓

Notes: The table shows the instrumental variables estimates of the effect of an unplanned cesarean birth on different indicators of neonatal health. The outcome variable in columns (1)-(2) is a dummy variable equal to one if the newborn was admitted to the intensive care unit; in columns (3)-(4), an indicator for whether the newborn needed reanimation (assisted ventilation), and in columns (5)-(6) an indicator of neonatal death. The endogenous variable, an indicator for an unplanned cesarean birth, is instrumented with a dummy variable equal to one for births between 11 pm to 4 am (early night). Panel A shows the second stage coefficients, while Panel B displays the corresponding first stage results. First-stage F statistics are reported at the bottom of the table. The first column for each outcome shows the results of this regression controlling for maternal characteristics, weekday and hospital fixed effects; in the second column pregnancy controls are also added. Maternal controls comprise: level of education, nationality, maternal weight, height, age, and marital status. Pregnancy controls include: an indicator for previous c-section, the trimester in which prenatal care began, an indicator for obstetric risk, an indicator for preterm birth, and an indicator for induced labor. Mean of Y refers to the average of the outcome variable in the sample. The sample is restricted to single births, unscheduled c-sections, and vaginal deliveries (excluding breech vaginal babies). Standard errors (in parentheses) are clustered at the hospital-shift level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

We are not able to estimate the effect for women who have a clear indication for a vaginal delivery or for women who receive c-sections that are medically indicated.

If we compare the results from the IV and OLS estimations, the IV coefficients are larger in absolute terms for Apgar scores. This can be explained by the fact that with the OLS estimation we include medically indicated c-sections, which reduce fetal distress and this partially offsets the negative effects of the non-medically indicated c-sections that we find when using our instrument.

However, if we compare the results for the other outcomes (see tables 2.3 and 2.5), we observe that in this case OLS coefficients are larger and significant: c-sections are associated with an increased probability of needing intensive care and reanimation. This suggests that these medically-indicated c-sections are performed in order to assist infants in distress who need immediate support. On the other hand, the IV estimates are not significant, arguably because the effects of non-medically indicated c-sections are short-lived: in spite of the worsening in Apgar scores, we do not find substantial evidence that these negative effects translate into needing intensive care, reanimation, or increased mortality risk.

To support the interpretation that our IV identifies the effect of non-medically indicated c-sections, we provide evidence that the c-sections captured by our instrument are not correlated with indications that should predict a medically necessary cesarean. In particular, we show that, while unplanned c-sections are in general strongly correlated with fetal distress, as measured by the level of intrapartum pH, we do not see any relationship when we focus on the predicted c-sections from our first stage. This comparison can be found in table 2.A.9 in the appendix.

So far, our analysis has compared c-sections with all vaginal births. The latter comprise two main categories: eutocic births – without any instrumentation – and operative (or instrumented) vaginal deliveries, which involve the use of forceps, vacuum or spatula. Medical studies have documented a negative association between operative vaginal deliveries and infant health (American College of Obstetricians and Gynecologists, 2015). Moreover, the decision to perform these procedures is also subject to variation at the provider level (Webb, 2002). For a cleaner comparison without the potential manipulation of the control group, we perform the same analysis comparing c-sections with eutocic deliveries. We would expect the effects of non-medically indicated c-sections to be stronger if compared with this group. The results in table 2.A.10 seem to confirm this hypothesis, and we also observe a slightly stronger first stage, suggesting that physician impatience might also lead to an increased use of instrumentation in the early hours of the night.

2.5 ROBUSTNESS CHECKS AND EXTENSIONS

2.5.1 *Exclusion restriction: variation within the night*

One potential concern of our identification strategy is that the quality of medical care could differ during the day compared to the night. Hence, it may be that the negative effects that we find on infant health are not due to the increased probability of having a c-section, but rather to a reduction in the quality of care during this time.

To further investigate this issue, we perform the same IV estimation but restricting the sample to mothers who gave birth during the night. We thus use variation in the probability of having a c-section during the night, holding the quality of care constant. As before, our instrument is an indicator variable equal to 1 if the woman gives birth during the early hours of the night (from 11 pm to 4 am). The sample is restricted to deliveries taking place from 8 pm to 8 am; i.e., during the last half of physicians' shifts, when healthcare professionals in the labor room – both obstetricians and midwives – do not change.

Results for the IV estimation using variation during the night can be found in Table 2.6. Despite the smaller sample size, we again find that a c-section reduces both Apgar scores 1 and 5. The coefficients remain large and significant, in particular so for Apgar 5. We interpret these results as evidence in favor of our exclusion restriction.

2.5.2 *Excluding inductions*

The comparison of maternal characteristics in Table 2.1 showed that mothers giving birth in the early hours of the night are more likely to have had their labor induced. Inductions can be scheduled, typically because the pregnancy has gone beyond full term and labor has not

Table 2.6: IV Estimation – Apgar Scores during the Night

	Apgar Score 1			Apgar Score 5		
	(1)	(2)	(3)	(1)	(2)	(3)
Panel A. 2SLS						
Unplanned CS	-1.530*	-1.524*	-1.413	-1.511**	-1.512**	-1.535**
	(0.814)	(0.830)	(0.964)	(0.653)	(0.663)	(0.766)
Mean of Y		8.879			9.790	
Panel B. First stage						
Early Night	0.054***	0.053***	0.044***	0.053***	0.053***	0.044***
	(0.013)	(0.013)	(0.012)	(0.013)	(0.013)	(0.012)
Observations	3023	3023	3023	3022	3022	3022
First-stage F	17.217	16.619	12.812	17.144	16.537	12.760
Maternal controls		✓	✓		✓	✓
Pregnancy controls			✓			✓

Notes: The table shows the instrumental variables estimates of the effect of an unplanned cesarean birth on Apgar scores 1 and 5, respectively, for births that took place between 8 pm and 8 am. The endogenous variable, an indicator for an unplanned cesarean birth, is instrumented with a dummy variable equal to one for births between 11 pm to 4 am (early night). Panel A shows the second stage coefficients, while Panel B displays the corresponding first stage results. First-stage F statistics are reported at the bottom of the table. The first column for each outcome shows the results of this regression controlling only for weekday and hospital fixed effects; in the second column maternal controls are added, and in the third column pregnancy controls are also included. Maternal controls comprise: level of education, nationality, maternal weight, height, age, and marital status. Pregnancy controls include: an indicator for previous c-section, the trimester in which prenatal care began, an indicator for obstetric risk, an indicator for preterm birth, and an indicator for induced labor. Mean of Y refers to the average of the outcome variable in the sample. The sample is restricted to single births, unscheduled c-sections, and vaginal deliveries (excluding breech vaginal babies) that took place during the night. Standard errors (in parentheses) are clustered at the hospital-shift level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

spontaneously started, or can be unscheduled if the mother's waters break but labor does not begin (NICE, 2008). If an induction is to be scheduled, the hospitals in our sample usually plan the latter for the morning, such that after progression of labor at average pace these women are expected to give birth in the evening or during the early hours of the night.

The relation between inductions and c-sections is a question where the medical literature and medical practice seem to differ. We observe in our sample that mothers with induced labor are more likely to have a c-section (see table 3.A.1). However, the recent medical literature finds that, while c-sections are conventionally regarded as the main potential complication of inductions, inductions at full term do not increase the risk of cesarean delivery (Saccone and Berghella, 2015) or even lower it (Mishanina et al., 2014), with no increased risks for the mother and some benefits for the fetus. All in all, it seems that whether or not a c-section is needed in cases of induced labor is likely to be dependent on the assessment of the obstetrician, such that mothers having had inductions probably fall into a "gray area" where we expect doctors' decisions to be more sensitive to external factors and incentives.

In any case, even if the decision to perform a c-section on mothers with induced labor was more dependent on doctors' routines or incentives than on the health conditions of the mother and the baby, if our analysis was driven by this type of mother alone, we would not be able to disentangle the effect of c-sections from the effect of medical inductions. In our main specifications we directly control for whether labor was induced, but in Table 2.7 we also repeat our analysis excluding inductions from our sample[16]. Here we see that, despite the reduction in the number of observations, our qualitative conclusions hold: births in the early night are still more likely to end up as cesarean sections, and these have a negative and significant impact on Apgar scores. We thus conclude that, although inductions seem to make our first stage stronger as they might offer room for discretionary behavior, our findings do not depend on including them.

2.5.3 *Falsification test*

In order to lend support to the credibility of our identification strategy, we run additional "placebo" regressions using an outcome variable that is predetermined when the mother goes into labor, and thus should not

[16]The results for both the specification without inductions and the specification with only births during the night for reanimation, ICU admission, and neonatal death are consistent with those of table 2.5. Results are available upon request.

Table 2.7: Robustness Check – Excluding Inductions

	Apgar Score 1			Apgar Score 5		
	(1)	(2)	(3)	(1)	(2)	(3)
Panel A. 2SLS						
Unplanned CS	-1.747	-1.769	-1.804	-1.804*	-1.847*	-1.921*
	(1.086)	(1.104)	(1.171)	(0.931)	(0.952)	(1.011)
Mean of Y		8.952			9.828	
Panel B. First stage						
Early Night	0.037***	0.037***	0.035***	0.037***	0.037***	0.035***
	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
Observations	4369	4369	4369	4367	4367	4367
First-stage F	10.720	10.663	10.179	10.677	10.614	10.319
Maternal controls		✓	✓		✓	✓
Pregnancy controls			✓			✓

Notes: The table shows the instrumental variables estimates of the effect of an unplanned cesarean birth on Apgar scores 1 and 5, respectively, for non-induced births. The endogenous variable, an indicator for an unplanned cesarean birth, is instrumented with a dummy variable equal to one for births between 11 pm to 4 am (early night). Panel A shows the second stage coefficients, while Panel B displays the corresponding first stage results. First-stage F statistics are reported at the bottom of the table. The first column for each outcome shows the results of this regression controlling only for weekday and hospital fixed effects; in the second column maternal controls are added, and in the third column pregnancy controls are also included. Maternal controls comprise: level of education, nationality, maternal weight, height, age, and marital status. Pregnancy controls include: an indicator for previous c-section, the trimester in which prenatal care began, an indicator for obstetric risk, and an indicator for preterm birth. Mean of Y refers to the average of the outcome variable in the sample. The sample is restricted to single births, unscheduled c-sections, and vaginal deliveries (excluding breech vaginal babies) that were not induced. Standard errors (in parentheses) are clustered at the hospital-shift level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

be affected by a c-section. In particular, we analyze birth weight and weeks of gestation. The results of this analysis are reported in Table 2.8. As in previous tables, the first column for each outcome presents the results without controls, the second column incorporates controls for maternal characteristics, and finally the third column adds information about the pregnancy. The results of this exercise suggest that there is no effect of c-sections on birth weight or gestational weeks. This provides further evidence in favor of our specification.

Table 2.8: Placebo Regressions: Birth Weight and Gestational Weeks

	Birth Weight (in logs)			Gestational weeks		
	(1)	(2)	(3)	(1)	(2)	(3)
Panel A. 2SLS						
Unplanned CS	-0.023	-0.027	0.042	0.250	0.203	0.081
	(0.077)	(0.076)	(0.077)	(0.774)	(0.772)	(0.866)
Mean of Y		8.080			39.266	
Observations	5782	5782	5782	5783	5783	5783
First-stage F	41.627	41.559	34.222	41.661	41.591	35.154
Maternal controls		✓	✓		✓	✓
Pregnancy controls			✓			✓

Notes: The table shows the instrumental variables estimates of the effect of an unplanned cesarean birth on birth weight (in natural logs) and gestational weeks, respectively. The endogenous variable, an indicator for an unplanned cesarean birth, is instrumented with a dummy variable equal to one for births between 11 pm to 4 am (early night). Panel A shows the second stage coefficients, while Panel B displays the corresponding first stage results. First-stage F statistics are reported at the bottom of the table. The first column for each outcome shows the results of this regression controlling only for weekday and hospital fixed effects; in the second column maternal controls are added, and in the third column pregnancy controls are also included. Maternal controls comprise: level of education, nationality, maternal weight, height, age, and marital status. Pregnancy controls include: an indicator for previous c-section, the trimester in which prenatal care began, an indicator for obstetric risk, an indicator for preterm birth (except in the regression of gestational weeks), and an indicator for induced labor. Mean of Y refers to the average of the outcome variable in the sample. The sample is restricted to single births, unscheduled c-sections, and vaginal deliveries (excluding breech vaginal babies). Standard errors (in parentheses) are clustered at the hospital-shift level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

2.5.4 *Time of admission and time of birth*

One potential concern with using time of birth as an instrument for the mode of delivery is that, given that cesarean sections by definition shorten labor, the exact time of birth will be influenced by the type of birth itself. In other words, one might be worried about reverse causality in the first stage. We argue that any potential bias should be alleviated by the specification of the instrument not as the time of birth itself, but as a relatively wide time interval (in particular, as a dummy equal to one for births between 11 pm and 4 am). Because the instrument is defined in this way, we do not need to assume that the exact time of birth is not influenced by the mode of delivery; it suffices that any impact of the decision about the type of birth on the time interval in which the delivery takes place is negligible.

In our context, if doctors' incentive is to perform a cesarean section to ongoing deliveries early at night that they expect to end up during their shift, it will likely be to mothers that are advanced in labor. Therefore, the counterfactual to the cesarean is expected to be a vaginal birth two or three hours later[17]; that is, for most c-sections in the early night, the counterfactual vaginal birth would have probably taken place in the early hours of the night as well. As a result, the change in the probability of giving birth between 11 pm and 4 am caused by having a c-section is expected to be small.

In order to assess empirically the magnitude of the potential bias, we use information about the time of admission of mothers to the hospital, which is only available for one of the hospitals in our sample. In particular, we want to see if our results are robust to substituting our instrument with one based on the time of admission. This alternative instrument should remove concerns about reverse causality since, for unscheduled deliveries, time of admission should not be affected by mode of delivery.

First, we explore the distribution of the c-section rate as a function of time of admission (see figure 2.A.7) and find that there is a similar peak to that in figure 3.1a, in this case for mothers admitted between 2 pm and 8 pm. Therefore, we define our new instrument to be equal to one for mothers admitted during this time interval[18]. Results using this new instrument can be found in table 2.9, which follows the usual table structure. Panel B displays the coefficients of the first-stage regressions: in the third column for each outcome, which shows the results of the

[17] See an explanation of the average time of each stage of labor in section 2.2.1.

[18] Following the same logic as in our main analysis, we select the interval in which the c-section rate is above 15%.

Table 2.9: Robustness check – IV Estimation with Admission Time Instrument

	Apgar Score 1			Apgar Score 5		
	(1)	(2)	(3)	(1)	(2)	(3)
Panel A. 2SLS						
Unplanned CS	-1.554**	-1.568*	-1.601*	-0.802	-0.791	-0.793
	(0.787)	(0.815)	(0.960)	(0.578)	(0.601)	(0.712)
Mean of Y		8.861			9.869	
Panel B. First stage						
Admission time 2pm-8pm	0.077***	0.074***	0.064***	0.077***	0.074***	0.063***
	(0.022)	(0.022)	(0.021)	(0.022)	(0.022)	(0.021)
Observations	2289	2289	2289	2287	2287	2287
First-stage F	12.079	11.601	9.465	12.029	11.550	9.423
Maternal controls		✓	✓		✓	✓
Pregnancy controls			✓			✓

Notes: The table shows the instrumental variables estimates of the effect of an unplanned c-section on Apgar scores 1 and 5, respectively. The endogenous variable, an indicator for an unplanned cesarean birth, is instrumented with a dummy variable equal to one for mothers admitted to the hospital between 2 pm and 8 pm. Panel A shows the second stage coefficients, while Panel B displays the corresponding first stage results. First-stage F statistics are reported at the bottom of the table. The first column for each outcome shows the results of this regression controlling only for weekday and hospital fixed effects; in the second column maternal controls are added, and in the third column pregnancy controls are also included. Maternal controls comprise: level of education, nationality, maternal weight, height, age, and marital status. Pregnancy controls include: an indicator for previous c-section, the trimester in which prenatal care began, an indicator for obstetric risk, an indicator for preterm birth, and an indicator for induced labor. Mean of Y refers to the average of the outcome variable in the sample. The sample is restricted to single births, unscheduled c-sections, and vaginal deliveries (excluding breech vaginal babies). Standard errors (in parentheses) are clustered at the hospital-shift level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

specification with the full set of controls, we can see that mothers that arrived at the hospital between 2 pm and 8 pm were around 6.3 percentage points more likely to have a c-section. This is the same result we found for mothers giving birth between 11 pm and 4 am: they are also 6.3 percentage points more likely to have a cesarean birth. Panel A shows the 2SLS coefficients: despite the reduced sample size, we find very similar point estimates to those in table 2.4. The resemblance of these results to those in our main analysis suggests that reverse causality, in practice, does not have a large influence in our setting, and supports the validity of our instrument.

2.5.5 *Another measure of neonatal health: umbilical cord pH*

In addition to Apgar scores, reanimation, ICU admission and neonatal death, we also study the impact of cesarean sections on the pH of the umbilical cord. Although it has not been used in the economics literature, this measure of neonatal health has been widely analyzed in medical studies, and it is considered to add objective information to the Apgar score regarding the status of the newborn. Due to its objective nature, it is used to support medico-legal claims (Skiold et al., 2017). As explained in Section 2.3.1, the examination of the umbilical artery provides a measure of fetal distress. Although the relationship between pH levels and Apgar scores is not one-to-one, they are positively correlated[19]. The medical literature recommendation is to consider pH levels together with Apgar scores in order to assess the well-being of the newborn (Hannah, 1989; Malin et al., 2010).

Table 2.10 shows the results from the estimation of the impact of a c-section on the probability of the pH level being below different thresholds (7.20, 7.15 and 7.10) for the different samples: the full specification (columns 1–3), during the night (columns 4–6) and excluding inductions (7–9). This outcome was only recorded in 3 out of the 4 hospitals in our sample, and thus the number of observations is lower. All our estimates go in the same direction: c-sections increase the probability of pH levels being below the different thresholds, suggesting the presence of a negative health effect as measured by this outcome. The most consistent results are found for the pH threshold of 7.15. Our first stage F-statistic is strong for the full specification (25.58) but becomes weaker as the sample drops. Overall, these findings go in line with the previous results of a negative effect of c-sections on neonatal health.

[19]Figure 2.A.8 in the appendix shows the distributions of umbilical cord pH for infants with Apgar scores 1 above and below 9 (first panel), and for infants with Apgar scores 5 above and below 9 (second panel). We observe that the distribution of pH levels for infants with Apgar scores below 9 is shifted to the left compared to that for babies with higher scores, with this being more salient for Apgar score 5.

Table 2.10: IV estimation — Umbilical cord pH level

	Full Specification			During the Night			Excluding Inductions		
pH threshold	7.20	7.15	7.10	7.20	7.15	7.10	7.20	7.15	7.10
Panel A. 2SLS									
Unplanned CS	0.303	0.341*	0.184	1.074*	0.857**	0.307	1.004	0.947*	0.573*
	(0.250)	(0.192)	(0.122)	(0.562)	(0.415)	(0.220)	(0.671)	(0.538)	(0.333)
Mean of Y	0.221	0.102	0.042	0.212	0.100	0.044	0.216	0.096	0.039
Panel B. First stage									
Early Night		0.063***			0.042***			0.033***	
		(0.012)			(0.014)			(0.012)	
Observations		4444			2316			3403	
First-stage F		25.589			8.567			6.992	

Notes: The table shows the instrumental variable estimates of the effect of an unplanned cesarean birth on the probability of the umbilical cord pH being below different thresholds (7.20, 7.15, and 7.10), for different samples. Columns (1)-(3) use the usual full sample, columns (4)-(6) use only births during the night, and columns (7)-(9) include only non-induced births. The endogenous variable, an indicator for an unplanned cesarean birth, is instrumented with a dummy variable equal to one for births between 11 pm to 4 am (early night). Panel A shows the second stage coefficients, while Panel B displays the corresponding first stage results. First-stage F statistics are reported at the bottom of the table. All specifications include maternal and pregnancy controls, and weekday and hospital fixed effects. Maternal controls comprise: level of education, nationality, maternal weight, height, age, and marital status. Pregnancy controls include: an indicator for previous c-section, the trimester in which prenatal care began, an indicator for obstetric risk, an indicator for preterm birth, and an indicator for induced labor (except in the last three columns). Mean of Y refers to the average of the outcome variable in the sample. The sample is in all cases restricted to single births, unscheduled c-sections, and vaginal deliveries (excluding breech vaginal babies). Standard errors (in parentheses) are clustered at the hospital-shift level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

2.6 CONCLUSIONS

This paper provides new credible evidence of the adverse effects of avoidable cesarean sections on newborn health. In order to overcome potential omitted variable bias and abstract from those cases in which c-sections respond to a clear clinical indication, we make use of a novel instrument that exploits variation in the probability of receiving a c-section that is unrelated to maternal and fetal health: variation in time of birth. Specifically, we document an increase in unplanned c-sections during the early hours of the night (from 11 pm to 4 am) that is not driven by different characteristics of mothers who give birth during this time, providing us with exogenous variation in the probability of the delivery ending up in a cesarean section.

Our findings suggest that these non-medically indicated c-sections lead to a significant worsening of newborn health, as measured by Apgar scores. According to the medical literature, deterioration in these outcomes might be capturing increased respiratory problems and reduced excitability and muscle tone (Hyde et al., 2012). However, the magnitude of our estimates suggests that these c-sections lead to a decrease of just around one point in Apgar scores 1 and 5 in otherwise healthy babies – the mean Apgar scores 1 and 5 are 8.9 and 9.8, respectively. Our analysis by thresholds of Apgar scores confirms that the effects of these c-sections are limited to the higher levels of these scales; in particular, we see an increased probability of having Apgar score 5 below 10, 9 and 8. It is worth noting that previous studies find worse long-run outcomes for newborns with these levels of Apgar, compared to their siblings with perfect scores, even if these levels are not generally considered to be concerning: Oreopoulos et al. (2008) find that individuals with Apgar scores of 7 or 8 are more likely to drop out or repeat a grade, and that those with Apgar scores between 7 and 9 are also more likely to receive social assistance after age 18.

In any case, we do not find evidence that these effects translate into a significant increase in the need for reanimation or intensive care, or into increased risk of neonatal death, which is consistent with the absence of significant impacts on lower levels of Apgar scores and on low thresholds of the pH of the umbilical cord. We can thus rule out very severe impacts at birth, as well as any short-run health benefit of these avoidable interventions. This is an important contribution, given that previous studies in the medical literature documented an association between c-sections and an increased risk of serious respiratory morbidity and subsequent admission to neonatal ICU (Grivell and Dodd, 2011). Their findings are consistent with the results of our OLS estimation, suggesting that former analysis might have been capturing the underlying health status of newborns who need a medically necessary cesarean.

However, it should also be pointed out that some effects of c-sections may not be visible at birth. In particular, medical studies suggest that the exposure of newborns to the maternal vaginal microbiota is interrupted with cesarean birthing, and that this could translate into increased risk for immune and metabolic disorders in the long run (Hyde et al., 2012; Dominguez-Bello et al., 2016). Any such effect need not be reflected in any of the short-run outcomes we are able to explore in this study, which limits the conclusions we can derive from our analysis. In this paper, however, we propose a new instrument that will make possible to examine this and other channels and gather evidence to obtain a more complete understanding of the causal effect of non-medically indicated c-sections on the health of the infant and the mother in the longer run.

Our results also highlight non-financial incentives as an important factor influencing the decision-making of health care providers. Although more work is needed to clearly understand the decisions of doctors driving the observed time variation in c-section rates, we have provided some suggestive evidence that stresses the potential role of leisure incentives in the context of public hospitals, and which is consistent with the findings of previous studies. In particular, our findings suggest that doctors may be less tolerant to the time-consuming natural progression of labor during times of day when leisure incentives are more salient, and thus are more willing to perform procedures that accelerate the delivery. Along this line, our results point to the need to revise the incentives created by the shift structure and long working hours of physicians, so as to reduce avoidable interventions.

A simple back-of-the-envolope calculation can shed some light on the potential gains that could result from such reduction. The first-stage coefficient from our main specification with all controls (column 3 in table 2.4) implies that, holding all other characteristics constant, during the early hours of the night the c-section rate increases by 6.3 percentage points compared to the rest of the day. Given that the c-section rate in our sample of hospitals is 16.5%, removing these excess c-sections would lower the c-section rate by 38.1% – or equivalently, a decrease of 245 c-sections per year[20]. Taking into account that the average cost of a

[20]This figure is calculated with data from 2015, when there were 644 cesareans out of 4027 births in the four hospitals of our sample.

c-section for the Spanish public health system is 1692.97 Euros higher than that of a vaginal delivery[21], by cutting these excessive c-sections, hospitals in our sample could achieve a cost reduction of around 675,500 Euros. Applying the same logic for all births that took place in Spanish public hospitals in 2014, this would result in savings of more than 47 million Euros for the Spanish health system[22]. To give some meaning to these numbers, given that the average annual salary for a speciality doctor is 45,970 Euros[23] and there are 453 public hospitals in Spain, these savings would enable each hospital to hire more than 2 additional doctors. An increase in the number of obstetricians could help, in turn, to alleviate the need for such long working hours. Importantly, these savings could be materialized without harming neonatal health, given the absence of benefits of these avoidable c-sections.

[21]The Spanish National Health System estimated that, for the year 2014, the average cost of a cesarean section without complications was 3,739.06 Euros, while that of a vaginal birth without complications was 2,046.09 Euros. See Ministerio de Sanidad, Servicios Sociales e Igualdad (2014).

[22]The c-section rate for all public hospitals in Spain in 2014 was 22.1%. Assuming that these hospitals have a similar time variation in the c-section rate, removing the excessive c-sections of the early hours of the night would result in a c-section rate of 13.68%. Given that there were 332,252 births, the number of c-sections would decrease from 73,411 to 45,452; that is, a reduction of 27,959 c-sections per year.

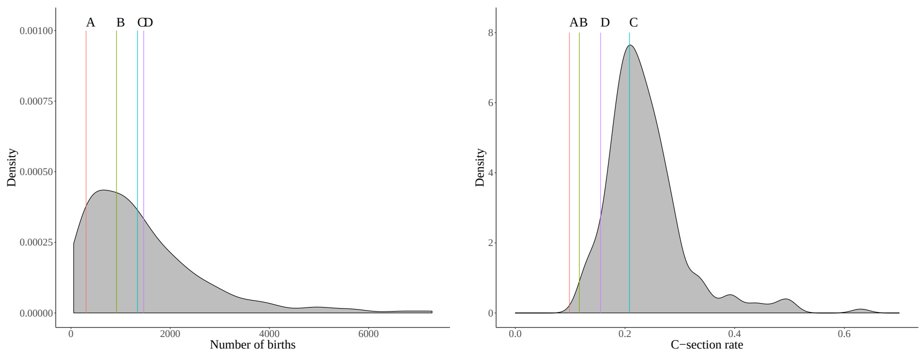
[23]Adecco Healthcare (2017)

2

APPENDIX

Appendix 2.A

Figure 2.A.1: Distribution of Number of Births and C-Section Rates in all Spanish Public Hospitals

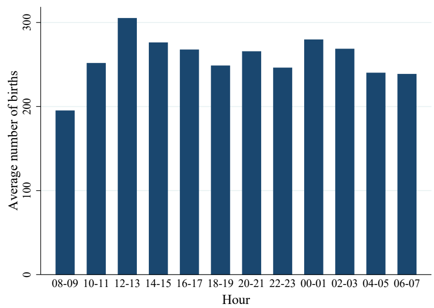


(a) Number of Births Attended in a Year

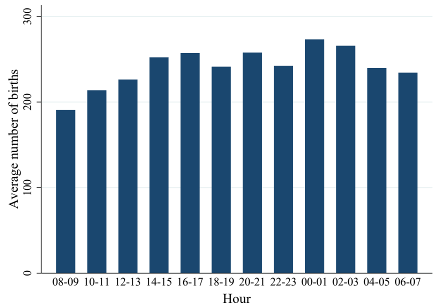
(b) C-Section Rate in a Year

Notes: Figure (a) shows the distribution of the number of births attended in one year for all Spanish Public Hospitals compared to hospitals in our sample (A, B, C and D). Figure (b) shows the distribution of c-section rates in a year for all Spanish Public Hospitals compared to hospitals in our sample (A, B, C and D). Source: our data (2015) and Estadística de Centros Sanitarios de Atención Especializada (2013).

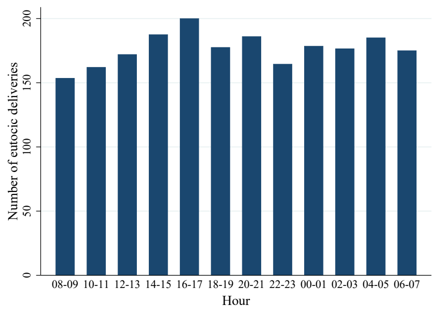
Figure 2.A.2: Distribution of Different Types of Births across Times of Day



(a) All Births



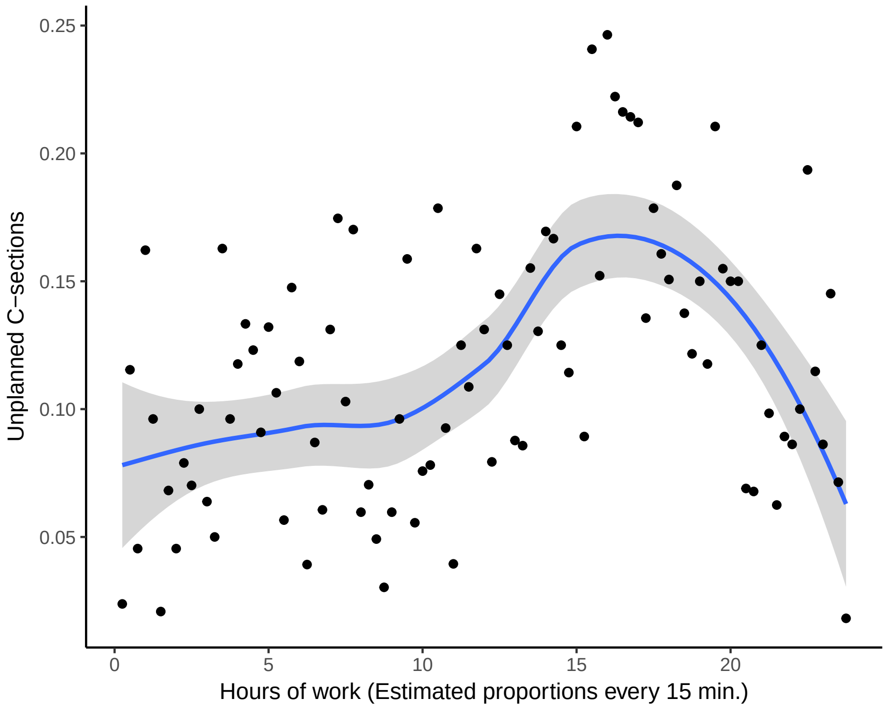
(b) Unplanned C-Sections and Vaginal Births



(c) Eutocic Deliveries

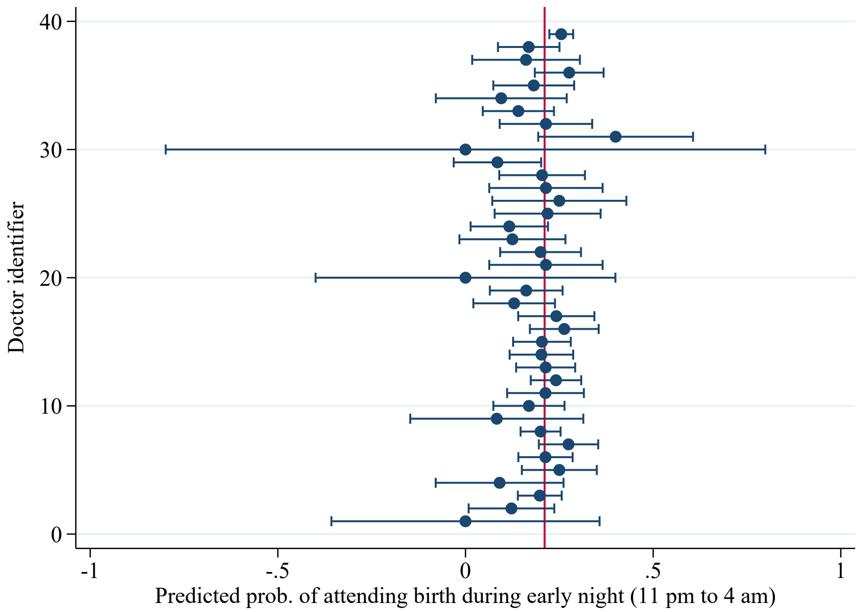
Notes: These figures represent the distribution of different types of births across times of day, grouped by intervals of two hours. Figure (a) represents the number of births per two hours using the full sample of 6,163 observations. Figures (b)-(c) use our usual sample of 5,783 observations. Figure (b) shows the number of births per two hours in this restricted sample, which includes only unplanned c-sections or vaginal births (excluding breech vaginal births), while figure (c) displays the number of eutocic deliveries.

Figure 2.A.3: Proportion of Unplanned C-Sections by Physicians' Hours Worked (Loess Estimate)



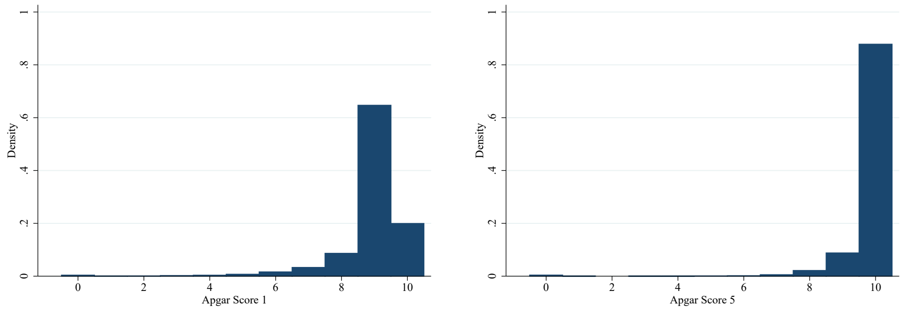
Notes: This figure shows the LOESS or local regression estimate of the proportion of observed unplanned c-sections as a function of a 24h shift, starting at 8 am and finishing at 8 am of the following day with a span of 15 minutes. The shaded area shows the 95% confidence interval.

Figure 2.A.4: Predicted Probability by Doctor of Attending Births during the Early Hours of the Night



Notes: The figure shows the probability of attending births during the early hours of the night across different doctors, for a subsample of births for which the doctor identifier was registered (N=3,018). Sample is further restricted to single births, unscheduled c-sections and vaginal births (excluding breech vaginal babies).

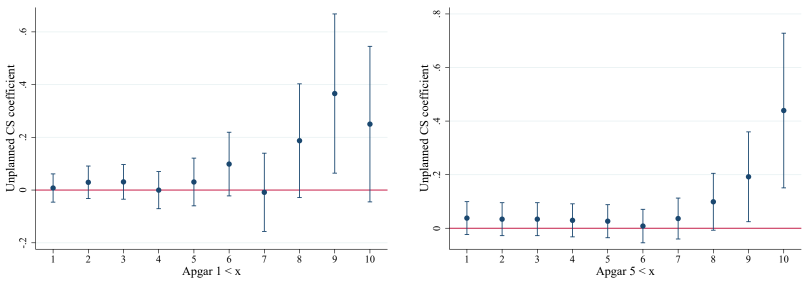
FIGURE 2.A.5: Distribution of Apgar Scores



(a) Apgar Score 1 (b) Apgar Score 5

Notes: These figures show the distribution of Apgar scores for all births. Figure (a) shows the distribution for Apgar scores at minute 1 after birth. Figure (b) shows the distribution for Apgar scores at minute 5 after birth. Sample is restricted to single births, unscheduled c-sections and vaginal births (excluding breech vaginal babies).

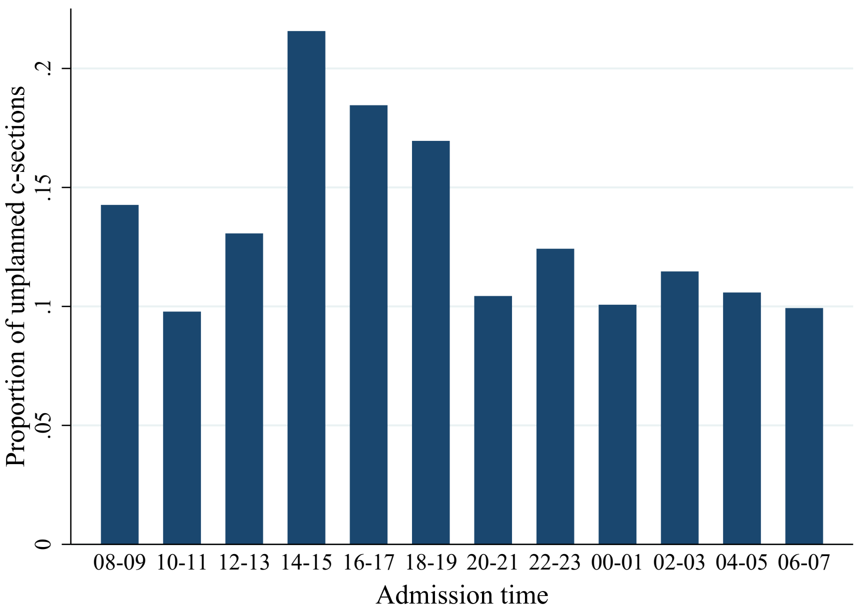
FIGURE 2.A.6: IV Coefficients by Apgar Threshold



(a) Apgar Score 1 (b) Apgar Score 5

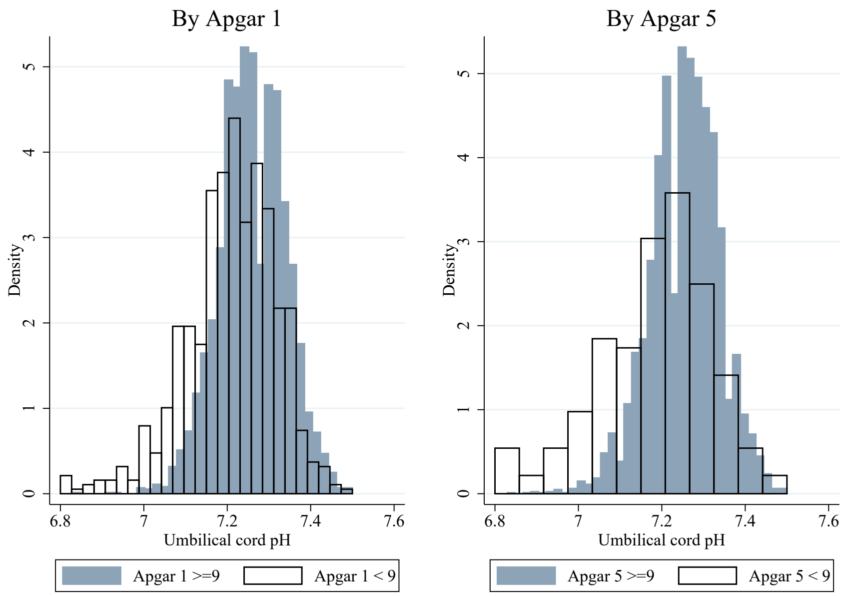
Notes: The figures show the second stage coefficients for the IV regressions of the effect of an unplanned c-section on the probability of Apgar scores being below different thresholds, in regressions with the full set of pregnancy and maternal controls. Figure (a) shows the coefficients for Apgar score at minute 1 after birth. Figure (b) shows the coefficients for Apgar score at minute 5 after birth. Sample is restricted to single births, unscheduled c-sections and vaginal births (excluding breech vaginal babies).

Figure 2.A.7: Proportion of Unplanned C-Sections by Time of Admission



Notes: The figure shows the proportion of unplanned c-sections over the sample of unplanned c-sections and vaginal births, by time of admission to the hospital. Sample is restricted to one hospital (C), single births, unscheduled c-sections and vaginal births (excluding breech babies).

Figure 2.A.8: Distribution of Umbilical Cord pH by Levels of Apgar 1 and 5



Notes: These figures show the distribution of values of umbilical cord pH by Apgar scores above or below 9. Figure (a) shows the distribution for Apgar scores at minute 1 after birth. Figure (b) shows the distribution for Apgar scores at minute 5 after birth. Sample is restricted to single births, unscheduled c-sections and vaginal births (excluding breech vaginal babies)

Table 2.A.1: Summary Statistics

	Mean	SD
A. Mother characteristics		
Mother's age	31.890	5.414
Level of education		
No school	0.032	0.175
Primary school	0.257	0.437
Secondary school	0.523	0.500
University education	0.188	0.391
Non-Spanish	0.250	0.433
Single	0.017	0.130
Mother's weight	65.715	14.536
Mother's height	1.638	2.087
B. Pregnancy characteristics		
Tobacco during pregnancy	0.122	0.327
Alcohol during pregnancy	0.004	0.062
Previous c-section	0.113	0.317
Gestation weeks	39.204	1.785
Multiple pregnancy	0.004	0.064
Obstetric Risk	0.406	0.491
Induction	0.227	0.419
C. Type of birth		
Planned c-section	0.053	0.224
Unplanned c-section	0.112	0.316
Spatula	0.007	0.084
Eutocic	0.687	0.464
Forceps	0.0141	0.118
Breech Vaginal	0.001	0.036
Vacuum	0.125	0.331
D. Newborn outcomes		
Apgar 1	8.884	1.117
Apgar 5	9.793	0.818
Birth weight (in gr.)	3267.970	519.988
Low birth weight (<2500 gr.)	0.068	0.252
Intensive care unit	0.064	0.244
Reanimation	0.084	0.277
Neonatal death	0.004	0.061
Umbilical cord pH	7.254	0.086
Intrapartum pH	7.273	0.073
Male	0.521	0.500
Observations	6163	

Notes: The table shows means and standard deviations for the outcome variables and a set of background variables for all births in our sample of public hospitals.

Table 2.A.2: Summary Statistics of All Births in Spanish Hospitals (2014-2015)

	Mean	SD
Mother's age	32.274	5.449
Non-Spanish	0.180	0.384
Gestation weeks	39.024	1.919
Multiple pregnancy	0.023	0.149
Birth weight (in gr.)	3227.344	531.320
Low birth weight (<2500 gr.)	0.069	0.253
Male	0.516	0.500
Observations	827,692	

Notes: This table shows descriptive statistics from all births in Spanish hospitals in 2014 and 2015. Source: Spanish National Statistics Institute, births microdata.

Table 2.A.3: First Stage: Busy vs. Non-Busy Nights

	(1) Single-birth nights	(2) Multiple-birth nights
Early Night	0.092***	0.054***
	(0.023)	(0.012)
Observations	1471	3733

Notes: The table shows the results of the first stage estimation on two different samples: single and multiple birth nights. The coefficients are OLS estimates of the regression of an indicator for an unplanned cesarean birth on an indicator for births during the early hours of the night (from 11 pm to 4 am). Single-birth nights are defined as days in which there is only one delivery from 8 pm to 8 am, whereas multiple-birth nights are those in which more than one delivery occurs during these times. All specifications include maternal and pregnancy controls, and weekday and hospital fixed effects. Maternal controls comprise: level of education, nationality, maternal weight, height, age, and marital status. Pregnancy controls include: an indicator for previous c-section, the trimester in which prenatal care began, an indicator for obstetric risk, an indicator for preterm birth, and an indicator for induced labor. The sample is in all cases restricted to single births, unscheduled c-sections, and vaginal deliveries (excluding breech vaginal babies). Standard errors (in parentheses) are clustered at the hospital-shift level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 2.A.4: Maternal Characteristics by Type of Birth

	Means		p-value
	Vaginal birth	Unplanned CS	for difference
A. Personal characteristics			
Mother's age	31.622	32.828	0.000
Level of education			
No school	0.033	0.022	0.126
Primary school	0.263	0.206	0.001
Secondary school	0.514	0.609	0.000
University education	0.191	0.164	0.083
Non-Spanish	0.255	0.199	0.001
Single	0.017	0.015	0.662
Mother's weight	65.312	67.830	0.000
Mother's height	1.646	1.595	0.559
B. Pregnancy characteristics			
Tobacco during pregnancy	0.120	0.134	0.277
Alcohol during pregnancy	0.003	0.007	0.089
Gestation weeks	39.320	38.863	0.000
Previous c-section	0.076	0.223	0.000
Obstetric risk	0.367	0.580	0.000
Intrapartum pH	7.288	7.245	0.000
Birth weight	3288.492	3181.038	0.000
Induction	0.214	0.431	0.000
Observations	5098	685	5783

Notes: The table shows means for a set of maternal and pregnancy characteristics by type of birth and the p-value for the difference between the means of the two groups. Sample is restricted to single births, unscheduled c-sections and vaginal births (excluding breech babies).

Table 2.A.5: Doctor Characteristics by Time of Day

	Means		p-value
	Not early night	Early night	for difference
Male doctor	0.205	0.217	0.538
Number of doctors	1.568	1.603	0.286
Observations	1827	511	2338

Notes: The table shows the mean proportion of male doctors and number of doctors by time of day and the p-value for the difference between the means of the two groups. Sample is restricted to single births, unscheduled c-sections and vaginal births (excluding breech vaginal babies).

Table 2.A.6: IV Estimation – Apgar Scores: Standard Errors Robustness

	Apgar Score 1			Apgar Score 5		
	(1)	(2)	(3)	(1)	(2)	(3)
Unplanned CS	-0.992*	-0.992*	-0.992*	-0.936**	-0.936**	-0.936**
	(0.577)	(0.572)	(0.568)	(0.461)	(0.464)	(0.465)
Mean of Y		8.895			9.798	
Observations		5783			5781	
Cluster (shift)	✓			✓		
Cluster (hospital-shift)		✓			✓	
Robust			✓			✓

Notes: The table shows the instrumental variables estimates of the effect of an unplanned cesarean birth on Apgar scores 1 and 5, respectively, comparing alternative standard error estimations. The endogenous variable, an indicator for an unplanned cesarean birth, is instrumented with a dummy variable equal to one for births between 11 pm to 4 am (early night). The first column for each outcome has clustered standard errors at the shift level; in the second column standard errors are clustered at the hospital-shift level, as in our main specification, and in the third column we estimate heteroscedasticity-robust standard errors. All specifications include maternal and pregnancy controls, and weekday and hospital fixed effects. Maternal controls comprise: level of education, nationality, maternal weight, height, age, and marital status. Pregnancy controls include: an indicator for previous c-section, the trimester in which prenatal care began, an indicator for obstetric risk, an indicator for preterm birth, and an indicator for induced labor. Mean of Y refers to the average of the outcome variable in the used sample. The sample is restricted to single births, unscheduled c-sections, and vaginal deliveries (excluding breech vaginal babies). * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 2.A.7: IV Estimation – Apgar Score < 10

	Apgar Score 1 <10			Apgar Score 5 <10		
	(1)	(2)	(3)	(1)	(2)	(3)
Panel A. 2SLS						
Unplanned CS	0.283*	0.285*	0.250	0.433***	0.445***	0.439***
	(0.157)	(0.158)	(0.182)	(0.146)	(0.147)	(0.170)
Mean of Y		0.801			0.122	
Panel B. First stage						
Early night	0.073***	0.073***	0.063***	0.073***	0.073***	0.063***
	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
Observations	5783	5783	5783	5781	5781	5781
First-stage F	41.661	41.591	34.234	41.570	41.487	34.159
Maternal controls		✓	✓		✓	✓
Pregnancy controls			✓			✓

Notes: The table shows the instrumental variables estimates of the effect of an unplanned cesarean birth on the probability of Apgar scores 1 and 5, respectively, being lower than 10. The endogenous variable, an indicator for an unplanned cesarean birth, is instrumented with a dummy variable equal to one for births between 11 pm to 4 am (early night). Panel A shows the second stage coefficients, while Panel B displays the corresponding first stage results. First-stage F statistics are reported at the bottom of the table. The first column for each outcome shows the results of this regression controlling only for weekday and hospital fixed effects; in the second column maternal controls are added, and in the third column pregnancy controls are also included. Maternal controls comprise: level of education, nationality, maternal weight, height, age, and marital status. Pregnancy controls include: an indicator for previous c-section, the trimester in which prenatal care began, an indicator for obstetric risk, an indicator for preterm birth, and an indicator for induced labor. Mean of Y refers to the average of the outcome variable in the used sample. The sample is restricted to single births, unscheduled c-sections, and vaginal deliveries (excluding breech vaginal babies). Standard errors (in parentheses) are clustered at the hospital-shift level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 2.A.8: IV Estimation – Apgar Score < 9

	Apgar Score 1 <9			Apgar Score 5 <9		
	(1)	(2)	(3)	(1)	(2)	(3)
Panel A. 2SLS						
Unplanned CS	0.380**	0.391**	0.366**	0.189**	0.192**	0.192*
	(0.158)	(0.159)	(0.183)	(0.088)	(0.089)	(0.103)
Mean of Y		0.154			0.034	
Panel B. First stage						
Early night	0.073***	0.073***	0.063***	0.073***	0.073***	0.063***
	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
Observations	5783	5783	5783	5781	5781	5781
First-stage F	41.661	41.591	34.234	41.570	41.487	34.159
Maternal controls		✓	✓		✓	✓
Pregnancy controls			✓			✓

Notes: The table shows the instrumental variables estimates of the effect of an unplanned cesarean birth on the probability of Apgar scores 1 and 5, respectively, being lower than 9. The endogenous variable, an indicator for an unplanned cesarean birth, is instrumented with a dummy variable equal to one for births between 11 pm to 4 am (early night). Panel A shows the second stage coefficients, while Panel B displays the corresponding first stage results. First-stage F statistics are reported at the bottom of the table. The first column for each outcome shows the results of this regression controlling only for weekday and hospital fixed effects; in the second column maternal controls are added, and in the third column pregnancy controls are also included. Maternal controls comprise: level of education, nationality, maternal weight, height, age, and marital status. Pregnancy controls include: an indicator for previous c-section, the trimester in which prenatal care began, an indicator for obstetric risk, an indicator for preterm birth, and an indicator for induced labor. Mean of Y refers to the average of the outcome variable in the used sample. The sample is restricted to single births, unscheduled c-sections, and vaginal deliveries (excluding breech vaginal babies). Standard errors (in parentheses) are clustered at the hospital-shift level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 2.A.9: Robustness Check: Fetal Distress and C-Sections

	Unplanned CS		Predicted CS	
	(1)	(2)	(1)	(2)
Intrapartum pH	-1.768***		0.018	
	(0.281)		(0.019)	
Intra. pH < 7.2		0.312***		-0.002
		(0.060)		(0.004)
Observations	425	425	425	425

Notes: The table shows the results of OLS regressions of all unplanned cesarean sections and the time-predicted c-sections on indicators of fetal distress. In the first two columns the dependent variable is an indicator equal to one for all unplanned c-sections, while in the last two columns the dependent variable takes the fitted values from the first-stage regression. In the first column for each outcome the explanatory variable is the level of intrapartum of fetal scalp pH, while in the second column is an indicator equal to one if the intrapartum pH is below 7.20. All specifications include maternal and pregnancy controls, and weekday and hospital fixed effects. Maternal controls comprise: level of education, nationality, maternal weight, height, age, and marital status. Pregnancy controls include: an indicator for previous c-section, the trimester in which prenatal care began, an indicator for obstetric risk, an indicator for preterm birth, and an indicator for induced labor. The sample is restricted to single births, unscheduled c-sections, and vaginal deliveries (excluding breech vaginal babies) for which we have information about the intrapartum pH. Standard errors (in parentheses) are clustered at the hospital-shift level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 2.A.10: IV Estimation – Apgar Scores: Comparing C-Sections with Eutocic Births

	Apgar Score 1			Apgar Score 5		
	(1)	(2)	(3)	(1)	(2)	(3)
Panel A. 2SLS						
Unplanned CS	-1.179***	-1.218***	-1.161**	-0.907**	-0.954**	-0.942**
	(0.448)	(0.459)	(0.514)	(0.372)	(0.382)	(0.426)
Mean of Y		8.945			9.809	
Panel B. First stage						
Early night	0.090***	0.088***	0.078***	0.090***	0.088***	0.078***
	(0.013)	(0.013)	(0.012)	(0.013)	(0.013)	(0.012)
Observations	4886	4886	4886	4884	4884	4884
First-stage F	45.329	43.974	39.192	45.222	43.852	39.102
Maternal controls		✓	✓		✓	✓
Pregnancy controls			✓			✓

Notes: The table shows the instrumental variables estimates of the effect of an unplanned cesarean birth on Apgar scores 1 and 5, respectively, compared to an eutocic birth (a vaginal birth without any instrumentation). The endogenous variable, an indicator for an unplanned cesarean birth, is instrumented with a dummy variable equal to one for births between 11 pm to 4 am (early night). Panel A shows the second stage coefficients, while Panel B displays the corresponding first stage results. First-stage F statistics are reported at the bottom of the table. The first column for each outcome shows the results of this regression controlling only for weekday and hospital fixed effects; in the second column maternal controls are added, and in the third column pregnancy controls are also included. Maternal controls comprise: level of education, nationality, maternal weight, height, age, and marital status. Pregnancy controls include: an indicator for previous c-section, the trimester in which prenatal care began, an indicator for obstetric risk, an indicator for preterm birth, and an indicator for induced labor. Mean of Y refers to the average of the outcome variable in the used sample. The sample is restricted to single births, unscheduled c-sections, and eutocic vaginal deliveries. Standard errors (in parentheses) are clustered at the hospital-shift level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Appendix 2.B

Table 2.B.1: IV Estimation – Full Regression Output Second Stage

	Apgar Score 1			Apgar Score 5		
	(1)	(2)	(3)	(1)	(2)	(3)
Unplanned CS	-1.122**	-1.147**	-0.992*	-0.956**	-0.987**	-0.936**
	(0.497)	(0.501)	(0.572)	(0.404)	(0.408)	(0.464)
Hospital B	0.188***	0.196***	0.179***	-0.080*	-0.059	-0.071*
	(0.048)	(0.049)	(0.048)	(0.041)	(0.040)	(0.039)
Hospital C	0.234***	0.262***	0.252***	0.170***	0.170***	0.178***
	(0.053)	(0.053)	(0.059)	(0.047)	(0.045)	(0.051)
Hospital D	0.481***	0.503***	0.483***	0.113**	0.152***	0.146***
	(0.058)	(0.060)	(0.062)	(0.046)	(0.047)	(0.048)
Tuesday	-0.005	-0.001	-0.003	-0.021	-0.019	-0.026
	(0.058)	(0.057)	(0.056)	(0.048)	(0.048)	(0.047)
Wednesday	0.043	0.046	0.047	0.063*	0.065*	0.062*
	(0.051)	(0.051)	(0.051)	(0.037)	(0.037)	(0.037)
Thursday	-0.026	-0.023	-0.022	-0.016	-0.015	-0.019
	(0.058)	(0.057)	(0.056)	(0.046)	(0.046)	(0.045)
Friday	0.089*	0.091*	0.093*	0.068*	0.071*	0.068*
	(0.054)	(0.054)	(0.053)	(0.039)	(0.040)	(0.039)
Saturday	0.052	0.056	0.057	0.047	0.050	0.045
	(0.055)	(0.055)	(0.054)	(0.043)	(0.043)	(0.042)
Sunday	0.004	0.007	0.010	0.005	0.007	0.009
	(0.055)	(0.056)	(0.055)	(0.045)	(0.045)	(0.044)
No studies		-0.105	-0.104		-0.088	-0.086
		(0.107)	(0.104)		(0.094)	(0.092)
Secondary school		-0.007	-0.010		0.070*	0.072**
		(0.046)	(0.046)		(0.037)	(0.036)
University education		0.060	0.052		0.089**	0.086**
		(0.048)	(0.047)		(0.038)	(0.037)
Non Spanish		0.057	0.060		0.012	0.015
		(0.042)	(0.042)		(0.032)	(0.032)
Mother weight		0.001	0.001		0.001	0.001
		(0.001)	(0.001)		(0.001)	(0.001)
Mother height		0.002	0.002		0.002	0.002
		(0.002)	(0.002)		(0.001)	(0.001)
Mother age		0.000	0.001		0.001	0.001
		(0.003)	(0.003)		(0.003)	(0.002)
Single		-0.170	-0.165		-0.189	-0.179
		(0.144)	(0.142)		(0.122)	(0.120)
Previous c-section			-0.021			0.052
			(0.105)			(0.083)
Prenatal Care 2T			-0.023			0.017
			(0.078)			(0.061)
Prenatal Care 3T			-0.024			-0.015
			(0.170)			(0.097)
Obstetric risk			-0.019			0.032
			(0.037)			(0.031)
Preterm			-0.468***			-0.452***
			(0.138)			(0.131)
Induction			-0.104			-0.014
			(0.076)			(0.063)
Constant	8.783***	8.699***	8.748***	9.831***	9.660***	9.696***
	(0.065)	(0.152)	(0.150)	(0.053)	(0.129)	(0.126)
Observations	5783	5783	5783	5781	5781	5781
First-stage F	41.661	41.591	34.234	41.570	41.487	34.159
Maternal controls		✓	✓		✓	✓
Pregnancy controls			✓			✓
Mean of Y		8.895			9.798	

Notes: The table shows the instrumental variables estimates of the effect of an unplanned cesarean birth on Apgar scores 1 and 5. The endogenous variable, an indicator for an unplanned cesarean birth, is instrumented with a dummy variable equal to one for births between 11 pm to 4 am (early night). The omitted category for the hospital indicators is Hospital A; for weekdays, it is Monday; for levels of education it is primary school, and for trimester in which prenatal care began it is the first trimester. The sample is restricted to single births, unscheduled c-sections, and vaginal deliveries (excluding breech vaginal babies). Standard errors (in parentheses) are clustered at the hospital-shift level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 2.B.2: IV Estimation – Full Regression Output First Stage

	Apgar Score 1			Apgar Score 5		
	(1)	(2)	(3)	(1)	(2)	(3)
Early Night	0.073***	0.073***	0.063***	0.073***	0.073***	0.063***
	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
Hospital B	-0.011	-0.006	0.006	-0.011	-0.006	0.006
	(0.012)	(0.012)	(0.011)	(0.012)	(0.012)	(0.011)
Hospital C	0.064***	0.056***	0.065***	0.064***	0.056***	0.065***
	(0.012)	(0.013)	(0.012)	(0.012)	(0.013)	(0.012)
Hospital D	0.024*	0.036**	0.044***	0.024*	0.036**	0.044***
	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)
Tuesday	0.010	0.009	0.004	0.010	0.009	0.004
	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)
Wednesday	-0.010	-0.011	-0.013	-0.010	-0.011	-0.013
	(0.015)	(0.015)	(0.014)	(0.015)	(0.015)	(0.014)
Thursday	-0.004	-0.004	-0.005	-0.004	-0.004	-0.005
	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)
Friday	-0.002	-0.002	-0.007	-0.002	-0.002	-0.007
	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)
Saturday	0.015	0.014	0.007	0.015	0.014	0.007
	(0.016)	(0.016)	(0.015)	(0.016)	(0.016)	(0.015)
Sunday	-0.009	-0.009	-0.009	-0.009	-0.009	-0.009
	(0.016)	(0.016)	(0.015)	(0.016)	(0.016)	(0.015)
No studies		-0.012	-0.005		-0.012	-0.005
		(0.022)	(0.022)		(0.022)	(0.022)
Secondary school		0.026**	0.029***		0.026**	0.029***
		(0.011)	(0.010)		(0.011)	(0.010)
University education		-0.002	0.007		-0.002	0.007
		(0.013)	(0.013)		(0.013)	(0.013)
Non Spanish		0.008	0.007		0.008	0.007
		(0.011)	(0.011)		(0.011)	(0.011)
Mother weight		0.001***	0.001***		0.001***	0.001***
		(0.000)	(0.000)		(0.000)	(0.000)
Mother height		-0.001*	-0.001***		-0.001*	-0.001***
		(0.001)	(0.000)		(0.001)	(0.000)
Mother age		0.004***	0.002**		0.004***	0.002**
		(0.001)	(0.001)		(0.001)	(0.001)
Single		0.003	0.014		0.003	0.014
		(0.031)	(0.032)		(0.031)	(0.032)
Previous c-section			0.151***			0.151***
			(0.021)			(0.021)
Prenatal Care 2T			0.001			0.001
			(0.021)			(0.021)
Prenatal Care 3T			-0.044			-0.044
			(0.028)			(0.028)
Obstetric risk			0.031***			0.031***
			(0.009)			(0.009)
Preterm			0.143***			0.143***
			(0.026)			(0.026)
Induction			0.110***			0.110***
			(0.012)			(0.012)
Constant	0.077***	-0.127***	-0.121***	0.076***	-0.127***	-0.121***
	(0.014)	(0.034)	(0.033)	(0.014)	(0.034)	(0.033)
Observations	5783	5783	5783	5781	5781	5781

Notes: The table shows the first stage coefficients of the IV regression of the effect of an unplanned cesarean birth on Apgar scores 1 and 5. The endogenous variable, an indicator for an unplanned cesarean birth, is instrumented with a dummy variable equal to one for births between 11 pm to 4 am (early night). The omitted category for the hospital indicators is Hospital A; for weekdays, it is Monday; for levels of education it is primary school, and for trimester in which prenatal care began it is the first trimester. The sample is restricted to single births, unscheduled c-sections, and vaginal deliveries (excluding breech vaginal babies). Standard errors (in parentheses) are clustered at the hospital-shift level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

3

THE LONG-RUN EFFECTS OF CESAREAN SECTIONS

Joint with Ana Costa-Ramón (UPF), Mika Kortelainen (VATT Institute for Economic Research and University of Turku) and Lauri Sääksvuori (National Institute for Health and Welfare)

3.1 INTRODUCTION

There is little doubt that prenatal health and early childhood circumstances can have long-term effects on mortality, morbidity and human capital development. The theory of the developmental origins of adult health and disease has proven to describe a surprisingly general phenomenon. The effects of prenatal health conditions and early-life events extend to a wide spectrum of educational, cognitive, behavioral and demographic outcomes (Almond et al., 2018).

In human development, the transition from fetal to newborn life at birth is an abrupt event that represents major physiological challenges for the neonates. There is accumulating evidence that many medical and operative interventions at birth are associated with long-term health. Most notably, cesarean delivery for low-risk pregnancies is associated with a wide variety of adverse short- and long-term health outcomes. However, the causal nature of these relationships has received little attention.

The most prominent mechanism thought to mediate the long-term effects of cesarean sections on health and disease emphasizes the importance of early exposure to a diverse range of microbes that adjust the human immune system to appropriately react to extrauterine environment. This general class of mechanisms is often dubbed either as the hygiene hypothesis (Strachan, 1989) or the old friends hypothesis (Scudellari, 2017). According to these hypotheses, children born by cesarean section lack the beneficial exposure to their mother's vaginal microbiome and are more prone to develop immune-mediated diseases.

Cesarean section is the most commonly performed major surgery in many countries. Understanding the consequences of cesarean sections on later-life health and human capital development is important from a number of perspectives varying from clinical decision making to economic and health policy. The rapidly growing incidence of cesarean sections across the globe suggests that even small increases in mortality and morbidity due to C-sections would lead to large reductions in life expectancy and substantial losses of human welfare.[1]

This paper provides new evidence on the effect of potentially avoidable cesarean sections on several relevant health outcomes. To identify the causal effect and abstract from cases where C-sections respond to a clear medical indication, we exploit variation in physician demand for leisure. We show that the probability of unscheduled C-section increases substantially during the normal working hours (8am – 4pm) on working days that precede a leisure day. Importantly, we find that these excess C-sections are not driven either by selection of different

[1] Cesarean section rates have increased in the US from 20.7 percent in 1996 to 32.9 percent in 2009 (Currie and Macleod, 2017). In OECD countries, the rate of cesarean sections has increased from 20 percent in 2000 to 25 percent in 2013 (OECD, 2013). Currently, the highest rates of cesarean sections are reported in many of the world's most populous countries including among others China (41.3 percent in 2016) and Brazil (55.6 percent in 2015). Boerma et al. (2018) review the disparities in C-section use around the world.

mothers giving birth at these times or by advancing births that would have been cesarean deliveries in any event.

Using fine grained data on birth times and intrapartum diagnoses, we show that the increased likelihood of cesarean sections during the normal working hours on days that precede a leisure day is coupled with the increased use of more discretionary diagnoses. Moreover, we observe that physician demand for leisure does not affect mothers who are in the medical profession. Our data lend substantial support for the contention that the excess numbers of unplanned cesarean deliveries observed during the normal working hours on days that precede a leisure day are largely driven by physician incentives. We use this time variation as an instrument for C-section. We provide a detailed discussion and numerous robustness checks to support the validity of the required identification assumptions.

We investigate the effects of cesarean sections on infant and children outcomes using a comprehensive and precise administrative data resource which includes birth and health records for all children born in Finland between 1990 and 2014. We follow entire birth cohorts from birth to teenage years and use detailed diagnosis data to study the causal effects of cesarean sections on children's health. We focus on outcomes whose onset is hypothesized to be influenced by cesarean delivery: asthma and other atopic diseases, type 1 diabetes and obesity. These are among the most common chronic conditions in childhood (Torpy, 2010).[2]

[2]Understanding and quantifying the potential contribution of C-sections to the development of these diseases is not limited to medical practice and health policy. Chronic health conditions cause an immense financial burden to households and public health care financing. The total cost of asthma in the working age population was estimated to be $24.7 billion during 1999-2002 in Europe (Global Asthma Network, 2018). The two other atopic diseases we investigate imply high costs: atopic dermatitis has been estimated to cost at least $5.3 billion (in 2015 USD) in the US (Drucker et al., 2017). The estimated annual cost of allergic rhinitis is in the range of $2–5 billion (in 2003 USD) (Reed et al., 2004). Type 1 diabetes has been found to cost $14.4 billion a year in medical costs and lost income in the US (Tao et al., 2010). Finally, childhood obesity, which has been on the rise in recent years, has been calculated to imply $19,000 per child in lifetime medical costs in the US (Finkelstein

Our instrumental variable estimates suggest that avoidable C-sections increase the probability of asthma diagnosis from early childhood onward. This effect is clinically and economically relevant. However, we do not find consistent evidence that cesarean sections affect the probability of developing atopic diseases at large, type 1 diabetes or obesity.

We complement our instrumental variables estimates using a differences-in-differences model with family fixed effects that compares the health gap between siblings in families where the second child was born by unplanned C-section with the health gap between siblings who were born by vaginal delivery. The results from our supplementary empirical strategy support our main findings. These estimates suggest that unplanned C-sections increase the risk of childhood asthma and enable to rule out meaningful effects on other atopic diseases, type 1 diabetes and obesity. We provide several sensitivity checks that suggest that the effect on asthma is unlikely to be explained by negative selection.

Our results are consistent with the hypothesis that the mode of delivery may influence the development of the immune system and have long-term effects on health and disease. However, our results paint a more nuanced picture about the long-term effects of cesarean deliveries than existing evidence based mostly on associations. Our findings suggest that C-sections cause a much narrower spectrum of diseases than currently hypothesized and call for a careful analysis on the relationships between the delivery mode and long-term health.

Our paper relates to an important literature estimating the effects of early interventions on long-term health and human capital development. Moreover, we contribute at least in three ways to a nascent economics literature on the effects of treatment choices at birth. First, we investigate the long-term effects of unplanned C-sections on children. To evaluate the costs and benefits of C-sections, it is crucial to investigate long-term effects, as potential alterations of the immune system and long-run consequences of C-sections are not necessarily visible at birth

et al., 2014).

and in early childhood. Moreover, we report age-by-age estimates for entire cohorts from birth to teenage years and provide evidence about the effects of early life events during the middle childhood, thus expanding our knowledge about the "missing middle" years.[3] Existing papers investigating the effects of potentially avoidable C-sections have concentrated on neonatal outcomes or short-term effects.[4] Costa-Ramón et al. (2018) investigate the effects of cesarean sections on neonatal health using time variation in unplanned C-section rates. Card et al. (2019) study the short-term health effects of hospital delivery practices using relative distance from a mother's home to hospitals with high and low C-sections rates.[5]

Second, we study the effects of discretionary unplanned C-sections that could potentially be avoided, while existing papers have not been able to separate planned (elective) and unplanned C-sections or have concentrated on C-sections with a clear medical indication. Hannah et al. (2000), Jensen and Wüst (2015) and Mühlrad (2017) show that breech babies can benefit from C-section delivery. However, these results concern medically necessary C-sections in a specific high-risk group and do not readily generalize to cesarean deliveries in general or for avoidable

[3]Almond et al. (2018) discuss that, due to data availability, most of the literature analyzes the effect of early life events on birth or adult outcomes. This implies that we have little knowledge about how developmental trajectories are affected by policies or shocks experienced over the life course. They refer to this gap in the literature as the "missing middle".

[4]To our knowledge, the only paper looking at longer-term effects is by Jachetta (2015), who explores the relation of cesarean delivery with hospitalizations using regional variation in medical malpractice insurance premia in the US as an instrument for C-sections. However, the instrument used in that paper does not necessarily allow for credible causal inference, since the author finds that higher premia also predict delayed prenatal care, lower birth weight and reduced gestational age.

[5]A few papers have also examined the effects of cesarean sections on mothers. Halla et al. (2016) study the effects of C-sections on fertility and maternal labor supply. Tonei (2019) studies the impact on mental health for mothers with breech babies who undergo a C-section. Our findings on children health complement these maternal results and contribute to obtaining a more complete picture of the effect of cesarean sections.

unplanned C-sections, in particular. While C-sections are often life-saving at the top of the risk distribution (Currie and Macleod, 2017), more evidence is required about the effects of discretionary C-sections that could be potentially avoidable.

Third, to evaluate causal effects of C-sections, we use two different identification strategies based on somewhat different assumptions. Our instrumental variable strategy builds on previous work using time variation in C-section rates in combination with high-quality administrative data. Moreover, we employ a differences-in-differences research design that has not been used in previous papers on C-sections. In addition, for both methods we provide several pieces of evidence that support the credibility of the identification assumptions. Thus, by using two different strategies, we hope to provide more reliable evidence on the causal effects of avoidable unscheduled interventions at birth on children both in the short and long run.

The paper is structured as follows. Section 3.2 provides background information about the biological mechanisms hypothesized to mediate the effects of mode of delivery on infant outcomes, about the different types of cesarean sections, and about the institutional context of our analysis. Section 3.3 introduces the data, provides key descriptive statistics and lays out our econometric approach. Section 3.4 reports our main results. Section 3.5 presents robustness checks and additional evidence to support our main conclusions. The last section concludes.

3.2 BACKGROUND

3.2.1 *Mechanisms*

A large body of literature documents the developmental origins of health and disease. The process of labor can be seen as one crucial step in adaptation to the extrauterine environment. The prevailing evidence highlights the role of vaginal delivery as an important early programming event with potentially life-long consequences (Hyde et al., 2012). While there is strong consensus that medically indicated cesarean sections decrease the risk of fetal death at birth, the absence or modification of vaginal delivery has been linked to several adverse health outcomes and anomalies in human development. In the following, we summarize some of the most widely acknowledged findings to understand how C-sections might have long-lasting effects on health and human development.

It is well-recognized that early exposure to microbes is necessary to train the human immune system to react appropriately to environmental stimulation. The original formulation of the theory, dubbed as the hygiene hypothesis, states that the lack of early childhood exposure to infectious agents and symbiotic microbes increases susceptibility to multiple autoimmune diseases by suppressing the natural development of the immune system (Strachan, 1989). Lately, refinements to the original formulation, known as the old friends hypothesis, have challenged the role of infectious pathogens and highlight the importance of early exposure to a diverse range of harmless microbes to strengthen the human immune system and combat the threat of environmental pathogens (Scudellari, 2017).

Mode of delivery may affect early exposure to microbes through several channels. First, bacteria from the mother and the surrounding environment colonize the infant's gut during birth (Neu and Rushing, 2011). Exposure to the maternal vaginal microbiota is interrupted in a cesarean birth and externally derived environmental bacteria play an important role for the infants' intestinal colonization. Consequently, infants delivered by C-sections acquire a microbiota that differs from that of vaginally delivered infants (Dominguez-Bello et al., 2016). Second, the transfer of microbiota continues through breastfeeding after birth. Breast milk contains a number of bioactive components that can have an important impact on infant's microbiota composition and health (Collado et al., 2015). The negative association between cesarean sections and the initiation of breastfeeding provides an additional mechanism to explain the differences in microbiota by type of birth (Prior et al., 2012).

The potential biological mechanisms are consistent with the reported associations between cesarean delivery and adverse infant outcomes. These studies relate cesarean deliveries to a marked increase in the susceptibility of multiple immune and metabolic conditions. Even though cesarean deliveries have been associated with a broad array of immune-mediated diseases, recent meta-analyses conclude that C-sections are most robustly related to asthma, atopic diseases, type 1 diabetes and obesity (Blustein and Liu, 2015; Keag et al., 2018; Cardwell et al., 2008; Thavagnanam et al., 2008; Peters et al., 2018; Bager et al., 2008).[6] However, the causal nature and clinical relevance of these relationships remains largely unknown.[7]

3.2.2 *Classification of Cesarean Sections*

Cesarean sections are performed for several indications at different stages of the pregnancy. Cesarean sections are classified either as scheduled (elective) or unscheduled operations. Scheduled C-sections occur without attempted labor and are agreed upon in advance. The large majority of scheduled C-sections are performed during the regular working hours (8am — 4pm) from Monday to Friday. Medical indications

[6] In addition to health outcomes, literature has associated cesarean sections with worse cognitive and emotional development (Bentley et al., 2016).

[7] Hyde et al. (2012) summarize evidence from 14 RCTs that compare the effects of cesarean and vaginal deliveries on infant health. All these studies are small RCTs conducted in populations of at risk babies (e.g. breech delivery). These studies have had exceptionally large problems to achieve target recruitment and do not include long-term follow-ups. Overall, there exist no RCTs to date that would enable to investigate the long-term effects of cesarean sections on infant health. Hyde and Modi (2012) report evidence from survey studies that investigate the perceived acceptability of randomizing the mode of delivery to address long-term health outcomes in low-risk pregnancies. The perceived acceptability of randomizing the mode of delivery in healthy, term, cephalic and singleton pregnancies remains low among obstetricians and mothers, suggesting that adequately powered large-scale RCTs to compare the effects of cesarean and vaginal deliveries on long-term outcomes may remain unrealized in the near future.

that make scheduled C-sections advisable include, among others, multiple pregnancies with non-cephalic presentation of the first fetus or placenta previa. We exclude all scheduled C-sections from our sample.

Most C-sections are performed with no scheduled intervention after spontaneous or medically induced onset of labor. Unscheduled C-sections are surgeries where an attempt of vaginal birth is transformed to a cesarean delivery after the mother has been admitted to a hospital. Unscheduled C-sections are classified by urgency. Emergency C-sections are performed within 30 minutes of the decision, due to an immediate threat to the life of the mother or the baby (NICE, 2011). However, most unscheduled C-sections are performed without such immediate threat. The optimal timing and indication for these operations are imprecise and give large discretion to the clinician. Slow progression of labor or cephalopelvic disproportion are examples of diagnoses that may require an unplanned non-urgent cesarean section. There is wide variation among clinicians in the use of discretionary diagnoses that justify C-sections (Barber et al., 2011; Fraser et al., 1987). Our data contains the registered diagnosis linked to the C-section for a subsample of births. These observations enable us to verify that the peaks in unplanned C-sections are coupled with the use of more discretionary diagnoses.

3.2.3 *Institutional Context*

Finland has universal public health coverage. Comprehensive pre- and postnatal care services are included in the publicly provided services. There are no private medical institutions running maternity wards. Consequently, all deliveries take place in public hospitals. All medical expenses related to prenatal care, delivery and postnatal care are fully covered by the public health care system.

Pregnant women usually give birth in the nearest hospital. Only high-risk pregnancies are systematically directed to a higher-level hospital for obstetric care and delivery. Expectant women do not have pre-assigned midwifes or physicians for the delivery. Midwives take care of the delivery in all hospitals, while physicians have the ultimate responsibility for obstetric care, decide on the type of delivery and perform C-sections. There are no delivery units led by midwifes. The C-section rate (15.5% in 2015) is relatively low from an international perspective (OECD, 2017).

The regular working shifts for physicians are from 8 am to 4 pm from Monday to Friday. The on-call hours for physicians may not exceed 24 hours during the regular working week and last typically from 8 am to 8 am. On weekends, the on-call hours for physicians are from 8 am to 9 am on next day.[8] Midwives follow the same rotation regardless of the type of day and work in three shifts of around 8 hours.[9]

3.3 Data and Methods

3.3.1 *Data*

The two main data sources used in our analysis are the Finnish Medical Birth Register and the Hospital Discharge Register. The Finnish Medical Birth Register was established in 1987. This administrative data resource includes data on all live births and on stillbirths of fetuses with a birth weight of at least 500 grams or with a gestational age of at least 22 weeks. The register includes information on maternal background, health care utilization, and medical interventions during pregnancy and delivery. It also includes mother's diagnoses during delivery (ICD-10 codes) and newborn outcomes until the age of 7 days. From 1990, the register

[8]Even though the statutes that govern on-call arrangements have changed in recent years, during most years covered in our data, small hospitals with less than 1000 annual births could autonomously decide their on call arrangements. In certain hospitals, physicians were allowed to be at home while on duty, if they could arrive to the hospital within 30 minutes from home.

[9]An example of midwives' schedules: (i) from 7 am to 3 pm, (ii) from 2 pm to 9.30 pm, and (iii) from 9.15 pm to 7.15 am.

contains detailed information about the type of C-section (scheduled vs. unscheduled). These data are collected at all delivery hospitals.

We exclude from our sample planned C-sections and multiple pregnancies. For our instrumental variable strategy, we focus only on first births.[10] Our analysis sample includes 392,560 deliveries that took place from 1990 to 2014. For the differences-in-differences analysis, we focus on both first and second births from families where the first child was born by vaginal delivery (more details are provided in section 3.3.2.2). The analysis sample consists of 645,292 children from 322,646 sibling pairs. There are 43 hospitals in our sample. Table 3.A.1 shows summary statistics for all births in Finland between 1990 and 2014.

We match the Finnish Medical Birth Register to the Finnish Hospital Discharge Register, which contains information about the diagnosed medical conditions, medical operations, and the date of diagnoses. This hospital register contains all inpatient consultations in Finland from 1990 to 2013. From 1998, the data include all outpatient visits to hospitals. All diagnoses are coded using the International Classification of Diseases (ICD) tool.[11]

We explore two sets of outcome variables. First, to test whether unplanned C-sections have an impact on neonatal health, we analyze indicators of neonatal health included in the birth register. We study

[10]We follow a common practice in literature and focus on first births, which also allows us to keep just one birth per mother, and abstract from a potential source of correlation between the observations. First-time mothers are also the group of mothers where we find larger variation. Given the faster pace of labor in higher-order births (NICE, 2014) and the high risk of repeated C-section, there is less room for discretion in the decision to perform an unplanned C-section in subsequent deliveries. Our results are qualitatively similar but less precise when we include higher order births.

[11]Diagnoses for years from 1990 to 1995 are recorded using ICD-9 classification. Diagnoses from 1996 onwards are recorded using ICD-10 classification. The quality and completeness of the Finnish Hospital Discharge Register has been assessed in multiple validation studies that have compared recorded data entries with external information. The completeness and accuracy of the data are found to be exceptionally high (Sund, 2012). We assess to what extent our data is able to identify the individuals with a certain diagnosis in the Results section.

Apgar scores one minute after birth, admission to intensive care unit (ICU), need of assisted ventilation and early neonatal mortality (defined as neonatal death in the first week of life).[12] Second, we study longer term outcomes using detailed inpatient and outpatient diagnosis data from the Finnish Hospital Discharge Register. We use primary diagnoses.[13] To maintain a relatively large sample size, we follow individuals from birth until age 15. We focus on the four metabolic and immune-related conditions that have been most robustly associated with cesarean delivery: asthma, atopic diseases (atopic dermatitis and allergic rhinitis), type 1 diabetes and obesity. Table 3.A.2 in the appendix provides more detail about each of these diagnoses.

3.3.2 *Empirical strategy*

We aim to estimate the impact of a cesarean delivery on child's health at birth and older ages. We define a binary variable CS_i that takes value 1 if the delivery is an unplanned C-section and 0 if it is a vaginal delivery. Thus, we aim to estimate the following equation:

$$Y_i = \beta_0 + \beta_1 CS_i + X_i' \beta_2 + \delta_m + \lambda_y + \phi_h + \varepsilon_i, \tag{3.1}$$

where Y_i is the health outcome of infant i, X_i is a vector of covariates and δ_m, λ_y, ϕ_h are fixed effects for the month, year, and hospital of birth, respectively.[14]

[12] Apgar scores result from the examination of the newborn by the midwife or pediatrician one minute after the birth. Five different dimensions are measured and graded from 0 to 2: appearance (skin color), pulse (heart rate), grimace (reflex irritability), activity (muscle tone), and respiration. The resulting score takes values from 1 to 10.

[13] We replicated all our analysis using both primary and secondary diagnoses. All results remain unchanged. Results are available upon request.

[14] The vector of covariates includes the gender of the baby, the mother's marital status, nationality, socioeconomic status, age and smoking status. In addition, we include a wide range of pregnancy and delivery related indicators that include in-vitro fertilization, amniocentesis during pregnancy, ultrasound during pregnancy,

The estimation of equation (3.1) is, however, likely to provide biased estimates of β_1 due to potential selection into cesarean birth.[15] To study the causal effects of cesarean delivery on health, we exploit two different empirical strategies.

3.3.2.1 IV strategy: Variation by time and type of day

Our instrumental variable strategy exploits the higher likelihood of being born by C-section during the normal working shift on pre-leisure days compared to regular working days. We use the interaction between the type of day and work shift as an instrument for the mode of delivery.

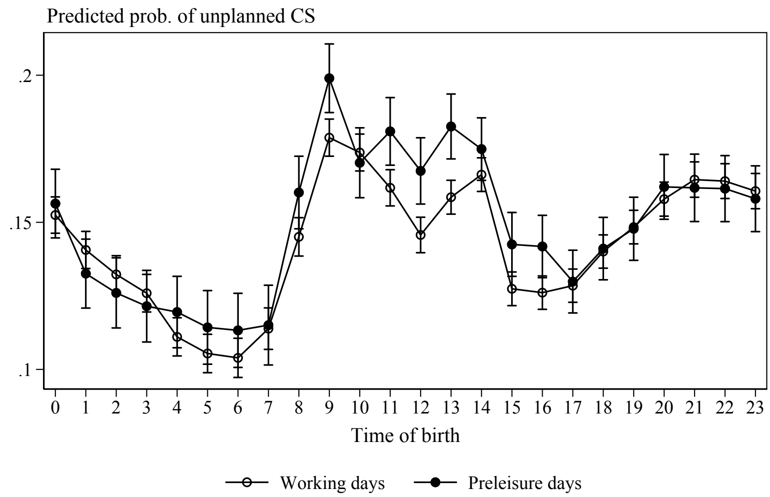
Figure 3.1 presents the predicted probability of unplanned C-section delivery by hour and type of day. We adjust for hospital, month and year of birth fixed effects. Figure 3.1a plots the distribution of C-sections over a 24-hour cycle for working days that precede a leisure day compared to other working days.[16] We find that substantially more C-sections are performed during regular working hours on days that precede a leisure day compared to the rest of working days. Figure 3.1b presents the predicted probability of having an unplanned C-section by work shift and type of day. We find that the gap in C-section rates between a day that precede a leisure day and the rest of working days emerges only during the regular working hours (from 8 am to 4 pm).

Importantly, we find that the excess C-sections performed in days that precede a leisure day are not driven by advancing births that would have been cesarean deliveries in any event. We do not observe any

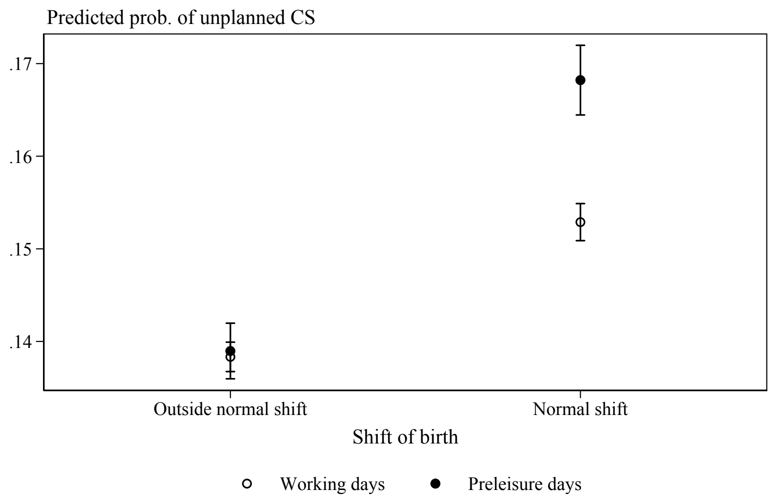
gestational diabetes, maternal hospitalization due to hypertension, maternal hospitalization due to placenta previa, maternal hospitalization due to eclampsia, gestational weeks, induced labor, prostaglandin pre-induction, epidural use, and laughing gas anesthesia.

[15]Figure 3.A.1 in the appendix shows that mothers and babies who undergo a C-section are very different from those mothers and babies who undergo a vaginal delivery.

[16]Working days that precede a leisure day include Fridays and days preceding public holidays. Table 3.A.3 documents all public holidays in Finland. Friday is not considered a working day that precedes a leisure day if it is a holiday.



(a) By time of birth



(b) By shift of birth

Figure 3.1: Predicted probability of unplanned C-section

Notes: Figure (a) presents the predicted probability of unplanned C-section by hour and type of day. Figure (b) shows the predicted probability of unplanned C-section by shift and type of day. Both figures adjust for hospital, month, and year of birth fixed effects. Pre-leisure days include working days that precede a Finnish public holiday or a weekend, while working days include the rest of working days. Sample is restricted to singleton first births which are either unscheduled C-sections or vaginal births.

relative fall in C-sections during the evening hours preceding a leisure day compared to the evenings of regular working days (Figure 3.1a) or during the leisure day (Figure 3.A.2 in the Appendix).[17] These observations suggest that physicians perform C-sections during the regular working hours on pre-leisure days that would not have been performed otherwise.

The time pattern of C-sections is consistent with previous work by Brown (1996) and Halla et al. (2016) that documents an increase in C-section rates on days that precede a leisure day. Halla et al. (2016) exploit this variation in an instrumental variable framework to study the impact of delivery mode on maternal fertility and labor supply. Like the existing literature, we attribute the pre-leisure anomaly in the time pattern of C-sections to physician demand for leisure. This incentive arises from the higher time cost and uncertainty of vaginal births. A cesarean section takes on average 30-75 minutes and is perceived as a relatively easy surgical intervention with low complication rates (NICE, 2011).The average duration of labor for first-time mothers who have a vaginal birth is 11 hours (NICE, 2014).

We provide two pieces of complementary evidence to validate that the excess rate of C-sections is not driven by medical factors. First, we build on previous evidence that some medical diagnoses linked to a cesarean birth are more discretionary than others. Dystocia (prolonged or obstructed labor), one of the most common indications for primary cesarean section, is believed to provide the greatest room for diagnostic discretion (Fraser et al., 1987). The number of dystocia diagnoses has been shown to strongly respond to physician incentives (Evans et al., 1984; Fraser et al., 1987; McCloskey et al., 1992). We examine if there is an excess number of dystocia diagnoses during regular working hours on pre-leisure days. Our results (Table 3.A.4) show that giving birth during

[17]This figure compares the predicted probability of unplanned C-section by hour separately for Saturdays or holidays (the leisure day following the pre-leisure day) and Sundays (a leisure day that is not preceded by a working day). We do not see any relative drop in the C-section rate on Saturdays compared to Sundays at any time of day.

the regular hours on a pre-leisure day increases the probability of having a dystocia diagnosis compared to other working days. Importantly, we do not find this temporal pattern for medical emergencies, for which there should not be any room for discretion. In particular, we find that our instrument does not predict additional examinations of the fetus during labor, which doctors should perform if there are any signs of fetal suffering.[18]

Our second piece of evidence builds on the literature showing that physician mothers are less likely to receive C-sections driven by financial incentives (Johnson and Rehavi, 2016). Consequently, we expect that the probability of having a C-section does not respond to physician demand for leisure among physician mothers and other medical professionals. Our results (Table 3.A.5) support this hypothesis. We do not find that medical professionals have an increased risk of having a C-section during the regular shift on pre-leisure days, while we do find this increase for non-medical mothers with an equivalent level of education.[19]

We exploit the variation in the probability of unplanned C-sections by time and type of day and adopt an instrumental variable approach. We first estimate a standard two-stage least squares (2SLS) with the following first stage:

$$CS_i = \gamma_0 + \gamma_1 NS_i + \gamma_2 Preleisure_i + \gamma_3 NS_i \times Preleisure_i + X_i'\gamma_4 + \delta_m + \lambda_y + \phi_h + \upsilon_i \qquad (3.2)$$

[18] We examine whether physicians take measurements of intrapartum or fetal scalp pH, which proxies the oxygen saturation of fetal blood during labor.

[19] Our definition of medical professionals includes physicians, midwifes and nurses. Our observation relates to a large literature on physician-induced demand in health care. Since the work of Arrow (1963), it has been recognized that asymmetric incentives between physicians and their patients are a central feature of the medical marketplace. The role of financial incentives on the supply of cesarean sections has been documented by Gruber and Owings (1996). Johnson and Rehavi (2016) observe that financial incentives have a particularly large effect on the probability of having a cesarean section among non-physicians. Our results complement the literature on physician-induced demand and show that the excess rate of C-section on pre-leisure days is restricted to non-medical professionals.

and the corresponding second stage:

$$Y_i = \alpha_0 + \alpha_1 NS_i + \alpha_2 Preleisure_i + \alpha_3 \widehat{CS}_i + X_i' \alpha_4 + \delta_m + \lambda_y + \phi_h + \varepsilon_i \tag{3.3}$$

where NS_i is a dummy that takes value 1 for births that take place during the normal shift (from 8 am to 4 pm) and 0 otherwise, $Preleisure_i$ takes value 1 for Fridays or working days preceding a Finnish public holiday and 0 for other working days, $\widehat{CS}_i$ in equation (3.3) are the predicted C-sections from the first stage, X_i is the vector of individual controls,[20] and δ_m, λ_y, ϕ_h are month, year, and hospital of birth fixed effects, respectively. The interaction between regular working hours and a day preceding a leisure day will serve as an instrument. As a result, we will be comparing mothers who give birth in the same hospital during the same shift, but on different types of days (working days preceding a leisure day or other working days). We expect a positive $\widehat{\gamma}_3$ due to increasing physician demand for leisure on days preceding a weekend or public holiday.

Our instrumental variables estimation needs to meet three conditions to yield valid estimates. First, the instrument should strongly influence the probability of C-section (first stage). Second, there should be no selection of mothers who give birth during the regular shift on different types of days. Finally, being born during the regular shift on pre-leisure days, compared to other working days, should only affect child outcomes through the increased probability of being born by C-section (exclusion restriction).

Table 3.1 shows the results from the estimation of the first stage. Column (1) shows the first stage estimates including month, year, and

[20] Gender of the baby, mother's marital status, nationality, socioeconomic status, age, smoking status, and the following pregnancy and delivery characteristics: gestational weeks and indicators for in-vitro fertilization, amniocentesis during pregnancy, ultrasound during pregnancy, gestational diabetes, maternal hospitalization due to hypertension, maternal hospitalization due to placenta previa, maternal hospitalization due to eclampsia, induced labor, prostaglandin pre-induction, epidural use, and laughing gas anesthesia.

hospital fixed effects. Column (2) includes a richer set of controls. These estimates show that being born during the normal shift increases the probability of C-sections for all working days. Moreover, being born during the normal shift on pre-leisure days increases the probability of C-section by 1.5 percentage points. The first stage F-statistics are larger than 25 in both specifications. Following the common critical values for weak instruments (Stock and Yogo, 2005), we can reject the null hypothesis that the instrument is weak.

Figure 3.2 shows that our instrument does not predict a large set of maternal and pregnancy characteristics, including medical conditions that could predict a C-section. This indicates that mothers giving birth during the regular shift on pre-leisure days compared to other working days are similar in observable characteristics, suggesting that the observed increase in C-sections at these times cannot be explained by selection.

Finally, regarding the exclusion restriction, we focus on births that take place on working days, when hospital resources and quality of care should be constant. Moreover, to compromise our empirical strategy, any change in the quality of care would need to happen on pre-leisure days only during the regular working hours. We provide numerous supplementary analyses in section 3.5.1 that reinforce the credibility of this assumption.

The two-stage least squares estimator enables us to identify a local average treatment effect (LATE). This is the effect of C-sections for infants whose mothers' mode of delivery is sensitive to the subjective assessment of the physician. More accurately, we capture births where the type of day affects the decision of the doctor to perform a C-section during the normal shift. The counterfactual for these births is unlikely to be exclusively a cesarean section later on, given that we do not find a relative drop in C-sections on pre-leisure days after the normal shift or during the following day. The LATE will not be informative of the effect of medically indicated C-sections, as those are not affected by

Table 3.1: First stage

	Unplanned CS	
	(1)	(2)
Normal shift	0.015***	0.017***
	(0.001)	(0.001)
Preleisure day	0.001	-0.002
	(0.002)	(0.002)
Normal shift× Preleisure	0.015***	0.014***
	(0.003)	(0.003)
Observations	392561	392561
Controls	NO	YES
$\overline{Y}$	0.145	0.145
First-stage F	26.650	25.209
Adjusted R^2	0.008	0.070

This table shows estimates from the first stage (see equation (3.2)). All specifications include hospital, year and month of birth fixed effects. Controls: gender, maternal age, marital status, nationality, mother occupation (long-term unemployed, high-skilled white collar, low-skilled white collar, manual worker, student, other), whether mother smoked during pregnancy, high/low number of prenatal visits, IVF, gestation weeks, induced labor, prostaglandin preinduction, epidural or laughing gas anesthesia. Robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

leisure incentives. Moreover, the LATE does not capture the effect of unplanned C-sections for babies who had a very fast delivery, leaving no room for physician discretion.

Our primary health outcomes and the endogenous variable are binary. Consequently, besides the 2SLS models we estimate (recursive) bivariate probit models. These specifications mirror equations (3.2) and (3.3) and assume that cesarean delivery (CS_i) and the binary indicator of health Y_i are determined by the following latent indices:

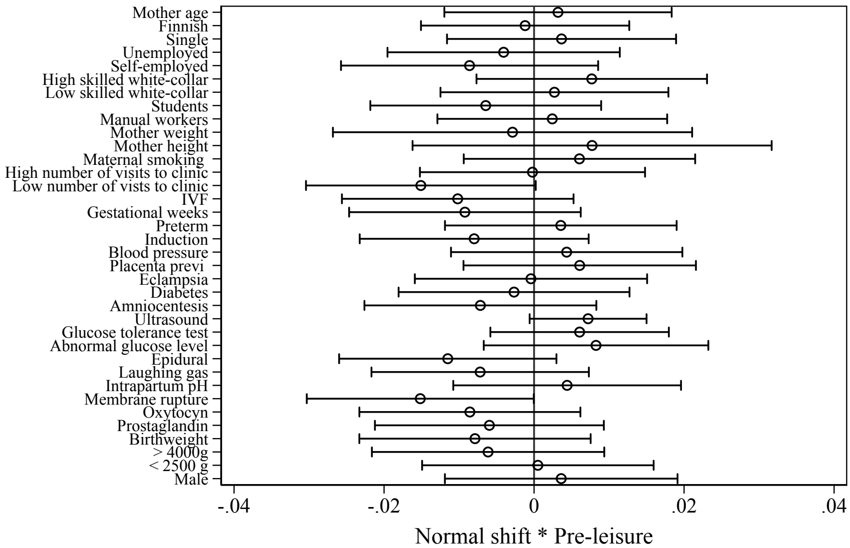


Figure 3.2: Instrument and baseline characteristics

Notes: The figure represents the coefficients and 95% CI from separate regressions of each (standardized) predetermined variable on the instrument (Normal shift * Pre-leisure), controlling for normal shift time, pre-leisure day, and hospital, month, and year of birth fixed effects. Sample is restricted to single births, unscheduled C-sections and vaginal births that take place on working days.

$$CS_i = \mathbb{1}\left[\rho_1 NS_i + \rho_2 Preleisure_i + \rho_3 NS_i \times Preleisure_i + X_i'\rho_4 + \delta_m + \lambda_y + \phi_h + \nu_i > 0\right] \tag{3.4}$$

$$Y_i = \mathbb{1}\left[\pi_1 NS_i + \pi_2 Preleisure_i + \pi_3 CS_i + X_i'\pi_4 + \delta_m + \lambda_y + \phi_h + \xi_i > 0\right] \tag{3.5}$$

where (ν_i, ξ_i) follow a bivariate standard normal distribution with unknown correlation. These equations can be estimated through maximum likelihood. Identification in this setting relies on the same assumptions that are needed to estimate the 2SLS model together with an additional assumption about the joint normality of the error terms.

Bivariate probit estimation is expected to present substantial advantages in the context of this paper. The bivariate probit estimation is shown to be more efficient and less biased than 2SLS when treatment and outcome probabilities are close to 0 or 1 (Chiburis et al., 2012; Bhattacharya et al., 2006; Nielsen et al., 2009). Given that we work in a low C-section rate setting and examine relatively rare outcomes, we expect bivariate probit to outperform 2SLS in terms of efficiency. In the results section, we report marginal effects for both estimators.[21]

3.3.2.2 Differences-in-differences

Our second empirical strategy applies a differences-in-differences approach to a sample of sibling pairs. We restrict the sample to families where the older sibling was born by vaginal delivery and compare the health gap between siblings in families where the second child was born by an unplanned C-section against families where the second child was born by vaginal delivery. This enables us to control for all time-invariant unobserved heterogeneity at the family level and the effect of birth order. Our empirical strategy builds on numerous papers that have used siblings fixed-effects to estimate the impact of health shocks while in-utero or after birth (e.g. Oreopoulos et al., 2008; Almond et al., 2009; Almqvist et al., 2012; Aizer et al., 2016) and extends the model to a difference-in-differences specification with family fixed-effects. A related approach is used by Black et al. (2017) to study the impact of child disability on sibling outcomes.

[21] Bivariate probit models estimate unconditional average causal effects. In contrast, 2SLS estimates the LATE. However, in practice, the average causal effects produced by bivariate probit are likely to be similar to 2SLS estimates (Angrist and Pischke, 2009).

We estimate the following equation:

$$Y_{if} = \psi_0 + \psi_1 Secondborn_{if} + \psi_2 Secondborn_{if} \times CS_{if} + X'_{if}\psi_3 + \gamma_f + \delta_m + \lambda_y + \phi_h + \eta_{if} \tag{3.6}$$

where Y_{if} is the health outcome of child i in family f, $Secondborn_{if}$ is a dummy variable equal to 1 for the second child and 0 for the first child, CS_{if} is an indicator equal to 1 for unplanned C-section and 0 for vaginal delivery, X_{if} is a vector with the same pregnancy and maternal controls of equation (3.3), except for maternal characteristics that are time-invariant, and diagnoses during delivery (prolonged and obstructed labor),[22] γ_f, δ_m, λ_y and ϕ_h are family, month, year, and hospital of birth fixed effects, respectively.[23] We cluster standard errors at the family level. Our parameter of interest is ψ_2, which identifies the change in the health gap between siblings in families where the first child was born by vaginal delivery and the second child by C-section compared to families where both children were born by vaginal delivery.

We do not include families whose older child was born by C-section for two reasons. First, mothers who have a C-section in the first delivery and a vaginal birth in the second delivery are a very selected sample, given the very high probability of having a repeat C-section.[24] Second, some studies find that having a C-section is associated with lower fertility (Halla et al., 2016; Keag et al., 2018). We abstract from these concerns by focusing on mothers whose first birth was a vaginal delivery.

Even though our rich data sources make it possible to control for a large set of observable characteristics, it could be that there are sibling-

[22] We do not include these diagnoses during labor as controls in the IV specification, given that we find evidence that they can be an outcome of the time and type of day.

[23] We cannot estimate the baseline effects of the CS_{if} indicator, which are absorbed by the interaction $Secondborn_{if} \times CS_{if}$, since by construction of our sample only second children have C-sections.

[24] In 2010, The American College of Obstetricians and Gynaecologists (ACOG) encouraged doctors to allow women to opt for a vaginal delivery after a C-section, but the number of vaginal births after C-section has remained low (American College of Obstetricians and Gynecologists, 2010).

specific unobservable differences that vary within family. In particular, younger siblings born by C-section could be negatively selected compared to their vaginally-delivered older siblings if the cesarean delivery is caused by complications, either during the pregnancy or delivery, that we cannot observe in our data. These unobservable complications could cause our estimates to be negatively biased. Thus, our difference-in-difference estimates could overestimate the impact of C-sections on the different diagnoses. In section 3.5.2 we assess the magnitude of the potential bias and provide evidence that it is relatively small. We will nonetheless keep the direction of this bias in mind when interpreting the results from this strategy.

3.4 RESULTS

3.4.1 *Neonatal outcomes*

We first estimate the impact of C-sections on neonatal outcomes. Table 3.2 shows our OLS (first panel), 2SLS (second panel), bivariate probit marginal effects (third panel) and differences-in-differences (fourth panel) estimates. We find that the OLS results replicate existing findings. Cesarean sections are associated with adverse outcomes at birth and higher neonatal mortality.[25] Our 2SLS estimates are not significant for any of the outcomes. However, the magnitude of coefficients and large standard errors suggest that we cannot reject that there is a (potentially large) effect on neonatal outcomes. As discussed in section 3.3.2.1, 2SLS estimates are expected to be particularly uninformative with low treatment and outcome probabilities.

Bivariate probit marginal effects are substantially more precisely estimated than the 2SLS coefficients. Yet, all point estimates from

[25]The OLS estimation is ran in a sample that only excludes planned C-sections and births for which we do not observe parity. The specification includes the full set of controls and fixed effects described in equation (3.1), as well as controls for birth order.

Table 3.2: Neonatal outcomes

	(1) Low Apgar 1	(2) ICU	(3) Assisted ventilation	(4) Neonatal mortality
OLS	0.068***	0.118***	0.027***	0.002***
	(0.001)	(0.001)	(0.001)	(0.000)
$\overline{Y}$	0.049	0.087	0.009	0.001
N	1119467	1120932	1120932	1119842
2SLS	-0.018	-0.088	-0.006	0.006
	(0.140)	(0.170)	(0.061)	(0.023)
$\overline{Y}$	0.066	0.106	0.012	0.002
N	392017	392560	392560	392173
Biprobit	0.104***	0.163***	0.017***	-0.001
	(0.008)	(0.009)	(0.005)	(0.005)
$\overline{Y}$	0.066	0.106	0.012	0.002
N	392017	392560	392560	392173
Diff-in-diff	0.053***	0.111***	0.036***	0.001
	(0.007)	(0.007)	(0.004)	(0.002)
$\overline{Y}$	0.038	0.070	0.006	0.001
N	644551	645292	645292	644746
First-stage F	24.996	25.216	25.216	26.007

This table shows the estimates of the marginal effect of an unplanned CS on different neonatal health indicators by OLS, 2SLS, bivariate probit and differences-in-differences estimation (see equations (3.2), (3.3), and (3.6)). Specifications as detailed in section 3.3.2, with the full set of fixed effects and controls. Robust standard errors (in parentheses) in panels 1-3, and standard errors clustered at the family level in the differences in differences panel. First-stage F statistic from 2SLS and bivariate probit specifications. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

the bivariate probit models are within the confidence intervals of the 2SLS estimates. The bivariate probit results suggest that unplanned C-sections increase the probability of having a low Apgar score (Apgar lower than 7), being admitted to the intensive care unit and receiving assisted ventilation. The magnitude of the bivariate probit marginal effect estimates are similar to OLS estimates. However, we do not find significantly increased mortality risk within seven days after birth. The results from the differences-in-differences models give support to these findings with similarly-sized and more precise coefficients. Overall, our results suggest that unplanned C-sections have a negative impact on neonatal health. However, these adverse effects do not translate into a higher probability of early neonatal mortality.

3.4.2 *Later infant health*

We now turn to the results of the long-run effects of C-sections on health outcomes. Table 3.3 shows the OLS (first panel), two-stage least squares (second panel), bivariate probit (third panel) and differences-in-differences (fourth panel) marginal effect estimates at ages 5 and 10. We analyze health conditions that have been extensively documented in the literature as being positively associated with cesarean deliveries: type 1 diabetes, obesity, asthma, and other atopic diseases (atopic dermatitis and allergic rhinitis). Given that we study health outcomes for children who are born from 1990 to 2014, the sample size decreases as we consider older ages. We report year by year bivariate probit and diff-in-diff estimates up to age 15 in Figures 3.3 and 3.4, respectively. We report our OLS estimates in Figure 3.A.3.

The OLS estimates suggest that cesarean sections are associated with a higher probability of asthma, obesity and atopic diseases. These findings are consistent with existing studies that have documented significant associations between cesarean sections and metabolic and immune-related conditions. However, we do not detect that C-sections

Table 3.3: Child diagnoses by age

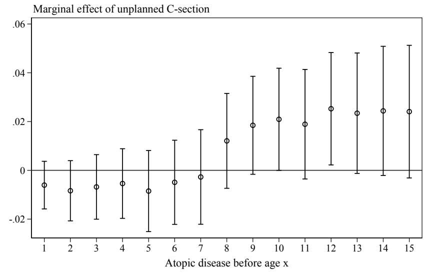
	Type 1 diabetes		Asthma		Obesity		Atopy	
By age:	5	10	5	10	5	10	5	10
OLS	0.000	0.000	0.007***	0.010***	0.000***	0.002***	0.002***	0.004***
	(0.000)	(0.000)	(0.001)	(0.001)	(0.000)	(0.000)	(0.001)	(0.001)
$\overline{Y}$	0.003	0.006	0.045	0.071	0.001	0.004	0.044	0.061
N	807035	556009	807035	556009	807035	556009	807035	556009
2SLS	0.089**	0.062	0.074	-0.121	0.001	0.000	0.022	0.110
	(0.036)	(0.044)	(0.113)	(0.139)	(0.013)	(0.034)	(0.112)	(0.127)
$\overline{Y}$	0.003	0.006	0.040	0.070	0.001	0.004	0.041	0.058
N	296998	217768	296998	217768	296998	217768	296998	217768
Biprobit	0.003	0.003	0.031***	0.015	0.001	0.003	-0.008	0.021
	(0.002)	(0.004)	(0.009)	(0.015)	(0.001)	(0.003)	(0.010)	(0.013)
$\overline{Y}$	0.003	0.006	0.040	0.070	0.001	0.004	0.041	0.058
N	296998	217768	296998	217768	296998	217768	296998	217768
Diff-in-diff.	-0.001	-0.001	0.014**	0.011	0.001	0.001	0.003	-0.001
	(0.001)	(0.003)	(0.006)	(0.009)	(0.001)	(0.002)	(0.005)	(0.007)
$\overline{Y}$	0.003	0.006	0.045	0.070	0.001	0.004	0.044	0.060
N	510075	366885	510075	366885	510075	366885	510075	366885
First-stage F	25.725	29.546	25.725	29.546	25.725	29.546	25.725	29.546

This table shows the estimates of the effect of an unplanned CS on the probability of the child having each diagnosis by age by OLS, 2SLS, bivariate probit (marginal effects) and differences-in-differences estimation (see equations (3.2), (3.3), and (3.6)). Specifications as detailed in section 3.3.2, with the full set of fixed effects and controls. Robust standard errors (in parentheses) in panels 1-3, and standard errors clustered at the family level in the differences in differences panel. First-stage F statistic from 2SLS and bivariate probit specifications. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

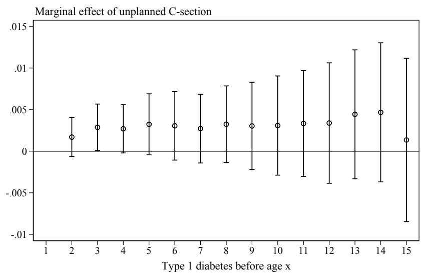
are associated with a higher probability of type 1 diabetes diagnosis.

The 2SLS results suggest that unplanned C-sections increase the probability of having a type 1 diabetes diagnosis before age 5, even though the effect is not significant by age 10. The effect size of the estimate is large, but very imprecise. Our results suggests 9 percentage point increase in the probability of type 1 diabetes, but are consistent with an increase ranging from 6.3 to 12.5 percentage points. The 2SLS estimates for asthma are not significant. However, the lack of precision does not enable us to rule out even very large (positive or negative) effects. For instance, the estimates by age 5 suggest that the impact of C-sections may range from -4.2 pp to 18.4 pp. Finally, the 2SLS estimates for obesity and atopic diseases are not significant, but also too imprecise to rule out very large effects.

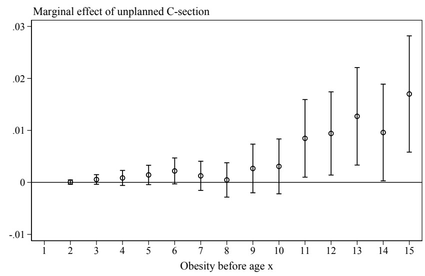
(a) Asthma



(b) Atopy



(c) Type 1 Diabetes



(d) Obesity

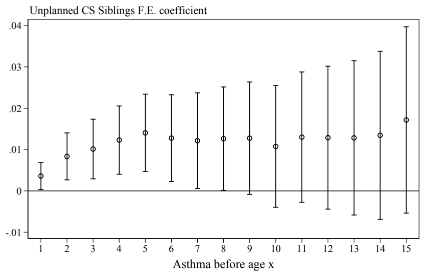
FIGURE 3.3: Bivariate probit estimation – Child diagnoses by age

Notes: The figure plots the marginal effects from the bivariate probit estimation of the effect of unplanned CS on the probability of each diagnosis by age, with our usual specification. All regressions include hospital, year and month of birth fixed effects and the full set of controls as described in Section 3.3.2.1.

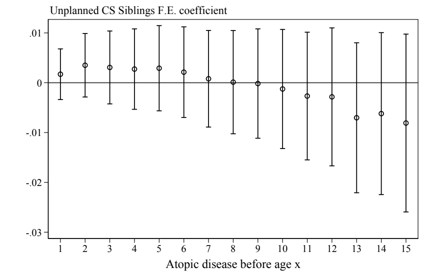
Similarly to our results for neonatal outcomes, the bivariate probit estimates (marginal effects) are substantially more precisely estimated than the 2SLS coefficients. Yet, practically all point estimates from the bivariate probit models are within the confidence intervals of the 2SLS estimates. For type 1 diabetes, the marginal effect is much smaller than the coefficient from the linear model and not significant. For asthma, the results suggest a significant increase in the probability of a diagnosis by age 5 of 0.031 (95% CI 0.022–0.04). Even though estimates are noisier and no longer significant by age 10, the results in Figure 3.3 show that unplanned C-sections significantly increase the probability of an asthma diagnosis for children as young as 2 years old. The effect is statistically significant up to age 9. For obesity, the bivariate probit results are precisely estimated at zero at age 5 (0.001, 95% CI 0.000–0.002) and 10 (0.003, 95% CI 0.000–0.006). However, the results in Figure 3.3 show a statistically detectable effect from age 11. Finally, we do not find a significant impact on atopic diseases at age 5 or 10.

The differences-in-differences results are very similar to the bivariate probit ones. We find that the second-born child has substantially greater risk of having an asthma diagnosis by age 5 than the first-born child in families where the second child is born by C-section. Similarly to the bivariate probit estimates, Figure 3.4 shows that this effect is significant from ages 1 to 8. Despite the fact that our differences-in-differences estimates could be negatively biased (Section 3.3.2.2), we do not find any significant effects on obesity, atopic diseases or type 1 diabetes. These results reinforce the conclusion that C-sections do not have impact on these outcomes.

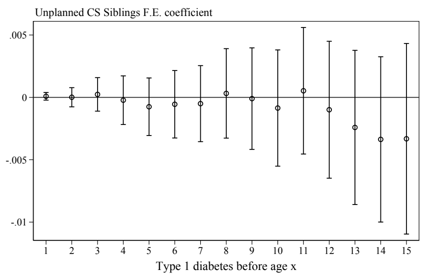
Overall, our results suggest that unplanned C-sections increase the probability of suffering from asthma during childhood. The magnitude of this effect differs slightly depending on the estimation method. The bivariate probit estimates indicate a slightly larger but more imprecisely estimated impact (around 2 pp on average for ages 5 to 10) than the estimates based on differences-in-differences analysis (1.3 pp). By comparing these estimates to the sample mean, we find that the less precise bivariate probit estimates suggest a 36% increase in the probability of having asthma diagnosis (compared to the mean of 5.5% over ages 5-10), while the differences-in-differences estimates suggest a 21% increase (compared to the sample mean of 5.8%). The latter is closer to the 20% increase in the risk of asthma that is documented in recent meta-analyses



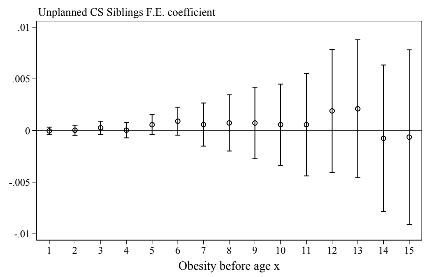
(a) Asthma



(b) Atopy



(c) Type 1 Diabetes



(d) Obesity

FIGURE 3.4: Diff-in-diff analysis – Child diagnoses by age

Notes: The figure plots the coefficient of unplanned C-section for each diagnosis by age in family fixed effects models. All regressions include family, hospital, year and month of birth fixed effects and the full set of controls as described in Section 3.3.2.2.

(Thavagnanam et al., 2008; Keag et al., 2018).

Our analysis indicates that C-sections do not increase the probability of type 1 diabetes or atopic diseases. For diabetes, we can rule out effects larger than 0.7 pp at age 5 using the bivariate probit model and larger than 0.1 pp using the differences-in-differences model. For atopic diseases, in turn, our results discard effects larger than 1.2-1.3 pp with both methods. Finally, bivariate probit results suggest there might be an effect of C-sections on obesity after age 11. This observation is consistent with the evidence that puberty is a vulnerable period for the development of overweight and obesity (Lobstein et al., 2004). However, our analysis is not conclusive in this regard, as the results from the differences-in-differences estimation do not corroborate this finding. For younger ages, all methods suggest that there is no impact on obesity. For instance, estimates at age 5 enable us to rule out effects larger than 0.3 pp.

One potential limitation of our analysis is that we study diagnoses made at inpatient or outpatient visits to a hospital. For some outcomes, these diagnoses may be a good approximation to the true prevalence of the disease, while for others hospital diagnoses may lead to underestimation. A previous study on type 1 diabetes documents that in Finland practically all new type 1 diabetes diagnoses are made in a hospital and listed in the Hospital Discharge Register (Harjutsalo, 2008). This evidence implies that we are able to observe practically all type 1 diabetes diagnoses in our population of interest. However, since 1994, diagnoses for asthma in Finland are often made by general practitioners (Tuomisto et al., 2010). Thus, we are likely to trace only the most severe cases of asthma. The same might be true for atopic disesase and obesity.[26] In any case, OLS results show that C-sections in general are associated even with these hospital diagnoses. Our analysis thus highlights the importance of dealing with the endogeneity of the delivery mode.

3.5 VALIDITY CHECKS

3.5.1 *Exclusion restriction and sensitivity checks*

Our instrumental variables strategy relies on the assumption that the interaction of regular working hours and days that precede a weekend or public holiday affects health outcomes only through its impact on the likelihood of cesarean sections. We argue that, in this setting, this is

[26] There is some evidence that, among children, ICD-coding underestimates the true prevalence of obesity. ICD-coded cases have a higher BMI and higher healthcare utilization than those not coded (Kuhle et al., 2011).

Table 3.4: Validity checks

	Birth weight	Asthma at age 5 for sample	
		Thursdays vs Fridays	Excluding inductions
Biprobit	-	0.023	0.036***
	-	(0.015)	(0.010)
$\overline{Y}$	-	0.040	0.039
N	-	117826	246933
Diff-in-diff.	-5.416	-	0.017**
	(7.617)	-	(0.007)
$\overline{Y}$	3566.117	-	0.044
N	645134	-	440291

This table shows, in column 1, a placebo regression where the outcome is birth weight; and in columns 2 and 3, the results from the bivariate probit (top) and the differences in differences (bottom) estimation of the impact of unplanned CS on the probability of asthma diagnosis by age 5 restricting the sample to births taking place on Thursdays or Fridays (col. 2) or to non-induced births (col. 3). Specifications as detailed in sections 3.3.2.1 and 3.3.2.2, with the full set of fixed effects and controls. Robust standard errors (in parentheses) for bivariate probit results, and standard errors clustered at the family level in the differences in differences panel. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

likely to hold, since a violation would require other changes to happen on days that precede a public holiday but only during the regular shift. In the following, we provide several pieces of evidence that support the credibility of this assumption.

First, we explore the overall activity at maternity wards across the different types of days. Figure 3.A.4 (the first panel) shows the proportion of planned cesarean sections by time of birth and type of day. We find that scheduled activity is organized very similarly during all working days. Moreover, we compare the number of births by type of day and weekday (Figure 3.A.4, second panel) and do not find any evidence of maternity ward crowding during the days that precede a public holiday.

Second, we explore the quality of care provided during different weekdays. The first panel of Figure 3.A.5 shows that the probability of having a low Apgar score (below 7) does not differ between weekdays or type of day, suggesting that the quality of care during labor and delivery does not differ by type of day. Figure 3.A.5 (second panel) shows the probability of early neonatal mortality, defined as death of a live-born baby within the first seven days of life, by weekday and type of day. We expect that this measure would capture changes in the quality of care after birth. We do not find evidence that early neonatal mortality is higher for babies born on days that precede a public holiday compared to other weekdays. Moreover, we do not find that mothers who have a C-section on a day that precedes a public holiday have a longer length of stay than mothers who have a C-section on other weekdays (Figure 3.A.6). We interpret these findings as evidence that the quality of care remains constant across all working days.

Third, since babies born on days that precede a public holiday or weekend stay in the hospital during the following non-working days, one could argue that their quality of post-natal care is worse compared to children born on other working days. This would be constant for both babies born during the regular shift and at other times, and hence would not necessarily comprise the exclusion restriction. Yet, in what follows we assess this concern. Table 3.4 shows the coefficients for IV regressions that restrict the sample to babies born on Thursdays or Fridays.[27] We find, despite the reduced sample size, that the results from this estimation are consistent with our main results.

Finally, we report in Figure 3.2 that mothers who give birth during

[27]The average length of stay in our sample is four days. The majority of babies born on Thursdays and Fridays are hospitalized during the weekend.

the regular working hours on days that precede a public holiday do not have higher probability of having induced labor. However, the induction of labor is likely to offer more room for discretionary behavior, in which case the decision to perform a C-section might be more sensitive to physician demand for leisure.[28] In other words, we expect that mothers whose labor has been artificially induced are more likely to be part of the complier population. Column 3 in Table 3.4 shows that our coefficients remain about the same if we exclude mothers whose labor was induced from our sample. The same conclusion holds if we exclude inductions from our differences-in-differences estimation. These results suggest that our findings are not driven by mothers whose labor has been induced after an admission to the maternity ward.

3.5.2 *Differences-in-differences validity checks*

The results from our differences-in-differences model with family fixed effects could be biased if there are unobservable characteristics correlated with the mode of delivery that vary within family and across siblings. Under this scenario, this methodology would yield upward biased estimates. However, as shown in Section 3.4, our differences-in-differences results suggest that C-sections do not increase the risk of developing various immune-mediated diseases that have previously been associated with cesarean births.

To assess the extent to which these results could be explained by selection, we first run a regression using birth weight as a placebo outcome, given that it cannot be affected by unplanned C-sections. Table 3.4 shows that our differences-in-differences model with family fixed effects does not predict birth weight. This result supports the

[28] Recent evidence casts doubt on the commonly-held belief that induction of labor increases the risk for cesarean delivery. In particular, recent studies show that inductions at full term do not increase the risk of cesarean delivery (Saccone and Berghella, 2015) or even lower it (Mishanina et al., 2014), with no increased risks for the mother and some benefits for the fetus.

validity of this strategy: family fixed-effects, jointly with the large set of controls, seem to be taking into account general health differences between siblings born by C-section and vaginal delivery.

Second, we compare our differences-in-differences estimates to those from other samples of sibling pairs where we expect the second child to be negatively selected with respect to their older sibling, but where none of them was born by C-section. These samples include (i) a sample of siblings where the first child is born by eutocic birth and the second child is born either by eutocic or by instrumented birth, and (ii) a sample of siblings where the first born had a low-risk pregnancy and the second born had either a low- or a high-risk pregnancy, while all children in the sample were born by vaginal delivery.[29] Consequently, we assess the health gap between siblings across families that had a complication during the second birth or during the second pregnancy compared to families where none of the siblings encountered any of these complications during pregnancy or birth.

Table 3.5 shows our differences-in-differences estimates using these samples of siblings. The first four columns show that, compared to families where both siblings were born by eutocic birth, second children born by instrumented vaginal delivery have worse neonatal health than their older siblings who had an eutocic birth. We find a significantly higher probability of having low Apgar scores and of being admitted to the ICU (top panel). In the bottom panel, we can see that children who experienced a high-risk pregnancy do not have significantly worse neonatal health by any of the indicators, even though all coefficients have a positive sign. In the last four columns, we explore if negative selection leading to instrumented birth or risk pregnancy is associated with a higher probability of having any of the diagnoses we analyze in section

[29] An eutocic delivery is a vaginal delivery with no instrumentation. We define a high-risk pregnancy as a pregnancy where the mother had at least one of these complications: a positive result in the glucose tolerance test, an hospitalization during pregnancy due to blood loss, hypertension, eclampsia or placenta previa. A low-risk pregnancy is defined as the absence of these issues.

Table 3.5: Validity of differences-in-differences

	Neonatal health				Diagnosis by age 5			
	Low Apgar	ICU	Assisted Ventilation	Neonatal mortality	Type 1 diabetes	Asthma	Obesity	Atopy
Instrumented	0.060***	0.020**	0.001	-0.001	-0.001	-0.004	-0.001	-0.006
	(0.009)	(0.009)	(0.003)	(0.002)	(0.002)	(0.008)	(0.001)	(0.009)
$\overline{Y}$	0.028	0.061	0.005	0.001	0.003	0.044	0.001	0.044
N	534119	534689	534689	534264	428392	428392	428392	428392
Risk pregnancy	0.001	0.016	0.003	0.001	-0.001	0.002	-0.000	0.005
	(0.007)	(0.010)	(0.004)	(0.002)	(0.002)	(0.009)	(0.001)	(0.009)
$\overline{Y}$	0.035	0.062	0.005	0.001	0.003	0.044	0.001	0.044
N	608688	609368	609368	608909	482536	482536	482536	482536

This table shows the results from sibling fixed effect models, following the specification in equation (3.6), for two different samples of children: in the top panel, for a sample of sibling pairs where the first child was born by eutocic birth, and the second child is born either by eutocic or instrumented vaginal birth; in the bottom panel, for vaginally delivered sibling pairs where the first child did not have a high-risk pregnancy and the second child had a low- or high-risk pregnancy. The top panel coefficient represents the change in the health gap between siblings in families where the second child was born by instrumented vaginal delivery, while the bottom panel coefficient represents the same for families where the second child had a high-risk pregnancy. All specifications include family, hospital, year and month of birth fixed effects and the controls described in section 3.3.2.2. Standard errors are clustered at the family level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

3.4. We do not find evidence that siblings born by instrumented vaginal delivery or those who had a high-risk pregnancy have an increased risk of type 1 diabetes, asthma, atopic diseases or obesity at age 5. These observations suggest that our differences-in-differences results for asthma are unlikely to be explained by negative selection.

3.6 Conclusions

This paper provides new evidence on the effects of avoidable cesarean sections on various short- and long-term health outcomes. We use a novel instrumental variable estimation strategy to overcome the potential endogenenity of birth mode and abstract from cases in which C-sections respond to a clear clinical indication. Our empirical strategy builds on the finding that unplanned C-sections are more common during regular working hours on Fridays and working days preceding public holidays. We complement this empirical strategy by estimating a differences-in-differences model with family fixed effects that compares the health gap between siblings in families where the second child was born by unplanned C-section with the health gap between siblings who were both born by vaginal delivery.

Our results suggest that C-sections have a substantial negative impact on neonatal health. However, these adverse effects are not severe enough to translate into a higher probability of increased neonatal mortality. Our long-run analysis follows children from birth to age 15 and investigates the impact of C-sections on four health outcomes that have been consistently associated with C-sections: type 1 diabetes, asthma, obesity, and atopic diseases. In contrast to the OLS estimates, our instrumental variable and differences-in-differences estimates show that unplanned C-sections do not have a significant effect on the probability of having a type 1 diabetes, obesity, or atopic disease diagnosis. However, we do find that being born by an unplanned C-section increases the probability of having asthma. This effect is detectable from ages 1-2 and of similar size to the associations reported by previous studies (Thavagnanam et al., 2008; Keag et al., 2018). Our results are consistent with the hypothesis that mode of delivery can affect the development of immune-related conditions, but suggest more nuanced effects of C-sections than previous work.

This paper provides first evidence on the long-term effects of unplanned C-sections that do not respond to a clear medical indication, using inpatient and outpatient data for all children born in Finland from 1990 to 2014. Although we are able to observe most of the cases of type 1 diabetes, for some diagnoses (asthma, atopic disease, and obesity) we might be only able to trace the most severe cases, given that these conditions are often treated by general practitioners. However, the fact that our OLS estimation, which includes a large set of controls, shows significant associations of cesarean birth with these outcomes, highlights the importance of dealing with omitted variable bias when analyzing the impact of mode of delivery. Future work should focus on analyzing the causal effect of C-sections on obesity and other metabolic disorders using primary care data and anthropometric measurements.

We make use of the detailed diagnosis data to show that variation by time and type of day can be a valid source of variation to investigate the impact of avoidable C-sections. First, we show that mothers who give birth at regular working hours on pre-leisure days are comparable in terms of a extensive list of pregnancy, health, and sociodemographic characteristics to mothers who give birth during these times on the rest of working days. Second, we show that during the normal shift on these pre-leisure days physicians make greater use of more discretionary diagnoses as justification for the C-section. We also show that these additional C-sections are not performed to mothers who are in the medical profession, and whose mode of delivery has been shown by the literature not to respond to doctors' incentives (Johnson and Rehavi, 2016).

All in all, our results suggest that the additional C-sections performed during regular working hours on pre-leisure days are not driven by medical factors. We provide this evidence in the context of Finland, one of the countries with the lowest C-section rate in the world (OECD, 2017). We would expect this variation to provide an even stronger source of identification in other countries with higher rates of medical interventionism during childbirth. Thus, this paper hopes to provide a solid base upon which future research on the effects of avoidable cesarean sections can be built.

3

APPENDIX

Appendix 3.A

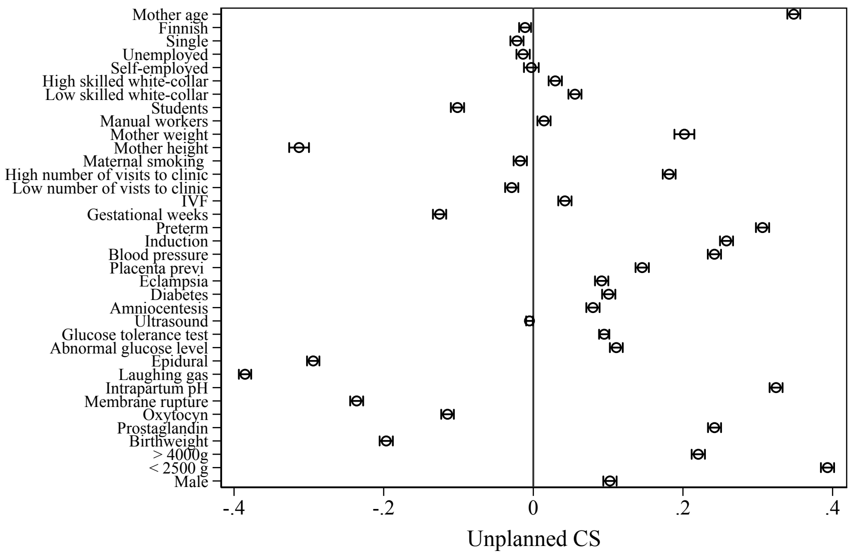


Figure 3.A.1: Difference in baseline characteristics by type of birth

Notes: The figure represents the coefficients and 95% CI from separate regressions of each (standardized) predetermined variable on an indicator taking value 1 if the mother had an unplanned C-section, and 0 if it was a vaginal delivery, controlling for normal shift time, pre-leisure day, and hospital, month, and year of birth fixed effects. Sample is restricted to single births, unscheduled C-sections and vaginal births that take place on working days.

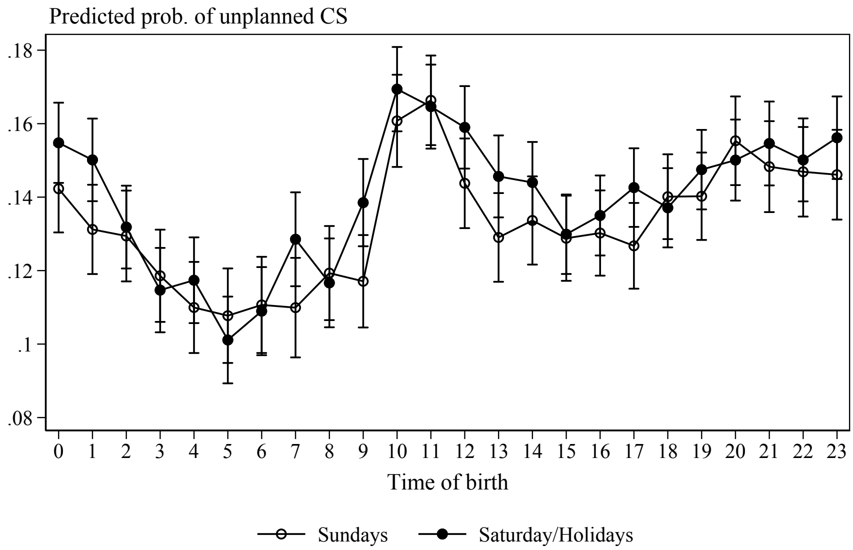
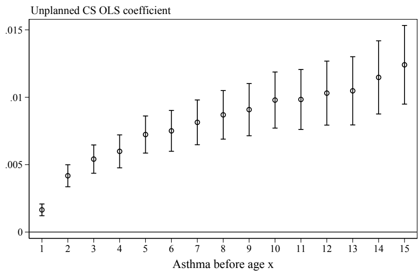
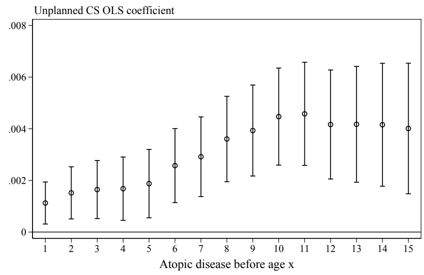


FIGURE 3.A.2: Predicted probability of unplanned C-section by time on weekends

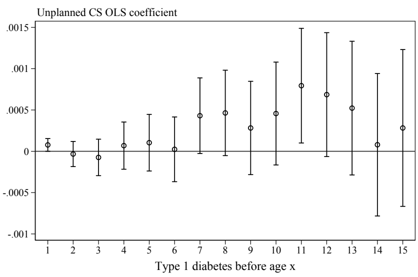
Notes: The figure represents the predicted probability of unplanned C-sections by time of birth for Sundays and for Saturdays or holidays, adjusting for hospital, month, and year of birth fixed effects. Sample is restricted to single births, unscheduled C-sections and vaginal births that take place on Saturdays or holidays and Sundays.



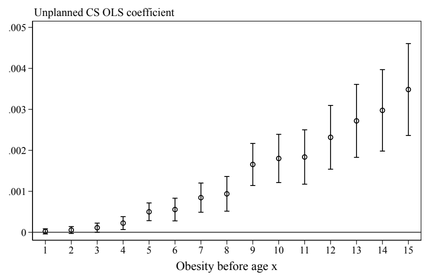
(a) Asthma



(b) Atopy



(c) Type 1 Diabetes

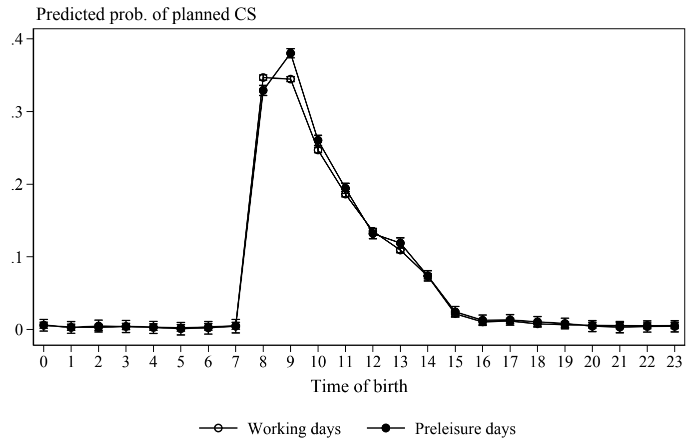


(d) Obesity

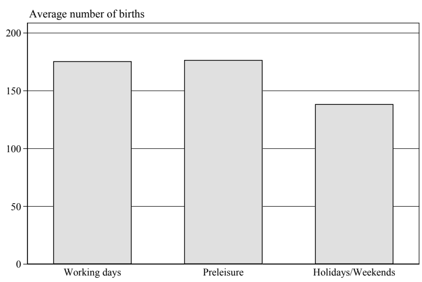
Figure 3.A.3: OLS estimation: Child diagnoses by age

Notes: The figure plots the results from the OLS estimation of the effect of unplanned CS on the probability of each diagnosis by age, with our usual specification. All regressions include hospital, year, and month of birth fixed effects and the full set of controls described in section 3.3.2.

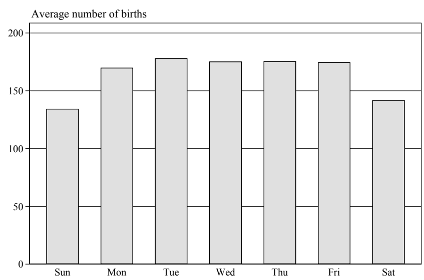
Planned C-sections



Number of births



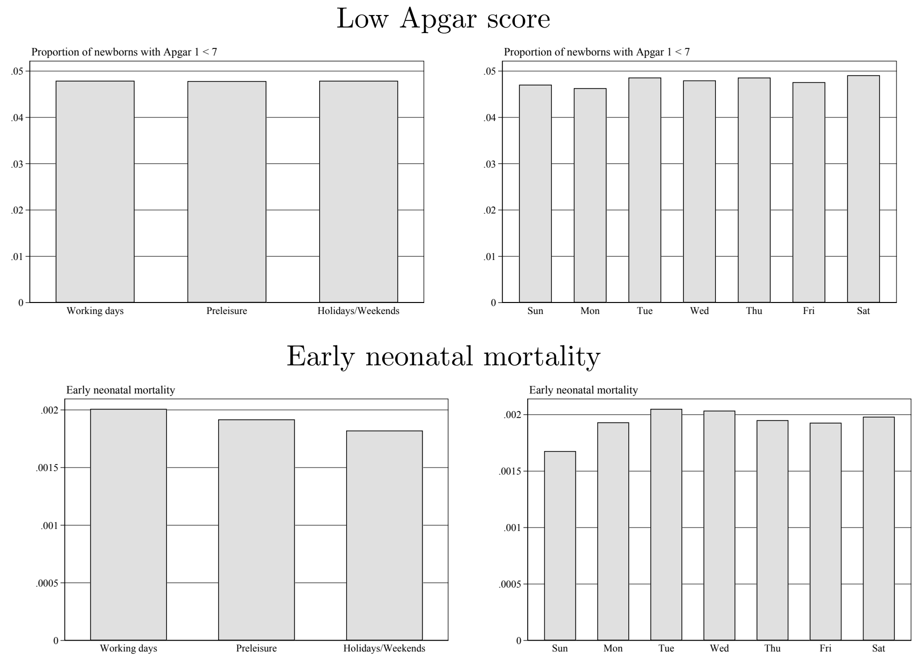
(b) Births by type of day



(c) Births by weekday

FIGURE 3.A.4: Activity at maternity wards by type of day

Notes: This figure plots, in the first panel, the probability of planned C-section by time of birth on pre-leisure working days and other working days, and in the second panel, the average number of births by type of day (column (a)) and by weekday (column (b)).



(b) By type of day (c) By weekday

Figure 3.A.5: Quality of care by type of day

Notes: This figure plots, in the first panel, the probability of the newborn having low Apgar score and in the second panel the probability of early neonatal mortality by type of day (column (a)) and by weekday (column (b)). Sample is restricted to single births, unscheduled C-sections and vaginal births.

(a) By type of day

(b) By weekday

Figure 3.A.6: Mother length of stay by type of day

Notes: This figure plots, in the left panel, the average length of stay of the mother for mothers who had a C-section by type of day, and in the right panel, by day of the week. Sample is restricted to single births and unscheduled C-sections.

Table 3.A.1: Summary statistics

	Full sample	
	Mean	SD
Background characteristics		
Mother's age	29.369	5.335
Finnish	0.958	0.200
Married	0.628	0.483
Unemployed	0.004	0.061
Selfemployed	0.017	0.128
High skilled white-collar	0.178	0.382
Low skilled white-collar	0.433	0.496
Student	0.095	0.294
Manual workers	0.180	0.384
Pregnancy characteristics		
Mother weight	66.780	14.033
Mother height	165.562	6.032
Tobacco during pregnancy	0.128	0.334
High visits clinic	0.239	0.426
Low visits clinic	0.190	0.392
IVF	0.003	0.057
Gestational weeks	39.702	1.853
Preterm	0.056	0.230
Previous CS	0.099	0.299
First birth	0.410	0.492
Blood pressure hospitalization	0.033	0.178
Placenta previa	0.003	0.052
Eclampsia	0.000	0.022
Gestational diabetes	0.007	0.085
Amniocentesis	0.029	0.168
Ultrasound	0.458	0.498
Glucose Tolerance Test	0.183	0.387
Glucose Tolerance Test Positive	0.049	
Childbirth characteristics		
Induction	0.165	0.372
Epidural	0.326	0.469
Laughing gas	0.453	0.498
Intrapartum pH	0.042	0.201
Membrane rupture	0.448	0.497
Oxytocyn	0.401	0.490
Prostaglandin	0.076	0.265
Birth weight	3520.736	571.55
Male	0.511	0.500
Mode of delivery		
Planned CS	0.071	0.257
Unplanned CS	0.101	0.301
Eutocic	0.763	0.425
Ventose	0.066	0.248
Forceps	0.001	0.033
Breech vaginal	0.005	0.073
Observations	1482884	

Table 3.A.2: Long-term outcome variables

Outcome	ICD-10 codes	Description
Asthma	J45, J46	Asthma is the most common chronic disease in children (Asher and Pearce, 2014). Asthma is an inflammatory disorder characterized by recurrent attacks of breathlessness and wheezing and can also cause cough, particularly in children. Recurrent asthma symptoms frequently cause sleeplessness, daytime fatigue, reduced activity levels and school and work absenteeism.[a] It is caused by a complex combination of genetic and environmental factors.
Atopic diseases	L20, J30.1-30.4, J30.8, J30.9	It includes atopic dermatitis and allergic rhinitis. Atopy is a predisposition toward developing certain allergic hypersensitivity reactions. Atopic dermatitis is a chronic inflammatory skin disease associated with cutaneous hyperreactivity to environmental trigger. It is believed to be the product of interactions between susceptibility genes, the environment and immunologic responses (Leung et al., 2004). Allergic rhinitis is characterized by one or more symptoms including sneezing, itching, nasal congestion, and rhinorrhea (Skoner, 2001).
Type 1 Diabetes	E10	Type 1 diabetes is a chronic auto-immune mediated disease. The body destroys beta cells, which are cells located in the pancreas that produce and segregate insulin, the hormone that regulates glucose levels in the blood. In type 1 diabetes patients, the body is unable to regulate glucose levels. This disease develops in genetically susceptible individuals, but the medical literature has recognized environmental factors as crucial in the triggering and development of the condition (Knip and Simell, 2012).
Obesity	E65-E68	It includes obesity, overweight, localized adiposity and other hyperalimentation. Obesity is defined as abnormal or excessive fat accumulation that may impair health. The prevalence of overweight and obesity among children and adolescents aged 5-19 has risen dramatically from just 4% in 1975 to just over 18% in 2016.[b] Although obesity is most commonly caused by excess energy consumption (dietary intake) relative to energy expenditure, the etiology of obesity is highly complex and includes genetic, physiologic, environmental, psychological, social and economic factors (Wright and Aronne, 2012). Recent research highlights the role of gut microbiota in the development of obesity (Ottosson et al., 2018).

[a] http://www.who.int/respiratory/asthma/en/
[b] http://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight

Table 3.A.3: Public Holidays in Finland (Year 1992)

Public holiday	Date (1992)	Weekday (1992)
New Year's Day	January, 1	Wednesday
Epiphany[a]	January, 6	Monday
Good Friday[b]	April, 17	Friday
Easter Sunday[c]	April, 19	Sunday
Easter Monday[d]	April, 20	Monday
May Day	May, 1	Friday
Ascension Day[e]	May, 28	Thursday
Whit Sunday[f]	June, 7	Sunday
Midsummer Eve[g*]	June, 19	Friday
Midsummer Day	June, 20	Saturday
Finnish Independence Day	December, 6	Sunday
Christmas Eve[*]	December, 24	Friday
Christmas Day	December, 25	Saturday
Boxing Day	December, 26	Sunday

[a] Epiphany was moved to January 6 in 1992. Previously, Epiphany was the Saturday following January 5. [b] Moveable Friday before Easter Sunday. [c] Moveable Sunday following the first full moon on or after March 21. [d] Moveable Monday after Easter Sunday. [e] Moveable Thursday 39 days after Easter Sunday. Until 1992, the Ascension Day was the Saturday before the Thursday1. [f] Moveable Sunday 49 days after Easter Sunday. [g] First Friday on or after June 19. [*] No legal status as a public holiday, but included in collective labor agreements.

Table 3.A.4: Relation of the instrument with discretionary diagnoses vs. medical emergencies

	(1) Dystocia	(2) Suspected fetal suffering
Preleisure day	-0.002**	-0.001
	(0.001)	(0.001)
Normal shift	-0.002**	0.005***
	(0.001)	(0.001)
Normal shift*Preleisure	0.005**	0.002
	(0.002)	(0.002)
Observations	392560	392560
Adjusted R^2	0.074	0.057
Controls	YES	YES
F-statistic	9.211	0.607

This table shows the results from our usual first-stage specification, but with the following dependent variables: in column 1, an indicator for prolonged or obstructed labor; in column 2, an indicator equal to 1 if fetal scalp pH measurements were taken during labor. All specifications include hospital, year, and month of birth fixed effects, and the full set of controls as described in equation (3.2). Robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 3.A.5: First stage – Medical Professional Mothers vs. Others

Sample:	All non-medical mothers		Non-medical mothers with university education		Medical mothers	
Unplanned CS	(1)	(2)	(3)	(4)	(5)	(6)
Normal shift	0.014***	0.017***	0.014***	0.018***	0.025***	0.025***
	(0.001)	(0.001)	(0.002)	(0.005)	(0.006)	(0.005)
Preleisure day	0.001	-0.002	-0.001	-0.002	0.000	-0.002
	(0.002)	(0.002)	(0.003)	(0.007)	(0.007)	(0.007)
Normal shift*Preleisure	0.016***	0.015***	0.015***	0.014***	-0.004	-0.003
	(0.003)	(0.003)	(0.005)	(0.012)	(0.012)	(0.012)
Observations	367825	367825	147463	147463	22526	22526
Adjusted R^2	0.008	0.071	0.008	0.072	0.006	0.068
Controls	NO	YES	NO	YES	NO	YES
Mean of Y	0.146	0.146	0.152	0.151	0.154	0.154
First-stage F	28.998	27.378	10.428	9.609	0.092	0.067

This table shows the usual first stage, with unplanned C-section as dependent variable, for different groups of mothers: all mothers not in the medical profession (columns 1-2), for mothers not in the medical profession with university education (columns 3-4), and for mothers in the medical profession (5-6). Medical mothers include doctors, nurses and midwives. All specifications include hospital, year, and month of birth fixed effects and the full set of controls as described in equation (3.2). Robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

4

THE IMPACT OF THE FEMALE ADVANTAGE IN EDUCATION ON THE MARRIAGE MARKET

4.1 INTRODUCTION

Recent decades have seen a decline and reversal of the traditional gender gap in education in favor of men in many countries. In the US, for instance, in 1960 there were above 50% more men than women with university degrees in the working-age population. This difference gradually declined during the second half of the twentieth century, and by the 2000's the gap had been reversed.[1] This period also witnessed dramatic changes to family structure. Fertility rates fell, dropping below replacement levels in many countries; the age at first birth increased, and marriage now takes place later and less often. This transformation of the family and the increases in female education and labor force participation have been studied as closely connected phenomena (Goldin, 2006; Goldscheider et al., 2015; Oláh et al., 2018).

The reversal of the gender gap in education and the emerging female advantage could have far-reaching consequences for the family. Traditional heterosexual mating patterns have been characterized by men marrying women at most as educated as themselves. These patterns are

[1]See this evolution in Figure 4.A.1 with data for the US for the OECD average.

likely to be challenged as women's education levels exceed those of men but, so far, we lack evidence on how family formation may be affected as a consequence.

The aim of this paper is to understand the causal impact of relative increases in women's educational attainment on marriage and fertility. While the direct consequences of educational attainment for women and men have been widely researched, the focus here is on the effects of changes in the gender gap in education in the marriage market. To investigate this question, I exploit the gradual implementation of a school reform in Finland that increased the female advantage in education.

Conditional on own educational attainment, changes in the educational composition of the marriage market might affect union formation and family outcomes, as these have been shown to depend on the availability of suitable partners (Abramitzky et al., 2011; Angrist, 2002). In the context of marriage models à la Becker (1973), a larger female advantage could potentially enable more specialization between spouses, and thus increase the gains from marriage.[2] On the other hand, if individuals prefer a partner with their same level of education, we would expect an increasing mismatch between the distributions of educational attainment of men and women to lower marriage rates, and potentially fertility. In particular, we might expect there to be an excess number of high-educated women and low-educated men who are unable to find a match. This effect would be reinforced in the presence of gender identity norms that make a situation where the wife has higher education than her husband particularly undesirable (Bertrand et al., 2015; Greitemeyer, 2007; Hitsch et al., 2010).

[2]These types of models predict positive assortative matching in education, but this only refers to the ranks of individuals in their gender-specific distribution of traits. Absolute differences in the education levels between men and women play no significant role in this context (Bertrand et al., 2015). Education is seen as an input for both market and non-market sectors. While the closing of the male-female gap in education could reduce the gains from specialization, if the new female advantage in education becomes larger than the former male advantage, gains from specialization could in principle increase, with an inversion of the role of spouses.

Finland implemented a large school reform in the 1970s, transforming the former selective school model, where students were separated into different tracks at age 11, into a comprehensive system where they were kept together until age 16. The choice between vocational and academic track was thus delayed from age 11 to age 16, and a national curriculum was introduced. This reform has been found to widen the gender differences in education, increasing the female advantage in pursuing the academic track and entering into university (Pekkarinen, 2008).[3]

The reform followed a gradual implementation plan, with different municipalities adopting the new system in different years during the period 1972-1977. This adoption path generates variation in exposure to the new school system within municipalities across cohorts, and within cohorts across municipalities, which can be used to identify the impact of individual exposure to the reform. Crucially, I can also exploit variation in the degree of exposure to the reform of a person's marriage market, even conditional on own exposure. This is because marriage markets do not coincide fully with municipality-cohort groups, given that individuals do not marry only within municipalities or within cohorts—men tend to marry slightly younger women.

Exploiting these sources of variation and using rich data from Finnish administrative registers, I first show that the reform increased the female advantage in educational attainment. I find that the female-male gap in continuing education beyond secondary school increased by 19%, and the gender gap in university education was reversed. I then estimate the impact of higher marriage market exposure to the reform, conditional on own exposure, on marriage and fertility patterns. In my baseline specification, marriage markets are based on region of birth and on the

[3]A potential explanation for the differential effect of this reform is related to the gender differences in the timing of puberty, with girls entering adolescence before boys. The gender gap in maturity by age 16 might exacerbate differences in academic performance and aspirations, and educational choices at this age might be affected as a result (Pekkarinen, 2008).

whole distribution of the age gap within couples in pre-reform cohorts. I measure marriage market exposure to the reform as the proportion of people in a person's marriage market who were enrolled in the new school system. In marriage markets with higher exposure there was thus a larger female advantage in education.

My results show that in marriage markets with a larger female advantage in education there were declines in marriage and fertility, so that men were more likely to be single by age 40 and had fewer children. In particular, a one standard deviation increase in marriage market exposure to the reform decreases the probability of being married or cohabiting by 1.4% and the number of children by 1.7% for men. These effects are sizeable compared to the changes in family structure that took place in Finland during this period. For instance, an increase in marriage market exposure from the 25th to the 75th percentile of the distribution can account for around 20% of the actual decline in the share of men who are in a relationship observed during these decades. Importantly, this increase in bachelorhood is not driven by a decrease in the propensity to marry of women who became more educated as a result of the reform, as the reform had if anything a positive direct effect on women's marriage and fertility.

These results are based on a reduced-form analysis, and do not rely on the assumption that only the gender gap in education changed in marriage markets more affected by the reform.[4] Rather, I claim that changes in the gender gap in education are an important channel driving these findings, and provide suggestive evidence supporting this interpretation. First, consistent with the effects being driven by a 'mismatch' between high-educated women and low-educated men, I find stronger negative effects for these two groups. Moreover, I exploit heterogeneity

[4]Previous studies have found that the Finnish comprehensive school reform increased intergenerational mobility and decreased inequality in mortality and cognitive skills by parental income (Kerr and Pekkarinen, 2013; Ravesteijn et al., 2017; Pekkarinen et al., 2009). We might thus expect that in more affected marriage markets there is also less social inequality.

in the baseline gender gap in education to show that marriage and fertility declined more in marriage markets where 'mismatch' increased the most as a result of the reform.

While the increasing mismatch between the educational distributions of men and women seems to be a key driver, the results are not solely explained by it: there are also negative effects in marriage markets where the male-female educational gap was reversed, but mismatch did not increase in absolute terms. My findings thus suggest that also the sign of the gender gap, and not only its size, matters, consistent with the importance of gender identity norms.

I also provide suggestive evidence that in marriage markets with a larger female advantage in education men became more likely to marry a woman more educated than themselves, and the average age gap within couples decreased. I do not find any effect on the probability of divorce. Lastly, my results suggest that these changes in family structure might have had negative consequences for men's mental health and health behaviors, especially for those with low level of education.

This paper contributes to several strands of the literature. It first contributes to the studies on the implications of the reversal of the gender gap in education. So far, these works have been descriptive in nature. For instance, Esteve et al. (2012, 2016) study the association between the reversal of the educational gender imbalance and patterns of assortative mating, and show that, as the female advantage in education increases, so does the prevalence of couples in which the wife has more education. Schwartz and Han (2014) document that, while in the past couples where the wife is more educated than her husband were more likely to divorce, this difference has attenuated over time. I contribute to this literature by providing causal estimates of the effect of an increase in the female advantage in education on a set of family outcomes.

Second, this paper speaks to the literature on the causal impact of women's education on fertility and marriage outcomes. This literature generally finds that, in developed countries, increases in educational attainment at the lower end of the distribution (such as those induced by extensions of compulsory schooling) decrease teenage births, but have small or even positive effects on completed fertility (Black et al., 2008; Fort et al., 2016; McCrary and Royer, 2011; Monstad et al., 2008).[5] Regarding marital outcomes, higher female education has been found not to affect the probability of marriage, but to improve spouse quality (Anderberg and Zhu, 2014; Lefgren et al., 2006; McCrary and Royer, 2011).[6]

My results on the effect of direct exposure to the reform are in line with this previous evidence. I find that being exposed to the new school system, which led to higher educational attainment for women, does not have significant effects on the probability of marriage, and has a small positive impact on fertility. More importantly, my findings show that, beyond the impact of individual changes in education, changes in the relative levels of education of men and women in a given marriage market also affect family outcomes. In this sense, this paper is also related to a broad literature on how changes in marriage market conditions, and in particular sex ratios, affect the family (Angrist, 2002; Charles and Luoh, 2010; Abramitzky et al., 2011; Mechoulan, 2011; Lafortune, 2013; Brainerd, 2017; Grosjean and Khattar, 2019). My work is more closely connected to the scarce papers within this literature which focused on education-level specific sex ratios (Negrusa and Oreffice, 2010), or even field-of-study specific ratios (Pestel, 2017).

Finally, this study is related to the literature exploring the consequences for the family of changes in the relative position of men and

[5]The relationship between schooling extensions and fertility seems to depend, at least in part, on the institutional context. For instance, Cygan-Rehm and Maeder (2013) find that extensions of compulsory schooling are related to decreases in total fertility in Germany, where the opportunity cost of childrearing is high. Similarly, Fort et al. (2016) finds that female education has a negative effect on fertility in England, but not in continental Europe.

[6]In developing countries, increased female education has been found to delay (and in some cases decrease) fertility, delay marriage and improve spouse quality (Heath and Jayachandran, 2017).

women that violate traditional gender norms. Bertrand et al. (2015) study the causes and implications of relative income within spouses, and find evidence consistent with social aversion to a situation in which the wife outearns her husband. Using a Bartik-style instrument, they show that when, in a given marriage market, women are more likely to earn more than men, marriage rates decline. Autor et al. (2019), in turn, exploit trade shocks to show that relative decreases in men's earnings lead to lower marriage rates and fertility, and to increased premature mortality among men.[7] Tur-Prats (2017) shows that relative decreases in female unemployment levels, compared to male's, increase the incidence of intimate-partner violence in Spanish regions with more traditional gender norms. Lastly, Folke and Rickne (2020) study the tension between women's career success and marital stability. They find that women's promotions, but not men's, increase their probability of divorce, based on the analysis of just-winning and just-losing candidates in parliamentarian and mayor elections in Sweden, and CEO promotions.[8]

In this paper, I study the implications of changes in the relative position of men and women in educational attainment. This is a closely-related but different dimension, which has been ignored so far, despite being highly relevant in the context of most developed countries.[9] My findings corroborate that relative advances in women's economic position

[7]In a related paper, Kearney and Wilson (2018) use the fracking boom and find that increases in men's earnings potential increase marital and non-marital births, but not marriage.

[8]Similarly, Stuart et al. (2018) find that winning a Best Actress Oscar increases actresses' probability of divorce, while the same is not true for Best Actor Oscar winners.

[9]The reversal of the gender gap in education has been a common phenomenon in most developed and some developing countries in recent years, certainly more common than the closing of the gender wage gap. In fact, while educational attainment is related to earnings potential, changes in the gender gap in educational might not necessarily lead to a reversal of the wage gap: education and labor market segregation, motherhood penalties, and gender norms might all complicate this relation (Klesment and Van Bavel, 2017).

can generate frictions in marriage markets.

The rest of the paper is structured as follows. In section 4.2 I describe the content and implementation of the Finnish comprehensive school reform. I section 4.3 I lay out the identification strategy. Section 4.4 describes the data used and provides descriptive statistics. Section 4.5 shows the results, section 4.6 provides supplementary analyses and robustness checks to corroborate the main findings, and section 4.7 concludes.

4.2 Background: Finnish Comprehensive School Reform

In the 1970s, Finland transformed its school system and adopted a comprehensive school model, with the aim of equalizing educational opportunities for all students. Similar reforms had taken place some years before in Sweden (Meghir and Palme, 2005; Meghir et al., 2018) and Norway (Aakvik et al., 2010; Monstad et al., 2008).

Before the reform, Finland had a selective school system. Children entered in primary school at age 7, and there were only four years of common education for all students. At age 11, they could choose to apply for admission to a general secondary school or to continue in primary school. Admission was based on teacher recommendations, an entrance exam, and primary school grades. Those admitted continued their education in a general secondary school for five more years, and at age 16 were eligible to attend an upper secondary school (for two years) and, later, university. Those who were not admitted, or did not apply, stayed in primary school for two more years. By the beginning of the 1970s, most primary schools offered continuation classes (civic schools), which offered a more practically-oriented education, such that virtually all students remained in school until age 16 (Pekkarinen, 2008). After civic school, students could finish their education or continue with vocational training, but could not attend upper secondary schools.

With the implementation of the reform, the former primary, general secondary and civic schools disappeared and were replaced by comprehensive schools. Comprehensive schools offered the same educational content to all students for nine years, from age 7 to 16. After this compulsory education, students could choose to either apply to an upper secondary school, apply to a vocational school,[10] or stop studying.

The reform thus implied several changes. First, it delayed the choice of academic or vocational track from age 11 to age 16. Second, it meant that all students would now be together in the same facilities and exposed to the same national curriculum for nine (instead of four) years. However, it did not, in practice, extend compulsory schooling, as most students were already enrolled in school for nine years before the reform (Pekkarinen, 2008).

The adoption of the reform was approved by parliament and legislated in the 1968 School Systems Act (467/1968). The reform was mandated to be implemented gradually from 1972 to 1977, with the order of adoption being determined geographically. It started with the northern municipalities, which had lower levels of educational attainment. The plan of adoption is described in Figure 4.1. The transition was overseen by regional school boards (Pekkarinen et al., 2009). In the year of implementation of the reform in a given municipality, all students in the first five grades were enrolled directly in the comprehensive school, while those in the sixth grade and above continued their education in the pre-reform system.

[10] Admission to either track was based on comprehensive school grades only.

Figure 4.1: Year of adoption of the reform by municipality

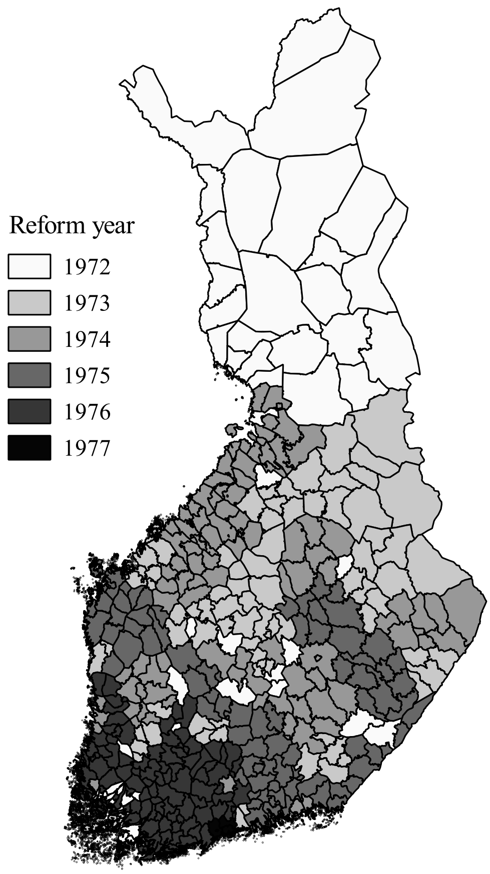


Table 4.1: Cohorts exposed to the new school system by reform year of municipality

	Reform year					
Year of birth	1972	1973	1974	1975	1976	1977
<=1960						
1961	X					
1962	X	X				
1963	X	X	X			
1964	X	X	X	X		
1965	X	X	X	X	X	
>=1966	X	X	X	X	X	X

4.3 IDENTIFICATION STRATEGY

4.3.1 *Effect of the reform on the gender gap in education*

The gradual adoption of the comprehensive school system, as described in section 4.2, generated variation in exposure to the new system across municipalities within cohorts, and across cohorts within municipalities. This variation is illustrated in Table 4.1. All students turning 11 in the year of adoption of the reform (who would start their fifth grade in that academic year) and all the younger ones were enrolled in the new system, while those turning 12 or more were never exposed. For instance, among students living in municipalities that implemented the reform in 1972, all those born in 1960 and before were never in the new system, while all those born in 1961 and afterwards were exposed to it.[11]

I will leverage this variation to first identify the impact of the reform

[11]All of them were exposed to the change in the tracking age from age 11 to 16. The years of exposure to the new curriculum depended on their age at the time of the reform. For instance, those that were in fifth grade when the reform was implemented were exposed to the new curriculum for four years, those in fourth grade were exposed to it for five years, and so on. This information is summarized in Table 4.A.1.

on individual educational attainment and the gender gap in education, using a differences-in-differences strategy:[12]

$$\begin{aligned} y_{imrc} =&\beta_0 + \beta_1 OwnExposure_{mc} + \mu_c + \gamma_r \times t + (\beta_2 \\ &+ \beta_3 OwnExposure_{mc} + \nu_c + \lambda_r \times t) \times F_i + \delta_m + \varepsilon_{imrc} \end{aligned} \tag{4.1}$$

where y_{imrc} is an indicator of educational attainment of individual i, born in municipality m (located in region r) in cohort c; $OwnExposure_{mc}$ takes value 1 if cohort c from municipality m was affected by the school reform; μ_c are cohort fixed effects; $\gamma_r \times t$ are region-specific linear trends (in cohort year); F_i is an indicator for female gender, and δ_m are municipality of birth fixed effects. Standard errors are clustered at the municipality of birth level. I will present results on the direct impact of the reform separately on men and women, and on the gender gap in education (which will be captured by $\widehat{\beta_3}$). In section 4.6.2 I also discuss and show results with alternative specifications, such as including municipality-specific trends instead, or partialling out region-specific linear pre-trends.

One necessary condition for the causal interpretation of these results is that the timing of the adoption of the reform for different municipalities was unrelated to trends in educational attainment. To study whether this assumption is likely to hold, I perform an event study exercise in which I estimate changes in educational attainment by cohort, with cohorts normalized with respect to the first exposure to the reform in each municipality. For example, in municipalities implementing the reform in 1972, the 1960 cohort would have value -1, as it was the last cohort not exposed to the reform; the 1961 cohort would have value 0, the 1962 cohort would have value 1, and so on. I run the following regression:

[12] Similar specifications have been used by papers studying the effects of the Finnish comprehensive school reform (Kerr and Pekkarinen, 2013; Pekkarinen, 2008; Pekkarinen et al., 2009) and other similar reforms in other Nordic countries (e.g. Meghir and Palme, 2005; Meghir et al., 2018; Monstad et al., 2008).

$$y_{imc} = \sum_{t \neq -1} \gamma_t + \mu_c + \delta_m + \varepsilon_{imc} \tag{4.2}$$

where y_{imc} is an indicator of educational attainment, γ_t are coefficients on indicators for number of cohorts relative to first exposure to the reform, and t runs from -10 to 4. The indicator for $t = -1$ is excluded, such that coefficients represent changes in educational attainment with respect to the last non-affected cohort in a municipality. μ_c and δ_m are cohort and municipality of birth fixed effects, respectively. The results of this exercise are shown in section 4.5.1. Figure 4.A.2 further shows that municipalities that adopted the reform earlier (in years 1972-74) and those that adopted it later (in 1975-1977) where following similar marriage and fertility trends in pre-reform cohorts.

There are some potential caveats when using the variation generated from the adoption of the comprehensive school system in a differences-in-differences setting (Pekkarinen, 2008). First, as shown in Figure 4.1, there were some municipalities in southern parts of the country which were assigned to implement the reform earlier than the rest of municipalities surrounding them. Although Table 4.A.2 shows that these localities did not present different educational characteristics than others within their region, one could still be worried that this choice might have been not random. Second, in the Helsinki region, which was assigned to implement the reform in 1977, some municipality-run general secondary schools deviated from the existing selective system by taking in whole cohorts of students already some years before the official creation of comprehensive schools. As a result, in this region the reform might have been redundant. This would potentially lead to underestimation of the effects of the reform, given that 'treated' units will serve as controls. To assess the impact these two features have on the results, in section 4.6.2 I perform robustness checks in which I exclude individuals from the Helsinki region and from these 'outlier' municipalities that implemented the reform before their surrounding localities did.

Recent work on differences-in-differences methods by Goodman-Bacon (2018) highlights other potential concerns with this type of estimators. He shows that, in models with variation in treatment timing, the diff-in-diff estimator can be seen as a weighted average of all two-way fixed effects diff-in-diffs that compare timing groups to each other (and to always-treated and never-treated units, if these exist). When treatment effects vary over time, relying on comparisons that use earlier-treated units as controls might bias the estimator. In order to assess whether, in my setting, this is likely to affect the estimates, I perform the Goodman-Bacon (2018) decomposition, which allows one to see what type of comparisons have the most weight for the aggregate estimator.[13] The results show that 84% of the weight comes from comparisons that use earlier-treated units as treatment and later-treated units as controls. Moreover, comparisons with earlier-treated units as controls, which account for the remaining 16% weight, give almost identical point estimates (see Table 4.A.3). In consequence, time-varying effects are unlikely to be a source of bias in my specification.

4.3.2 *Effect of marriage market exposure to the reform on family outcomes*

In order to study how reform-induced changes in the gender gap in education in the pool of potential mates affect marriage and fertility, I regress different family outcomes on a measure of marriage market exposure to the reform. Marriage market exposure to the reform is calculated as the proportion of people in a person's marriage market who were enrolled in the new school system.

Crucially, these regressions also control for whether a given person was herself enrolled in the new system, as this in itself could affect

[13] The decomposition was performed using the `bacondecomp` Stata package (Goodman-Bacon et al., 2019).

their family outcomes. We can separate marriage market exposure from own exposure to a certain extent, given that marriage markets do not coincide fully with municipality-cohort groups, as individuals do not marry only within cohorts—in particular, men tend to marry slightly younger women, while women tend to marry slightly older men—and because marriage patterns are broader than municipalities in geographical terms. For instance, among those who marry from pre-reform cohorts, only 24% of people marry someone born in the same municipality, while 53% of them marry someone born in the same region; less than 12% are married to someone from the same cohort, while more than 50% are in couples where the husband is from 0 to 3 years older than the wife.[14] The gradual implementation of the reform, together with these standard features of the marriage market, generate variation in the degree to which someone's marriage market is exposed to the reform, conditional on that person's individual exposure.[15]

I thus run the following type of regressions:

$$\begin{aligned} y^g_{imrc} = \alpha_0 + \alpha_1 MarriageMarketExposure^g_{rc} + \alpha_2 OwnExposure_{mc} \\ + \mu_c + \delta_m + \gamma_r \times t + \upsilon^g_{imrc} \end{aligned} \tag{4.3}$$

where y^g_{imrc} is the outcome of individual i, of gender g, born in municipality m of region r in cohort c; $MarriageMarketExposure^g_{rc}$

[14] The distribution of the age difference within couples, calculated as husband's minus wife's age, for men and women in pre-reform cohorts is shown in Figure 4.A.3.

[15] To see this, consider for instance the case of men born in 1960. These men were not exposed to the reform in any part of Finland. However, in municipalities that implemented the reform in 1972, women born in 1961 or later were enrolled in the new system. Hence, the marriage market of 1960 men was substantially exposed to the reform. This exposure was lower in municipalities that adopted the reform later. For example, in municipalities that implemented the reform in 1977, the marriage market of the 1960 cohort of men was barely affected by the reform. Moreover, the fact that not all contiguous municipalities implemented the reform in the same year gives rise to additional variation in marriage market exposure. Figure 4.A.4 shows how even within regions (with borders marked in thicker lines) there is variation in reform timing.

indicates the proportion of women (men) in a man's (woman's) marriage market who were exposed to the new school system; $OwnExposure_{mc}$ takes value 1 if cohort c from municipality m was affected by the school reform; μ_c are cohort fixed effects; δ_m are municipality of birth fixed effects; $\gamma_r \times t$ are region-specific linear time (cohort) trends, and standard errors are clustered at the municipality of birth level. These regressions are run separately for men and women.

I measure marriage market exposure in different ways. In my preferred measure, I consider individuals born in the same region as belonging to the same marriage market.[16] I then use the distribution of the age difference between couples in pre-reform cohorts, separately for men and women (see Figure 4.A.3), to impute the probability that person j belongs to person i's marriage market based on the age gap between the two. These probabilities are used as the weight that person j has on i's marriage market.[17] Specifically, marriage market exposure for individuals of gender g, born in region r in cohort c, is calculated as a weighted average of exposure to the reform in their marriage market, as follows:

$$MarriageMarketExposure^{g}_{rc} = \sum_{m' \in r} \sum_{c'} (\widehat{\omega}^{g'}_{c',c} \times w^{Pop}_{m'c'}) OwnExposure_{m'c'} \tag{4.4}$$

where $\widehat{\omega}^{g'}_{c',c}$ is the estimated probability that an individual of gender g' and from cohort c' belongs to the marriage market of individuals of gender g from cohort c, based on the age difference between the two (and their gender); $w^{Pop}_{m'c'}$ are weights for the population size of cohort

[16]There are currently 19 regions in Finland, with the number of municipalities per region varying from 9 to 57 (median of 27). Figure 4.A.4 shows the map of Finland with the delimitation of regions and municipalities, together with the reform implementation year.

[17]Figure 4.A.5 shows, as an example, the resulting weights that men have for 1960 women's marriage market (in panel a) and that women have for 1960 men's marriage market (panel b) based on their year of birth.

c' in municipality m', and $OwnExposure_{m'c'}$ is an indicator equal to 1 if individuals from cohort c' and municipality m' in region r were exposed to the reform (where c' can be equal to c, and m' can be equal to m). Figure 4.A.6 shows the distribution of marriage market exposure separately for those exposed and not exposed to the reform themselves.

One potential concern is that the definition of the relevant marriage market changes as a result of the reform itself. In Table 4.A.4 I explore whether this is likely to be the case. Using the specification in equation (4.1), I check if exposure to the reform changed the average age gap within the couples, the probability of marrying someone from the same region, or the probability of living by age 40 in a different region than that of birth. The results show that the reform did not significantly affect any of these aspects.

Nevertheless, I also explore the sensitivity of the results to using alternative marriage market definitions, including the following: a) considering only individuals born in the same region and with an age difference of 0 to 3 years in favor of the man; b) using the weights based on the age difference as in the baseline definition, and also weights based on the distance between municipalities of birth;[18] c) using weights based on age difference (as in baseline definition), together with weights for the surrounding municipalities of birth based on the frequency of marriage of people from those municipalities in pre-reform cohorts. In section 4.6.1 I discuss how results vary with these different measures of exposure.

[18]In particular, I calculate the probability that a person from municipality m' belongs to the marriage market of a person from municipality m as the (normalized) inverse of the distance between the two municipalities. Figure 4.A.7 shows, as an example, the weight that individuals from each municipality have in the marriage market of people from Tampere depending on the distance.

4.4 Data and Descriptive Statistics

4.4.1 *Data*

The main data source for the analysis is the FLEED-FOLK (Finnish Longitudinal Employer– Employee Data) dataset provided by Statistics Finland. It contains rich information about all individuals permanently living in Finland at the end of a given year. For the main part of the analysis, I use the files for years 1988-2006 and select all individuals born in Finland and aged 40 in each year. Hence, my sample consists of all Finnish-born individuals from cohorts 1948-1966 who are still living in Finland by age 40.[19] The region of Åland islands is excluded from the sample due to lack of information about the year of adoption of the reform. As a result, my sample consists of 1,460,448 individuals from 430 municipalities in 18 different regions.

The database contains basic information about the year, municipality and region of birth, as well as the following variables regarding each statistical year: municipality of residence, civil status and family structure, educational attainment, and labor market status, among others. Besides the basic file, I use the supplementary marriage and family modules, which contain more detail about the history of marriages and divorces (including the spouse identifier), and about children (including their year of birth and identifiers). I supplement the information about children using the Finnish Medical Birth register, which contains information about all births taking place in Finland from 1987.

I combine the information about the year and municipality of birth with the year of adoption of the reform by municipality (as depicted in Figure 4.1) to construct a binary variable indicating if individuals were exposed to the new school system or not. Since I only know the municipality of birth, rather than the municipality where children were living at school age, estimates of this exposure variable could be affected

[19] I use information for cohorts up to 1970 for descriptive statistics.

by measurement error, likely leading to underestimation of the effects. For each person, I then construct a measure of exposure to the reform of their marriage market as a weighted average of the individual exposure indicators of those people in their marriage market, as explained in section 4.3.2.

In order to study the impact of the reform on educational attainment, I construct an indicator variable for having more than secondary education, and an indicator for having at least a bachelor's degree or equivalent level.[20] In terms of marriage outcomes, I use the history of marriages to construct indicators for having married and for having divorced by age 40, to construct an indicator for being married or cohabiting at this age, and to obtain the identifier of the first spouse. Using the spouse identifier I collect information about their year and place of birth and their educational attainment. This allows me to construct indicators for whether a person is equally, more, or less educated than their spouse, and for the age difference between them. The analysis focuses on heterosexual couples, given that there are virtually no same-sex couples in the data for the cohorts of the sample.[21] I also examine the following fertility-related variables: the number of children a person has by age 40, and an indicator for childlessness at this age.[22]

In supplementary analyses I also explore the following labor market outcomes: annual labor earnings and an indicator for being employed, both at age 30 and age 40. Finally, I combine these datasets with the Finnish Hospital Discharge Register, which contains information about

[20]The available variables for educational attainment are left-censored, and only distinguish among education levels starting from the upper secondary level. As a result, for lower levels, one can only know that a person did not achieve upper secondary education, but one cannot tell whether they finished compulsory schooling or dropped out.

[21]Registered partnerships for same-sex couples were introduced in Finland in 2002, and same-sex marriage was not legalized until 2017.

[22]Information on biological children is only available in the register from 1989 onwards. In the analysis of fertility outcomes I thus focus on cohorts from 1949 to 1966.

the diagnosed medical conditions coded in ICD10, medical operations, and the date of diagnoses. I use data from outpatient and inpatient visits from 1998 to 2011 and construct individual indicators for having a visit with a given diagnosis at ages 40-45.[23] This analysis is thus restricted to individuals born from 1958 onward.[24] I look at the following groups of diagnoses: mental health problems and abnormal emotional symptoms (ICD10 F09-F99 and R45), alcoholic liver disease and cirrhosis (K70, K74), and drug overdoses (T36-T51).

4.4.2 *Descriptive statistics*

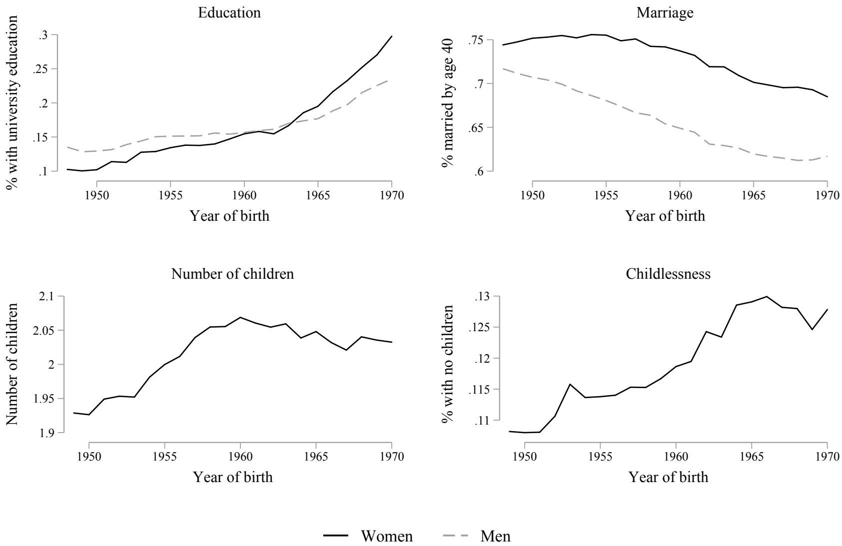
Figure 4.2 presents the aggregate trends in education and family structure in Finland from 1948 to 1970. While at the beginning of this period there were more men than women with university degrees, the gender gap in university education closed with the cohorts born around 1960, and for cohorts born by 1965 there was already a female advantage, which continued to grow thereafter. At the same time, there were substantial changes to family structure. Marriage rates declined over this period: the percentage of men who ever married by age 40 declined by 14%, while there was a 8% decrease for women. The average number of children per woman, which was increasing until the 1960 cohort, plateaued and then started to decrease for younger cohorts. Finally, we see an increase in the proportion of women who do not have any children by age 40 over the whole period.

Figure 4.A.8 shows the distribution of educational attainment for men and women just before (cohorts 1956-60) and just after the reform (cohorts 1966-70). It plots the percentage of men and women in each cohort group with three levels of education: basic (with at most upper secondary education), medium (more than secondary education, but

[23] I consider not only age 40 but the ages 40-45 in order not to have such rare outcomes.

[24] This restriction is due to the data not containing information on outpatient visits prior to 1998.

FIGURE 4.2: Aggregate education and family trends in Finland



Notes: This figure plots the percentage of men and women with university education, the percentage of men and women who were ever married by age 40, the average number of biological children per woman, and the percentage of women who do not have any children by age 40 in Finland by year of birth.

less than university degree), and high (university degree or higher). In the pre-reform cohorts, there were substantially more men than women with low level of education, but also slightly more men than women with university degree. Post-reform cohorts had in general higher educational attainment, with decreases in the percentage of men and women with low education and increasing prevalence of university degrees. This increase was larger for women: the gender gap in having low educational attainment increased from 9.8 to 16.3 percentage points, and the gap in university education was reversed, such that in post-reform cohorts there is a 4 percentage point female advantage.

Finally, Figure 4.A.9 shows the frequency of different types of couples, by relative level of education, in the same pre- and post-reform cohorts. Couples are classified into four groups: couples where none has university education (L-L), couples were both have university education (H-H), couples where only the husband has university education, and couples where only the wife has university education. The most remarkable changes from the pre-to the post-reform cohorts are the decrease in the frequency of low-educated couples, and the increased prevalence of couples where both have university education, and of couples where the wife is more educated than her husband.

4.5 RESULTS

4.5.1 *Impact of the reform on the gender gap in education*

The estimates of the impact of the reform on educational attainment for women and men and on the resulting gender gap, using the specification of equation (4.1), are shown in Table 4.2. The first three columns show the results for the probability of having more than secondary education, while the last three columns have an indicator for having at least university education as dependent variable.

The results show that the reform had a positive effect on women's educational attainment, but virtually no impact on men's education. Women exposed to the reform had a 1.4 pp higher probability of having more than secondary schooling, a 3.6% increase with respect to the pre-reform average, and 0.9 pp higher probability of having university education (a 6% increase). As a result, the female advantage in having more than secondary education increased by 1.7 pp (a 19% increase). The former gender gap in university education in favor of men (1 pp) was reversed, as the female advantage increased by 1.1 pp.

These findings are consistent with previous results by Pekkarinen (2008) showing that the reform increased the female advantage in choosing the academic track and in entering into tertiary education. He

Table 4.2: Reform impact on gender gap in education

	Post-secondary			University		
	Women	Men	Female adv.	Women	Men	Female adv.
Own exposure	0.014***	-0.002	0.017***	0.009***	-0.002	0.011***
	(0.004)	(0.004)	(0.005)	(0.003)	(0.003)	(0.003)
Observations	1460448	1460448	1460448	1460448	1460448	1460448
Adjusted R^2	0.034	0.034	0.034	0.016	0.016	0.016
Pre-reform mean	0.39	0.30	0.09	0.14	0.15	-0.01

This table shows estimates for the impact of direct exposure to the reform on the educational attainment of women and men, and on the female advantage in education. The first three columns have as dependent variable an indicator for more than secondary education, and the last three columns an indicator for university degree. Standard errors (in parentheses) are clustered at the municipality of birth level. The specification includes cohort and municipality of birth F.E., as well as region-specific linear trends. Own exposure takes value 1 for cohorts and municipalities affected by the reform. Pre-reform means refers to average of the dependent variable in the sample of each column for cohorts born in 1956-1960. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

discusses that this differential effect on boys and girls is likely related to gender differences in the timing of puberty, with girls entering into adolescence before boys. While up to age 11 boys and girls have on average developed at the same pace, around this age their trajectories temporarily diverge, and by age 16 the gender gap in maturity might exacerbate the gender differences in academic performance and educational choices. This is consistent with studies showing that late pubertal development is associated with worse academic performance at age 16 and lower total educational attainment (Koerselman and Pekkarinen, 2018).

As discussed in section 4.3, to evaluate the extent to which the timing of the adoption of the reform for different municipalities was unrelated to trends in educational attainment, I perform an event study exercise. In particular, I estimate changes in female educational attainment by cohort, with cohorts normalized with respect to the first exposure to the reform in each municipality.[25]

The results of this exercise are shown in Figure 4.3. Panel (a) presents

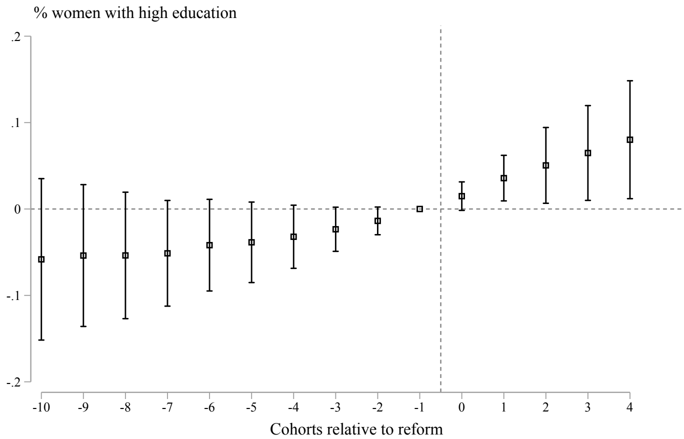
[25]Cohorts up to 1970 are used in order to have a balanced sample.

the results for the whole sample. While none of the pre-trend coefficients are significant, there seems to be an upward trend in female education before the reform took place. This finding could be related to the fact that schools in the Helsinki region, in spite of being scheduled to be among the last to implement the reform, had in practice already started to adopt it some years before. To check if this explains the observed pre-trends, in panel (b) I repeat this exercise excluding observations from the Helsinki region. In this case one cannot see any clear patterns for the cohorts preceding exposure to the reform, and the increases in female education start clearly only after its implementation. This suggests that an important robustness check will be to test the sensitivity of the results to excluding the capital region.

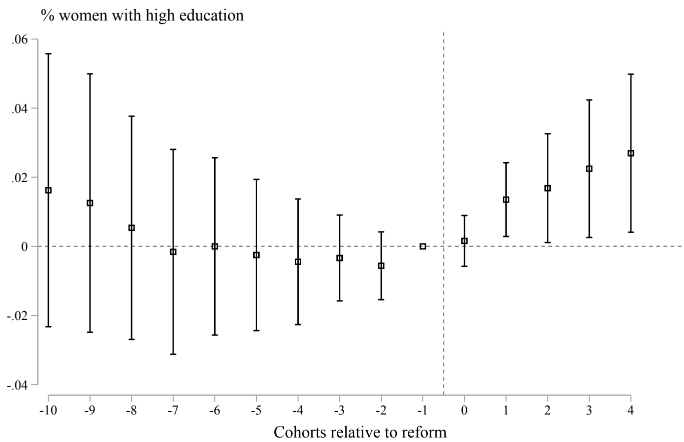
Finally, Table 4.A.5 explores the effect of the reform on gender gaps in the labor market by age 30, using the same specification as in (4.1).[26] The first three columns show results for the effect on earnings. We see that women affected by the reform earned on average 180 euro more by age 30. While no significant effect is found for men, the gender wage gap decreased as a result by around 280 euro (a 5% decrease). The last three columns show that the reform did not affect the probability of being employed by age 30 for either women or men. These results are again consistent with the findings by Pekkarinen (2008) that the changes in the gender gap in education induced by the reform also translated to a certain extent into gender differences in earnings. These changes in relative earnings might thus be part of the channel through which marriage market exposure to the reform affects family outcomes (Autor et al., 2019; Bertrand et al., 2015).

[26] Ideally we would like to observe labor market outcomes as early as possible, before individuals “enter” into the marriage market. However, labor and marriage decisions are likely to be almost simultaneous in many cases, and due to data limitations the earliest the 1960 cohort is observed is at age 28.

Figure 4.3: Percentage of women with high education by cohorts relative to first exposure to the reform



(a) Full sample



(b) Without Helsinki region

Notes: These figures plot the coefficients of a regression of women with high (post-secondary) education on indicators for number of cohorts relative to the one first exposed to the reform in a municipality, following the specification in (4.2). The coefficient on $t = -1$ is omitted, such that coefficients represent changes with respect to the last non-exposed cohort. Panel (a) shows results for the full sample, while in panel (b) the Helsinki region is excluded.

4.5.2 *Impact of marriage market exposure to the reform on marriage and fertility*

This section presents the main results for the impact of marriage market exposure to the reform on family outcomes. I first show that marriage market exposure to the reform, conditional on individual exposure, does not itself affect a person's own level of education. The results are shown in Table 4.A.6: the coefficient of own exposure is not affected by the inclusion of marriage market exposure in the regression, and marriage market exposure does not have any significant impact on men's and women's level of education. The results in this section can thus be interpreted as the effect of changes in the educational composition of the marriage market, separate from changes in own level of education.

The first two columns in Table 4.3 show the estimates of the effect of marriage market exposure on men's marriage outcomes: on the probability of having ever married by age 40 (column 1) and on the probability of being in a couple, either married or cohabiting, at this age (column 2). The results show that marriage market exposure to the reform did not significantly effect the probability of having been in a formal marriage, but decreased the probability of being in a couple by age 40: a one standard deviation increase in marriage market exposure to the reform decreases the probability of being in a couple by 1 pp (a 1.4% decrease). Own exposure to the reform, on the other hand, does not seem to have affected these outcomes.

The last two columns of Table 4.3 show results for the impact of marriage market exposure on the probability of not having had any children by age 40 and on the number of children by this age. Men whose marriage market was more affected by the reform had on average fewer children: a one standard deviation increase in marriage market exposure decreases the number of children by 1.7%. The probability of having at least one child, in turn, does not seem to be affected.

Table 4.A.7 in the Appendix shows that conclusions are similar if I

instead examine women's outcomes. There are also negative, although not significant, effects on the probability of having ever married and the probability of being in a couple at age 40, and a significant increase in the probability of not having any children. Interestingly, the estimates for own exposure show that women who where directly exposed to the reform, and had thus on average higher education, were if anything more likely be in a couple and had more children. These results are in line with previous literature showing that increases in women's education have small effects on completed fertility in industrialized countries, which are even positive in some cases (Fort et al., 2016). This suggests that the negative effects of marriage market exposure for men are not simply driven by the high-educated women in these more affected marriage markets being less likely to marry and having lower fertility. The 'mismatch' between the distributions of educational attainment of men and women seems a more plausible explanation, which I explore further in section 4.5.3.

In order to put the magnitude of these effects in context, I compare the effect sizes with the observed change during the period of study, and with the estimates from related papers analyzing the impact on the family of changes in the gender gap in earnings. Among men born in 1950 in Finland, 78% of them where married or cohabiting at age 40. This number declined to 71% for men born in 1970. An increase in marriage market exposure to the reform from the 25th to the 75th percentile of the distribution, which would lead to an 8% increase in the female educational advantage, can account for around 20% of this decrease. Compared to the results by Bertrand et al. (2015), in turn, I find that the effect on the share of married males of a one standard deviation increase in marriage market exposure to the reform would be roughly equivalent to the effect of a 2.8 pp increase in the probability that a woman earns more than a man in the marriage market.[27]

[27]The definitions of the outcome variables in Bertrand et al. (2015) differ slightly from mine. In their case, the share of married males refers to the proportion of

Table 4.3: Marriage market exposure impact on men's family outcomes by age 40

	Marriage		Fertility	
	Ever married	Married/cohab	Childless	Num children
Marriage market exposure (sd)	-0.003	-0.010**	0.004	-0.031**
	(0.006)	(0.005)	(0.004)	(0.014)
Own exposure	0.004	0.000	-0.002	0.003
	(0.004)	(0.003)	(0.003)	(0.010)
Observations	743911	743911	638569	638569
Adjusted R^2	0.010	0.008	0.011	0.015
Pre-reform mean	0.66	0.72	0.20	1.81

Standard errors (in parentheses) are clustered at the municipality of birth level. This table shows the effect of higher marriage market exposure to the reform on men's outcomes: the probability of having ever been married by age 40 (column 1), the probability of being either married or cohabiting at this age (column 2), the probability of not having had any children by age 40, and the total number of children by this age in the last column. The specification includes cohort and municipality of birth F.E., as well as region-specific linear trends. Marriage market exposure (in standard deviations) indicates the proportion of people in someone's marriage market affected by the reform. Own exposure takes value 1 for cohorts and municipalities affected by the reform. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

4.5.3 *Interpretation of results*

The results from the last subsection show that, on average, higher marriage market exposure to the reform leads to decreases in marriage and cohabitation and fertility. Due to reduced-form nature of the analysis, these findings do not rely on the claim that only the gender gap in education is changing in more affected marriage markets. I argue, however, that changes in the female advantage in education in these markets are an important driver of these effects. This subsection provides several pieces of evidence that support this interpretation.

First, if in more affected marriage markets there is a larger mismatch

males who are currently married in each marriage market, which is defined for broad age groups (e.g. men aged 24-33), so the estimate refers to an average effect across different ages. In my analysis, in turn, this estimate refers to the probability for men of being in a couple (married or cohabiting) by age 40.

between the educational distributions of men and women, such that it becomes more difficult to find a partner with the same level of education as oneself, we would expect larger declines in marriage and fertility for high-educated women and low-educated men. In order to see if this is the case, I explore heterogeneous effects by level of education. Because the reform had a direct effect on educational attainment, at least for women, conditioning on level of education for the whole sample would lead to biased estimates. I will therefore focus on cohorts not exposed to the reform themselves, and exploit variation in degree of exposure in their marriage market only.

Table 4.4 shows, separately for high- and low-educated men and women, respectively, the effect of higher marriage market exposure on the probability of having ever married, on the probability of being married or cohabiting by age 40, on childlessness and on the number of children. The sample is restricted to those who were never exposed to the new school system.[28] Low-educated individuals are defined as those with at most secondary education, while the rest are classified as highly educated. Results defining high educated individuals as those with university degree yield similar results, and are shown in Table 4.A.8.

The results suggest that, among those not directly exposed to the reform, higher marriage market exposure leads to decreases in the probability of having ever married among women with high level of education, but not among the low-educated ones. Results are similar, although a bit smaller, for the probability of being married or cohabiting by age 40. For men, higher marriage market exposure leads to a (non-significantly) higher probability of having married for those with high level of education, but to a slightly lower probability for low-educated ones. Similarly, we see a small decrease in the probability of being married or cohabiting

[28]Individuals from Helsinki region are also excluded given that, as discussed in section 4.3, some were exposed to the new system before the date assigned in the adoption plan.

by 40, which is larger for those with low education. Consistent with this, albeit not always significant, the estimates for fertility outcomes suggest that both increases in childlessness and decreases in the number of children are concentrated in women with high level of education and men with low level of education.

Given that men's level of education does not seem to have been affected by the reform, the same heterogeneity analysis could be performed for them including both those directly and not directly exposed to the reform, in order to have a larger sample and more variation in marriage market exposure. Results from this exercise show significant negative effects on the probability of being in a couple and the number of children only for low-educated men (Table 4.A.9).

All in all, this evidence is consistent with high-exposure marriage markets having a larger mismatch among the educational distributions of men and women, such that there are 'excess' numbers of high-educated women and low-educated men who are unable to find a suitable match.[29]

Following this same line of reasoning, we would expect stronger effects in marriage markets where the absolute gender difference in education increased more as a result of the reform. The male-female gap in (university) education before the reform varied across regions: while in some regions men had a large advantage, in others women had already caught up to a great extent. As a result, the increase in women's education induced by the reform led, in absolute terms, to decreases in the educational mismatch in some markets, to increases in others, and to little change in some (but to a reverse of the gap). I classify regions into two groups: regions in which the gender educational mismatch increased in absolute terms after the reform, and regions in which it did not change or it decreased.[30]

[29]This is despite the fact that a match between high-educated women and low-educated men would give rise to large specialization gains, to the extent that education predicts market productivity (Becker, 1973).

[30]Specifically, I define the change in the gender gap as the difference between the gender gap in absolute value after the reform, and the gender gap in absolute value

Table 4.4: Heterogeneous effects of marriage market exposure by level of education – sample not directly exposed

	Women		Men	
	Low	High	Low	High
A. Marriage outcomes				
Ever married by 40	0.000	-0.012*	-0.002	0.008
	(0.004)	(0.006)	(0.009)	(0.012)
Mean of Y	0.73	0.78	0.62	0.78
Married/cohabiting by 40	0.001	-0.007	-0.010	-0.004
	(0.004)	(0.006)	(0.007)	(0.012)
Mean of Y	0.74	0.78	0.69	0.81
N	329638	166352	374126	139927
B. Fertility outcomes				
Childless	-0.000	0.001	0.015*	-0.004
	(0.003)	(0.005)	(0.008)	(0.014)
Mean of Y	0.11	0.12	0.22	0.16
Number of children	0.007	0.004	-0.007	-0.000
	(0.014)	(0.018)	(0.029)	(0.042)
Mean of Y	2.11	1.99	1.77	1.93
N	287439	146516	314773	122854

This table shows the coefficients of marriage market exposure in separate regressions where the dependent variable is the one indicated in each row. Sample is restricted to individuals not directly exposed to the reform, and divided into men and women with low (at most secondary education) and high (more than secondary education) education level. Standard errors (in parentheses) are clustered at the municipality of birth level. The specification includes cohort and municipality of birth F.E., as well as region-specific linear trends. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

It should be noted that, if the increase in educational mismatch was the only force driving the results, we would not expect to see negative effects on marriage or fertility in marriage markets where the mismatch did not increase. In those markets, the only change induced by the reform was making women more educated than men. The presence of negative effects also in those regions would suggest that not only the size of the gender gap, but also its sign, matter, consistent with the importance of gender identity norms.[31]

I explore heterogeneity by the change in the gender gap in education induced by the reform at the marriage market level in Table 4.5. The first column displays the estimates for regions in which the gender gap in education did not increase in absolute terms, while the last column shows results for those in which it increased. Each row presents estimates of the effect of marriage market exposure from separate regressions with the different dependent variables. In general, we see that the effects are stronger in marriage markets where the reform led to an increase in educational mismatch: higher marriage exposure leads in these regions to declines in the probability of having ever married, and to a lower probability of being in a couple by age 40. However, even in regions where the reform did not lead to an increase in mismatch, higher marriage market exposure has negative effects. In particular, we see that the increase in female childlessness is the same in both groups of regions.

The results from this exercise suggest that, even though increases in educational mismatch seem to be an important driving force, they are not enough to explain the main findings. The fact that higher exposure to the reform in the marriage market has a negative impact on fertility, even where mismatch did not increase, suggests that gender identity

before the reform. To do so I estimate the impact of the reform on the gender gap in university education separately for each region.

[31] Akin to the social norms about relative earnings discussed by Bertrand et al. (2015), there might be a resistance to a situation in which the wife has higher education than her husband.

norms might also play a role.

Table 4.5: Heterogeneous effects of marriage market exposure by change in educational mismatch

	No increase		Increase	
	Women	Men	Women	Men
A. Marriage outcomes				
Ever married by 40	-0.003	0.000	-0.009**	-0.003
	(0.006)	(0.007)	(0.004)	(0.011)
Mean of Y	0.74	0.66	0.75	0.67
Married/cohabiting by 40	-0.000	-0.006	-0.011*	-0.018**
	(0.006)	(0.007)	(0.006)	(0.007)
Mean of Y	0.74	0.72	0.74	0.73
N	364908	378492	351629	365419
B. Fertility outcomes				
Childless	0.010***	-0.002	0.009**	0.007
	(0.003)	(0.005)	(0.004)	(0.006)
Mean of Y	0.12	0.21	0.12	0.20
Number of children	-0.012	-0.017	-0.001	-0.036
	(0.015)	(0.018)	(0.014)	(0.024)
Mean of Y	2.01	1.77	2.08	1.84
N	322464	324535	310729	314034

This table shows the coefficients of marriage market exposure in separate regressions where the dependent variable is the one indicated in each row. The sample in the first two columns consists of regions where the gender gap in university education decreased or did not change as a result of the reform, while the last two columns show results for regions where it increased. Standard errors (in parentheses) are clustered at the municipality of birth level. The specification includes cohort and municipality of birth F.E., as well as region-specific linear trends. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

4.5.4 *Assortative mating and marital dissolution*

Higher marriage market exposure to the reform might also affect other family-related outcomes, such as assortative mating or the probability of marital dissolution. However, the causal pathway to these outcomes is mediated by the impact of marriage market exposure on the probability of marriage itself. With these caveats in mind, in this section I first provide some suggestive evidence about the relationship between marriage market exposure and assortative mating by education and age, and marital dissolution.

The first three columns of Table 4.6 present estimates of the impact of marriage market exposure on the relative level of education within married couples. Higher marriage market exposure is related to an increased probability for men of being less educated than their spouse. This is consistent with previous descriptive evidence by Esteve et al. (2012, 2016) showing that, as the female advantage in education increases in the population, so does the prevalence of couples where the wife is more educated.

The fourth column shows results for the age difference within couples, expressed as husband's minus wife's age, such that it is on average positive. The estimates suggest that higher marriage market exposure decreases the average age difference, and the inspection of different margins reveals that this comes from a decrease in the number of couples where the wife is 4 or more years younger than her husband.[32] Finally, the last column shows that marriage market exposure does not affect the probability of (formal) divorce for men. It might thus be that the decreased probability of being in a couple comes from separations, instead of divorces, or that it is driven by couples that would have never been formally married in the first place.

[32] These results are available upon request.

Table 4.6: Marriage market exposure impact on assortative mating and divorce

	Relative level of education			Age difference	Divorced
	Equal	More	Less	with spouse	by 40
Marriage market exposure (sd)	-0.008 (0.005)	-0.004 (0.003)	0.007** (0.003)	-0.124** (0.060)	0.003 (0.004)
Observations	743911	743911	743911	570897	743911
Adjusted R^2	0.010	0.003	0.007	0.004	0.011
Pre-reform mean	0.45	0.08	0.13	1.74	0.14

Standard errors (in parentheses) are clustered at the municipality of birth level. The dependent variable in columns 1-3 is an indicator equal to 1 if the person's level of education is equal, higher, or lower than that of their spouse, respectively. The dependent variable in column 4 is the age difference between the husband and the wife, and in column 5, an indicator equal to 1 if the man has divorced by age 40. The specification includes cohort and municipality of birth F.E., as well as region-specific linear trends. Marriage market exposure (in standard deviations) indicates the proportion of people in someone's marriage market affected by the reform. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

4.5.5 *Health implications*

Finally, changes in family structure and social values have been shown to be closely related to labor market outcomes (Autor et al., 2019; Coile and Duggan, 2019). Declines in men's value in labor and marriage markets, in turn, have been associated with negative health consequences, like increases in premature mortality, especially from "deaths of despair"—i.e., suicides, and alcohol and drug related problems (Autor et al., 2019; Case and Deaton, 2017; Coile and Duggan, 2019). The combination of data from labor and hospital registers allows me to explore whether in marriage markets with a larger female advantage in education men's labor and health outcomes are affected, and to look at a less extreme health measure than mortality.

The results are shown in Table 4.7. The first three rows show the coefficient of marriage market exposure to the reform from regressions with indicators for different health problems as dependent variables: mental health problems, alcoholic liver disease, and drug overdose. These indicators take value 1 if the person had a hospital visit (inpatient or outpatient) at ages 40-45 with one of these diagnoses. The last row shows results for the probability of being employed at age 40. The first column shows results for all men, while columns 2-3 present heterogeneous results by level of education. We would expect low-educated men to be the most affected, given that the effects on family outcomes were stronger for them.

The estimates suggest that in marriage markets with a higher exposure to the reform, and thus with a larger female advantage in education, men have on average a higher probability of having mental health and alcohol problems, but do not present more hospital visits with substance abuse diagnoses. In line with this, they are also less likely to be employed at age 40. The heterogeneity analysis reveals that these effects are entirely driven by low-educated men.

One limitation of this analysis is that it does not allow us to disentangle the direction of causality; that is, we do not know whether the lower probability of being in a couple affects health and employability, or whether, on the contrary, the lower probability of working makes men less attractive as partners. Coile and Duggan (2019) suggest that these factors are likely to all affect each other. In any case, the analysis in Table 4.A.5 showed that men's earnings or their probability of working at age 30 was not affected by own (and same-cohort) exposure to the school reform. If men's family outcomes are responding to labor market outcomes, it is to an increase in women's earnings, rather than to a direct decline in men's labor market opportunities, as in Autor et al. (2019)'s setting.[33]

[33] The analysis in Table 4.A.5 is based on direct exposure to the reform of a person's cohort and municipality. While it is true that labor markets are likely to be broader than that, and might be very correlated with marriage markets as defined here, there is a priori no reason to expect the relative level of education with respect to younger cohorts to be more determinant of one's labor market outcomes than the relative level of education within one's own cohort.

Table 4.7: Marriage market exposure impact on men's health and employment outcomes

	(1) All	(2) Low educated	(3) High educated
Mental health	0.007*	0.011**	-0.001
	(0.004)	(0.005)	(0.005)
Mean of Y	0.08	0.09	0.04
N	329408	225024	104384
Alcoholic liver	0.006*	0.010*	-0.002
	(0.003)	(0.006)	(0.004)
Mean of Y	0.05	0.06	0.03
N	329408	225024	104384
Substance abuse	0.000	-0.000	0.001
	(0.001)	(0.002)	(0.001)
Mean of Y	0.01	0.01	0.00
N	329408	225024	104384
Employed	-0.020***	-0.027***	-0.001
	(0.005)	(0.006)	(0.008)
Mean of Y	0.80	0.76	0.91
N	743911	528571	215340

This table shows the coefficients of marriage market exposure in separate regressions where the dependent variable is the one indicated in each row. Mental health, alcoholic liver, and substance abuse are indicators equal to 1 if the person had any hospital visit with those groups of diagnoses between ages 40-45. Employed takes value 1 if the person is employed at age 40. Standard errors (in parentheses) are clustered at the municipality of birth level. The specification includes cohort and municipality of birth F.E., as well as region-specific linear trends. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

4.6 Supplementary Analyses

4.6.1 *Measuring marriage market exposure*

As discussed in section 4.3, in my baseline estimation the definition of marriage market exposure consists of a weighted average of individuals' exposure to the reform in someone's marriage market, geographically defined as their region of birth. The weight that different individuals have for someone's marriage market depends on the age difference between them (and gender), based on the distribution of the age gap within couples in pre-reform cohorts.

In this section I discuss how the main results differ when alternative specifications of the marriage market are used. In particular, I consider the following alternatives: 1) focusing only on individuals born in the same region and within the most common age gap, that is, 0-3 years in favor of the man; 2) using weights for the probability that j belongs to i's marriage market based on their age difference (as in the baseline) and their municipality of birth, using the frequency of marriages across different municipalities in pre-reform cohorts; and 3) using weights for the probability that j belongs to i's marriage market based on their age difference (as in the baseline) and the inverse distance of their municipalities of birth.

Results for the different family outcomes using the baseline (column 1) and these alternatives definitions of marriage market exposure are compared in Table 4.8. The main conclusions are not affected by changing the definition of marriage market. The measure of exposure that yields the most different results is the one that uses the age distribution from pre-reform cohorts (as in the baseline) and the normalized inverse distance between municipalities of birth as weights. The estimates using this measure are in most specifications substantially larger than the baseline estimates. On the contrary, the definition that restricts the marriage market to those born in the same region and within an age gap of 0-3 years gives consistent, yet slightly smaller estimates. Part of this difference could be explained by the rigidity of this definition, which captures effects only for a part of the marriage market. This is likely to introduce measurement error that biases the estimates downwards. Overall, however, using one or another definition of marriage market does not affect the qualitative conclusions.

Table 4.8: Marriage market exposure coefficient with alternative marriage market definitions – men's outcomes

	Baseline	Region & 0-3 years	Age dist. & freq. marriage	Age dist. & inv. distance
	(1)	(2)	(3)	(4)
A. Marriage outcomes				
Ever married by 40	-0.003	-0.001	-0.008	-0.009
N=743911	(0.006)	(0.003)	(0.008)	(0.009)
Married/cohabiting by 40	-0.010**	-0.005**	-0.015**	-0.013**
N=743911	(0.005)	(0.002)	(0.006)	(0.005)
B. Fertility outcomes				
Childless	0.004	0.002	-0.002	0.008
N=638569	(0.004)	(0.002)	(0.007)	(0.005)
Number of children	-0.031**	-0.013*	-0.025	-0.025
N=638569	(0.014)	(0.007)	(0.024)	(0.019)

This table shows the coefficients of marriage market exposure in separate regressions where the dependent variable is the one indicated in each row. Different columns use different definitions of the marriage market, as indicated by column titles. Standard errors (in parentheses) are clustered at the municipality of birth level. The specification includes cohort and municipality of birth F.E., as well as region-specific linear trends.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

4.6.2 *Robustness tests*

In this section I check the sensitivity of the main results to alternative control strategies and sample choices. Table 4.9 compares the coefficient of marriage market exposure (expressed in standard deviations) in the baseline specification (column 1) with several alternatives. Each row shows results from separate regressions with different dependent variables. The first column also shows the Romano-Wolf stepdown adjusted p-values to correct for multiple hypothesis testing in the baseline specification. All main results survive this adjustment.

Table 4.9: Robustness of marriage market exposure impact on men's family outcomes

	Baseline (1)	Municipality trends (2)	Region pre-trends (3)	W/o Helsinki (4)	W/o outliers (5)
A. Marriage outcomes					
Ever married by 40	-0.003	-0.005	-0.001	0.001	-0.004
RW p-value=0.604	(0.006)	(0.006)	(0.004)	(0.006)	(0.006)
Married/cohabiting by 40	-0.010**	-0.010**	-0.010**	-0.008	-0.010**
RW p-value=0.039	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
N	743911	743911	743911	671868	711116
B. Fertility outcomes					
Childless	0.004	0.004	0.004	0.002	0.003
RW p-value=0.574	(0.004)	(0.004)	(0.003)	(0.005)	(0.004)
Number of Children	-0.031**	-0.029**	-0.007	-0.025	-0.029**
RW p-value=0.039	(0.014)	(0.014)	(0.011)	(0.016)	(0.014)
N	638569	638569	638569	577743	610398

This table shows the coefficients of marriage market exposure in separate regressions where the dependent variable is the one indicated in each row. Standard errors (in parentheses) are clustered at the municipality of birth level, and bootstrapped in column (3). RW p-value refers to the Romano-Wolf stepdown adjusted p-value to correct for multiple hypothesis testing in the baseline specification. All specifications include cohort and municipality of birth F.E., and additional controls as indicated in column titles. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

In column 2 region-specific linear trends are substituted by municipality-specific linear trends. The results remain virtually unaltered. In column 3, instead of including linear trends, I instead de-trend the dependent variable of region-specific linear pre-trends. To do so, I follow Goodman-Bacon (2018) and estimate pre-trends by regressing the dependent variable on region-specific linear trends for cohorts up to 1960. These trends are then substracted from the full panel. The specification then includes only municipality and cohort of birth fixed effects. Standard errors are bootstrapped to account for the two-step estimation. Using this method has no visible effect on most results, except for the coefficient on the number of children.

In the last two columns I show results using the baseline specification but restricting the sample in different ways. First, as discussed in section 4.5.1, municipalities in Helsinki region had started to implement the reform before they were supposed to according to the adoption plan. To check whether this affects the results, in column 4 I exclude individuals from this region. In spite of the reduced sample size, the estimates remain consistent, albeit a bit smaller, suggesting that the potentially different trends of the capital region are not completely driving the results. Finally, some municipalities were assigned to adopt the reform earlier than most of their surrounding localities (see section 4.3). As discussed by Pekkarinen (2008), the choice of these municipalities is unlikely to have been random. In column 5 I drop individuals from these municipalities and find that results are unaffected. This indicates that the combination of municipality fixed effects and region-specific trends effectively controls for any potential differences in levels or trends.

4.7 CONCLUSION

This paper provides evidence on the effects of the female educational advantage on marriage and fertility outcomes. Exploiting changes in the gender gap in education in the marriage market induced by the Finnish comprehensive school reform, I show that in marriage markets with a higher female educational advantage men are more likely to be single by age 40, and have fewer children. The size of these effects is substantial. An increase in marriage market exposure from the 25th to the 75th percentile of the distribution can explain 20% of the decline in the share of men who are in a couple by age 40 that took place in Finland during this period.

My findings suggest that an important driver of the effects is the increasing mismatch between the distributions of educational attainment of men and women resulting from the reform. As such, the effects are stronger for low-educated men and high-educated women, and larger in marriage markets where the reform increased mismatch more. However, my analysis also highlights that the sign of the gender gap in education, and not only its size, matters. In particular, there are negative effects on family outcomes even in marriage markets where the absolute size of the educational mismatch did not increase, but women became more educated than men. This is consistent with recent work highlighting the importance of gender identity norms (Bertrand et al., 2015; Folke and Rickne, 2020; Tur-Prats, 2017), and with previous evidence from online dating sites showing that men shy away from women more educated than themselves (Hitsch et al., 2010).

One limitation of this analysis is that it does not allow us to quantify the extent to which the estimated effects are driven by an advancement of women's position in the labor market, as opposed to other changes in social status induced by an increase in the level of education. Overall, these results are consistent with the sociological hypothesis that changes in the economic roles of men and women have profound implications for family structure (Goldscheider et al., 2015), and with previous evidence showing that relative advances by women can generate frictions in marriage markets (Bertrand et al., 2015).

Finally, even though a welfare assessment is outside the scope of this paper, the results suggest that the changes in family structure affecting, in particular, low-educated men, might have had negative consequences in terms of their health behaviors and mental health. The question remains as to whether these effects would persist in the future, as social norms evolve towards more egalitarian gender attitudes.

4

APPENDIX

Appendix 4.A

Table 4.A.1: Years of exposure to new curriculum by year of birth and reform year of municipality

	Reform year					
Year of birth	1972	1973	1974	1975	1976	1977
<=1960						
1961	5					
1962	6	5				
1963	7	6	5			
1964	8	7	6	5		
1965	9	8	7	6	5	
1966	9	9	8	7	6	5

Table 4.A.2: 'Outlier' municipalities' education levels for pre-reform cohorts

	Post-secondary		University	
	(1)	(2)	(3)	(4)
Outlier==1	0.002	-0.001	0.006**	0.004
	(0.004)	(0.004)	(0.003)	(0.003)
Observations	430	430	430	430
Adjusted R^2	0.002	0.177	0.007	0.170
Region F.E.	No	Yes	No	Yes

Standard errors in parentheses. This table compares the education levels of pre-reform cohorts (1956-1960) in 'outlier' municipalities and the rest. The dependent variable is the proportion of people in the municipality with more than secondary education in columns 1-2, and the proportion of people with at least a university degree in columns 3-4. Outlier is an indicator equal to 1 if the municipality implemented the reform in a different year than most municipalities of the same region. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 4.A.3: Bacon decomposition results

DD Comparison	Weight	Avg DD estimate
Earlier T vs Later C	0.842	0.017
Later T vs Earlier C	0.158	0.016
Diff-in-diff estimate:		0.017

T=Treatment, C=Control. This table shows the results from the Goodman-Bacon decomposition, performed using the `bacondecomp` Stata package (Goodman-Bacon et al., 2019).

Table 4.A.4: Reform impact on definition of marriage market

	Age gap within couple		Spouse from same region		Living in different region	
	Women	Men	Women	Men	Women	Men
Own exposure	-0.006	-0.046	0.000	-0.003	0.006	0.006
	(0.033)	(0.031)	(0.003)	(0.003)	(0.005)	(0.004)
Observations	1460448	1460448	1460448	1460448	1460448	1460448
Adjusted R^2	0.003	0.003	0.021	0.019	0.074	0.074
Mean of Y	2.39	1.74	0.39	0.36	0.34	0.34

Standard errors (in parentheses) are clustered at the municipality of birth level. This table shows the effect of own exposure to the reform on the average age gap within couples (columns 1-2), on the probability of having a spouse born in the same region (columns 3-4), and on the probability of living in a different region than the region of birth by age 40 (last two columns), for men and women. The specification includes cohort and municipality of birth F.E., as well as region-specific linear trends. Own exposure takes value 1 for cohorts and municipalities affected by the reform * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 4.A.5: Impact of the reform on labor market outcomes at age 30

	Earnings			Working		
	Women	Men	Female adv.	Women	Men	Female adv.
Own exposure	184.020***	-97.125	281.146***	-0.001	0.004	-0.004
	(64.083)	(77.252)	(99.713)	(0.003)	(0.003)	(0.004)
Observations	263500	282148	545648	317145	327439	644584
Pre-reform mean	11062.13	16398.02	-5335.89	0.80	0.89	-0.09

Standard errors (in parentheses) are clustered at the municipality of birth level. This table shows the effect of own exposure to the reform on annual earnings by age 30 (columns 1-3) and on the probability of being employed by this age (columns 4-6), for men and women and the gender gap (expressed as the interaction of female with own exposure). The specification includes cohort and municipality of birth F.E., as well as region-specific linear trends. Own exposure takes value 1 for cohorts and municipalities affected by the reform * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 4.A.6: Own vs. Marriage market exposure: impact on high level of education

	Women		Men	
	(1)	(2)	(3)	(4)
Own exposure	0.014***	0.014***	-0.002	-0.001
	(0.004)	(0.004)	(0.004)	(0.004)
Marriage market exposure (sd)		-0.000		-0.002
		(0.002)		(0.003)
Observations	716537	716537	743911	743911
Adjusted R^2	0.038	0.038	0.016	0.016
Pre-reform mean	0.39		0.30	

Standard errors (in parentheses) are clustered at the municipality of birth level. This table shows that marriage market exposure to the reform does not affect individuals' level of education, once own exposure to the reform is accounted for. The dependent variable is a dummy taking value 1 if the person has more than secondary schooling, and 0 otherwise. The specification includes cohort and municipality of birth F.E., as well as region-specific linear trends. Marriage market exposure (in standard deviations) indicates the proportion of people in someone's marriage market affected by the reform. Own exposure takes value 1 for cohorts and municipalities affected by the reform. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 4.A.7: Marriage market exposure impact on women's marriage and fertility by age 40

	Marriage		Fertility	
	Ever married	Married/cohab	Childless	Num children
Marriage market exposure (sd)	-0.004	-0.004	0.009***	-0.008
	(0.004)	(0.005)	(0.002)	(0.010)
Own exposure	0.002	0.005*	-0.001	0.017*
	(0.003)	(0.003)	(0.002)	(0.009)
Observations	716537	716537	633193	633193
Adjusted R^2	0.007	0.009	0.012	0.019
Pre-reform mean	0.74	0.74	0.12	2.05

Standard errors (in parentheses) are clustered at the municipality of birth level. This table shows the effect of higher marriage market exposure to the reform on women's marriage (probability of having ever married by age 40 and probability of being currently married or cohabiting by this age) and fertility outcomes (probability of not having any children by age 40 and number of children by this age). The specification includes cohort and municipality of birth F.E., as well as region-specific linear trends. Marriage market exposure (in standard deviations) indicates the proportion of people in someone's marriage market affected by the reform. Own exposure takes value 1 for cohorts and municipalities affected by the reform. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 4.A.8: Heterogeneous effects of marriage market exposure by level of education – sample not directly exposed

	Women		Men	
	Low	Uni	Low	Uni
A. Marriage outcomes				
Ever married by 40	-0.001	-0.015	0.003	0.009
	(0.004)	(0.010)	(0.009)	(0.016)
Mean of Y	0.75	0.77	0.64	0.80
Married/cohabiting by 40	-0.001	-0.002	-0.009	0.008
	(0.004)	(0.011)	(0.006)	(0.018)
Mean of Y	0.75	0.77	0.71	0.82
N	434851	61139	442937	71116
B. Fertility outcomes				
Childless	-0.000	0.005	0.010	0.005
	(0.003)	(0.010)	(0.008)	(0.019)
Mean of Y	0.11	0.14	0.21	0.15
Number of children	0.008	-0.018	-0.002	-0.001
	(0.012)	(0.034)	(0.026)	(0.059)
Mean of Y	2.08	1.96	1.79	1.96
N	381327	52628	375334	62293

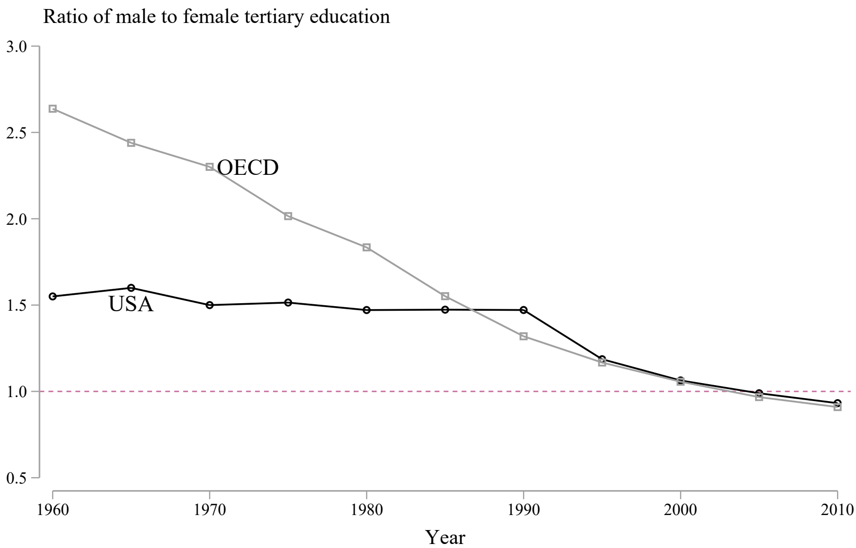
This table shows the coefficients of marriage market exposure in separate regressions where the dependent variable is the one indicated in each row. Sample is restricted to individuals not directly exposed to the reform, and divided into men and women with low level of education (less than university degree) and those with university education. Standard errors (in parentheses) are clustered at the municipality of birth level. The specification includes cohort and municipality of birth F.E., as well as region-specific linear trends. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 4.A.9: Heterogeneous effects of marriage market exposure by level of education for all men

	Married/cohabiting		Number of children	
	Low	High	Low	High
Marriage market exposure (sd)	-0.011*	-0.005	-0.035*	-0.018
	(0.006)	(0.007)	(0.018)	(0.023)
Observations	528571	215340	448030	190539
Adjusted R^2	0.011	0.007	0.017	0.014

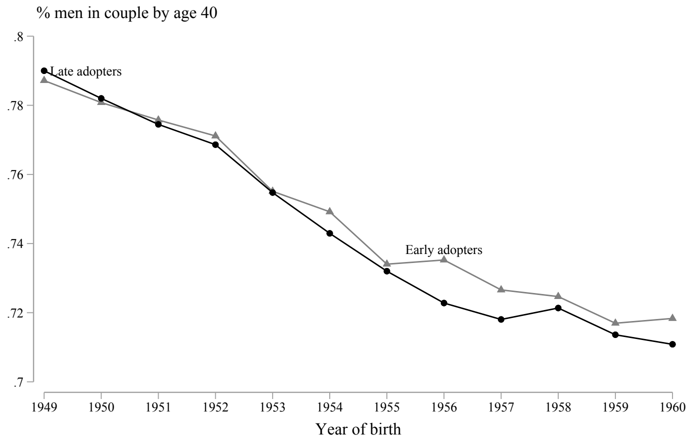
Standard errors (in parentheses) are clustered at the municipality of birth level. This table shows heterogeneous effects of marriage market exposure by level of education for all men (those directly exposed and not directly exposed to the reform). The dependent variable in columns 1-2 is an indicator for being married or cohabiting at age 40, and in columns 3-4 the number of children by age 40. The specification includes cohort and municipality of birth F.E., as well as region-specific linear trends. Marriage market exposure (in standard deviations) indicates the proportion of people in someone's marriage market affected by the reform. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Figure 4.A.1: Ratio of percentage of men to percentage of women (ages 20-64) with tertiary education in the US and on average in the OECD

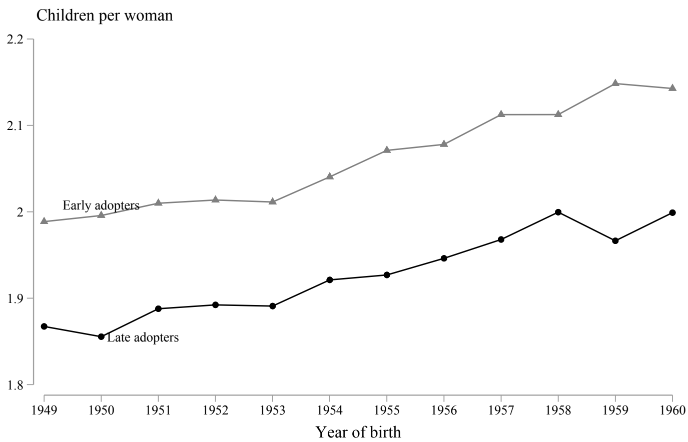


Notes: This figure shows the evolution of the ratio of the percentage of men to the percentage of women with tertiary education among the population aged 20-64 in the US (black line) and on average for OECD countries (gray line). Data from Barro and Lee (2013).

Figure 4.A.2: Trends in family outcomes in pre-reform cohorts – early vs. late reform municipalities



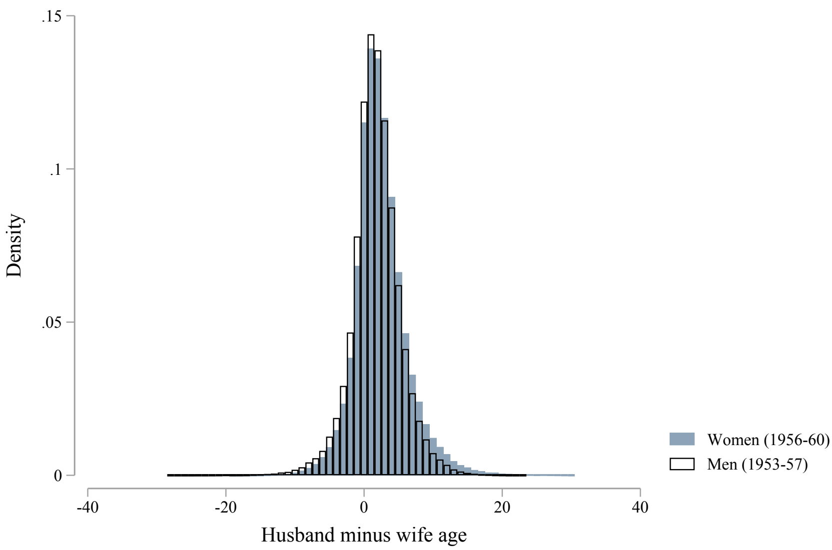
(a) Percentage of men married or cohabiting by age 40



(b) Average number of children per woman

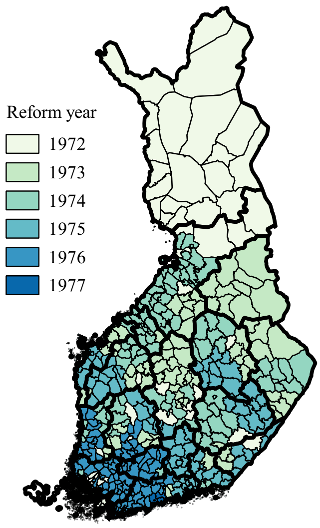
Notes: This figure presents the evolution of trends in fertility and marriage outcomes in early-adopter municipalities (those that implemented the reform in 1972-1974) and in late-adopter municipalities (those that implemented it in 1975-1977). Panel (a) shows the the percentage of men who were married or cohabiting by age 40 by cohort, and panel (b) shows the average number of children per woman by cohort.

Figure 4.A.3: Distribution of age difference between husband and wife in pre-reform cohorts



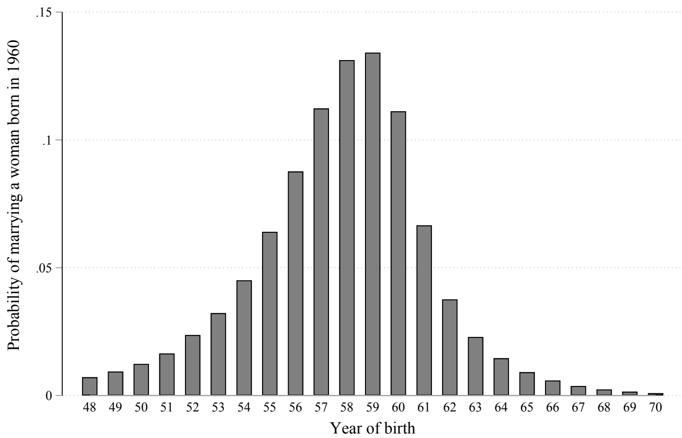
Notes: This figure shows the distribution of the age difference within married couples in pre-reform cohorts (1956-60 for women and 1953-57 for men).

Figure 4.A.4: Variation in year of reform implementation by municipality and region

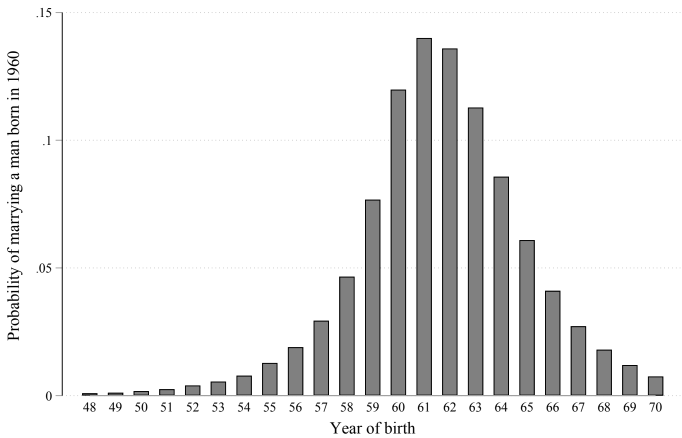


Notes: This map shows the year of adoption of the reform by municipality. Thicker lines indicate region boundaries.

Figure 4.A.5: Example of imputed probability of belonging to the marriage market – 1960 cohort



(a) Probability of belonging to the marriage market of a woman born in 1960



(b) Probability of belonging to the marriage market of a man born in 1960

Notes: This figure represents the weight given to men and women of each cohort for constructing the marriage market of 1960 women in panel (a), and of 1960 men in panel (b). The calculation is based on the distribution of the age difference within couples in pre-reform cohorts (1956-60 for women and 1953-57 for men), which is shown in Figure 4.A.3.

FIGURE 4.A.6: Variation in the proportion of an individual's marriage market exposed to the reform for individuals affected and not affected by the reform themselves

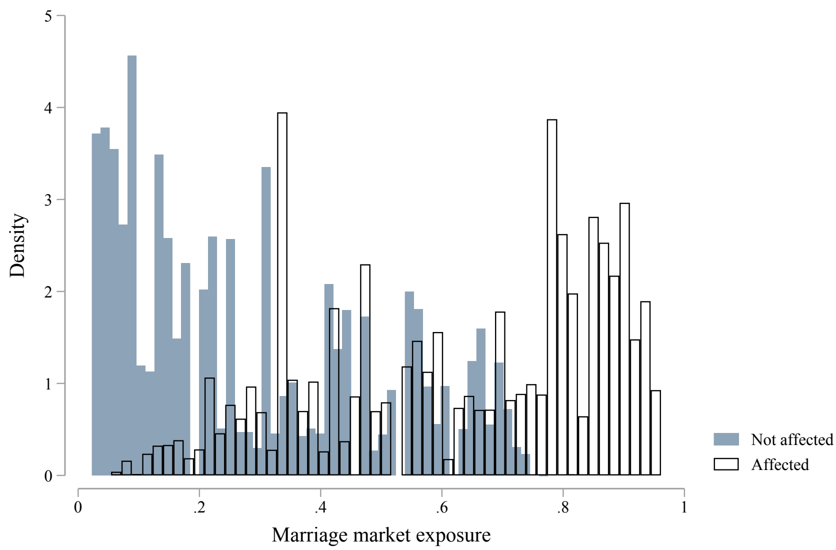
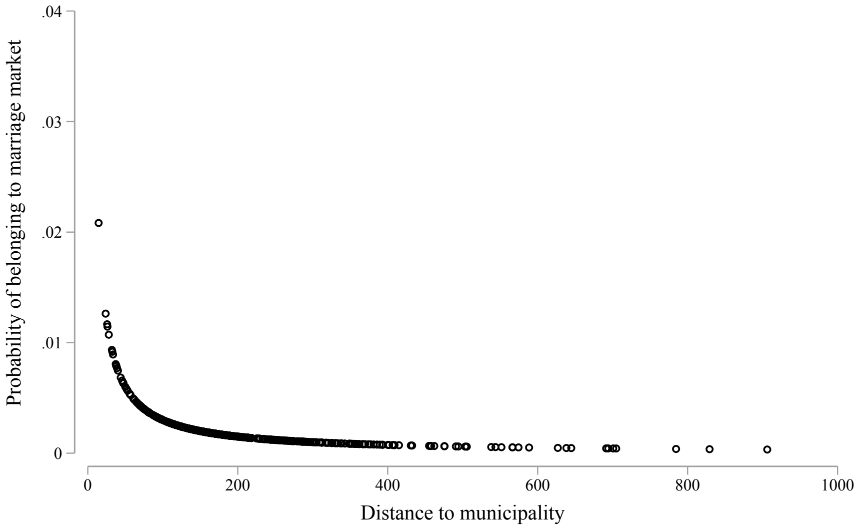
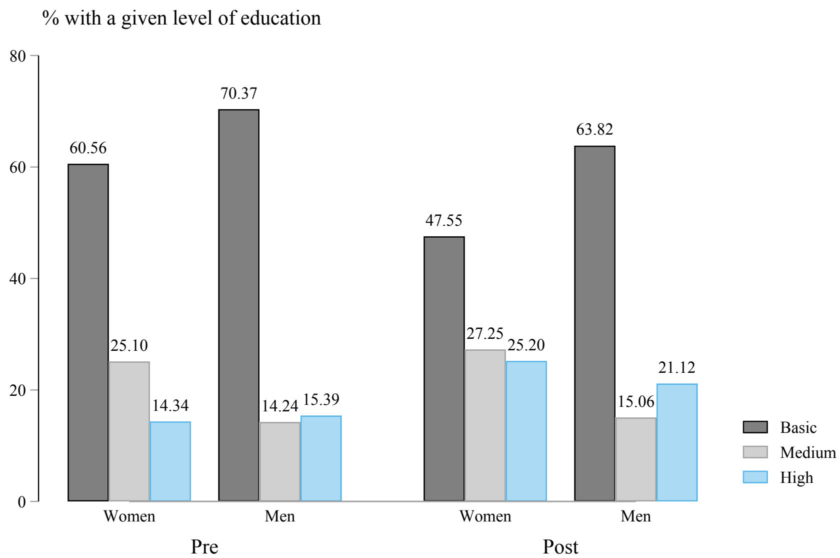


Figure 4.A.7: Probability of belonging to marriage market by distance between municipalities: Tampere (example)



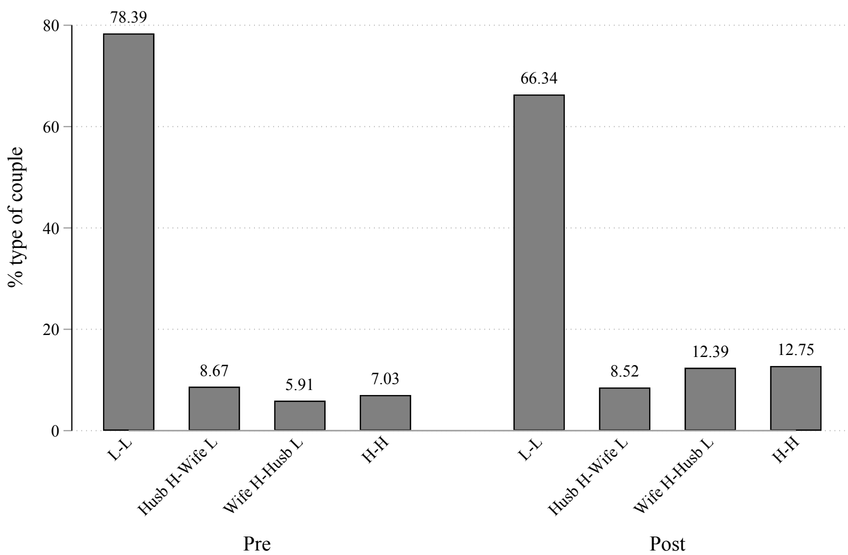
Notes: This figure plots the imputed probability for people in each municipality of belonging to the marriage market of individuals from Tampere (as an example). This probability is based on the inverse of the distance between each municipality and Tampere. Inverse distance probabilities are rescaled such that they add up to 1.

Figure 4.A.8: Distribution of educational attainment by gender and cohorts



Notes: This figure plots the percentage of men and women with basic, medium, and high level of education in pre-reform (1956-60) and post reform (1966-70) cohorts. Basic education is defined as upper secondary education at most; medium education is defined as more than secondary, but less than university education, and high education refers to university degree or higher.

Figure 4.A.9: Frequency of different types of couples by relative education – pre- and post-reform cohorts



Notes: This figure plots the frequency of different types of couples, by relative level of education, in pre-reform (1956-60) and post reform (1966-70) cohorts. Couples are classified into four groups: couples where none of the spouses have a university degree (L-L), those in which both spouses have a university degree (H-H), couples where the husband has a university degree and the wife does not (Husb H-Wife L), and couples where the wife has a university degree and the husband does not (Wife H-Husb L).

5

CHANGES IN INEQUALITY IN MORTALITY: NEW EVIDENCE FOR SPAIN

Joint with Libertad González (UPF and Barcelona GSE)

5.1 INTRODUCTION

Spain is projected to become the country with the highest life expectancy in the world by 2040 (Foreman et al., 2018). Life expectancy has been increasing steadily over the past few decades, fueled by declines in mortality in all age groups. However, averages hide potential variation in health and longevity by socioeconomic status (SES), and it is relevant to understand whether the sustained improvements come from the bottom or the top of the income distribution.

We analyze the evolution of inequality in mortality in Spain during 1990-2014. We focus on age-specific mortality, and consider inequality across narrowly defined geographical areas, ranked by average socioeconomic status. This approach allows us to explore changes in inequality for all age groups, including children, compared to alternative methods using SES indicators at the individual level, such as education or occupation. By considering inequality across areas ranked in terms of their relative socioeconomic level, we also avoid concerns of compositional changes within socioeconomic groups over time.

Our results show substantial decreases in mortality over the past 25 years for all age groups, which were particularly pronounced for men, resulting in a sizeable reduction in the gender gap in mortality. Inequality in mortality across Spanish localities was very low during the whole period, except for the elderly, and it remained essentially unchanged over time, including during the recent recession. During the same period, income inequality was decreasing, albeit with short setbacks, such as after the 2008 crisis (Anghel et al., 2018; Ferrer-i Carbonell et al., 2013).

Compared to other countries where similar analyses have been recently performed, we find that inequality in mortality is lower in Spain than in the US and Canada, and comparable to that in France (Baker et al., 2017; Currie et al., 2018). We find essentially no change in inequality among middle-aged women and the elderly, in contrast to the increase found in the US and Canada (Baker et al., 2017; Currie and Schwandt, 2016a).

Recent research found increases in mortality rates among (older) adults of lower socio-economic status (SES) in the United States (Case and Deaton, 2015, 2017). This increase was driven by drug overdoses, suicides, and alcohol-related liver mortality, and some have suggested it might be related to economic factors (Case and Deaton, 2017; Ruhm, 2018), leading to the label "deaths of despair." These findings raised concerns about increasing inequality in life expectancy (Chetty et al., 2016; Cutler et al., 2011; National Academies of Sciences Engineering and Medicine, 2015) .

Recent work on the US (Currie and Schwandt, 2016a,b) revisits these concerns, highlighting the relevance of studying trends in inequality separately for younger and older ages, as well as accounting for compositional changes within socioeconomic groups over time. Their analysis is based on comparing the evolution of age-specific mortality rates during the period 1990-2010 across groups of counties ranked by poverty rates. This approach reveals that, while for older groups (particularly for women), decreases in mortality have been larger in better-off areas, inequality has in fact decreased substantially among children and youth. Thus, the documented increases in inequality in mortality are concentrated in the group of older women, with inequality actually decreasing at younger ages. These results suggest that inequality in life expectancy may in fact be expected to fall in the long term for the younger cohorts.

Two recent studies comparing trends in the US with other countries reinforce the idea that increases in income inequality need not necessarily lead to increases in inequality in mortality, and stress the role of public policy and public health insurance in mediating this relationship. Parallel analyses of inequality in mortality in Canada (Baker et al. 2017) and France (Currie et al. 2018) find substantial decreases in mortality and low levels of inequality in these countries from 1990 to 2010, particularly in France and among younger cohorts, despite increases in income inequality during this period.

We perform an exercise similar to Currie and Schwandt (2016b) using data for Spain, with municipalities as the geographical unit of interest. Similarly to them, we split municipalities into groups that account for fixed proportions of the total population, so that we can compare mortality rates for the 5% of the population living in richer areas with those for the 5% of the population living in poorer areas in each period.

We contribute to the international literature by providing evidence on changes in inequality in mortality by age and gender for a European country. The only similar analysis for Europe is a recent working paper analyzing French data (Currie et al., 2018), where their level of analysis is the *département* (there are 96 of them in France). Such an aggregate level of analysis, however, may miss some within-unit inequality and thus obscure some of the recent trends. In this paper, we are able to check the robustness of our findings to the level of aggregation by comparing the results by province (52 units) to those by municipality (around 400 units).

Spain offers an interesting case for comparison with the other countries where analyses of this kind have been performed. On the one hand, it resembles France and Canada in having population-wide health insurance coverage which, together with other elements of the welfare system, might mitigate the potential impact of income inequality on inequality in health and mortality. On the other hand, Spain has been particularly affected by the recent economic crisis, and income inequality has increased more than in Canada and France, although inequality levels are still far from those in the US. In this context, it is remarkable that we find steady declines in mortality for all age groups during the period of analysis, including the recent recession years. Our results also show no increase in inequality at older ages, and essentially flat slopes in 2010-2014 for most age groups below 50. We only find a slight increase in inequality among middle-aged men, which seems to be driven by differential patterns of smoking cessation for different income groups. Overall, our results show no indication that inequality in life expectancy may increase in the near future.

Several previous studies have documented the extent of inequality in health and mortality rates in Spain, but these have been mainly limited to specific regions, namely Madrid, Barcelona and the Basque Country, where availability of better data allows for analysis of mortality inequality at the individual level (Borrell et al., 1999, 2008; Martínez et al., 2009; Regidor et al., 2003). These three regions have been the only representation of Spain in international studies comparing socioeconomic inequalities in mortality across European countries. Those studies found that both Spain and Italy (represented by the city of Turin) have lower inequality than their northern neighbors[1] (Huisman et al., 2005;

[1] A recent exception is Bohácek et al. (2018), where they use harmonized panel data from 10 European countries, England, and the US and study inequalities in life expectancy at age 50 by level of education and gender across countries during the period 2002-2015. Their results differ from previous studies that had used data only from some regions in Italy and Spain in that they do not find inequalities to be

Kulhánová et al., 2014; Mackenbach et al., 2008), and that, contrary to what has happened in other European countries, inequalities have not increased in recent decades (Mackenbach et al., 2015).[2]

Due to the lack of appropriate data, there are no papers studying the evolution of inequality in mortality over long periods of time using individual indicators for the whole of Spain. The most comprehensive study is Regidor et al. (2016), where the authors compare trends in mortality for the population aged 10-74 before (2004-2007) and after the crisis (2008-2011) in three socioeconomic groups, defined by indicators of household wealth (household floor space and number of cars). They find larger reductions in mortality after the crisis than in the years preceding it, with larger declines in the lowest SES group.

Evidence of the evolution of inequality across geographical areas is also scarce; again, most analyses of this kind have focused on specific regions or urban areas (Dalmau-Bueno et al., 2010; Marí-Dell'Olmo et al., 2016; Nolasco et al., 2009; Rodríguez-Fonseca et al., 2013; Ruiz-Ramos et al., 2006). The only paper covering the whole country is Regidor et al. (2014b), where they analyze the evolution of inequality in mortality rates at the province level, ranking provinces by income, from 1970 to 2010, a period when income inequality across provinces declined in Spain. They consider both infant mortality and premature mortality, defined as mortality among the population younger than 75. They find decreasing inequality in infant mortality, but increasing inequality in premature mortality, especially among women. A caveat in the interpretation of these results is that they could be confounded by selective migration of healthier individuals from poorer to richer provinces over time.

To the best of our knowledge, ours is the first analysis documenting the evolution of inequality in mortality in Spain for all age groups

smallest in these Southern countries, but in Scandinavian ones

[2]According to Kulhánová et al. (2014), these smaller socioeconomic differences arise from lower inequality in cardiovascular disease among men, and in cancer among women, together with lower inequality in smoking and sedentary lifestyle, due to the less healthy behaviors of higher educated individuals.

that accounts for compositional changes within socioeconomic groups, while also addressing selective migration concerns by always comparing mortality rates across areas of similar population size.

The remainder of the paper is organized as follows. We describe our data sources and the methodology in Section 5.2. Section 5.3 describes our main results for changes in inequality in mortality by sex and age. Then we discuss our findings in relation with the recent literature for other countries and provide some additional analysis on the main causes of death in Section 5.4, and Section 5.5 concludes.

5.2 DATA AND METHODOLOGY

5.2.1 *Data*

We use three main sources of data: death certificate data, the decennial Census, and the local population registry.

In order to construct (five-year) mortality rates, we use death counts by municipality, gender and age group, obtained from the National Statistical Institute (INE) death certificate microdata, from 1990 to 2014. We also use the restricted version of the death certificate data, which contains information on the cause of death. We assign individuals to their municipality of residence. These data only allow one to identify municipalities with at least 10,000 inhabitants; for the smaller ones, only the province is available.

For the denominator of the mortality rates we need population counts by municipality, gender, and age group, for years 1990, 1995, 2000, 2005, and 2010 (the initial year for each five-year interval). Both the Census and the local population registry provide this information, but they cover different periods. The Census is available for 1991, 2001 and 2011, while the population registry data are available annually from 1996 onwards. We use the population registry for all years starting in 1996, and supplement it with Census data for population counts in

1991. When we explore finer age groups for the younger ages, we use the Census data for 2011, since it allows for a more disaggregated analysis.

Finally, in order to rank municipalities by (proxies of) average socioeconomic status, we use the Census data of 1991, 2001 and 2011 to obtain municipality characteristics. We construct high school dropout rates (proportion of the population aged 19 and older that does not have a high school diploma), employment rates (proportion of the population 16 to 65 that is employed), and unemployment rates (proportion of unemployed among the active population) for each year. We approximate these measures for the intermediate years (1995 and 2005) using linear interpolation. Similar to the death certificate data, the Census only identifies municipalities of more than 20,000 inhabitants, while we only know the province of the smaller ones.

We check that our socioeconomic proxies are correlated with per capita income at the municipality level (which we cannot use directly since data on income by municipality is only available for recent years). To that end, we use estimates of median income by municipality constructed by FEDEA from tax records for the year 2006 (see section 5.2.2).[3]

5.2.2 *Methodology*

We first analyze the evolution of age-specific mortality by sex at the national level. Then, in order to study the evolution of inequality in mortality, we follow the methodology of Currie and Schwandt (2016a,b) and construct 5-year mortality rates by sex, age group, and 5-year period, at the municipality level.[4] We then rank municipalities according to our proxy for average socioeconomic level, with lower rank always indicating better outcomes, and group municipalities into bins, each accounting for approximately 5% of the total Spanish population in the base year.

[3] Data available at http://www.fedea.net/renta/renta.html.

[4] Currie and Schwandt (2016a,b) run their analysis at the county level and construct 3-year mortality rates.

This allows us to compare the average mortality rate, for a given sex and age group in a given period, for the 5% of the population living in the richest municipalities with the 5% of the population living in the poorest municipalities, and see how this comparison evolves over time.

For each of the periods 1990-94, 1995-99, 2000-04, 2005-09, and 2010-14, we construct mortality rates for each gender and age group at the national (municipality) level by adding up all deaths for that sex and age group that took place in Spain (or in that municipality) during each period, and dividing over the population of that sex and age group in Spain (or in the municipality) in the starting year of the period.[5] For instance, the 5-year mortality rate for females aged 0-4 in a given municipality in 2010-2014 would be the sum of all deaths for females younger than 5 that took place in that municipality between January 1st, 2010 and December 31st, 2014, divided by the number of girls younger than 5 that lived in that municipality in 2010.

As explained in section 5.2.1, the data only allow us to identify the municipality of residence of the deceased for municipalities of at least 10,000 inhabitants, and to obtain municipality socioeconomic characteristics for municipalities with at least 20,000 inhabitants. In order not to limit our analysis to the larger municipalities, we group, for each province, the values from all the smaller (unidentified) municipalities separately for municipalities with less than 10,000 inhabitants and for municipalities with between 10,001 and 20,000 inhabitants. By doing so, we reduce measurement error problems that normally arise when constructing mortality rates for small areas (Schmertmann and Gonzaga, 2018). As a result of these restrictions, the total number of municipalities that we can observe in our data, including the ones representing averages from smaller towns, ranges from 380 in 1990 to 489 in 2010. This increase in the number of municipalities is due to migration from

[5]Due to limitations in data availability, for the first two periods we have to use as denominator for the mortality rates the population in years 1991 and 1996, respectively.

smaller towns to larger municipalities over time, as shown by Figure 5.A.1, which displays the percentage of the total population in each year living in localities of different size.

We proxy socioeconomic status at the municipality level with the share of high school dropouts, as well as employment and unemployment rates, as described in section 5.2.1. Figure 5.A.2 shows the correlation between median per capita income in each municipality and our three proxies for the year 2006, when income data are available. Median income shows a strong negative correlation with high school dropout rates (-0.77) and unemployment rates (-0.54), and a positive one with local employment rates (0.48), all of them significant at the 99% confidence level.

We combine the information of these three variables in an index that weighs them equally. To construct the index, we first standardize the three variables by substracting the mean of each period and dividing by the standard deviation. We reorder the variables such that increasing values mean worse outcomes. Then we take the average of the three standardized scores. The resulting index is also strongly correlated with median income in 2006 (correlation of -0.67, significant at the 99% confidence level). For our main results, we rank municipalities using this index, but in the Appendix we also report results using each dimension separately.

We first order municipalities from higher to lower socioeconomic rank, and then group municipalities into “bins” using ventiles of the distribution of the socioeconomic variable, such that each bin contains approximately 5% of the total Spanish population. This process is done separately for each year of analysis. As argued by Currie and Schwandt (2016b), the advantage of this procedure over using municipalities directly is that we avoid selection problems that could arise from shrinking or growing municipalities. Given the skewness in the distribution of population size across municipalities, we make some adjustments to achieve similar-sized bins. In particular, we split the two largest municipalities

(Madrid and Barcelona) into five, each with 1/5 of the population of the original municipality and identical values for mortality and the ranking variable. Figure 5.A.3 in the appendix shows the resulting population per bin for the different years. The variation across bins is relatively small and is not systematically related to the socioeconomic rank.

We also replicate our main analysis using data at the province level, in order to compare the results at different levels of granularity. There are 50 provinces in Spain, plus two additional autonomous cities. We thus aggregate mortality rates and socioeconomic status proxies at the province level. We then follow the same procedure as with municipalities, ranking provinces by socioeconomic status and grouping them into bins of approximately 5% of the total Spanish population in each year.

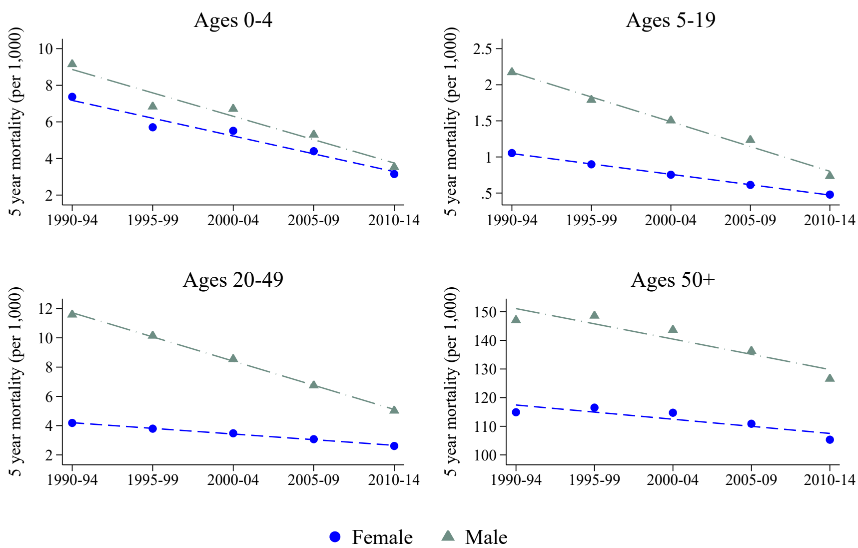
There is large variation in SES across bins in all periods, as shown in Figure 5.A.4 in the appendix. For instance, high school dropout rates range from 40% to 80% in 1990, and from 15% to 45% in 2010. This figure also shows that, while in 2010 there is a strong linear relation between the socioeconomic index and all variables, the relationship, particularly with dropout rates, was less linear in the 1990's, because the correlation between education and employment rates was weaker. This highlights the importance of combining all available information in a single index, given that different indicators might be capturing slightly different dimensions of the socioeconomic level of an area.

5.3 Main Results

5.3.1 *National-level results*

Before exploring changes in inequality using the geographical variation, Figure 5.1 shows 5-year national mortality rates by age group and sex, from 1990-94 to 2010-14. There are three main takeaways. First, during the 25-year period under analysis, there are important declines in mortality in all age groups (illustrated by the negative slopes), including young adults and children. Second, mortality rates are lower for women than men, in all periods and age groups (although the difference is smaller among children 0-4). Third, the declines over time are more pronounced for men. In all but the oldest age group, the mortality rate falls by more than half for men between the early 1990's and the early 2010's. The negative slope is much less pronounced for women (except for children 0-4), so that the gender gap in mortality falls significantly over the period. In the early 1990's, the mortality rate for men 20-49 was about 12 per 1,000, compared with 4 for women. By the 2010's, the corresponding numbers were 5 and 3.

FIGURE 5.1: Evolution of national mortality rates by gender and age group



Notes: This figure plots the evolution of 5-year mortality rates for each gender and age group at the national level from 1990-94 to 2010-14.

The 1990's, as well as the 2010's, were periods of high unemployment and low growth, compared with the 2000's (at least up to 2008). In spite of the changing economic conditions, the declines in mortality seem to follow an approximately linear trend, with no apparent deceleration during the recent recession. Our results are consistent with the findings by several recent studies that mortality and health indicators continued to improve in the first years after the crisis (Regidor et al., 2016, 2014a; Tapia Granados and Ionides, 2017).[6]

Mortality rates are very low for all age groups except the oldest one. National five-year mortality rates are below 10 per 1,000 population for children under 5 over the whole period, and below 2.5 for ages 5 to 19. They are below 12 for both men and women 20 to 49. Mortality rates are an order of magnitude higher for adults over 49, ranging between 100 and 150, so that the behavior of the older group drives the overall evolution of mortality and life expectancy.

5.3.2 *Results by socioeconomic rank*

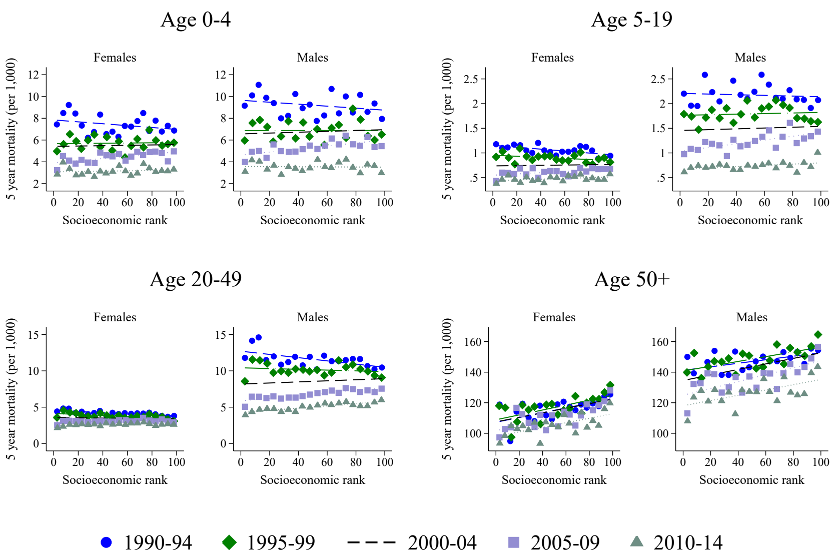
We next present the results for mortality rates by socioeconomic status (SES), measured at the municipality level (Figures 5.2 and 5.3). We rank municipalities by our proxy of per capita income, and group them in 20 bins that include approximately 5% of total population each. Bins are ordered from higher to lower SES. The slopes of the lines measure the degree of inequality in each 5-year period. Positive slopes indicate inequality, in the sense that richer municipalities have lower mortality rates.

The results for the same four age groups as in the previous section are displayed in Figure 5.2 (and Appendix Table 5.A.1). The top dots (and lines) describe the degree of inequality in mortality in 1990-94. The lines are essentially flat for all groups younger than 20, indicating that

[6]There is a large literature exploring the relationship between mortality and the business cycle. Following the seminal work by Ruhm (2000), many studies found mortality to be procyclical. However, analyses of more recent recessions led to different results (Ruhm, 2015). In the case of Spain, papers analyzing both older recessions and the recent one have found mortality to be procyclical (Cervini-Plá and Vall-Castelló, 2018; Tapia Granados, 2005).

mortality rates were similar in poorer and richer areas. The slopes are positive for older ages (50+); i.e., mortality rates were lower in richer areas. Mortality rates were in fact higher in richer areas among men aged 20-49, as illustrated by the negative slope. We explore the sources of this counterintuitive pattern in section 5.4.2.

FIGURE 5.2: Age-specific mortality rates by socioeconomic rank



Notes: This figure shows the evolution of 5-year mortality rates for each gender and age group by socioeconomic level, proxied by an equal-weights index of high school dropout, unemployment and employment rates. Each dot ("bin") represents average values for groups of municipalities accounting for approximately 5% of the total Spanish population in that given year. Bins are ordered from higher to lower SES in each period, so that a positive slope implies lower mortality in richer areas. Different colors of lines and dots are for different groups of 5 years.

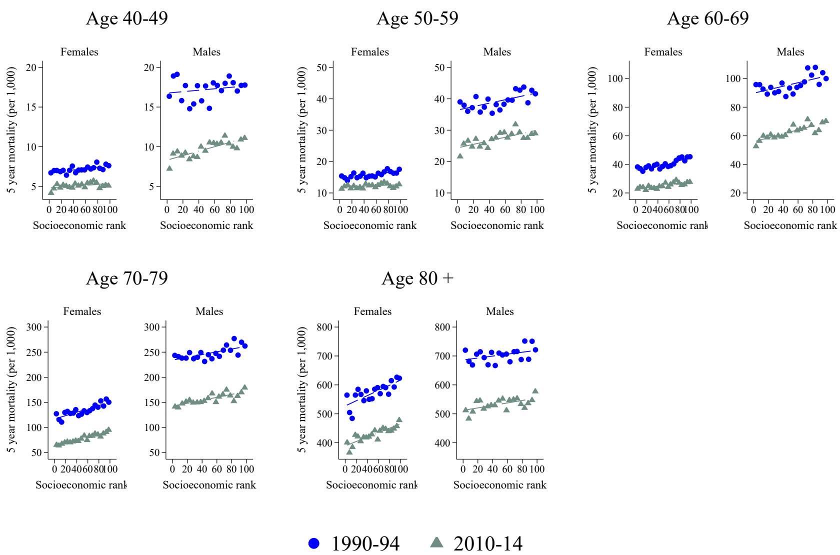
We report mortality rates for the highest and lowest SES bins in 1990-94, as well as the slope (the socio-economic gradient) in Appendix Table 5.A.1. Column 5 shows that mortality rates were essentially the same across SES levels for men and women under 50 (since all coefficients are very small), while a sizeable degree of inequality is observed in the older group, with positive, large, and significant coefficients.

The bottom triangles (and corresponding lines) in Figure 5.2 illustrate the degree of inequality in mortality for the different age groups in 2010-14. The slopes are essentially flat or very small for both men and women in all age groups below 50, indicating that by the early 2010's mortality rates are very similar in poorer and richer areas. The slopes are however significantly positive and large in magnitude for older men and women (see also column 6 in Appendix Table 5.A.1). Inequality in mortality did not decline among the older age group between the early 1990's and the 2010's.

Since the 50+ age group is very broad and there could be changes in age composition across geographical areas, Figure 5.3 disaggregates this group into four finer ones (see also Appendix Table 5.A.2). This figure also includes the group aged 40-49 in order to have a closer look at middle-aged men and women, who have experienced increases in inequality in the US and Canada (Baker et al., 2017; Currie and Schwandt, 2016a). The slopes in 2010-14 are essentially flat for women 40-69, while they remain positive for all groups of men older than 40, and for women 70 and older. We still see that the slopes for both men and women older than 50 are similar in 1990-94 and 2010-14, suggesting no change in inequality over time. However, there has been an increase in inequality for men aged 40-49: while the SES gradient was flat in 1990-94, mortality decreased more in richer than in poorer areas. In section 5.4.2 we explore the causes that have contributed to this phenomenon.

In Figure 5.A.5 in the Appendix we compare the analysis for the four bigger age groups using measures of mortality and socioeconomic status at the province level and at the municipality level. The qualitative conclusions are robust to the level of aggregation, as with either unit we see virtually no inequality among younger groups by 2010-14, and a positive gradient among older men and women.

Figure 5.3: Age-specific mortality rates by socioeconomic rank, finer age groups



Notes: This figure shows the evolution of 5-year mortality rates for each gender and finer age group by socioeconomic level, proxied by an equal-weights index of high school dropout, unemployment and employment rates. Each dot ("bin") represents average values for groups of municipalities accounting for approximately 5% of the total Spanish population in that given year. Bins are ordered from higher to lower SES in each period, so that a positive slope implies lower mortality in richer areas. Blue circles and lines represent values in 1990-94, while green triangles and lines represent values in 2010-14.

The results reported in this section are also robust to using each dimension of the socioeconomic index as alternative ranking variables. Appendix Figure 5.A.6 replicates Figure 5.2 using employment rates as an alternative socioeconomic proxy.[7] The conclusions are similar: we see flat gradients among all young groups by the 2010's and parallel positive trends for the older groups.

Finally, Figure 5.4 shows the gender gap in mortality by socioeconomic rank in 1990-94 and 2010-14. The gender gap declined across all levels of income for all age groups, but the drop was more pronounced in richer than in poorer areas for adults (20-49 and 50+). By 2010-14, we see a flat gradient among those older than 50, with male mortality being 20% higher than female mortality across the whole socioeconomic spectrum.

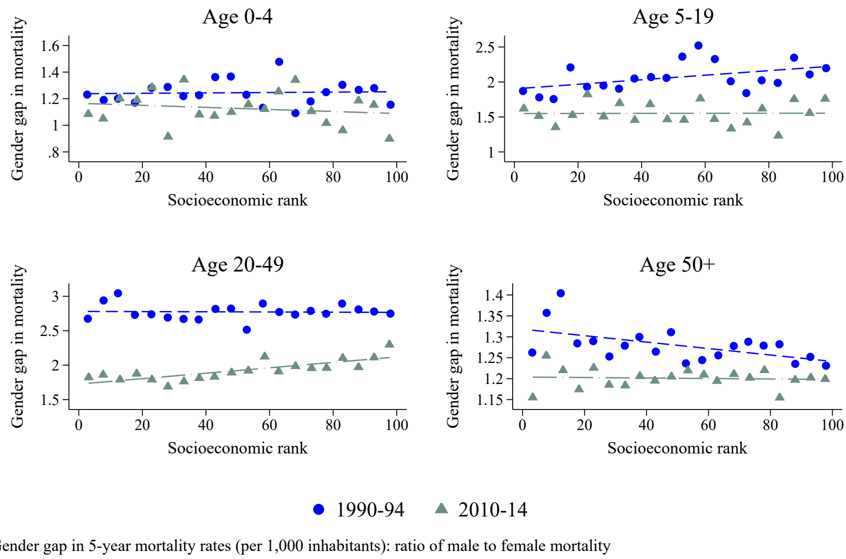
5.4 Discussion: International Comparison and Main Causes of Death

5.4.1 *International comparison*

In this section we provide a brief comparison of our results with recent findings of similar analyses in the US, Canada, and France (Baker et al., 2017; Currie and Schwandt, 2016a,b; Currie et al., 2018). Figures 5.A.7 and 5.A.8 in the Appendix compare macroeconomic conditions in Spain and these other countries during the period of study using OECD data. Regarding income inequality (Figure 5.A.7), we can only compare indicators for all countries for a limited number of years. The available data suggests that Spain has higher income inequality than Canada and France, but lower than the US, and that inequality increased during the crisis and recession years, while it remained relatively stable in Canada

[7] Results using high school dropout rates and unemployment rates as socioeconomic proxies, while not reported here for the sake of brevity, are similar and available upon request.

FIGURE 5.4: Gender gap in mortality rates by socioeconomic rank



Notes: This figure shows the evolution of the gender gap in 5-year mortality rates (ratio of male to female mortality) for each age group by socioeconomic level, proxied by an equal-weights index of high school dropout, unemployment and employment rates. Each dot ("bin") represents average values for groups of municipalities accounting for approximately 5% of the total Spanish population in that given year. Bins are ordered from higher to lower SES in each period, so that a positive slope implies a smaller gender gap in richer areas. Different colors of lines and dots are for different groups of 5 years.

(and declined from 2012 to 2014 in France). Spain was also particularly affected by the crisis in terms of unemployment and GDP growth, as shown in the first two panels of Figure 5.A.8. However, health spending as a percentage of the GDP continued to increase in all four countries in 2010-2014. Finally, the last panel of Figure 5.A.8 shows the stark difference between Canada, France and Spain on the one hand, and the US on the other in terms of health insurance coverage, with the latter being the only one without population-wide health insurance.

Turning to the results of the analysis, compared to the findings for the US by Currie and Schwandt (2016a,b), our results suggest that mortality rates were lower in Spain than in the US in 1990 for all age groups, except for children younger than 5 and for the elderly. In terms of the evolution over time, decreases in mortality were larger in Spain for children, adult women, and the elderly. The results for Spain are similar to those found for France (Currie et al., 2018) in that adult women experienced substantial decreases in mortality in both countries, in contrast to the little improvement seen in the poorer areas of the US and Canada (Baker et al., 2017). For adult men, declines have been smaller in the poorest places in Spain when compared to the poorest places in the US, but higher in the richest areas. As a result, by the 2010's Spain has lower mortality rates than the US, and comparable to those in Canada and France for most age groups.

Declines in inequality were greater in the US for children and for men until age 49. For older men there was not much change in inequality in either country. However, inequality among adult women increased both in the US and Canada, due to the stagnation of mortality in the poorest places. In France, similar to Spain, inequality changed little in all age groups and remained at low levels throughout the period (Currie et al., 2018). As a consequence of these developments, in 2010 Spain presents overall low levels of inequality, comparable to those of France and lower than those in the US and Canada. This finding is consistent with previous literature that, although based only on data from three regions in Spain, found inequality in mortality to be relatively low in Spain compared to other countries (Huisman et al., 2005; Kulhánová et al., 2014; Mackenbach et al., 2008).

5.4.2 *Analysis by cause of death*

In this section we turn to the analysis of mortality by cause of death to further investigate the patterns in overall mortality. We first explore the main drivers of the large declines in mortality documented for all ages at the national level, and then discuss the sources of the socioeconomic inequalities identified for certain groups.

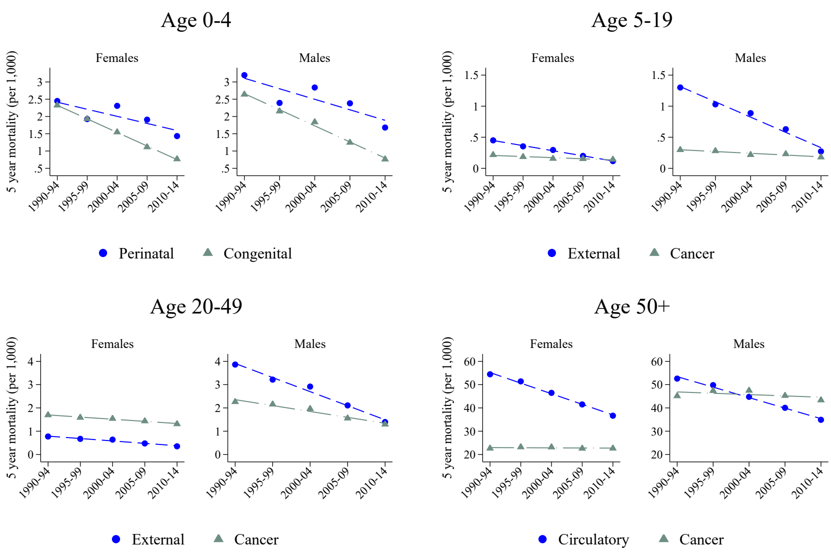
<u>*National level*</u>

Figure 5.5 shows the evolution of mortality rates from the two main causes of death in each age group. Among children younger than 5, complications arising in the perinatal period were and remain the leading cause of death, particularly in the first year of life. Mortality from this cause has decreased notably for both boys and girls during this period, but not as much as mortality from congenital malformations, the second cause of death, which fell to one third of its 1990 level for girls and even more so for boys.

For older children (ages 5-19), external injuries (mostly from car accidents) were the main cause of death in the 1990's, but have decreased sharply over time, particularly for boys, and account for roughly the same number of deaths as cancer by 2010. Mortality from cancer, the second cause of death in this group, remained stable over the period and at low levels, comparable to those in the US and Canada in 2011 (Baker et al., 2017).

Among adults younger than 50, cancer remains the first cause of death among women, and showed only moderate improvement over the last two decades. Among men, on the other hand, external injuries (again, derived chiefly from traffic accidents) accounted for the most deaths in this age group in the 1990's, but have been reduced by more

Figure 5.5: Evolution of main causes of mortality



Notes: This figure shows the evolution of 5-year mortality rates from the two main causes of death for each gender and age group from 1990-94 to 2010-14 at the national level.

than 60%, although they are still the leading cause of death together with cancer.

Finally, for older ages, there were large decreases in mortality from diseases of the circulatory system, both for men and women. Mortality from cancer has changed little for either sex, but is twice as large for males than for females, becoming the leading cause of death among men older than 50 by the 2010's.[8]

Our analysis highlights three main drivers of the large overall decreases in mortality: perinatal and congenital-related diseases among

[8]If we look at finer age groups, cancer is the first cause of death for all men older than 50, except for ages older than 80, where circulatory-related diseases are more important.

young children, traffic accidents among children and young adults, and circulatory diseases among the elderly. Declines in mortality from perinatal and congenital-related conditions are likely to be the result of advances in screening, delivery attendance, and perinatal care (Alonso et al., 2006; Zeitlin et al., 2016). These declines have been larger in Spain than those observed in US or Canada, where mortality from congenital anomalies also decreased substantially, but where deaths from perinatal factors actually increased (Baker et al., 2017).

Traffic accidents and their associated fatalities were higher in Spain than in other Western European countries in the early 1990's, particularly among youth, but have decreased strongly since then due to improvements in safety technologies and regulation (Redondo Calderón et al., 2000; Villalbí and Pérez, 2006). As a result, in the most recent period of analysis mortality from external injuries is lower in Spain than in the US and Canada, both for children and for younger adults (Baker et al., 2017).

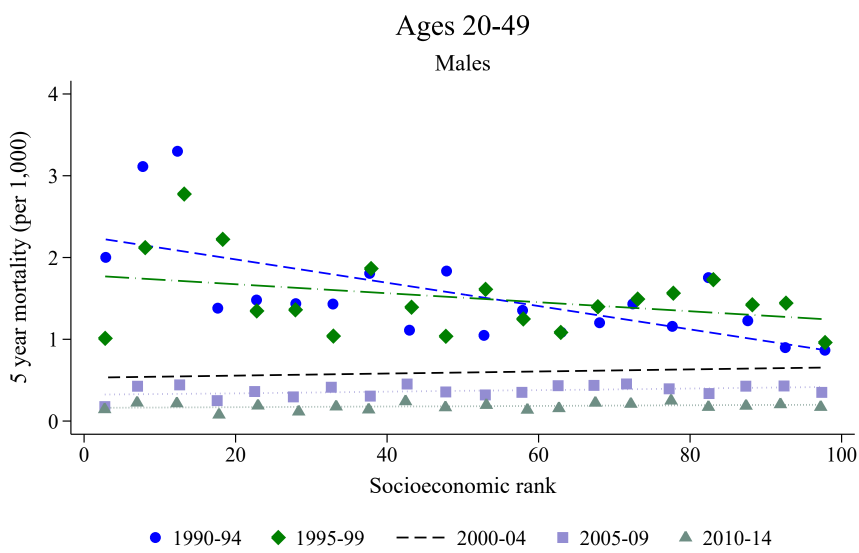
Lastly, mortality from circulatory conditions (cardio and cerebrovascular disease) decreased at a faster pace in Spain compared to the OECD average (OECD, 2015), partly due to better initial treatment, and partly due to decreases in risk factors like high cholesterol or high blood pressure (Flores-Mateo et al., 2011). By the 2010's, mortality from these causes was lower in Spain than that in the US, and similar to that of Canada for women, but lower for men (Baker et al., 2017).

Results by SES rank

We next turn to the discussion of the causes of death behind some of the patterns observed in the analysis of inequality in mortality.

The results in Figure 5.2 revealed that mortality was higher in richer areas in the 1990's among men aged 20 to 49. This surprising pattern turns out to be driven by AIDS mortality. Although AIDS accounted for fewer deaths than traffic accidents or cancer, it presented a marked negative gradient, as shown in Figure 5.6. In 1990, the 5-year mortality rate in some of the richest groups of municipalities, as proxied by our index, was higher than 3 deaths per 1,000 inhabitants, while it was less than 1 in the poorest groups. After 1995, mortality from AIDS fell vastly in richer areas, and by 2000 it presented a completely flat gradient and low levels.

Figure 5.6: Evolution of AIDS mortality for males aged 20-49



Notes: This figure shows the evolution of 5-year mortality rates from AIDS and HIV by socioeconomic level, proxied by an equal-weights index of high school dropout, unemployment and employment rates, for males aged 20 to 49. Each dot ("bin") represents average values for groups of municipalities accounting for approximately 5% of the total Spanish population in that given year. Bins are ordered from higher to lower SES in each period, so that a positive slope implies lower mortality in richer areas. Different colors of dots and lines are for different groups of 5 years.

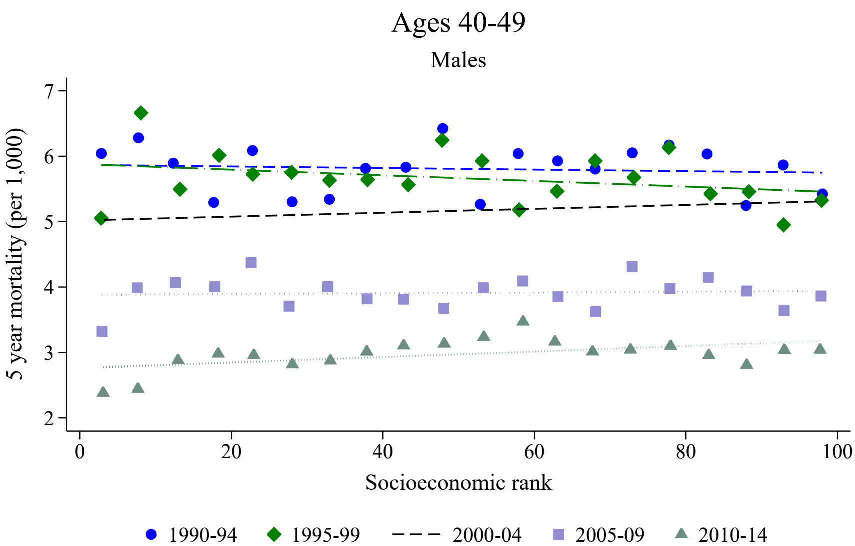
HIV infection spread rapidly in Spain during the 1980s, mostly among intravenous drug users, and linked to the "heroin boom" of this decade (Valdes, 2013). Although this epidemic affected both rural and urban areas, it was particularly problematic in poor neighborhoods within overall rich cities, explaining the observed negative gradient (Gamella, 1997). AIDS mortality reached its peak in Spain in 1994, with higher levels than in other European countries, and then fell from 1995 on thanks to combination therapy and prevention and awareness campaigns (Valdes, 2013).

Our analysis by socioeconomic rank also showed that inequality increased for men in their 40's because mortality fell more in richer areas (see Figure 5.3). This finding bears some resemblance to the documented increases in inequality among middle-aged men and women in the US and Canada (Baker et al., 2017; Currie and Schwandt, 2016a). However, we find that this emerging inequality is due to larger decreases in AIDS mortality in richer areas, as just discussed, and also in cancer mortality, while it does not seem to be related to "deaths of despair", which received much recent attention in the US (Case and Deaton, 2015; Ruhm, 2018).

Figure 5.7 shows the evolution of mortality rates from cancer for men aged 40 to 49 by socioeconomic level. While cancer presented a flat gradient in the 1990's, improvements over this period were larger in richer areas. This widening inequality was mostly driven by lung cancer, the most common type of cancer for this group, and resembles the pattern observed in smoking cessation among men during the 1990's, with higher educated men being more likely to quit smoking than lower educated ones (Fernandez et al., 2001).

Figure 5.8, in turn, examines the evolution of inequality in "deaths of despair" in this age group: we look at mortality from suicides, drug or alcohol poisoning, and alcoholic liver diseases and cirrhosis. Suicides and poisonings have both increased over this period, but without much change in inequality. Suicides present a marked positive gradient, with mortality rates in low SES areas being around twice those in high SES ones. Poisonings do not differ much along the socioeconomic spectrum. On the other hand, mortality from alcoholic liver diseases and cirrhosis has decreased in the last two decades, without much change in its

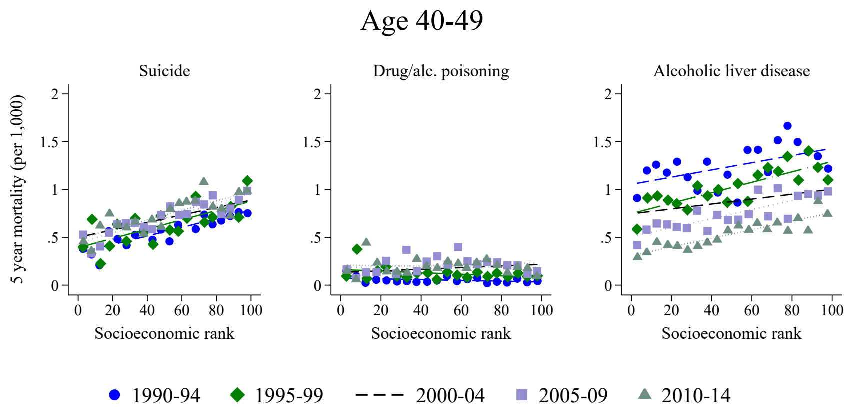
Figure 5.7: Evolution of cancer mortality for males aged 40-49



Notes: This figure shows the evolution of 5-year mortality rates from cancer by socioeconomic level, proxied by an equal-weights index of high school dropout, unemployment and employment rates, for males aged 40 to 49. Each dot ("bin") represents average values for groups of municipalities accounting for approximately 5% of the total Spanish population in that given year. Bins are ordered from higher to lower SES in each period, so that a positive slope implies lower mortality in richer areas. Different colors of dots and lines are for different groups of 5 years.

positive gradient. Overall, the increase in mortality from suicides and poisonings among middle-aged men resembles the trend observed in the US and Canada (Baker et al., 2017). However, the magnitude of the increase has been smaller in Spain, so that none of these "despair" causes affect significantly the general patterns in mortality.

Figure 5.8: Evolution of mortality from "deaths of despair" for males aged 40-49



Notes: This figure shows the evolution of 5-year mortality rates from three types of "deaths of despair" by socioeconomic level, proxied by an equal-weights index of high school dropout, unemployment and employment rates, for males aged 40 to 49. The left panel represents mortality from suicides; the center panel shows mortality from alcohol or drug poisoning, and the right one shows mortality from alcoholic liver diseases and cirrhosis. Each dot ("bin") represents average values for groups of municipalities accounting for approximately 5% of the total Spanish population in that given year. Bins are ordered from higher to lower average SES in each period, so that a positive slope implies lower mortality in richer areas. Different colors of dots and lines are for different groups of 5 years

5.5 CONCLUSIONS

We analyze the evolution of inequality in age-specific mortality across Spanish municipalities, ranked by socioeconomic status, from 1990 to 2014. We document substantial decreases in mortality over the past 25 years for all age groups, which were particularly marked for males, resulting in a sizeable reduction in the gender gap in mortality. We find that the degree of inequality across locations was low in the 1990's, and remained so during the whole period, for most age groups (except the elderly). Our results are consistent with the results from Coveney et al. (2016) showing that health inequality by income did not increase in Spain after the 2008 crisis (in fact, it decreased). Compared to the US and Canada, inequality is lower in Spain, comparable to that in France. We find essentially no change in inequality among middle-aged women and the elderly, in contrast to the increase found in the US and Canada.

Compared to previous literature on inequality in mortality in Spain, which used mainly socioeconomic measures at the individual level, such as education or occupation, the analysis of inequality across small geographical areas allows us to explore changes in inequality for the whole of Spain separately for different age groups, including children. This approach provides relevant insights: like Currie and Schwandt (2016a,b) in the US, we see lower inequality among younger cohorts, which may be anticipating lower inequality in life expectancy in the future.

The low levels of inequality observed across municipalities by 2010 do not imply that inequality does not exist between individuals within municipalities. For instance, a recent study for the region of Catalonia using individual data found substantial socioeconomic inequalities in health care utilization even among young children in 2015 (Observatori del Sistema de Salut de Catalunya, 2017).

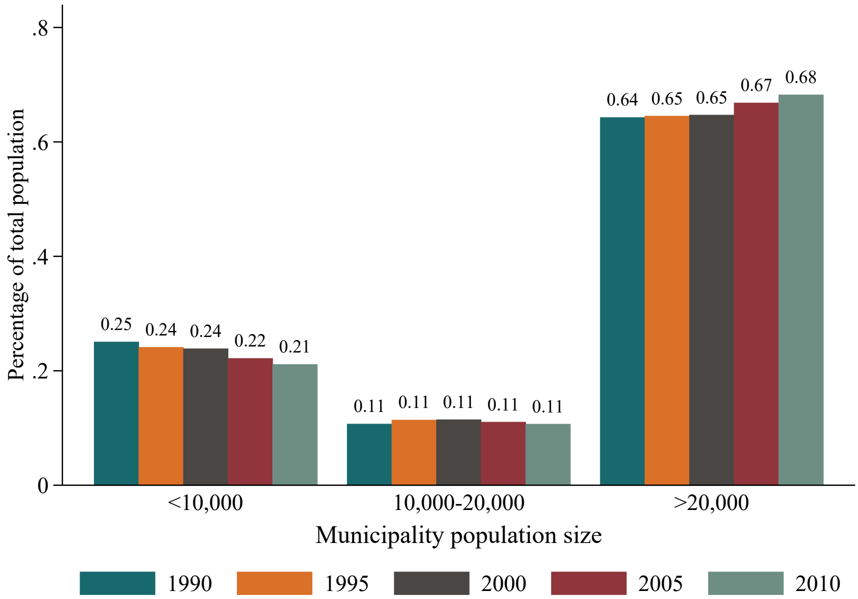
In conclusion, we show that the decreases in mortality experienced during the last twenty-five years in Spain were accompanied by little change in inequality, in spite of the increase in income inequality that followed the recent economic crisis. These findings support the idea that increases in income inequality do not necessarily translate, at least in the short run, into more inequality in mortality, in the context of a country with public health insurance and a European welfare system.

5

APPENDIX

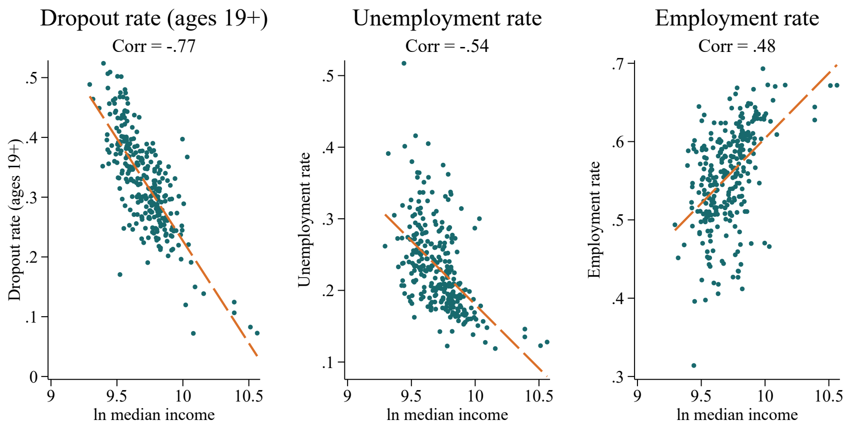
Appendix 5.A

Figure 5.A.1: Distribution of total population by size of municipality of residence



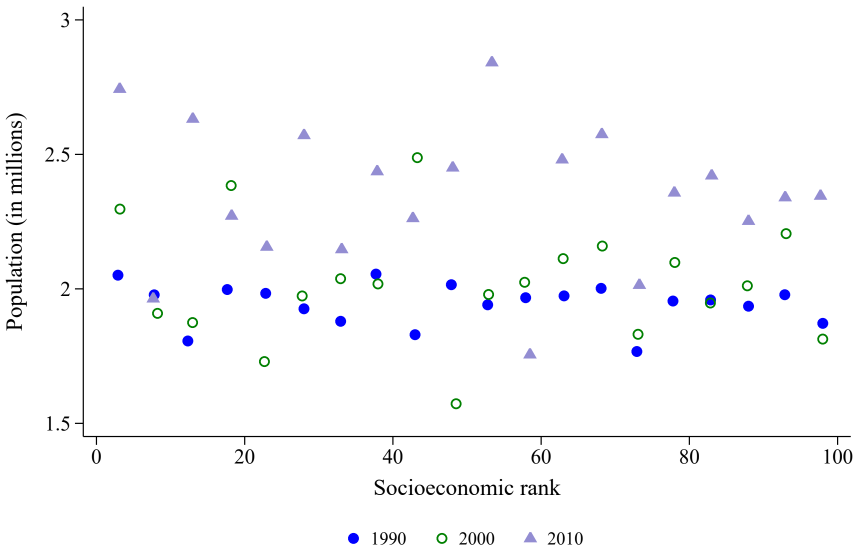
Notes: This figure shows the proportion of the total Spanish population in each year living in municipalities of less than 10,000 inhabitants, in those of more than 10,000 and less than 20,000 inhabitants, and in those of more than 20,000 inhabitants

Figure 5.A.2: Relationship between median income and socioeconomic indicators in 2006



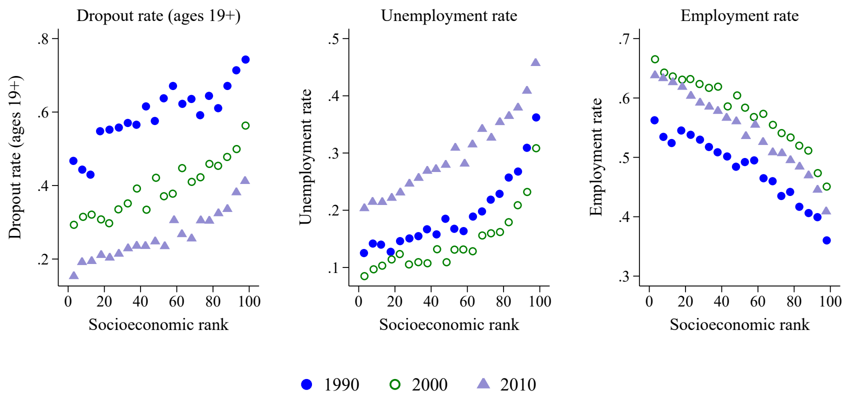
Notes: This figure shows the relationship of high school dropout rates (left), unemployment rates (center) and employment rates (right) with median income (in natural logarithms) per tax payer by municipality in 2006. Source: FEDEA for income data, and 2001 and 2011 Census for socioeconomic status proxies (2006 values are linear interpolations of these two years).

Figure 5.A.3: Population by bin



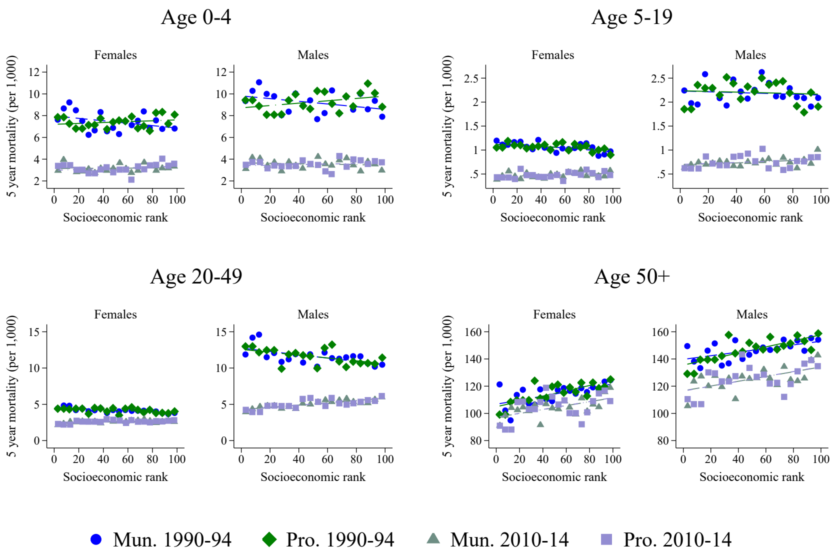
Notes: This figure shows the population included in each bin (in millions), with bins ordered by from higher to lower average socioeconomic status, proxied by an equal weights index of high school dropout, unemployment and employment rates. Each dot ("bin") represents values for groups of municipalities accounting for approximately 5% of the total Spanish population in that given year. Different colors of dots are for different years.

FIGURE 5.A.4: Bin characteristics



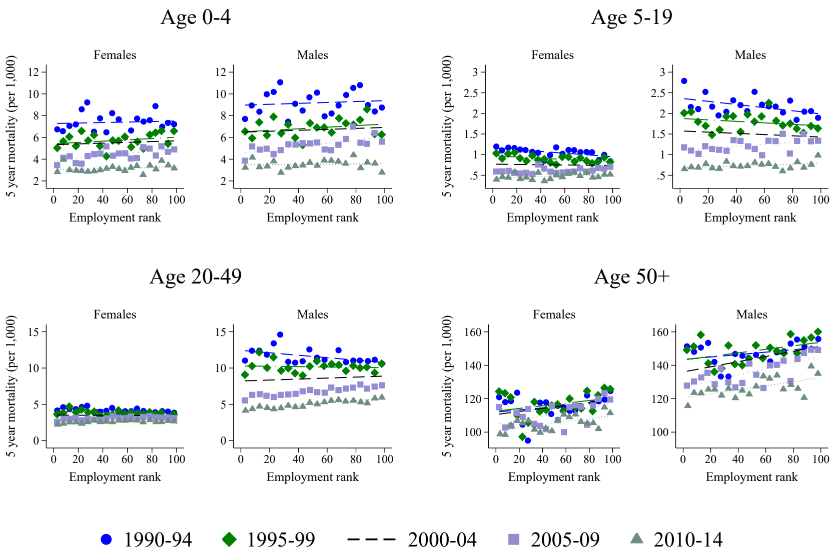
Notes: This figure shows average high school dropout rates (left), unemployment rates (center), and employment rates (left) by bin, with bins ordered by average socioeconomic status, proxied by an equal-weights index of high school dropout, unemployment and employment rates. Each dot ("bin") represents values for groups of municipalities accounting for approximately 5% of the total Spanish population in that given year. Blue colors are for 1990, green hollow circles for 2000, and lilac triangles for 2010.

Figure 5.A.5: Age-specific mortality rates by socioeconomic rank at the province and municipality level



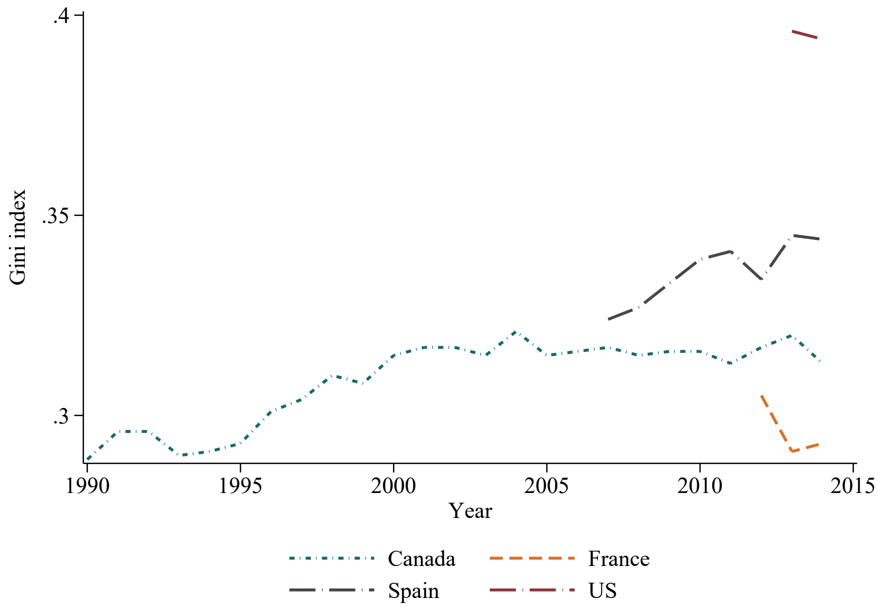
Notes: This figure shows the evolution of 5-year mortality rates for each gender and age group by socioeconomic level, separately for "bins' constructed from province-level data (rhombus and squares) and from municipality-level data (circles and triangles). Each dot ("bin") represents values for groups of provinces or municipalities accounting for approximately 5% of the total Spanish population in that given year. Bins are ordered from higher to lower average socioeconomic status in each period, so that a positive slope implies lower mortality in richer areas. Blue and green dots represent values in 1990-94, while light green and mauve ones represent values in 2010-2014.

FIGURE 5.A.6: Age-specific mortality rates by employment rate rank



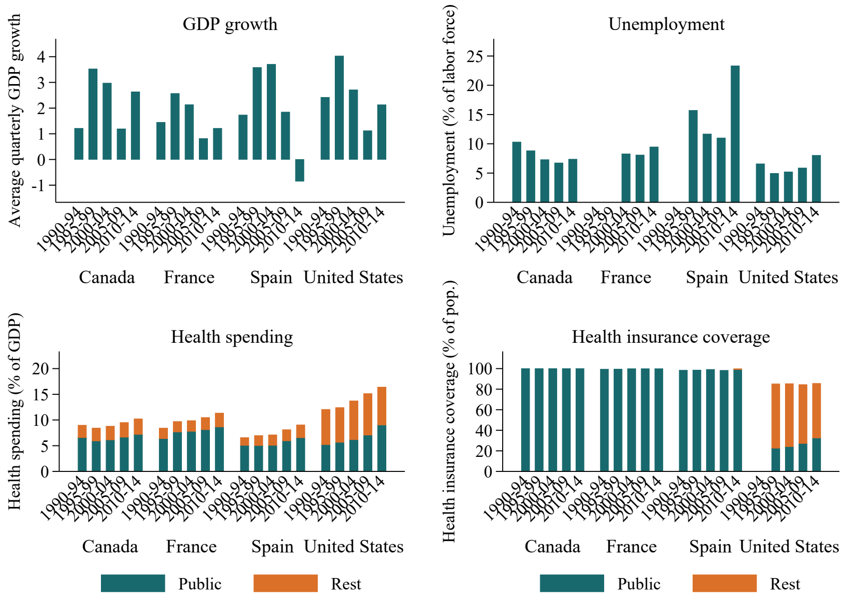
Notes: This figure shows the evolution of 5-year mortality rates for each gender and finer age group by socioeconomic level, proxied by employment rates. Each dot ("bin") represents average values for groups of municipalities accounting for approximately 5% of the total Spanish population in that given year. Bins are ordered from high to low employment rates in each period, so that a positive slope implies lower mortality in areas with more employment. Different colors of dots and lines are for different groups of 5 years.

Figure 5.A.7: Income inequality for Spain, Canada, France, and the US over time



Notes: This figure shows the evolution of the Gini index in Canada, France, Spain, and the US from 1990 to 2014. Only for Canada is data available for the whole period. Data for Spain is available from 2007 to 2014; for France, from 2012 to 2014, and for the US from 2013 to 2014. The Gini index takes values from 0 to 1, with higher values indicating higher inequality. Data from OECD (2018), "Income inequality" (indicator), https://doi.org/10.1787/459aa7f1-en.

FIGURE 5.A.8: Macroeconomic and social spending indicators for Spain, Canada, France, and the US over time



Notes: This figure shows averages of quarterly GDP growth (percent change with respect to the same quarter of the previous year); the unemployment rate as a percentage of the labor force; health spending as percentage of GDP, and health insurance coverage as percentage of the total population, for each 5-year period of analysis for Canada, France, Spain, and the US. In the case of unemployment, data for France is only available from 2003 to 2014, while data for Spain is available from 1999 to 2014. For health spending, the height of the bar refers to the total, while the blue part indicates the fraction of government spending and compulsory health insurance. Similarly, for health insurance the height of the bar refers to total public and private primary health insurance coverage, while the blue part represents the fraction covered by public insurance. Data from OECD (2018): Quarterly GDP (indicator). doi: 10.1787/b86d1fc8-en; Unemployment rate (indicator). doi: 10.1787/997c8750-en; Health spending (indicator). doi: 10.1787/8643de7e-en, and "Social protection," OECD Health Statistics (database), https://doi.org/10.1787/data-00544-en.

Table 5.A.1: Age-specific mortality in municipalities of highest and lowest SES

	Lowest SES		Highest SES		Slope of regression line		
	1990-94	2010-14	1990-14	2010-14	1990-94	2010-14	p-value diff
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Males							
0-4	9.155	3.099	7.939	2.978	-0.009	-0.001	0.001
	(2.718)	(1.044)	(1.712)	(0.861)			
5-19	2.202	0.613	2.071	1.008	-0.001	0.001*	0.221
	(0.740)	(0.274)	(0.471=	(0.346)			
20-49	11.791	3.922	10.464	5.995	-0.022***	0.016***	0.000
	(2.445)	(0.913)	(2.044)	(0.654)			
50 +	150.014	107.981	154.394	143.591	0.116***	0.172***	0.433
	(9.566)	(19.951)	(11.513)	(13.226)			
Females							
0-4	7.437	2.856	6.875	3.312	-0.009	0.002	0.097
	(2.222)	(1.478)	(2.280)	(0.867)			
5-19	1.177	0.378	0.942	0.572	-0.002***	0.001*	0.000
	(0.385)	(0.235)	(0.367)	(0.276)			
20-49	4.409	2.150	3.807	2.609	-0.008***	0.003**	0.000
	(0.912)	(0.482)	(0.697)	(0.342)			
50 +	118.828	93.471	125.416	119.735	0.156***	0.149***	0.922
	(14.680)	(13.340)	(9.524)	(12.581)			

Columns (1)-(4) report the means (and standard deviation in parentheses) of 5-year mortality rates for each gender and age group in 1990 and 2010, in the bin of municipalities with lowest and highest average socioeconomic status, respectively. Columns (5) and (6) report the coefficient of the fitted regression line in each year, and column (7) reports the p-value for the null hypothesis that the slopes are equal in both years.

Table 5.A.2: Age-specific mortality in municipalities of highest and lowest SES – older groups

	Lowest SES		Highest SES		Slope of regression line		
	1990	2010	1990	2010	1990	2010	p-value diff
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Males							
40-49	16.360	7.214	17.772	11.137	0.006	0.032***	0.034
	(3.661)	(2.161)	(2.576)	(1.513)			
50-59	39.007	21.638	41.604	29.018	0.054***	0.065***	0.625
	(4.784)	(4.474)	(6.556)	(2.437)			
60-69	95.840	52.661	100.016	70.277	0.112**	0.158***	0.366
	(8.426)	(6.907)	(13.517)	(5.933)			
70-79	243.553	141.540	262.282	179.838	0.275***	0.333***	0.579
	(26.963)	(13.583)	(29.240)	(16.609)			
80+	719.971	512.294	721.291	578.237	0.317	0.527***	0.460
	(95.060)	(43.257)	(60.342)	(45.533)			
Females							
40-49	6.711	4.166	7.606	5.086	0.009***	0.007**	0.889
	(1.199)	(1.267)	(1.488)	(0.894)			
50-59	15.407	11.315	17.530	12.764	0.025***	0.010**	0.392
	(2.106)	(2.052)	(2.847)	(2.190)			
60-69	38.206	22.904	45.345	27.856	0.094***	0.054***	0.683
	(4.935)	(2.772)	(7.667)	(3.473)			
70-79	127.364	65.403	150.407	95.377	0.363***	0.316***	0.983
	(20.700)	(6.717)	(16.996)	(11.157)			
80+	564.620	400.828	622.945	478.241	0.992***	0.865***	0.176
	(88.538)	(41.581)	(40.942)	(31.121)			

Columns (1)-(4) report the means (and standard deviation in parentheses) of 5-year mortality rates for each gender and age group in 1990 and 2010, in the bin of municipalities with lowest and highest average socioeconomic status, respectively. Columns (5) and (6) report the coefficient of the fitted regression line in each year, and column (7) reports the p-value for the null hypothesis that the slopes are equal in both years.

BIBLIOGRAPHY

Aakvik, A., Salvanes, K. G., and Vaage, K. (2010). Measuring heterogeneity in the returns to education using an education reform. *European Economic Review*, 54(4):483–500.

Abramitzky, R., Delavande, A., and Vasconcelos, L. (2011). Marrying Up: the Role of Sex Ratio in Marriage by Class. *American Economic Journal: Applied Economics*, 3(July):124–157.

Adecco Healthcare (2017). Empleo en el sector sanitario. Adecco Group. https://adecco.es/wp-content/uploads/2017/07/NdP-Las-ofertas-de-empleo-en-el-sector-sanitario-se-incrementarán-un-40-según-Adecco-Healthcare.pdf. Accessed: 26-12-2017.

AEPED (2014). Test de apgar. Asociación Española de Pediatría (Spanish Pediatrics Association). http://www.aeped.es/rss/en-familia/test-apgar. Accessed: 20-12-2017.

Aizer, A., Stroud, L., and Buka, S. (2016). Maternal Stress and Child Outcomes: Evidence from Siblings. *Journal of Human Resources*, 51(3):523–555.

Almond, D., Chay, K. Y., and Lee, D. S. (2005). The costs of low birth weight. *The Quarterly Journal of Economics*, 120(3):1031–1083.

Almond, D. and Currie, J. (2011). Killing me softly: The fetal origins hypothesis. *Journal of Economic Perspectives*, 25(3):153–72.

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

Almond, D., Edlund, L., and Palme, M. (2009). Chernobyl's subclinical legacy: Prenatal exposure to radioactive fallout and school outcomes in Sweden. *The Quarterly Journal of Economics*, 124(4):1729–1772.

Almqvist, C., Cnattingius, S., Lichtenstein, P., and Lundholm, C. (2012). The impact of birth mode of delivery on childhood asthma and allergic diseases–a sibling study. *Clinical & Experimental Allergy*, 42(9):1369–1376.

Alonso, V., Fuster, V., and Luna, F. (2006). Causes of neonatal mortality in Spain (1975-98): Influence of sex, rural-urban residence and age at death. *Journal of Biosocial Science*, 38(4):537–551.

American College of Obstetricians and Gynecologists (2006). Mode of Term Singleton Breech Delivery. ACOG Commitee Opinion No. 340. *Obstetrics and Gynecology*, 108:235–237.

American College of Obstetricians and Gynecologists (2010). ACOG Practice bulletin no. 115: Vaginal birth after previous cesarean delivery. *Obstetrics and Gynecology*, 116(2):450–63.

American College of Obstetricians and Gynecologists (2015). Practice Bulletin No. 154: Operative Vaginal Delivery. *Obstetrics and Gynecology*, 126(5):e56–e65.

Anderberg, D. and Zhu, Y. (2014). What a difference a term makes: the effect of educational attainment on marital outcomes in the UK. *Journal of Population Economics*, 27:387–419.

Anghel, B., Basso, H. S., Bover, O., Casado, J. M., Hospido, L., Izquierdo, M., Kataryniuk, I., Lacuesta, A., Montero, J. M., and Vozmediano, E. (2018). Income, Consumption and Wealth Inequality in Spain. *SERIEs*, 9:351–387.

Angrist, J. (2002). How do sex ratios affect marriage and labor markets? Evidence from America's second generation. *Quarterly Journal of Economics*, 117(3):997–1038.

Angrist, J. and Pischke, J.-S. (2009). *Mostly Harmless Econometrics: An Empiricist's Companion.* Princeton University Press.

APA (2017). Cesarean procedure. American Pregnancy Association. http://americanpregnancy.org/labor-and-birth/cesarean-procedure/. Accessed: 11-12-2017.

Arrow, K. J. (1963). Uncertainty and the Welfare Economics of Medical Care. *American Economic Review*, 53(5):941–973.

Asher, I. and Pearce, N. (2014). Global burden of asthma among children. *The International Journal of Tuberculosis and Lung Disease*, 18(11):1269–1278.

Autor, D., Dorn, D., and Hanson, G. (2019). When work disappears: Manufacturing decline and the falling marriage market value of young men. *American Economic Review: Insights*, 1(2):161–78.

Bager, P., Wohlfahrt, J., and Westergaard, T. (2008). Caesarean delivery and risk of atopy and allergic disesase: Meta-analyses. *Clinical & Experimental Allergy*, 38(4):634–642.

Baicker, K., Buckles, K. S., and Chandra, A. (2006). Geographic variation in the appropriate use of cesarean delivery. *Health Affairs*, 25(5):w355–w367.

Baker, M., Currie, J., and Schwandt, H. (2017). Mortality Inequality in Canada and the U.S.: Divergent or Convergent Trends? *NBER Working Paper*, 23514.

Barber, E. L., Lundsberg, L. S., Belanger, K., Pettker, C. M., Funai, E. F., and Illuzzi, J. L. (2011). Indications Contributing to the Increasing Cesarean Delivery Rate. *Obstetrics and Gynecology*, 118(1):29–38.

Barro, R. J. and Lee, J. W. (2013). A new data set of educational attainment in the world, 1950–2010. *Journal of Development Economics*, 104:184–198.

Becker, G. S. (1973). A Theory of Marriage: Part I. *Journal of Political Economy*, 81(4):813–846.

Bentley, J. P., Roberts, C. L., Bowen, J. R., Martin, A. J., Morris, J. M., and Nassar, N. (2016). Planned Birth Before 39 Weeks and Child Development: A Population-Based Study. *Pediatrics*, 138(6):e20162002–e20162002.

Bertrand, M., Kamenica, E., and Pan, J. (2015). Gender Identity and Relative Income within Households. *The Quarterly Journal of Economics*, pages 571–614.

Bhattacharya, J., Goldman, D., and McCaffrey, D. (2006). Estimating probit models with self-selected treatments. *Statistics in Medicine*, 25(3):389–413.

Black, S. E., Breining, S., Figlio, D. N., Guryan, J., Karbownik, K., Nielsen, H. S., Roth, J., and Simonsen, M. (2017). Sibling spillovers. Working Paper 23062, National Bureau of Economic Research.

Black, S. E., Devereux, P. J., and Salvanes, K. G. (2008). Staying in the classroom and out of the maternity ward? The effect of compulsory schooling laws on teenage births. *Economic Journal*, 118(2004):1025–1054.

Blustein, J. and Liu, J. (2015). Time to consider the risks of caesarean delivery for long term child health. *BMJ*, 350.

Boerma, T., Ronsmans, C., Melesse, D. Y., Barros, A. J., Barros, F. C., Juan, L., Moller, A. B., Say, L., Hosseinpoor, A. R., Yi, M., de Lyra Rabello Neto, D., and Temmerman, M. (2018). Global epidemiology of use of and disparities in caesarean sections. *The Lancet*, 392(10155):1341–1348.

Bohácek, R., Bueren, J., Crespo, L., Mira, P., and Pijoan-Mas, J. (2018). Inequality in Life Expectancies across Europe. *CEMFI Working Paper*, 1810.

Boksa, P. and Zhang, Y. (2008). Epinephrine administration at birth prevents long-term changes in dopaminergic parameters caused by cesarean section birth in the rat. *Psychopharmacology*, 200(3):381–391.

Borrell, C., Azlor, E., Rodríguez-Sanz, M., Puigpinós, R., Cano-Serral, G., Pasarín, M. I., Martínez, J. M., Benach, J., and Muntaner, C. (2008). Trends in socioeconomic mortality inequalities in a southern European urban setting at the turn of the 21st century. *Journal of Epidemiology and Community Health*, 62(3):258–266.

Borrell, C., Regidor, E., Arias, L. C., Navarro, P., Puigpinós, R., Domínguez, V., and Plasència, A. (1999). Inequalities in mortality according to educational level in two large Southern European cities. *International Journal of Epidemiology*, 28(1):58–63.

Brainerd, E. (2017). The lasting effect of sex ratio imbalance on marriage and family: evidence from world war II in russia. *The Review of Economics and Statistics*, 99(May):229–242.

Brown, H. (1996). Physician demand for leisure: Implications for cesarean section rates. *Journal of Health Economics*, 15(2):233 – 242.

Card, D., Fenizia, A., and Silver, D. (2019). The impacts of hospital delivery practices on infant health. *NBER Working Paper, 25986*.

Cardwell, C. R., Stene, L. C., Joner, G., Cinek, O., Svensson, J., Goldacre, M. J., Parslow, R. C., Pozzilli, P., Brigis, G., Stoyanov, D., Urbonaité, B., Šipetić, S., Schober, E., Ionescu-Tirgoviste, C., Devoti, G., de Beaufort, C. E., Buschard, K., and Patterson, C. C. (2008). Caesarean section is associated with an increased risk of childhood-onset type 1 diabetes mellitus: a meta-analysis of observational studies. *Diabetologia*, 51(5):726–735.

Case, A. and Deaton, A. (2015). Rising morbidity and mortality in midlife among white non-Hispanic Americans in the 21st century. *Proceedings of the National Academy of Sciences*, 112(49):15078–15083.

Case, A. and Deaton, A. (2017). Mortality and Morbidity in the 21st Century. *Brookings Papers on Economic Activity*, pages 397–476.

Cervini-Plá, M. and Vall-Castelló, J. (2018). Mortality and business cycle in Spain.

Charles, K. K. and Luoh, M. C. (2010). Male Incarceration and and Female Marriage Outcomes. *The Review of Economics and Statistics*, 92(3):614–627.

Chetty, R., Stepner, M., Abraham, S., Lin, S., Scuderi, B., Turner, N., Bergeron, A., and Cutler, D. (2016). The association between income and life expectancy in the United States, 2001-2014. *JAMA - Journal of the American Medical Association*, 315(16):1750–1766.

Chiburis, R. C., Das, J., and Lokshin, M. (2012). A practical comparison of the bivariate probit and linear iv estimators. *Economics Letters*, 117(3):762 – 766.

Coile, C. C. and Duggan, M. G. (2019). When Labor's Lost: Health, Family Life, Incarceration, and Education in a Time of Declining Economic Opportunity for Low-Skilled Men. *Journal of Economic Perspectives*, 33(2):191–210.

Collado, M. C., Rautava, S., Isolauri, E., and Salminen, S. (2015). Gut microbiota: a source of novel tools to reduce the risk of human disease? *Pediatric Research*, 77:182 – 188. Review.

Costa-Ramón, A. M., Rodríguez-González, A., Serra-Burriel, M., and Campillo-Artero, C. (2018). It's about time: Cesarean sections and neonatal health. *Journal of Health Economics*, 59:46 – 59.

Coveney, M., García-Gómez, P., Van Doorslaer, E., and Van Ourti, T. (2016). Health disparities by income in Spain before and after the economic crisis. *Health Economics*, 25(Suppl.:141–158.

Currie, J. and MacLeod, W. B. (2008). First Do No Harm? Tort Reform and Birth Outcomes. *Quarterly Journal of Economics*, 123(2):795–830.

Currie, J. and Macleod, W. B. (2017). Diagnosing Expertise: Human Capital, Decision Making, and Performance among Physicians. *Journal of Labor Economics*, 35(1):1–43.

Currie, J. and Schwandt, H. (2016a). Inequality in mortality decreased among the young while increasing for older adults, 1990 – 2010. *Science*, 352(6286).

Currie, J. and Schwandt, H. (2016b). Mortality Inequality: The Good News from a County-Level Approach. *Journal of Economic Perspectives*, 30(2):29–52.

Currie, J., Schwandt, H., and Thuilliez, J. (2018). Pauvreté, Egalité, Mortalité: Mortality (In)Equality in France and the United States. *NBER Working Paper*, 24623.

Cutler, D. M., Lange, F., Meara, E., Richards-Shubik, S., and Ruhm, C. J. (2011). Rising educational gradients in mortality: The role of behavioral risk factors. *Journal of Health Economics*, 30(6):1174–1187.

Cygan-Rehm, K. and Maeder, M. (2013). The effect of education on fertility: Evidence from a compulsory schooling reform. *Labour Economics*, 25:35–48.

Dalmau-Bueno, A., García-Altés, A., Marí-Dell'Olmo, M., Pérez, K., Kunst, A. E., and Borrell, C. (2010). Veintidós años de evolución de las desigualdades socioeconómicas en la mortalidad en la ciudad de Barcelona. *Gaceta Sanitaria*, 24(1):20–27.

Davidson, R., Roberts, S. E., Wotton, C. J., and Goldacre, M. J. (2010). Influence of maternal and perinatal factors on subsequent hospitalisation for asthma in children: evidence from the Oxford record linkage study. *BMC Pulmonary Medicine*, 10(1):14.

Dominguez-Bello, M. G., Jesus-laboy, K. M. D., Shen, N., Cox, L. M., Amir, A., Gonzalez, A., Bokulich, N. A., Song, S. J., Rivera-vina, J. I., Mendez, K., Knight, R., and Clemente, J. C. (2016). Partial restoration of the microbiota of cesarean-born infants via vaginal microbial transfer. *Nature Medicine*, 22(3):250–253.

Drucker, A. M., Wang, A. R., Li, W. Q., Sevetson, E., Block, J. K., and Qureshi, A. A. (2017). The Burden of Atopic Dermatitis: Summary of a Report for the National Eczema Association. *Journal of Investigative Dermatology*, 137(1):26–30.

Esteve, A., García-Román, J., and Permanyer, I. (2012). The gender-gap reversal in education and its effect on union formation: The end of hypergamy? *Population and Development Review*, 38(3):535–546.

Esteve, A., Schwartz, C. R., van Bavel, J., Permanyer, I., Klesment, M., and García-Román, J. (2016). The End of Hypergamy: Global Trends and Implications. *Population and Development Review*, 42(4):615–625.

Evans, M. I., Richardson, D. A., Sholl, J. S., and Johnson, B. A. (1984). Cesarean section. Assessment of the convenience factor. *J Reprod Med*, 29(9):670–676.

Fernandez, E., Schiaffino, A., García, M., and Borràs, J. M. (2001). Widening social inequalities in smoking cessation in Spain, 1987-1997. *Journal of epidemiology and community health*, 55(10):729–30.

Ferrer-i Carbonell, A., Ramos, X., and Oviedo, M. (2013). Growing Inequalities and Their Impacts in Spain. Technical report.

Figlio, D., Guryan, J., Karbownik, K., and Roth, J. (2014). The effects of poor neonatal health on children's cognitive development. *American Economic Review*, 104(12):3921–55.

Finkelstein, E. A., Graham, W. C. K., and Malhotra, R. (2014). Lifetime Direct Medical Costs of Childhood Obesity. *Pediatrics*, 133(5):854–862.

Flores-Mateo, G., Grau, M., O'Flaherty, M., Ramos, R., Elosua, R., Violan-Fors, C., Quesada, M., Martí, R., Sala, J., Marrugat, J., and Capewell, S. (2011). Analyzing the Coronary Heart Disease Mortality Decline in a Mediterranean Population: Spain 1988-2005. *Revista Española de Cardiología (English Edition)*, 64(11):988–996.

Folke, O. and Rickne, J. (2020). All the Single Ladies: Job Promotions and the Durability of Marriage. *American Economic Journal: Applied Economics*, 12(1):260–87.

Foreman, K. J., Marquez, N., Dolgert, A., Fukutaki, K., Fullman, N., McGaughey, M., Pletcher, M. A., Smith, A. E., Tang, K., Yuan, C. W., Brown, J. C., Friedman, J., He, J., Heuton, K. R., Holmberg, M., Patel, D. J., Reidy, P., Carter, A., Cercy, K., Chapin, A., Douwes-Schultz, D., Frank, T., Goettsch, F., Liu, P. Y., Nandakumar, V., Reitsma, M. B., Reuter, V., Sadat, N., Sorensen, R. J., Srinivasan, V., Updike,

R. L., York, H., Lopez, A. D., Lozano, R., Lim, S. S., Mokdad, A. H., Vollset, S. E., and Murray, C. J. (2018). Forecasting life expectancy, years of life lost, and all-cause and cause-specific mortality for 250 causes of death: reference and alternative scenarios for 2016–40 for 195 countries and territories. *The Lancet*, 392(10159):2052–2090.

Fort, M., Schneeweis, N., and Winter-Ebmer, R. (2016). Is Education Always Reducing Fertility? Evidence from Compulsory Schooling Reforms. *The Economic Journal*, 126:1823–1855.

Fraser, W., Usher, R. H., McLean, F. H., Bossenberry, C., Thomson, M. E., Kramer, M. S., Smith, L., and Power, H. (1987). Temporal variation in rates of cesarean section for dystocia: Does "convenience" play a role? *American Journal of Obstetrics and Gynecology*, 156(2):300 – 304.

Gamella, J. F. (1997). Heroína en España, 1977-1996. Balance de una crisis de drogas. *Claves de Razón Práctica*, 72:20–30.

Global Asthma Network (2018). *The Global Asthma Report 2018*. Global Asthma Network, Auckland, New Zealand.

Goldin, C. (2006). The Quiet Revolution That Transformed Women's Employment, Education, and Family. *AEA Papers and Proceedings*, 96(2):1–21.

Goldscheider, F., Bernhardt, E., and Lappegård, T. (2015). The Gender Revolution: A Framework for Understanding Changing Family and Demographic Behavior. *Population and Development Review*, 41(2):207–239.

Goodman-Bacon, A. (2018). Difference-in-Differences with Variation in Treatment Timing. *NBER Working Paper*, 25018.

Goodman-Bacon, A., Goldring, T., and Nichols, A. (2019). Bacondecomp: Stata module to perform a bacon decomposition of difference-in-differences estimation. *Statistical Software Components.*

Grant, D. (2009). Physician financial incentives and cesarean delivery: New conclusions from the healthcare cost and utilization project. *Journal of Health Economics*, 28(1):244 – 250.

Greitemeyer, T. (2007). What do men and women want in a partner? Are educated partners always more desirable? *Journal of Experimental Social Psychology*, 43(2):180–194.

Grivell, R. M. and Dodd, J. M. (2011). Short- and long-term outcomes after cesarean section. *Expert Review of Obstetrics and Gynecology*, 6(2):205–215.

Grosjean, P. and Khattar, R. (2019). It's Raining Men! Hallelujah? The Long-Run Consequences of Male-Biased Sex Ratios. *Review of Economic Studies*, 86:723–754.

Gruber, J. and Owings, M. (1996). Physician Financial Incentives and Cesarean Section Delivery. *The RAND Journal of Economics*, 27(1):99.

Halla, M., Mayr, H., Pruckner, G. J., and Garcia-Gomez, P. (2016). Cutting fertility? The effect of cesarean deliveries on subsequent fertility and maternal labor supply. IZA Discussion Paper.

Hannah, M. (1989). Birth asphyxia: does the Apgar score have diagnostic value? *Obstetrics and Gynecology*, 73(2):299–300.

Hannah, M. E., Hannah, W. J., Hewson, S. A., Hodnett, E. D., Saigal, S., Willan, A. R., and Collaborative, T. B. T. (2000). Planned caesarean section versus planned vaginal birth for breech presentation at term: a randomised multicentre trial. *The Lancet*, 356(9239):1375–1383.

Hansen, A. K., Wisborg, K., Uldbjerg, N., and Henriksen, T. B. (2008). Risk of respiratory morbidity in term infants delivered by elective caesarean section: cohort study. *BMJ: British Medical Journal*, 336:85–87.

Harjutsalo, V. (2008). Time trends in the incidence of type 1 diabetes in Finnish children: a cohort study. *Lancet (London, England)*, 371(9626):1777–82.

Heath, R. and Jayachandran, S. (2017). The Causes and Consequences of Increased Female Education and Labor Force Participation in Developing Countries. In Averett, S. L., Argys, L. M., and Hoffman, S. D., editors, *The Oxford Handbook of Women and the Economy*, pages 345–367. Oxford University Press.

Hitsch, G. J., Hortaçsu, A., and Ariely, D. (2010). Matching and sorting in online dating. *American Economic Review*, 100(1):130–163.

Hueston, W. J., McClaflin, R. R., and Claire, E. (1996). Variations in cesarean delivery for fetal distress. *The Journal of Family Practice*, 43(5):461–467.

Huisman, M., Kunst, A. E., Bopp, M., Borgan, J.-K., Borrell, C., Costa, G., Deboosere, P., Gadeyne, S., Glickman, M., Marinacci, C., Minder, C., Regidor, E., Valkonen, T., and Mackenbach, J. P. (2005). Educational inequalities in cause-specific mortality in middle-aged and older men and women in eight western European populations. *The Lancet*, 365(9458):493–500.

Hyde, M. J. and Modi, N. (2012). The long-term effects of birth by caesarean section: The case for a randomised controlled trial. *Early Human Development*, 88(12):943–949.

Hyde, M. J., Mostyn, A., Modi, N., and Kemp, P. R. (2012). The health implications of birth by caesarean section. *Biological Reviews*, 87(1):229–243.

Jachetta, C. (2015). Cesarean sections and later health outcomes. Unpublished, Available at http://www.christinejachetta.com.

Jensen, V. M. and Wüst, M. (2015). Can caesarean section improve child and maternal health? The case of breech babies. *Journal of Health Economics*, 39:289 – 302.

Johnson, E. M. and Rehavi, M. M. (2016). Physicians treating physicians: Information and incentives in childbirth. *American Economic Journal: Economic Policy*, 8(1):115–41.

Keag, O. E., Norman, J. E., and Stock, S. J. (2018). Long-term risks and benefits associated with cesarean delivery for mother, baby, and subsequent pregnancies: Systematic review and meta-analysis. *PLOS Medicine*, 15(1):1–22.

Kearney, M. S. and Wilson, R. (2018). Male earnings, marriageable men, and nonmarital fertility: evidence from the fracking boom. *The Review of Economics and Statistics*, 100(October):678–690.

Kerr, S. P. and Pekkarinen, T. (2013). School Tracking and Development of Cognitive Skills. *Journal of Labor Economics*, 31(3):577–602.

Klasko, S. K., Cummings, R. V., Balducci, J., DeFulvio, J. D., and Reed, J. F. (1995). The impact of mandated in-hospital coverage on primary cesarean delivery rates in a large nonuniversity teaching hospital. *American Journal of Obstetrics and Gynecology*, 172(2):637 – 642.

Klesment, M. and Van Bavel, J. (2017). The Reversal of the Gender Gap in Education, Motherhood, and Women as Main Earners in Europe. *European Sociological Review*, 33(3):465–481.

Knip, M. and Simell, O. (2012). Enviornmental Triggers of Type 1 Diabetes. *Cold Spring Harbor perspectives in medicine*, 2(7):a007690.

Koerselman, K. and Pekkarinen, T. (2018). Cognitive consequences of the timing of puberty. *Labour Economics*, 54(September 2017):1–13.

Kuhle, S., Kirk, S. F. L., Ohinmaa, A., and Veugelers, P. J. (2011). Comparison of ICD code-based diagnosis of obesity with measured obesity in children and the implications for health care cost estimates. *BMC Medical Research Methodology*, 11:173.

Kulhánová, I., Bacigalupe, A., Eikemo, T. A., Borrell, C., Regidor, E., Esnaola, S., Mackenbach, J. P., and Consortium, E. (2014). Why does Spain have smaller inequalities in mortality ? An exploration of potential explanations. *European Journal of Public Health*, 24(3):370–377.

Lafortune, J. (2013). Making Yourself Attractive: Pre-Marital Investments and the Returns to Education in the Marriage Market. *American Economic Journal: Applied Economics*, 5(2):151–178.

Lefgren, L., Mcintyre, F., Lefgren, L., and Mcintyre, F. (2006). The Relationship between Women's Education and Marriage Outcomes. *Journal of Labor Economics*, 24(4):787–830.

Lefèvre, M. (2014). Physician induced demand for c-sections: does the convenience incentive matter? Health, Econometrics and Data Group (HEDG) Working papers, HEDG, c/o Department of Economics, University of York.

Leung, D. Y., Boguniewicz, M., Howell, M. D., Nomura, I., and Hamid, Q. A. (2004). New insights into atopic dermatitis. *The Journal of Clinical Investigation*, 113:651–657.

Lobstein, T., Baur, L., and Uauy, R. (2004). Obesity in children and young people: a crisis in public health. *Obesity Reviews*, 5(s1):4–85.

Mackenbach, J. P., Kulhánová, I., Menvielle, G., Bopp, M., Borrell, C., Costa, G., Deboosere, P., Esnaola, S., Kalediene, R., Kovacs, K., Leinsalu, M., Martikainen, P., Regidor, E., Rodriguez-Sanz, M., Strand, B. H., Hoffmann, R., Eikemo, T. A., Östergren, O., and Lundberg, O. (2015). Trends in inequalities in premature mortality: a study of 3.2 million deaths in 13 European countries. *Journal of epidemiology and community health*, 69(3):207–217.

Mackenbach, J. P., Stirbu, I., Roskam, A.-j. R., Schaap, M. M., Menvielle, G., Leinsalu, M., and Kunst, A. E. (2008). Socioeconomic Inequalities in Health in 22 European Countries. *New England Journal of Medicine*, 359(12):1290–1291.

Malin, G. L., Morris, R. K., and Khan, K. S. (2010). Strength of association between umbilical cord pH and perinatal and long term outcomes: systematic review and meta-analysis. *BMJ: British Medical Journal*, 340:c1471.

Marcos, J. C. M. (2008). Cesarea a Demanda. Sociedad Española de Ginecología y Obstetricia.

Marí-Dell'Olmo, M., Gotsens, M., Palència, L., Rodríguez-Sanz, M., Martinez-Beneito, M. A., Ballesta, M., Calvo, M., Cirera, L., Daponte, A., Domínguez-Berjón, F., Gandarillas, A., Goñi, N. I., Martos, C., Moreno-Iribas, C., Nolasco, A., Salmerón, D., Taracido, M., and Borrell, C. (2016). Trends in socioeconomic inequalities in mortality in small areas of 33 Spanish cities. *BMC Public Health*, 16(1):1–13.

Martínez, C., Regidor, E., Sánchez, E., Pascual, C., and De La Fuente, L. (2009). Heterogeneity by age in educational inequalities in cause-specific mortality in women in the Region of Madrid. *Journal of Epidemiology and Community Health*, 63(10):832–838.

McCloskey, L., Petitti, D. B., and Hobel, C. J. (1992). Variations in the use of cesarean delivery for dystocia: lessons about the source of care. *Med Care*, 30(2):126–135.

McCrary, J. and Royer, H. (2011). The Effect of Female Education on Fertility and Infant Health: Evidence from School Entry Policies Using Exact Date of Birth. *American Economic Review*, 101(February):158–195.

Mechoulan, S. (2011). The External Effects of Black Male Incarceration on Black Females. *Journal of Labor Economics*, 29(1):1–35.

Meghir, C. and Palme, M. (2005). Educational Reform, Ability, and Family Background. *American Economic Review*, 95(1):414–424.

Meghir, C., Palme, M. M. M., and Simeonova, E. (2018). Education and Mortality: Evidence from a Social Experiment. *American Economic Journal: Applied Economics*, 10(2):234–256.

Ministerio de Sanidad, Servicios Sociales e Igualdad (2014). Costes hospitalarios - contabilidad analítica. Spanish Ministry of Health, Social Services and Equality. http://www.msssi.gob.es/estadEstudios/estadisticas/inforRecopilaciones/anaDesarrolloGDR.htm. Accessed: 13-01-2017.

Ministerio de Sanidad, Servicios Sociales e Igualdad (2015). Informe anual del sistema nacional de salud. Spanish Ministry of Health, Social Services and Equality. http://www.msssi.gob.es/estadEstudios/estadisticas/sisInfSanSNS/tablasEstadisticas/InfAnSNS.htm. Accessed: 13-01-2017.

Ministerio de Sanidad y Política Social (2009). Informe maternidad hospitalaria: Estándares y recomendaciones. Spanish Ministry of Health and Social Policy. http://www.msc.es/organizacion/sns/planCalidadSNS/docs/AHP.pdf. Accessed: 14-12-2017.

Ministerio de Trabajo y Asuntos Sociales (1997). BOE-A-1997-26857. https://www.boe.es/diario_boe/txt.php?id=BOE-A-1997-26857.

Mishanina, E., Rogozinska, E., Thatthi, T., Uddin-Khan, R., Khan, K. S., and Meads, C. (2014). Use of labour induction and risk of cesarean delivery: a systematic review and meta-analysis. *Canadian Medical Association Journal*, 186(9):665–673.

Monstad, K., Propper, C., and Salvanes, K. G. (2008). Education and Fertility: Evidence from a Natural Experiment. *Scandinavian Journal of Economics*, 110(September 2007):827–852.

Mühlrad, H. (2017). Cesarean section for high-risk births: short and long term consequences for breech births. https://www.dropbox.com/s/xto1qlsf1qtxf7d/C-section_Muhlrad.pdf. Accessed: 26-12-2017.

National Academies of Sciences Engineering and Medicine (2015). The Growing Gap in Life Expectancy by Income.

Negrusa, B. and Oreffice, S. (2010). Quality of available mates, education, and household labor supply. *Economic Inquiry*, 48(3):558–574.

Neu, J. and Rushing, J. (2011). Cesarean versus vaginal delivery: Long-term infant outcomes and the hygiene hypothesis. *Clinics in Perinatology*, 38(2):321 – 331. Delivery After Previous Cesarean.

NICE (2008). Inducing labour. NICE guideline CG70, National Institute for Health and Care Excellence. https://www.nice.org.uk/guidance/cg70. Accessed: 08-06-2017.

NICE (2011). Cesarean section. NICE clinical guideline CG190, National Institute for Health and Care Excellence. https://www.nice.org.uk/guidance/cg132/chapter/1-guidance. Accessed: 06-12-2017.

NICE (2014). Intrapartum care for healthy women and babies. NICE guideline CG190, National Institute for Health and Care Excellence. https://www.nice.org.uk/guidance/cg190. Accessed: 06-02-2017.

NICE (2016). Deciding whether to offer caesarean section. NICE pathways, National Institute for Health and Care Excellence. http://pathways.nice.org.uk/pathways/caesarean-section. Accessed: 04-10-2016.

Nielsen, H. S., Smith, N., and Çelikaksoy, A. (2009). The effect of marriage on education of immigrants: Evidence from a policy reform restricting marriage migration. *Scandinavian Journal of Economics*, 111(3):457–486.

Nolasco, A., Melchor, I., Pina, J. A., Pereyra-Zamora, P., Moncho, J., Tamayo, N., García-Senchermes, C., Zurriaga, O., and Martínez-Beneito, M. A. (2009). Preventable avoidable mortality: Evolution of socioeconomic inequalities in urban areas in Spain, 1996-2003. *Health and Place*, 15(3):702–711.

Observatori del Sistema de Salut de Catalunya (2017). *Desigualtats socioeconòmiques en la salut i la utilització de serveis sanitaris públics en la població de Catalunya. Observatori sobre els efectes de la crisi en la salut de la població.* Agència de Qualitat i Avaluació Sanitàries de Catalunya. Departament de Salut. Generalitat de Catalunya, Barcelona.

OECD (2013). *Health at a Glance 2013: OECD Indicators.* OECD Publishing, Paris.

OECD (2015). OECD Health Policy Studies Cardiovascular Disease and Diabetes: Policies for better health and quality of care. (June):page no. 87.

OECD (2017). *Health at a Glance 2017: OECD Indicators.* OECD Publishing, Paris.

Oláh, L. S., Kotowska, I. E., and Richter, R. (2018). Olah et al 2018 - The new roles of men and women and implications for families and societies.pdf. In Doblhammer, G. and Gumà, J., editors, *A Demographic perspective on gender, family and health in Europe*, pages 41–64. Springer, Cham.

Oreopoulos, P., Stabile, M., Walld, R., and Roos, L. L. (2008). Short-, medium-, and long-term consequences of poor infant health: An analysis using siblings and twins. *Journal of Human Resources*, 43:88–138.

Otamiri, G., Berg, G., Ledin, T., Leijon, I., and Lagercrantz, H. (1991). Delayed neurological adaptation in infants delivered by elective cesarean section and the relation to catecholamine levels. *Early Human Development*, 26(1):51 – 60.

Ottosson, F., Brunkwall, L., Ericson, U., Nilsson, P. M., Almgren, P., Fernandez, C., Melander, O., and Orho-Melander, M. (2018). Connection between bmi-related plasma metabolite profile and gut microbiota. *The Journal of Clinical Endocrinology & Metabolism*, 103(4):1491–1501.

Pekkarinen, T. (2008). Gender differences in educational attainment: Evidence on the role of tracking from a Finnish quasi-experiment. *Scandinavian Journal of Economics*, 110(4):807–825.

Pekkarinen, T., Uusitalo, R., and Kerr, S. (2009). School tracking and intergenerational income mobility: Evidence from the Finnish comprehensive school reform. *Journal of Public Economics*, 93(7-8):965–973.

Pestel, N. (2017). Searching on Campus? Marriage Market Effects of the Student Gender Composition. *IZA Discussion Paper*, 11175.

Peters, L. L., Thornton, C., de Jonge, A., Khashan, A., Tracy, M., Downe, S., Feijen-de Jong, E. I., and Dahlen, H. G. (2018). The effect of medical and operative birth interventions on child health outcomes in the first 28 days and up to 5 years of age: A linked data population-based cohort study. *Birth*, pages 1–11.

Prior, E., Santhakumaran, S., Gale, C., Philipps, L. H., Modi, N., and Hyde, M. J. (2012). Breastfeeding after cesarean delivery: a systematic review and meta-analysis of world literature. *The American Journal of Clinical Nutrition*, 95(5):1113–1135.

Ravesteijn, B., van Kippersluis, H., Avendano, M., Martikainen, P., Vessari, H., and van Doorslaer, E. (2017). The Impact of Later Tracking on Mortality by Parental Income in Finland. *Tinbergen Institute Discussion Paper*.

Redondo Calderón, J. L., De Dios Luna Del Castillo, J., Jiménez Moleón, J. J., Lardelli Claret, P., and Gálvez Vargas, R. (2000). Evolución de la mortalidad por accidentes de tráfico en España, 1962-1994. *Gaceta sanitaria / S.E.S.P.A.S*, 14(1):7–15.

Reed, S. D., Lee, T. A., and McCrory, D. C. (2004). The Economic Burden of Allergic Rhinitis. *PharmacoEconomics*, 22(6):345–361.

Regidor, E., Barrio, G., Bravo, M. J., and De La Fuente, L. (2014a). Has health in Spain been declining since the economic crisis? *Journal of Epidemiology and Community Health*, 68(3):280–282.

Regidor, E., Calle, M. E., Navarro, P., and Dominguez, V. (2003). The size of educational differences in mortality from specific causes of death in men and women. *European Journal of Epidemiology*, 18(5):395–400.

Regidor, E., Santos, J. M., Ortega, P., Calle, M. E., Astasio, P., and Martínez, D. (2014b). Decreasing income inequality and emergence of the association between income and premature mortality: Spain, 1970-2010. *Health and Place*, 27:30–37.

Regidor, E., Vallejo, F., Granados, J. A., Viciana-Fernández, F. J., de la Fuente, L., and Barrio, G. (2016). Mortality decrease according to socioeconomic groups during the economic crisis in Spain: a cohort study of 36 million people. *The Lancet*, 388(10060):2642–2652.

Rodríguez-Fonseca, M., Palència, L., Marí-Dell'Olmo, M., Gandarillas, A., Domínguez-Berjón, M. F., Gotsens, M., Rodríguez-Sanz, M., and Borrell, C. (2013). Evolution of socio-economic inequalities in mortality in small geographical areas of the two largest cities in Spain (Barcelona and Madrid), 1996-2007. *Public Health*, 127(10):916–921.

Ruhm, C. J. (2000). Are Recessions Good for your Health? *Quarterly Journal of Economics*, (May).

Ruhm, C. J. (2015). Recessions, healthy no more? *Journal of Health Economics*, 42:17–28.

Ruhm, C. J. (2018). Deaths of Despair or Drug Problems? *NBER Working Paper*, 24188(January).

Ruiz-Ramos, M., Escolar Pujolar, A., Sánchez Perea, J., and Garrucho Rivero, G. (2006). Evolución de las desigualdades sociales en la mortalidad general de la ciudad de Sevilla (1994-2002). *Gaceta Sanitaria*, 20(4):303–310.

Saccone, G. and Berghella, V. (2015). Induction of labor at full term in uncomplicated singleton gestations: a systematic review and metaanalysis of randomized controlled trials. *American Journal of Obstetrics and Gynecology*, 213(5):629 – 636.

Schmertmann, C. P. and Gonzaga, M. R. (2018). Bayesian Estimation of Age-Specific Mortality and Life Expectancy for Small Areas With Defective Vital Records. *Demography*, 55:1363–1388.

Schwartz, C. R. and Han, H. (2014). The reversal of the gender gap in education and trends in marital dissolution. *American Sociological Review*, 79(4):605–629.

Scudellari, M. (2017). News feature: Cleaning up the hygiene hypothesis. *Proceedings of the National Academy of Sciences*, 114(7):1433–1436.

SEGO (2005). Monitorización fetal intraparto. *Progresos de Obstetricia y Ginecología*, 48(4):207–216.

Sevelsted, A., Stokholm, J., Bønnelykke, K., and Bisgaard, H. (2015). Cesarean section and chronic immune disorders. *Pediatrics*, 135(1):e92–e98.

Shurtz, I. (2013). The impact of medical errors on physician behavior: Evidence from malpractice litigation. *Journal of Health Economics*, 32(2):331 – 340.

Skiold, B., Petersson, G., Ahlberg, M., Stephansson, O., and Johansson, S. (2017). Population-based reference curve for umbilical cord arterial pH in infants born at 28 to 42 weeks. *Journal of Perinatology*, 37(3):254–259.

Skoner, D. P. (2001). Allergic rhinitis: Definition, epidemiology, pathophysiology, detection, and diagnosis. *Journal of Allergy and Clinical Immunology*, 108(1, Supplement):S2 – S8.

Spetz, J., Smith, M. W., and Ennis, S. F. (2001). Physician incentives and the timing of cesarean sections: evidence from California. *Medical care*, 39(6):536–550.

Spong, C. Y., Berghella, V., Wenstrom, K. D., Mercer, B. M., and Saade, G. R. (2012). Preventing the first cesarean delivery: Summary of a joint Eunice Kennedy Shriver National Institute of Child Health and Human Development, Society for Maternal-fetal Medicine, and American College of Obstetricians and Gynecologists workshop. *Obstetrics and Gynecology*, 120(5).

Stock, J. and Yogo, M. (2005). Testing for weak instruments in linear IV regression. In Andrews, D. W., editor, *Identification and Inference* *for Econometric Models*, pages 80–108. Cambridge University Press, New York.

Strachan, D. P. (1989). Hay fever, hygiene, and household size. *BMJ (Clinical research ed.)*, 299(6710):1259–60.

Stuart, H. C., Moon, S. H., and Casciaro, T. (2018). Penalty for Success? Career Achievement and Gender Differences in Divorce. In *The Work-Family Interface: Spillover, Complications, and Challenges*, pages 265–289. Emerald Publishing Limited.

Sund, R. (2012). Quality of the Finnish Hospital Discharge Register: A systematic review. *Scandinavian Journal of Public Health*, 40(6):505–515.

Tao, B., Pietropaolo, M., Atkinson, M., Schatz, D., and Taylor, D. (2010). Estimating the cost of type 1 diabetes in the U.S.: A propensity score matching method. *PLoS ONE*, 5(7).

Tapia Granados, J. A. T. (2005). Recessions and Mortality in Spain, 1980–1997. *European Journal of Population / Revue européenne de Démographie*, 21(4):393–422.

Tapia Granados, J. A. T. and Ionides, E. L. (2017). Population health and the economy: Mortality and the Great Recession in Europe. *Health Economics*, 26(December 2016):e219–e235.

Thavagnanam, S., Fleming, J., Bromley, A., Shields, M. D., and Cardwell, C. R. (2008). A meta-analysis of the association between Caesarean section and childhood asthma. *Clinical & Experimental Allergy*, 38(4):629–633.

Tonei, V. (2019). Mother's mental health after childbirth: Does the delivery method matter? *Journal of Health Economics*, 63:182–196.

Torpy, J. M. (2010). Chronic diseases of children. *JAMA - Journal of the American Medical Association*, 303(7):682.

Tuomisto, L., Erhola, M., Luukkaala, T., Puolijoki, H., Nieminen, M., and Kaila, M. (2010). Asthma programme in Finland: Did the use of secondary care resources become more rational? *Respiratory Medicine*, 104(7):957 – 965.

Tur-Prats, A. (2017). Unemployment and intimate-partner violence: A gender-identity approach. *UPF Economic Working Paper Series*, 1564.

Valdes, B. (2013). Analyse démographique de la mortalité par sida en Espagne. *Population*, 68(3):539.

Villalbí, J. R. and Pérez, C. (2006). Evaluación de políticas regulatorias: Prevención de las lesiones por accidentes de tráfico. *Gaceta Sanitaria*, 20(SUPPL. 1):79–87.

Webb, D. A. (2002). Time of day variation in rates of obstetric intervention to assist in vaginal delivery. *Journal of Epidemiology & Community Health*, 56(8):577–578.

WHO (2015). *WHO Statement on Caesarean Section Rates*. World Health Organization, Geneva.

Wright, S. M. and Aronne, L. J. (2012). Causes of obesity. *Abdominal Radiology*, 37(5):730–732.

Zanardo, V., Simbi, A., Franzoi, M., Soldà, G., Salvadori, A., and Trevisanuto, D. (2004). Neonatal respiratory morbidity risk and mode of delivery at term: influence of timing of elective caesarean delivery. *Acta Paediatrica*, 93(5):643–647.

Zeitlin, J., Mortensen, L., Cuttini, M., Lack, N., Nijhuis, J., Haidinger, G., Blondel, B., Hindori-Mohangoo, A. D., Alexander, S., Pavlou, P., Velebil, P., Andersen, A. M. N., Sakkeus, L., Gissler, M., Antsaklis, A., Berbik, I., Ólafsdóttir, H. S., Bonham, S., Misins, J., Jaselioniene, J., Wagener, Y., Gatt, M., van der Pal, K., Klungsoyr, K., Szamotulska, K., Barros, H., Horga, M., Cap, J., Mandic, N. T., Bolúmar, F., Gottvall, K., Berrut, S., and Macfarlane, A. (2016). Declines in stillbirth and neonatal mortality rates in Europe between 2004 and 2010: Results from the Euro-Peristat project. *Journal of Epidemiology and Community Health*, 70(6):609–615.