UNIVERSITA' DELLA CALABRIA

Dipartimento di Economia, Statistica e Fianza

Dottorato di Ricerca in

Scienze Economiche e Aziendali

CICLO

XXXVI

Parental Divorce and Children's Outcomes: Some Evidence from Ireland

Settore Scientifico Disciplinare: SECS-P/01

Coordinatore: Ch.mo Prof. Fabio Piluso

FABIO PILUSO
25.03.2024
13:39:30
GMT+00:00

Supervisore/Tutor: Ch.mo Prof. Vincenzo Scoppa

Dottoranda: Dott.ssa Idola Francesca Spanò

Sommario

Introduction

There is an extensive literature documenting the effects of dynamics within different family types on a wide range of children's outcomes. These outcomes of interest include child development and well-being, behavioural problems, educational achievements, and demographic indicators (such as the dissolution of unions among children from divorced families, marriage rates, and fertility choices). Some researchers have directed their attention towards single-parent families resulting from separation and divorce, while others have explored the differences between children raised by married parents and those raised by cohabiting parents.

In this thesis, we examine the situation in Ireland and explore the impact of parental divorce on children using data from the Growing Up in Ireland (GUI) study. This is a national longitudinal study of children and young people, a joint project of the Department of Children, Equality, Disability, Integration and Youth (DCEDIY) and the Central Statistics Office (CSO). It tracks the progress of two cohorts of children: more than 8,000 9-year-olds (referred to as "Child Cohort") and around 10,000 9-month-olds (referred to as “Infant Cohort”).

Just to give an overview of households and families in the Irish context, we briefly describe the most relevant results of the Census of Population 2022. Similar to other nations, the demography of Irish families has undergone various changes in family structure over the past few decades. According to the Central Statistics Office (CSO)'s Census of Population 2022, there has been a 5% increase in the overall number of families since 2016, resulting in a total of 1,279,951. When compared to 1996, this increase represents a significant rise of 59%. Conversely, the average number of children per family has consistently decreased from 1.82 in 1996 to 1.34 in 2022. The number of families without children has experienced a 14% increase since 2011, while families with four or more children have decreased by 10% compared to 2011. In terms of family types, there has been a 12.7% increase in cohabiting couples with children compared to the data from the 2016 Census. Lastly, there has been a 13% increase in one-parent families headed by fathers, while the number of one-parent households headed by mothers has seen a slight decrease of 1%.

The thesis is divided into three chapters.

In the first chapter, we use three waves of data from the GUI - Child Cohort. Children in this cohort were 9 years old at the time of the first data collection, 13 years old at the time of the second survey, and 17 years old at the time of the third one. We assess the impact of parental divorce on children’s socioemotional problems, as well as their physical and dental health.

Concerning the socioemotional aspect of the child, we use the emotional subscale from the Strength and Difficulties Questionnaire (SDQ) (A. Goodman and R. Goodman 2009; R. Goodman 1997). Physical health is evaluated by considering the child's risk of being overweight and their attitude towards engaging in physical activity. Finally, dental health is assessed by measuring the frequency of the child's visits to the dentist.

Importantly, the available data allows to discern heterogeneity in the effects according to the age of the child at the time of parental divorce. Specifically, we distinguish between two age ranges for the child when the event occurs: between the ages of 9 and 13, and between the ages of 13 and 17.

Regarding the methodology, in the first part of the analysis we employ a Difference-in-Differences approach. Thereafter, to address potential issues related to the endogeneity of parental divorce, we include household's fixed effect to control for the unobserved source of heterogeneity.

Our results show that children who experience parental divorce between the ages of 9 and 13 do not appear to be affected by it. However, when parental divorce occurs during the adolescent's age of 13 and 17, it leads to adverse effects on all the outcomes under consideration.

In the second chapter, we investigate the impact of variations in the family structure due to parental divorce on child's later educational outcomes measured at the age of 20, by using data from the third and fourth waves of the GUI – Child Cohort.

We need to restrict our analysis on these two periods because the educational outcomes are only measured in the fourth wave. Therefore, we employ the information gathered in the third wave when the adolescent is 13 years old to make predictions about the outcomes in the fourth wave when the adolescent is 20 years old.

To identify changes in family structure, we rely on the availability of previous waves of data, which enables us to track changes that occurred over the time span of the study. It is important to note that all transitions identified are predetermined relative to the measured educational outcomes.

The novelty of this chapter lies in the implementation of a recently developed double-machine learning estimator, which facilitates a "data-driven" selection of variables. Basically, we compare the results obtained from estimations based on a traditional "ad-hoc" variable selection procedure with those obtained from the machine learning technique. Additionally, we account for the heterogeneity in the effects according to the age of the children, as done previously.

Our findings indicate that experiencing a change in family structure between the ages of 9 and 13 has an adverse effect on educational achievements when the adolescents reach the age of 20. Undergoing such a transition before the age of 9 or between the ages of 13 and 17 does not exert any influence on their educational outcomes.

In the third chapter, we use data from the GUI – Infant Cohort. Children in this cohort were 9 months old, 3 years old, 5 years old, 7 years old, and 9 years old during the initial, second, third, fourth, and fifth data collections, respectively. Implementing a Difference-in-Differences approach, we study the impact of a change in the family structure, specifically the transition from a two-parent to a one-parent configuration, on the externalizing and internalizing problems experienced by children. Additionally, we assess how the quality of the parent-child relationship is affected by this change. We identify a transition that takes place when children are aged 3-5 and measure its effects both at the age of 5 and 9 respectively.

Unlike the first chapter, a preliminary inspection of the data reveals the presence of statistically significant differences in pre-existing characteristics for two-parent households (control group) and one-parent ones (the treated ones). This might point out that families who undergo a structural change differ from families that do not and that they are a selected group. To deal with potential selection bias, we adopt a propensity score matching as preliminary stage before implementing a Difference-in-Differences approach. Specifically, we use children from traditional two-parent families as control group and children from one-parent families as treated units. Through a Kernel matching procedure, we match treated children with untreated children that have a similar likelihood of experiencing a family structure change. Thereafter, we identify the effect of family structure change measuring the differences in the average outcomes for those two groups of children.

Our results show that children who experience family structure between 3-5 years old have worse internalizing and externalizing problems and the parent-child relationship deteriorates. These effects do not disappear when the child is 9 years old.

Chapter I

The Effects of Parental Divorce on Children Well-Being: A Difference-in-Differences Evaluation

Abstract

In this article, we analyse data from Growing Up in Ireland (GUI) to evaluate whether experiencing parental divorce impacts the well-being of children. Specifically, we examine the impact of parental divorce on children's socioemotional behaviour and physical health. To address the issue of potential endogeneity of parental divorce, we employ a Difference-in-Differences identification strategy and compare the average variations in outcomes experienced by children exposed to parental divorce with the average changes observed in children not exposed to it. Our results reveal the critical role of the age of children at the time of parental divorce in mitigating or exacerbating the effects. Notably, our evidence indicate that parental divorce does not have a negative impact on children's outcomes if it occurs between the ages of 9 and 13. However, if parental divorce takes place when children are between the ages of 13 and 17, it produces negative effects across all measured outcomes. Compared to adolescents from intact families, adolescents exposed to parental divorce between 13 and 17 years of age have greater level of emotional problems, with a 0.6-point increase and a 15-percentage point greater likelihood of experiencing severe emotional problems. Their physical health is negatively affected by parental divorce, with an increase of 16 percentage points in their risk of being overweight and a decrease of 14 percentage points in the likelihood of engaging in physical activity. Lastly, their probability of visiting the dentist at least once a year is 11 percentage points lower. Accounting for time-invariant household's characteristics by estimating a fixed-effects model produces similar results. A robustness check in which we assess the impact of the breakup of cohabiting parents confirms our initial findings. We also document heterogeneous effects by child's gender, primary caregiver's age at child's birth and level of education, as well as household income and area of residence. Overall, our study contributes to the understanding of the role of children's age in the process of adjusting to parental divorce.

JEL Classification: C31, C33, I12, I30, J12

Keywords: parental divorce, age at divorce, child well-being, behavioural problems, overweight risk, dental care

I. Introduction

Decades of research show that family structural change is a key determinant of several child development outcomes. International evidence Amato (2000, 2014), Amato and Keith (1991) and researchers from different fields of study suggest that children in divorced families perform worse than children in stably married families on a variety of outcomes, including educational achievement (Bernardi and Boertien, 2016; Sun and Li, 2011), physical health and well-being (Rasmussen, 2009; Schmeer, 2012; Anderson, 2014; Augustine and Kimbro, 2017), behavioural and conduct problems, as well as emotional problems (Cherlin *et al.*, 1991; Amato and Sobolewski, 2001; Cooper *et al.*, 2008; Fomby and Osborne, 2010).

Early evidence suggested very detrimental effects of divorce on children's outcomes. However, one of main criticisms of early studies is the use of cross-sectional data to compare outcomes for children in stable married families and those of children whose parents are divorced, leading to an overestimate of the actual effects. The main issue related to these findings is whether they can be interpreted as causal as parental divorce is likely to be related to other characteristics (from socioeconomics to child's, mother's, and household's characteristics) that are also likely to affect children's outcomes (Gruber, 2004; Sanz-De-Galdeano and Vuri, 2007; Amato and Anthony, 2014).

To overcome this problem, the use of longitudinal data and the adoption of methodologies that allow to control for unobservable factors has been greatly suggested. Fixed-effect models eliminate the unobserved sources of heterogeneity that are time invariant. Some studies adopt sibling fixed-effect model which is based on observing two siblings within the same family exposed to parental divorce for different lengths of time. This approach allows to remove unobserved family characteristics hold in common by the siblings. Results are mixed, with some reporting no evidence of an association between parental divorce and children's outcomes (Aughinbaugh, Pierret and Rothstein, 2005; Powers, 2005) and others documenting a negative causal impact of divorce (Cherlin, Chase-Lansdale and McRae, 1998; Ermisch *et al.*, 2001; Gruber, 2004).

Another approach employed involves conducting longitudinal comparisons of children's outcomes before and after divorce. The idea is to measure the same outcome before and after the occurrence of divorce. The design is similar to fixed-effect models, although without accounting for time-invariant unobservable factors. As with the previous models, findings fail to yield definitive conclusion, as some studies assess a causal detrimental impact of parental divorce (Strohschein 2005; Sun 2001) and others reject it (Sanz-De-Galdeano and Vuri, 2007).[1]

In this paper we investigate the impact of parental divorce on children's emotional problems, physical health and exercise, and dentist frequency, using a nationally representative sample of children from the *Growing Up in Ireland (GUI)*, the National Longitudinal Study of Children and Young people in Ireland. We use the first three waves of the Child Cohort, a representative sample of 8,568 children aged 9. The Cohort, that is the children and their families, has been re-surveyed through follow-ups when the children were 13 and 17 years old, respectively.

To assess the impact of parental divorced on child's emotional problems we use data from the Strength and Difficulties Questionnaire (SDQ), a brief behavioural screening questionnaire (Goodman, 1997; Goodman and Goodman, 2009). It contains 25 items divided into 5 scales of 5 items each and it measures the frequency of the mother-reported behaviours on a three-point scale marked as "Not true", "Somewhat true", and "Certainly true" (scored 0, 1, and 2 consecutively). Responses are summed to produce a scale ranging from 0 to 10. The measure has been frequently used and validated to assess psychological difficulties in children (Stone *et al.*, 2010).

Child's physical health is assessed using two variables: the likelihood of being overweight/obese; and the likelihood of being physically active. Using child's height and weight measured at the end of each survey by the interviewer, we first calculate child's body mass index (BMI). Then, based on specific BMI thresholds[2], we define a categorical variable Overweight/Obese if the child falls in this category. Both weight and height were recorded using medically approved devices (SECA 761 flat mechanical scales and Leicester portable height measure) in kilograms and centimetres respectively. Finally, we evaluate child's dental health through a variable which measures the likelihood of visiting the dentist at least once/twice a year in contrast to never visit it or only when there is a problem. Note that we create our response variables using the information provided by the primary caregiver to ensure comparability among surveys.

The longitudinal nature of the data allows to control for potential endogeneity of parental divorce through the implementation of the Difference-in-Differences (DiD thereafter) model and of the Two-Way-Fixed-Effect (TWFE) model. The DiD approach relies on observing the outcomes of interest both for children who experience parental divorce and children who do not before and after divorce takes place. The TWFE model allows to include household and time-fixed effects.

We conduct separate estimations of the impact of parental divorce according to the child's age at the time of the event. Specifically, we examine the effects of experiencing parental divorce

[1] Many of these methodologies are reviewed in Amato (2014)

[2] The World Health Organisation suggests a threshold BMI of 25 for "overweight", a threshold of 30 for "obesity" and a threshold of 40 for "severely obese".

between the ages of 9 and 13 on outcomes at 13 and 17 years old, and the effects of experiencing parental divorce between the ages of 13 and 17 on outcomes at 17 years old. Further details regarding our methodology for identifying the event of interest are provided in the dedicated section below. For the first group of children, we are able to analyse the effects of parental divorce both in the short term and in a medium-long term perspective. However, for children who experience parental divorce between the ages of 13 and 17, we can only observe the effects in a shorter time frame, as the outcomes are only observed for one period after the event.

It is important to acknowledge that there does not exist a singular definition of short- and medium-long term effects of divorce in the existing literature, as different time frames are employed. The length of time for which children's paths can be observed after parental divorce plays a significant role in this definition. To prevent any confusion, we will refer to short-term effects in our paper when analysing children's outcomes immediately following parental divorce, and we will refer to medium-long term effects when we are able to observe children's outcomes for multiple periods after parental divorce has occurred.

Our findings show that age plays a significant role in the adjustment process of children to divorce. In this regard, a consistent strand of literature argues that parental divorce has particularly harmful effects on children who experience it during early childhood, specifically before the age of 5 (Cherlin, Chase-Lansdale and McRae, 1998; Ermisch *et al.*, 2001; Cavanagh and Huston, 2006; Amato and James, 2010; Kravdal and Grundy, 2019).

Consistent with this perspective, we do not observe any statistically significant effects on almost all the measured outcomes for children who experience parental divorce between the ages of 9 and 13, except for a slight impact on the likelihood of visiting the dentist frequently in the medium-long term, that is at the age of 17.

Accordingly, we should not expect to find any statically significant effects of parental divorce on adolescents who experience it between the ages of 13 and 17. However, the main contribution of our paper is that, among children who experience parental divorce in early adolescence, the age of the children significantly shapes these effects. Indeed, our results provide evidence of a strong negative effect of parental divorce on adolescents who undergo it between the ages of 13 and 17.

Results show that these adolescents present higher levels of emotional problems on average, with a difference of 0.6 points compared to adolescents from households with stably married parents. Furthermore, they have an almost 15 percentage points higher likelihood of experiencing severe emotional problems. In terms of health-related outcomes, adolescents with divorced parents face a higher risk of being overweight or obese by on average 16 percentage points, and are less likely to engage in physical activity by 14 percentage points. Lastly, their probability of frequent dental visits is 11 percentage points lower compared to adolescents from intact families.

These results are in line with other studies that documented an adverse impact of divorce on children's socioemotional behaviour (Paul R. Amato and Keith, 1991; Amato, 2000; Sigle-Rushton, Hobcraft and Kiernan, 2005; Rasmussen, 2009), and on health-related outcomes (StØrksen *et al.*, 2006; Rasmussen, 2009; Augustine and Kimbro, 2013; Bright *et al.*, 2015; Goisis, Özcan and Van Kerm, 2019).

In our main analysis, we identify changes in family structure by considering the legal marital status of the primary caregiver, excluding from the sample cohabiting parents who never married. As a robustness check, we include those households made of partners who never married, but cohabit, estimating consequently the impact of a breakup between cohabiting parents, regardless of their marital status. We find very similar results with statistically significant effects only for the group of adolescents who experience the breakup of their parents at a later age.

Although our analysis mainly focuses on the average effect of parental divorce, it might be interesting study whether the impact changes according to specific characteristics. To this end, we run our analysis on subgroups based on child's gender, primary caregiver's age at child's birth and level of education, household income, and area of residence. Our results document large heterogeneous effects of parental divorce occurring between the ages of 13 and 17, with the worse outcomes observed for girls, children of low educated and younger primary caregiver, children in households with disadvantage economic conditions, and for those living in rural area.

The remainder of the paper is organized as follows. Section II briefly reviews the literature. Section III describes the data and the criteria adopted to build our sample. Section IV lays out the empirical strategy. In Section V and VI we investigate the impact of parental divorce between 9 and 13 years and between 13 and 17 years, respectively, children's outcomes using an OLS estimator. In Section VII we propose the results from a fixed-effect model. In Section VIII we run a robustness check using breakup of cohabiting parents as event on interest. Section IX assesses potential heterogeneity in the impact of parental divorce. Section X concludes.

II. Literature Review

The impact of parental divorce on child's emotional and behavioural traits is well documented in the literature. Using a child fixed-effects model and using two national data sets, the Early Childhood Longitudinal Study - Kindergarten Cohort (ECLS-K), and the National Educational Longitudinal Study (NELS), Amato and Anthony (2014) found that children of divorced parents reported higher internalizing and externalizing problems (0.3 standard deviation) than children whose parents remained married.

Using a sample of 8,984 Norwegian adolescents (13–19 years) from the HUNT study – a longitudinal population health study in Norway, StØrksen et al. (2006) show that children whose parents divorced reported greater symptoms of anxiety and depression, poorer self-reported subjective well-being, and higher academic problems (which essentially refer to externalising problems) relative to the group of children whose parents stayed together, and these effects are stable over time. Girls show enduring symptoms of anxiety and depression than boys.

Others focus on the role of gender and age at parental divorce on the adjustment process of the child. Using the Norwegian Population Register and Norwegian Prescription Database, Kravdal and Grundy (2019) provide evidence of more detrimental effects of parental divorce if experienced at a younger age: children aged 15-19 when their parents divorced were 12 per cent less likely to purchase antidepressants in adulthood compared to children who aged 0-4 at parental divorce; for those aged 20 or more at parental divorce this likelihood decreases by 19 percentage points than those aged 0-4.

The impact of parental divorce is found to be worsened by the number of transitions experienced within the household. Using an administrative register dataset containing information on the entire population of Danish children born in 1985, Rasmussen (2009) analyses the impact of family structure changes due to parental relationship dissolution on educational, behavioural, and health children' outcomes ages 12 or more. He shows that children who experienced family structure changes have worse outcomes than children who did not in all the outcomes considered. These detrimental effects exacerbate as the number of transitions increases and if parental separation occurs before child turns 8.

Using the Strength and Difficulties Questionnaire (SDQ) to evaluate the well-being of children aged 8 or more, Bream and Buchanan (2003) found that children who experienced parental divorce present higher level of distress relative to children in intact families both in the short and in the long term with gender differences. In particular, 52 % of boys and 48 % of girls present significant adjustment problems immediately after the court proceeding and these problems persist one year later, mainly for boys than girls (62 % and 32 % respectively).

Although overweight and obesity have important implications for children's future health (Berenson, 2012), few studies have investigated the relationship between family disruption and children physical health. Using the Early Childhood Longitudinal Survey-Birth Cohort (ECLS-B), a nationally representative sample of children born in 2001, Augustine and Kimbro (2017) show that children in single and cohabiting families have a significantly higher likelihood of being obese relative to children with married parents (24% vs 18% respectively), with girls having worse outcomes than boys.

Goisis, Özcan, and Van Kerm (2019) assess the impact of parental separation on children's body mass index (BMI) and overweight/obesity risk using the UK Millennium Cohort Study (MCS), a nationally representative longitudinal data set of children born in 2001 and followed until age 14. Estimating child fixed-effects models, they do not find evidence of anticipation effects or of statistically significant effects immediately after the separation in children's physical health. Instead, they do find significant effects in the long term, with the BMI of children whose parents separate significantly deviates from the BMI of children whose parents stay married. In addition, the differences are even larger if separation occur before children turn 6.

Overall, the main prevalent result is that children in single or unmarried-parent families face a higher risk of being overweight or obese than children in two-parent or married families (Gibson *et al.*, 2007; Chen and Escarce, 2010; Bzostek and Beck, 2011; Schmeer, 2012).

Physical exercise plays an important role in child's physical well-being. According to the literature, parents significantly influence their children's physical activity levels, including the type of activity they engage in (Golan and Crow, 2004; Beets, Cardinal and Alderman, 2010). However, the impact of family formation is less clear. Lower level of stability, often associated with single parent families, leads to poor child-exercise behaviour (Chen and Escarce, 2010). Yelick (2017) shows that youth from single-parent families are more likely to have low-levels of exercise patterns compared to youth from two-parent families.

Child's dental health has been found susceptible to adverse childhood experience as well. Poor dental health has important implications for children's general health and quality life. It hinders weight gain, growth, and cognitive development of young children (Sheiham, 2006). Dental health can interfere with the ability of children to succeed in school, with a primary negative impact on the number of absences at school due to oral health problems; in turn, missed school time lead to a decline in school performance (Jackson *et al.*, 2011).

Bright et al. (2015) investigate the impact of adverse child' stressors, into which parental divorce falls, on dental health measured through tooth condition and the presence of caries. Their results show that even the presence of one of these stressful events increases the child's likelihood of having poor dental health. Listl (2011) assesses the association between family composition and children's dental health behaviour. He shows that children growing up in either family with the birth mother and a stepfather or with adoptive parents have a lower likelihood of accessing dental services in comparison with children who grow up with both biological parents.

III. Data and descriptive statistics

To analyse the impact of divorce on the emotional well-being of children, their physical health and activity levels, as well as their frequency of dental visits., we use data from the *Growing Up in Ireland* (GUI), the National Longitudinal Study of Children and Young people in Ireland. The study aims at gathering information on the factors that contribute to or undermine children development and well-being.[3]

The present study focuses on the Child Cohort, a cohort of 8,568 nine-year olds children born in 1998. After the first wave of data collection, which took place between September 2008 and March 2009 when the children were nine-year-olds, a second follow-up took place between August 2011 and March 2012, when the children were thirteen-year-olds; a third interview took place between April 2015 and August 2016, when they were seventeen/eighteen-year-olds; finally, a fourth wave took place between August 2018 and June 2019 at age 20. At wave 1 a total of 8,568 9-year-olds and their families were interviewed. 7,525 families of these participated at wave 2, giving a response rate of 89 per cent. The response rate decreases reaching a 76 per cent (6,216 households) at wave 3 and 65 per cent (5,190 households) at wave 4.

Table 1 below provides a summary of all the relevant information outlined above regarding the study design.

Table 1. Study design and response of the Child Cohort

Wave	I	II	III	IV
Data collection date	Sept 2008 – Mar 2009	Aug 2011 – Mar 2012	Apr 2015 – Aug 2016	Aug 2018 – Jun 2019
Child's age	9	13	17	20
Involved families	8,568	7,525	6,216	5,190

In all surveys, the home-based questionnaire, administered at the household level, was filled in by the Primary Caregiver – the person who provided most care and the one with most knowledge about the Study Child - the Secondary Caregiver – the spouse or partner of the Primary Caregiver - and the Study Child him/herself. We recover information on the household composition, socio-demographics, child's health and wellbeing, and mother's marital status from the main questionnaire filled in by the primary caregiver. We do not include information on the secondary caregiver, in most of the cases the child's father, because of a substantial number of missing values.

We follow a few criteria in the cleaning process of the dataset. First, in our study we consider children whose families participated in the first three interviews because the outcomes of interest are not measured in the fourth one. Second, to ensure that the data are comparable over time, we include in our sample families whose primary caregiver does not change across surveys (it

[3] The GUI involves two cohorts of children: the Child Cohort and the Infant Cohort, a group of 9-year-olds and 9-month-olds children, respectively, at the time of the first interview.

only changes in 966 households). More important, to be able to identify the event of interest, i.e. parental divorce, we only consider households in which the marital status of the primary caregiver is married, separated, or divorce, i.e., we do not include the categories never married or widow. Never married parents have been removed since it has been pointed out as a limitation of studies that mix children with divorced parents and children with never married parents (Amato, 2010). The occurrence of never married primary caregiver interests 1,864 households in our sample, of which 800 report living with a partner at home. These aspects will be discussed later in the robustness check section. Likewise, removing children who experience parental death is a strategy commonly adopted in the literature to avoid that results are confounded (Amato and Anthony, 2014). In this case, 245 primary caregivers are widow in our sample, and only 15 of them live with a partner.

Parental divorce is identified through the primary caregiver's marital status reported at each interview. It is important to highlight that the data at hand do not allow to recover detailed information on the marital history of the parent, such as the date of marriage or the date of divorce. The only way for us to code a family transition is to observe a change in the marital status from being married in the first survey to be separated/divorce in the following ones. Hence, we can identify a parental breakup between two waves, but not the exact moment at which it occurs. Note that we aggregate into one category both separated and divorced parents and thereafter we will refer to it simply as divorced parents.

We classify families into four mutually exclusive categories: a) *intact families*, a group of households in which parents remain married for the entire period; b) *already divorced families,* in which parents were already divorced at the time of the first survey; c) *a first group of non-intact families,* in which parental divorce occurs between age 9 and 13 of their children (i.e., the change in the marital status is observed between the first and the second wave); d) *a second group of non-intact families* where parents stay married both in the first and second survey, and then divorce when the child is between 13 and 17 years old. Note that once we observe parental divorce, we do not consider further potential transitions, i.e., parents who reunite or have a new partner.

This cleaning process results in a balanced panel of 14,556 observations, with 4,852 households followed from when the child is 9 years old to 17 years old. Table 2 reports the summary statistics. About 49 % of children in our sample are boys. On average, they are 9, 13, and 17 years old in the first, second, and third wave respectively. The average child's birthweight is 3.54 kgs. 57% of children in our sample have been breastfed. Twins represent about 3 % of children in our sample. Average height and weights are approximately 157 cms and 52 kgs respectively. A child has on average 1.93 siblings[4] and 5.2 % of children in our sample is an only child. Almost 4 % of them are not Irish.[5]

Table 2. Descriptive Statistics

	Mean	Std. Dev.	Min	Max	Obs
Outcomes:					
Emotional problems	1.843	1.955	0	10	14334
Severe emotional problems	.191	.393	0	1	14556
Overweight/Obese	.224	.417	0	1	9544
Physical activity	.939	.24	0	1	14509
Dentist frequency	.829	.376	0	1	14556
Child's Characteristics:					
Boy	.488	.5	0	1	14556
Age	12.999	3.268	8	18	14556
Birthweight	3.54	.595	1.7	6.1	14412
Been breastfed	.565	.496	0	1	14448
Twin	.031	.173	0	1	14556
Height in cms	156.904	16.428	121	192	14235
Weight in kgs	51.601	17.498	21	110	14227
Number of siblings	1.927	1.002	0	5	14556
Only Child	.052	.222	0	1	14556
Not Irish	.04	.197	0	1	14556
Primary Caregiver's Characteristics:					
Female	.994	.079	0	1	14556
Age	44.746	5.633	26	58	14556
Non biological parent	.003	.056	0	1	14556
Pcg's years of education	14.733	2.211	8	18	14555
Employed	.646	.478	0	1	14556
Unemployed	.015	.123	0	1	14556
Currently outside labour force	.339	.473	0	1	14556
Household's Characteristics:					
Predicted real income	17501.287	4060.672	362.957	31012.26	14555
Rural area	.568	.495	0	1	14556
Family Structure:					
Intact families	.899	.301	0	1	14556
Divorced before child's 9	.051	.221	0	1	14556
Divorced between child's 9-13	.026	.16	0	1	14556
Divorced between child's 13-17	.023	.15	0	1	14556

Notes: Dataset GUI Child Cohort, wave 1 – wave 3. Note that the number of observations for the variable overweight/obese is remarkably lower than the others because we set the variable equal to 0 if the child is underweight.

Regarding primary caregivers' characteristics, in 99.4 % of cases they are female; the average age is 41, 45, and 49, respectively in the first, second, and third wave; only 0.3% of those interviewed is not the biological parent of the child. On average, primary caregiver has acquired nearly 14.73 years of education[6]. About 65 % are employed, 1 % are unemployed and 34 % are outside the labour force.

[4] Unfortunately, the data does not allow us to establish sibling's gender and age because of the way information was gathered. At each wave, the grid containing information on family members' characteristics other than Study Child and mother refers to a different person. We could only recover the number of siblings by taking the highest number of occurrences of "full brother/sister of Study Child" within each household and over all the surveys.

[5] Regarding child's nationality, the data allow us to assess only whether the child is an Irish citizen.

[6] Since the GUI contains information only on the highest educational qualification obtained, we compute the number of years of education using the legal duration of different educational grades as follows: 8 for primary education or none (unfortunately the data do not allow to separate this information); 11 for lower secondary education (Junior Secondary

In terms of household characteristics, to avoid a potential issue of endogeneity wherein income might be affected by divorce itself, we adopted an alternative approach following De Paola and Gioia (2017). Specifically, instead of using the level of income reported at each wave, we compute household predicted income in the absence of divorce by considering age and years of education of the primary caregiver, a dummy for being employed, a dummy for residing in a rural area or not, the size of the family, and temporal indicators to capture income variations over the course of the survey.[7] The annual predicted income is on average 17,501 €.

Finally, about 57 % of the households in our sample lives in a rural area[8]. Note that, concerning the area of residence, data were collected only in first two surveys, not in the third one. Since we observe for some households a change in residence, to avoid to mistakenly impute a value to the variable in the third wave, and since the relocation might be itself a result of divorce, we only use information provided at the first wave and keep it constant on the next ones.

With regards family structure, 90 % of households in our sample is represented by intact families in which parents stay married throughout the surveys; 5.1 % of them were already divorce at the time of the first survey when children are 9 years old; 2.6 % incurs in parental divorce when the child is between 9 and 13 years old; in the remaining 2.3 %, parental divorce takes place when the child is between 13 and 17 years old.

III.I Trends in Children's Outcomes: A Preliminary Evidence of the Impact of Parental Divorce

In this section we first describe the outcomes of the children measured in our analysis and then, by means of a graphical representation, we compare the evolution of these variables for children in the four types of families identified above.

Our first dependent variable is the *Emotional Problem Subscale* from the Strength and Difficulties Questionnaire (Goodman, 1997; Goodman and Goodman, 2009). The subscale consists of the following five items pertinent to socioemotional and behavioural disorders in children. These items include statements such as "Often complains of headaches, stomach aches or sickness"; "Many worries, often seems worried"; "Often unhappy, downhearted or tearful"; "Nervous or clingy in new situations, easily loses confidence"; and "Many fears, easily scared". For each item,

---

School); 13 for upper secondary education (Senior Secondary School); 16 for non-degree e primary degree (the former refers to technical and vocational offered by Institute of Technology and specialized colleges, while the latter refers to Bachelor's Degree); 18 for postgraduate education (even in this case, we are not able to distinguish between Master's Degree or PhD Degree).

[7] It is worth noting that before predicting household income, we adjust it for inflation using the Consumer Price Index.

[8] The data provide information only on whether a family lives in a rural or an urban area, but do not allow to access to more detailed information.

the primary caregiver provides an evaluation of child's behaviour on a three-point scale marked as "Not true", "Somewhat true", and "Certainly true" (scored 0, 1, and 2 respectively). The final score is obtained by summing up the points for each item; therefore *Emotional Problem* is a numeric variable ranging from 0 to 10. Higher values on this scale indicate more intense emotional problems.

Based on specific thresholds, it is possible to classify the emotional subscale into three classes: normal, borderline, and abnormal. The thresholds that define the three classes are the following: 0-3 for the normal one, 4 for borderline one, and 5-10 for the abnormal one. Scoring beyond the normal category constitutes a warning signal and may imply the existence of clinically significant behavioural and socioemotional problems. Therefore, we use as our second dependent variable *Severe Emotional Problems,* a dummy variable taking the value of 1 if the child obtains a score above the normal threshold, and 0 otherwise.

The health-related measures used in the analysis are the following: *Overweight/Obese*, a dummy equal 1 if the child is overweight/obese, and 0 if s/he is in good health, meaning in a normal weight status. The use of the GUI data has the advantage of providing a measure of child's weight and height recorded at each wave by the interviewer through medically approved devices. Using this information, we derive our dependent variable based on child's body mass index BMI.

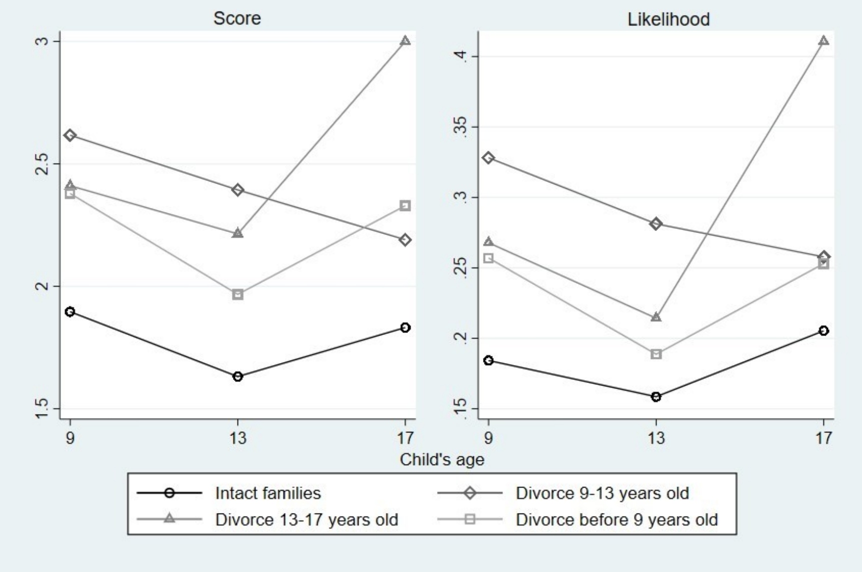
Physical activity, which is a dummy variable that takes the value of 1 if the child engages in physical activity for a minimum of 20 minutes within the previous 14-day period, and 0 otherwise. *Dentist frequency*, a dummy variable equal 1 if the child attends dental visits at least once a year, and 0 if the occurrence of the visits reduces to almost never or only when necessary. The information pertaining to the child's level of physical activity and dentist frequency is derived from responses provided by the primary caregiver.

As shown in Table 1 in the previous section, children in our sample present on average an emotional problems score of 1.84, 19% of them exhibit emotional problems above the normal threshold, 83% of them visit the dentist at least once a year, 22.4% is at risk of being overweight/obese, and 94% of them perform at least 20 minutes of physical exercise.

Figure 1 displays the mean values of children's emotional problems in the four groups of families identified: children from intact families; children whose parents divorce before they are 9 years old, when they are between the ages 9 and 13, and when they are between the ages 13 and 17.

The values presented in the left and right panels are expressed as scores and likelihood of encountering such problems, respectively.

Figure 1: Mean Emotional Problems by Household Categories



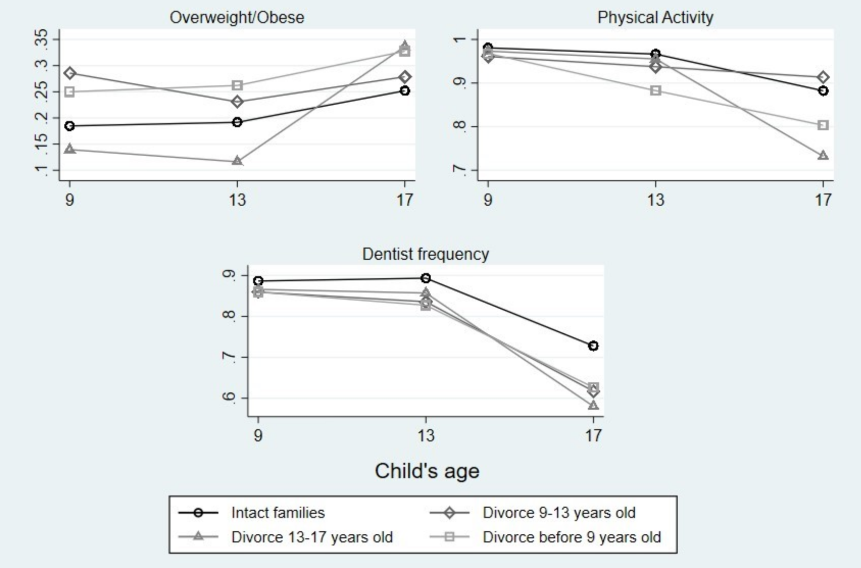
Important insights come up from this graph. First, emotional problems tend to decrease on average for all children between 9 and 13 years of age. After the age of 13, the emotional difficulties increase for children from both household types, intact and disrupted, except for children whose parents divorce when they are between 9 and 13 years old. Children in intact families always fare better than children with divorced parents, presenting the lowest emotional problems at each age. On the contrary, children who experience parental divorce present higher emotional problems in all the observed periods, even before parental divorce takes place.

Overall, the age range of 13-17 appears to present a challenging phase for all adolescents. However, for those whose parents undergo divorce before reaching 9 years of age or during the ages of 13-17, the worsening is remarkable. The circumstances differ for children whose parents divorce between the ages of 9 and 13: initially, they exhibit the most adverse outcomes, but subsequently, it appears that following the occurrence of parental divorce, there is a consistent tendency for emotional problems to diminish. These variations may stem from dynamics or factors related to parental divorce that we cannot observe, such as parental conflict, quality of parenting, and in general stress associated with unhappy family life.

Figure 2 shows the mean values for the health-related outcomes. The picture presented here is quite different from the one proposed above. Initially, it is worth noting that there are no pre-existing variations in the probability of performing physical activity or visiting frequently the dentist among children of different family types. Instead, there are initial divergences in the likelihood of being overweight.

Focusing on the likelihood of being overweight, children from families that have experienced parental divorce, either before the age of 9 or between the ages of 9 and 13, present the highest level of risk at the age of 9. Children in intact families are ranked as the second lowest risk category, while children whose parents divorce between the ages of 13 and 17 years have the lowest risk of being overweight. There are some minor variations observed between the ages of 9 and 13. Similar to the emotional problems, once the child reaches the age of 13, there is a reversal in the trend, resulting in an increase in the likelihood of being overweight. However, the magnitude of this increase varies depending on the family type. More specifically, at the age of 17, children from intact families have the lowest overweight risk, while children in disrupted families fare worse in comparison. Furthermore, the group of children who experience parental divorce between the ages of 13 and 17 experiences the greatest increase in the probability of being overweight.

Figure 2. Mean Health Measures by Household Categories



As regard the trend in child's level of engagement in physical activity, it remains relatively stable between ages 9 and 13, except for children whose parents divorced prior to age 9. Once reached the age of 13, adolescent’s likelihood of exercising decreases, and those exposed to parental divorce between ages 13 and 17 exhibit the most noticeable decline.

Although the probability of visiting the dentist frequently is relatively uniformed distributed among children from all household types at the age of 9, the variable presents a decreasing trend for all children after the age of 13, but for those from disadvantaged families the decline is more pronounced.

IV. The Empirical Strategy

We estimate the impact of parental divorce on children’s outcomes by implementing a Difference-in-Differences (DiD) study design (Meyer, 1995). The DiD approach relies on observing children’s outcomes before and after the divorce takes place, both for children exposed to the divorce (the so called treated) and for those not exposed (the so called untreated). Then, its basic idea is to compare the difference in the average outcomes for children in the treated group with those of children in the untreated group.

As a starting point, let us consider the simplest form of the DID design in which there are only two groups, $g = 1$ and $g = 2$, observed in two periods, $t = 1$ and $t = 2$ (later on, we will refer to this design as DID^{2x2}). In the first period, both groups are exposed to the same condition, that is children grow up in families in which parents are married. Between these two periods, parents in some households go through divorce while others stay married, exposing the former to the treatment condition.

Let $Disrupted_g = 1\ [g = 2]$ be a dummy variable identifying observations in the treated group. Let $Post_t = 1\ [t = 2]$ be a dummy variable identifying observations from period 2, that is the post-divorce period. The variable that identifies the treatment condition is the dummy $Divorce_{gt}$ which takes value of 1 if child in group g experiences parental divorce between period 1 and period 2, and the value of 0 if parents are still married in period 2. Since parental divorce takes place after period 1, by definition $Divorce_{g1} = 0$ for both groups in period 1. Hence, the treatment variable is the product of the two dummy variables stated above: $Divorce_{gt} = Disrupted_g * Post_t$ and measures the causal impact of parental divorce.

The main assumption behind the DiD strategy is the *parallel trend assumption* which states that, in the absence of parental divorce, changes in the outcome variable over time would have been the same both for children who experience parental divorce and for those who do not.

One way to make this clearer is to think about the average outcomes for the two groups in terms of potential outcomes that describe the same unit under different (hypothetical) treatment conditions (Wing, Simon and Bello-Gomez, 2018). Let $Y(1)_{gt}$ be the outcome of interest for unit in group g in time t under the hypothetical scenario in which units are exposed to the treatment condition; $Y(0)_{gt}$ is the outcome of the same unit and time under the alternative scenario in which the control condition is in force. The treatment effect for a specific unit and time period is given by the difference in the outcome across the two scenarios, $\Delta_{gt} = Y(1)_{gt} - Y(0)_{gt}$. In practice, we cannot observe the same unit under two different scenarios since each unit is exposed to only one treatment condition in a specific time period. What we observe for a given unit and time is $Y_{gt} = Y(0)_{gt} + [Y(1)_{gt} - Y(0)_{gt}]D_{gt}$.

In our context, the untreated potential outcome is $Y(0)_{gt} = \beta_0 + \beta_1 Disrupted_g + \beta_2 Post_t + \varepsilon_{gt}$. In the absence of the treatment, the average outcome for group 1 is β_0 in period 1 and $\beta_0 + \beta_2$ in period 2. The average untreated outcome for group 2 is $\beta_0 + \beta_1$ in period 1 and $\beta_0 + \beta_1 + \beta_2$ in period 2. Therefore, under the common trend assumption, β_1 captures the so-called group effect, that is the time-invariant difference in outcomes between the two groups; β_2 measures the time trend, that is the group-invariant difference in outcomes between the two periods.

The treated potential outcome is given by the sum of the untreated potential outcome plus a treatment effect, $Y(1)_{gt} = Y(0)_{gt} + \beta_3$. Combining the two potential outcome specifications, we get a general formula for the outcome: $Y_{gt} = Y(0)_{gt} + D_{gt}[Y(1)_{gt} - Y(0)_{gt}]$. Replacing the potential outcomes with the model specification gives $Y_{gt} = \beta_0 + \beta_1 Disrupted_g + \beta_2 Post_t + \varepsilon_{gt} + Divorce_{gt}[Y(0)_{gt} + \beta_3 - Y(0)_{gt}]$.

Since in two-group two-period setting $Divorce_{gt} = Disrupted_g * Post_t$, by rearranging the above equation we obtain the standard DiD estimating equation:

$$Y_{gt} = \beta_0 + \beta_1 Disrupted_g + \beta_2 Post_t + \beta_3 (Disrupted_g * Post_t) + \varepsilon_{gt}.$$

The data available to us present multiple groups and time periods, but we decide to perform the analysis in subsamples in order to accommodate the requirements for the simple DID^{2x2}. To this end, we restrict our analysis to families with married parents in the first survey, excluding the ones who were already divorced. In addition, we evaluate separately the impact of parental divorce when this takes place when the children are between 9 and 13 years old and when it takes place between 13 and 17 years old. In the former case, we restrict our analysis on the subsample made of by intact families and families in the first treatment group and we use all the three waves available. This allows us to measure both the impact in a shorter period and after a longer period from the incidence of parental divorce. In the latter case, we include in our sample intact families and families from the second treatment group, and we use only the last two waves of data. More details are provided in the specific section.

V. Parental Divorce between age 9-13 and Child's Outcomes: OLS estimates

In this section we examine the relationship between the occurrence of parental divorce when children are aged between 9 and 13 and child's emotional problems using a simple Ordinary Least Squares (OLS) estimator. Therefore, we set our attention to the first group of treated children and compare the outcomes for those children with the outcomes observed for children in intact families.

We measure the outcomes both at age 13 and 17 slightly modifying the above reported standard DID equation and adding a further interaction term as follows:

$$Y_{it} = \beta_0 + \delta_1 Disrupted_i + \beta_1(Disrupted_i * age_{13}) + \beta_2(Disrupted_i * age_{17}) + \beta_3 \boldsymbol{C}_{it} + \beta_4 \boldsymbol{PCG_{it}} + \beta_5 \boldsymbol{H}_{it} + \gamma_1 age_{13} + \gamma_2 age_{17} + \varepsilon_{it}$$

where Y_{it} is the outcome of interest for the child in household *i* at time *t*; $Disrupted_i$ takes value of 1 if parents in household i divorced between 9 and 13 years of age of the child, and 0 otherwise (i.e., parents stay married); $Disrupted_i * age_{13}$ measures the impact of divorce when the child is 13 years old (we would get the same result if we would had included the variable $Divorce_{it}$ instead); $Disrupted_i * age_{17}$ measures the change in the impact of divorce when the child is 17 years old; $\boldsymbol{C}_{it}$ is a vector of child's characteristics (gender, birthweight, been breasted, not Irish, twin birth, number of siblings); the vector $\boldsymbol{PCG}_{it}$ includes primary caregiver's characteristics (gender, not being the biological parent, age, age squared, years of education, employed, unemployed); $\boldsymbol{H}_{it}$ is a vector of household characteristics (predicted income and rural area); age_{13} and age_{17} are two dummies equal 1 if the child is 13 and 17 years old, respectively, and 0 otherwise; ε_{it} is an error term.

We estimate four different specifications, starting with the simplest in which we only include the event of interest, namely parental divorce. Subsequently, we progressively add the sets of control variables. Child's age dummies are included in all the specifications to assess variation in our dependent variable over time. The results are reported in Table 3.

Results from the first column, in which we do not add any control variables, reveal that prior to the occurrence of parental divorce, children in the two groups exhibit different level of emotional problems. Specifically, for children in the treated group, the emotional score is 0.721 points higher than the one reported by untreated children. This difference is statistically significant at the 1 percent level (as indicated by the coefficient of the variable *Disrupted*).

Once parental divorce took place, the difference in the emotional problems remains almost unchanged: the impact of parental divorce is not statistically significant different from zero (first row in column 1). As children reach the age of 17, the difference actually decreases, yet the impact of parental divorce remains statistically insignificant.

The result is robust to the inclusion of child, primary caregiver, and households' characteristics as shown in the second, third, and fourth column respectively. Across all the specifications, the impact of divorce continues to be statistically insignificant.

Table 3. OLS Estimates. Effects of Parental Divorce on Child's Emotional Problems

	(1)	(2)	(3)	(4)
Disrupted * age13	.041	.047	.119	.016
	(.203)	(.206)	(.207)	(.21)
Disrupted * age17	-.362	-.347	-.277	-.354
	(.237)	(.238)	(.237)	(.241)
Boy		-.54***	-.526***	-.521***
		(.045)	(.045)	(.045)
Birthweight		-.047	-.034	-.028
		(.042)	(.042)	(.042)
Been breastfed		-.263***	-.14***	-.121**
		(.046)	(.048)	(.048)
Not Irish		-.068	-.09	-.145
		(.105)	(.103)	(.105)
Twin		.287*	.309**	.337**
		(.154)	(.154)	(.154)
Number of siblings		-.079***	-.152***	-.056*
		(.023)	(.025)	(.034)
Predicted Income			-.059***	.03
			(.007)	(.022)
Rural area			-.176***	.004
			(.048)	(.064)
Male pcg				.181
				(.277)
Non biological parent				.102
				(.782)
Pcg's age				-.073*
				(.041)
Pcg's age squared				.047
				(.046)
Pcg's years of education				-.108***
				(.03)
Employed				-.194***
				(.055)
Unemployed				.084
				(.169)
Age 13	-.265***	-.268***	-.413***	-.025
	(.029)	(.029)	(.033)	(.088)
Age 17	-.065*	-.067**	-.271***	.301**
	(.034)	(.034)	(.041)	(.128)
Disrupted	.721***	.753***	.676***	.633***
	(.186)	(.188)	(.189)	(.19)
Observations	13264	13124	13123	13123
R-squared	.006	.034	.044	.047

Notes: The Table reports OLS estimates. Sample: GUI dataset, children aged 9, 13, and 17 years old. The dependent variable is *SDQ Emotional Problems Score.* Standard errors, reported in parentheses, are corrected for heteroskedasticity, and allow for clustering at the household level. To make easier the interpretation of the coefficients, in all the regressions we have rescaled the variables *Income*, dividing it by 1000, and *Age2,* dividing it by 100, thus reducing the number of zeros after the decimal points. In all specifications we include *age13* and *age17* dummies. The symbols ***, **, * indicate that the coefficients are statistically significant at the 1, 5, and 10 percent level respectively.

In column (2) we add the vector of child's characteristics, explained in section III above. Among these controls, the child's gender, whether they were breastfed, whether they were born from a twin birth, and the number of siblings they have all have a statistically significant effect on the child's emotional problems. On average, boys tend to have emotional problems that are 0.5 points lower than girls; for those who were breastfed, emotional problems decrease on average by 0.26 points; twins exhibit greater emotional problems, with an increase by on average 0.29; children with siblings have lower emotional problems compared to those who are only children, with an average decrease of 0.08 point for each additional sibling.

The third specification enriches our analysis by including household's characteristics. The findings indicate that children residing in rural areas, as well as those in higher income households, exhibit lower levels of emotional distress, with reductions of 0.06 and 0.18, respectively, compared to children living in urban areas and living in families who experience poorer economic conditions.

Finally, in column (4) we also take into account primary caregiver's characteristics. It is important to note that neither being male nor not being the biological parent has any discernible impact on the emotional well-being of the children. Parent's level of education is statistically significant as expected: on average, for each additional year of education completed by the parent, the child's emotional problems decrease by 0.11. The impact of age is decreasing, reaching its minimum at 77. It is worth mentioning that this value lies outside the observed range of primary caregiver ages in our sample. Children whose primary caregiver is employed have lower emotional problems compared to those whose parent is not part of the labour market, with a difference of 0.19. Conversely, there are no significant differences between children of unemployed parents and those whose parents are outside of the labour force.

In Table 4 we estimate the impact of experiencing parental divorce on the likelihood of obtaining an emotional problem score higher than the normal category (thereafter, we will refer to it as the probability of having severe emotional problems) using a Linear Probability Model. Similar to our previous estimations, we first include only parental divorce as explanatory variable. Thereafter, we include child', household', and primary caregiver's characteristics respectively.

To prevent overwhelming the reader, we choose not to display the coefficients of the control variables in Table 4 below.

In all specifications, we find that the occurrence of parental divorce between the ages of 9-13 does not have an impact on the child's probability of experiencing severe emotional problems at the age of 13. In columns (1) and (2), there is a weakly significant effect observed at the age of 17. Nevertheless, when additional explanatory variables are introduced, this effect becomes statistically insignificant.

Table 4. Linear Probability Model. Effects of Parental Divorce on Child's Likelihood of Having Severe Emotional Problems

	(1)	(2)	(3)	(4)
Disrupted * age13	-.021	-.013	-.002	-.013
	(.052)	(.052)	(.053)	(.053)
Disrupted * age17	-.091*	-.084*	-.073	-.08
	(.049)	(.049)	(.049)	(.05)
Age 13	-.026***	-.026***	-.048***	-.008
	(.007)	(.007)	(.007)	(.018)
Age 17	.021***	.02***	-.01	.045*
	(.008)	(.008)	(.009)	(.025)
Disrupted	.144***	.145***	.134***	.129***
	(.042)	(.042)	(.042)	(.042)
Observations	13473	13332	13331	13331
R-squared	.005	.02	.025	.026
Child controls	NO	YES	YES	YES
Household controls	NO	NO	YES	YES
Primary caregiver controls	NO	NO	NO	YES

Notes: Linear Probability Model. Sample: GUI dataset, children aged 9, 13, and 17 years old. The dependent variable is *Severe Emotional Problems,* which takes value of 1 if the child has an Emotional Problem Score higher than the normal class, and 0 otherwise. Control variables are the same used in Table 3, where *Child controls* include gender, birthweight, been breastfed, not Irish, twin birth, number of siblings; *Household controls* include predicted income, and rural area; *Primary caregiver controls* include gender, not being biological parent, age, age squared, years of education, employed, unemployed. Standard errors, reported in parentheses, are corrected for heteroskedasticity, and allow for clustering at the household level. In all specifications we include *age13* and *age17* dummies. The symbols ***, **, * indicate that the coefficients are statistically significant at the 1, 5, and 10 percent level respectively.

We now turn our attention to the health-related outcomes. Results presented in Table 5 and Table 6 show the effects of experiencing parental divorce on child's physical health, as measured through the child's likelihood of being overweight/obese and their likelihood of performing at least 20 minutes of physical activity.

As before, we estimate four distinct specifications, starting with one that does not include control variables, and subsequently adding characteristics relating to the child, household, and primary caregiver.

Results in Table 5 make it clear that the occurrence of parental divorce during the ages 9-13 does not exert any effects on the likelihood of a child being overweight. None of the interaction terms is statistically significant different from zero.

Likewise, any statistically significant effect of parental divorce is found on child's level of physical activity, as shown in Table 6.

Table 5. Linear Probability Model. Effects of Parental Divorce on Child's Overweight Risk

	(1)	(2)	(3)	(4)
Disrupted * age13	-.062	-.067	-.057	-.073
	(.068)	(.068)	(.068)	(.068)
Disrupted * age17	-.074	-.078	-.066	-.079
	(.078)	(.078)	(.078)	(.079)
Age 13	.007	.01	-.013	.041
	(.012)	(.012)	(.012)	(.026)
Age 17	.067***	.073***	.041***	.114***
	(.012)	(.012)	(.013)	(.035)
Disrupted	.101	.103	.09	.087
	(.07)	(.07)	(.07)	(.071)
Observations	8787	8693	8692	8692
R-squared	.006	.013	.02	.021
Child controls	NO	YES	YES	YES
Household controls	NO	NO	YES	YES
Primary caregiver controls	NO	NO	NO	YES

Notes: Linear Probability Model. Sample: GUI dataset, children aged 9, 13, and 17 years old. The dependent variable is *Overweight/Obese,* which takes value of 1 if the child is in the overweight/obese category, and 0 if s/he is in the normal weight category. Control variables are the same used in Table 3, where *Child controls* include gender, birthweight, been breastfed, not Irish, twin birth, number of siblings; *Household controls* include predicted income, and rural area; *Primary caregiver controls* include gender, not being biological parent, age, age squared, years of education, employed, unemployed. Standard errors, reported in parentheses, are corrected for heteroskedasticity, and allow for clustering at the household level. In all specifications we include *age13* and *age17* dummies. The symbols ***, **, * indicate that the coefficients are statistically significant at the 1, 5, and 10 percent level respectively.

Table 6. Linear Probability Model. Effects of Parental Divorce on Child's Physical Activity

	(1)	(2)	(3)	(4)
Disrupted * age13	-.009	-.01	-.014	-.012
	(.028)	(.029)	(.029)	(.029)
Disrupted * age17	.051	.05	.046	.048
	(.032)	(.032)	(.032)	(.032)
Age 13	-.014***	-.014***	-.005	-.014
	(.003)	(.003)	(.004)	(.012)
Age 17	-.098***	-.098***	-.086***	-.1***
	(.005)	(.005)	(.006)	(.017)
Disrupted	-.02	-.019	-.015	-.014
	(.017)	(.017)	(.017)	(.017)
Observations	13428	13287	13286	13286
R-squared	.034	.044	.047	.047
Child controls	NO	YES	YES	YES
Household controls	NO	NO	YES	YES
Primary caregiver controls	NO	NO	NO	YES

Notes: Linear Probability Model. Sample: GUI dataset, children aged 9, 13, and 17 years old. The dependent variable *Physical Activity* takes value of 1 if the child exercised at least 20 minutes, and 0 otherwise. Control variables are the same used in Table 3: *Child controls* include gender, birthweight, been breastfed, not Irish, twin birth, number of siblings; *Household controls* include predicted income, and rural area; *Primary caregiver controls* include gender, not being biological parent, age, age squared, years of education, employed, unemployed. Standard errors, reported in parentheses, are corrected for heteroskedasticity, and allow for clustering at the household level. In all specifications we include *age13* and *age17* dummies. The symbols ***, **, * indicate that the coefficients are statistically significant at the 1, 5, and 10 percent level.

Table 7 shows the impact of experiencing parental divorce between 9 and 13 years of age on the child's likelihood of visiting the dentist frequently, that is visit it at least once a year.

Children who experience parental divorce and children who do not exhibit a very similar probability of visiting the dentist: the occurrence of parental divorce does not exert any impact on this likelihood. The effects when the child is aged 13 are never statistically significant in any of the specifications (row 1). Although the effect is slightly statistically significant at the ages of 17 (row 2, column 1), the interaction terms turn statistically insignificant as many explanatory variables are controlled for (row 2, columns 2-4).

Table 7. Linear Probability Model. Effects of Parental Divorce on Child's Dental Care

	(1)	(2)	(3)	(4)
Disrupted * age13	-.03	-.031	-.041	-.021
	(.042)	(.043)	(.043)	(.044)
Disrupted * age17	-.084*	-.08	-.09	-.077
	(.049)	(.049)	(.049)	(.049)
Age 13	.007	.007	.028***	-.035**
	(.006)	(.006)	(.007)	(.018)
Age 17	-.159***	-.158***	-.128***	-.217***
	(.008)	(.008)	(.009)	(.026)
Disrupted	-.027	-.022	-.01	-.004
	(.031)	(.031)	(.031)	(.032)
Observations	13473	13332	13331	13331
R-squared	.044	.05	.056	.059
Child controls	NO	YES	YES	YES
Household controls	NO	NO	YES	YES
Primary caregiver controls	NO	NO	NO	YES

Notes: Linear Probability Model. Sample: GUI dataset, children aged 9, 13, and 17 years old. The dependent variable is *Dentist Frequency,* which takes value of 1 if the child visits at least once a year the dentist, and 0 otherwise. Control variables are the same used in Table 3, where *Child controls* include gender, birthweight, been breastfed, not Irish, twin birth, number of siblings; *Household controls* include predicted income, and rural area; *Primary caregiver controls* include gender, not being biological parent, age, age squared, years of education, employed, unemployed. Standard errors, reported in parentheses, are corrected for heteroskedasticity, and allow for clustering at the household level. In all specifications we include *age13* and *age17* dummies. The symbols ***, **, * indicate that the coefficients are statistically significant at the 1, 5, and 10 percent level respectively.

VI. A "late" Parental Divorce and Young Adult's Outcomes: OLS estimates

In this section we shift our focus to the group of children who experience the breakup of their parents when they are older (specifically between the ages of 13 and 17) in comparison to children in the previous section (whose parents divorced when they were between the ages of 9 and 13). Thereafter, we will refer to this group as adolescent. In line with the strand of the literature documenting more detrimental effects if the event occurs during childhood age, our findings from the previous section provide evidence of an insignificant effect of divorce on children aged between 9 and 13.

The aim of this part of the research is to examine whether there exists heterogeneity among teenagers in the effects of divorce based on their age at the time of occurrence.

Similar to before, we evaluate the impact of parental divorce as difference in the outcomes for adolescents from intact families with those of adolescents from disrupted families. However, for this analysis only two time periods are available: wave 2 and wave 3. In wave 2, when the adolescents are 13 years old, none of the household experiences parental divorce; between these two periods, parents in some households undergo the process of divorce.

The equation estimated is the following:

$$Y_{it} = \beta_0 + \delta_1 Disrupted_i + \beta_1(Divorce_{it}) + \beta_2 \boldsymbol{C}_{it} + \beta_3 \boldsymbol{PCG_{it}} + \beta_4 \boldsymbol{H}_{it} + + \gamma_1 age_{17}\, \varepsilon_{it}$$

where Y_{it} is the outcome of interest for adolescent in household i at time t; $Disrupted_i$ takes value of 1 if parents in household i divorced when the adolescent is aged between 13 and 17, and 0 otherwise; $Divorce_{it}$ is equivalent to the interaction term $Disrupted_{it} * age_{17}$ and measures the impact of divorce when the adolescent is 17 years old; $\boldsymbol{C}_{it}$ is a vector of child's characteristics (gender, birthweight, been breasted, not Irish, twin birth, number of siblings); the vector $\boldsymbol{PCG}_{it}$ includes primary caregiver's characteristics (gender, not being the biological parent, age, age squared, years of education, employed, unemployed); $\boldsymbol{H}_{it}$ is a vector of household characteristics (predicted income and rural area); age_{17} is a dummy variable equal 1 if the adolescent is 17 years old, and 0 otherwise; ε_{it} is an error term.

Table 8 reports the estimates of the effect using an OLS estimator for four different specifications. The first column displays the results from a very basic regression in which we only include the event of interest as explanatory variable. In the following ones, we progressively include the sets of control variables, as previously done. For reasons of brevity, the coefficients on the explanatory variables included are not presented.

In column (1) we find that the effect of experiencing parental divorce between 13-17 years of age is statistically significant and positive: on average adolescents' emotional problems increase by 0.585 points (0.3 SD) relative to adolescent from intact families.

This adverse effect remains robust when control variables are included, as demonstrated in columns (2)-(4). Not only does the impact remain statistically significant, but it also increases in both magnitude and level of significance.

Table 9 presents the results obtained from a Linear Probability Model that assesses the effect of a late parental divorce on the adolescents' probability of experiencing severe emotional problems.

Table 8. OLS Estimates. Effects of Parental Divorce on Adolescent's Emotional Problems

	(1)	(2)	(3)	(4)
Disrupted * age17	.585**	.583**	.644***	.605***
	(.228)	(.228)	(.227)	(.229)
Age 17	.2***	.201***	.145***	.261***
	(.031)	(.031)	(.032)	(.057)
Disrupted	.583***	.489**	.447**	.426*
	(.218)	(.218)	(.215)	(.217)
Constant	1.632***	2.551***	3.639***	4.898***
	(.028)	(.171)	(.22)	(1.489)
Observations	8743	8654	8653	8653
R-squared	.009	.046	.056	.057
Child controls	NO	YES	YES	YES
Household controls	NO	NO	YES	YES
Primary caregiver controls	NO	NO	NO	YES

Notes: OLS estimates. Sample: GUI dataset, children aged 13 and 17. The dependent variable is *Emotional Problems Score. Child controls* include gender, birthweight, been breastfed, not Irish, twin birth, number of siblings; *Household controls* include predicted income, rural area; *Primary caregiver controls* include gender, not being biological parent, age, age squared, years of education, employed, unemployed. Standard errors, reported in parentheses, are corrected for heteroskedasticity, and allow for clustering at the household level. In all specifications we include *age17* dummy. The symbols ***, **, * indicate that the coefficients are statistically significant at the 1, 5, and 10 percent level respectively.

In all the specifications, the effect is statistically significant at the 1 percent level and remarkably strong: those who experience parental divorce between the ages of 13 and 17 display, on average, a 15-percentage point higher likelihood of experiencing severe emotional problems compared to adolescents whose parents remain married.

Table 9. Linear Probability Model. Effects of Parental Divorce on Adolescent's Likelihood of Having Severe Emotional Problems

	(1)	(2)	(3)	(4)
Disrupted * age17	.15***	.15***	.159***	.153***
	(.049)	(.049)	(.049)	(.049)
Disrupted	.056	.043	.036	.038
	(.039)	(.039)	(.039)	(.039)
Age 17	.047***	.047***	.039***	.043***
	(.007)	(.007)	(.007)	(.012)
Constant	.159***	.264***	.425***	.391
	(.006)	(.031)	(.041)	(.278)
Observations	8950	8860	8859	8859
R-squared	.008	.03	.035	.035
Child controls	NO	YES	YES	YES
Household controls	NO	NO	YES	YES
Primary caregiver controls	NO	NO	NO	YES

Notes: Linear Probability Model. Sample: GUI dataset, children aged 13 and 17. The dependent variable, *Severe Emotional Problems,* takes value of 1 if the child has an Emotional Problem Score higher than the normal class, and 0 otherwise. *Child controls* include gender, birthweight, been breastfed, not Irish, twin birth, number of siblings; *Household controls* include predicted income, rural area; *Primary caregiver controls* include gender, not being biological parent, age, age squared, years of education, employed, unemployed. Standard errors, reported in parentheses, are corrected for heteroskedasticity, and allow for clustering at the household level. In all specifications we include *age17* dummy. The symbols ***, **, * indicate that the coefficients are statistically significant at the 1, 5, and 10 percent level respectively.

Parental divorce occurring at a later stage in life has a detrimental effect on the physical health-related outcomes, as demonstrated in Table 10 and Table 11 below.

The impact of parental divorce on adolescent's likelihood of being overweight shown in Table 10 stands out particularly. Adolescents from disrupted families initially exhibit a lower overweight risk prior to the occurrence of divorce, and this difference is statistically significant. Thereafter, parental divorce leads to a sharp increase in their risk of being overweight by 16 percentage points relative to adolescents from intact families. These effects are statistically significant at the 1 percent level and remain robust even after controlling for several explanatory variables.

Table 10. Linear Probability Model. Effects of Parental Divorce on Adolescent's Overweight Risk

	(1)	(2)	(3)	(4)
Disrupted * age17	.16***	.161***	.176***	.162***
	(.048)	(.048)	(.048)	(.048)
Age 17	.06***	.064***	.053***	.085***
	(.008)	(.008)	(.008)	(.014)
Disrupted	-.075**	-.081**	-.091***	-.096***
	(.035)	(.036)	(.035)	(.036)
Constant	.192***	.178***	.396***	1.132***
	(.007)	(.041)	(.052)	(.358)
Observations	7261	7186	7185	7185
R-squared	.007	.015	.026	.029
Child controls	NO	YES	YES	YES
Household controls	NO	NO	YES	YES
Primary caregiver controls	NO	NO	NO	YES

Notes: Linear Probability Model. Sample: GUI dataset, children aged 13 and 17 years old. The dependent variable is *Overweight/Obese,* which takes value of 1 if the child is in the overweight/obese category, and 0 if s/he is in the normal weight category. *Child controls* include gender, birthweight, been breastfed, not Irish, twin birth, number of siblings; *Household controls* include predicted income, and rural area; *Primary caregiver controls* include gender, not being biological parent, age, age squared, years of education, employed, unemployed. Standard errors, reported in parentheses, are corrected for heteroskedasticity, and allow for clustering at the household level. In all specifications we include *age17* dummy. The symbols ***, **, * indicate that the coefficients are statistically significant at the 1, 5, and 10 percent level respectively.

The likelihood of adolescents engaging in physical activity is adversely affected by parental divorce, with the impact being highly statistically significant at the 1 percent level. Experiencing parental divorce between the ages of 13 and 17 leads to an average reduction of 14 percentage points in the probability of exercising at least 20 minutes, compared to adolescents from intact families. It is important to note that there are no preexisting differences between these two groups of adolescents.

Table 11. Linear Probability Model. Effects of Parental Divorce on Adolescent 's Physical Activity

	(1)	(2)	(3)	(4)
Disrupted * age17	-.139***	-.139***	-.144***	-.142***
	(.045)	(.045)	(.045)	(.046)
Age 17	-.084***	-.085***	-.079***	-.09***
	(.005)	(.005)	(.005)	(.009)
Disrupted	-.011	-.004	0	.003
	(.02)	(.019)	(.019)	(.019)
Constant	.966***	.934***	.838***	.592***
	(.003)	(.018)	(.026)	(.19)
Observations	8907	8817	8816	8816
R-squared	.031	.044	.048	.049
Child controls	NO	YES	YES	YES
Household controls	NO	NO	YES	YES
Primary caregiver controls	NO	NO	NO	YES

Notes: Linear Probability Model. Sample: GUI dataset, children aged 13 and 17. The dependent variable, *Physical Activity,* takes value of 1 if the child exercised at least 20 minutes, and 0 otherwise. *Child controls* include gender, birthweight, been breastfed, not Irish, twin birth, number of siblings; *Household controls* include predicted income, rural area; *Primary caregiver controls* include gender, not being biological parent, age, age squared, years of education, employed, unemployed. Standard errors, reported in parentheses, are corrected for heteroskedasticity, and allow for clustering at the household level. In all specifications we include *age17* dummy. The symbols ***, **, * indicate that the coefficients are statistically significant at the 1, 5, and 10 percent level respectively.

Table 12 shows the impact of parental divorce on the child's dental care. Despite the absence of preexisting differences, parental divorce significantly affects adolescents in the treated groups, resulting in an average decrease of 11 percentage points in their likelihood of visiting the dentist at least once a year compared to adolescents in intact families.

Table 12. Linear Probability Model. Effects of Parental Divorce on Child's Dental Care

	(1)	(2)	(3)	(4)
Disrupted * age17	-.111**	-.112**	-.124**	-.107**
	(.05)	(.05)	(.05)	(.05)
Age 17	-.165***	-.165***	-.154***	-.187***
	(.008)	(.008)	(.008)	(.012)
Disrupted	-.036	-.036	-.028	-.022
	(.033)	(.033)	(.033)	(.033)
Constant	.893***	.878***	.667***	.198
	(.005)	(.029)	(.04)	(.278)
Observations	8950	8860	8859	8859
R-squared	.048	.055	.063	.067
Child controls	NO	YES	YES	YES
Household controls	NO	NO	YES	YES
Primary caregiver controls	NO	NO	NO	YES

Notes: Linear Probability Model. Sample: GUI dataset, children aged 13 and 17. The dependent variable, *Dentist Frequency,* takes value of 1 if the child visits at least once a year the dentist, and 0 otherwise. *Child controls* include gender, birthweight, been breastfed, not Irish, twin birth, number of siblings; *Household controls* include predicted income, rural area; *Primary caregiver controls* include gender, not being biological parent, age, age squared, years of education, employed, unemployed. Standard errors, reported in parentheses, are corrected for heteroskedasticity, and allow for clustering at the household level. In all specifications we include *age17* dummy. The symbols ***, **, * indicate that the coefficients are statistically significant at the 1, 5, and 10 percent level respectively.

VII. Fixed-effects models

A second main assumption behind the DiD design is the *exogeneity assumption*, according to which the assignment of the treatment, parental divorce, is exogenous with respect to the outcomes. A possible concern with the results in the previous sections is that OLS might be biased if *parental divorce* is correlated with a number of unobservable household factors that can directly affect child's outcomes as well (Amato, 2010).

Therefore, to further validate the previous results and to support the plausibility of this assumption, we add in our model household-fixed effects, in addition to the time-fixed effects already controlled for. Household fixed-effects allow to account for unobservable time-invariant family characteristics which are constant within the family but do change across units. We do not include control variables that are time invariant.

The idea is that, once these unobserved sources of heterogeneity that are time invariant are controlled for, the difference in the outcomes for the two groups of children should represent the causal effect of parental divorce.

The estimated model is the following:

$$Y_{it} = \delta_i + \lambda_t + \beta_0 \boldsymbol{X}_{it} + \beta_1 Divorce_{it} + \varepsilon_{it}$$

where Y_{it} is the outcome of interest for Study Child in household *i; $Divorce_{it}$* is a dummy variable taking the value of 1 if child in household i is exposed to the treatment (parental divorce between two consecutive surveys) in time *t*, and the value of 0 otherwise, i.e. parents stay married; δ_i is household-level fixed effect; λ_t is the time-level fixed effect; $\boldsymbol{X}_{it}$ is a vector of time-variant household-level covariates which includes predicted income and the employment situation of the primary caregiver; ε_{it} is an i.i.d. normally distributed stochastic error term.

In Table 13 and Table 14 we report the effects of experiencing parental divorce between ages 9-13 and between ages 13-17, respectively, on children's outcomes measured at the age of 17.

The fixed-effect models reveal no detrimental impact of parental divorce on children's outcomes when it occurs between the ages of 9 and 13 as shown in Table 13. It is important to emphasize that, in contrast to the previous findings for the first group of treated children, the coefficient of *Divorce* represents the overall average impact of divorce and does not distinguish between the effects measured at the age of 13 and those measured at the age of 17 as done in the estimations above.

Table 13. Fixed-Effects Estimates. Effects of Parental Divorce on Child's Outcomes

	Emotional problems score	Severe emotional problems	Overweight Risk	Physical Activity	Dentist Frequency
Parental divorce	-.132	-.056	-.074	.02	-.058
	(.197)	(.044)	(.075)	(.025)	(.039)
Age 13	-.297***	-.029***	.175***	-.014**	.005
	(.043)	(.01)	(.018)	(.006)	(.01)
Age 17	-.133**	.014	.305***	-.094***	-.163***
	(.057)	(.013)	(.022)	(.009)	(.014)
Constant	2.023***	.203***	-.001	.973***	.918***
	(.242)	(.054)	(.083)	(.039)	(.059)
Observations	13263	13472	8786	13427	13472
R-squared	.01	.005	.119	.053	.073
Controls	YES	YES	YES	YES	YES
Household FE	YES	YES	YES	YES	YES

Notes: The Table reports Fixed-Effects estimates. Sample: GUI dataset, children aged 9, 13, and 17. The dependent variables are the following: *Emotional Problems Score, Severe Emotional Problems,* which takes the value of 1 if the child has emotional problems higher than the normal class, and 0 otherwise; *Overweight*, equals 1 if the child is overweight/obese, and 0 otherwise; *Physical Activity*, equals 1 if the child exercised at least 20 minutes, and 0 otherwise; *Dentist Frequency*, that equals 1 if the child visits the dentist at least once a year, and 0 otherwise. The control variables included in each model are *Predicted income*, and *Primary caregiver's employment status.* Standard errors, reported in parentheses, are corrected for heteroskedasticity, and allow for clustering at the household level. Each estimate includes *age13* and *age17* dummies. The symbols ***, **, * indicate that the coefficients are statistically significant at the 1, 5, and 10 percent level respectively.

As regards to adolescents aged 13-17, the fixed-effects models confirm the negative impact of parental divorce on all the outcomes considered, as illustrated in Table 14.

Table 14. Fixed-Effects Estimates. Effects of Late Parental Divorce on Adolescent's Outcomes

	Emotional problems score	Severe emotional problems	Overweight Risk	Physical Activity	Dentist Frequency
Parental divorce	.632***	.153***	.157***	-.141***	-.11**
	(.228)	(.049)	(.054)	(.046)	(.05)
Age 17	.13***	.041***	.136***	-.082***	-.164***
	(.037)	(.009)	(.01)	(.006)	(.009)
Constant	2.269***	.239***	.18**	.955***	.947***
	(.298)	(.067)	(.079)	(.052)	(.073)
Observations	8742	8949	7260	8906	8949
R-squared	.013	.015	.089	.06	.104
Controls	YES	YES	YES	YES	YES
Household FE	YES	YES	YES	YES	YES

Notes: The Table reports Fixed-Effects estimates. Sample: GUI dataset, children aged 13, and 17. The dependent variables are the following: *Emotional Problems Score, Severe Emotional Problems,* which takes the value of 1 if the child has emotional problems higher than the normal class, and 0 otherwise; *Overweight*, equals 1 if the child is overweight/obese, and 0 otherwise; *Physical Activity*, equals 1 if the child exercised at least 20 minutes, and 0 otherwise; *Dentist Frequency*, that equals 1 if the child visits the dentist at least once a year, and 0 otherwise. The control variables included in each model are *Predicted income*, and *Primary caregiver's employment status.* Standard errors, reported in parentheses, are corrected for heteroskedasticity, and allow for clustering at the household level. Each estimate includes *age17* dummy. The symbols ***, **, * indicate that the coefficients are statistically significant at the 1, 5, and 10 percent level respectively.

The coefficient of parental divorce is strongly statistically significant in each model and preserves the same direction and almost the same magnitude of the impacts indicated by the OLS specifications. Adolescents who experience parental divorce between the ages of 13 and 17 exhibit higher emotional problems by on average 0.63 points and are approximately 15 percentage points more likely to experience such problems relative to adolescents who grow up with stably married parents. Furthermore, their risk of being overweight is almost 16 percentage points higher, their likelihood of engaging in regular physical activity decreases by on average 14 percentage points, and their likelihood of visiting the dentist at least once a year decreases by on average 11 percentage points.

VIII. Robustness check: Breakup of Cohabiting Parents and Children's Outcomes

In the main analysis, we use variations in the legal marital status of the primary caregiver to identify child's exposure to the event of interest, parental divorce. Households wherein the primary caregiver asserts oneself as either never married or a widow are not considered. Nevertheless, it should be noted that a substantial number of households within our dataset are formed by cohabiting parents who are not bound by marriage, as explained earlier.

In this section, we aim at estimating the impact of a breakup of cohabiting parents, regardless of their marital status, on children's outcomes. To determine the event of interest, namely the *breakup of cohabiting parents*, we use the variable *household type*, which categorizes families as either single-parent families or couples. Parental breakup is identified if we observe a variation in the composition of household from being formed by a couple to being a single-parent family across surveys.

We classify families into four mutually exclusive categories: a) *intact families*, households whose type remains as couples for the entire duration of the study; b) *already single-parent families,* households whose type has been a single-parent family since the first survey; c) *a first group of non-intact families,* in which parental breakup occurs between age 9 and 13 of children; and d) *a second group of non-intact families* where parents cohabit both in the first and second surveys, and then dissolve their relationship when the child is between 13 and 17 years old. Note that, once we observe parental breakup, we do not consider further potential transitions, such parents who reunite or form new partnership.

Similarly to the main analysis, we do not consider in our sample households in which the primary caregiver changes across surveys to guarantee consistency and comparability in the response provided. We also exclude households where primary caregiver is widow. This cleaning process results in a panel dataset of 15,768 observations, with 5,256 households taking part at the survey in each wave.

Table 15 above shows the descriptive statistics for the variables used in the regression models. Slight differences in the mean values are observed due to variations in sample size, but they are not remarkable.

Table 15. Descriptive Statistics

	Mean	Std.Dev.	Min	Max	Obs
Outcomes:					
Emotional Problems	1.886	1.975	0	10	15542
Borderline/Abnormal emotional categ.	.198	.398	0	1	15768
Overweight/Obese	.229	.42	0	1	10400
Exercise	.937	.243	0	1	15712
Dentist frequency	.824	.381	0	1	15768
Child's Characteristics:					
Boy	.485	.5	0	1	15768
Birthweight	3.531	.6	1.7	6.1	15597
Been breastfed	.553	.497	0	1	15645
Twin	.029	.169	0	1	15768
Height in cms	156.832	16.39	121	192	15406
Weight in kgs	51.669	17.505	21	110	15389
Number of siblings	1.857	1.037	0	5	15768
Only Child	.079	.269	0	1	15768
Not Irish	.04	.196	0	1	15768
Primary Caregiver's Characteristics:					
Female pcg	.994	.078	0	1	15768
Pcg's age	44.317	5.943	26	58	15768
Non biological parent	.004	.059	0	1	15768
Pcg's years of education	14.662	2.241	8	18	15767
Employed	.646	.478	0	1	15768
Unemployed	.018	.133	0	1	15768
Currently outside labour force	.336	.472	0	1	15768
Pcg is roman catholic	.934	.249	0	1	14814
Other religion	.066	.249	0	1	14814
Household's Characteristics:					
Predicted real income	17457.909	4127.055	362.957	30165.523	15767
Rural area	.558	.497	0	1	15768
Household Type:					
Intact families	.872	.334	0	1	15768
Parental breakup before 9	.071	.257	0	1	15768
Parental breakup between 9-13	.028	.166	0	1	15768
Parental breakup between 13-17	.028	.166	0	1	15768

Notes: Dataset GUI Child Cohort, wave 1 – wave 3

The differences in the households type are worth to be noted: a smaller percentage of families in the analysis is now represented by stably cohabiting parents, on average 87%; 7% of them went through parental breakup before child turned 9; for 3 % of the households parental breakup takes place when the child is between 9 and 13 years old; the remaining 3% experiences parental breakup when the child is between the ages of 13 and 17.

As before, we implement a DiD strategy to assess the causal impact of the breakup of cohabiting parents on child's outcomes. We achieve this by subsampling our population based on the timing of the event and by considering children who grow up in households stably formed by coupled parents as the control group.

Table 16 shows the results from OLS estimates for the cohort of children whose parents break up when the children are between 9 and 13 years old.

Table 16. OLS Estimates. Effects of Parental Breakup between 9-13 years on Child's Outcomes

	Emotional problems score	Severe emotional problems	Overweight Risk	Physical Activity	Dentist Frequency
Parental breakup * age13	.048	-.002	-.017	-.022	0
	(.186)	(.047)	(.067)	(.027)	(.04)
Parental breakup * age17	-.282	-.07	-.084	-.017	-.093*
	(.217)	(.045)	(0.071)	(.034)	(.071)
Parental breakup	.462***	.083**	.081	-.008	-.008
	(.175)	(.039)	(.063)	(.015)	(.03)
Observations	13841	14051	9175	14000	
R-squared	.048	.025	.022	.05	
Controls	YES	YES	YES	YES	YES
Household FE	YES	YES	YES	YES	YES

Notes: The Table reports OLS estimates. Sample: GUI dataset, children aged 9, 13, and 17. The dependent variables are the following: *Emotional Problems Score; Severe Emotional Problems,* which takes value of 1 if the child has emotional problems higher than the normal score, and 0 otherwise; *Overweight*, equals 1 if the child is overweight/obese, and 0 for normal weight; *Physical Activity*, equals 1 if the child exercised at least 20 minutes, and 0 otherwise; *Dentist Frequency*, that equals 1 if the child visits at least once a year the dentist, and 0 otherwise. *Child controls* include gender, birthweight, been breastfed, not Irish, twin birth, number of siblings; *Household controls* include predicted income, and rural area; *Primary caregiver controls* include gender, not being biological parent, age, age squared, years of education, employed, unemployed. Standard errors, reported in parentheses, are corrected for heteroskedasticity, and allow for clustering at the household level. Each estimate includes *age13* and *age17* dummies. The symbols ***, **, * indicate that the coefficients are statistically significant at the 1, 5, and 10 percent level respectively.

Although we do not present the coefficients for the control variables, the specifications presented are estimated using the vector of characteristics pertaining to the child, household, and primary caregiver as explanatory variables.

The results support what shown in the main section, as they reveal no statistically significant impact of parental breakup on child's emotional and health related problems. This holds true both when the child is 13 years old and when they reach the age of 17. As before, pre-existing significant differences exist only in terms of emotional problems between the two groups of children. Additionally, there is a weakly statistically significance decrease in the likelihood of frequently visiting the dentist when the child is 17 years old.

Table 17 reports estimations from fixed-effect models. The results confirm a lack of significance in the impact of parental breakup experienced between the ages of 9 and 13 on children's outcomes. All the specifications include as time-variant explanatory variables predicted income and primary caregivers' employment status.

Table 17. Fixed-Effects Estimates. Effects of Parental Breakup between 9-13 on Child's Outcomes

	Emotional problems score	Severe emotional problems	Overweight Risk	Physical Activity	Dentist Frequency
Parental breakup	-.062	-.041	-.016	-.017	-.053
	(.175)	(.039)	(.07)	(.024)	(.037)
Age 13	.293***	-.028***	.169***	-.012**	.004
	(.041)	(.009)	(.018)	(.006)	(.01)
Age 17	-.136***	.014	.298***	-.096***	-.165***
	(.055)	(.012)	(.022)	(.09)	(.014)
Constant	2.168***	.213***	.002	.960***	.908***
	(.229)	(.052)	(.081)	(.038)	(.057)
Observations	13984	14195	9272	14144	14195
R-squared	.009	.004	.115	.056	.075
Controls	YES	YES	YES	YES	YES
Household FE	YES	YES	YES	YES	YES

Notes: The Table reports Fixed-effects estimates. Sample: GUI dataset, children aged 9, 13, and 17. The dependent variables are in the following order: *Emotional Problems Score*; *Severe Emotional Problems,* which takes value of 1 if the child has emotional problem higher than the normal score, and 0 otherwise; *Dentist Frequency*, equals 1 if child visits at least the dentist once a year, and 0 otherwise; *Overweight/Obese*, that takes value of 1 if the child is overweight/obese, and 0 if normal weight; *Exercise*, which takes value of 1 if the child exercised at least once, and 0 otherwise. The control variables included in each model are *Predicted income*, and *primary caregiver's employment status.* Standard errors, reported in parentheses, are corrected for heteroskedasticity, and allow for clustering at the household level. In the estimates we include *age13* and *age17* dummies. The symbols ***, **, * indicate that the coefficients are statistically significant at the 1, 5, and 10 percent level respectively.

When we shift our attention to the sample of adolescents who experience parental breakup between the ages of 13 and 17, we do find a significant effect on all the outcomes considered.

Table 18 and 19 below show the results from OLS and FE estimations respectively.

Regarding OLS estimations, presented in Table 18 below, adolescents who experience parental breakup have a higher prevalence of emotional problems compared to those who do not on average by 0.334 points. These adolescents face a greater probability of having such problems by 11 percentage points. Adolescents exposed to parental breakup face a greater risk of being overweight on average by 9 percentage points, and a decrease in their likelihood of exercising by 9 percentage points, relative to their peers whose parents remain together. Parental breakup also leads to a decrease in the adolescent's likelihood of visiting the dentist at least once a year by on average 11 percentage points.

Fixed effects models produce very similar results, although the effects are slightly lower in magnitude. For emotional problems score and the risk of being overweight, the effects present weak statistical significance, with p-values of 0.101 and 0.11 respectively.

Table 18. OLS Estimates. Effects of a Late Parental Breakup on Adolescents' Outcomes

	Emotional problems	Severe emotional problems	Overweight Risk	Physical Activity	Dentist frequency
Parental breakup*age17	.334*	.112**	.09**	-.089**	-.112**
	(.2)	(.044)	(.046)	(.038)	(.045)
Parental breakup	.344*	.019	-.009	-.002	.002
	(.177)	(.033)	(.037)	(.017)	(.027)
Observations	9165	9373	7608	9325	9373
R-squared	.06	.035	.03	.049	.067
Child controls	YES	YES	YES	YES	YES
HH controls	YES	YES	YES	YES	YES
PCG controls	YES	YES	YES	YES	YES

Notes: Table reports OLS estimates. Sample: GUI dataset, children aged 13, and 17. Dependent variables are the following: *Emotional Problems Score*; *Severe Emotional Problems,* equals 1 if the child has Emotional Problem higher than the normal class, and 0 otherwise ; *Overweight/Obese*, equals 1 if the child is in the overweight/obese category, and 0 if s/he is in the normal weight category; *Physical Activity*, equals 1 if the child exercised at least 20 minutes, and 0 otherwise; *Dentist Frequency*, equals 1 if the child visits at least once a year the dentist, and 0 otherwise. *Child controls* include gender, birthweight, been breastfed, not Irish, twin birth, number of siblings; *Household controls* include predicted income, rural area; *Primary caregiver controls* include gender, not being biological parent, age, age squared, years of education, employed, unemployed. Standard errors, reported in parentheses, are corrected for heteroskedasticity, and allow for clustering at the household level. In the estimates we include *age17* dummy. The symbols ***, **, * indicate that the coefficients are statistically significant at the 1, 5, and 10 percent level respectively.

Table 19 below reports Fixed-Effects estimates.

Table 19. FE Estimates. Effects of a Late Parental Breakup on Adolescents' Outcomes

	Emotional problems	Severe emotional problems	Dentist frequency	Overweight Risk	Physical Activity
Parental breakup * age17	.327*	.109**	-.109**	.077	-.091**
	(.199)	(.044)	(.044)	(.048)	(.038)
Age 17	.176***	.046***	-.166***	.142***	-.086***
	(.030)	(.007)	(.007)	(.008)	(.005)
Constant	1.69***	.165***	.889***	.145***	.965***
	(.015)	(.003)	(.004)	(.005)	(.003)
Observations	9255	9464	9464	7684	9416
R-squared	.009	.012	.103	.087	.06
Controls	YES	YES	YES	YES	YES
Household FE	YES	YES	YES	YES	YES

Notes: Table reports Fixed-effects estimates. Sample: GUI dataset, children aged 13, and 17 years old. Dependent variables are: *Emotional Problems Score*; *Severe Emotional Problems, taking* value of 1 if child has emotional problems higher than normal class, and 0 otherwise; *Overweight/Obese*, equals 1 if child is in the overweight/obese category, and 0 if in the normal weight category; *Physical Activity*, equals 1 if the child exercised at least 20 minutes, and 0 otherwise; *Dentist Frequency*, equals 1 if the child visits at least once a year the dentist, and 0 otherwise. Control variables included in each model are p*redicted income*, and *primary caregiver's employment status.* Standard errors, reported in parentheses, are corrected for heteroskedasticity, and allow for clustering at the household level. In the estimates we include *age17* dummy. The symbols ***, **, * indicate that the coefficients are statistically significant at the 1, 5, and 10 percent level respectively.

IX. Heterogeneous Effects

In this section we investigate whether the impact of parental divorce on children's outcomes differs in relation to certain characteristics. Firstly, we present the results pertaining to the group of children who experience parental divorce between the ages of 9 and 13. Subsequently, we analyze the outcomes for the other group of children exposed to the event at a later age, specifically between 13 and 17 years old.

We first focus on child's gender and estimate two regressions separately for boys and girls. As shown in Table 20, we do not find any statistically significant effects for neither group at the age of 13. Once reached the age of 17, some positive effects are fund only for boys, as shown in Panel A. On average, boys who experience parental divorce have lower emotional problems by 0.89 points, and their likelihood of having severe emotional problems declines by 18 percentage points. Similarly, boys' likelihood of being physically active significantly increases by about 7 percentage points on average.

Table 20. Heterogeneity by Child's Gender

	Emotional problems	Severe emotional problems	Overweight risk	Physical activity	Dentist frequency
Panel A: Boys					
Disrupted	0.990***	0.175***	0.150	-0.0187	-0.0701
	(0.272)	(0.0593)	(0.0974)	(0.0216)	(0.0486)
Disrupted * age13	-0.0706	0.0209	-0.0158	-0.0119	-0.00237
	(0.273)	(0.0744)	(0.0991)	(0.0378)	(0.0622)
Disrupted * age17	-0.892***	-0.181***	-0.0847	0.0676**	-0.0650
	(0.274)	(0.0604)	(0.109)	(0.0273)	(0.0723)
Observations	6577	6666	4218	6641	6666
R-squared	0.0167	0.00789	0.00698	0.0136	0.0464
Panel B: Girls					
Disrupted	0.451*	0.113*	0.0363	-0.0213	0.0171
	(0.253)	(0.0593)	(0.0991)	(0.0271)	(0.0376)
Disrupted * age13	0.157	-0.0640	-0.0910	-0.00605	-0.0589
	(0.301)	(0.0728)	(0.0916)	(0.0423)	(0.0576)
Disrupted * age17	0.224	0.00435	-0.0492	0.0324	-0.102
	(0.356)	(0.0730)	(0.106)	(0.0562)	(0.0644)
Observations	6687	6807	4569	6787	6807
R-squared	0.0124	0.0106	0.00724	0.0591	0.0426

Notes: Each column reports estimates from OLS regression. GUI dataset, children aged 9, 13, and 17. The dependent variables are *Emotional Problems Score*, *Severe Emotional Problems*, *Overweight Risk*, *Physical Activity*, and *Dentist Frequency* respectively. Standard errors, reported in parentheses, are corrected for heteroskedasticity, and allow for clustering at the household level. The symbols ***, **, * indicate that the coefficients are statistically significant at the 1, 5, and 10 percent level respectively.

In Table 21 we look at the heterogeneous effects of parental divorce according to the level of education of the primary caregiver. We distinguish between households with low educated primary caregiver (Panel A), defined as those who have acquired 13 years of education or less, equivalent to a secondary level of education, and households with high educated primary caregiver (Panel B), defined as those who have obtained more than 13 years of education.

Results from Panel A indicate that parental divorce does not have any effects on the outcomes considered for children whose primary caregiver has a low level of education: none of the interaction terms is statistically significant. In contrast, a high level of education of the primary caregiver is helpful in mitigating the impact of experiencing parental divorce reducing the adolescent's emotional problems at the age of 17 by 0.95 and 17 percentage points respectively. For the other outcomes, experiencing a parental divorce between the ages of 9 and 13 does not result in any statistically significant effects.

Table 21. Heterogeneity by Primary Caregiver's Education

	Emotional problems	Severe emotional problems	Overweight risk	Physical activity	Dentist frequency
Panel A: Low educated					
Disrupted	0.521*	0.108*	0.163*	-0.0378	-0.0357
	(0.266)	(0.0564)	(0.0945)	(0.0276)	(0.0452)
Disrupted * age13	-0.122	-0.0640	-0.152	0.00397	-0.00837
	(0.274)	(0.0666)	(0.0976)	(0.0444)	(0.0614)
Disrupted * age17	0.178	-0.0142	-0.0391	0.0868*	-0.0187
	(0.350)	(0.0696)	(0.112)	(0.0469)	(0.0705)
Observations	5674	5744	3914	5727	5744
R-squared	0.00511	0.00286	0.0126	0.0426	0.0553
Panel B: High educated					
Disrupted	0.872***	0.173***	0.00139	0.00227	-0.00700
	(0.255)	(0.0629)	(0.0987)	(0.0176)	(0.0411)
Disrupted * age13	0.180	0.0157	0.0504	-0.0224	-0.0512
	(0.299)	(0.0821)	(0.0907)	(0.0329)	(0.0602)
Disrupted * age17	-0.948***	-0.174**	-0.0934	0.0147	-0.148*
	(0.318)	(0.0694)	(0.102)	(0.0403)	(0.0761)
Observations	7589	7728	4872	7700	7728
R-squared	0.00755	0.00611	0.00376	0.0272	0.0344

Notes: Each column reports estimates from OLS regression. GUI dataset, children aged 9, 13, and 17. The dependent variables are *Emotional Problems Score, Sever Emotional Problems, Overweight Risk*, *Physical Activity*, and *Dentist Frequency* respectively. The level of education is measured in years. Pcg has a low level of education if s/he obtained 13 years of education or less. Standard errors, reported in parentheses, are corrected for heteroskedasticity, and allow for clustering at the household level. The symbols ***, **, * indicate that the coefficients are statistically significant at the 1, 5, and 10 percent level respectively.

We further investigate whether the effects parental divorce vary according to the age of the primary caregiver when the child is born. We split our sample in two groups: one formed by household in which the primary caregiver's age at the time of child's birth was below the median age of 32 and those with primary caregiver's age at child's birth above the median age. Table 22 below shows the results.

Table 22. Heterogeneity by Primary Caregiver's Age at Child's Birth (below/above median)

	Emotional problems	Severe emotional problems	Overweight risk	Physical activity	Dentist frequency
Panel A: Age at child's birth ≤32					
Disrupted	0.745***	0.144***	0.126	-0.0240	-0.0177
	(0.227)	(0.0500)	(0.0870)	(0.0210)	(0.0362)
Disrupted * age13	0.0401	-0.0191	-0.0962	-0.00167	-0.0480
	(0.243)	(0.0587)	(0.0829)	(0.0339)	(0.0505)
Disrupted * age17	-0.285	-0.0655	-0.124	0.0629*	-0.0719
	(0.288)	(0.0589)	(0.0915)	(0.0376)	(0.0553)
Observations	7313	7398	4894	7382	7398
R-squared	0.00913	0.00610	0.00866	0.0363	0.0538
Panel B: Age at child's birth >32					
Disrupted	0.424	0.108	0.0452	-0.00931	-0.0421
	(0.303)	(0.0761)	(0.118)	(0.0292)	(0.0612)
Disrupted * age13	0.135	0.00301	0.0102	-0.0215	0.0209
	(0.365)	(0.111)	(0.122)	(0.0509)	(0.0782)
Disrupted * age17	-0.456	-0.132	0.0194	0.0302	-0.0768
	(0.409)	(0.0877)	(0.148)	(0.0585)	(0.101)
Observations	5951	6075	3893	6046	6075
R-squared	0.00306	0.00392	0.00342	0.0322	0.0332

Notes: Each column reports estimates from OLS regression. GUI dataset, children aged 9, 13, and 17. The dependent variables are *Emotional Problems Score, Sever Emotional Problems, Overweight Risk, Physical Activity,* and *Dentist Frequency* respectively. The median primary caregiver's age at birth is 32. Standard errors, reported in parentheses, are corrected for heteroskedasticity, and allow for clustering at the household level. The symbols ***, **, * indicate that the coefficients are statistically significant at the 1, 5, and 10 percent level respectively.

Our findings indicate that parental divorce occurring between the ages of 9 and 13 for children does not exert any significant effects, regardless of the primary caregiver's age at the time of their birth.

Another important issue is related to the children's adjustment process to divorce according to the household income. To shed light on this aspect, we estimate separate regressions for children in household with household income below and above the median (Panel A and Panel B in Table 23) respectively. Bear in mind that in our analysis we use the household income predicted in the absence of parental divorce. Surprisingly, we do not find any statistically significant effect of parental divorce for children in household with income lower than the median value. Positive effects are found for adolescents' emotional problems with a decrease of 1 point and 24 percentage points respectively (Panel B, columns (1) and (2)) statistically significant at the 5 percent level. A decrease in the likelihood of visiting the dentist frequently at the age of 17 is found for children whose parents divorced between the ages of 9 and 13 in household with income higher than the median.

Table 23. Heterogeneity by Household Income (below/above median)

	Emotional problems	Severe emotional problems	Overweight risk	Physical activity	Dentist frequency
Panel A: Household Income ≤ median					
Disrupted	0.546*	0.107*	0.143	-0.0467	-0.0104
	(0.311)	(0.0627)	(0.101)	(0.0335)	(0.0485)
Disrupted * age13	0.00763	-0.0317	-0.124	0.0129	-0.0460
	(0.316)	(0.0710)	(0.0997)	(0.0479)	(0.0625)
Disrupted * age17	0.127	0.0144	-0.0746	0.0659	-0.0271
	(0.365)	(0.0714)	(0.109)	(0.0501)	(0.0662)
Observations	6718	6825	5015	6796	6825
R-squared	0.00673	0.00353	0.00812	0.0321	0.0400
Panel B: Household Income > median					
Disrupted	0.783***	0.161***	0.0602	0.00357	-0.0299
	(0.224)	(0.0566)	(0.0977)	(0.0144)	(0.0404)
Disrupted * age13	0.227	0.0176	0.00523	-0.0270	-0.0296
	(0.324)	(0.0849)	(0.102)	(0.0322)	(0.0563)
Disrupted * age17	-1.007***	-0.236***	-0.104	0.0503	-0.217***
	(0.326)	(0.0649)	(0.111)	(0.0354)	(0.0825)
Observations	6546	6648	3772	6632	6648
R-squared	0.0106	0.00748	0.00265	0.0283	0.0360

Notes: Each column reports estimates from OLS regression. GUI dataset, children aged 9, 13, and 17. The dependent variables are *Emotional Problems Score*, *Severe Emotional Problems, Overweight Risk, Physical Activity,* and *Dentist Frequency* respectively. Standard errors, reported in parentheses, are corrected for heteroskedasticity, and allow for clustering at the household level. The symbols ***, **, * indicate that the coefficients are statistically significant at the 1, 5, and 10 percent level respectively.

Finally, in Table 24 we assess whether the effect of parental divorce changes based on the area of residence of the family. Results do not show any statistically significant effect of parental divorce, except for the likelihood of being overweight in both areas of residence.

Table 24. Heterogeneity by Area of residence

	Emotional problems	Severe emotional problems	Overweight risk	Physical activity	Dentist frequency
Panel a: Rural area					
Disrupted	0.672***	0.158***	0.259***	-0.0327	-0.0446
	(0.240)	(0.0550)	(0.100)	(0.0258)	(0.0437)
Disrupted * age13	0.278	0.0024	-0.240***	-0.0403	-0.0479
	(0.234)	(0.0677)	(0.0907)	(0.0454)	(0.0622)
Disrupted * age17	-0.257	-0.0845	-0.206*	0.0718*	-0.103
	(0.301)	(0.0607)	(0.109)	(0.0424)	(0.0657)
Observations	7614	7746	5081	7716	7746

R-squared	0.00864	0.00582	0.00796	0.0353	0.0427

Table 24. Heterogeneity by Area of residence

Panel B: Urban area					
Disrupted	0.792***	0.124*	-0.132**	-0.0009	-0.0003
	(0.295)	(0.0647)	(0.0593)	(0.0193)	(0.0415)
Disrupted * age13	-0.295	-0.0554	0.200**	0.0362*	-0.00631
	(0.362)	(0.0816)	(0.0841)	(0.0198)	(0.0517)
Disrupted * age17	-0.511	-0.101	0.117	0.0206	-0.0563
	(0.382)	(0.0818)	(0.0769)	(0.0466)	(0.0710)
Observations	5650	5727	3706	5712	5727
R-squared	0.00424	0.00336	0.00617	0.0344	0.0482

Notes: Each column reports estimates from OLS regression. GUI dataset, children aged 9, 13, and 17. The dependent variables are *Emotional Problems Score, Severe Emotional Problems, Overweight Risk, Physical Activity,* and *Dentist Frequency* respectively. Standard errors, reported in parentheses, are corrected for heteroskedasticity, and allow for clustering at the household level. The symbols ***, **, * indicate that the coefficients are statistically significant at the 1, 5, and 10 percent level respectively.

Specifically, for children exposed to parental divorce and living in rural area the risk of being overweight decreases by on average 24 percentage points at the age of 13, whereas for children of divorce who live in urban area that risk increases by almost 20 percentage points.

We now present the results for the group of children who experience the divorce of their parents when older, namely between the ages of 13 and 17.

In Table 25 we investigate whether the effects diverge between boys (Panel A) and girls (Panel B).

Table 25. Heterogeneity by Child's Gender – Late parental divorce

	Emotional problems	Severe emotional problems	Overweight risk	Physical activity	Dentist frequency
Panel A: Boys					
Disrupted	0.653	0.0590	-0.0233	0.0309***	-0.0506
	(0.402)	(0.0598)	(0.0616)	(0.00374)	(0.0567)
Disrupted * age17	0.419	0.0675	0.109	-0.152**	-0.0411
	(0.365)	(0.0696)	(0.0781)	(0.0597)	(0.0711)
Observations	4313	4401	3478	4378	4401
R-squared	0.00587	0.00168	0.00652	0.0105	0.0457
Panel B: Girls					
Disrupted	0.470*	0.0424	-0.114***	-0.0364	-0.0293
	(0.251)	(0.0515)	(0.0422)	(0.0315)	(0.0411)
Disrupted * age 17	0.573**	0.185***	0.192***	-0.113*	-0.156**
	(0.287)	(0.0654)	(0.0605)	(0.0635)	(0.0672)
Observations	4430	4549	3783	4529	4549
R-squared	0.0202	0.0170	0.00747	0.0532	0.0498

Notes: Each column reports estimates from OLS regression. GUI dataset, children aged 13, and 17. The dependent variables are *Emotional Problems Score, Severe Emotional Problems, Overweight Risk*, *Physical Activity*, and *Dentist Frequency* respectively. Standard errors, reported in parentheses, are corrected for heteroskedasticity, and allow for clustering at the household level. The symbols ***, **, * indicate that the coefficients are statistically significant at the 1, 5, and 10 percent level respectively.

Our results demonstrate that the effects of parental divorce are more detrimental for girls. While for boys experiencing parental divorce negatively affects only their likelihood of engaging in physical exercise, for girls the impact of parental divorce is always statistically significant at a level that ranges between the 1 and 5 percent. More specifically, girls whose parents divorced between the ages of 13 and 17 have higher emotional problems compared to those whose parents stay married, with a difference of 0.586 points and 16 percentage points respectively. On average, their likelihood of being overweight increases by 14 percentage points. Finally, they are less likely to perform physical activity and visit the dentist frequently, with a decrease of 18 percentage points and 14 percentage points respectively.

As regards heterogeneous effects based on the level of education of the primary caregiver, lowed and high educated, we find that children in both groups are negatively affected by parental divorce.

Table 26. Heterogeneity by Primary Caregiver's Education – Late parental divorce

	Emotional problems	Severe emotional problems	Overweight risk	Physical activity	Dentist frequency
Panel A: Low educated					
Disrupted	0.511	-0.0019	-0.104**	0.00620	-0.0944*
	(0.318)	(0.0540)	(0.0512)	(0.0253)	(0.0571)
Disrupted * age17	0.502	0.181**	0.229***	-0.161**	-0.0547
	(0.349)	(0.0714)	(0.0747)	(0.0634)	(0.0741)
Observations	3693	3764	3175	3748	3764
R-squared	0.00585	0.00527	0.00809	0.0334	0.0544
Panel B: High educated					
Disrupted	0.575*	0.0985*	-0.0625	-0.0253	0.0334
	(0.300)	(0.0568)	(0.0483)	(0.0303)	(0.0306)
Disrupted * age17	0.721**	0.127*	0.0997*	-0.115*	-0.164**
	(0.292)	(0.0670)	(0.0578)	(0.0646)	(0.0708)
Observations	5049	5185	4085	5158	5185
R-squared	0.00981	0.00945	0.00478	0.0266	0.0384

Notes: Each column reports estimates from OLS regression. GUI dataset, children aged 13, and 17. The dependent variables are *Emotional Problems Score, Severe Emotional Problems, Overweight Risk, Physical Activity,* and *Dentist Frequency* respectively. The level of education is measured in years. Pcg has a low level of education if s/he obtained 13 years of education or less. Standard errors, reported in parentheses, are corrected for heteroskedasticity, and allow for clustering at the household level. The symbols ***, **, * indicate that the coefficients are statistically significant at the 1, 5, and 10 percent level respectively.

Results in Table 26 show that adolescents whose primary caregiver has a low level of education exhibit a greater probability of having severe emotional problems by on average 18 percentage points. In addition, they present a higher risk of being overweight and a lower likelihood of performing physical activity by 23 and 16 percentage points respectively. Interestingly, for adolescent with a high educated primary caregiver, parental divorce exerts a negative impact on all the outcomes considered. These adolescents present higher emotional problems, with an increase of 0.72 points and 13 percentage points respectively. Their likelihood of being overweight increases by nearly 10 percentage points, while their likelihood of being engaged in physical activity decrease by 12 percentage points. Lastly, their probability of visiting the dentist at least once a year decreases on average by 16 percentage points.

Regarding the age of the primary caregiver at child's birth, we find that parental divorce exerts very detrimental consequences for those adolescents with younger primary caregiver. The effects are statistically significant at the 1 and 5 percent levels, as evidenced in Table 27.

Compared to adolescents whose parents stay married, treated adolescents with younger primary caregiver experience greater levels of emotional problems, on average by 0.59 points and 16 percentage points more. Additionally, their risk of being overweight increases by 14 percentage points, while their likelihood of performing physical activity decreases by 18 percentage points. Finally, their probability of visiting the dentist at least once a year reduces by 14 percentage points on average.

Table 27. Heterogeneity by Primary Caregiver's Age at Child's Birth - Late parental divorce

	Emotional problems	Severe emotional problems	Overweight risk	Physical activity	Dentist frequency
Panel A: Age at Child's Birth ≤32					
Disrupted	0.505**	0.0568	-0.0386	-0.0106	-0.0312
	(0.243)	(0.0464)	(0.0471)	(0.0241)	(0.0396)
Disrupted * age17	0.586**	0.161***	0.140**	-0.182***	-0.140**
	(0.271)	(0.0574)	(0.0586)	(0.0578)	(0.0583)
Observations	4825	4908	4027	4893	4908
R-squared	0.00911	0.00964	0.00936	0.0354	0.0588
Panel B: Age at Child's Birth >32					
Disrupted	0.668	0.0463	-0.185***	-0.0054	-0.0369
	(0.477)	(0.0735)	(0.0106)	(0.0330)	(0.0625)
Disrupted *age17	0.594	0.119	0.186**	-0.0188	0.00429
	(0.425)	(0.0959)	(0.0835)	(0.0553)	(0.0918)
Observations	3918	4042	3234	4014	4042
R-squared	0.00709	0.00556	0.00484	0.0259	0.0343

Notes: Each column reports estimates from OLS regression. GUI dataset, children aged 13, and 17 The dependent variables are *Emotional Problems Score, Severe Emotional Problems, Overweight Risk, Physical Activity,* and *Dentist Frequency* respectively. Standard errors, reported in parentheses, are corrected for heteroskedasticity, and allow for clustering at the household level. The symbols ***, **, * indicate that the coefficients are statistically significant at the 1, 5, and 10 percent level respectively.

In Table 28 we estimate separate regressions for adolescents in household with income below and above the median, respectively. Results show that the adverse effects are predominantly concentrated among adolescents in households with income lower than the median.

These adolescents face higher emotional problems, 0.78 points and 20 percentage points respectively. On average, their risk of being overweight significantly increases by 16 percentage points, while their likelihood of being engaged in physical activity decreases by 22 percentage points. Finally, for adolescents in household with income above the median value, parental divorce negatively affects their likelihood of being overweight, 17 percentage points higher, and their probability of visiting the dentist at least once a year, 16 percentage points lower on average.

Table 28. Heterogeneity by Household Income (below/above median) – Late parental divorce

	Emotional problems	Severe emotional problems	Overweight risk	Physical activity	Dentist frequency
Panel A: Household Income ≤ median					
Disrupted	0.483*	0.0114	-0.105**	0.0105	-0.0511
	(0.272)	(0.0479)	(0.0439)	(0.0200)	(0.0464)
Disrupted * age17	0.776**	0.203***	0.161**	-0.216***	-0.0934
	(0.305)	(0.0640)	(0.0625)	(0.0615)	(0.0639)
Observations	5256	5363	4487	5335	5363
R-squared	0.00711	0.00560	0.00641	0.0338	0.0428
Panel B: Household Income > median					
Disrupted	0.675*	0.118*	-0.0346	-0.0458	0.00294
	(0.368)	(0.0675)	(0.0594)	(0.0409)	(0.0411)
Disrupted * age17	0.346	0.0718	0.173**	-0.005	-0.158*
	(0.360)	(0.0778)	(0.0749)	(0.0641)	(0.0828)
Observations	3487	3587	2774	3572	3587
R-squared	0.00938	0.0113	0.00566	0.0237	0.0468

Notes: Each column reports estimates from OLS regression. GUI dataset, children aged 13, and 17 The dependent variables are *Emotional Problems Score, Severe Emotional Problems, Overweight Risk, Physical Activity,* and *Dentist Frequency* respectively. Standard errors, reported in parentheses, are corrected for heteroskedasticity, and allow for clustering at the household level. The symbols ***, **, * indicate that the coefficients are statistically significant at the 1, 5, and 10 percent level respectively.

Finally, we focus on the area of residence and divide our sample into two groups: households living in rural area and household living in urban one.

As shown in Table 29, the area of residence shapes the impact of parental divorce, with negative consequences for those residing in rural area.

Adolescents who experience parental divorce between the ages of 13 and 17 and live in rural area present greater emotional problems than adolescents with stably married parents, on average more 0.94 points and 24 percentage points respectively. Furthermore, their physical health is worse: their likelihood of being overweight increases by 12percentage points, while their likelihood of performing physical activity decreases by 17 percentage points. Finally, their probability of visiting the dentist frequently decreases by 10 percentage points on average.

Table 29. Heterogeneity by Area of residence - Late parental divorce

	Emotional problems	Severe emotional problems	Overweight risk	Physical activity	Dentist frequency
Panel A: Rural area					
Disrupted	0.241	-0.0286	-0.121***	0.0164	-0.0379
	(0.254)	(0.0426)	(0.0411)	(0.0161)	(0.0445)
Disrupted * age 17	0.943***	0.238***	0.122**	-0.167***	-0.101*
	(0.301)	(0.0696)	(0.0612)	(0.0597)	(0.0611)
Observations	5006	5138	4182	5109	5138
R-squared	0.00859	0.00839	0.00586	0.0328	0.0488
Panel B: Urban area					
Disrupted	1.019***	0.164**	-0.0159	-0.0464	-0.0341
	(0.366)	(0.0676)	(0.0602)	(0.0394)	(0.0505)
Disrupted * age 17	0.125	0.0359	0.219***	-0.103	-0.124
	(0.338)	(0.0645)	(0.0760)	(0.0693)	(0.0825)
Observations	3737	3812	3079	3798	3812
R-squared	0.00969	0.00897	0.0106	0.0283	0.0458

Notes: Each column reports estimates from OLS regression. GUI dataset, children aged 13, and 17. The dependent variables are *Emotional Problems Score, Severe Emotional Problems, Overweight Risk, Physical Activity,* and *Dentist Frequency* respectively. Standard errors, reported in parentheses, are corrected for heteroskedasticity, and allow for clustering at the household level. The symbols ***, **, * indicate that the coefficients are statistically significant at the 1, 5, and 10 percent level respectively.

X. Conclusion

Parental divorce has been shown to have detrimental impact on children's development, emotional and psychological wellbeing, educational outcomes, and physical health (Amato and James, 2010; Sun and Li, 2011; Amato, 2014).

This paper studies the impact of parental divorce on a bunch of behavioral and health-related children's outcomes using three waves of data from the Growing in Up in Ireland (GUI), the National Longitudinal Study of Children and Young people in Ireland. The longitudinal nature of the data allows us to account for the endogeneity of parental divorce by implementing a simple 2x2 Difference-in-Differences study design and a Two-way-fixed-effects model when both household and time fixed-effects are included.

The main contribution of our paper is that, among the same cohort of adolescents, age is a key factor in mitigating or exacerbating the effects of divorce. Indeed, in line with the strand of literature according to which parental divorce has very detrimental effects on children if it takes place in a very early stage of their life (Cherlin, Chase-Lansdale and McRae, 1998; Cavanagh and Huston, 2006), our results show that parental divorce does not harm mental and physical children's health if they are between 9 and 13 years when it occurs. However, strongly significant and negative effects are found if divorce takes place when children are between 13 and 17 years of age, providing evidence of an heterogenous role of age on adolescent's adjustment to parental divorce. Young who experience parental divorce between 13 and 17 years have on average 0.6 points more in the emotional problems score; they show a higher likelihood of having severe emotional problems by 15 percentage points; their likelihood of visiting the dentist at least once a year sharply decreases by 11 percentage points; their physical health and exercise activity are also affected with an increase in the risk of being overweight by 16 percentage points and a decrease in the likelihood of exercising by 14 percentage points. These results are in line with other studies who documented an adverse impact of divorce on children's emotional well-being (Sigle-Rushton, Hobcraft and Kiernan, 2005; Rasmussen, 2009), and on health and well-being (Chen and Escarce, 2010; Augustine and Kimbro, 2013; Bright *et al.*, 2015; Goisis, Özcan and Van Kerm, 2019).

As a robustness check, we analyse the impact of parental breakdown on children's outcomes including unmarried but cohabiting parents in the analysis. Results are very similar with the previous ones, with no evidence of a statistically significant effect if parental dissolution takes place when children are aged between 9-13, and strong negative effects if the dissolution takes place at a later age, between 13-17 years old.

Furthermore, we document large heterogeneous effects by child's gender, primary caregiver's age at child's birth and level of education measured in year, household's income, and area of residence.

Note that preexisting differences between the two groups of children can be observed before divorce occurs in socioemotional outcomes and the risk of being overweight. Moreover, for the group of children whose parents divorce during the last stage of adolescence, i.e., between the ages of 13 and 17, not only do these differences persist after the occurrence of parental divorce, but the event significantly increases the gap between these two groups of children.

Although the impact of parental divorce on children is statistically insignificant when divorce occurs between the age of 9 and 13, this group exhibits the highest initial emotional scores. However, contrary to the others group of children, it is the only one in which the emotional problems show a decreasing trend over time. While it is not possible to examine the dynamics of this group during the pre-divorce period due to the absence of data, it is reasonable to suppose that parental divorce may have had a positive effect on the well-being of these children by alleviating unobserved factors commonly associated with divorce, such as stress and parental conflict.

It is noteworthy that the available data allows us to examine the effects of divorce on children in both a short and a moderately medium-long term period, but only for the first group of treated children. For the second group of children, who experience parental divorce between the ages of 13 and 17, we can only focus on the short-term effects, as mentioned in the introduction. With access to more data on post-divorce periods, further investigations could determine whether and to what extent this detrimental impact observed in older children persists over time.

Our study is not without limitations. The first limitation relates to the lack of information regarding the exact moment at which parental divorce occurs. We refer to the first post-divorce period as short-term, although our findings may uncover medium-long term effects if divorce occurred shortly after the initial round of interview or even shorter-term if divorce took place shortly before the second survey.

This should not pose an issue in the first treated groups since we do have two post-divorce periods available, and our models – run over specific subsamples and with all the interaction terms needed - should be able to catch any effects at work regardless of how we define it. Conversely, some problems may arise with the group of children who undergoes parental divorce at a more advanced age, potentially leading to an underestimation of effects or an inability to detect any if it occurs with short advance relative to the time of the interview.

Another limitation arises in relation to the outcomes of emotional problems, as they are constructed based on the information provided by the primary caregiver. We hypothesize that the child's perspective may differ from those of the primary caregiver.

References

Abufhele, A., Behrman, J. and Bravo, D. (2017) 'Parental preferences and allocations of investments in children's learning and health within families', *Social Science and Medicine*, 194, pp. 76–86. Available at: https://doi.org/10.1016/j.socscimed.2017.09.051.

Amato, P.R. (1996) 'Explaining the Intergenerational Transmission of Divorce', *Journal of Marriage and the Family*, 58(3), p. 628. Available at: https://doi.org/10.2307/353723.

Amato, P.R. (2000) 'The consequences of divorce for adults and children', *Journal of Marriage and Family*. National Council on Family Relations, pp. 1269–1287. Available at: https://doi.org/10.1111/j.1741-3737.2000.01269.x.

Amato, P.R. (2005) 'The Impact of Family Formation Change on the Cognitive, Social, and Emotional Well-Being of the Next Generation', *The Future of Children*, 15(2), pp. 75–96.

Amato, P.R. (2010) 'Research on divorce: Continuing trends and new developments', *Journal of Marriage and Family*, 72(3), pp. 650–666. Available at: https://doi.org/10.1111/j.1741-3737.2010.00723.x.

Amato, P.R. (2014) 'The consequences of divorce for adults and children: an update', *Drustvena Istrazivanja*, 23(1), pp. 5–24. Available at: https://doi.org/10.5559/di.23.1.01.

Amato, P.R. (2016) 'The Consequences of Divorce for Adults and Children Published by : National Council on Family Relations Linked references are available on JSTOR for this article : The Consequences of Divorce for Adults and Children', *Journal of Marriage and Family*, 62(4), pp. 1269–1287.

Amato, P.R. and Anthony, C.J. (2014) 'Estimating the effects of parental divorce and death with fixed effects models', *Journal of Marriage and Family*, 76(2), pp. 370–386. Available at: https://doi.org/10.1111/jomf.12100.

Amato, P.R. and Cheadle, J.E. (2008) 'Parental divorce, marital conflict and children's behavior problems: A comparison of adopted and biological children', *Social Forces*, 86(3), pp. 1139–1161. Available at: https://doi.org/10.1353/sof.0.0025.

Amato, P.R. and DeBoer, D.D. (2001) 'The transmission of marital instability across generations: Relationship skills or commitment to marriage?', *Journal of Marriage and Family*, 63(4), pp. 1038–1051. Available at: https://doi.org/10.1111/j.1741-3737.2001.01038.x.

Amato, P.R. and James, S. (2010) 'Divorce in Europe and the United States: Commonalities and differences across nations', *Family Science*, 1(1), pp. 2–13. Available at: https://doi.org/10.1080/19424620903381583.

Amato, Paul R and Keith, B. (1991) 'Linked references are available on JSTOR for this article : Parental Divorce and Adult Well-being : A Meta-analysis', 53(1), pp. 43–58.

Amato, Paul R. and Keith, B. (1991) 'Parental divorce and the well-being of children: A meta-analysis', *Psychological Bulletin*, 110(1), pp. 26–46. Available at: https://doi.org/10.1037/0033-2909.110.1.26.

Amato, P.R. and Sobolewski, J.M. (2001) 'The effects of divorce and marital discord on adult children's psychological well-being', *American Sociological Review*, 66(6), pp. 900–921. Available at: https://doi.org/10.2307/3088878.

Anderson, J. (2014) 'The impact of family structure on the health of children: Effects of divorce', *Linacre Quarterly*, 81(4), pp. 378–387. Available at: https://doi.org/10.1179/0024363914Z.00000000087.

Andini, M. *et al.* (2018) 'Targeting with machine learning: An application to a tax rebate program in Italy', *Journal of Economic Behavior and Organization*, 156(March), pp. 86–102. Available at: https://doi.org/10.1016/j.jebo.2018.09.010.

Antulov-Fantulin, N., Lagravinese, R. and Resce, G. (2021) 'Predicting bankruptcy of local government: A machine learning approach', *Journal of Economic Behavior and Organization*, 183, pp. 681–699. Available at: https://doi.org/10.1016/j.jebo.2021.01.014.

Arlot, S. and Celisse, A. (2010) 'A survey of cross-validation procedures for model selection', *Statistics Surveys*, 4, pp. 40–79. Available at: https://doi.org/10.1214/09-SS054.

Arpino, B., Moglie, M. Le and Mencarini, L. (2018) 'Machine-Learning techniques for family demography : An application of random forests to the analysis of divorce determinants in Germany', *Annual Meeting Paa*, (56), p. 32.

Aughinbaugh, A., Pierret, C.R. and Rothstein, D.S. (2005) 'The impact of family structure transitions on youth achievement: Evidence from the children of the NLSY79', *Demography*, 42(3), pp. 447–468. Available at: https://doi.org/10.1353/dem.2005.0023.

Augustine, J.M. and Kimbro, R.T. (2013) 'Journal of Applied Research on Children: Informing Policy for Children at Risk', *Journal of Applied Research on Children: Informing Policy for Children at Risk*, 4(1), pp. 0–24. Available at: http://digitalcommons.library.tmc.edu/childrenatrisk%5Cnhttp://digitalcommons.library.tmc.edu/childrenatrisk/vol4/iss1/5.

Augustine, J.M. and Kimbro, R.T. (2017) 'Associations and Intervening Mechanisms Between Family Structure and Young Children's Obesity', *Journal of Family Issues*, 38(16), pp. 2277–2302. Available at: https://doi.org/10.1177/0192513X15621344.

Austin, P.C. (2011) 'An introduction to propensity score methods for reducing the effects of confounding in observational studies', *Multivariate Behavioral Research*, 46(3), pp. 399–424. Available at: https://doi.org/10.1080/00273171.2011.568786.

Beets, M.W., Cardinal, B.J. and Alderman, B.L. (2010) 'Parental social support and the physical activity-related behaviors of youth: A review', *Health Education and Behavior*, 37(5), pp. 621–644. Available at: https://doi.org/10.1177/1090198110363884.

Belloni, A. *et al.* (2012) 'Sparse Models and Methods for Optimal Instruments With an Application to Eminent Domain', *Econometrica*, 80(6), pp. 2369–2429. Available at: https://doi.org/10.3982/ecta9626.

Belloni, A., Chernozhukov, V. and Hansen, C. (2014) 'on Structural and Treatment Effects †', *Journarl of economci perspectives*, 28(2), pp. 29–50.

Berenson, G.S. (2012) 'Health Consequences of Obesity', *Pediatric Blood and Cancer*, 58(1), pp. 117–121. Available at: https://doi.org/10.1002/pbc.23373.

Bernardi, F. and Boertien, D. (2016) 'Understanding heterogeneity in the effects of parental separation on educational attainment in Britain: Do children from lower educational backgrounds have less to lose?', *European Sociological Review*, 32(6), pp. 807–819. Available at: https://doi.org/10.1093/esr/jcw036.

Biblarz, T.J. and Gottainer, G. (2000) ‘Family structure and children’s success: A comparison of widowed and divorced single-mother families’, *Journal of Marriage and Family*, 62(2), pp. 533–548. Available at: https://doi.org/10.1111/j.1741-3737.2000.00533.x.

de Blasio, G. and D’ignazio, A. (2021) ‘Regional policy in the machine learning era: Opportunities and challenges’, *Scienze Regionali*, 20(2), pp. 205–220. Available at: https://doi.org/10.14650/99817.

Bradley RH, C.R. (2008) ‘Infant temperament, parenting, and externalizing behavior in first grade: a test of the differential susceptibility hypothesis’, *J Child Psychol Psychiatry*, 49(2), pp. 124–31.

Bream, V. and Buchanan, A. (2003) ‘Distress among Children whose Separated or Divorced Parents cannot Agree Arrangements for Them’, *British Journal of Social Work*, 33(2), pp. 227–238. Available at: https://doi.org/10.1093/bjsw/33.2.227.

Bright, M.A. *et al.* (2015) ‘Adverse childhood experiences and dental health in children and adolescents’, *Community Dentistry and Oral Epidemiology*, 43(3), pp. 193–199. Available at: https://doi.org/10.1111/cdoe.12137.

Brown, S.L. (2004) ‘Family structure and child well-being: The significance of parental cohabitation’, *Journal of Marriage and Family*, 66(2), pp. 351–367. Available at: https://doi.org/10.1111/j.1741-3737.2004.00025.x.

Bzostek, S.H. and Beck, A.N. (2011) ‘Familial instability and young children’s physical health’, *Social Science and Medicine*, 73(2), pp. 282–292. Available at: https://doi.org/10.1016/j.socscimed.2011.04.014.

Carrieri, V., Lagravinese, R. and Resce, G. (2021) ‘Predicting vaccine hesitancy from area-level indicators: A machine learning approach’, *Health Economics (United Kingdom)*, 30(12), pp. 3248–3256. Available at: https://doi.org/10.1002/hec.4430.

Cavanagh, S. and Fomby, P. (2019) ‘Family Instability in the Lives of American Children’, *Annual Review of Sociology*, 45(1), pp. 493–513. Available at: https://doi.org/10.1146/annurev-soc-073018-022633.

Cavanagh, S.E. and Huston, A.C. (2006) ‘Family instability and children’s early problem behavior’, *Social Forces*, 85(1), pp. 551–581. Available at: https://doi.org/10.1353/sof.2006.0120.

Cerqua, A. *et al.* (2021) ‘Local mortality estimates during the COVID-19 pandemic in Italy’, *Journal of Population Economics*, 34(4), pp. 1189–1217. Available at: https://doi.org/10.1007/s00148-021-00857-y.

Chen, A.Y. and Escarce, J.J. (2010) ‘Family structure and childhood obesity, Early Childhood Longitudinal Study - Kindergarten Cohort’, *Preventing Chronic Disease*, 7(3). Available at: http://www.cdc.gov/pcd/issues/2010/may/09_0156.htm.Accessed[date].

Chen, X., Liu, M. and Li, D. (2000) ‘Parental warmth, control, and indulgence and their relations to adjustment in Chinese children: A longitudinal study.’, *Journal of Family Psychology*, 14, pp. 401–419.

Chen, Y.-C., Fan, E. and Liu, J.-T. (2019) ‘Understanding the mechanisms of parental divorce effects on child’s higher education’, *NBER Working Papers* [Preprint]. Available at: http://www.nber.org/papers/w16019.

Cherlin, A.J. *et al.* (1991) ‘Longitudinal studies of effects of divorce on children in Great Britain and the United States', *Science*, 252(5011), pp. 1386–1389. Available at: https://doi.org/10.1126/science.2047851.

Cherlin, A.J., Chase-Lansdale, P.L. and McRae, C. (1998) 'Effects of parental divorce on mental health throughout the life course', *American Sociological Review*, 63(2), pp. 239–249. Available at: https://doi.org/10.2307/2657325.

Chernozhukov, V. *et al.* (2018) 'Double/debiased machine learning for treatment and structural parameters', *Econometrics Journal*, 21(1), pp. C1–C68. Available at: https://doi.org/10.1111/ectj.12097.

Christina J. Cross (2019) 'Racial/Ethnic Differences in the Association Between Family Structure and Children's Education', *Journal of Marriage and Family*, Vol. 90(November), pp. 1–22.

Cid, A. and Stokes, C.E. (2013) 'Family Structure and Children's Education Outcome: Evidence from Uruguay', *Journal of Family and Economic Issues*, 34(2), pp. 185–199. Available at: https://doi.org/10.1007/s10834-012-9326-z.

Cooper, C.E. *et al.* (2008) 'Partnership Instability and Child Wellbeing during the Transition to Elementary School', (April), pp. 1–37.

Duflo, E., Bertrand, M. and Mullainathan, S. (2004) 'How Much Should We Trust Differences-in-Differences Estimates ? Author (s): Marianne Bertrand , Esther Duflo and Sendhil Mullainathan Published by : Oxford University Press Stable URL : http://www.jstor.org/stable/25098683 HOW MUCH SHOULD WE TRUST DIFFE', *The Quarterly Journal of Economics*, 119(1), pp. 249–275.

E. Leuven and Sianesi, B. (2003) 'Psmatch2: Stata module to perform full Mahalanobis and propensity score matching, common support graphing, and covariate impalance testing'. Available at: https://www.pep-net.org/sites/pep-net.org/files/typo3doc/pdf/Training_Material/statadoc.pdf.

Ermisch, J.F. *et al.* (2001) *Family structure and children's achievements*.

Fischer, T. (2007) 'Parental divorce and children's socio-economic success: Conditional effects of parental resources prior to divorce, and gender of the child', *Sociology*, 41(3), pp. 475–495. Available at: https://doi.org/10.1177/0038038507076618.

Fomby, P. and Cherlin, A.J. (2007) 'Family instability and child well-being', *American Sociological Review*, 72(2), pp. 181–204. Available at: https://doi.org/10.1177/000312240707200203.

Fomby, P. and Osborne, C. (2010) 'The influence of union instability and union quality on children's aggressive behavior', *Social Science Research*, 39(6), pp. 912–924. Available at: https://doi.org/10.1016/j.ssresearch.2010.02.006.

Francesconi, M., Jenkins, S.P. and Siedler, T. (2010) 'Childhood family structure and schooling outcomes: Evidence for Germany', *Journal of Population Economics*, 23(3), pp. 1073–1103. Available at: https://doi.org/10.1007/s00148-009-0242-y.

Geisser, S. (1975) 'The predictive sample reuse method with applications', *Journal of the American Statistical Association*, 70(350), pp. 320–328. Available at: https://doi.org/10.1080/01621459.1975.10479865.

Gibson, L.Y. *et al.* (2007) 'The role of family and maternal factors in childhood obesity', *Medical Journal of Australia*, 186(11), pp. 591–595. Available at: https://doi.org/10.5694/j.1326-5377.2007.tb01061.x.

Goisis, A., Özcan, B. and Van Kerm, P. (2019) ‘Do Children Carry the Weight of Divorce?’, *Demography*, 56(3), pp. 785–811. Available at: https://doi.org/10.1007/s13524-019-00784-4.

Golan, M. and Crow, S. (2004) ‘Targeting parents exclusively in the treatment of childhood obesity: Long-term results’, *Obesity Research*, 12(2), pp. 357–361. Available at: https://doi.org/10.1038/oby.2004.45.

Gonzalez, L. (2011) ‘The Long Term Effects of Legalizing Divorce on Children’, *SSRN Electronic Journal* [Preprint], (3789). Available at: https://doi.org/10.2139/ssrn.1374206.

Goodman, A. and Goodman, R. (2009) ‘Strengths and difficulties questionnaire as a dimensional measure of child mental health’, *Journal of the American Academy of Child and Adolescent Psychiatry*, 48(4), pp. 400–403. Available at: https://doi.org/10.1097/CHI.0b013e3181985068.

Goodman, R. (1997) ‘The strengths and difficulties questionnaire: A research note’, *Journal of child psycology and psychiatry*, 38(5), pp. 581–586. Available at: https://doi.org/10.4324/9781315258324-21.

Gruber, J. (2004) ‘Is making divorce easier bad for children? the long-run implications of unilateral divorce’, *Journal of Labor Economics*, 22(4), pp. 799–833. Available at: https://doi.org/10.1086/423155.

Hannan, C. and Halpin, B. (2014) ‘The influence of family structure on child outcomes: Evidence for Ireland’, *Economic and Social Review*, 45(1), pp. 1–24.

Hassen, T.A. *et al.* (2021) ‘The association between birth weight and proxy-reported health-related quality of life among children aged 5 − 10 years old: A linked data analysis’, *BMC Pediatrics*, 21(1), pp. 1–8. Available at: https://doi.org/10.1186/s12887-021-02882-y.

Ho, D.E. *et al.* (2007) ‘Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference’, *Political Analysis*, 15(3), pp. 199–236. Available at: https://doi.org/10.1093/pan/mpl013.

Hveem, M.R., Faulconer, S.C.M. and Dufur, M.J. (2022) ‘Comparing Children’s Behavior Problems in Biological Married, Biological Cohabitating, and Stepmother Families in the UK’, *International Journal of Environmental Research and Public Health*, 19(24). Available at: https://doi.org/10.3390/ijerph192416543.

J.D., A. and Pischke, J.S. (2009) *Mostly Harmless Econometrics : An Empiricist ’ s Companion*.

Jackson, S.L. *et al.* (2011) ‘Impact of poor oral health on children’s school attendance and performance’, *American Journal of Public Health*, 101(10), pp. 1900–1906. Available at: https://doi.org/10.2105/AJPH.2010.200915.

King, V. (2009) ‘Stepfamily formation: Implications for adolescent ties to mothers, nonresident fathers, and stepfathers’, *Journal of Marriage and Family*, 71(4), pp. 954–968. Available at: https://doi.org/10.1111/j.1741-3737.2009.00646.x.

Kleinberg, J. *et al.* (2015) ‘Prediction Policy Problems’, *American Economics Review*, pp. 1–6.

Kravdal, Ø. and Grundy, E. (2019) ‘Children’s age at parental divorce and depression in early and mid-adulthood’, *Population Studies*, 73(1), pp. 37–56. Available at: https://doi.org/10.1080/00324728.2018.1549747.

Lisa Strohschein (2005) ‘Parental Divorce and Child Mental Health Trajectories’, 67(5), pp. 1286–

1300.

Listl, S. (2011) ‘Family composition and children’s dental health behavior: Evidence from Germany’, *Journal of Public Health Dentistry*, 71(2), pp. 91–101. Available at: https://doi.org/10.1111/j.1752-7325.2010.00205.x.

McCarty, C.A. *et al.* (2005) ‘Parental Emotional Support and Subsequent Internalizing and Externalizing Problems Among Children’, *Journal of Developmental & Behavioral Pediatrics*, 26(4), pp. 267–275.

McNamara, E., Murray, A. and O’Mahony, D. (2019) ‘Growing up in Ireland. Design, Instrumentation and Procedures for Cohort ’08 at 9 Years Old (Wave 5)’, *Longitudinal and Life Course Studies* [Preprint]. Available at: https://doi.org/10.1332/175795919X15699445254797.

Meesters, C. and Muris, P. (2004) ‘Perceived parental rearing behaviours and coping in young adolescents’, *Personality and Individual Differences*, 37(3), pp. 513–522. Available at: https://doi.org/10.1016/j.paid.2003.09.022.

Meyer, B.D. (1995) ‘Natural and quasi-experiments in economics’, *Journal of Business and Economic Statistics*, 13(2), pp. 151–161. Available at: https://doi.org/10.1080/07350015.1995.10524589.

Osborne, C. and Mclanahan, S. (2007) ‘Partnership Instability and Child Well-Being’, *Journal of Marriage and Family*, 69(4), pp. 1065–1083.

De Paola, M. and Gioia, F. (2017) ‘Does patience matter in marriage stability? Some evidence from Italy’, *Review of Economics of the Household*, 15(2), pp. 549–577. Available at: https://doi.org/10.1007/s11150-014-9275-4.

Parcel, T.L., Campbell, L.A. and Zhong, W. (2012) ‘Children’s Behavior Problems in the United States and Great Britain’, *Journal of Health and Social Behavior*, 53(2), pp. 165–182. Available at: https://doi.org/10.1177/0022146512436742.

Paul R. Rosenbaum and Donald B. Rubin (1983) ‘The Central Role of the Propensity Score in Observational Studies for Causal Effects Author (s): Paul R . Rosenbaum and Donald B . Rubin Published by : Oxford University Press on behalf of Biometrika Trust Stable URL : http://www.jstor.’, 70(1), pp. 41–55.

Pianta, R.C. (1992) ‘Child-parent relationship scale.’, *Unpublished measure, University of Virginia* [Preprint], (427).

Powers, D.A. (2005) ‘Effects of family structure on the risk of first premarital birth in the presence of correlated unmeasured family effects’, *Social Science Research*, 34(3), pp. 511–537. Available at: https://doi.org/10.1016/j.ssresearch.2004.04.007.

Rasmussen, A.W. (2009) *Family Structure Changes and Children ′ s Health , Behavior , and Educational Outcomes Department of Economics Family Structure Changes and Children ’ s Health , Behavior , and Educational Outcomes*.

Resce, G. (2022) ‘The impact of political and non-political officials on the financial management of local governments’, *Journal of Policy Modeling*, 44(5), pp. 943–962. Available at: https://doi.org/10.1016/j.jpolmod.2022.08.004.

Resce, G. and Vaquero-Piñeiro, C. (2022) ‘Predicting agri-food quality across space: A Machine Learning model for the acknowledgment of Geographical Indications’, *Food Policy*, 112(September), p. 102345. Available at: https://doi.org/10.1016/j.foodpol.2022.102345.

Resce, G. and Vaquero-Piñeiro, C. (2023) 'Taste of home: Birth town bias in Geographical Indications', *Available at SSRN 4437094* [Preprint], (089). Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4437094%0Ahttp://web.unimol.it/progetti/repec/mol/ecsdps/ESDP23089.pdf.

Rubin, D.B. (2001) 'Using propensity scores to help design observational studies: Application to the tobacco litigation', *Health Services & Outcomes Research Methodology*, (2), pp. 169–188. Available at: https://doi.org/10.1017/CBO9780511810725.030.

Rubin, D.B. (2006) *Matched sampling for causal inference*. Edited by U.C.U.P. Cambridge.

Sandra L . Hofferth (2006) 'Residential Father Family Type and Child Well-Being : Investment versus Selection', *Springer on behalf of the Population Association of America*, 43(1), pp. 53–77.

Sanz-De-Galdeano, A. and Vuri, D. (2007) 'Parental divorce and students' performance: Evidence from longitudinal data', *Oxford Bulletin of Economics and Statistics*, 69(3), pp. 321–338. Available at: https://doi.org/10.1111/j.1468-0084.2006.00199.x.

Schmeer, K.K. (2012) 'Family structure and obesity in early childhood', *Social Science Research*, 41(4), pp. 820–832. Available at: https://doi.org/10.1016/j.ssresearch.2012.01.007.

Sharon H. Bzostek, L.M.B. (2017) 'Family Structure Experiences and Child Socioemotional Development During the First Nine Years of Life: Examining Heterogeneity by Family Structure at Birth', *Demography*, 54(2), pp. 513–540. Available at: https://doi.org/https://doi.org/10.1007/s13524-017-0563-5.

Sheiham, A. (2006) 'Dental caries affects body weight, growth and quality of life in pre-school children', *British Dental Journal*, 201(10), pp. 625–626. Available at: https://doi.org/10.1038/sj.bdj.4814259.

Sigle-Rushton, W., Hobcraft, J. and Kiernan, K. (2005) 'Parental divorce and subsequent disadvantage: A cross-cohort comparison', *Demography*, 42(3), pp. 427–446. Available at: https://doi.org/10.1353/dem.2005.0026.

Smetana, J.G. (2017) 'Current research on parenting styles, dimensions, and beliefs', *Current Opinion in Psychology*, 15, pp. 19–25. Available at: https://doi.org/10.1016/j.copsyc.2017.02.012.

Sohail, S. *et al.* (2018) 'Implementation of machine learning algorithm on factors effecting divorce rate', *2018 International Conference on Engineering and Emerging Technologies, ICEET 2018*, 2018-Janua, pp. 1–5. Available at: https://doi.org/10.1109/ICEET1.2018.8338618.

StataCorp LLC (2023) *Stata 2018 Lasso Reference Manual*.

Steinberg, L., Fletcher, A. and Darling, N. (1994) 'Parental monitoring and peer influences on adolescent substance use', *Pediatrics*, 93(6 II), pp. 1060–1064. Available at: https://doi.org/10.1542/peds.93.6.1060.

Stone, L.L. *et al.* (2010) 'Psychometric properties of the parent and teacher versions of the strengths and difficulties questionnaire for 4- to 12-Year-olds: A review', *Clinical Child and Family Psychology Review*, 13(3), pp. 254–274. Available at: https://doi.org/10.1007/s10567-010-0071-2.

StØrksen, I. *et al.* (2006) 'Adolescent adjustment and well-being: Effects of parental divorce and distress', *Scandinavian Journal of Psychology*, 47(1), pp. 75–84. Available at:

https://doi.org/10.1111/j.1467-9450.2006.00494.x.

Sun, Y. (2001) ‘Family environment and adolescents’ well-being before and after parents’ marital disruption: A longitudinal analysis’, *Journal of Marriage and Family*, 63(3), pp. 697–713. Available at: https://doi.org/10.1111/j.1741-3737.2001.00697.x.

Sun, Y. and Li, Y. (2011) ‘Effects of family structure type and stability on children’s academic performance trajectories’, *Journal of Marriage and Family*, 73(3), pp. 541–556. Available at: https://doi.org/10.1111/j.1741-3737.2011.00825.x.

Tendulkar, S.A.; Buka, S.; Dunn, E.C.; Subramanian, S.V.; Koenen, K.C. (2010) ‘A multilevel investigation of neighborhood effects on parental warmth.’, *J. Community Psychol.*, 38, pp. 557–573.

Thornton, M. *et al.* (2013) ‘Growing up in Ireland. Design, Instrumentation and Procedures for the Infant Cohort at wave 1 (9 Months)’, *Longitudinal and Life Course Studies* [Preprint]. Available at: https://doi.org/10.1332/175795919X15699445254797.

Tibshirani (1996) ‘Regression Shrinkage and Selection via the Lasso’, *Journal of the Royal Statistical Society. Series B (Methodological)*, 58(1), pp. 267–288. Available at: jstor.org/stable/2346178.

Varian, H.R. (2014) ‘Big data: New tricks for econometrics’, *Journal of Economic Perspectives*, 28(2), pp. 3–28. Available at: https://doi.org/10.1257/jep.28.2.3.

Varian, H.R. (2016) ‘Causal inference in economics and marketing’, *Proceedings of the National Academy of Sciences of the United States of America*, 113(27), pp. 7310–7315. Available at: https://doi.org/10.1073/pnas.1510479113.

Waldfogel, J., Craigie, T.A. and Brooks-Gunn, J. (2010) ‘Fragile families and child wellbeing’, *Future of Children*, 20(2), pp. 87–112. Available at: https://doi.org/10.1353/foc.2010.0002.

Williams, J. *et al.* (2015) ‘Growing Up in Ireland. Design, Instrumentation and Procedures for Cohort ’08 at wave 3 (5 Years)’. Available at: www.dcya.gov.ie.

Wing, C., Simon, K. and Bello-Gomez, R.A. (2018) ‘Designing Difference in Difference Studies: Best Practices for Public Health Policy Research’, *Annual Review of Public Health*, 39, pp. 453–469. Available at: https://doi.org/10.1146/annurev-publhealth-040617-013507.

Yang, Z., Zhang, R. and Liu, L. (2022) ‘The Effect of Parent-Child Relationship on Child Development’, *Academic Journal of Humanities & Social Sciences*, 5(15), pp. 118–122. Available at: https://doi.org/10.25236/ajhss.2022.051515.

Yelick, A. (2017) ‘The Effects of Family Structure on Consumption and Exercise Patterns for Adolescent Youth’, *Child and Adolescent Social Work Journal*, 34(4), pp. 381–395. Available at: https://doi.org/10.1007/s10560-016-0468-y.

Chapter II

Family Structure Change and Children’s Educational Outcomes: A Double-Machine Learning Approach

Abstract

The traditional nuclear family has been acknowledged as the better environment for the development of children. However, an increasing number of children grow up in non-traditional families. In this article, we analyse data from the Growing Up in Ireland study to evaluate whether undergoing a change in the family structure affects child’s later educational achievement. Using a recently developed double-machine learning estimator, we show that the traditional “ad-hoc” variable selection procedure can overestimate the impact of such a transition, and we describe the advantages of using this new tool. Controlling for the age of the children, our findings suggest that experiencing a family structure change between the ages of 9 and 13 has a detrimental effect on educational achievements measured when the adolescent reaches the age of 20; undergoing such a transition before the age of 9 and between the ages of 13 and 17 does not exert any influence on the educational outcomes.

JEL Classification: C21, I21, J12, J24,

Keywords: family structure, divorce, education, double-machine learning

I. Introduction

The implications of family structure on the growth of children have been a central issue of research for numerous years. The traditional family, which consists of two biological parents cohabiting in the same house, has been recognized as the better environment to foster children's development in various aspects compared to alternative non-traditional family paradigms. These aspects include health and well-being, educational achievement, fertility choice, transmission of instability in relationships and marriages across generations, as well as future earnings and employment status (Paul R Amato and Keith, 1991; Amato, 1996, 2000; Amato and DeBoer, 2001).

Many researchers have directed their attention towards the challenge of interpreting these differences as causal, since family structure change is associated with other factors that also exert their influence on children's outcomes (Gruber, 2004; Sanz-De-Galdeano and Vuri, 2007; Amato, 2010).

In contrast to earlier studies that relied on the use of cross-sectional data, the availability of longitudinal studies has allowed researchers to adopt methodologies that account for the potential endogeneity of changes in family structure. The most extensively studied type of family structure change is parental divorce, whose effects has been analyzed using mainly difference-in-differences (DiD) and fixed-effect models. The DiD approach relies on observing the outcome of interest both before and after the transition occurs for at least two groups of children: untreated children, who grow up with stably married parents, and treated children, who experience a change in the family structure at some point in their lives. Fixed-effects models have the advantage of controlling for all time-invariant and unobservable characteristics.

When examining children's educational achievements, family structure change has been proved to negatively affect children's academic performance.

Ermisch *et al.* (2001) estimate the effects of experiencing a family disruption during childhood on several outcomes, including later educational attainment. Using the first five waves of the British Household Panel Survey (BHPS), they show that such event adversely affects the probability of achieving A-level qualification.[9] On average, children who go through a variation in the family structure experience a decrease in their likelihood of achieving an A-level qualification between 13.7 and 14.6 percentage points. Experiencing a family disruption in early childhood (ages 0-5) has the most disadvantageous consequences on later educational achievements.

Consistent with the prevailing literature and using data from the Family Survey Dutch Population 1998 and 2000, Fischer (2007) shows that parental divorce has negative effects on

[9] In the British education system, "A (Advanced) level" corresponds to education beyond high school, but short of a university degree. At least one A level is necessary to be admitted to a university.

children's education both in the short and in the long term. Children whose parents divorced achieve lower level of education both at age 15 and at age 30, -0.07 and -0.10, respectively compared to children not exposed to parental divorce.

Rasmussen (2009) delves into the impact of family structure shocks resulting from parental relationship dissolution on Danish children. Results indicate that children who experience family structure changes during childhood have poorer educational outcomes, as indicated by lower high school enrolment rates and lower high school grade point averages (GPA). Interestingly, the number of transitions appeared to be more influential on child outcomes than the mere fact of growing up in a single-parent household, with a higher number of family structure changes correlating with worse outcomes.

Using data from the Early Childhood Longitudinal Study, Kindergarten Cohort (ECLS-K) and the National Educational Longitudinal Study (NELS), Amato and Anthony (2014) find that parental divorce is significantly associated with a decline in scores on reading and mathematics tests, -0.145 and -0.126 SD respectively.

Similarly, Chen, Fan and Liu (2019) examine the impact of parental divorce occurring between the ages of 13 and 18 on educational achievement for the Taiwanese children. Their findings show that experiencing parental divorce resulted in a 10.6 percentage point decrease in the likelihood of university admission at the age of 18.

Sanz-De-Galdeano and Vuri (2007) employ double and triple differences models to estimate the impact of parental divorce on students' performance on standardized tests using data from the National Education Longitudinal Study of 1998 (NELS:88). In contrast to the studies discussed above, their results do not provide evidence of a negative causal effect of parental divorce on teenager's cognitive development, nor do they suggest that divorce has a more detrimental effect on teenagers if it occurred at a younger age.

Although in this field of research the use of cross-sectional data has raised some reservations regarding its interpretability in terms of causality, when it comes to educational outcomes, we often lack the possibility of employing sophisticated methodologies that rely on longitudinal data, as this type of data is not available. Except for outcomes derived from standardized tests, most of the time educational outcomes are observable only once "by definition", such as the grade of a school-leaving diploma or the decision to enrol in college. Consequently, this restricts the range of methodologies that can be implemented.

In addition, aware of the fact that families that experience a structural change may be a selected group, it is common to include covariates in cross-sectional estimations to control for the selection. If the coefficient for family structure change becomes statistically insignificant once additional explanatory variables are controlled for, this is interpreted as evidence for selection being the driver of the results depicted, and thus the effects cannot be interpreted in causal terms. Conversely, if the significance of the coefficient persist as many control variables are included in the model, then there is no evidence for selection being the underlying reason for the relationship between children's outcomes and family structure change.

More formally, within empirical economics, the causal effect of a parameter of interest is typically evaluated based on the *conditional mean assumption (CIA)* – also known as *selection-on-observable* (J.D. and Pischke, 2009). Conditional on observed characteristics, the parameter of interest is conditionally exogenous, meaning that it is uncorrelated with the error term. This requires the availability and plausibility of a large set of control variables to be used.

In general, one can refer to economic theory and existing literature to identify the set of variables needed to assume the assumption holds. As an alternative to this "ad-hoc" variable selection procedure, researchers can resort to recently developed machine learning tools which enable them to conduct causal analysis and perform valid statistical inference based on observational data. Double-Machine Learning (Chernozhukov *et al.*, 2018) facilitates an objective and data-driven selection of control variables, taking into account that the final goal of the researcher is to make causal inference on specific parameters of interest.

In this paper, we are interested in estimating the impact of experiencing a family structure change resulting from parental divorce on child's later educational attainments. Using the last two waves of Growing Up in Ireland (GUI) – Child Cohort-, we study whether such transitions affect the adolescent's probability of obtaining a high grade on their high-school diploma and their likelihood of enrolling in college at the age of 20.

The GUI – Child Cohort is a longitudinal study that tracks a sample of Irish children from the age of 9 until they reach 20 years old. The cohort consists of four waves of data collection, occurring at children's age of 9, 13, 17, and 20, respectively.

Two aspects need more attention: first, the data at our disposal does not allow to identify the exact timing of the structural change, but only whether it occurs between consecutive surveys. Second, the educational achievements are measured in the last data collection.

Taking these in mind, we proceed as follows. Initially, we exploit the longitudinal nature of our data to observe whether and at what age children in our sample experience a family structural change during their development trajectory up until the penultimate available period, that is until when they are 17 years old. Since the data allows us to identify only whether a structural change takes place between two periods, we intentionally choose not to consider any structural changes that occurs between the last two periods available, i.e. when adolescents are between the ages of 17 and

20. Secondly, because educational achievements are measured in the final period available, we employ the information gathered at the second to last survey, i.e. when adolescents are 17 years old, and the identified structural changes up until that point to estimate the effects of experiencing such family structure change on educational outcomes when the adolescent is 20 years old.

We conduct our analysis employing both the "ad-hoc" variable selection procedure and the "data-driven" feature selection approach. Our results show that adolescents who experience a family structure change are less likely to achieve a high grade on their school-leaving diploma and to enrol in college compared to adolescents who do not experience any structural changes.

Although there is growing attention towards the double machine learning estimator, its practical implementation is rather scarce (Knaus 2021; Skoufias and Vinha 2021; Vanneste and Gulati 2022). The literature is basically lacking papers with practical guidance on the use of such method in applied empirical studies. Furthermore, there is a need for further research to address practical challenges that hinder their adoption in the field of economics.

While recent papers used Machine Learning Algorithms to predict the determinants of divorce (Arpino, Moglie and Mencarini, 2018; Sohail *et al.*, 2018), to the best knowledge of the authors, this is the first paper that employs a double machine learning approach to estimate the impact of family structural change resulting from parental divorce on children's educational attainments at an older age.

The remainder of the paper is organized as follows. Section II provides a brief overview of the adoption of Machine Learning techniques in economic applications. Section III describes the data used in the empirical analysis. Section IV outlines the methodologies adopted. Section V and VI present the results derived from the "ad-hoc" and data-driven variable selection approaches, respectively, regarding the probability of obtaining a high grade at the high-school diploma and enrolling in college. Section VII concludes.

II. A short overview of Machine Learning and its applications in Economics

Machine learning (ML) algorithms have been developed in the field of computer science and statistics with the specific purpose of making predictions. In a typical supervised ML scenario, our objective is to predict the value of y – target variable in the ML jargon - based on a set of observed characteristics x – features. The set of data is generally divided into two subsets: the training dataset and the test dataset. The training dataset is used to build a prediction model, also known as a learner, that allows us to predict the outcome for new, unseen data points in the test dataset. A good learner is an algorithm that accurately predicts the target variable.

Unlike many economic applications and standard estimation problems, where the focal purpose reverses around unbiased parameter estimation, ML algorithms prioritize accurate predictions, aiming at minimizing the out-of-sample prediction error. This prediction error is influenced by the complexity of the model and follows a U-shaped pattern. Adding additional explanatory variables improves the fit of the model to the estimation training dataset and increases its prediction accuracy for out-of-sample data points up to a certain point. Once reached that point, the inclusion of new explanatory variables increases the fit of the model in the training set, but it also significantly decreases its prediction accuracy on the test dataset, leading to the so-called overfitting problem. To address this issue, ML tools rely on different regularization rules to prevent overfitting and determine an optimal level of model complexity for prediction.

The difference in the final objective, namely accurate predictions versus unbiased estimates, has restrained the adoption of ML methods in economics applications. Thanks to the contributions of prominent economists, the use of these methods has gradually gained popularity in the field of social sciences as well (Varian, 2014). Their primary application has been for prediction purposes (Kleinberg *et al.*, 2015; Andini *et al.*, 2018; Antulov-Fantulin, Lagravinese and Resce, 2021; Carrieri, Lagravinese and Resce, 2021; Resce and Vaquero-Piñeiro, 2022). More recently, the literature on causal machine learning has been enriched with new evidence on the use of such tools in counterfactual analysis (Varian, 2016; Cerqua *et al.*, 2021; de Blasio and D'ignazio, 2021; Resce, 2022; Resce and Vaquero-Piñeiro, 2023).

III. Data

The analysis draws on the Growing Up in Ireland (GUI), the National Longitudinal Study of Children and Young people in Ireland. The GUI study serves as a valuable source of data, with its primary objective being to provide information for government policies concerning children, young people, and families.

Our focus is on the Child Cohort study, a cohort of 8,568 nine-year olds children. This cohort, including the children and their families, has been surveyed at various intervals when the children were thirteen, seventeen/eighteen, and twenty years old. The first survey, known as the first wave, was conducted between August 2007 and May 2008. For the second wave, data collection occurred between August 2011 and March 2012, with 7,525 households completing the follow-up interviews. The third wave took place between April 2015 and August 2016, resulting in 6,216 completed cases. Lastly, the fourth wave was conducted between August 2018 and June 2019, resulting in a complete set of data with 5,190 cases.

In all the surveys, the primary caregiver, who is the individual providing the most care and possessing the most knowledge about the study child, filled in the home-based questionnaire. Additionally, the secondary caregiver, who is the spouse or partner of the primary caregiver, and the study child themselves also participated in filling out the questionnaire.

We recover information on household composition, socio-demographics, primary caregiver’s marital status, and child’s characteristics and educational outcomes from the main questionnaire completed by the primary caregiver and/or the child him/herself. Unfortunately, due to a significant number of missing values, information on the secondary caregiver was not included in our analysis.

The cleaning process of the dataset involves several criteria. First, we consider children whose families participated in all the four interviews to be able to track the evolution of the family structure over time. More important, our focus is on households where the primary caregiver's marital status is classified as married, separated, or divorced. We do not consider categories such as never married or widow as they do not align with our objective of identifying a change the family structure resulting from parental divorce, which leads to a shift from a two-parent household to a one-parent household. In addition, we combine the categories of separation and divorce into a single category.

To discern a change in family structure, we make use of the primary caregiver’s marital status reported at each interview. If there is a variation in the primary caregiver’s marital status between surveys, then a family structure change occurred. It should be noted that the available data does not allow us to recover the exact moment of the event, but only if it takes place between two surveys.

We distinguish four potential changes in family structure: 1) *absence of change*, where the primary caregiver consistently states their marital status as married throughout the entire period; 2) *already occurred change,* if at least one structure change had taken place prior to the first survey. This is assumed when the main caregiver's claimed marital status is divorced since the beginning of the study; 3) *structure change between 9-13,* in case we observe a change in the primary caregiver’s marital status from being married at the first wave - when the child is 9 years old – to being divorced in the second one – when the child is 13 years old; 4) *structure change between 13-17,* where the primary caregiver marital status is married in both the first and second surveys, and then shifts to divorced in the third one, when the adolescent is between the ages of 13 and 17.

For the purpose of our analysis, we leverage the longitudinal nature of the data to ascertain whether any of the aforementioned structural changes occurs within the timeframe of the surveys. However, our actual analysis restricts to the use of the third and fourth waves because the educational outcomes of interest are only measured in the final survey, when the adolescents reach the age of 20.

Our outcomes of interest are *adolescents' likelihood of attaining a high grade in their school-leaving diploma*[10], and their *likelihood of enrolling in college/university.* To measure these probabilities at the age of 20, we use the information collected during the third wave, when the individuals are 17 years old.

Given that we can only observe whether a structural change occurs between two surveys without knowing the exact moment, we do not consider any changes that occur between the third and fourth waves. This is due to our inability to distinguish, in that case, which event occurred first between the event of interest and the measured outcomes. Therefore, we only include structural changes in our model that we are certain took place prior to the adolescent achieving the educational measurements under consideration. In other words, the structural changes observed are predetermined with respect to the outcome variables, allowing us to infer their impacts.

The school-leaving grade is measured through an ordinal variable which takes on a value of 1 for grades below 200 points, a value of 2 for grades between 201 and 300, a value of 3 for grades between 301 and 400, a value of 4 for grades between 401 and 500, and a value of 5 for grades exceeding 500 points. We create our first dependent variable, namely *High grade*, as a dummy variable which takes the value of 1 if the adolescent obtains a grade higher than 400 points, and the value of 0 for lower grades.

Our second dependent variable, *College enrolment*, is also a dummy variable which takes the value of 1 if the young enrol in college/university, and the value of 0 otherwise.

Table 1 reports summary statistics for the variables used in the analysis. On average, about 61% of adolescents in our sample attain a school-leaving diploma grade higher that 400 points, with prevailing category falling within the range of 301-400 points. Roughly 79% of them enrol in college. The sample is balanced with respect to adolescent's gender: 49% of them are boys. The average birthweight is 3.54 kilograms, and approximately 56% of adolescents in our sample has been breastfed. Moreover, an average of 3% of the individuals were born from a twin birth. On average, only 3.2% of them do not hold Irish citizenship. [11] Each adolescent has on average 1.93 siblings.[12] Almost the totality of them, 96.5% attended an academically oriented secondary school. In terms of primary caregiver' s characteristics, it is found that, on average, 97% of them are women and they have an average age of 49 years. They obtain 14.72 years of education on

---

[10] It is called Leaving Certificate in the Irish educational system.

[11] Unfortunately, the data at hand do not allow to recover detailed information on young's nationality.

[12] The data provides limited information on siblings' characteristics. Indeed we could not recover either sibling's gender or age.

average.[13] In relation to their position in the labour market, approximately 71% of them are employed, 2% are unemployed, and 27% are classified as being out of the labour force.

Table 1. Descriptive Statistics

	Mean	SD	Min	Max	N
High Leaving Certificate grade	.61	.488	0	1	3353
Leaving Certificate total points	3.693	1.062	1	5	3353
College/university enrollment	.786	.41	0	1	3572
Boy	.489	.5	0	1	3572
Birthweight	3.532	.593	1.7	6.1	3516
Been breastfed	.565	.496	0	1	3555
Not Irish	.032	.175	0	1	3572
Twin	.032	.175	0	1	3572
Number of siblings	1.926	1.001	0	5	3572
Academically oriented secondary school	.965	.184	0	1	3526
Female pcg	.968	.176	0	1	3572
Pcg's age	48.669	4.66	34	58	3571
Pcg's years of education	14.715	2.22	8	18	3571
Pcg - Employed	.709	.454	0	1	3572
Pcg - Unemployed	.02	.14	0	1	3572
Pcg - Outside of labour force	.271	.444	0	1	3572
Annual income adj for inflation	16133.457	8493.082	5000	56000	3222
Predicted Income	16162.62	3692.327	362.957	26904.78	3570
Rural area	.542	.498	0	1	3572
Absence of change	.908	.288	0	1	3572
Already occurred change	.044	.206	0	1	3572
Structure change between 9-13	.022	.145	0	1	3572
Structure change between 13-17	.026	.158	0	1	3572

Notes: Dataset GUI - Child Cohort. The dependent variables (High Leaving Certificate grade, Leaving Certificate total points, College/university enrollment) are measured in wave 4. Control variables are measured in wave 3. For some variables the number of observations is slightly lower due to missing values.

In our estimations, we use two measures of income: the annual income reported by the primary caregiver, which we adjust for inflation using the Consumer Price Index, and the annual income predicted in the absence of any structural change.[14] This is done to address any potential endogeneity issue, following De Paola and Gioia (2017). The average annual income amounts to 16133 €, while the predicted income is, on average, 16163 €. Approximately 54% of the households in our sample reside in rural area. Lastly, with regards to household structure, about 91% of families do not undergo any structural changes, 4% of them experience at least one structural change before

[13] The GUI contains information on the primary caregiver's highest educational qualification obtained. We compute the number of years of education using the legal duration of several educational grades as follows: 8 for primary education or none (unfortunately the data do not allow to discern this information); 11 for lower secondary education (Junior Secondary School); 13 for upper secondary education (Senior Secondary School); 16 for non-degree e primary degree (the former refers to technical and vocational offered by Institute of Technology and specialized colleges, while the latter refers to Bachelor's Degree); 18 for postgraduate education (even in this case, we are not able to distinguish between Master's Degree or PhD Degree).

[14] We use as control variables primary caregiver's age and years of education, their occupational situation, area of residence, and family size.

the child reaches the age of 9 (i.e., prior to the initial survey), 2% of households experience a structural change when the child is between the ages of 9 and 13, and the remaining 3% of households go through a structural change when the child is between the ages of 13 and 17.

IV. Methodologies

In this paper we want to estimate the impact of a family structural change resulting from parental divorce on adolescent's educational achievements. To this end, we can estimate a classic regression model such as

$$Y_i = \beta d_i + f(\boldsymbol{W_i}) + \varepsilon_i, \qquad \text{with } E(\varepsilon_i \mid \boldsymbol{W_i}, d_i) = 0 \qquad (1)$$

where Y_i is the outcome of interest, d_i is a dummy variable which identifies whether one of the above-described structural change takes place and when, and $f(\boldsymbol{W_i})$ is a possibly nonlinear function of the confounding factors or "nuisance" control variables $\boldsymbol{W_i}$.

Conventional ad-hoc approach in the confounding variables selection

Traditionally, the choice of the control variables in economics and social sciences applications is determined in an "ad-hoc manner". This means that the researcher chooses the set of variables based on common sense, economic intuition, and/or well-established literature.

In this paper, we first estimate the relationship between the educational outcomes and the change in the structure of the family using this "ad-hoc approach" by assuming that a small set of control variables enter equation (1) linearly as follows:

$$Y_i = \beta d_i + \theta \boldsymbol{X_{ik}} + \varepsilon_i \qquad (2)$$

Note that the control variables in equation (2) are denoted by $\boldsymbol{X_{ik}}$ with $k = 1, \dots, K$ instead of $\boldsymbol{W_i}$ to explicitly clarify that we use a subset of selected control variables from the wider set of controls available in the dataset and denoted in equation (1) as $\boldsymbol{W_i}$.

The set of control variables selected in this "ad-hoc" approach and included in the regressions is made of 16 variables listed in Table 1 above.

In empirical applications, the coefficient of interest estimated has a causal interpretation if the conditional mean independence assumption (CIA) holds. The assumption requires that d_i is conditionally exogenous, which implies that d_i is as good as randomly assigned conditional on the covariates **X**. In other words, it requires that we observe all the variables that influence changes in family structure and the outcomes of interest simultaneously. Although one can rely on what is common in the empirical literature or on assumption and economic intuition to determine the subset of control variables to include in the model, there exists no definitive answer as to which variables should be part of the final set of controls. Given that only a limited number of controls are included in **X**, the estimated coefficient β might suffer from omitted variable bias if, for example, d_i is correlated with the error term ε_i, thereby violating the CIA assumption. Note also that we assume, as is commonly done in most applied papers, a linear functional form for the controls $\boldsymbol{X_{ik}}$, implicitly assuming that interactions and non-linearity are irrelevant. But, again, most of the time there is no theoretical justification for such assumption. Overall, one should be careful in interpreting the estimated coefficient from a regression model based on an "ad hoc" selection of few controls as causal estimate of the effect of d_i on Y_i , and instead prefer an interpretation in terms of association or partial correlation of Y_i and d_i.

A data-driven approach in the selection of control variables

With this consideration in mind, we resort to a specific ML technique that first performs a data-driven selection of variables from a wider set of control variables. Then, it allows to draw inferences on the parameter of a few variables of interest as an alternative to the ad-hoc procedure. Enlarging the set of controls to a much larger one reduces potential omitted variable bias, making the conditional mean assumption required for causal inference more reasonable than in the case of a few selected variables.

The approach employed in our analysis is based on the lasso regression, one of regularized regression models. Therefore, we start with an introduction of the lasso approach itself, followed by an explanation of how the lasso regression is used to make causal inference.

Regularized linear regression minimizes the sum of squared deviations between observed and model predicted values, imposing a regularization penalty aimed at limiting model complexity. The main advantage of regularized regression over least squares is rooted in the bias-variance trade-off. The prediction error can be decomposed into the unknown error variance (which cannot be reduced), the squared estimation bias and the variance of the predictor. As the complexity of the model increases, so does the variance of the estimated predictor, whereas the bias tends to decrease. By reducing model complexity and introducing a shrinkage bias, regularized regression tends to outperform OLS in terms of out-of-sample prediction performance, addressing the well-known problem of overfitting. Since the primary goal of regularized regression is prediction, its estimates cannot be interpreted as causal.

The LASSO (Least Absolute Shrinkage and Selection Operator) (Tibshirani, 1996) minimizes the mean squared error subject to a penalty on the absolute value of the coefficient estimates:

$$\beta_{lasso}(\lambda) = \arg\min \sum_{i=1}^{n}\left(y_i - \beta_0 - \sum_{j=1}^{p}\beta_j \boldsymbol{x}'_{ij}\right)^2 + \lambda\sum_{j=1}^{p}|\beta_j|$$

The Lasso can be seen as an OLS estimator at which a penalty λ on the absolute value of the coefficients is added. The tuning parameter λ, called *shrinkage penalty*, has the effect of shrinking the estimates of β_j towards zero. When $\lambda = 0$, the penalty has no effect, and the lasso regression will simply produce least squares estimates. When λ is sufficiently large, it has the effect of forcing (shrinking) some of the coefficient estimates to be exactly zero to satisfy the constraint. Therefore, λ has a direct impact on the size of the model. For this reason, the lasso has the additional advantage of serving as a ***variable selection model***: it shrinks the variables with little or no predictive power to zero.

The choice of the tuning parameter plays a crucial role and is often accomplished using cross-validation. The general aim of cross-validation is to assess the performance of a model on unseen data. For this purpose, the available data is split into two distinct sets: a training and validation data sets. The models are fit to the training data and their predictive performance is assessed on the validation set. In the context of regularized regression, the aim of cross-validation is to select λ that optimize the out-of-sample prediction performance, that is the smaller out-of-sample mean squared prediction error. The most popular method used is K-fold cross validation (Geisser, 1975), according to which the data is randomly divided into K groups, or folds, of approximately equal size. The first fold is treated as a validation set and the model is fit on the remaining K-1 folds. The procedure is repeated for each fold so that every data point is used for validation once. The mean squared prediction error for each fold is computed as

$$MSPE_k(\lambda) = \frac{1}{n_k}\sum_{i\in K_k}(y_i - \boldsymbol{x}'_i\hat{\beta}_k(\lambda))^2$$

The K-fold cross-validation estimate of the MSPE is computed by averaging these values:

$$\hat{f}^{CV} = \frac{1}{K}\sum_{k=1}^{K} MSPE_k(\lambda)$$

The choice of K in general falls between 5 and 10, arguing that the performance of CV rarely improves for K larger than 10 (Arlot and Celisse, 2010).

In conjunction with model selection purpose, Lasso is also suitable for drawing inferences about the parameters of the model. However, one should pay attention when the goal is causal inference.

A naïve approach consists in using the set of variables selected by LASSO and conducting inference about parameters of interest using ordinary least squares. The core of the problem in this procedure is rooted in the predictive task of the regularized regressions, which is what they are designed for. Because Lasso minimizes the prediction error subject to the constraint of limiting model complexity and optimize the out-of-sample prediction accuracy as explained above, it might omit covariates with small, but nonzero coefficients. Firstly, this can result in errors in the model selection. Those mistakenly omitted variables can be cause of omission bias for other variables selected by the model if there is a correlation between them.

Furthermore, the model selection that is performed is based on a single sample. Drawing repeated samples from a population and performing variables selection on each one would result in different sets of covariates in each sample. Standard errors from ordinary least square do not account for the sample-to-sample variability in the variable selection process. Therefore, the problem in following this naïve approach is twofold.

In contrast to this naive procedure, more sophisticated and robust approaches have been developed: Double-selection (Belloni, Chernozhukov and Hansen, 2014), Partialing out (Belloni *et al.*, 2012), and Cross-fit partialing out (Chernozhukov *et al.*, 2018). These methods can be applied to linear, logistic and Poisson models and are all applicable in Stata using the inferential lasso models (StataCorp LLC, 2023).

In our study, we use the Cross-fit partialing out approach, also known as Double Machine Learning (DML) approach. It makes use of the technique of sample splitting to overcome the ignored sample-to-sample variability in the variable selection task when it is performed on a single sample.

The approach provides an estimate of the parameter of interest in each subsample, and subsequently, the resulting estimators are averaged. Estimating the parameter of interest in a single subsample result in a loss of efficiency as we only use a subset of the available data. By flipping the role of training and auxiliary samples to obtain multiple estimates and averaging the results, we can obtain additional information about the bias and efficiency of the fitted model, thereby enabling us to regain full efficiency. This sample-splitting procedure is called 'cross-fitting'.

In what follows, we present the algorithm using two samples, but it works splitting the sample in K=10 parts.

Having in mind equation (1) from above as our starting point, the DML approach involves the following steps:

Step 1: Split the data into roughly k = 2 equal-sized subsamples.

Step 2: In the first sample:

- Fit a lasso regression predicting d_i on W_{ik} :

$$d_i = \boldsymbol{W}'_{ik} Q_{ik} + u_i$$

Let $\tilde{W}_{d1}$ be the variables with non-zero coefficients selected.

- Fit a linear regression of d_i on $\tilde{W}_{d1}$:

$$d_i = \tilde{\boldsymbol{W}}'_{d1} Q_{d1} + u_i$$

- Fit a lasso regression predicting Y_i on W_{ik} :

$$Y_i = \boldsymbol{W}'_{ik} Q_{ik} + v_i$$

Let $\tilde{W}_{y1}$ be the covariates with non-zero coefficients selected.

- Fit a linear regression of Y_i on $\tilde{W}_{y1}$:

$$Y_i = \tilde{\boldsymbol{W}}'_{y1} Q_{y1} + v_i$$

Step 3: Swap to the second sample: use the coefficients of the covariates selected in the previous step in the first sample to construct the "residualized" or partialled-out $\tilde{d}_i$ and $\tilde{Y}$ as follows:

- Substitute $\tilde{d}_i = d_i - \tilde{\boldsymbol{W}}'_{d1} \hat{Q}_{d1}$
- Substitute $\tilde{Y}_i = Y_i - \tilde{\boldsymbol{W}}'_{y1} \hat{Q}_{y1}$

Step 4: Still in the second sample:

- Fit a lasso regression predicting d_i on W_{ik} :

$$d_i = \boldsymbol{W}'_{ik} Q_{ik} + u_i$$

Let $\tilde{W}_{d2}$ be the variables with non-zero coefficients selected.

- Fit a linear regression of d_i on $\tilde{W}_{d2}$:

$$d_i = \tilde{\boldsymbol{W}}'_{d2} Q_{d2} + u_i$$

- Fit a lasso regression predicting Y_i on W_{ik} :

$$Y_i = \boldsymbol{W}'_{ik} Q_{ik} + v_i$$

Let $\tilde{W}_{y2}$ be the covariates with non-zero coefficients selected.

- Fit a linear regression of Y_i on $\tilde{W}_{y2}$:

$$Y_i = \tilde{\boldsymbol{W}}'_{y2} Q_{y2} + v_i$$

Step 5: Swap to the first sample: as before, use the coefficients of the covariates selected in the above step in the second sample to construct the partialled-out $\tilde{d}_i$ and $\tilde{Y}$ as follows:

- Substitute $\tilde{d}_i = d_i - \tilde{\boldsymbol{W}}'_{d2} \hat{Q}_{d2}$
- Substitute $\tilde{Y}_i = Y_i - \tilde{\boldsymbol{W}}'_{y2} \hat{Q}_{y2}$

Step 6: In the full sample: Fit a linear regression of $\tilde{Y}$ on $\tilde{d}$ $\tilde{Y}_i = \alpha \tilde{d}_i$

The estimate of $\hat{\alpha}$ from the above regression is then the coefficient on $\tilde{d}_i$

V. Family structural change and children's secondary school performance

In this section, we examine the relationship between changes in family structural and the probability of achieving a high grade on the high-school diploma, referred to as Leaving Certificate in Ireland. We first present the results using the "ad-hoc" set of selected control variables, followed by the ones resulting from the implementation of the DML approach.

Our dependent variable is the dummy variable *High grade*, which takes the value of 1 if the adolescent gets a grade exceeding 400 points, and 0 otherwise. Considering the binary nature of the outcome of interest, we first estimate a linear probability model (LPM), and then a logit regression model. The latter model, overcoming the well-known limitations of the former, provides a more realistic value for the probability (between 0 and 1) and model a non-linear relationship between the probability and the explanatory variables.

In the case of the LPM, we use the following regression model:

$$P(High\ Leaving\ Certificate\ Grade_{it} = 1 \mid X_{it-1}) = \beta_0 + \beta_1 AlreadyOccuredChange_i + \beta_2 StructureChange\ 9-13_i + \beta_3 StructureChange\ 13-17_i + \gamma X_{it-1} + \varepsilon_{it}$$

where P denotes the likelihood for adolescent in household i of getting a high grade at time t, that is at age 20; $Already\ Occured\ Change_i$ is a dummy variable which takes the value of 1 if a structural change in family i had already taken place before the child turned 9, and 0 if no change occurred; $Structure\ Change\ 9-13_i$ is a dummy variable taking the value of 1 if child in household i experiences a family structure change between the ages of 9 and 13, and 0 otherwise; $Structure\ Change\ 13-17_i$ is a dummy variable which takes the value of 1 if the structural change is experienced between 13 and 17 years old, and 0 otherwise; X_{t-1} is a vector of control variables all measured at time t-1, i.e. when the adolescent is 17 years old, including primary caregiver's, adolescent's, and household's characteristics; ε_{it} is an i.i.d. normally distributed stochastic error term.

In the case of the logit model, the regression is the following:

$$P(High\ Leaving\ Certificate\ Grade_{it} = 1 \mid X_{it-1}) = f(Z)$$

where $f(Z) = \frac{e^z}{1+e^z}$ is the logistic distribution function and $Z = \beta_0 + \beta_1 AlreadyOccuredChange_i + \beta_2 StructureChange\ 9-13_i + \beta_3 StructureChange\ 13-17_i + \gamma X_{it-1} + \varepsilon_{it}$.

Table 2 below shows the linear probability estimates for five different specifications.[15] The first column examines the effect of having experienced a family structure change before the age 9,

[15] Note that the sample size differs among specifications because of some missing values in some controls.

between the ages of 9 and 13, and between ages 13 and 17 on the likelihood of getting a high grade, without including other explanatory variables. We find a strong negative relationship between family structural changes and educational performance: those who experiences a change in their family composition have a lower probability of attaining a high grade at their Leaving Certificate by on average 21 percentage points (p.p. thereafter) if the event occurs before turning 9, by 31 p.p. if the event is experienced between the ages of 9 and 13, and by almost 15 p.p. if the transition occurs between 13 and 17 years old, relative to adolescents who do not encounter any structural changes. The effects are statistically significant at the 1 percent level.

These first results are important not only because they provide evidence of a detrimental impact of family structure change on educational outcomes, but they show that these effects considerably persist over time. Although we do not observe exactly when each of the event considered takes place, concerning the changes that occurs before the age of 9, and between the ages of 9 and 13, we can affirm that their consequences remain relevant after at least 11 and 7 years respectively (referring to the time span between the survey in which we identify the structural change and the latest survey in which the educational outcome is measured).

In the second column we add as controls some adolescent's characteristics, such as gender, birthweight used as proxy for unobserved ability and later-life children's outcomes (Abufhele, Behrman and Bravo, 2017; Hassen *et al.*, 2021), number of siblings, primary caregiver's age and level of education, and annual income. When controlling for these variables, the results confirm the negative association between family structure changes and educational attainments.

The effects are still statistically significant at the 1 percent level for the first two structural changes, although slightly lower in magnitude: experiencing a structural change before the age of 9 causes a reduction in the likelihood of obtaining a high grade by on average 16 p.p.; a reduction of 26 p.p. is observed if the change occurs between the ages of 9 and 13 relative to adolescents whose family structure remains stable. The transition that takes place when the adolescent is between 13 and 17 years of age is weakly statistically significant with a p-value of 0.12: for these adolescents, the likelihood of achieving a high grade decreases on average by 7 p.p. Boys have a lower probability of obtaining a high grade by on average 7 p.p.; birthweight exerts a positive and statistically significant effect, increasing the likelihood of interest by nearly 2.5 p.p.; having one more sibling raises the likelihood by 1.84 p.p. on average. Both the age and level of education of the primary caregiver bear positive and significant effects, resulting in an increase in the likelihood of obtaining a high grade by on average 1 and 5 p.p., respectively, in response to a unitary increase in the caregiver's age and education. The effect of an increase in the income level is statistically significant and positive, albeit scarcely detectable.

Table 2. Linear Probability Model. High-Grade Leaving Certificate

	(1)	(2)	(3)	(4)	(5)
Already occurred change	-0.209***	-0.162***	-0.184***	-0.152***	-0.158***
	(0.0432)	(0.0432)	(0.0441)	(0.0433)	(0.0455)
Structure change between 9-13	-0.310***	-0.260***	-0.269***	-0.268***	-0.250***
	(0.0593)	(0.0587)	(0.0593)	(0.0590)	(0.0612)
Structure change between 13-17	-0.146***	-0.0694	-0.0871	-0.0701	-0.0700
	(0.0562)	(0.0573)	(0.0567)	(0.0589)	(0.0578)
Boy		-0.0745***	-0.0663***	-0.0747***	-0.0662***
		(0.0168)	(0.0161)	(0.0166)	(0.0160)
Birthweight		0.0250*	0.0256*	0.0226	0.0230
		(0.0141)	(0.0135)	(0.0147)	(0.0141)
Been breastfed			0.0872***	0.0694***	0.0839***
			(0.0179)	(0.0187)	(0.0179)
Not Irish			0.00184	0.0173	0.000411
			(0.0505)	(0.0499)	(0.0505)
Twin				-0.0195	-0.0353
				(0.0512)	(0.0489)
Number of siblings		0.0184**	-0.0127	0.0132	-0.0190*
		(0.00860)	(0.00960)	(0.00870)	(0.0105)
Academically oriented secondary school				0.269***	0.266***
				(0.0476)	(0.0472)
Female pcg		-0.0300	-0.0345	-0.0321	-0.0362
		(0.0478)	(0.0455)	(0.0480)	(0.0459)
Pcg's age		0.0119***	0.0627**	0.0481*	0.0637**
		(0.00188)	(0.0271)	(0.0277)	(0.0271)
Pcg's age squared			-0.0499*	-0.0383	-0.0484*
			(0.0279)	(0.0287)	(0.0281)
Pcg's years of education		0.0515***	0.0716***	0.0467***	0.0839***
		(0.00399)	(0.0102)	(0.00435)	(0.0131)
Employed				-0.0290	0.00971
				(0.0198)	(0.0236)
Unemployed				-0.0187	-0.139*
				(0.0664)	(0.0730)
Predicted Income			-0.0130*		-0.0239**
			(0.00708)		(0.0101)
Annual income adjusted for inflation		0.00857***		0.00838***	
		(0.00102)		(0.00102)	
Rural area			-0.0112	0.0406**	-0.0258
			(0.0207)	(0.0169)	(0.0252)
Constant	0.628***	-0.920***	-2.130***	-1.972***	-2.450***
	(0.00873)	(0.119)	(0.653)	(0.665)	(0.655)
R-squared	0.0159	0.142	0.123	0.157	0.132
Observations	3353	2971	3304	2970	3303

Notes: The Table reports Linear probability model estimates. Sample: GUI dataset, children aged 20 years old. The dependent variable is *High Leaving Certificate Grade*, which takes the value of 1 if the young get a grade higher than 400, and 0 otherwise. Standard errors, reported in parentheses, are corrected for heteroskedasticity, and allow for clustering at the household level. To make easier the interpretation of the coefficients, in all the regressions we have rescaled the variables *Income,* dividing it by 1000, and *Age2,* dividing it by 100, thus reducing the number of zeros after the decimal points. The symbols ***, **, * indicate that the coefficients are statistically significant at the 1, 5, and 10 percent level respectively.

In column (3) we include additional adolescent's characteristics, area of residence dummy, and primary caregiver's age squared. Results reaffirm the previously observed negative effects associated with structural changes. Note that in this specification, we use the income predicted in the absence of any structural change. The adverse effects of structural transitions persist and remain statistically significant at the 1 percent level. Specifically, experiencing the change in the family structure prior to the age of 9 and between the ages of 9 and 13 leads to a decrease of 18 and 27 p.p., respectively, in the likelihood of attaining a high grade relative children not exposed to such change.

The fourth specification enriches the analysis by including the labour market situation of the primary caregiver among controls. Moreover, we include a variable which controls for the type of secondary school the adolescent attended The negative effect of experiencing a change in family structure remains significant: the probability of achieving a high grade decreases by on average of 15 p.p. if the change occurs before the age of 9, and by 27 p.p. if it takes place between the ages of 9 and 13, compared to adolescents in intact families. A transition in family structure between the ages of 13 and 17 does not exert any statistically significant effect. Other control variables retain the same effects, except for the number of siblings which turns statistically insignificant. Primary caregiver's occupational status seems to not significantly influence adolescent's grade. Adolescents who attend an academically oriented secondary school have a greater probability of obtaining a top grade at their diploma by on average 27 p.p. relative to adolescents who attend a vocational/technical oriented school.

Finally, in column (5) we replace the actual income with the predicted one. The effects are robust to the inclusion of such controls: family structural changes retain their negative effects on education. On average, those who experience such change before turning 9 have a lower probability of getting a high grade by 16 p.p. Similarly, the probability drops by 25 p.p. if the transition occurs between 9 and 13 years of age, relative to adolescent with unchanged family structure.

In Table 3 we report the logit model estimates of the adolescent's probability of obtaining a high grade at their high-school diploma.

As before, we report five different specifications controlling for an increasing number of control variables each time. The reported coefficients represent the average marginal effects of each regressor on the probability of interest. In all the specifications, the results confirm the detrimental effect of experiencing a family structure change either before the age of 9 and between 9 and 13 years olds in the probability of achieving a high grade. The effects are statistically significant at the 1 percent level. The structure transition that takes place between 13 and 17 years of age turns statistically insignificant. On average, adolescents who face such a variation in their family structure before turning 9 see their probability decreasing between 16 and 20 p.p., while those for which the variation takes place between 9 and 13 years of age, the decrease in the likelihood ranges between 28 and 31 p.p. relative to adolescents who grow up within a permanent two-parent family structure. The impacts of the control variables are roughly the same, with boys showing a lower probability of getting a high grade; having been breastfed as well as primary caregiver's age and level of education and having attended an academically oriented school exert their positive effects.

Table 3. Logit Estimates. High Leaving Certificate Grade

	(1)	(2)	(3)	(4)	(5)
Already occurred change	-0.202***	-0.169***	-0.198***	-0.162***	-0.172***
	(0.0423)	(0.0472)	(0.0481)	(0.0479)	(0.0500)
Structure change between 9-13	-0.306***	-0.288***	-0.299***	-0.302***	-0.281***
	(0.0650)	(0.0772)	(0.0729)	(0.0777)	(0.0753)
Structure change between 13-17	-0.142***	-0.0673	-0.0921	-0.0710	-0.0746
	(0.0536)	(0.0604)	(0.0601)	(0.0631)	(0.0623)
Boy		-0.0857***	-0.0751***	-0.0870***	-0.0752***
		(0.0193)	(0.0181)	(0.0195)	(0.0182)
Birthweight		0.0280*	0.0290*	0.0264	0.0263
		(0.0161)	(0.0153)	(0.0171)	(0.0161)
Been breastfed			0.0941***	0.0757***	0.0915***
			(0.0191)	(0.0206)	(0.0192)
Not Irish			-0.00344	0.0192	-0.00540
			(0.0571)	(0.0572)	(0.0573)
Twin				-0.0173	-0.0396
				(0.0586)	(0.0541)
Number of siblings		0.0227**	-0.0147	0.0169*	-0.0217*
		(0.00994)	(0.0108)	(0.0102)	(0.0118)
Academically oriented secondary school				0.312***	0.301***
				(0.0658)	(0.0605)
Female pcg		-0.0347	-0.0404	-0.0370	-0.0442
		(0.0555)	(0.0522)	(0.0564)	(0.0536)
Pcg's age		0.0132***	0.0638**	0.0478	0.0653**
		(0.00213)	(0.0310)	(0.0328)	(0.0314)
Pcg's age squared			-0.0495	-0.0367	-0.0482
			(0.0321)	(0.0342)	(0.0327)
Pcg's years of education		0.0552***	0.0788***	0.0515***	0.0933***
		(0.00472)	(0.0114)	(0.00510)	(0.0150)
Employed				-0.0404*	0.0101
				(0.0230)	(0.0265)
Unemployed				-0.0172	-0.152*
				(0.0726)	(0.0800)
Predicted Income			-0.0148*		-0.0270**
			(0.00780)		(0.0114)
Annual income adjusted for inflation		0.0112***		0.0112***	
		(0.00151)		(0.00156)	
Rural area			-0.0144	0.0462**	-0.0305
			(0.0232)	(0.0198)	(0.0285)
P-seudo R-squared	0.012	0.114	0.096	0.127	0.103
Observations	3353	2971	3304	2970	3303

Notes: Table reports marginal effects from logit estimates. Sample: GUI dataset, children aged 20 years old. The dependent variable is *High Leaving Certificate Grade*, which takes the value of 1 if the young get a grade higher than 400, and 0 otherwise. Standard errors, reported in parentheses, are corrected for heteroskedasticity, and allow for clustering at the household level. To make easier the interpretation of the coefficients, in all the regressions we have rescaled the variables *Income,* dividing it by 1000, and *Age2,* dividing it by 100, thus reducing the number of zeros after the decimal points. The symbols ***, **, * indicate that the coefficients are statistically significant at the 1, 5, and 10 percent level respectively.

As an additional analysis, we study the effect of family structure changes on adolescent's education making use of the original ordinal variable, *Leaving Certificate Total Points*. It takes the value of 1 for a final grade lower than 200 points, value of 2 for a total point between 201 and 300, value of 3 if between 301-400, value of 4 if between 401-500, and the value of 5 for a grade higher than 500 points.

Table 4. Ordered Logit Models. Leaving Certificate Grade

	(1)	(2)	(3)	(4)	(5)
Already occurred change	.425***	.494***	.444***	.49***	.492***
	(.064)	(.08)	(.075)	(.08)	(.083)
Structure change between 9-13	.285***	.342***	.315***	.321***	.346***
	(.062)	(.082)	(.073)	(.074)	(.082)
Structure change between 13-17	.508***	.697	.659*	.7	.716
	(.109)	(.159)	(.15)	(.166)	(.165)
Boy		.741***	.75***	.737***	.747***
		(.051)	(.049)	(.051)	(.048)
Birthweight		1.114*	1.126**	1.117*	1.126**
		(.064)	(.061)	(.068)	(.064)
Been breastfed			1.469***	1.379***	1.464***
			(.103)	(.103)	(.103)
Not Irish			.99	1.128	.977
			(.199)	(.234)	(.199)
Twin				1.026	.951
				(.223)	(.19)
Number of siblings		1.078**	.938*	1.05	.898**
		(.037)	(.036)	(.037)	(.038)
Academically oriented secondary school				3.58***	3.85***
				(.727)	(.765)
Female pcg		.983	.955	.981	.948
		(.205)	(.184)	(.206)	(.188)
Pcg's age		1.056***	1.402***	1.304**	1.402***
		(.008)	(.153)	(.149)	(.154)
Pcg's age squared			.757**	.801*	.767**
			(.085)	(.096)	(.088)
Pcg's years of education		1.266***	1.415***	1.239***	1.52***
		(.021)	(.061)	(.022)	(.082)
Employed				.862*	1.084
				(.069)	(.103)
Unemployed				.805	.452***
				(.23)	(.138)
Predicted Income			.934**		.877***
			(.027)		(.035)
Annual income adjusted for inflation		1.047***		1.045***	
		(.005)		(.005)	
Rural area			.861*	1.115	.785**
			(.072)	(.079)	(.08)
P-seudo R-squared	0.007	0.072	0.062	0.079	0.067
Observations	3353	2971	3304	2970	3303

Notes: Ordered logit model estimates. Sample: GUI dataset, children aged 20. The dependent variable, *Leaving Certificate Grade* equals 1 for grade lower than 200, 2 for grade between 201-300, 3 if between 301-400, 4 if between 401-500, and 5 if higher than 400. Standard errors, reported in parentheses, are corrected for heteroskedasticity, and allow for clustering at the household level. To make easier the interpretation of the coefficients, we have rescaled the variables *Income* and *Age2,* dividing by 1000 and 100, respectively, reducing the number of zeros after the decimal points. The symbols ***, **, * indicate that the coefficients are statistically significant at the 1, 5, and 10 percent level.

Table 4 reports ordered logit model estimates for five specifications using the same set of control variables employed before. The reported coefficients are the ordered odds ratios which represents the impact on the dependent variable of one unit increase in the explanatory variable. Since we are not interest in this type of interpretation, in this context we only consider the direction of the effects of the regressors of our interest and whether they are statistically significant.

In all the specifications, family structure changes occurred before turning 9 and between the ages of 9 and 13 are statistically significant at the 1 percent level and the odds ratios are lower than 1, meaning that for adolescents experiencing such transitions the likelihood of falling into the lowest category (that is, getting the lowest grade) increases, while the probability of being into the highest category (i.e., getting the highest grade) decreases in comparison to adolescents who do not experience such family structure change. The effect of experiencing a structure change between 13 and 17 years of age turns to be weakly significant in column (2), p-value equals 0.114, and statistically insignificant in the last two columns, p-values of 0.133 and 0.147 respectively.

Table 5 shows the marginal effects from the ordered logit model in Table 4 only for the specification in column (5), which includes all the explanatory variables, and for the parameters of interest: *already occurred change, structural change between 9-13,* and *structural change between 13-17*.

Table 5. Marginal Effects from Ordered Logit Model. Leaving Certificate Grade

	Already occurred change	Structure change between 9-13	Structure change between 13-17
Grade:			
200 or less	0.0208***	0.0304***	0.00973
	(0.00531)	(0.00690)	(0.00668)
201 - 300	0.0527***	0.0768***	0.0246
	(0.0132)	(0.0171)	(0.0168)
301 - 400	0.0784***	0.114***	0.0366
	(0.0194)	(0.0252)	(0.0250)
401 - 500	-0.0316***	-0.0460***	-0.0147
	(0.00826)	(0.0108)	(0.0102)
500 or more	-0.120***	-0.175***	-0.0562
	(0.0295)	(0.0381)	(0.0383)
P-seudo R-squared			0.067
Observations			3303

Notes: The Table reports marginal effects only for the structural changes of interest from an ordered logit model. Sample: GUI dataset, children aged 20 years old. The dependent variable is *Leaving Certificate Grade*, which takes the value of 1 if the young get a grade lower than 200, the value of 2 if s/he gets a grade between 201 and 300, the value of 3 for a grade between 301 and 400, the value of 4 for a grade between 401 and 500, and the value of 5 if s/he gets a grade higher than 500. Standard errors, reported in parentheses, are corrected for heteroskedasticity, and allow for clustering at the household level. The symbols ***, **, * indicate that the coefficients are statistically significant at the 1, 5, and 10 percent level respectively.

Adolescents who experience the structural change before turning 9 present the following results: an increase of nearly 2.1 p.p. in the probability of obtaining the lowest grade (less than 200 points), an increase of 5.3 p.p. in the likelihood of obtaining a grade between 201-300 (next to last), and 7.8 p.p. more for the probability of falling into the third category, grade between 301-400 points. Conversely, the likelihood of getting the highest grades decreases on average by 3.2 and 12 p.p. for a score between 401-500 and higher than 500 points, respectively, relative to adolescents who do not have experience of these changes in their family structure.

Those exposed to the structural change when aged 9-13 have a higher probability of, on average, 3 p.p. of obtaining the lowest grade, of almost 7.7 p.p. of falling into the second to last category, that is a grade between 201 and 300 points, 11.4 p.p. more of getting a grade between 301 and 400 points (third category). Accordingly, the likelihood of getting a grade between 401-500 points decreases on average by 4.6 p.p., and the probability of getting the highest grade, more than 500 points, decreases by 17.5 p.p. relative to adolescents not exposed to household structural variations. Statistically insignificant effects are found for the effects of the structural change experienced between the ages of 13 and 17.

All things considered, our results from the estimations in which we use the "ad-hoc" chosen set of controls show a negative effect of undergoing a change in the family structure on adolescent's educational attainment and provide evidence of a significant persistence of these effects over time: adolescents whose families shift from being a two-parent type to a one-parent type both before the age 9 and between the ages of 9 and 13 exhibit a lower probability of getting a high grade in all the models estimated. The magnitude of the effects is greater when the event takes place between 9 and 13 years than when it takes place before 9 years of age, suggesting a more pronounced impact when the children approach their adolescent phase coinciding with the occurrence of the change. At the same time, the results suggest that the effects are improbable to dissolve.

Experiencing a structure change between the ages of 13 and 17 becomes statistically insignificant as a greater number of control variables are included in the model. This might suggest either that such transition at an older age does not detrimentally affect adolescent's educational outcomes or that it might take time for such transition to produce its effects.

In Table 6 we present the results of the impact of changes in the family structure on adolescent's educational attainment, but instead of using the subset of "ad-hoc" selected control variables as done above, we let the DML algorithm to identify, among a larger set of controls, the variables needed for estimating the effects of interest. We report three specifications of the DML models: the first consists of a set of 55 potential controls; the subsequent two models exponentially increase the set of potential controls including interactions terms and polynomials up to the second order, respectively.

As a preliminary step, we estimate a simple lasso logit model with the exclusive aim of performing features selection, that is determine the variables the most contribute to the educational outcomes' prediction. This step is based on our intuition that family structural changes might play a significant role in predicting the educational outcomes of the adolescents. As supposed, our intuition finds support: the three family structural changes are recognized as important variables among the set of selected features.[16] Results are not provided.

To draw causal inference conclusions, we employ the cross-validation partialing out lasso logit model (or DML). One of the disadvantages of the model, at least when using the package available in Stata, is the absence of estimates for the control variables selected by the model. It only displays estimated coefficients, standard errors, and test statistics for the parameters of interest specified in the model. The reported coefficients are the odds ratios of the variables for which we seek to establish causality. In addition, column (1) displays odds ratios from a logistic model in which we include all the "ad-hoc" selected controls used in the previous estimations.

Table 6. Double-Machine Learning Estimates: High Leaving Certificate Grade

	Logit	DML	p-value	DML	p-value	DML	p-value
Already occurred change	0.481***	0.566***	0.014	0.708	0.157	0.731	0.200
	(0.102)	(0.130)		(0.172)		(0.179)	
Structure change between 9-13	0.303***	0.331***	0.001	0.340***	0.004	0.342***	0.004
	(0.097)	(0.111)		(0.126)		(0.127)	
Structure change between 13-17	0.728	0.753	0.308	0.797	0.448	0.817	0.500
	(0.193)	(0.21)		(0.238)		(0.245)	
Interaction terms		NO		YES		YES	
Second order polynomials		NO		NO		YES	
# potential controls		55		4584		4590	
# selected controls	16	15		339		324	
Observations	3303	2531		2489		2488	

Notes: The Table reports the estimated odds ratios using Logit model in column (1) and Cross-fit partialing out logit in the following columns. Sample: GUI dataset, children aged 20 years old. The dependent variable is *High Leaving Certificate Grade*, which takes the value of 1 if the young enrolled in college/university, and 0 otherwise. The variables of interests are *Already occurred change*, which takes the value of 1 if the event took place before the child turned 9, and 0 otherwise; *Structure change between 9-13* , which takes the value of 1 if such change took place between 9 and 13 years of age, and 0 otherwise; *Structure change between 13-17*, which takes the value of 1 if the transition took place between 13 and 17 years of age, and 0 otherwise. Standard errors, reported in parentheses, are corrected for heteroskedasticity, and allow for clustering at the household level. The symbols ***, **, * indicate that the coefficients are statistically significant at the 1, 5, and 10 percent level respectively.

The results from the DML approach are fully consistent with the ones obtained using the logit model only in the case of a "naïve" specification, that does not include any interactions or polynomials, displayed in column 2. Results from this model provide a statistically significant impact of family structure changes occurring before the age of 9 and between the ages of 9 and 13: adolescents who experience such changes have a lower probability of obtaining a high grade on

[16] Note that we perform two lasso regressions using two different approaches to identify the optimal parameter λ: cross-validation and adaptive method. The first approach determines a set of 27 variables; the second approach identifies a set of 24 variables as important ones for the prediction task.

average by 43 and 67 p.p., respectively, than those who do not experience any structural changes. Experiencing a change in family structure between the ages of 13 and 17 does not negatively affect adolescent's educational achievement.

When the number of potential control variables increases exponentially exceeding the sample size (columns 3 and 4), the family structure change that takes place before 9 years old turns statistically insignificant. The effect of experiencing such a change between 9 and 13 years old remains statistically significant at the 1 percent level and nearly identical in magnitude. This structural change is associated with a decrease in likelihood of achieving a high grade on the final exam of approximately 66 p.p. Experiencing such event between 13 and 17 years old does not exert any statistically significant effects, similar to the previous model.

VI. Family structural change and children's college enrolment

In this section we estimate the impact of family structure changes on adolescent's likelihood of enrolling in university. Consistent with our previous approach, we begin by presenting the results from the models estimated using the "ad-hoc" set of selected control variables, and subsequently, the ones resulting from the employment of the DML methodology.

Our dependent variable is *College enrolment*, a dummy variable which takes the value of 1 if the adolescent enrols in college, and 0 otherwise. The binary nature of the dependent variable leads us to estimate a linear probability model (LPM) and a logit regression model as done previously.

In the case of the LPM, we use the following regression model:

$$P(College\ enrolment_{it} = 1 \mid X_{it-1}) = \beta_0 + \beta_1 AlreadyOccuredChange_i + \beta_2 StructureChange\ 9-13_i + \beta_3 StructureChange\ 13-17_i + \gamma X_{it-1} + \varepsilon_{it}$$

where P denotes the adolescent's likelihood of enrolling in college at time *t*, that is at age 20; $Already\ Occured\ Change_i$ is a dummy that takes the value of 1 if in household *i* a structural change took place before the child turns 9, and 0 otherwise; $Structure\ Change\ 9-13_i$ is a dummy variable taking the value of 1 if adolescent in household *i* experienced a family structure change between the ages of 9 and 13, and 0 otherwise; $Structure\ Change\ 13-17_i$ is a dummy variable taking the value of 1 if in household *i* the structural change takes place between the ages of 13 and 17 of the adolescent, and 0 otherwise; X_{t-1} is a vector of control variables all measured at time *t*-1, i.e. when the adolescent is 17 years old, including primary caregiver's, adolescent's, and household's characteristics; ε_{it} is an i.i.d. normally distributed stochastic error term.

In the case of the logit model, the regression is the following:

$$P(College\ enrolment_{it} = 1 \mid X_{it-1}) = f(Z)$$

where $f(Z) = \frac{e^z}{1+ e^z}$ is the logistic distribution function and

$$Z = \beta_0 + \beta_1 AlreadyOccuredChange_i + \beta_2 StructureChange\ 9 - 13_i + \beta_3 StructureChange\ 13 - 17_i + \gamma X_{it-1} + \varepsilon_{it}$$

Table 7 below shows the estimated coefficients from five different specifications of the linear probability models.

In the first column, we examine the effect of encountering one of these transitions on the likelihood of enrolling in college, without including other control variables. All the structural changes have adverse effects on the probability of enrolling in college and these impact are statistically significant at the 1 percent level. Relative to adolescents who do not encounter in family structure changes, those who experience it before turning 9 present a lower likelihood of enrolling in college on average by 16 p.p.; adolescents exposed to it between the ages of 9 and 13 are the ones mostly affected with a decrease in the likelihood in object of almost 23 p.p.; finally, those exposed to the transition between the ages of 13 and 17 present a lower probability of enrolling in college by 14 p.p.

In column (2) we include adolescent's, primary caregiver's, and households' characteristics as explanatory variables. When controlling for these variables, the impact of a structural change occurring between the ages of 13 and 17 is weakly statistically significant, with a p-value of 0.114. The other two transitions remain statistically significant at the 1 percent level, although slightly smaller in magnitude. Those who experience a family structure change prior to the age of 9 see their likelihood of enrolling in college decreases by 12 p.p. on average, while for those exposed to the structural change between the ages of 9 and 13 the reduction is approximately 19 p.p. relative to adolescent who do not encounter in any change. There is no difference in the likelihood of signing up for college between girls and boys. Birthweight continues to exert its positive influence, as do age and education level of the primary caregiver, and household income.

In the third specification we include further adolescent's and primary caregiver's characteristics, along with an indicator variable for the area of residence. The effects of the structural changes remain negative and statistically significant at the 1 percent level for the change occurred prior to the age of 9 and the one occurred between 9 and 13 years. For the change that took place between 13 and 17 years, the significance level is 10 percent. The effect of the other explanatory variables is almost the same.

Table 7. Linear Probability Model. College enrollment

	(1)	(2)	(3)	(4)	(5)
Already occurred change	-0.156***	-0.121***	-0.135***	-0.112***	-0.110***
	(0.0387)	(0.0378)	(0.0380)	(0.0380)	(0.0388)
Structure change between 9-13	-0.230***	-0.181***	-0.190***	-0.159***	-0.145***
	(0.0569)	(0.0568)	(0.0559)	(0.0549)	(0.0550)
Structure change between 13-17	-0.139***	-0.0793	-0.0892*	-0.0613	-0.0603
	(0.0498)	(0.0502)	(0.0493)	(0.0484)	(0.0485)
Boy		-0.0212	-0.0163	-0.0177	-0.0132
		(0.0141)	(0.0133)	(0.0138)	(0.0130)
Birthweight		0.0270**	0.0266**	0.0226*	0.0226*
		(0.0125)	(0.0117)	(0.0133)	(0.0124)
Been breastfed			0.0255*	0.00957	0.0174
			(0.0146)	(0.0153)	(0.0144)
Not Irish			0.00694	0.0209	0.0125
			(0.0415)	(0.0414)	(0.0403)
Twin				-0.000401	-0.00314
				(0.0425)	(0.0394)
Number of siblings		0.00111	-0.0155*	-0.00374	-0.0218**
		(0.00752)	(0.00822)	(0.00751)	(0.00886)
Academically oriented secondary school				0.292***	0.307***
				(0.0458)	(0.0433)
Female pcg		0.0201	0.00915	0.0117	0.0115
		(0.0421)	(0.0396)	(0.0413)	(0.0386)
Pcg's age		0.00777***	0.123***	0.120***	0.120***
		(0.00167)	(0.0238)	(0.0248)	(0.0238)
Pcg's age squared			-0.118***	-0.118***	-0.112***
			(0.0245)	(0.0257)	(0.0246)
Pcg's years of education		0.0344***	0.0434***	0.0298***	0.0535***
		(0.00356)	(0.00887)	(0.00381)	(0.0110)
Employed				-0.0255	0.00291
				(0.0167)	(0.0196)
Unemployed				-0.0993*	-0.158**
				(0.0601)	(0.0643)
Predicted Income			-0.00583		-0.0174**
			(0.00610)		(0.00835)
Annual income adjusted for inflation		0.00400***		0.00337***	
		(0.000856)		(0.000845)	
Rural area			0.00837	0.0383***	-0.00908
			(0.0174)	(0.0142)	(0.0210)
Constant	0.802***	-0.260**	-3.001***	-3.104***	-3.207***
	(0.00700)	(0.108)	(0.575)	(0.594)	(0.574)
R-squared	0.0148	0.0801	0.0829	0.102	0.0994
Observations	3572	3167	3515	3125	3472

Notes: The Table reports Linear probability model estimates. Sample: GUI dataset, children aged 20 years old. The dependent variable is *College enrollment*, which takes the value of 1 if the young enrolled in college/university, and 0 otherwise. Standard errors, reported in parentheses, are corrected for heteroskedasticity, and allow for clustering at the household level. To make easier the interpretation of the coefficients, in all the regressions we have rescaled the variables *Income,* dividing it by 1000, and *Age2,* dividing it by 100, thus reducing the number of zeros after the decimal points. The symbols ***, **, * indicate that the coefficients are statistically significant at the 1, 5, and 10 percent level respectively.

The fourth specification improve the analysis by further controlling for the primary caregiver's employment status and the type of secondary school attended by the young. The structural changes exert their negative impact on the educational attainment as seen in previous estimations. Additionally, the results indicate a decrease of 10 p.p. in the likelihood of an adolescent enrolling in college if the primary caregiver is currently unemployed. Conversely, attended an academically oriented secondary school increases that likelihood by almost 29 p.p.

Finally, column (5) differs from the specification in column (4) regarding the income variable included. Here, we control for the predicted income, instead of the actual income, which turns to be statistically significant. The effects of the structural changes of interest are robust to the inclusion of such control, preserving their negative and statistically significant effects. On average, the probability of enrolling in college decreases by 11 and 14 p.p. if the transition from a two-parent to a one-parent family type takes place before the age of 9 and between the ages of 9 and 13 respectively.

The estimates for the probability of enrolling in college using a logit model are presented in Table 8. Similar to the linear probability model, we estimate five different specifications by increasing the number of explanatory variables included.

Estimates from the logistic model validate the negative impacts of undergoing such transitions on adolescents' educational attainment at the age of 20.

The specification in column (1) shows the effects of the three types of family changes, without including other explanatory variables. On average, adolescent's probability of enrolling in college decreases by 13 p.p., 18 p.p., and 12 p.p. if the transition occurred before the age of 9, between the ages of 9 and 13, and between the ages of 13 and 17 respectively, in comparison to adolescents not exposed to variations in the structure of their families.

When adding control variables, specifications (2)-(5), the effects of the changes that took place between the ages of 13 and 17 eventually turn statistical insignificant. The structural changes occurred before the age of 9 and between the ages of 9 and 13 continue to exhibit statistically significant and detrimental effects. These effects result in a decrease in the likelihood of enrolling in college ranging from 9 to 11 p.p. if the structural change takes place before the child turned 9, and from 11 to 14 p.p. if the change takes place between the ages of 9 and 13, relative to adolescents not exposed to structural changes.

Table 8. Logit Estimates. College enrollment

	(1)	(2)	(3)	(4)	(5)
Already occurred change	-0.133***	-0.0983***	-0.114***	-0.0901***	-0.0895***
	(0.0286)	(0.0285)	(0.0288)	(0.0287)	(0.0300)
Structure change between 9-13	-0.185***	-0.136***	-0.150***	-0.121***	-0.112***
	(0.0390)	(0.0409)	(0.0402)	(0.0399)	(0.0403)
Structure change between 13-17	-0.120***	-0.0629	-0.0762**	-0.0501	-0.0480
	(0.0374)	(0.0383)	(0.0382)	(0.0376)	(0.0383)
Boy		-0.0221	-0.0171	-0.0198	-0.0148
		(0.0141)	(0.0134)	(0.0140)	(0.0133)
Birthweight		0.0258**	0.0266**	0.0218	0.0221*
		(0.0125)	(0.0119)	(0.0133)	(0.0126)
Been breastfed			0.0253*	0.00957	0.0178
			(0.0143)	(0.0151)	(0.0143)
Not Irish			0.00267	0.0196	0.00871
			(0.0416)	(0.0417)	(0.0408)
Twin				-0.00186	-0.00805
				(0.0414)	(0.0384)
Number of siblings		0.00337	-0.0140*	-0.00156	-0.0202**
		(0.00732)	(0.00783)	(0.00742)	(0.00847)
Academically oriented secondary school				0.207***	0.220***
				(0.0315)	(0.0297)
Female pcg		0.0152	0.00528	0.00990	0.00595
		(0.0379)	(0.0360)	(0.0370)	(0.0348)
Pcg's age		0.00748***	0.0979***	0.0951***	0.0951***
		(0.00154)	(0.0201)	(0.0211)	(0.0201)
Pcg's age squared			-0.0931***	-0.0927***	-0.0875***
			(0.0210)	(0.0222)	(0.0212)
Pcg's years of education		0.0317***	0.0421***	0.0282***	0.0524***
		(0.00340)	(0.00818)	(0.00363)	(0.0104)
Employed				-0.0292*	0.00208
				(0.0165)	(0.0193)
Unemployed				-0.0722*	-0.132***
				(0.0430)	(0.0500)
Predicted Income			-0.00583		-0.0171**
			(0.00556)		(0.00792)
Annual income adjusted for inflation		0.00517***		0.00447***	
		(0.00118)		(0.00116)	
Rural area			0.00578	0.0372***	-0.0104
			(0.0169)	(0.0143)	(0.0205)
P-seudo R-squared	0.013	0.078	0.077	0.095	0.091
Observations	3572	3167	3515	3125	3472

Notes: The Table reports marginal effects from logit estimates. Sample: GUI dataset, children aged 20 years old. The dependent variable is *College enrollment*, which takes the value of 1 if the young enrolled in college/university, and 0 otherwise. Standard errors, reported in parentheses, are corrected for heteroskedasticity, and allow for clustering at the household level. To make easier the interpretation of the coefficients, in all the regressions we have rescaled the variables *Income,* dividing it by 1000, and *Age2,* dividing it by 100, thus reducing the number of zeros after the decimal points. The symbols ***, **, * indicate that the coefficients are statistically significant at the 1, 5, and 10 percent level respectively.

As accomplished in the preceding section, Table 9 provides a comparison between the findings derived from the "ad-hoc" chosen subset of explanatory variables and the results obtained using the DML approach, which identifies a data-driven subset of control variables.

As a preliminary step, we proceed to estimate a simple lasso logit model with the merely aim of performing feature selection and determining whether the change structure indicators are identified as important feature. In this case, the only variable chosen by the model is the absence of any change in family structure. [17] Results are not presented.

We implement cross-validation partialing out lasso logit model (or DML) to make inference about the effects of experiencing a family structure change on the probability of signing up in college. To be consisted with the analysis performed using the ad-hoc selected controls, we indicate the same structural change indicator as variables of interest. In addition, the first column presents estimates from a logistic model in which we control for all the "ad-hoc" selected explanatory variables used in the estimations presented above.

The coefficients displayed in Table 9 are the odds ratios of the variables for which we want to infer causality. We estimate three DML specifications. The first specification does not take into account non-linearity and interactions within the control variables. The second and third specifications include interactions between variables and polynomial terms up to the second order of the base variables, respectively.

Table 9. Double-Machine Learning Estimates: College enrollment

	Logit	DML	p-value	DML	p-value	DML	p-value
Already occurred change	0.549***	0.639***	0.042	0.807	0.364	0.798	0.336
	(0.111)	(0.141)		(0.189)		(0.187)	
Structure change between 9-13	0.472***	0.556***	0.037	0.538**	0.035	0.573*	0.066
	(0.128)	(0.157)		(0.158)		(0.173)	
Structure change between 13-17	0.725	0.634*	0.084	0.654	0.122	0.657	0.125
	(0.186)	(0.167)		(0.179)		(0.179)	
Interaction terms		NO		YES		YES	
Second order polynomials		NO		NO		YES	
# potential controls		55		4584		4590	
# selected controls	16	13		285		285	
Observations	3472	2661		2656		2656	

Notes: The Table reports the estimated odds ratios using Logit model in column (1) and Cross-fit partialing out logit in the following columns. Sample: GUI dataset, children aged 20 years old. The dependent variable is *College enrollment*, which takes the value of 1 if the young enrolled in college/university, and 0 otherwise. The variables of interests are *already occurred change*, which takes the value of 1 if the event took place before the child turned 9, and 0 otherwise; *Structure change between 9-13*, which takes the value of 1 if such change took place between 9 and 13 years of age, and 0 otherwise; *Structure change between 13-17*, which takes the value of 1 if the transition took place between 13 and 17 years of age, and 0 otherwise. Standard errors, reported in parentheses, are corrected for heteroskedasticity, and allow for clustering at the household level. The symbols ***, **, * indicate that the coefficients are statistically significant at the 1, 5, and 10 percent level respectively.

Similar to previous findings, the results from the first specification of the DML approach are similar to those obtained when using the set of "ad-hoc" selected controls. In the high-dimensionality setting, where the number of potential controls exceeds the sample size (columns

[17] As before, we perform two lasso logit models, using cross-validation and the adaptive method to select the optimal value of the penalized parameter λ^* . The subset of important feature is made of 26 variables in the first model and of 15 variables in the second.

(3)-(4)), the coefficient of the change in family structure occurred before the age of 9 turns statistically insignificant. The effects of experiencing a structural change between the ages of 9 and 13 are statistically significant at the 5 percent level. Adolescents who experience such change between the ages of 9 and 13 have, on average, a lower probability of enrolling in college by 46 and 43 percentage points (columns 3 and 4 respectively) compared to adolescents who do not experience any structural change in their family. Experiencing a structural change between the ages of 13 and 17 does not have a significant effect on the likelihood of enrolling in college for adolescents.

VII. Conclusions

This paper studies the extent to which a family structure change resulting from parental divorce would produce adverse consequences for adolescent's educational outcomes as measured by their final high school grade and their likelihood of enrolling in college. In the analysis we use data from the Growing Up in Ireland data – Child Cohort, the Irish longitudinal study that tracks a cohort of children from 9 to 20 years old.

Previous research provides evidence of a negative association between the transition from a two-parent family type to a one-parent type on several children's outcomes. However, the use of different types of data and methodologies has produced mixed findings. Earlier studies based on cross-sectional data have been criticized for not considering the endogeneity of family structure change and for potentially overestimating its impact. Therefore, it is strongly recommended to employ methodologies that account for endogeneity. Nonetheless, when examining some educational achievements which are observed only once by definition, the range of methodologies that can be utilized is limited.

To take into account potential heterogeneous effects that the age at which child is exposed to the structural change may have on the educational achievement, we use all the available waves in our data, with the exception of the last one in which the educational outcomes are measured, to determine whether in each household a structural change took place.

Additionally, we compare the results obtained from a model that includes a subset of "ad-hoc" selected control variables with the results obtained from a model that employs Machine Learning tools for the selection procedure of control variables. These tools enable to make causal inference by applying a data-driven approach that allows us to address the issue of omitted variables bias that could arise if some of the predictors excluded from the model were correlated with the ones included and affect the outcome.

Overall, our results provide evidence of a negative effect of experiencing a family structure change on later educational outcomes and prove that children's age at the time of the transition plays an important role in mitigating this impact.

However, it is crucial to pay attention to the subset of control variables included the model. On one hand, when employing the "ad-hoc" variables selection approach, the findings bring to light a statistically significant impact of experiencing a change in family structure both prior to the age of 9 and between the ages of 9 and 13 on the educational outcomes, namely the adolescent's likelihood of obtaining a high grade at their high-school diploma and of enrolling in college. Conversely, the study reveals that undergoing such a transition between the ages of 13 and 17 does not yield statistically significant results for adolescents' education achievements.

On the other hand, the use of the data-driven approach partly confirms these adverse effects. In particular, the first DML model - where we propose a "small" set of control variables from which the algorithm can select the important ones to make inference about the indicated regressors of interest, and we implicitly allow the model examine a linear relationship among the controls, the parameters of interest, and the outcomes – produces results that are consistent with those produced by the "ad-hoc" controls selection procedure. All the parameters show the same level of statistical significance and direction of the impact; only the magnitude of the effects slightly differ in the DML estimates.

As we progressively increase the number of control variables and the model complexity by including interactions and polynomials of the base variables up to the second order, we end up in a high-dimensional setting where the dimension of the control variables vector exceeds the sample size, and where the ML approaches demonstrate their main strength. In these specifications the effects of experiencing a structural change before the ages of 9 turns statistically insignificant.

Therefore, the results obtained from the DML approach allow us to assert that the results from the "ad-hoc" variable selection procedure tend to overestimate the impact of such a change and are likely to suffer from omitted variable bias.

Overall, resorting to ML approach for the purpose of making causal inference is extremely useful in situations where there is a lack of an underlying economic theory that defines exactly which controls need to be included in the model, freeing the researcher from the necessity to select a subset of control variables or to specify a distinct functional form for the model. In a different context, however, employing a double machine learning (DML) approach not only fails to diverge from the estimates produced by the standard "ad-hoc" approach, but it does not solve the omitted variable bias issue. It is crucial to conduct further research to promote the adoption of machine learning techniques with the aim of performing causal inference analysis.

References

Abufhele, A., Behrman, J. and Bravo, D. (2017) ‘Parental preferences and allocations of investments in children’s learning and health within families’, *Social Science and Medicine*, 194, pp. 76–86. Available at: https://doi.org/10.1016/j.socscimed.2017.09.051.

Amato, P.R. (1996) ‘Explaining the Intergenerational Transmission of Divorce’, *Journal of Marriage and the Family*, 58(3), p. 628. Available at: https://doi.org/10.2307/353723.

Amato, P.R. (2000) ‘The consequences of divorce for adults and children’, *Journal of Marriage and Family*. National Council on Family Relations, pp. 1269–1287. Available at: https://doi.org/10.1111/j.1741-3737.2000.01269.x.

Amato, P.R. (2005) ‘The Impact of Family Formation Change on the Cognitive, Social, and Emotional Well-Being of the Next Generation’, *The Future of Children*, 15(2), pp. 75–96.

Amato, P.R. (2010) ‘Research on divorce: Continuing trends and new developments’, *Journal of Marriage and Family*, 72(3), pp. 650–666. Available at: https://doi.org/10.1111/j.1741-3737.2010.00723.x.

Amato, P.R. (2014) ‘The consequences of divorce for adults and children: an update’, *Drustvena Istrazivanja*, 23(1), pp. 5–24. Available at: https://doi.org/10.5559/di.23.1.01.

Amato, P.R. (2016) ‘The Consequences of Divorce for Adults and Children Published by : National Council on Family Relations Linked references are available on JSTOR for this article : The Consequences of Divorce for Adults and Children’, *Journal of Marriage and Family*, 62(4), pp. 1269–1287.

Amato, P.R. and Anthony, C.J. (2014) ‘Estimating the effects of parental divorce and death with fixed effects models’, *Journal of Marriage and Family*, 76(2), pp. 370–386. Available at: https://doi.org/10.1111/jomf.12100.

Amato, P.R. and Cheadle, J.E. (2008) ‘Parental divorce, marital conflict and children’s behavior problems: A comparison of adopted and biological children’, *Social Forces*, 86(3), pp. 1139–1161. Available at: https://doi.org/10.1353/sof.0.0025.

Amato, P.R. and DeBoer, D.D. (2001) ‘The transmission of marital instability across generations: Relationship skills or commitment to marriage?’, *Journal of Marriage and Family*, 63(4), pp. 1038–1051. Available at: https://doi.org/10.1111/j.1741-3737.2001.01038.x.

Amato, P.R. and James, S. (2010) ‘Divorce in Europe and the United States: Commonalities and differences across nations’, *Family Science*, 1(1), pp. 2–13. Available at: https://doi.org/10.1080/19424620903381583.

Amato, Paul R and Keith, B. (1991) ‘Linked references are available on JSTOR for this article : Parental Divorce and Adult Well-being : A Meta-analysis’, 53(1), pp. 43–58.

Amato, Paul R. and Keith, B. (1991) ‘Parental divorce and the well-being of children: A meta-analysis’, *Psychological Bulletin*, 110(1), pp. 26–46. Available at: https://doi.org/10.1037/0033-2909.110.1.26.

Amato, P.R. and Sobolewski, J.M. (2001) ‘The effects of divorce and marital discord on adult children’s psychological well-being’, *American Sociological Review*, 66(6), pp. 900–921. Available at: https://doi.org/10.2307/3088878.

Anderson, J. (2014) 'The impact of family structure on the health of children: Effects of divorce', *Linacre Quarterly*, 81(4), pp. 378–387. Available at: https://doi.org/10.1179/0024363914Z.00000000087.

Andini, M. *et al.* (2018) 'Targeting with machine learning: An application to a tax rebate program in Italy', *Journal of Economic Behavior and Organization*, 156(March), pp. 86–102. Available at: https://doi.org/10.1016/j.jebo.2018.09.010.

Antulov-Fantulin, N., Lagravinese, R. and Resce, G. (2021) 'Predicting bankruptcy of local government: A machine learning approach', *Journal of Economic Behavior and Organization*, 183, pp. 681–699. Available at: https://doi.org/10.1016/j.jebo.2021.01.014.

Arlot, S. and Celisse, A. (2010) 'A survey of cross-validation procedures for model selection', *Statistics Surveys*, 4, pp. 40–79. Available at: https://doi.org/10.1214/09-SS054.

Arpino, B., Moglie, M. Le and Mencarini, L. (2018) 'Machine-Learning techniques for family demography : An application of random forests to the analysis of divorce determinants in Germany', *Annual Meeting Paa*, (56), p. 32.

Aughinbaugh, A., Pierret, C.R. and Rothstein, D.S. (2005) 'The impact of family structure transitions on youth achievement: Evidence from the children of the NLSY79', *Demography*, 42(3), pp. 447–468. Available at: https://doi.org/10.1353/dem.2005.0023.

Augustine, J.M. and Kimbro, R.T. (2013) 'Journal of Applied Research on Children: Informing Policy for Children at Risk', *Journal of Applied Research on Children: Informing Policy for Children at Risk*, 4(1), pp. 0–24. Available at: http://digitalcommons.library.tmc.edu/childrenatrisk%5Cnhttp://digitalcommons.library.tmc.edu/childrenatrisk/vol4/iss1/5.

Augustine, J.M. and Kimbro, R.T. (2017) 'Associations and Intervening Mechanisms Between Family Structure and Young Children's Obesity', *Journal of Family Issues*, 38(16), pp. 2277–2302. Available at: https://doi.org/10.1177/0192513X15621344.

Austin, P.C. (2011) 'An introduction to propensity score methods for reducing the effects of confounding in observational studies', *Multivariate Behavioral Research*, 46(3), pp. 399–424. Available at: https://doi.org/10.1080/00273171.2011.568786.

Beets, M.W., Cardinal, B.J. and Alderman, B.L. (2010) 'Parental social support and the physical activity-related behaviors of youth: A review', *Health Education and Behavior*, 37(5), pp. 621–644. Available at: https://doi.org/10.1177/1090198110363884.

Belloni, A. *et al.* (2012) 'Sparse Models and Methods for Optimal Instruments With an Application to Eminent Domain', *Econometrica*, 80(6), pp. 2369–2429. Available at: https://doi.org/10.3982/ecta9626.

Belloni, A., Chernozhukov, V. and Hansen, C. (2014) 'on Structural and Treatment Effects †', *Journarl of economci perspectives*, 28(2), pp. 29–50.

Berenson, G.S. (2012) 'Health Consequences of Obesity', *Pediatric Blood and Cancer*, 58(1), pp. 117–121. Available at: https://doi.org/10.1002/pbc.23373.

Bernardi, F. and Boertien, D. (2016) 'Understanding heterogeneity in the effects of parental separation on educational attainment in Britain: Do children from lower educational backgrounds have less to lose?', *European Sociological Review*, 32(6), pp. 807–819. Available at: https://doi.org/10.1093/esr/jcw036.

Biblarz, T.J. and Gottainer, G. (2000) 'Family structure and children's success: A comparison of widowed and divorced single-mother families', *Journal of Marriage and Family*, 62(2), pp. 533–548. Available at: https://doi.org/10.1111/j.1741-3737.2000.00533.x.

de Blasio, G. and D'ignazio, A. (2021) 'Regional policy in the machine learning era: Opportunities and challenges', *Scienze Regionali*, 20(2), pp. 205–220. Available at: https://doi.org/10.14650/99817.

Bradley RH, C.R. (2008) 'Infant temperament, parenting, and externalizing behavior in first grade: a test of the differential susceptibility hypothesis', *J Child Psychol Psychiatry*, 49(2), pp. 124–31.

Bream, V. and Buchanan, A. (2003) 'Distress among Children whose Separated or Divorced Parents cannot Agree Arrangements for Them', *British Journal of Social Work*, 33(2), pp. 227–238. Available at: https://doi.org/10.1093/bjsw/33.2.227.

Bright, M.A. *et al.* (2015) 'Adverse childhood experiences and dental health in children and adolescents', *Community Dentistry and Oral Epidemiology*, 43(3), pp. 193–199. Available at: https://doi.org/10.1111/cdoe.12137.

Brown, S.L. (2004) 'Family structure and child well-being: The significance of parental cohabitation', *Journal of Marriage and Family*, 66(2), pp. 351–367. Available at: https://doi.org/10.1111/j.1741-3737.2004.00025.x.

Bzostek, S.H. and Beck, A.N. (2011) 'Familial instability and young children's physical health', *Social Science and Medicine*, 73(2), pp. 282–292. Available at: https://doi.org/10.1016/j.socscimed.2011.04.014.

Carrieri, V., Lagravinese, R. and Resce, G. (2021) 'Predicting vaccine hesitancy from area-level indicators: A machine learning approach', *Health Economics (United Kingdom)*, 30(12), pp. 3248–3256. Available at: https://doi.org/10.1002/hec.4430.

Cavanagh, S. and Fomby, P. (2019) 'Family Instability in the Lives of American Children', *Annual Review of Sociology*, 45(1), pp. 493–513. Available at: https://doi.org/10.1146/annurev-soc-073018-022633.

Cavanagh, S.E. and Huston, A.C. (2006) 'Family instability and children's early problem behavior', *Social Forces*, 85(1), pp. 551–581. Available at: https://doi.org/10.1353/sof.2006.0120.

Cerqua, A. *et al.* (2021) 'Local mortality estimates during the COVID-19 pandemic in Italy', *Journal of Population Economics*, 34(4), pp. 1189–1217. Available at: https://doi.org/10.1007/s00148-021-00857-y.

Chen, A.Y. and Escarce, J.J. (2010) 'Family structure and childhood obesity, Early Childhood Longitudinal Study - Kindergarten Cohort', *Preventing Chronic Disease*, 7(3). Available at: http://www.cdc.gov/pcd/issues/2010/may/09_0156.htm.Accessed[date].

Chen, X., Liu, M. and Li, D. (2000) 'Parental warmth, control, and indulgence and their relations to adjustment in Chinese children: A longitudinal study.', *Journal of Family Psychology*, 14, pp. 401–419.

Chen, Y.-C., Fan, E. and Liu, J.-T. (2019) 'Understanding the mechanisms of parental divorce effects on child's higher education', *NBER Working Papers* [Preprint]. Available at: http://www.nber.org/papers/w16019.

Cherlin, A.J. *et al.* (1991) 'Longitudinal studies of effects of divorce on children in Great Britain and the United States', *Science*, 252(5011), pp. 1386–1389. Available at: https://doi.org/10.1126/science.2047851.

Cherlin, A.J., Chase-Lansdale, P.L. and McRae, C. (1998) 'Effects of parental divorce on mental health throughout the life course', *American Sociological Review*, 63(2), pp. 239–249. Available at: https://doi.org/10.2307/2657325.

Chernozhukov, V. *et al.* (2018) 'Double/debiased machine learning for treatment and structural parameters', *Econometrics Journal*, 21(1), pp. C1–C68. Available at: https://doi.org/10.1111/ectj.12097.

Christina J. Cross (2019) 'Racial/Ethnic Differences in the Association Between Family Structure and Children's Education', *Journal of Marriage and Family*, Vol. 90(November), pp. 1–22.

Cid, A. and Stokes, C.E. (2013) 'Family Structure and Children's Education Outcome: Evidence from Uruguay', *Journal of Family and Economic Issues*, 34(2), pp. 185–199. Available at: https://doi.org/10.1007/s10834-012-9326-z.

Cooper, C.E. *et al.* (2008) 'Partnership Instability and Child Wellbeing during the Transition to Elementary School', (April), pp. 1–37.

Duflo, E., Bertrand, M. and Mullainathan, S. (2004) 'How Much Should We Trust Differences-in-Differences Estimates ? Author (s): Marianne Bertrand , Esther Duflo and Sendhil Mullainathan Published by : Oxford University Press Stable URL : http://www.jstor.org/stable/25098683 HOW MUCH SHOULD WE TRUST DIFFE', *The Quarterly Journal of Economics*, 119(1), pp. 249–275.

E. Leuven and Sianesi, B. (2003) 'Psmatch2: Stata module to perform full Mahalanobis and propensity score matching, common support graphing, and covariate impalance testing'. Available at: https://www.pep-net.org/sites/pep-net.org/files/typo3doc/pdf/Training_Material/statadoc.pdf.

Ermisch, J.F. *et al.* (2001) *Family structure and children's achievements*.

Fischer, T. (2007) 'Parental divorce and children's socio-economic success: Conditional effects of parental resources prior to divorce, and gender of the child', *Sociology*, 41(3), pp. 475–495. Available at: https://doi.org/10.1177/0038038507076618.

Fomby, P. and Cherlin, A.J. (2007) 'Family instability and child well-being', *American Sociological Review*, 72(2), pp. 181–204. Available at: https://doi.org/10.1177/000312240707200203.

Fomby, P. and Osborne, C. (2010) 'The influence of union instability and union quality on children's aggressive behavior', *Social Science Research*, 39(6), pp. 912–924. Available at: https://doi.org/10.1016/j.ssresearch.2010.02.006.

Francesconi, M., Jenkins, S.P. and Siedler, T. (2010) 'Childhood family structure and schooling outcomes: Evidence for Germany', *Journal of Population Economics*, 23(3), pp. 1073–1103. Available at: https://doi.org/10.1007/s00148-009-0242-y.

Geisser, S. (1975) 'The predictive sample reuse method with applications', *Journal of the American Statistical Association*, 70(350), pp. 320–328. Available at: https://doi.org/10.1080/01621459.1975.10479865.

Gibson, L.Y. *et al.* (2007) 'The role of family and maternal factors in childhood obesity', *Medical Journal of Australia*, 186(11), pp. 591–595. Available at: https://doi.org/10.5694/j.1326-5377.2007.tb01061.x.

Goisis, A., Özcan, B. and Van Kerm, P. (2019) 'Do Children Carry the Weight of Divorce?', *Demography*, 56(3), pp. 785–811. Available at: https://doi.org/10.1007/s13524-019-00784-4.

Golan, M. and Crow, S. (2004) 'Targeting parents exclusively in the treatment of childhood obesity: Long-term results', *Obesity Research*, 12(2), pp. 357–361. Available at: https://doi.org/10.1038/oby.2004.45.

Gonzalez, L. (2011) 'The Long Term Effects of Legalizing Divorce on Children', *SSRN Electronic Journal* [Preprint], (3789). Available at: https://doi.org/10.2139/ssrn.1374206.

Goodman, A. and Goodman, R. (2009) 'Strengths and difficulties questionnaire as a dimensional measure of child mental health', *Journal of the American Academy of Child and Adolescent Psychiatry*, 48(4), pp. 400–403. Available at: https://doi.org/10.1097/CHI.0b013e3181985068.

Goodman, R. (1997) 'The strengths and difficulties questionnaire: A research note', *Journal of child psycology and psychiatry*, 38(5), pp. 581–586. Available at: https://doi.org/10.4324/9781315258324-21.

Gruber, J. (2004) 'Is making divorce easier bad for children? the long-run implications of unilateral divorce', *Journal of Labor Economics*, 22(4), pp. 799–833. Available at: https://doi.org/10.1086/423155.

Hannan, C. and Halpin, B. (2014) 'The influence of family structure on child outcomes: Evidence for Ireland', *Economic and Social Review*, 45(1), pp. 1–24.

Hassen, T.A. *et al.* (2021) 'The association between birth weight and proxy-reported health-related quality of life among children aged 5 − 10 years old: A linked data analysis', *BMC Pediatrics*, 21(1), pp. 1–8. Available at: https://doi.org/10.1186/s12887-021-02882-y.

Ho, D.E. *et al.* (2007) 'Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference', *Political Analysis*, 15(3), pp. 199–236. Available at: https://doi.org/10.1093/pan/mpl013.

Hveem, M.R., Faulconer, S.C.M. and Dufur, M.J. (2022) 'Comparing Children's Behavior Problems in Biological Married, Biological Cohabitating, and Stepmother Families in the UK', *International Journal of Environmental Research and Public Health*, 19(24). Available at: https://doi.org/10.3390/ijerph192416543.

J.D., A. and Pischke, J.S. (2009) *Mostly Harmless Econometrics : An Empiricist ' s Companion*.

Jackson, S.L. *et al.* (2011) 'Impact of poor oral health on children's school attendance and performance', *American Journal of Public Health*, 101(10), pp. 1900–1906. Available at: https://doi.org/10.2105/AJPH.2010.200915.

King, V. (2009) 'Stepfamily formation: Implications for adolescent ties to mothers, nonresident fathers, and stepfathers', *Journal of Marriage and Family*, 71(4), pp. 954–968. Available at: https://doi.org/10.1111/j.1741-3737.2009.00646.x.

Kleinberg, J. *et al.* (2015) 'Prediction Policy Problems', *American Economics Review*, pp. 1–6.

Kravdal, Ø. and Grundy, E. (2019) 'Children's age at parental divorce and depression in early and mid-adulthood', *Population Studies*, 73(1), pp. 37–56. Available at: https://doi.org/10.1080/00324728.2018.1549747.

Lisa Strohschein (2005) 'Parental Divorce and Child Mental Health Trajectories', 67(5), pp. 1286–

1300.

Listl, S. (2011) ‘Family composition and children’s dental health behavior: Evidence from Germany’, *Journal of Public Health Dentistry*, 71(2), pp. 91–101. Available at: https://doi.org/10.1111/j.1752-7325.2010.00205.x.

McCarty, C.A. *et al.* (2005) ‘Parental Emotional Support and Subsequent Internalizing and Externalizing Problems Among Children’, *Journal of Developmental & Behavioral Pediatrics*, 26(4), pp. 267–275.

McNamara, E., Murray, A. and O’Mahony, D. (2019) ‘Growing up in Ireland. Design, Instrumentation and Procedures for Cohort ’08 at 9 Years Old (Wave 5)’, *Longitudinal and Life Course Studies* [Preprint]. Available at: https://doi.org/10.1332/175795919X15699445254797.

Meesters, C. and Muris, P. (2004) ‘Perceived parental rearing behaviours and coping in young adolescents’, *Personality and Individual Differences*, 37(3), pp. 513–522. Available at: https://doi.org/10.1016/j.paid.2003.09.022.

Meyer, B.D. (1995) ‘Natural and quasi-experiments in economics’, *Journal of Business and Economic Statistics*, 13(2), pp. 151–161. Available at: https://doi.org/10.1080/07350015.1995.10524589.

Osborne, C. and Mclanahan, S. (2007) ‘Partnership Instability and Child Well-Being’, *Journal of Marriage and Family*, 69(4), pp. 1065–1083.

De Paola, M. and Gioia, F. (2017) ‘Does patience matter in marriage stability? Some evidence from Italy’, *Review of Economics of the Household*, 15(2), pp. 549–577. Available at: https://doi.org/10.1007/s11150-014-9275-4.

Parcel, T.L., Campbell, L.A. and Zhong, W. (2012) ‘Children’s Behavior Problems in the United States and Great Britain’, *Journal of Health and Social Behavior*, 53(2), pp. 165–182. Available at: https://doi.org/10.1177/0022146512436742.

Paul R. Rosenbaum and Donald B. Rubin (1983) ‘The Central Role of the Propensity Score in Observational Studies for Causal Effects Author (s): Paul R . Rosenbaum and Donald B . Rubin Published by : Oxford University Press on behalf of Biometrika Trust Stable URL : http://www.jstor.’, 70(1), pp. 41–55.

Pianta, R.C. (1992) ‘Child-parent relationship scale.’, *Unpublished measure, University of Virginia* [Preprint], (427).

Powers, D.A. (2005) ‘Effects of family structure on the risk of first premarital birth in the presence of correlated unmeasured family effects’, *Social Science Research*, 34(3), pp. 511–537. Available at: https://doi.org/10.1016/j.ssresearch.2004.04.007.

Rasmussen, A.W. (2009) *Family Structure Changes and Children ′ s Health , Behavior , and Educational Outcomes Department of Economics Family Structure Changes and Children ’ s Health , Behavior , and Educational Outcomes*.

Resce, G. (2022) ‘The impact of political and non-political officials on the financial management of local governments’, *Journal of Policy Modeling*, 44(5), pp. 943–962. Available at: https://doi.org/10.1016/j.jpolmod.2022.08.004.

Resce, G. and Vaquero-Piñeiro, C. (2022) ‘Predicting agri-food quality across space: A Machine Learning model for the acknowledgment of Geographical Indications’, *Food Policy*, 112(September), p. 102345. Available at: https://doi.org/10.1016/j.foodpol.2022.102345.

Resce, G. and Vaquero-Piñeiro, C. (2023) 'Taste of home: Birth town bias in Geographical Indications', *Available at SSRN 4437094* [Preprint], (089). Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4437094%0Ahttp://web.unimol.it/progetti/repec/mol/ecsdps/ESDP23089.pdf.

Rubin, D.B. (2001) 'Using propensity scores to help design observational studies: Application to the tobacco litigation', *Health Services & Outcomes Research Methodology*, (2), pp. 169–188. Available at: https://doi.org/10.1017/CBO9780511810725.030.

Rubin, D.B. (2006) *Matched sampling for causal inference*. Edited by U.C.U.P. Cambridge.

Sandra L . Hofferth (2006) 'Residential Father Family Type and Child Well-Being : Investment versus Selection', *Springer on behalf of the Population Association of America*, 43(1), pp. 53–77.

Sanz-De-Galdeano, A. and Vuri, D. (2007) 'Parental divorce and students' performance: Evidence from longitudinal data', *Oxford Bulletin of Economics and Statistics*, 69(3), pp. 321–338. Available at: https://doi.org/10.1111/j.1468-0084.2006.00199.x.

Schmeer, K.K. (2012) 'Family structure and obesity in early childhood', *Social Science Research*, 41(4), pp. 820–832. Available at: https://doi.org/10.1016/j.ssresearch.2012.01.007.

Sharon H. Bzostek, L.M.B. (2017) 'Family Structure Experiences and Child Socioemotional Development During the First Nine Years of Life: Examining Heterogeneity by Family Structure at Birth', *Demography*, 54(2), pp. 513–540. Available at: https://doi.org/https://doi.org/10.1007/s13524-017-0563-5.

Sheiham, A. (2006) 'Dental caries affects body weight, growth and quality of life in pre-school children', *British Dental Journal*, 201(10), pp. 625–626. Available at: https://doi.org/10.1038/sj.bdj.4814259.

Sigle-Rushton, W., Hobcraft, J. and Kiernan, K. (2005) 'Parental divorce and subsequent disadvantage: A cross-cohort comparison', *Demography*, 42(3), pp. 427–446. Available at: https://doi.org/10.1353/dem.2005.0026.

Smetana, J.G. (2017) 'Current research on parenting styles, dimensions, and beliefs', *Current Opinion in Psychology*, 15, pp. 19–25. Available at: https://doi.org/10.1016/j.copsyc.2017.02.012.

Sohail, S. *et al.* (2018) 'Implementation of machine learning algorithm on factors effecting divorce rate', *2018 International Conference on Engineering and Emerging Technologies, ICEET 2018*, 2018-Janua, pp. 1–5. Available at: https://doi.org/10.1109/ICEET1.2018.8338618.

StataCorp LLC (2023) *Stata 2018 Lasso Reference Manual*.

Steinberg, L., Fletcher, A. and Darling, N. (1994) 'Parental monitoring and peer influences on adolescent substance use', *Pediatrics*, 93(6 II), pp. 1060–1064. Available at: https://doi.org/10.1542/peds.93.6.1060.

Stone, L.L. *et al.* (2010) 'Psychometric properties of the parent and teacher versions of the strengths and difficulties questionnaire for 4- to 12-Year-olds: A review', *Clinical Child and Family Psychology Review*, 13(3), pp. 254–274. Available at: https://doi.org/10.1007/s10567-010-0071-2.

StØrksen, I. *et al.* (2006) 'Adolescent adjustment and well-being: Effects of parental divorce and distress', *Scandinavian Journal of Psychology*, 47(1), pp. 75–84. Available at:

https://doi.org/10.1111/j.1467-9450.2006.00494.x.

Sun, Y. (2001) ‘Family environment and adolescents’ well-being before and after parents’ marital disruption: A longitudinal analysis’, *Journal of Marriage and Family*, 63(3), pp. 697–713. Available at: https://doi.org/10.1111/j.1741-3737.2001.00697.x.

Sun, Y. and Li, Y. (2011) ‘Effects of family structure type and stability on children’s academic performance trajectories’, *Journal of Marriage and Family*, 73(3), pp. 541–556. Available at: https://doi.org/10.1111/j.1741-3737.2011.00825.x.

Tendulkar, S.A.; Buka, S.; Dunn, E.C.; Subramanian, S.V.; Koenen, K.C. (2010) ‘A multilevel investigation of neighborhood effects on parental warmth.’, *J. Community Psychol.*, 38, pp. 557–573.

Thornton, M. *et al.* (2013) ‘Growing up in Ireland. Design, Instrumentation and Procedures for the Infant Cohort at wave 1 (9 Months)’, *Longitudinal and Life Course Studies* [Preprint]. Available at: https://doi.org/10.1332/175795919X15699445254797.

Tibshirani (1996) ‘Regression Shrinkage and Selection via the Lasso’, *Journal of the Royal Statistical Society. Series B (Methodological)*, 58(1), pp. 267–288. Available at: jstor.org/stable/2346178.

Varian, H.R. (2014) ‘Big data: New tricks for econometrics’, *Journal of Economic Perspectives*, 28(2), pp. 3–28. Available at: https://doi.org/10.1257/jep.28.2.3.

Varian, H.R. (2016) ‘Causal inference in economics and marketing’, *Proceedings of the National Academy of Sciences of the United States of America*, 113(27), pp. 7310–7315. Available at: https://doi.org/10.1073/pnas.1510479113.

Waldfogel, J., Craigie, T.A. and Brooks-Gunn, J. (2010) ‘Fragile families and child wellbeing’, *Future of Children*, 20(2), pp. 87–112. Available at: https://doi.org/10.1353/foc.2010.0002.

Williams, J. *et al.* (2015) ‘Growing Up in Ireland. Design, Instrumentation and Procedures for Cohort ’08 at wave 3 (5 Years)’. Available at: www.dcya.gov.ie.

Wing, C., Simon, K. and Bello-Gomez, R.A. (2018) ‘Designing Difference in Difference Studies: Best Practices for Public Health Policy Research’, *Annual Review of Public Health*, 39, pp. 453–469. Available at: https://doi.org/10.1146/annurev-publhealth-040617-013507.

Yang, Z., Zhang, R. and Liu, L. (2022) ‘The Effect of Parent-Child Relationship on Child Development’, *Academic Journal of Humanities & Social Sciences*, 5(15), pp. 118–122. Available at: https://doi.org/10.25236/ajhss.2022.051515.

Yelick, A. (2017) ‘The Effects of Family Structure on Consumption and Exercise Patterns for Adolescent Youth’, *Child and Adolescent Social Work Journal*, 34(4), pp. 381–395. Available at: https://doi.org/10.1007/s10560-016-0468-y.

Chapter III

The Influence of Family Structure on Child Well-Being: A Quasi-Experimental Design with Kernel Propensity Score Matching and Difference-in-Differences

Abstract

The type of family structure in which children are raised has important implications for children's development. It is common to estimate the impact of family structure change comparing outcomes obtained from children raised in two-parent families with outcomes of children nurtured in single-parent families. However, selection bias may be a critical issue when assessing such effect. Many important household' and parents' characteristics are linked to the structure change and to the children's outcomes as well. Failing to consider such aspect would lead to biased estimates of the contribution of family transition. In this paper, we estimate the impact of family structure change experienced during childhood (children aged 3-5) both in the short and long run (at 5 and 9 years old respectively) and we try to deal with the potential selection bias adopting a propensity score matching approach. Specifically, we use children from traditional two-parent families as control group and children from non-traditional families as units in the treated group. Once we match treated children with untreated children that have a similar likelihood of experiencing a family structure change, our results show that children who experience family structure between 3-5 years old have worse emotional and behavioural outcomes, measured by externalizing and internalizing problems, and the parent-child relationship deteriorates, with an increase in the level of conflict between the parent and the child, and a decrease in the level of closeness relative to children in two-parent households. These effects do not disappear when the child is 9 years old.

JEL Classification: C14, C31, C90, I30, J12

Keywords: Family structure, Chil well-being, Quasi-experimental methods, Difference-in-Differences, Propensity Score Matching

I. Introduction

Over the past decades several researchers have linked family structure, or parents' union status, with children development outcomes (Amato, 2000, 2010, 2016; Amato and James, 2010; Anderson, 2014). The traditional nuclear family structure, the one in which the two biological parents live together with their children, has become less common and the number of children raised in less traditional family structure, mainly single-parent households, has been continuously increasing.

International studies mainly suggest that family structure change has unfavourable implications for children under different spheres, including emotional and behavioural disorders, mental health, difficulties with social relationships and interactions, and educational achievements (Amato, 2000; Ermisch *et al.*, 2001; StØrksen *et al.*, 2006; Amato and Cheadle, 2008; Parcel, Campbell and Zhong, 2012).

In this paper we focus on the impact of family structure change on the socioemotional and behavioural child's development domains, and on the quality of the parent-child relationship.

It should be noted that in the literature there is not a unique definition of family structure change; a lot is contingent on the level of detail of available information and the purpose of the study. In general, it is conceptualized as either the transition between qualitatively distinct states or the repeated occurrence of disruption.

As regards the former concept, abundant research has documented the positive effects of living in a family consisting of two parents as opposed to family with only one parent on children's behavioural traits. For instance, children in biological married families tend to exhibit fewer behaviour problems compared to other family structures. Conversely, children in single-parent families present the worse behaviour problems (Amato, 2005; Sandra L . Hofferth, 2006; Waldfogel, Craigie and Brooks-Gunn, 2010; Sharon H. Bzostek, 2017; Cavanagh and Fomby, 2019).

Amato (2005) examines the impact of family structure on children's well-being. His findings show that children growing up in stable, two-parent families are less likely to experience cognitive, emotional, and social problems than children growing up in single-parent households. He claims that children in the former type of family are better off than those in the latter due to a higher standard of living, effective parenting, cooperative co-parenting, and fewer stressful events they are exposed to.

Along the same line, Hveem, Faulconer, and Dufur (2022) show that children living with stepmother and cohabiting biological parents exhibit more externalizing problems compared to children living with married biological parents, while they do not detect any impact of family structure on children's internalizing behaviour problems.

With regards to the second concept, the number of changes in parents' union status is another aspect that garnered scholars' attention. Osborne and Mclanahan (2007) show that offspring born to unmarried parents and that are of a minority background face a greater number of changes in their parents' partnerships, which are connected to a rise in children's behavioural issues. After evaluating the adverse impact stemming from family structure change on child's educational attainment, behaviour, and health outcomes, Rasmussen (2009) shows that any further modification in familial arrangements implies even more unfavourable consequences for children.

Family type at the time of child's birth has been proved to shape the impact of family structure change (Cavanagh and Huston, 2006; Sandra L . Hofferth, 2006; Fomby and Cherlin, 2007; Amato and James, 2010). Overall, the results provide clear indication of a greater effect of alterations in family structure on the socioemotional development of children born to married parents. In contrast, little evidence is found for children born to cohabiting or noncoresident parents.

A vast body of research shed light on the role of children's age at the time of structure change to mitigate or worsen the effects of the structure alteration. In this context, it has been shown that children who experience a family structure change across the early life course present more adverse outcomes relative to children who face such change when older (Ermisch *et al.*, 2001; Cavanagh and Huston, 2006; Rasmussen, 2009; Kravdal and Grundy, 2019). In addition, when investigating the continuity of these effects in the course of time, several studies demonstrate that for some children these harmful effects are active for only a short period of time, while others document a persistent impact even into adulthood (Paul R Amato and Keith, 1991; Amato and Sobolewski, 2001; Ermisch *et al.*, 2001; Amato, 2005).

In parallel with the configuration of the family, the responsiveness of parents towards their children's needs within the familial environment emerges as a crucial element for their development. Overall, evidence shows that children who live in warm family environments tend to have better social adjustment scores. Moreover, it is important to note that even during adolescence, parental responsiveness exerts favourable effects on various aspects of child's development (Steinberg, Fletcher and Darling, 1994; Chen, Liu and Li, 2000; Meesters and Muris, 2004; McCarty *et al.*, 2005; Bradley RH, 2008; Smetana, 2017; Yang, Zhang and Liu, 2022).

The degree of parental warmth and the quality of parent-child relationships are inclined to differ amongst various family structures. In particular, family structures that deviate from the traditional two-parent biological unit tend to exhibit lower rates of parental warmth (Brown, 2004; Tendulkar, S.A.; Buka, S.; Dunn, E.C.; Subramanian, S.V.; Koenen, 2010).

Also the entrance or exit of a romantic partner within the household can introduce complexities and disruptions to the dynamics between parents and children. As shown by King (2009), the level of closeness between a mother and her child declines once a cohabiting stepfather entered the household compared to parent-child relationship of mothers who remained single.

Drawing on data from the Growing Up in Ireland (GUI) – Infant Cohort, in this study we evaluate the impact of family structure change experienced in early childhood (between 3-5 years) on children's behavioural outcomes, assessed by externalizing and internalizing problems, and on the quality of the parent-child relationship, measured through the positive and negative subscales of the Child–Parent Relationship Scale (CPRS).

We use the term 'family structure' to indicate the instability of parents' union within the household, referring to the former concept outlined above that contemplates the transition between qualitatively different family types. A great advantage of the longitudinal data at our disposal is the opportunity to observe family dynamics within each household for a relatively large period, covering the developmental stage of the child through teenagerhood. More details on how we assess family structure change is provided in section II below.

As a first step, we employ a Difference-in-Differences (DiD) approach which relies on observing the outcomes of interest both before and after family structure change takes place, and for two groups of children: on one side, children who undergo the transition; on the other side, children who do not experience any transitions in their household's composition. We consider the group of children exposed to a structure change in their household as treated, and children in households who do not encounter any structural changes as control group. The causal impact of the structure change is identified as difference of the average outcomes for the two groups of children subtracting the variations documented over time.

Results from this stage of the analysis show that family structure change experienced between 3-5 years produces negative consequences both on children's emotional and behavioural outcomes and quality of parent-child relationship. When aged 5, children who experience a transition in their family have higher internalizing and externalizing problems (on average 0.486 and 0.899 respectively) relative to children in intact families. The structure change weakens the bond between parent and child, with an increase in the negative subscale (1.428) and a reduction in the positive one (-0.4). To enhance the credibility of our findings, we control for unobservable and time-invariant household's characteristics estimating a fixed-effect model. By employing this estimator, we are able to validate the adverse impact of changes in family structure, thus confirming previous findings.

A crucial assumption in the Difference-in-Differences estimation strategy is the *exogeneity assumption*, according to which the assignment of the treatment, the family structural change, is exogenous with respect to the children's outcomes. However, the validity of this assumption may be questioned due to the potential presence of selection bias. This bias arises from the fact that the occurrence of a structural change within a family may depend on household-specific characteristics[18] that are likely to have an impact on children's outcomes as well. Consequently, the estimated differences between treated and untreated children may also be affected. It is therefore strongly advised to control for the influence of confounding factors using experimental or quasi-experimental study designs (Varian, 2016).

As done in other studies, we implement propensity score matching (PSM) to tackle the selection into a family who experiences a structure change and obtain a counterfactual group, i.e. a group of untreated children, more similar, and then comparable, to the treated ones (Cid and Stokes, 2013; Hannan and Halpin, 2014; Christina J. Cross, 2019). The similarity between groups here refers to the likelihood of experiencing a family structure change conditional on pre-disruption observable characteristics, that is the propensity score. Children in families that experience structure change (treatment group) are matched to those in two-parent families (control group), based on their propensity score. Once treated and untreated children are matched, the impact of family structure is measured as difference in average outcomes of the two groups.

In this stage of the analysis, we actually combine two quasi-experimental methodologies by using the propensity score as an initial measure to obtain two groups of children, the treated and the untreated, that are balanced in terms of observable attributes; thereafter, we estimate the causal effect of family structure change as difference in the outcomes between the two groups, employing a Difference-in-Difference approach akin to the one used previously. Overall, we implement a Kernel Propensity Score Matching Difference-in-Differences approach.

Even results from this point of the study provide evidence of a detrimental impact of family structure change experienced between 3-5 years on children's behavioural outcomes and the quality of parent-child relationship at 5 years old.

To assess the robustness of our Kernel PSM DiD estimates, we change the algorithm and some of its parameters when performing the match between units. In particular, we use as alternative methods the Nearest Neighbour, the Radius, and the Normal density Kernel Matching. Estimates from all the specifications confirm the negative effects previously found.

Finally, we measure whether the adverse impact detected when the child is 5 years old is still in force when the child is aged 9. We find that after at least 4 years, family transition exerts a

[18] We use the term household-specific characteristics, but it may embrace a wide range of parents' characteristics as well.

negative impact on all the outcomes considered. The effects estimated are only slightly lower than the ones depicted when the child is 5 years old, apart from internalizing problems which actually increases. On average, children who experience a family structure change in early childhood present internalizing and externalizing problems higher by 0.984 and 0.721 respectively, and a worse relationship with the parent as the negative scale is 1.268 points greater and the positive one is 0.337 points lower than children in two-parent families when aged 9.

The paper proceeds as follows. Section II describes the data and provides some descriptive statistics. In Section III we illustrate our first econometric model implemented in the empirical analysis and present the related results. Section IV describes an alternative empirical strategy that overcomes issues related to the first one, present the main results and the robustness checks. In section V we present the impact of structure change measured after a longer period. Section VI offers some concluding remarks.

II. Data and Descriptive Statistics

In this study, we use data from the *Infant Cohort* of the panel study *Growing Up in Ireland (GUI).* The sample consists of 11,134 infants (aged nine months) born between December 2007 and June 2008. Given the longitudinal nature of the study, the families of the infant are interviewed again when the children are three, five, seven/eight, and finally when they are nine years old. The first data collection took place between September 2008 and April 2009; the second survey has been hold between December 2010 and July 2011 with 9,793 households participating; the third one took place between March and September 2013, resulting in 9,001 completed cases; the fourth wave of data collection took place in the spring of 2016 and was completed by 5,344 families; the fifth one took place between June 2017 and February 2018, resulting in 8,032 completed cases (Thornton *et al.*, 2013; Williams *et al.*, 2015; McNamara, Murray and O'Mahony, 2019). All this information is summarised in Table 1 below.

Table 1. Study design and response of the Infant Cohort of Growing Up in Ireland

Wave	I	II	III	IV	V
Data collection date	Sept 2008 - Apr 2009	Dec 2010 - Jul 2011	Mar - Sept 2013	Sept - Nov 2016	Jun 2017 - Feb 2018
Child's age	9 months	3 years	5 years	7/8 years	9 years
Participating households	11,134	9,793	9,001	5,344	8,032

The informants in the household were in all cases the Primary Caregiver (usually the mother) and, where relevant, the resident spouse/partner of the Primary Caregiver.

The first survey of data serves of little use in our empirical analysis since the outcomes of our interest were not observable at the time the child was 9 months old, but we take advantage of some relevant child's characteristics which are constant over time and recorded in that wave (such as a measure of their birthweight) as well as other household' and primary caregiver's information that will be useful in recovering the event of interest. We also decide to not use the fourth wave of data due to the limited number of records, as shown in Table 1, to avoid losing a lot of observations.

The focus of our research is to investigate how family structural change affects children's well-being and the quality of the relationship with their parents.

As dependent variables, we use *Externalizing* and *Internalizing problems scores* as measures of the child's emotional and behaviour attitudes. Both measures rely on a subset of questions from the Strength and Difficulties Questionnaire (Goodman, 1997). The externalizing problems reflects whether the child presents difficulties in interacting with others or if s/he turns problems outward through hyperactivity and misbehaviour. The internalizing problems assess whether the child feels fearful or too dependent, tends to isolate her/himself from other children, may be victim of bullying. Note that, the questionnaire was filled by the Primary caregiver, therefore our variables are a parent-report of their child's behaviour.

The quality of the parent-child relationship is assessed using the Child–Parent Relationship Scale (CPRS) proposed by Pianta (1992). This is a widely used self-report instrument completed by the parents to assess the perceptions of the relationship with their child. The scale measures both positive and negative aspects of that relationship through three dimensions: closeness, conflict, and dependency. The positive subscale assesses the degree to which the relationship is characterized by affection, warmth, instrumental support, positive involvement, and open communication, while the negative subscales measure to what extent a parent feels his/her relationship with the child is characterized by disagreement, anger and hostility, and conflict. Therefore, we use as dependent variables *Positive Subscale*, with ranges from 7 to 35 points, and *Negative Subscale*, which ranges between 8 and 40. In our sample, the Child–Parent Relationship Scale (CPRS) was administered to the primary caregiver. Consequently, all our outcomes result from a primary-caregiver viewpoint.

We embrace a few important criteria to define the family structure change. Firstly, we classify households in our sample into two family structure types based on whether the primary caregiver was living with a partner at the time of the interview: *two-parent type,* characterized by the presence of a partner cohabiting with the primary caregiver; *lone-parent type*, characterized by primary caregiver living alone with their children. Within these two types, we identify families that do not experience structural changes and families that do experience a family structure change from a two-parent type to a one-parent type based on variations in the presence of a partner observed across surveys. In particular, on one side we have households that do not encounter in any structural alterations as either the presence or the absence of a partner is consistently confirmed at each survey by the primary caregiver; on the other side, we have households that experience a structural change moving from a two-parent to a one-parent type between wave 2 and wave 3, that is when the child is aged between 3 and 5 years. We refer to the former group as untreated households and to the latter as treated households. Note that in our empirical analysis we do not consider households classified as lone-parent type at the time of the first data collection.

We decide to not base our detection of the family type on the primary caregiver's marital status because the data in question do not adequately make possible to distinguish between divorced and widow primary caregiver (both statuses are classified into one category), and, more important, because in many cases in which the primary caregiver declared never having been married, there was a partner living at home. Therefore, the analysis relates to structure change taking place in households made by cohabiting parents, regardless of their actual marital status. Note also that we do not consider households in which the reverse structural change takes place, that is household in which the primary caregiver reports not living with a partner in the first wave, and then the situation changes in the following interviews.

We perform our analysis using children from *two-parent family type* as control units and children from *one-parent family type* as treated units to measure the effect of a transition in the family structure by comparing the difference in the average outcomes for children who undergo the transition relative to the outcomes reported by children who do not.

Our final sample is made of almost 23,199 observations. Descriptive statistics are shown in Table 2. On average, children present an internalizing problems score of almost 2.56 points, an externalizing problems score of 4.65, Pianta negative scale is on average 15, while the positive scale is around 33.67 points. Almost 51% of children in our sample are boys, their average birthweight is 3.5 kilograms, only 3.5 % of them is born from a twin birth, 62% of children has been breastfed, and each child has on average 1.77 siblings. Approximately 100% of primary caregiver in our sample are women, they are 36.88 years old on average, with a high level of education (about 14.6 years). Almost 62% of them are employed, 3% unemployed, and the remaining 35% are outside the labour force. About 82% of primary caregiver are Irish, 3% are black, and 15% are of other ethnicities. 97% of the secondary caregiver are men, they are on average 39.6 years old, with 14.3 years of education. 83% are Irish, black ethnicity accounts for almost 3%, and the remaining 14% are of other ethnicities. Almost 84% of them are employed, 12% are unemployed, and 4% are outside of the labour force. On average, households in our sample fill in the third income quintile, while the income predicted in the absence of a structural change is slightly higher, 3.2, and the 58% of them live in a rural area. On average, 92% of families in are sample is made of two parents that cohabit in all the surveys, nearly 6 % are one-parent households' type since the time of the first survey, and the remaining 2% encounter in a structural change when the child is between 3 and 5 years old.

Table 2. Descriptive Statistics

Variable	Mean	Std. Dev.	Min	Max	Obs
Outcomes:					
Internalizing Problems Score	2.556	2.451	0	19	23199
Externalizing Problems Score	4.646	3.337	0	20	23199
Pianta Negative Scale - Pcg	15.046	5.569	8	40	23142
Pianta Positive Scale - Pcg	33.674	2.083	7	35	23160
Study Child Characteristics:					
Boy	.505	.5	0	1	23199
Birthweight	3.504	.534	1.5	4.6	22946
Twin birth	.035	.183	0	1	23199
Been breastfed	.623	.485	0	1	23193
Number of siblings	1.771	1.012	0	5	23199
Primary Caregiver Characteristics:					
Female	.997	.057	0	1	23199
Age	36.881	5.509	18	48	23199
Years of education	14.64	2.123	8	18	23197
Employed	.617	.486	0	1	23190
Unemployed	.033	.177	0	1	23190
Outside labour force	.35	.477	0	1	23190
Irish ethnicity	.826	.379	0	1	23199
Black ethnicity	.027	.162	0	1	23199
Other ethnicities	.147	.354	0	1	23199
Secondary Caregiver Characteristics:					
Male	.970	.170	0	1	22727
Age	39.607	5.984	18	68	21701
Years of education	14.289	2.264	8	18	20958
Irish ethnicity	.834	.372	0	1	20661
Black ethnicity	.025	.157	0	1	20661
Other ethnicities	.141	.348	0	1	20661
Employed	.839	.368	0	1	21675
Unemployed	.118	.323	0	1	21675
Outside labour force	.043	.203	0	1	21675
Household Characteristics:					
Annual Income - Quintiles	3.107	1.395	1	5	21702
Predicted Income - Quintiles	3.252	.917	1	5	20412
Rural area	.579	.494	0	1	23194
One-parent family type	.063	.243	0	1	23199
Two-parent family type	.921	.27	0	1	23199
Structural change between 3-5	.016	.125	0	1	23199

Notes: Dataset GUI - Infant Cohort (waves 2, 3, and 5)

III. Results from a Standard Difference-in-Differences Design

To provide evidence on the impact of family structural alteration on children's behavioural outcomes at age 5 and their relationship with the primary caregiver, we use a *Difference-in-*

Differences approach considering children exposed to the family structure change between 3-5 as treated and children not exposed to such a change, i.e. in a two-parent family type, as controls. The basic idea is to compare the difference in the average outcomes for treated children with those of untreated children before and after the transition takes place.

We use a linear estimator to estimate the following model:

$$Y_{it} = \alpha + \gamma Structural\ change_i + \theta age5 + \delta Structural\ change_i * age5 + \beta \boldsymbol{X}_{it} + \varepsilon_{it} \quad (1)$$

Where:

- Y_{it} represents the outcome of interest for the Study Child in household *i* at time *t*.
- $Structural\ change_i$ is a dummy variable which takes the value of 1 if in household i a structural change occurs between 3-5 child's age, and 0 otherwise. The coefficient γ measures the difference in the outcomes between children in one-parent and children in two-parent families before the transition takes place.
- $age5$ is a dummy taking the value of 1 the child is 5 years old, and 0 otherwise (i.e., it refers to the period after the change occurred). The coefficient θ measures the difference in the outcomes before and after the structural change.
- $Structural\ change_i * age5$ is the interaction term whose coefficient δ measures the treatment effect of our interest.
- X_{it} is a vector of household level covariates that could affect the family structural change. It includes primary and secondary caregivers' socio-demographics and economic resource of the household, as well as study child's characteristics (child's gender, birthweight, been breastfed, number of siblings; primary caregiver's age, age squared, years of education, Irish ethnicity, employment status; secondary caregiver's age, age squared, years of education, Irish ethnicity; area of residence, annual income, and income predicted in the absence of a structural change).
- ε_{it} is an error term.

Table 3 below shows the results obtained using an OLS estimator. The specification in column (1) reports the results of a very basic regression in which we only use as regressors $Structural\ change_{it}$, $age5$, and the interaction term $Structural\ change_{it} * age5$ and we do not add any control variables. We find that experiencing a family structural change produces negative and statistically significant effects on child's internalizing problems: children exposed to a family structure change when aged between 3 and 5 years present at the age of 5 higher internalizing problems by on average 0.486 relative to children in stably two-parent families.

In the column (2), we control for some child', primary caregiver', and secondary caregiver's characteristics to avoid unbalanced comparisons between treated and untreated children. Our results are robust to the inclusion of such explanatory variables since the effect of experiencing a family structural change is still detrimental and statistically significant at the 5 percent level: for children in the treated group, their internalizing problems increase on average by 0.533 relative to children who grow up in intact two-parent families. With regards to control variables, our findings show that: boys tend to have on average higher internalizing problems by 0.155, birthweight has a positive effect, since a higher birthweight significantly reduces the problems by on average 0.148 for each kilogram more, being born from a birth of twins raises the score by on average 0.239, having siblings reduces the internalizing problems almost by 0.093 for each sibling. Primary caregiver's age and education significantly reduce the child's internalizing problems by on average 0.042 and 0.049 respectively, while the age and education of the secondary caregiver do not exert any effect.

In column (3) we further control for whether the child has been breastfed, primary caregiver' and secondary caregiver's age squared, and household characteristics such as area of residence and income predicted in the absence of a structural change. We find that, of the included explanatory variables, living in a rural area decreases the internalizing problems by 0.151, and secondary caregiver's age has a U-shaped impact reaching a minimum at the age of 42.95, while the impact of the primary caregiver turns insignificant once the quadratic term has been included. The impact of the structural change remains statistically significant and positive, although at the 10 percent level and slightly decreases in magnitude: children exposed to a structural change in their family see their internalizing problems increase by on average 0.421 compared to children not exposed to such a change.

The fourth and fifth specifications enrich our analysis including as control variables primary and secondary caregivers' ethnicity and employment status. In addition, in the last column we substitute the predicted income with the actual one. When controlling for these new explanatory variables, the impact of the structural change is slightly larger (0.428 and 0.509 in columns (4) and (5) respectively) and statistically significant at the 5 percent level.

Our results show that children of Irish parents tend to have lower internalizing problems. While the employment status of the primary caregiver is important to mitigate child's problems, the employment status of the secondary caregiver does not produce any impact: children of employed primary caregiver have lower internalizing problems by on average 0.229 and 0.174 respectively than children whose primary caregiver is outside of the labour force. An increase in the annual income reduces child's problems by on average 0.07 (column 5).

Table 3. OLS estimates. Internalizing problems

	(1)	(2)	(3)	(4)	(5)
Structural change	-0.166	-0.305*	-0.303*	-0.302*	-0.385**
	(0.189)	(0.181)	(0.181)	(0.181)	(0.176)
Age 5	-0.0676**	-0.00941	0.108***	-0.0104	0.0357
	(0.0285)	(0.0329)	(0.0373)	(0.0438)	(0.0357)
Structural change*age 5	0.486**	0.533**	0.421*	0.428*	0.509**
	(0.236)	(0.240)	(0.234)	(0.234)	(0.249)
Boy		0.155***	0.149***	0.148***	0.146***
		(0.0438)	(0.0441)	(0.0441)	(0.0449)
Birthweight		-0.148***	-0.134***	-0.127***	-0.123***
		(0.0446)	(0.0451)	(0.0452)	(0.0458)
Twin		0.239*	0.288**	0.289**	0.335**
		(0.132)	(0.133)	(0.133)	(0.138)
Been breastfed			0.0654	0.0354	0.0499
			(0.0477)	(0.0489)	(0.0501)
Number of siblings		-0.0933***	-0.141***	-0.0810***	-0.0991***
		(0.0226)	(0.0241)	(0.0270)	(0.0242)
Pcg's age		-0.0415***	0.0421	0.0471	0.0302
		(0.00666)	(0.0639)	(0.0640)	(0.0649)
Pcg's age squared			-0.989	-1.074	-0.852
			(0.892)	(0.893)	(0.908)
Pcg's years of education		-0.0489***	-0.0118	-0.0392***	-0.0228*
		(0.0115)	(0.0135)	(0.0144)	(0.0127)
Irish ethnicity - Pcg				-0.177**	-0.119
				(0.0813)	(0.0779)
Employed – Pcg				-0.229***	-0.174***
				(0.0643)	(0.0520)
Unemployed - Pcg				0.111	0.102
				(0.118)	(0.119)
Scg's age		-0.000564	-0.0914**	-0.0881**	-0.0654
		(0.00554)	(0.0419)	(0.0421)	(0.0425)
Scg's age squared			1.064**	1.031**	0.778
			(0.513)	(0.515)	(0.523)
Scg's years of education		-0.00773	0.0148	-0.0171	-0.00663
		(0.0104)	(0.0115)	(0.0130)	(0.0112)
Irish ethnicity - Scg				-0.214***	-0.215***
				(0.0818)	(0.0800)
Employed - Scg				-0.127	-0.0844
				(0.109)	(0.106)
Unemployed - Scg				0.0935	0.0242
				(0.125)	(0.123)
Rural area			-0.151***	-0.0812*	-0.102**
			(0.0461)	(0.0481)	(0.0472)
Predicted income			-0.238***	0.0305	
			(0.0329)	(0.0643)	
Annual Income					-0.0714***
					(0.0209)
Observations	15221	14458	14088	14088	13515
Adjusted R^2	.000187	.0161	.0225	.0242	.0260

Notes: Dataset GUI - Infant Cohort (waves 2 and 3). The Table reports OLS estimates. The dependent variable is

Internalizing problems score. Standard errors, reported in parentheses, are corrected for heteroskedasticity, and allow for clustering at the household level. To make the reading fluent, in all the regressions we have rescaled the variables Age*2* dividing them by 1000, thus reducing the number of zeros after the decimal points. The symbols ***, **, * indicate that the coefficients are statistically significant at the 1, 5, and 10 percent level respectively.

In Table 4 we present the results for the impact of the structural change on children's externalizing problems. As before, the first column examines the effect of our interest without other controls; then, in the following columns, we progressively increase the number of control variables as done in Table 2. For the sake of brevity, we do not show the coefficients for these explanatory variables, but only for the parameters of interest.

We find that, in all specifications, children exposed to a change in their family structure and children not exposed to it do not present statistically significant difference in their externalizing problems before the occurrence of the change (the coefficient of *Structural change* is always statistically insignificant). Internalizing problems significantly tend to decrease over time (*Age5* is negative and strongly significant at the 1 percent level), however experiencing a family structure change adversely affects children's externalizing problems: treated children present greater externalizing problems relative to untreated children between 0.8 and 0.9 points. The effect is robust to the inclusion on explanatory variables whose significance level ranges between the 1 and the 5 percent.

Table 4. OLS estimates. Externalizing problems

	(1)	(2)	(3)	(4)	(5)
Structural change	0.0435	-0.215	-0.228	-0.234	-0.289
	(0.275)	(0.290)	(0.295)	(0.294)	(0.295)
Age 5	-0.612***	-0.585***	-0.502***	-0.702***	-0.508***
	(0.0374)	(0.0437)	(0.0501)	(0.0595)	(0.0480)
Structural change*age 5	0.899***	0.925***	0.810**	0.784**	0.856**
	(0.320)	(0.333)	(0.328)	(0.328)	(0.340)
Controls	NO	YES	YES	YES	YES
Observations	15221	14458	14088	14088	13515
Adjusted R^2	.009	.0515	.0513	.0537	.0558

Notes: Dataset GUI - Infant Cohort (waves 2 and 3). The Table reports OLS estimates. The dependent variable is *Externalizing problems score*. In column (2) we control for child's gender, birthweight, twin births, number of siblings, age and level of education of primary and secondary caregivers. In column (3) we further include a dummy variable for whether the child was breastfed, the squared age of both primary and secondary caregivers, rural area of residence and predicted income. Column (4) controls for the ethnicity and the employment status of primary and secondary caregivers. Finally, in column (5) we substitute the predicted income with the actual annual income. Standard errors, reported in parentheses, are corrected for heteroskedasticity, and allow for clustering at the household level. The symbols ***, **, * indicate that the coefficients are statistically significant at the 1, 5, and 10 percent level respectively.

Table 5 shows the OLS estimates of the impact of experiencing a structural change on the Pianta negative score for the five different specifications delineated above.

Results from column (1) highlight a strong significant and detrimental impact of the occurrence of the structural change on the quality of the parent-child relationship: children who experience a family structure change between the ages of 3 and 5 present a higher conflict score by 1.428.

Table 5. OLS estimates. Pianta negative score

	(1)	(2)	(3)	(4)	(5)
Structural change	0.417	0.164	0.122	0.159	0.168
	(0.467)	(0.469)	(0.470)	(0.472)	(0.472)
Age 5	-0.695***	-0.530***	-0.304***	-0.666***	-0.498***
	(0.0646)	(0.0754)	(0.0869)	(0.101)	(0.0825)
Structural change*age 5	1.428***	1.532***	1.347**	1.303**	1.334**
	(0.544)	(0.562)	(0.556)	(0.558)	(0.577)
Controls	NO	YES	YES	YES	YES
Observations	15181	14423	14057	14057	13490
Adjusted R^2	.0046	.0117	.0181	.0218	.0221

Notes: Dataset GUI - Infant Cohort (waves 2 and 3). Table reports OLS estimates. The dependent variable is *Pianta negative score*. In column (2) we control for child's gender, birthweight, twin births, number of siblings, age and level of education of primary and secondary caregivers. In column (3) we further include a dummy variable for whether the child was breastfed, the squared age of both primary and secondary caregivers, rural area of residence and predicted income. Column (4) controls for the ethnicity and the employment status of primary and secondary caregivers. Finally, in column (5) we substitute the predicted income with the actual annual income. Standard errors, reported in parentheses, are corrected for heteroskedasticity, and allow for clustering at the household level. The symbols ***, **, * indicate that the coefficients are statistically significant at the 1, 5, and 10 percent level respectively.

When we progressively control for a variety of explanatory variables, columns (2) – (5), the effects remain positive and statistically significant at the 1 and 5 percent levels in all the specifications: children exposed to a structural change face an increase in their conflict score with the primary caregiver between 1.5 and 1.3 points on average.

Finally, Table 6 reports the estimates of the impact of experiencing a structural change on the Pianta positive score.

Table 6. OLS estimates. Pianta positive score

	(1)	(2)	(3)	(4)	(5)
Structural change	0.155	0.166	0.206	0.215	0.190
	(0.151)	(0.157)	(0.156)	(0.155)	(0.158)
5 years old	-0.0588**	-0.0856***	-0.130***	-0.0745**	-0.104***
	(0.0261)	(0.0295)	(0.0332)	(0.0379)	(0.0323)
Structural change*age 5	-0.400**	-0.337*	-0.356*	-0.374*	-0.337*
	(0.203)	(0.196)	(0.195)	(0.194)	(0.198)
Controls	NO	YES	YES	YES	YES
Observations	15191	14433	14064	14064	13498
Adjusted R^2	.0003	.0095	.0108	.0115	.0114

Notes: Dataset GUI - Infant Cohort (waves 2 and 3). Table reports OLS estimates. The dependent variable is *Pianta positive score*. In column (2) we control for child's gender, birthweight, twin births, number of siblings, age and level of education of primary and secondary caregivers. In column (3) we further include a dummy variable for whether the child was breastfed, the squared age of both primary and secondary caregivers, rural area of residence and predicted income. Column (4) controls for the ethnicity and the employment status of primary and secondary caregivers. Finally, in column (5) we substitute the predicted income with the actual annual income. Standard errors, reported in parentheses, are corrected for heteroskedasticity, and allow for clustering at the household level. The symbols ***, **, * indicate that the coefficients are statistically significant at the 1, 5, and 10 percent level respectively.

Results in all the specifications show a statistically significant and negative impact of the structural change on the positive aspect of the parent-child relationship for treated children: the effect is found to vary between - 0.4 and -0.3 points; it is robust to the inclusion of control variables, although the magnitude and the level of significance moderately decline (columns 2-5).

All in all, our results show a detrimental impact of experiencing a family structural change when aged between 3 and 5 years on both children's behavioral problems and on the quality of their relationship with the primary caregiver measured when the child is 5 years old.

In table 7 below, to further validate our results, we estimate fixed-effect models including household-fixed effects which allow to account for unobservable time-invariant household characteristics that are constant within the household but change across units. The fixed-effect models do not allow to estimate the effect of time-invariant variables; therefore we could not include such variables as controls in our models.

The new estimated model is the following:

$$Y_{it} = \delta_i + \lambda_t + \text{Q}\boldsymbol{X}_{it} + \delta Structural\ change_i + \varepsilon_{it} \quad (2)$$

Where, compared to the previous model, we include household-fixed effects δ_i, and time-fixed effects λ_t. $\boldsymbol{X}_{it}$ is a vector of time variant characteristics including primary and secondary caregiver's age squared and employment status, and household income predicted in the absence of a family transition. $Structural\ change_i$ is a dummy variable taking the value of 1 if in household *i* a change in family structural occurs between two surveys, and the value of 0 otherwise. Note that, since we restrict our analysis only to the group of children who do not experience any structural change and the one that does, not only $Structural\ change_i = 0$ for both groups in the first period, but also it equals the interaction term $Structural\ change_{it} * age5$ used in Equation (1) above. The specification in Eq. (2) defines the so-called generalized DiD (Duflo, Bertrand and Mullainathan, 2004).

The results are presented in Table 7 where, for each outcome, the first column shows the effect when only the variable of interest is included in our model, and the second column shows the effect when the explanatory variables $\boldsymbol{X}_{it}$ are added.

Reassuringly, results in all the specifications paint a similar picture as done by the results in the earlier models, highlighting a statistically significant and detrimental effect of experiencing a family structural change on all the outcomes considered. The magnitude of the effects is somewhat smaller relative to the ones depicted above for all the outcomes but for the Pianta positive score, which is slightly greater; the direction of the effect is the same as before, and the level of significance is almost identical. The effects are robust to the inclusion of the explanatory variables.

Table 7. Fixed-effects estimates

	Internalizing		Externalizing		Negative		Positive	
Structural change	0.477**	0.459**	0.860***	0.740**	1.320**	1.308**	-0.391*	-0.391**
	(0.236)	(0.233)	(0.320)	(0.328)	(0.544)	(0.565)	(0.203)	(0.193)
Age 5	-0.0587**	-0.207	-0.573***	-1.249***	-0.687***	-2.573***	-0.0684***	-0.218
	(0.0284)	(0.255)	(0.0369)	(0.331)	(0.0635)	(0.560)	(0.0258)	(0.222)
Controls	NO	YES	NO	YES	NO	YES	NO	YES
Observations	15221	14247	15221	14247	15557	14562	15569	14570
Adjusted R^2	.001	.001	.0328	.0324	.0159	.0168	.0016	.0031

Notes: Dataset GUI – Infant Cohort (waves 2 and 3). The Table reports Fixed-effect estimates. The dependent variables are *Internalizing and Externalizing problems score, Pianta negative and positive scores*. Control variables included in the second specification for each outcome include the squared age and employment status of primary and secondary caregivers, as well as household predicted income. Standard errors, reported in parentheses, are corrected for heteroskedasticity, and allow for clustering at the household level. The symbols ***, **, * indicate that the coefficients are statistically significant at the 1, 5, and 10 percent level respectively.

IV. Results from a Propensity Score Matching Approach

Within the DiD strategy, the identification of the effect of the treatment of interest revolves around the satisfaction of two primary assumptions: the *parallel trend assumption,* according to which, in the absence of any structural change, the average outcome for the treated children would have followed the same path as the average outcome for the untreated ones, and the *exogeneity assumption*, which assumes that the assignment of the treatment, the family structural change, is exogenous with respect to the children outcomes. In general, the first assumption cannot be directly tested since the counterfactual is not observable. Moreover, in our study it is not possible to verify it indirectly since we do not have other pre-treatment periods available before the transition takes place. With regards to the second assumption, it might be implausible for the assumption to hold as there might be selection bias: the fact that a family encounters in a structural change might depend on several factors that are also likely to influence children's outcome and, therefore, the estimated differences between treated and untreated children.

It is rather controversial in the literature whether the differences observed in children's outcomes across different family types are to a great extent, or even entirely, determined by selection bias (Gruber, 2004; Amato, 2010; Amato and Anthony, 2014). Families that face a change in their structure are likely to differ in many characteristics from families that do not alter their composition, i.e. these families are a selected group. Accordingly, children from the two groups would have different outcomes both before and after the transition takes place. Our results partially reassure us in this sense since we do not find pre-existing statistically significance differences for three out of four of our outcomes between the two groups of children. However, it should be bearded in mind that we observe our outcomes only in one pre-family disruption period. Some differences are instead observed in terms of observable characteristics.

To address the potential selection bias due to differences in family type's attributes, in this section we make use of matching approach in the building of the counterfactual group (Paul R. Rosenbaum and Donald B. Rubin, 1983; Rubin, 2001, 2006; Austin, 2011). The general idea of matching methods is to find a large group of untreated units that are "similar" to the treated ones on almost all relevant observed pre-treatment characteristics **X**. Thereafter, since treated and matched units now only differ in their treatment status, it is possible to estimate the impact of a family structure change as difference between the average outcome for the treated and the average outcome for the matched comparison units.

Matching can be done on a range of variables, but it is limited in the case of a high dimensional vector **X.** For this reason, Paul R. Rosenbaum and Donald B. Rubin (1983) suggest to carry out the matching on the propensity score (thereafter referred to as PSM), which is the probability of receiving the "treatment" as a function of the observed covariates **X**. Once the matching has been performed and its quality has been assessed, we proceed with the estimation of the effect of interest on the difference observed in the outcomes between the treated group of children and the matched one.

We estimate the predicted probability of receiving the treatment, that is the propensity score, using a logit model. The analysis at this stage entails a few practical decisions. The first choice concerns the control variables to include in the model. We select a variety of covariates among child', primary and secondary caregiver', and household's characteristics measured prior to the family structure change (Rubin, 2001; Ho *et al.*, 2007) or that are time invariant to minimise reverse causality. In other words, we use variables measured when the child is 3 years old or fixed, as an example their birthweight, to predict the probability of being exposed to a family structural change when aged between 3 and 5. Results are reported Table 1A in the Appendix. It is important to include variables that affect both the probability of receiving the treatment and the outcome, but these variables should not be affected by the treatment.

Second, to match treated children with untreated ones we use as main algorithm the Kernel matching technique as recommended by E. Leuven and Sianesi (2003) focusing on observations that fell in the region of common support and using a biweight kernel function. The ***kernel matching*** is a non-parametric matching estimator which uses weighted averages of all observations in the control group to construct counterfactual outcome. It places higher weight on observations close in terms of propensity score (i.e., with similar probability of receiving the treatment) and lower weight on more distant observations (i.e., with very different probability of receiving the treatment). Quality of post matching covariates balance is assessed using mean differences between treated and matched observations along with t test and p-value. Results are shown in Table 8.

Table 8. Differences between treatment and control units with balancing property

Covariate Variables			Mean		T-test	
			Treated	Control	t	p-value
Child:	Boy	U	0.551	0.504	0.88	0.381
		M	0.545	0.515	0.41	0.684
	Birthweight	U	3.394	3.515	-2.11	0.035
		M	3.401	3.490	-0.99	0.323
	Been breastfed	U	0.663	0.638	0.49	0.624
		M	0.670	0.645	0.36	0.720
	Twin	U	0.079	0.035	2.17	0.030
		M	0.068	0.047	0.61	0.546
Pcg:	Age	U	33.596	34.808	-2.53	0.011
		M	33.739	34.581	-1.10	0.272
	Age2	U	1.159	1.232	-2.23	0.026
		M	1.167	1.218	-1.00	0.321
	Irish ethnicity	U	0.865	0.844	0.55	0.581
		M	0.864	0.850	0.25	0.804
	Has 2 children	U	0.483	0.341	2.80	0.005
		M	0.477	0.372	1.41	0.160
	Has 3 children	U	0.236	0.380	-2.79	0.005
		M	0.239	0.353	-1.66	0.099
	Has 4 children	U	0.101	0.163	-1.57	0.117
		M	0.102	0.151	-0.97	0.331
	Has 5 children	U	0.056	0.060	-0.14	0.890
		M	0.057	0.059	-0.07	0.943
	Roman Catholic	U	0.899	0.883	0.47	0.640
		M	0.898	0.887	0.22	0.824
	BMI	U	26.442	25.867	1.14	0.255
		M	26.467	25.992	0.58	0.564
Scg:	Age	U	36.315	37.156	-1.47	0.140
		M	36.466	37	-0.63	0.531
	Age2	U	1.354	1.409	-1.27	0.204
		M	1.363	1.399	-0.56	0.578
	Irish ethnicity	U	0.831	0.844	-0.32	0.747
		M	0.830	0.843	-0.24	0.581
	BMI	U	26.570	25.824	0.83	0.405
		M	26.631	26.034	0.55	0.581
Family:	Age gap	U	3.910	3.340	1.64	0.101
		M	3.932	3.451	0.91	0.362
	Education gap	U	1.921	1.593	1.78	0.076
		M	1.921	1.663	0.89	0.377
	Same economic status	U	0.483	0.560	-1.44	0.149
		M	0.489	0.545	-0.75	0.453
	Same ethnicity	U	0.865	0.901	-1.12	0.264
		M	0.864	0.893	-0.58	0.559
	Rural area	U	0.528	0.596	-1.29	0.198
		M	0.534	0.583	-0.66	0.511

Notes: Dataset GUI – Infant Cohort (Wave 2). U: Unmatched, M: Matched. Matching algorithm: kernel biweight function; common support, bandwidth: 0.06.

A good covariate balances is met since the differences that were statistically significant when comparing the untreated units turn statistically insignificant after matched. It is also evident that the two types of families differ in some important characteristics: in particular, primary caregiver in treated households are younger, have lower children, and a higher BMI. Similarly, children in treated families weighted significantly less at birth, and a higher percentage of them come from a twin birth. The secondary caregiver is typically younger in treated households. Instead of considering the education level of primary and secondary caregivers in years, we use a measure of the absolute difference in their level of education, as well as of their age. In the treated households, both the age and education gaps between the two parents are higher, although only the latter is statistically significant. Furthermore, we create an indicator of whether primary and secondary caregivers have the same employment status and ethnicity. Although not statistically significant, both measures are lower in the treated households, confirming that there is more heterogeneity among couples in families that experience a structure change.

These differences emphasize that, indeed, treated households present worse characteristics than untreated households that may negatively affect children's outcomes.

The final step of the analysis consists in estimating the average treatment effect on the treated (ATT) using as before a Difference-in-Differences approach, except that the two groups of children are now formed on one hand by the analogous treated children, and on the other hand by the matched ones which present a similar probability of experiencing a structural change. Essentially, we combined the two methodologies exploiting the benefit of the PSM to identify more comparable groups of children as a preliminary step to the implementation of the DiD approach with the purpose of mitigating the potential selection bias.

Table 9 displays the results from the Kernel Propensity Score Matching Difference-in-Differences of the effects of family structure change on all the outcomes considered. The first and second columns show the raw differences, i.e. before matching, and the PSM estimates respectively for each child outcomes exclusively for observations that fall into the region of common support. The last two columns show the groups' size.

The PSM estimates confirm a significant and adverse effect of family structure change on all the children's outcomes: experiencing such a transition causes children's internalizing and externalizing problems to increase by 0.475 and 0.812 points, respectively, compared to untreated matched children; the quality of the relationship with the primary caregiver declines, with the Pianta positive score decreasing by 0.448, and the conflict score increasesing by 1.045 on average. The significance level of these effects ranges between 1 and 10 percent.

Table 9. PSM estimates. Differences in Outcomes for Children in One-parent Families and Children in Two-parent Families

Outcomes	Difference Unmatched	Difference ATT	Treated Units	Control Units
Internalizing problems	0.426*	0.475*	88	5864
	(0.252)	(0.269)		
Externalizing problems	0.906***	0.812***	88	5864
	(0.327)	(0.372)		
Pianta positive	-0.414*	-0.448*	88	5843
	(0.228)	(0.231)		
Pianta negative	0.973*	1.045*	88	5839
	(0.566)	(0.651)		

Notes: Dataset GUI – Infant Cohort (waves 2 and 3). The Table reports the ATT (Average treatment effect for the treated) from the Kernel PSM DiD analysis. The dependent variables are the difference in the outcomes between treated and matched children. Standard errors do not take into account that the propensity score is estimated. The symbols ***, **, * indicate that the coefficients are statistically significant at the 1, 5, and 10 percent level respectively.

IV.I Robustness Checks

In this section, we evaluate the robustness of the results from the PSM approach by changing the matching algorithm or by changing the parameters of a given algorithm. This is useful to strengthen the reliability of the results by showing that the estimations do not depend crucially on the particular methodology chosen.

We estimate the average treatment effect on the treated (ATT) as before using the following algorithms: ***Nearest Neighbour (NN) Matching*** (with five neighbours), according to which each unit from the treated group is matched with the closest units in terms of propensity score from the comparison group. We implement the NN matching with replacement, allowing an untreated unit to be used more than once as a match, and we impose a tolerance level on the maximum propensity score distance (caliper) equal to one-quarter of the propensity score standard deviation. ***Radius Matching,*** which is a variant of the caliper matching, that uses not only the nearest neighbours within each caliper, but all of the comparison units within the caliper with the advantage of resorting to the using of extra units in case good matches are not available. Finally, we change in the ***Kernel Matching*** the kernel function and the bandwidth parameters compared to the ones used in in the paragraph above, choosing a normal density function and narrowing the bandwidth to 0.05.

Table 10 presents the results. Our results confirm the significant and detrimental effect of experiencing a structural change on children's behavioural outcomes and quality of their relationship with the parent, implying that our findings measure the causal effect of the transition in the family organization and do not depend on a specific matching algorithm implemented or on some of its parameters.

Table 10. PSM estimates. Differences in Outcomes for Children in One-parent Families and Children in Two-parent Families – Robustness Checks

Outcomes	(5) Nearest Neighbor	Radius	Normal Kernel
Internalizing problems	0.528*	0.401	0.462*
	(0.298)	(0.276)	(0.269)
Externalizing problems	0.754*	0.643*	0.817***
	(0.408)	(0.380)	(0.371)
Pianta positive	-0.41*	-0.368*	-0.448*
	(0.262)	(0.238)	(0.231)
Pianta negative	1.587***	1.114*	1.058*
	(0.687)	0.649)	(0.651)

Notes: Dataset GUI - Infant Cohort (waves 2 and 3). Each column reports the ATT (Average treatment effect for the treated) from a different matching algorithm: (1) K-Nearest Neighbor using k=5 nearest neighbors; (2) Radius matching with a caliper of one-quarter of the standard deviation of the propensity score; (3) Kernel matching using Normal density function. A caliper of one-quarter of the standard deviation of the propensity has been employed also in columns (1) and (2). Standard errors do not take into account that the propensity score is estimated. The symbols ***, **, * indicate that the coefficients are statistically significant at the 1, 5, and 10 percent level respectively.

V. Long-lasting Effects of Family Structure Change

In the above sections, we have estimated the impact of a family structure change occurring between the ages of 3-5 on children's behavioral outcomes and quality of the relationship with the parent when the child is 5 years old, that is in a relative short period of time.

In this section, we want to verify whether a structural change experienced in early childhood still exerts its effects when the child is aged 9. That is, we aim at examining the long-term effects of a family structure change, ensuring at least 4-year lag between the occurrence of the event and the measurement of the children's outcomes.

Table 11 presents estimates of the effect of a family structural change experienced between 3 and 5 years of age on children's outcomes when aged 9 for all the outcomes considered in the previous analysis.

For greater convenience in the reading, we present only the coefficients of interest, that is the one that measures the difference in the outcomes between children in two-parent households and children in one-parent households which experience the family structure in question. For all the outcomes, we estimate the DiD with both OLS and fixed-effects models, without and with covariates respectively, as presented in the first and second columns in the Table below.

For the OLS estimates, the vector of explanatory variables includes child's gender, birthweight, been breastfed, number of siblings; primary and secondary caregiver's age and age squared, years of education, Irish ethnicity, employment status; area of residence, income predicted in the absence of a structural change. The set of control variables included in the FE model consists of primary and secondary caregivers age squared and predicted income.

The results emphasize that experiencing a family structure change when aged between 3-5 has strong adverse long-lasting effects. The impact on child's internalizing problems almost doubles relative to the one depicted at the age of 5; the effects on the other outcomes maintain almost the same magnitude.

Table 11. Long-lasting Effects of Family Structure Change on Children's Outcomes

	OLS Estimates		FE Estimates	
	No covariates	With covariates	No covariates	With covariates
Internalizing	1.002***	1.066***	0.984***	1.015***
	(0.279)	(0.314)	(0.279)	(0.285)
Externalizing	0.827**	0.942***	0.721**	0.712**
	(0.329)	(0.357)	(0.329)	(0.343)
Pianta positive	-0.359***	-0.566***	-0.337*	-0.315*
	(0.195)	(0.050)	(0.195)	(0.193)
Pianta negative	1.333**	1.348**	1.268*	1.216*
	(0.640)	(0.708)	(0.647)	(0.663)

Notes: Dataset GUI - Infant Cohort (waves 2 and 5). The Table reports OLS and Fixed effect estimates of the long-term effects of family structure change on children's outcomes, that is when children are 9 years old. The dependent variables are *Internalizing* and *Externalizing problems scores*, *Pianta negative and positive scores*. Standard errors, reported in parentheses, are corrected for heteroskedasticity, and allow for clustering at the household level. The symbols ***, **, * indicate that the coefficients are statistically significant at the 1, 5, and 10 percent level respectively.

On average, experiencing a family structure change in early childhood produces an increase in child's internalizing and externalizing problems by 1.07 and 0.942 points respectively; it also adversely affects the quality of the relationship with the primary caregiver, causing a decrease in the positive scale by 0.57 and an increase in the negative one on the negative scale 1.35 (OLS estimates). When including fixed effects, the coefficients are slightly lower, but direction and significance level are preserved.

In what follow, we present the results of the long-lasting effects of experiencing a family structure change estimated using the PSM-DiD approach to mitigate the potential selection bias as done before.

Column (1) in Table 12 shows the estimates from the main PSM algorithm employed, the Kernel Matching with a biweight kernel function; the following columns present the robustness checks of the PSM estimates in which we change the PSM algorithm or modify some parameters. Only for the main estimates we show both the raw differences and the matched one; for the others, we present only the matched differences.

The results confirm what found before, showing that experiencing a family structure change between 3-5 years of age exerts strongly negative effects even when the child is 9 years old. The results are robust to variations of the matching algorithm or of its parameters.

Table 12. PSM estimates. Differences in Outcomes for Children in One-parent Families and Children in Two-parent Families. Long-lasting Effects

Outcomes	Biweight Kernel Difference Unmatched	Biweight Kernel Difference ATT	Nearest Neighbor (5)	Radius	Normal Kernel
Internalizing problems	1.059***	1.052***	1.095***	1.065***	1.066***
	(0.306)	(0.319)	(0.351)	(0.316)	(0.318)
Externalizing problems	0.895***	0.869***	0.973***	0.882***	0.880***
	(0.368)	(0.406)	(0.442)	(0.403)	(0.406)
Pianta positive	-0.309*	-0.334*	-0.359	-0.340*	-0.340*
	(0.266)	(0.207)	(0.241)	(0.204)	(0.207)
Pianta negative	1.121*	1.185*	1.802***	1.195	1.200*
	(0.635)	(0.804)	(0.850)	(0.799)	(0.804)

Notes: Dataset GUI - Infant Cohort (waves 2 and 5). Each column reports the ATT (Average treatment effect for the treated) from a different matching algorithm: (1) Kernel Matching using a biweight function – main estimates; (2) K-Nearest Neighbor using k=5 nearest neighbors; (3) Radius matching with a caliper of one-quarter of the standard deviation of the propensity score; (4) Kernel matching using Normal density function and a bandwidth of 0.05. A caliper of one-quarter of the standard deviation of the propensity has been employed also in columns (1) and (2). Standard errors do not take into account that the propensity score is estimated. The symbols ***, **, * indicate that the coefficients are statistically significant at the 1, 5, and 10 percent level respectively.

VI. Concluding remarks

In this paper we investigate the consequences of experiencing a family structure change for children aged between 3 and 5 at the time of the transition both after a relatively short period after the event took place, when children are aged 5, and after a longer period, when children are 9 years old, ensuring in the latter case at least 4-year lag between the event and the evaluation.

To this end, we exploit two quasi-experimental methodologies: Propensity Score Matching and Difference-in-Differences to estimate the average treatment effect of family structure on children's behavioral outcomes, as measured in terms of *Externalizing* and *Internalizing problems,* and parent-child relationship quality, as measured by the *Pianta positive* and *negative scores*.

In the first part of the study, we employ a standard DiD strategy and compare the variations over time in the outcomes between children who experience a family structure change when aged between 3-5 with the variations reported by children who do not experience such a change. We document a strong and persistent detrimental effect of family structure on all the outcomes measured: the internalizing problems for 5-year-old children in disrupted families increase on average by 0.486 and the difference doubles reaching 1 more point at the age of 9; externalizing problems grow by 0.9 and 0.827 in the short- and long post-structure change period, respectively. The quality of the relationship with the parent deteriorates under both the dimensions, with a higher score for the negative subscale by 1.428 and 1.33, and a lower one for the positive subscale by -0.337 and -0.36 when aged 5 and 9 respectively. The results remain robust when we control for a variety of child', primary and secondary caregivers' characteristics, as well as household's attributes. Furthermore, the effects detected hold even when we include household-fixed effects in our model.

In the second part of the study, we tackle potential selection bias that may arise from the fact that households who experience a family structure change might be a selected group of families characterized by specific qualities and, presumably, worse than the ones of intact two-parent households. These differences may affect children's outcomes beforehand the transition itself. As a way of dealing with this, we implement a Propensity Score Matching with which children in treated households are matched with children in untreated households, but that present a similar probability of being exposed to the family structure change, measured by the so-called propensity score. We estimate that probability of being exposed to a family structure change with a logit model, and then match the children using a Kernel matching procedure. Following the match and the covariates balance check, we employ a DiD using the sample of treated and matched untreated children. Results confirm the detrimental impact of family structure change on all the outcomes considered, with a modest upward bias in the estimates obtained without matching.

We strengthen our results by checking the robustness of our Kernel PSM-DiD estimates by changing the algorithm used and some of its parameters. The results are robust to such variations.

Overall, our findings shed light on the persistent consequences of family structure change on children who experience it when they are between 3 and 5 years old. Our results are in line with findings from the existing literature that assess a detrimental impact of family structural change if experienced in early childhood (Amato, 2010; Amato and James, 2010). We further show that these adverse effects do not attenuate over time but, conversely, tend to persist or even worse as the child grows up in line with other studies (Paul R Amato and Keith, 1991; Amato and Sobolewski, 2001; Ermisch *et al.*, 2001; Amato, 2005).

References

Abufhele, A., Behrman, J. and Bravo, D. (2017) 'Parental preferences and allocations of investments in children's learning and health within families', *Social Science and Medicine*, 194, pp. 76–86. Available at: https://doi.org/10.1016/j.socscimed.2017.09.051.

Amato, P.R. (1996) 'Explaining the Intergenerational Transmission of Divorce', *Journal of Marriage and the Family*, 58(3), p. 628. Available at: https://doi.org/10.2307/353723.

Amato, P.R. (2000) 'The consequences of divorce for adults and children', *Journal of Marriage and Family*. National Council on Family Relations, pp. 1269–1287. Available at: https://doi.org/10.1111/j.1741-3737.2000.01269.x.

Amato, P.R. (2005) 'The Impact of Family Formation Change on the Cognitive, Social, and Emotional Well-Being of the Next Generation', *The Future of Children*, 15(2), pp. 75–96.

Amato, P.R. (2010) 'Research on divorce: Continuing trends and new developments', *Journal of Marriage and Family*, 72(3), pp. 650–666. Available at: https://doi.org/10.1111/j.1741-3737.2010.00723.x.

Amato, P.R. (2014) 'The consequences of divorce for adults and children: an update', *Drustvena Istrazivanja*, 23(1), pp. 5–24. Available at: https://doi.org/10.5559/di.23.1.01.

Amato, P.R. (2016) 'The Consequences of Divorce for Adults and Children Published by : National Council on Family Relations Linked references are available on JSTOR for this article : The Consequences of Divorce for Adults and Children', *Journal of Marriage and Family*, 62(4), pp. 1269–1287.

Amato, P.R. and Anthony, C.J. (2014) 'Estimating the effects of parental divorce and death with fixed effects models', *Journal of Marriage and Family*, 76(2), pp. 370–386. Available at: https://doi.org/10.1111/jomf.12100.

Amato, P.R. and Cheadle, J.E. (2008) 'Parental divorce, marital conflict and children's behavior problems: A comparison of adopted and biological children', *Social Forces*, 86(3), pp. 1139–1161. Available at: https://doi.org/10.1353/sof.0.0025.

Amato, P.R. and DeBoer, D.D. (2001) 'The transmission of marital instability across generations: Relationship skills or commitment to marriage?', *Journal of Marriage and Family*, 63(4), pp. 1038–1051. Available at: https://doi.org/10.1111/j.1741-3737.2001.01038.x.

Amato, P.R. and James, S. (2010) ‘Divorce in Europe and the United States: Commonalities and differences across nations’, *Family Science*, 1(1), pp. 2–13. Available at: https://doi.org/10.1080/19424620903381583.

Amato, Paul R and Keith, B. (1991) ‘Linked references are available on JSTOR for this article : Parental Divorce and Adult Well-being : A Meta-analysis’, 53(1), pp. 43–58.

Amato, Paul R. and Keith, B. (1991) ‘Parental divorce and the well-being of children: A meta-analysis’, *Psychological Bulletin*, 110(1), pp. 26–46. Available at: https://doi.org/10.1037/0033-2909.110.1.26.

Amato, P.R. and Sobolewski, J.M. (2001) ‘The effects of divorce and marital discord on adult children’s psychological well-being’, *American Sociological Review*, 66(6), pp. 900–921. Available at: https://doi.org/10.2307/3088878.

Anderson, J. (2014) ‘The impact of family structure on the health of children: Effects of divorce’, *Linacre Quarterly*, 81(4), pp. 378–387. Available at: https://doi.org/10.1179/0024363914Z.00000000087.

Andini, M. *et al.* (2018) ‘Targeting with machine learning: An application to a tax rebate program in Italy’, *Journal of Economic Behavior and Organization*, 156(March), pp. 86–102. Available at: https://doi.org/10.1016/j.jebo.2018.09.010.

Antulov-Fantulin, N., Lagravinese, R. and Resce, G. (2021) ‘Predicting bankruptcy of local government: A machine learning approach’, *Journal of Economic Behavior and Organization*, 183, pp. 681–699. Available at: https://doi.org/10.1016/j.jebo.2021.01.014.

Arlot, S. and Celisse, A. (2010) ‘A survey of cross-validation procedures for model selection’, *Statistics Surveys*, 4, pp. 40–79. Available at: https://doi.org/10.1214/09-SS054.

Arpino, B., Moglie, M. Le and Mencarini, L. (2018) ‘Machine-Learning techniques for family demography : An application of random forests to the analysis of divorce determinants in Germany’, *Annual Meeting Paa*, (56), p. 32.

Aughinbaugh, A., Pierret, C.R. and Rothstein, D.S. (2005) ‘The impact of family structure transitions on youth achievement: Evidence from the children of the NLSY79’, *Demography*, 42(3), pp. 447–468. Available at: https://doi.org/10.1353/dem.2005.0023.

Augustine, J.M. and Kimbro, R.T. (2013) ‘Journal of Applied Research on Children: Informing Policy for Children at Risk’, *Journal of Applied Research on Children: Informing Policy for*

Children at Risk, 4(1), pp. 0–24. Available at: http://digitalcommons.library.tmc.edu/childrenatrisk%5Cnhttp://digitalcommons.library.tmc.edu/childrenatrisk/vol4/iss1/5.

Augustine, J.M. and Kimbro, R.T. (2017) 'Associations and Intervening Mechanisms Between Family Structure and Young Children's Obesity', *Journal of Family Issues*, 38(16), pp. 2277–2302. Available at: https://doi.org/10.1177/0192513X15621344.

Austin, P.C. (2011) 'An introduction to propensity score methods for reducing the effects of confounding in observational studies', *Multivariate Behavioral Research*, 46(3), pp. 399–424. Available at: https://doi.org/10.1080/00273171.2011.568786.

Beets, M.W., Cardinal, B.J. and Alderman, B.L. (2010) 'Parental social support and the physical activity-related behaviors of youth: A review', *Health Education and Behavior*, 37(5), pp. 621–644. Available at: https://doi.org/10.1177/1090198110363884.

Belloni, A. *et al.* (2012) 'Sparse Models and Methods for Optimal Instruments With an Application to Eminent Domain', *Econometrica*, 80(6), pp. 2369–2429. Available at: https://doi.org/10.3982/ecta9626.

Belloni, A., Chernozhukov, V. and Hansen, C. (2014) 'on Structural and Treatment Effects †', *Journarl of economci perspectives*, 28(2), pp. 29–50.

Berenson, G.S. (2012) 'Health Consequences of Obesity', *Pediatric Blood and Cancer*, 58(1), pp. 117–121. Available at: https://doi.org/10.1002/pbc.23373.

Bernardi, F. and Boertien, D. (2016) 'Understanding heterogeneity in the effects of parental separation on educational attainment in Britain: Do children from lower educational backgrounds have less to lose?', *European Sociological Review*, 32(6), pp. 807–819. Available at: https://doi.org/10.1093/esr/jcw036.

Biblarz, T.J. and Gottainer, G. (2000) 'Family structure and children's success: A comparison of widowed and divorced single-mother families', *Journal of Marriage and Family*, 62(2), pp. 533–548. Available at: https://doi.org/10.1111/j.1741-3737.2000.00533.x.

de Blasio, G. and D'ignazio, A. (2021) 'Regional policy in the machine learning era: Opportunities and challenges', *Scienze Regionali*, 20(2), pp. 205–220. Available at: https://doi.org/10.14650/99817.

Bradley RH, C.R. (2008) 'Infant temperament, parenting, and externalizing behavior in first grade:

a test of the differential susceptibility hypothesis', *J Child Psychol Psychiatry*, 49(2), pp. 124–31.

Bream, V. and Buchanan, A. (2003) 'Distress among Children whose Separated or Divorced Parents cannot Agree Arrangements for Them', *British Journal of Social Work*, 33(2), pp. 227–238. Available at: https://doi.org/10.1093/bjsw/33.2.227.

Bright, M.A. *et al.* (2015) 'Adverse childhood experiences and dental health in children and adolescents', *Community Dentistry and Oral Epidemiology*, 43(3), pp. 193–199. Available at: https://doi.org/10.1111/cdoe.12137.

Brown, S.L. (2004) 'Family structure and child well-being: The significance of parental cohabitation', *Journal of Marriage and Family*, 66(2), pp. 351–367. Available at: https://doi.org/10.1111/j.1741-3737.2004.00025.x.

Bzostek, S.H. and Beck, A.N. (2011) 'Familial instability and young children's physical health', *Social Science and Medicine*, 73(2), pp. 282–292. Available at: https://doi.org/10.1016/j.socscimed.2011.04.014.

Carrieri, V., Lagravinese, R. and Resce, G. (2021) 'Predicting vaccine hesitancy from area-level indicators: A machine learning approach', *Health Economics (United Kingdom)*, 30(12), pp. 3248–3256. Available at: https://doi.org/10.1002/hec.4430.

Cavanagh, S. and Fomby, P. (2019) 'Family Instability in the Lives of American Children', *Annual Review of Sociology*, 45(1), pp. 493–513. Available at: https://doi.org/10.1146/annurev-soc-073018-022633.

Cavanagh, S.E. and Huston, A.C. (2006) 'Family instability and children's early problem behavior', *Social Forces*, 85(1), pp. 551–581. Available at: https://doi.org/10.1353/sof.2006.0120.

Cerqua, A. *et al.* (2021) 'Local mortality estimates during the COVID-19 pandemic in Italy', *Journal of Population Economics*, 34(4), pp. 1189–1217. Available at: https://doi.org/10.1007/s00148-021-00857-y.

Chen, A.Y. and Escarce, J.J. (2010) 'Family structure and childhood obesity, Early Childhood Longitudinal Study - Kindergarten Cohort', *Preventing Chronic Disease*, 7(3). Available at: http://www.cdc.gov/pcd/issues/2010/may/09_0156.htm.Accessed[date].

Chen, X., Liu, M. and Li, D. (2000) 'Parental warmth, control, and indulgence and their relations to adjustment in Chinese children: A longitudinal study.', *Journal of Family Psychology*, 14, pp. 401–419.

Chen, Y.-C., Fan, E. and Liu, J.-T. (2019) 'Understanding the mechanisms of parental divorce effects on child's higher education', *NBER Working Papers* [Preprint]. Available at: http://www.nber.org/papers/w16019.

Cherlin, A.J. *et al.* (1991) 'Longitudinal studies of effects of divorce on children in Great Britain and the United States', *Science*, 252(5011), pp. 1386–1389. Available at: https://doi.org/10.1126/science.2047851.

Cherlin, A.J., Chase-Lansdale, P.L. and McRae, C. (1998) 'Effects of parental divorce on mental health throughout the life course', *American Sociological Review*, 63(2), pp. 239–249. Available at: https://doi.org/10.2307/2657325.

Chernozhukov, V. *et al.* (2018) 'Double/debiased machine learning for treatment and structural parameters', *Econometrics Journal*, 21(1), pp. C1–C68. Available at: https://doi.org/10.1111/ectj.12097.

Christina J. Cross (2019) 'Racial/Ethnic Differences in the Association Between Family Structure and Children's Education', *Journal of Marriage and Family*, Vol. 90(November), pp. 1–22.

Cid, A. and Stokes, C.E. (2013) 'Family Structure and Children's Education Outcome: Evidence from Uruguay', *Journal of Family and Economic Issues*, 34(2), pp. 185–199. Available at: https://doi.org/10.1007/s10834-012-9326-z.

Cooper, C.E. *et al.* (2008) 'Partnership Instability and Child Wellbeing during the Transition to Elementary School', (April), pp. 1–37.

Duflo, E., Bertrand, M. and Mullainathan, S. (2004) 'How Much Should We Trust Differences-in-Differences Estimates ? Author (s): Marianne Bertrand , Esther Duflo and Sendhil Mullainathan Published by : Oxford University Press Stable URL : http://www.jstor.org/stable/25098683 HOW MUCH SHOULD WE TRUST DIFFE', *The Quarterly Journal of Economics*, 119(1), pp. 249–275.

E. Leuven and Sianesi, B. (2003) 'Psmatch2: Stata module to perform full Mahalanobis and propensity score matching, common support graphing, and covariate impalance testing'. Available at: https://www.pep-net.org/sites/pep-net.org/files/typo3doc/pdf/Training_Material/statadoc.pdf.

Ermisch, J.F. *et al.* (2001) *Family structure and children's achievements*.

Fischer, T. (2007) 'Parental divorce and children's socio-economic success: Conditional effects of parental resources prior to divorce, and gender of the child', *Sociology*, 41(3), pp. 475–495. Available at: https://doi.org/10.1177/0038038507076618.

Fomby, P. and Cherlin, A.J. (2007) ‘Family instability and child well-being’, *American Sociological Review*, 72(2), pp. 181–204. Available at: https://doi.org/10.1177/000312240707200203.

Fomby, P. and Osborne, C. (2010) ‘The influence of union instability and union quality on children’s aggressive behavior’, *Social Science Research*, 39(6), pp. 912–924. Available at: https://doi.org/10.1016/j.ssresearch.2010.02.006.

Francesconi, M., Jenkins, S.P. and Siedler, T. (2010) ‘Childhood family structure and schooling outcomes: Evidence for Germany’, *Journal of Population Economics*, 23(3), pp. 1073–1103. Available at: https://doi.org/10.1007/s00148-009-0242-y.

Geisser, S. (1975) ‘The predictive sample reuse method with applications’, *Journal of the American Statistical Association*, 70(350), pp. 320–328. Available at: https://doi.org/10.1080/01621459.1975.10479865.

Gibson, L.Y. *et al.* (2007) ‘The role of family and maternal factors in childhood obesity’, *Medical Journal of Australia*, 186(11), pp. 591–595. Available at: https://doi.org/10.5694/j.1326-5377.2007.tb01061.x.

Goisis, A., Özcan, B. and Van Kerm, P. (2019) ‘Do Children Carry the Weight of Divorce?’, *Demography*, 56(3), pp. 785–811. Available at: https://doi.org/10.1007/s13524-019-00784-4.

Golan, M. and Crow, S. (2004) ‘Targeting parents exclusively in the treatment of childhood obesity: Long-term results’, *Obesity Research*, 12(2), pp. 357–361. Available at: https://doi.org/10.1038/oby.2004.45.

Gonzalez, L. (2011) ‘The Long Term Effects of Legalizing Divorce on Children’, *SSRN Electronic Journal* [Preprint], (3789). Available at: https://doi.org/10.2139/ssrn.1374206.

Goodman, A. and Goodman, R. (2009) ‘Strengths and difficulties questionnaire as a dimensional measure of child mental health’, *Journal of the American Academy of Child and Adolescent Psychiatry*, 48(4), pp. 400–403. Available at: https://doi.org/10.1097/CHI.0b013e3181985068.

Goodman, R. (1997) ‘The strengths and difficulties questionnaire: A research note’, *Journal of child psycology and psychiatry*, 38(5), pp. 581–586. Available at: https://doi.org/10.4324/9781315258324-21.

Gruber, J. (2004) ‘Is making divorce easier bad for children? the long-run implications of unilateral divorce’, *Journal of Labor Economics*, 22(4), pp. 799–833. Available at:

https://doi.org/10.1086/423155.

Hannan, C. and Halpin, B. (2014) ‘The influence of family structure on child outcomes: Evidence for Ireland’, *Economic and Social Review*, 45(1), pp. 1–24.

Hassen, T.A. *et al.* (2021) ‘The association between birth weight and proxy-reported health-related quality of life among children aged 5 − 10 years old: A linked data analysis’, *BMC Pediatrics*, 21(1), pp. 1–8. Available at: https://doi.org/10.1186/s12887-021-02882-y.

Ho, D.E. *et al.* (2007) ‘Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference’, *Political Analysis*, 15(3), pp. 199–236. Available at: https://doi.org/10.1093/pan/mpl013.

Hveem, M.R., Faulconer, S.C.M. and Dufur, M.J. (2022) ‘Comparing Children’s Behavior Problems in Biological Married, Biological Cohabitating, and Stepmother Families in the UK’, *International Journal of Environmental Research and Public Health*, 19(24). Available at: https://doi.org/10.3390/ijerph192416543.

J.D., A. and Pischke, J.S. (2009) *Mostly Harmless Econometrics : An Empiricist ’ s Companion*.

Jackson, S.L. *et al.* (2011) ‘Impact of poor oral health on children’s school attendance and performance’, *American Journal of Public Health*, 101(10), pp. 1900–1906. Available at: https://doi.org/10.2105/AJPH.2010.200915.

King, V. (2009) ‘Stepfamily formation: Implications for adolescent ties to mothers, nonresident fathers, and stepfathers’, *Journal of Marriage and Family*, 71(4), pp. 954–968. Available at: https://doi.org/10.1111/j.1741-3737.2009.00646.x.

Kleinberg, J. *et al.* (2015) ‘Prediction Policy Problems’, *American Economics Review*, pp. 1–6.

Kravdal, Ø. and Grundy, E. (2019) ‘Children’s age at parental divorce and depression in early and mid-adulthood’, *Population Studies*, 73(1), pp. 37–56. Available at: https://doi.org/10.1080/00324728.2018.1549747.

Lisa Strohschein (2005) ‘Parental Divorce and Child Mental Health Trajectories’, 67(5), pp. 1286–1300.

Listl, S. (2011) ‘Family composition and children’s dental health behavior: Evidence from Germany’, *Journal of Public Health Dentistry*, 71(2), pp. 91–101. Available at: https://doi.org/10.1111/j.1752-7325.2010.00205.x.

McCarty, C.A. *et al.* (2005) ‘Parental Emotional Support and Subsequent Internalizing and Externalizing Problems Among Children’, *Journal of Developmental & Behavioral Pediatrics*, 26(4), pp. 267–275.

McNamara, E., Murray, A. and O’Mahony, D. (2019) ‘Growing up in Ireland. Design, Instrumentation and Procedures for Cohort ’08 at 9 Years Old (Wave 5)’, *Longitudinal and Life Course Studies* [Preprint]. Available at: https://doi.org/10.1332/175795919X15699445254797.

Meesters, C. and Muris, P. (2004) ‘Perceived parental rearing behaviours and coping in young adolescents’, *Personality and Individual Differences*, 37(3), pp. 513–522. Available at: https://doi.org/10.1016/j.paid.2003.09.022.

Meyer, B.D. (1995) ‘Natural and quasi-experiments in economics’, *Journal of Business and Economic Statistics*, 13(2), pp. 151–161. Available at: https://doi.org/10.1080/07350015.1995.10524589.

Osborne, C. and Mclanahan, S. (2007) ‘Partnership Instability and Child Well-Being’, *Journal of Marriage and Family*, 69(4), pp. 1065–1083.

De Paola, M. and Gioia, F. (2017) ‘Does patience matter in marriage stability? Some evidence from Italy’, *Review of Economics of the Household*, 15(2), pp. 549–577. Available at: https://doi.org/10.1007/s11150-014-9275-4.

Parcel, T.L., Campbell, L.A. and Zhong, W. (2012) ‘Children’s Behavior Problems in the United States and Great Britain’, *Journal of Health and Social Behavior*, 53(2), pp. 165–182. Available at: https://doi.org/10.1177/0022146512436742.

Paul R. Rosenbaum and Donald B. Rubin (1983) ‘The Central Role of the Propensity Score in Observational Studies for Causal Effects Author (s): Paul R . Rosenbaum and Donald B . Rubin Published by : Oxford University Press on behalf of Biometrika Trust Stable URL : http://www.jstor.’, 70(1), pp. 41–55.

Pianta, R.C. (1992) ‘Child-parent relationship scale.’, *Unpublished measure, University of Virginia* [Preprint], (427).

Powers, D.A. (2005) ‘Effects of family structure on the risk of first premarital birth in the presence of correlated unmeasured family effects’, *Social Science Research*, 34(3), pp. 511–537. Available at: https://doi.org/10.1016/j.ssresearch.2004.04.007.

Rasmussen, A.W. (2009) *Family Structure Changes and Children ′ s Health , Behavior , and*

Educational Outcomes Department of Economics Family Structure Changes and Children ' s Health , Behavior , and Educational Outcomes.

Resce, G. (2022) 'The impact of political and non-political officials on the financial management of local governments', *Journal of Policy Modeling*, 44(5), pp. 943–962. Available at: https://doi.org/10.1016/j.jpolmod.2022.08.004.

Resce, G. and Vaquero-Piñeiro, C. (2022) 'Predicting agri-food quality across space: A Machine Learning model for the acknowledgment of Geographical Indications', *Food Policy*, 112(September), p. 102345. Available at: https://doi.org/10.1016/j.foodpol.2022.102345.

Resce, G. and Vaquero-Piñeiro, C. (2023) 'Taste of home: Birth town bias in Geographical Indications', *Available at SSRN 4437094* [Preprint], (089). Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4437094%0Ahttp://web.unimol.it/progetti/repec/mol/ecsdps/ESDP23089.pdf.

Rubin, D.B. (2001) 'Using propensity scores to help design observational studies: Application to the tobacco litigation', *Health Services & Outcomes Research Methodology*, (2), pp. 169–188. Available at: https://doi.org/10.1017/CBO9780511810725.030.

Rubin, D.B. (2006) *Matched sampling for causal inference*. Edited by U.C.U.P. Cambridge.

Sandra L . Hofferth (2006) 'Residential Father Family Type and Child Well-Being : Investment versus Selection', *Springer on behalf of the Population Association of America*, 43(1), pp. 53–77.

Sanz-De-Galdeano, A. and Vuri, D. (2007) 'Parental divorce and students' performance: Evidence from longitudinal data', *Oxford Bulletin of Economics and Statistics*, 69(3), pp. 321–338. Available at: https://doi.org/10.1111/j.1468-0084.2006.00199.x.

Schmeer, K.K. (2012) 'Family structure and obesity in early childhood', *Social Science Research*, 41(4), pp. 820–832. Available at: https://doi.org/10.1016/j.ssresearch.2012.01.007.

Sharon H. Bzostek, L.M.B. (2017) 'Family Structure Experiences and Child Socioemotional Development During the First Nine Years of Life: Examining Heterogeneity by Family Structure at Birth', *Demography*, 54(2), pp. 513–540. Available at: https://doi.org/https://doi.org/10.1007/s13524-017-0563-5.

Sheiham, A. (2006) 'Dental caries affects body weight, growth and quality of life in pre-school children', *British Dental Journal*, 201(10), pp. 625–626. Available at: https://doi.org/10.1038/sj.bdj.4814259.

Sigle-Rushton, W., Hobcraft, J. and Kiernan, K. (2005) 'Parental divorce and subsequent disadvantage: A cross-cohort comparison', *Demography*, 42(3), pp. 427–446. Available at: https://doi.org/10.1353/dem.2005.0026.

Smetana, J.G. (2017) 'Current research on parenting styles, dimensions, and beliefs', *Current Opinion in Psychology*, 15, pp. 19–25. Available at: https://doi.org/10.1016/j.copsyc.2017.02.012.

Sohail, S. *et al.* (2018) 'Implementation of machine learning algorithm on factors effecting divorce rate', *2018 International Conference on Engineering and Emerging Technologies, ICEET 2018*, 2018-Janua, pp. 1–5. Available at: https://doi.org/10.1109/ICEET1.2018.8338618.

StataCorp LLC (2023) *Stata 2018 Lasso Reference Manual*.

Steinberg, L., Fletcher, A. and Darling, N. (1994) 'Parental monitoring and peer influences on adolescent substance use', *Pediatrics*, 93(6 II), pp. 1060–1064. Available at: https://doi.org/10.1542/peds.93.6.1060.

Stone, L.L. *et al.* (2010) 'Psychometric properties of the parent and teacher versions of the strengths and difficulties questionnaire for 4- to 12-Year-olds: A review', *Clinical Child and Family Psychology Review*, 13(3), pp. 254–274. Available at: https://doi.org/10.1007/s10567-010-0071-2.

StØrksen, I. *et al.* (2006) 'Adolescent adjustment and well-being: Effects of parental divorce and distress', *Scandinavian Journal of Psychology*, 47(1), pp. 75–84. Available at: https://doi.org/10.1111/j.1467-9450.2006.00494.x.

Sun, Y. (2001) 'Family environment and adolescents' well-being before and after parents' marital disruption: A longitudinal analysis', *Journal of Marriage and Family*, 63(3), pp. 697–713. Available at: https://doi.org/10.1111/j.1741-3737.2001.00697.x.

Sun, Y. and Li, Y. (2011) 'Effects of family structure type and stability on children's academic performance trajectories', *Journal of Marriage and Family*, 73(3), pp. 541–556. Available at: https://doi.org/10.1111/j.1741-3737.2011.00825.x.

Tendulkar, S.A.; Buka, S.; Dunn, E.C.; Subramanian, S.V.; Koenen, K.C. (2010) 'A multilevel investigation of neighborhood effects on parental warmth.', *J. Community Psychol.*, 38, pp. 557–573.

Thornton, M. *et al.* (2013) 'Growing up in Ireland. Design, Instrumentation and Procedures for the Infant Cohort at wave 1 (9 Months)', *Longitudinal and Life Course Studies* [Preprint]. Available at: https://doi.org/10.1332/175795919X15699445254797.

Tibshirani (1996) ‘Regression Shrinkage and Selection via the Lasso’, *Journal of the Royal Statistical Society. Series B (Methodological)*, 58(1), pp. 267–288. Available at: jstor.org/stable/2346178.

Varian, H.R. (2014) ‘Big data: New tricks for econometrics’, *Journal of Economic Perspectives*, 28(2), pp. 3–28. Available at: https://doi.org/10.1257/jep.28.2.3.

Varian, H.R. (2016) ‘Causal inference in economics and marketing’, *Proceedings of the National Academy of Sciences of the United States of America*, 113(27), pp. 7310–7315. Available at: https://doi.org/10.1073/pnas.1510479113.

Waldfogel, J., Craigie, T.A. and Brooks-Gunn, J. (2010) ‘Fragile families and child wellbeing’, *Future of Children*, 20(2), pp. 87–112. Available at: https://doi.org/10.1353/foc.2010.0002.

Williams, J. *et al.* (2015) ‘Growing Up in Ireland. Design, Instrumentation and Procedures for Cohort ’08 at wave 3 (5 Years)’. Available at: www.dcya.gov.ie.

Wing, C., Simon, K. and Bello-Gomez, R.A. (2018) ‘Designing Difference in Difference Studies: Best Practices for Public Health Policy Research’, *Annual Review of Public Health*, 39, pp. 453–469. Available at: https://doi.org/10.1146/annurev-publhealth-040617-013507.

Yang, Z., Zhang, R. and Liu, L. (2022) ‘The Effect of Parent-Child Relationship on Child Development’, *Academic Journal of Humanities & Social Sciences*, 5(15), pp. 118–122. Available at: https://doi.org/10.25236/ajhss.2022.051515.

Yelick, A. (2017) ‘The Effects of Family Structure on Consumption and Exercise Patterns for Adolescent Youth’, *Child and Adolescent Social Work Journal*, 34(4), pp. 381–395. Available at: https://doi.org/10.1007/s10560-016-0468-y.

Appendix

Table A1. Logit model estimates Predicting Family Structure Change: Unmatched Sample (Comparison: Two-parent Families)

	(1)
Pcg's Characteristics	
Age	-0.593**
	(0.244)
Age2	8.451**
	(3.518)
Irish ethnicity	0.758
	(0.467)
Has 2 children	-0.245
	(0.355)
Has 3 children	-1.010**
	(0.398)
Has 4 children	-1.040**
	(0.491)
Has 5 children	-0.675
	(0.558)
Roman Catholic	0.141
	(0.468)
BMI	0.0210
	(0.0252)
Study Child's Characteristics	
Boy	0.211
	(0.213)
Birthweight	-0.229
	(0.220)
Been breastfed	-0.348
	(0.250)
Twin	0.824*
	(0.440)
Secondary Caregiver's Characteristics	
Age	0.154
	(0.168)
Age2	-2.459
	(2.173)
Irish ethnicity	0.223
	(0.373)
BMI	0.0158
	(0.0122)
Household's Characteristics	
Age gap between Primary and Secondary caregiver	0.0615
	(0.0463)
Education gap between Pcg and Scg	0.0881
	(0.0658)

Table A1. Logit model estimates Predicting Family Structure Change: Unmatched Sample (Comparison: Two-parent Families) (Contd.)

Household's Characteristics:	
Pcg and Scg have the same economic status	-0.287
	(0.221)
Pcg and Scg are of the same ethnicity	-0.526
	(0.364)
Rural area	-0.275
	(0.233)
Constant	3.112
	(3.720)
Observations	6610

Notes: The Table reports Logit estimates. Sample: GUI dataset - Infant Cohort. All the information refers to wave 2, when children are aged 3 years old. The dependent variable is *Family Structure Change* which takes value of 1 if the family will experience a family structure change in the next wave, and 0 otherwise. Standard errors, reported in parentheses, are corrected for heteroskedasticity, and allow for clustering at the household level. The symbols ***, **, * indicate that the coefficients are statistically significant at the 1, 5, and 10 percent level respectively.

Conclusion

This thesis studies the impact of parental divorce on a bunch of behavioral, educational, and health-related children's outcomes using data from the Growing in Up in Ireland (GUI), the National Longitudinal Study of Children and Young people in Ireland. The study follows two cohorts of children: Child Cohort, a cohort of 8,568 nine-year olds children born in 1998, and Infant Cohort, 11,134 aged nine-month-olds infants born in 2008.

In our opinion, the samples used in our study are highly representative of the population since, when the two cohorts are considered together, the age span of children included ranges from 9 months to 20 years.

As main contribution, our results shed light on the importance of the age of the child at the time of parental divorce or the change in family structure in the child's adjustment process. Surprisingly, among the same cohort of children – Children Cohort -, we find a great heterogeneity in the effects for all the outcomes considered, according to the age at which the event took place.

Among the cohort of children who experience parental divorce after the age of 9, our results show how different outcomes are variously affected at different child's age. Specifically, results from the first chapter show that socioemotional problems of the child and their physical and dental health are negatively affected by parental divorce if the event takes between the ages of 13-17. Parental divorce adversely impacts on child's later educational achievements if experienced between the ages of 9 and 13, as shown in the second chapter.

Moving our attention to the cohort of younger children who experience a change in their family structure between the ages of 3 and 5, our results provide evidence of negative impact on both the behavioral traits of the child and the quality of the relationship between the child and the parent. The event exerts negative effects both in a short period of time, detected when the child reaches the age of 5, and over a longer period of time, observed when the child is 9 years old.

In each chapter we tackle a particular issue through the use of more sophisticated or recently developed methodologies that overcome the OLS under certain conditions.

In the first chapter, we address the potential endogeneity of parental divorce using household fixed-effects model in conjunction with the DiD strategy. In the second one, we account for the potential omitted bias and the use of cross-sectional data, applying a double-machine learning approach. Finally, in the third one, we deal with the potential selection bias implementing a kernel propensity score matching DiD.