

Assignment cover sheet

COVER SHEET

Department of Social Sciences

Masters assignment

Candidate no.	KKRJ3		
Course	MSc. Social Policy and Social Research	Full time/ Part time/ Module only	Full time
Module Title	Dissertation	Module Code	SOCS0011
Assignment Title	Is father absence during early childhood associated with children's future outcomes? A study using the Peruvian Young Lives Study sample		
Date of submission	17 September 2021		

I have uploaded one copy of my work on Moodle on the deadline date

☐

I have included my word count and my work falls within the required word length

☒

This is my first submission for this module OR

☒

This is my resubmission for this module

☐

I have read the guidelines on plagiarism and this is entirely my own work.

☒

Is father absence during early childhood associated with children's future outcomes? A study using the Peruvian Young Lives Study sample

University College London

SOCS0011 - MSc Social Policy and Social Research Dissertation

Student Number: KKRJ3

October 2020

Word Count: 11,812

This dissertation may be made available to the general public for borrowing, photocopying or consultation without the prior consent of the author.

ABSTRACT

Academic research during the past decades has agreed that human skills formation is a process that starts very early in life. Moreover, experiences lived during early childhood can make a difference over future life trajectories. Therefore, parents have a fundamental role in skills formation as those in charge making the necessary investments so that their children can develop adequately. However, despite their importance, a relatively high share of children is born and raised in household without the presence of their fathers. This work studies the associations between early father absence and future children's cognitive and non-cognitive outcomes. We use the Peruvian sample of the Young Lives Survey and compare the trajectories of children who experienced and did not this event during their first five years of life making use of a pre-processing weighting strategy called balancing entropy. This technique allows to minimise observable differences between both groups that might be correlated with early father absence and the outcomes of interest, reducing the selection bias. Our results suggest negative long-term associations between our definition of absence and some cognitive (math and reading skills) and non-cognitive (self-esteem) outcomes. We also explored mechanisms that might explain such associations and found suggestive evidence that mothers who went through this separation may have had more difficulties for undertaking monetary and non-monetary investments in child skills. We also explored between-sex differences and observed potentially stronger associations for boys in some cognitive outcomes and girls in some non-cognitive outcomes. After discussing these results, we proposed a series of policy recommendations, mostly aiming at supporting single mothers throughout the challenging task of raising children on their own. Finally, we detail some limitations of our work and suggest ways in which future studies can enhance research over this topic.

Table of contents

1. Introduction	6
2. Literature review	9
2.1. Theoretical frameworks: early childhood, skills formation and the impact of father absence	9
2.2. Empirical findings: father absence and child skills	11
3. Methodology	15
3.1. Study sample	15
3.2. Variables	17
3.3. Statistical approach	22
4. Descriptive evidence	27
4.1. Comparing children who experienced and did not early father absence	27
5. Results	30
5.1. Main results	30
5.2. Mechanisms	32
5.3. Sensitivity analysis	42
6. Discussion and conclusion	43
References	47
Appendix A	59
Appendix B	60
Appendix C	61
Appendix D	70
Appendix E	71
Appendix F	72

Tables and figures

Table 1. Evolution of the YL young cohort sample (Peru)	16
Table 2. Sample loss in the study	17
Table 3. Outcomes of interest by round	20
Table 4. Comparing children who experience and did not early father absence	30
Table 5. ATT of early father absence on cognitive outcomes (complete sample)	36
Table 6. ATT of early father absence on non-cognitive outcomes (complete sample)	36

Table 7. ATT of early father absence on cognitive outcomes (differences by sex) ..	38
Table 8. ATT of early father absence on non-cognitive outcomes (differences by sex)	39
Table 9. Potential mechanisms to explain the effects of early father absence	40
Table 10. ATT of father absence on children's time use (differences by sex)	41
Table B1. Non-cognitive outcomes, instruments, and items	60
Table C1. Propensity score calculation	67
Table D1. Cognitive outcomes	70
Table D2. Non-cognitive outcomes	70
Table E1. Sample of interest: comparing complete and missing observations	71
Figure 1. Propensity score distribution by group	27
Figure 2. Performance of the matching and weighting alternatives (summary)	29
Figure 3. Comparison of control group weights (EB and IPW)	42
Figure A1. Distribution of the ratio of single-parent and biparental households by regions	59
Figure C1. Comparison of matching and weighting alternatives.....	69

1. Introduction

Birth works like a lottery: some children are “winners” and others are not depending on the household they are born into, which is entirely out of their control (Heckman, 2011).¹ For instance, approximately 320 million children lived in single-parent households in 2016 (Chamie, 2016). Considering the crucial role of parents in children’s skills formation (Cunha & Heckman, 2007), these children might have a higher propensity of experiencing more obstacles to develop skills that are necessary to thrive as adults. We could expect, for instance, that single parenthood may generate a series of complications in areas such as income and caregiver’s time availability and psychological well-being, all aspects that affect how much a parent can invest in their children’s skills. Thus, the absence of a parent can produce negative externalities to their children. Children born and raised during their *early childhood* under this condition can be especially at risk. UNICEF defines early childhood as the period starting in conception and ending when a child begins school (UNICEF, 2018). This moment has been considered a fundamental period in human skills development since, during this time, the foundations of brain architecture are being established (National Scientific Council on the Developing Child, 2007).

This work studies whether father absence during the first years of children’s lives influences their cognitive and non-cognitive skills development. Based on the frequency of father-child interactions and the importance of early childhood, we defined as “absent fathers” those who did not tend to see their children every day during the first five years of their lives. Besides, we categorised as “present fathers” those who did see them daily during this time. Population statistics support our choice of focusing only on fathers: single mothers represent, on average, 82% of single-parent households in Europe (OECD, 2016), 92% of single-parent households with children younger than five years old in OECD countries (OECD, 2018). Furthermore, we focus on Peru because Latin America has one of the highest presences of single-parent households according to data from the United Nations Population Division (see [Figure A1](#) in Appendix A) and the Gallup World Poll (Crabtree & Kluch, 2020).

¹ The view has been also commonly referred to as the “accident of birth”.

Specifically, the share of mother-leading households in Peru was around 83% of single-parent households in 2017 (Peruvian National Institute of Statistics, 2019).

This research uses data from the Young Lives Study (YLS). This cohort study has followed two samples (a "younger" and an "older" cohort) of children in Peru, Ethiopia, India, and Vietnam during the last 19 years. It had collected so far information in five rounds between 2002 (when children from the Peruvian younger cohort mainly were between 6 and 17 months old) and 2016 (when they were between 14 and 16 years old). Each wave allows observing cognitive (vocabulary, reading, and math) and non-cognitive outcomes (agency, self-efficacy, and self-esteem) and represents different stages of children's lives, beginning in early childhood and ending in adolescence. YLS also permits to study father absence during very early years.

We compared outcomes at several posterior life stages using non-experimental techniques to estimate more trustworthy associations between early father absence and children's outcomes. After testing several approaches, we selected entropy balancing (Hainmueller, 2012) and kernel propensity score matching to minimise the selection bias on a series of observable characteristics. Our aggregate results suggested few and (relatively) reduced adverse associations between early father absence and some outcomes during late childhood and adolescence. However, these general findings showed some degree of heterogeneity. For example, father absence had somewhat larger and more detrimental associations in mathematics for boys. Meanwhile, larger negative relationships with non-cognitive outcomes were observed in girls, mainly in self-esteem during adolescence.

Finally, we explore some potential mechanisms that might correlate to child skills investments following an early father absence. For instance, such an event was negatively associated with household income. Also, mothers affected by this situation seemed to have had steadily less time available to dedicate to their children, were less likely to have cohabitated with other adults who may have supported them, and had consistently lower levels of agency. Even though we should not make strictly causal interpretations of our results, these mechanisms may still be relevant to understand better the dynamics caused by early father absence.

This work contributes to the literature in several ways. Firstly, few studies show how parent absence affects children's trajectories by observing outcomes at several stages of childhood and adolescence.² Furthermore, it raises attention to the importance of going further from sample estimates since father absence can have differential impacts based on characteristics such as children's sex. Also, its focus on early father absence has been scarcely explored in previous research. Studying absence during this period is relevant considering the importance of timing on child skills investment (Cunha et al., 2006) and the particular neurodevelopmental significance of the first years of life. Fourthly, it illustrates different ways that can help to overcome some of the methodological challenges that arise when trying to make causal claims regarding parent absence. This event does not occur randomly, single-parent and biparental households differ systematically in observable and unobservable characteristics correlated with children outcomes (Sigle-Rushton & McLanahan, 2004). Thus, efforts that rely solely on ordinary least squares (OLS) regressions may produce biased results even after controlling for many covariates (Ribar, 2004). Even though the approaches used in this work still rely on strong assumptions, they may be more effective for diminishing selection bias than OLS regressions. Finally, it adds more evidence of the potential consequences of father absence in non-developed countries. The dynamics caused by this event deserve more attention in these settings due to their weaker government capabilities, lower access and quality of public services, and lower household income levels.

The paper contains six sections, including this introduction. Section 2 presents a literature review about the importance of early childhood and parents for human skills formation, together with a summary of empirical studies regarding father absence. Then, Section 3 details the sample, outcomes, and variables used in the analysis and the statistical approaches implemented to reduce the selection bias. Section 4 presents descriptive evidence about the sample. Section 5 presents the results, starting with an overview of the general effects, followed by a focus on heterogeneous effects and an exploration of some potential mechanisms that suggest causal pathways whereby early father absence might affect children. Afterwards, Section 6 concludes with a discussion over the results, policy implications, and limitations.

² See, for instance, Himaz (2020); Nguyen, (2016); Vera Rueda (2021)

2. Literature review

2.1. Theoretical frameworks: early childhood, skills formation and the impact of father absence

Cunha and Heckman (2007) present a framework to understand skills formation. They define skills as cumulative outputs generated by the continuous interaction of genes, environmental factors, and investments throughout life. Their model suggests that the process of skills generation has *sensitive* and *critical* periods during childhood, which aligns with several studies and discussions on child development.³ The first term implies moments wherein some investments produce more efficiently specific skills, while the second refers to the only time when some skills can be developed through investments.⁴ Sensitive periods were empirically observed in Cunha and Heckman (2008) and took place at earlier ages for cognitive than for non-cognitive outcomes.

Additionally, they indicate that the stock of skills in one period influences skills production in the next period. This effect occurs in two ways: *self-productivity* (higher current stocks lead to higher future stocks) and *dynamic complementarity* (higher current stocks lead to higher productivity of future investments).⁵ These four elements highlight the importance of investing in skills during early childhood and continuing investing throughout all stages of childhood and adolescence. The relevance of early investments was also observed in Cunha et al. (2010). This work found that it is more difficult to remediate lower initial levels of cognitive outcomes with investments at later childhood stages than to address them through investments that start earlier. In contrast, the results of investing earlier or later were more similar when building non-cognitive skills.

On their side, the medical literature argues that some fundamental biological changes occur during early childhood that allow children to build complex capabilities later in life (Britto & Pérez-Escamilla, 2013). For instance, developments in the prefrontal cortex during infancy are necessary for achieving subsequently satisfactory levels of

³ See, for instance, Knudsen (2004); Knudsen et al. (2006).

⁴ The authors also expanded this strict definition of critical periods to refer to periods wherein, although technically possible, some skills cannot be built below a certain economic cost of investments.

⁵ Empirical evidence of both features was also found in Cunha and Heckman (2008).

executive functioning (Hodel, 2018).⁶ Generally, early childhood is considered a fundamental moment in terms of brain architecture since, during this time, millions of brain connections are created and are more malleable to changes (National Scientific Council on the Developing Child, 2007). Therefore, early experiences can have powerful consequences in brain development (Bick & Nelson, 2016; Fox et al., 2010). Furthermore, some authors emphasise the particular importance within early childhood development of the “first 1,000 days” (a period that goes from conception to the second year of life) due to the higher levels of plasticity that the brain experiences at this moment.⁷ Nevertheless, others indicate that, although these three years are important, it is misleading to claim that the only “window of opportunity” for skills formation occurs during this period. This is because some areas of the brain in charge of several high functions are not yet completely developed at this point (National Scientific Council on the Developing Child, 2007). For instance, brain connections related to high cognitive functions of the prefrontal cortex keep maturing until adolescence (National Research Council, 2000).

The role of parents throughout the process of skills formation is vital in Cunha and Heckman (2007) 's model because they are in charge of making decisions that affect children and investing in them. Even though they might be driven by altruism, parents are also subject to several constraints and are limited by their skills. Therefore, they might invest below the optimal levels that would allow their children to reach their full potential, which the authors consider a market failure. Even though their model primarily focuses on monetary investments, others have also discussed parents' influence through non-financial investments. A traditional approach has focused on the time they spend with their children (Lamb et al., 1985). However, more recent works recommend concentrating on the *quality* over the quantity of time (Pleck, 2010).

Following the logic of this model, we can find several mechanisms whereby father absence can affect child skills formation. Fathers moving out of the household may be forced to increase their expenditure on housing, increasing their budget constraints for

⁶ Executive functioning are the high-cognitive abilities needed to regulate emotions and adapt behaviours when pursuing goal-oriented actions, mainly working memory, inhibitory control, cognitive flexibility, strategy application, and multitasking (Gilbert & Burgess, 2008). These skills have been widely associated with higher school performance, better mental health and social competence, less adoption of risky behaviours, among other outcomes (Robson et al., 2020).

⁷ See, for example, Georgieff et al. (2015); Schwarzenberg et al. (2018); Thompson and Nelson (2001)

investing on their children or keeping supporting their children's household. Besides, their access to spending time with their children decreases as they become excluded from their previous household's daily dynamics, especially after work hours. These changes can discourage them from being involved in childcaring.

Additionally, their absence can also have an effect through its influence on mothers, who usually end with a higher share of childcare responsibility while also facing the negative economic consequences of this event on household income (de Vaus et al., 2017). In such cases, they can appeal to coping strategies that might offset, at least partially, some of the adverse effects of father absence but may also lead to others (Mortelmans, 2020). For example, they might enter the labour market or increase their work hours, which reduces the time available for spending with their children. Furthermore, economically vulnerable mothers usually lean on informal sources of childcare while working, which sometimes might not be an adequate substitution for their dedication, especially during early childhood (Bernal & Keane, 2011; Herbst, 2013; Zhang et al., 2021). Besides, they can decide to repartner, leading to periods of instability in terms of household arrangements. Coresidential and dating transitions in single-mother households have been associated with harmful parenting practices (Beck et al., 2010) and child outcomes (Cooper et al., 2011). Additionally, as Mooney et al. (2009) point out, father absence can influence mothers' mental health directly whenever it causes parental conflicts; or indirectly through the pressures produced by economic hardship. These authors indicate that mothers' mental health is crucial since it affects their approach to parenting and childcaring.

2.2. Empirical findings: father absence and child skills

Father absence has been studied in several ways. Several studies have concentrated on family structure shocks, such as divorce and cohabitation dissolution.⁸ The consequences of these events depend, theoretically, on the extent to which fathers decide to be involved in their children's development. Some studies focused primarily on the United States (US) have found that separations are associated with harmful effects mainly in non-cognitive dimensions such as self-esteem, internalising and

⁸ Many do not differentiate between mothers' and fathers' absence. However, since fathers have a much higher proclivity of separating than mothers (as the statistics presented in the introduction have shown), the results of those works might be mainly driven by their absence.

externalising problem behaviours, and self-control.⁹ However, others outside the US have not found such results in other settings. For example, Reynolds et al. (2018) found no associations in cognitive and non-cognitive outcomes in Chile even though they observed a negative income effect related to father separation. This led them to propose that income might not be a channel as strong as, for instance, time dedication, which they assumed did not change much considering how employment arrangements in Chile leave fathers with short time to spend with their children.¹⁰ Bucheli and Vigorito (2021) found no differences in children's socio-emotional well-being in Uruguay. Nevertheless, they did observe negative associations in educational outcomes mainly during adolescence and concentrated on boys.¹¹ Another relevant aspect of this study was their analysis of mechanisms. As expected, separation was associated with a negative income effect and higher and more intense participation of mothers in the labour market, which might have reduced their time availability for spending with their children.¹² They also tested whether father dedication (in terms of monetary contribution and interaction with children) could be another mediator and found suggestive evidence although with a lack of robustness.

Another studied approximation to father absence is work migration, which might positively affect income via remittances and negatively impact time dedication. Some studies have found positive associations with school outcomes.¹³ There is also evidence of impact heterogeneity based on children's sex, with girls having more benefits (Antman, 2012). The effects of migration might also be context-dependent, as most of the revised studies of the Chinese setting have not shared these positive effects.¹⁴ Regarding mechanisms, positive results have been associated with elements consistent with the income effect, such as increases in access to private tuition (Sarma & Parinduri, 2016) and a rise in expenditures related to children's

⁹ See Amato & Anthony (2014); Kim (2011); Lee & McLanahan, (2015); Mitchell et al., (2015). Among few studies that have not observed negative differences, see Foster and Kalil (2007).

¹⁰ However, they were unable to test the hypothesis of the higher importance of time dedication and the supposition that those times remained unaltered after the separation.

¹¹ They also tested heterogeneity based on the timing of the separation but found no differences.

¹² Other studies that have observed negative income effects after separation are Aassve et al. (2007), Jarvis & Jenkins (1999), and Manting & Bouma (2006).

¹³ See Antman, (2012); Bai et al. (2018); Intemann and Katz (2014); Jamil (2017); Sarma and Parinduri (2016); Vikram (2021)

¹⁴ See Lu (2012); Li et al. (2017); Meng and Yamauchi (2017); Xu et al. (2019); Zhang et al. (2014). Only Bai et al. (2018) showed positive effects.

education (Vikram, 2021). On the other side, negative effects have been associated with less father involvement or supervision, such as a reduction of the time children dedicate to study at home (Meng & Yamauchi, 2017). Nonetheless, other studies have not found differences in children's time use (Nguyen & Vu, 2016).

Finally, other works have used the frequency of interactions between fathers and their children without differentiating the reasons for their absence, which is the approach that we use in our study. Using the YLS Peruvian sample, Dearden et al. (2013) classified fathers as "present" if they saw their children daily or weekly and analysed nutritional outcomes at age 5. As expected, children who saw their fathers regularly during infancy (1 year of age) and childhood (5 years of age) had better outcomes, consistent with a negative income effect of separation. Our work differs from this study in some ways. Firstly, it does not look at nutritional outcomes but focuses instead on cognitive skills and socio-emotional abilities. Also, we study changes in several moments of children's lives while Dearden et al. (2013) focused on one period. Thirdly, we use a different approach to reduce selection bias. While Dearden et al. (2013) controlled by a list of covariates that may correlate with both the outcomes and father absence, we use techniques that may yield less biased estimators.

Vera Rueda (2021) used the samples from all YLS countries to calculate the impact of father absence. She used a more similar measure to ours: not seeing the children every day or not living with them during the first two rounds of the YLS survey, which corresponds to most of the early childhood period (ages 1-5). The author used the younger cohorts as the counterfactuals of their even younger siblings born between the first and second surveys. Simultaneously, she also compared outcomes between households who did and did not experience father absence to discard that any effect is due to contextual factors such as macroeconomic shocks. This study found harmful effects on nutritional outcomes, mixed socioemotional results, and, perhaps surprisingly, positive cognitive measures. It also explored mechanisms and found suggestive evidence regarding a negative income effect and supporting family structures that may have substituted fathers' time dedication. However, the author found much between-country heterogeneity, making it difficult to establish a general conclusion.

Among several differences between Vera Rueda (2021) and our work, it is worth highlighting that both use somewhat different subsets of the YLS Peruvian sample. For instance, the identification strategy of the first work led to the exclusion of those children who had no siblings born during the first and second rounds of the survey, which in the case of Peru meant 75% of the YLS young cohort. Meanwhile, our strategy does not require such an exclusion. Additionally, we are only interested in two groups of children: those who were not exposed to father absence throughout the two survey rounds and those exposed in both cases. Another differential element are the differences in the identification strategies. Hence, the results of both studies should be seen as complementary in the discussion of the potential consequences of early father absence.

To sum up, the associations of father absence and children's outcomes depend on how this phenomenon affects the investments necessary for their skills formation, which will depend on its influence on both parents. However, it is essential to acknowledge that father absence has different causes, leading to different effects. For example, work migration might have positive results as it usually combines household income gains with losses of time dedication from fathers. Meanwhile, separation can entail both economic and time dedication losses. Additionally, a third category, not mentioned in this review, is father death, which can lead to high reductions in both aspects and a more significant psychological impact on both children and mothers because it usually has a higher degree of unpredictability than separation (Amato & Anthony, 2014).¹⁵ Thus, the effects of father absence observed in studies that do not focus on a specific cause are a combination of the impacts of these three cases weighted by their frequency. Additionally, it is also relevant to consider that studies showed high levels of heterogeneity based on the studied settings.

¹⁵ This category was not included in this literature review due to its low frequency in the Peruvian YLS sample: only 13 out of 2,052 children from the young cohort (0.64% of the sample) had experienced father death by the moment of the first round of the YL survey. This number barely increased to 27 by the time of the third round. In any case, very few studies have analysed the effects of father death on children's cognitive and non-cognitive outcomes: see, for instance, Himaz (2013, 2020).

3. Methodology

3.1. Study sample

This study uses data from the Peruvian sample of the YLS, a longitudinal cohort study that has followed two groups of children (a "younger" and "older" cohort) in Ethiopia, India, Peru, and Vietnam to analyse poverty and its consequences on children's developmental trajectories. The samples in all four countries were derived through a multi-stage procedure that ended with selecting twenty clusters ("sentinel sites") in each country and a random selection of households within each one. Additionally, they over-represented poor regions and were not intended to be representative of each country. In fact, the sentinel sites in Ethiopia, India, and Vietnam were selected purposefully. However, Peru's strategy differed. First, it followed a multi-stage random-sampling approach using the country's by then 1,818 districts (the minor political-administrative areas) as the sampling frame and information about poverty to discard the five per cent most affluent districts. After the sentinel sites were randomly chosen, it used data from the national census to select the blocks where households with children in the predefined age ranges would be invited to participate.¹⁶

YLS has gathered five rounds of information between 2002 and 2016, with timing that considered different schooling and neurocognitive development phases.¹⁷ Additionally, the study was supposed to collect the sixth round in 2020, but it was not possible due to the Covid-19 pandemic. Instead, they performed several rounds of phone surveys to gather information related to the pandemic. We do not use them because their content is entirely out of this research's scope. [Table 1](#) summarises information about the rounds used in our study. The YLS has successfully retained the great majority of children from the young cohort: less than 10% of the sample was no longer part of the study after 13 years and five survey rounds, including all cases of deceased children (25 children as of round 5, meaning 1.23% of the sample and 13.02% of the total attrition).

¹⁶ More detailed information about the sampling strategy can be found in Escobal et al. (2003).

¹⁷ According to Dawes (2020), some milestones that were considered when selecting the timing of the rounds were: end of preschool, which is also the period when early development skills are being formed (Round 2); the end of primary school and middle childhood, which is also the moment when concrete operational thinking is well established and puberty begins (Round 3); middle adolescence (Round 4); and late adolescence (Round 5).

Table 1. Evolution of the YL young cohort sample (Peru)

Round	Ages	Responses	Attrition	
			Num.	% of the original sample
1	6-18 months	2,052	-	-
2	4-6 years	1,963	89	4.34
3	7-8 years	1,943	119	5.31
4	11-12 years	1,902	150	7.31
5	14-15 years	1,860	192	9.36

Source: own elaboration

For several reasons, the sample used in our study (from now on, the "effective sample") is a subset of the whole YLS Peruvian young cohort. In the first place, not all children were part of our groups of interest. We will elaborate more on this in the variables section. Secondly, we also discarded those children who attrited at some point because we would not have been able to observe their complete trajectories. Thirdly, we also dropped all who had missing values in at least one of the relevant variables in this study. These cases included those without information in at least one of the covariates used to calculate the propensity score (see subsection 3.3) and those with missing values in any outcome throughout the rounds. [Table 2](#) summarises the steps taken when selecting our effective sample size, which ended up being 1,360 children (66% of the YLS younger cohort sample and 86% of the sample of interest). Finally, the data in our study is anonymised and publicly available in the UK Data Service under the following authors: Boyden (2018a, 2018b); Jones and Huttly (2018); Sanchez et al. (2018); and Woldehanna et al. (2018).

Table 2. Sample loss in the study

Ages	Size	%
YLS young cohort	2,052	100
Out of sample of interest	435	21.20
Sample of interest	1,617	78.80
Attrition	101	6.25
Incomplete information for propensity score estimation	13	0.80
Incomplete information of outcomes	143	8.84
Sample lost	257	15.89
Effective sample	1,360	84.11
Early father absence	196	14.41
Early father presence	1,164	85.59

Source: own elaboration

3.2. Variables

Cognitive skills:

The YLS measured several cognitive skills beginning in round 2 for the case of the young cohort. The following paragraphs describe the measures used in this study, all listed by round [Table 3](#):

Receptive vocabulary (RV) refers to "the words that [a person] can understand rather than the words that he or she normally uses" (American Psychological Association). The primary importance of this outcome, especially in early childhood, is its correlation with future reading capabilities and that it is a skill that children start to develop before the first year of life (Burger & Chong, 2011).¹⁸ Some studies have concluded that RV influences reading skills through its effect on early higher-level discourse skills such as inference and integration (Oakhill & Cain, 2012; Silva & Cain, 2015). On their side, child reading skills have also been related to essential adult outcomes such as earnings (Crawford & Cribb, 2013, 2015) and socioeconomic status (Ritchie & Bates,

¹⁸ Some examples of studies linking RV with reading skills are Hemphill and Tivnan (2008); Muter et al. (2004); Scarborough (2009); Suggate et al. (2018); Verhoeven and van Leeuwe (2008)

2013). RV was measured since round 2 using the Hispanic adaptation of the Peabody Picture Vocabulary Test – Revised (PPVT-R) and using a Quechuan adaption in a minority of cases.¹⁹ We used its standardised score because it makes it possible to determine children's performance relative to their age-specific expected mean.²⁰ It is also worth highlighting that this instrument showed high levels of reliability within the YLS sample (Cueto & Leon, 2012; Cueto et al., 2009).

Early quantitative reasoning was the second cognitive outcome in round 2. Measured using the quantity subscale of the Cognitive Development Assessment (CDA) instrument, it assessed quantity-related notions such as "few", "half", "equal", among others, by indicating which from a set of pictures best fits in the description of a given statement (Cueto et al., 2009).²¹ Compared to the PPVT, the CDA showed lower reliability, although its levels were acceptable (Cueto et al., 2009).

Early grade literacy was measured in round 3 using the Early Grade Reading Assessment (EGRA) test. This test is not a direct measure of reading capabilities since expecting such an outcome would still not be appropriate given the developmental stage of the children. Instead, it allows a "nuanced identification of the prereading and reading skills that a specific child has already acquired and those that the child still must master to become a competent reader in the language in which he or she is being taught to read" (Gove & Wetterberg, 2011, p. 11). In other words, EGRA allows measuring how are children progressing in acquiring the foundations needed for reading.²² The YLS adapted three subtasks of this test: "familiar word reading", "listening comprehension", and "oral reading fluency with comprehension" (Cueto & Leon, 2012).²³ Like the PPVT, the tests showed acceptable, although slightly lower, levels of reliability (Cueto & Leon, 2012).

¹⁹ Detailed information about the test can be found in Dunn et al. (1986).

²⁰ We must, nevertheless, point out a limitation of our decision. According to Cueto et al. (2009), it was possible to standardise this instrument because its Spanish version had been already validated in a Latin-American sample. However, we lack details about this process such as whether part of that validated sample came from Peru or when did the validation take place.

²¹ For instance, one of the statements was, "Look at the bowls of eggs. Point to the bowl that has **the most** eggs".

²² Such foundations include recognizing letters of the alphabet, reading simple words, and understanding sentences and paragraphs (Gove & Wetterberg, 2011).

²³ A description of these and all other subtasks of the EGRA can be found in Gove & Wetterberg (2011).

Mathematical skills and *reading comprehension* were also measured through tests. Math was introduced in round 3 after adapting released items of the 2003 Trends in International Mathematics and Science Study (TIMSS) and the Programme for International Student Assessment (PISA) tests. In round 3, the adapted mathematics test showed higher reliability than the EGRA and slightly lower than the PPVT (Cueto & Leon, 2012). Meanwhile, reading comprehension started at round 4 using a combination of items from the PISA test, the Literacy Assessment and Monitoring Programme (LAMP), and others that were created for YLS's school survey implemented in 2011 (Dawes, 2020).

Non-cognitive skills:

The YLS also measured several non-cognitive dimensions from round 3 onwards. [Table 3](#) also lists those that we included in this study, and [Table B1](#) shows the items of each dimension. In all cases, these outcomes are measured through Likert scales.

Self-esteem refers to "the degree to which the qualities and characteristics contained in one's self-concept are perceived to be positive" (American Psychological Association). It has been highlighted in the psychology literature for its protective role against stress, allowing people to become less affected by negative consequences, and its influence on emotional stability (Brown, 2010; Park & Crocker, 2008). As Zeigler-Hill (2013) indicates, self-esteem has been associated with several outcomes across dimensions such as mental and physical health, interpersonal relationships, academic performance, and crime. Moreover, following the review of recent studies by Orth and Robins (2014), it appears to be relatively stable since adolescence and a strong predictor of future personal metrics of well-being and success in life. In line with the intentions of our study, other works have found negative associations between father absence and self-esteem (Krauss et al., 2020; Luo et al., 2012). We hypothesise that this relationship could be mediated by parenting skills (Yang & Schaninger, 2010) and the quality of parent-child interactions (Orth, 2018). Self-esteem was measured during rounds 4 and 5 using a series of items related to children's pride, mainly regarding their appearance, together with other items from the Self-Description Questionnaire (SDQ) (Marsh, 1990), covering more general aspects of own self-worth. Both instruments showed good levels of reliability (Ogando & Yorke, 2018).

Agency is related to Rotter (1966) internal locus of control, which is the belief that a person has on influencing the outcomes of their lives through their means and efforts. This is a relevant concept because people with high levels of agency perceive having more control over their lives, give less importance to external factors, such as luck or fate, set personal goals, and dedicate more efforts to achieve them (Mirowsky & Ross, 2007). Meanwhile, people with low levels of personal control are less motivated to engage actively in problem-solving behaviours when facing challenges in life (Caplan & Schooler, 2007). Agency was measured using the same items since round 3, all of which looked to state children's perceptions of their control over their lives and future.²⁴ This scale was retained throughout the study even though it showed weak levels of reliability (Ogando & Yorke, 2018).

Self-efficacy is a dimension closely related to agency that refers to the subjective perception of achieving a result in a given situation (American Psychological Association). It has been studied as a relevant outcome due to its importance in translating knowledge and skills into accomplishments through its influence on cognitive motivation (Bandura, 1993). This dimension was incorporated in round 4 using the Generalised Self-Efficacy Scale (GSES) (Schwarzer & Jerusalem, 1995), which showed adequate levels of reliability (Ogando & Yorke, 2018).

Table 3. Outcomes of interest by round

Round	Cognitive outcomes	Non-cognitive outcomes
2	Receptive vocabulary Early quantitative reasoning	-
3	Receptive vocabulary Early grade literacy Mathematics	Agency
4	Receptive vocabulary Reading comprehension Mathematics	Agency Self-esteem Self-efficacy
5	Receptive vocabulary Reading comprehension Mathematics	Agency Self-esteem Self-efficacy

Source: own elaboration

²⁴ For instance, one item was: "If I study hard, I will be rewarded with a better job in the future".

Early father absence:

The literature review already mentioned the importance of investments during early childhood development, as this is a moment of relevant changes in the brain structure. Taking this into account, our measure of father absence uses information from rounds 1 and 2. Based on the frequency of father-child interactions during the six months prior to each round, we defined “absence” as not having interacted with the child every day in both rounds and “presence” as having interacted with the child on a daily basis in the two periods.²⁵ We excluded every children that did not fall in any of those two groups: those who experienced early father absence in round 1 but were seeing their fathers regularly at round 2 (31%); and those who saw their fathers every day during round 1 but did not keep the same contact by the time of round 2 (13%).

As indicated in [Table 1](#), children during round 2 were aged between 4 and 6 years old, most being in the age of 5. We chose to study father absence during the first five years and not only in round 1, when children were between 6 and 15 months, old because we considered that this more prolonged exposure to absence is more appropriate for observing associations of this event and children’s outcomes. For instance, claims regarding the developmental window of opportunity closing after the first year or the so-called 1,000 days period do not seem to be supported by scientific evidence (National Scientific Council on the Developing Child, 2007). Therefore, studying a longer exposure to absence appears to be relevant. Additionally, defining five years as the period of interest for father absence has other advantages. According to the National Research Council (2000), children at the age of five should have already developed adult levels of low-level brain connections such as receptive language and speech production, together with a higher share of connections related to higher cognitive functions. These authors also indicate that children should have already acquired an expanded range of emotions and even some levels of self-understanding and social awareness at this age. Thus, the first five years seem to be a relevant window of emotional development where the role of parents seems to be fundamental.

On other side, we decided to discard children who did not maintain the same status during both rounds because keeping them would have add some noise to our results.

²⁵ The frequencies indicated in the surveys were "daily", "weekly", "monthly", "less than monthly", or "not in the last six months".

Otherwise, the findings would have been a combination of prolonged father absence with these two other cases of less extended absence. Nonetheless, it is worth mentioning that our measure of father absence still has some degree of uncertainty because we could not observe what happened between both rounds. It could be, for instance, that some fathers decided to be involved for only a short period in the middle of the rounds, and we would not have been able to know it.

Our motivation for not using a more extreme definition of absence, such as focusing on those who did not see their children in the last six months in both rounds, is related to our interpretation of the relationship between parental inputs and child skill formation. We assume that fathers' daily presence is crucial for child skills investments because it is commonly a sign of biparental household structure. From a financial perspective, fathers living with their children can concentrate their income only on one household and, therefore, make higher contributions to investments in child skills. This situation also makes it easier for fathers to contribute through their use of time in aspects such as childcaring and supporting mothers with other household responsibilities. Furthermore, even if some fathers who see their children every day do not cohabitate in the same household, this frequency of interactions could signify their willingness to contribute monetarily, engage in childcaring, and support mothers. Thus, independently of the cohabitation situation, present fathers (according to our definition of presence) might also impact children directly via father-child interactions and indirectly through their potentially positive effects on mothers. Regarding this last point, mothers should be in a better state of mental health if they feel more supported by fathers and less exposed to financial hardships.

3.3. Statistical approach

As mentioned in the introduction, this study aims to add evidence to the academic research about the associations of early father absence and children's outcomes. Therefore, it is helpful to present this work under Rubin (1974) causal model of potential outcomes as a starting point. Following Holland (1986), we would be interested in the following estimator:

$$\tau_{\text{treat}} = E[Y_i(1) - Y_i(0) | W_i = 1]$$

τ_{treat} is the *average treatment effect on the treated* (ATT) or, in our case, the effect of early father absence (from now on the "treatment") on children who experienced this event. This estimate is the expected value of the difference in an outcome (for instance, self-esteem) if they were treated $[Y_i(1)]$ against the same outcome if they were not treated $[Y_i(0)]$ conditional on having been treated ($W_i = 1$). Considering this conditionality, the second term of the subtraction $[Y_i(0)]$ is impossible to observe and represents "the fundamental problem of causal inference".

However, we could observe outcomes in another group of untreated children (from now on the "control group") and compare them with our treated group to retrieve an observable estimate:

$$\tau_{\text{treat}} = E[Y_i(1)|W_i = 1] - E[Y_i(0)|W_i = 0] + \varepsilon_i$$

Whether this alternative allows recovering a reasonable approximation of the ATT depends on the existence and size of the selection bias (ε_i) that may arise from substituting the unobservable term $[Y_i(0)|W_i = 1]$ with the feasible alternative $[Y_i(0)|W_i = 0]$. For example, early father absence might be a consequence of unplanned pregnancies, which might be related to parents' age, as many unwanted pregnancies can occur more frequently during adolescence. Age may also be associated with the level of parenting skills, as teen parents might not have enough knowledge or be in an adequate emotional state to engage in parenthood properly. Thus, if we observe a negative effect on self-esteem, it would be difficult to determine how much is attributable to father absence and how much to the exposition to mothers' low parenting skills. Even though we are aware of this issue, the main problem is that just like $[Y_i(0)|W_i = 1]$, the selection bias is directly unobservable. Nevertheless, we can somewhat understand its existence and magnitude by determining if the children from the control group differ systematically from those whose fathers were absent (from now on, the "treatment group").

As Imbens (2015) states, extracting an unbiased estimate of the ATT in the presence of selection bias requires the validity of the *strong ignorability* condition, which is composed of two assumptions. Firstly, the *unconfoundedness* assumption implies that

conditional on a set of relevant covariates, the treatment status and outcomes. However, this is a strong assumption to hold because it means that we should incorporate in the analysis all covariates that are correlated with the outcome and treatment status and that have not been affected by this last variable. Achieving this is quite challenging because we would need to identify all these variables and measure them. Secondly, the *overlap* condition implies that both groups should share the same range of values in relevant covariates.

One of the contributions of this work is the use of *matching* and *weighting* methods to study the potential associations of early father absence. The basic idea behind matching is to group treated and very similar control units in terms of specific relevant characteristics, calculate the differences among those subgroups and then average them. Thereby, matching methods try "finding experimental data hidden within an observational data set" (King & Nielsen, 2019, p. 9). Weighting methods, on their side, assign values (weights) to each observation so that units more similar to each other gain more importance in the analysis than those that are more different. Since we are interested in calculating the ATT, we will reweight control units while keeping the same unitary weight to all treated units.

Simulations and empirical studies have shown no universal answer regarding the best metrics and algorithms among the vast set of matching and weighting alternatives for estimating ATT.²⁶ Instead, the optimal choice depends on elements related to the data generation process in each case. For this reason, we evaluated several options before deciding which to use. We tested the two metrics most used in the academic literature, the propensity score and the Mahalanobis distance, together with two matching algorithms, one-to-one nearest neighbours with replacement and kernel density. Besides, we also tested two weighting strategies: inverse probability weighting and entropy balancing (EB). All these alternatives are briefly described in Section 1 of [Appendix C](#).

We used the same set of covariates when testing each option. Following Caliendo and Kopeinig (2008), our covariates were either time-invariant characteristics or other

²⁶ See, for example, Augurzky and Kluve (2007); Austin (2014); Bai (2011); Busso et al. (2014); Dong et al., (2020); Galdo et al. (2008); Huber et al. (2013); Zhao (2004).

retrospective information about events that happened *before* the birth of the YLS children. Most were extracted from round 1 and the rest from round 2. All variables were related to our outcomes and early father absence without having been affected by this event. The selection of covariates was challenging because our treatment variable reflected father-child interactions during the last six months. In this sense, many variables that are theoretically relevant to predicting early father absence or our outcomes could have been already influenced by this event when round 1 was collected.²⁷ As Gertler et al. (2016) discuss, including such variables might produce invalid or biased estimations. For instance, if we had included mothers' participation in the labour market during round 1 and found a positive association between the variable and early father absence, we would not have been able to discern the correct direction of this association. It could have been that working mothers had a higher propensity to separate from fathers because they had higher levels of empowerment than mother who did not work. However, it could have also been the case that separated mothers had a higher proclivity to work due to the economic difficulties related to the separation. We, therefore, opted to exclude those types of variables.

Not using all the available and relevant information produces a significant limitation when interpreting the results of this work. As mentioned earlier, making causal claims under potential outcomes requires assuming strong ignorability, which is already challenging under normal circumstances. This means presuming that the groups do not differ in terms of *unobservable* characteristics. In other words, we need to assume the absence of any omitted variable bias. In our case, this supposition is even more challenging to support since we excluded relevant variables from our analysis. Consequently, even though our results might be *closer* to the true effects of early father absence, they should still be interpreted cautiously and taken as associations rather than causal effects. This issue is further discussed in Section 6.

The logistic regression results displayed in [Table C1](#) show that most of our variables help to predict early father absence. Households that experienced adverse productive shocks such as the death of livestock or failure of crops had higher odds of early father

²⁷ For instance, mother participation in the labour market is an example of a variable that is theoretically relevant to explain early father absence and that we could not include when analysing matching and weighting alternatives. Meanwhile, other variables correlated with children's outcomes that were available but could not be included are children's health and household income.

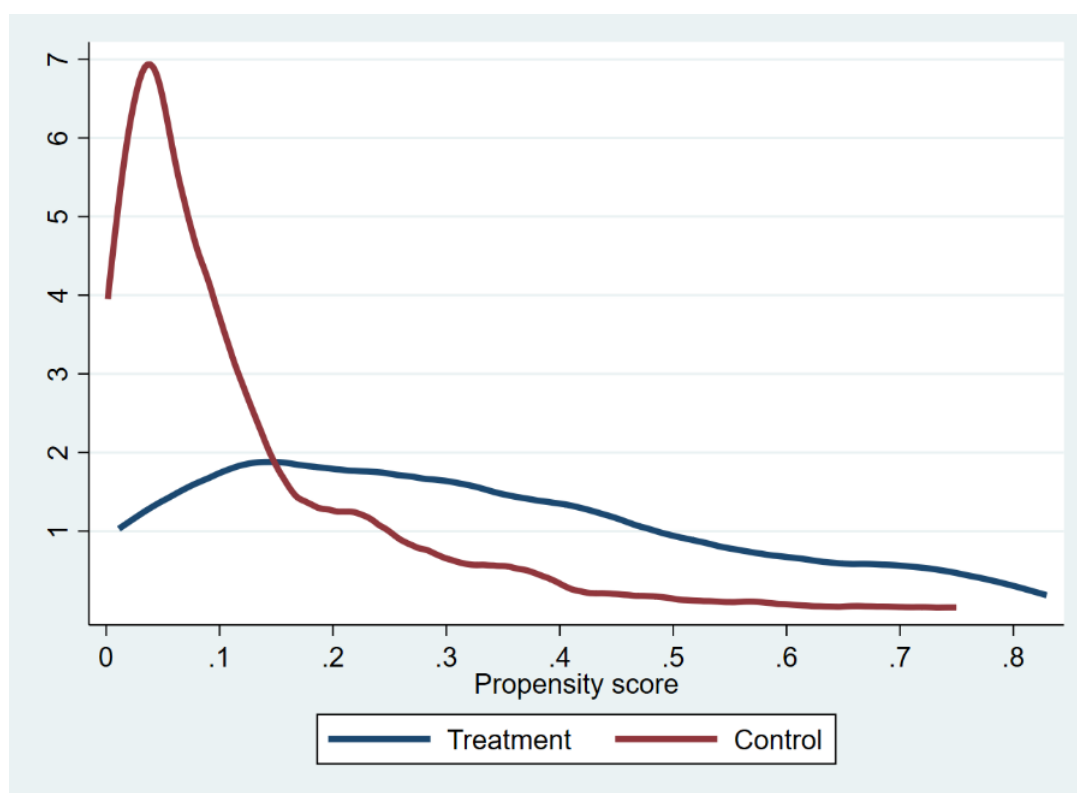
absence. This could be related to a higher disposition of fathers for engaging in work migration to offset the economic consequences of those events. Variables that are somewhat related to the levels of stability in family structures also showed the expected relationships. On one side, lower odds were observed in variables that show higher stability, such as having had children before the YLS child, YLS pregnancies that were wanted, and mothers who decided to move for cohabitating before the YLS children's birth. On the other side, variables related to weaker stability were associated with higher odds. For instance, YLS pregnancies that happened during adolescence (although not statistically significant), mothers who had experienced family disputes, and mothers who reported having had a difficult time during their YLS pregnancies had higher likelihood of having separated from the YLS fathers. Similarly, not having a male offspring is also related to higher odds, which aligns with the findings from Amato and Patterson (2017) and might be related to some fathers' frustrations of not having a son. Finally, mothers that reported having been beaten during their childhood are also less likely to experience absence.

[Figure 1](#) shows the distribution of the propensity score in both groups. The common support condition holds for the great majority of the sample: only 2% of treated and 5% of control units fell outside this zone. This is positive because we will only lose few observations during the analysis. [Figure 2](#) summarises all tested alternatives.²⁸ We chose two metrics to evaluate the disparities between the treatment and control groups in our set of covariates: the mean standardised differences (left side of the figure) and variance ratio (right side). The first row in both plots corresponds to the raw values (before applying any matching or weighting technique). All options enhanced the comparability: they reduced the standardised mean differences by at least 66% and variance ratios by at least 27%. Entropy balancing (EB) successfully eliminated all observable differences, which made it, logically, one of our chosen algorithms. Our second choice was the kernel matching with propensity score (KPS), considering that

²⁸ More information about how well each alternative reduced the differences between the groups can be found in the Section 3 of [Appendix C](#).

it showed the second-highest gains in bias while performing similarly in variance ratio reduction as other alternatives.

Figure 1. Propensity score distribution by group



Source: own elaboration with data from YLS

4. Descriptive evidence

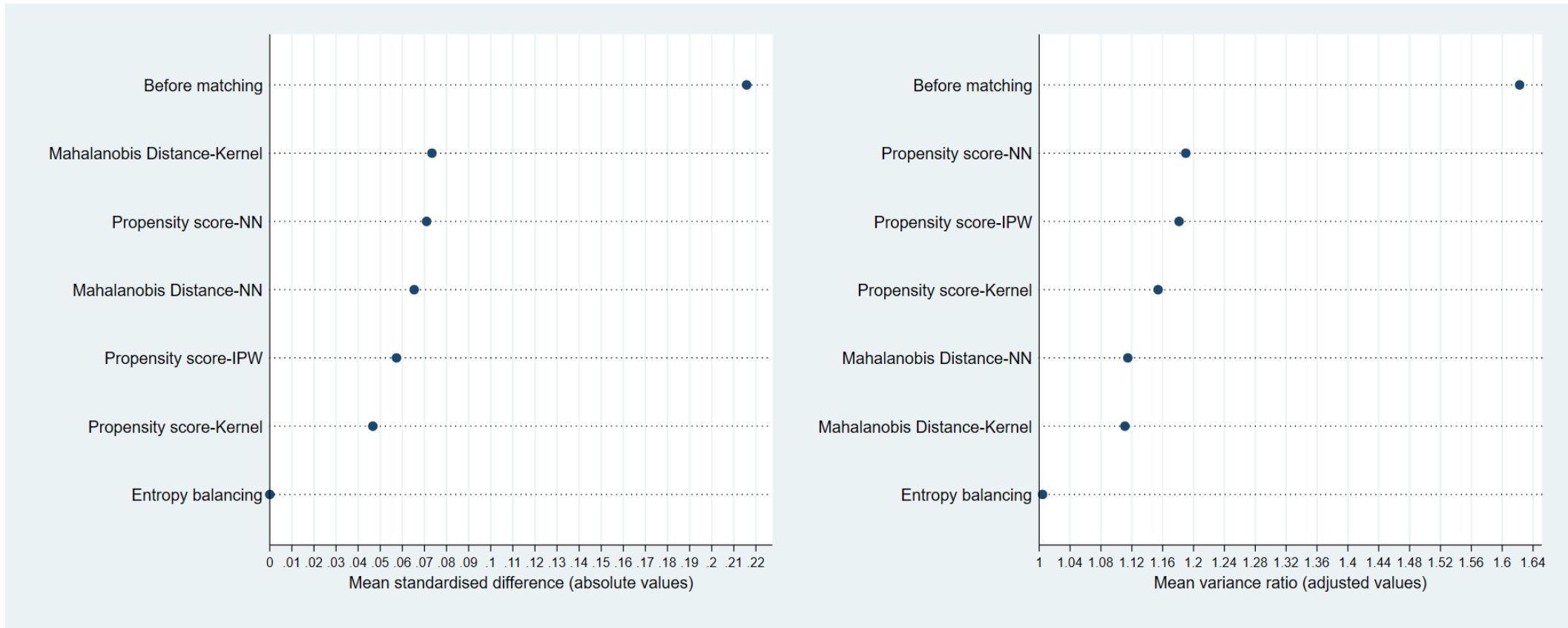
4.1. Comparing children who experienced and did not early father absence

[¡Error! No se encuentra el origen de la referencia.](#) showed that father absence is related to characteristics mainly consistent with weaker family structures. This is corroborated in [Table 4](#), which shows the raw standardised mean differences between both groups in the fourth column. YLS mothers of children who experienced early father absence differed consistently from those who did not go through this event. They seemed to have been less prepared to face the YLS pregnancy. They were, on average, younger, had a higher likelihood of having been teenagers during that period and having been going through their first pregnancy, and were less likely to have wanted to be pregnant at that time. Perhaps a sign of this weaker readiness is their lower odds of attending at least one antenatal health visit. Age differentials might also be determinants of the differences in pregnancy readiness and could also lead to

discrepancies in other aspects like childcaring, parenting, and household income. Moreover, mothers who experienced early father absence also had a lower propensity to cohabitate with the YLS children's fathers before and during pregnancy. This decision could be a sign of already low interest of fathers during that time. Therefore, not only were the treated YLS children exposed to early father absence, but they may have also been born after more challenging pregnancies and may have been raised by women who were less prepared to become mothers. These additional conditions could be related to other pathways that might have also affected the skills investments they received during their first years.

However, other aspects might act, in contrast, as positive confounders for children's outcomes. Mothers in the treatment group were more likely to live in a house owned by a household member. This can be related, at least in some cases, to more stability when raising their children, lower expenses in housing, and potentially higher levels of support from other household members. Hence, as McLanahan et al. (2013) mentioned, comparing the outcomes of children who experienced and did not father absence by performing simple mean differences or an OLS regression might yield biased results. In our case, this bias could go any way depending on the strength of these opposite factors. This corroborates the relevance of using other alternatives such as matching and weighting methods.

Figure 2. Performance of the matching and weighting alternatives (summary)



Source: own elaboration with data from YLS. The standardised differences are expressed in absolute values so that negative and positive values do not cancel out when calculating the mean values. The variance ratios were also adjusted considering that values smaller and larger than 1 would have also cancelled each other out when calculating the mean values. This second adjustment consisted in inverting all ratios that were below 1. For example, ratios of 0.90 and 0.50 were re-expressed, respectively, as 1.11 ($1/0.90$) and 2 ($1/0.50$). *NN* stands for 1-1 Nearest Neighbours with Replacement and *IPW* for Inverse Probability Weighting. The entropy balancing was performed targeting balance in means and variances. The kernel algorithms used a cross-validated bandwidth (see [Appendix C](#))

Table 4. Comparing children who experience and did not early father absence

Covariates	Treatment (T)	Control (C)	Difference (T-C)
Household setting: urban	-0.09	-0.00	-0.09
External bad event index	0.16	0.00	0.16**
Mum education level	0.10	0.00	0.10
YLS pregnancy: mum age	-0.35	0.00	-0.35***
YLS-pregnancy: teen	0.42	-0.00	0.42***
Experienced death of a child	-0.23	-0.00	-0.23***
Experienced abortion/miscarriage	-0.19	-0.00	-0.19**
YLS-parents have no boys	0.40	0.00	0.40***
YLS-child not first child	-0.67	-0.00	-0.67***
YLS-pregnancy was wanted	-0.18	0.00	-0.18**
Cohabitation during YLS-pregnancy	-0.33	0.00	-0.33***
Cohabitation before YLS-pregnancy	-0.04	-0.00	-0.04
Separation before YLS-pregnancy	0.03	-0.00	0.03
Family dispute before YLS-pregnancy	-0.01	0.00	-0.01
Never moved before YLS-pregnancy	-0.30	0.00	-0.30***
Some antenatal visit	-0.21	-0.00	-0.21***
YLS-pregnancy was difficult	0.05	-0.00	0.05
Mum beaten during childhood	-0.59	-0.00	-0.59***
Gave birth at home	-0.12	-0.00	-0.12
No health professionals at birth	-0.09	-0.00	-0.09
Grandmother's first language not Spanish	-0.21	-0.00	-0.21***
Mother has difficulties reading in Spanish	-0.14	0.00	-0.14*
Own house	0.16	0.00	0.16**

Source: own elaboration. * $p < 0.1$, ** $p < .05$, *** $p < 0.01$. Covariates were standardised using the control group's mean and standard deviation.

5. Results

5.1. Main results

[Table 5](#) and [Table 6](#) show, respectively, the ATT of early father absence in cognitive and non-cognitive outcomes during several stages in YLS children's lives. The first column below each outcome corresponds to simple differences in means using OLS

regressions, the second to the estimates generated by EB, and the third by KPS. In the case of EB, we decided to comply with the common support condition to enhance comparability with the KPS estimators.²⁹

Early father absence seems to be related to some losses, especially in cognitive outcomes. We observe negative differences in math skills that become statistically significant at conventional levels from round 4 onwards with coefficients between -0.11 and -0.15 standard deviations (SD). We also find negative associations between -0.17 and -0.18 SD in reading during the last round, also statistically significant. In terms of non-cognitive outcomes, we observe a relatively high drop in self-esteem (between -0.17 and -0.23 SD). The rest of the coefficients in both types of outcomes do not have the required precision and magnitude to be highlighted as noteworthy associations. The only exception is the case of agency, which shows a positive KPS estimate of 0.15 SD during adolescence. Nevertheless, the estimations using EB do not corroborate this magnitude.

Another point worth highlighting is that raw differences seem to be yielding somewhat upwardly biased estimates, especially in the case of cognitive outcomes. After applying the matching and weighting methods, most of the positive OLS coefficients are either cut by half (or more) or have their signs reverted, while most negative OLS coefficients double their magnitudes. Therefore, even though children who experienced early father absence had to face other difficulties that may have also hindered their skills development, it seems that positive confounders such as living in a place owned by a household member may have had a more substantial influence. However, this cannot be proven.

[Table 7](#) and [Table 8](#) show the results of dividing the sample by children's sex. For brevity, the tables only display the coefficients generated by EB. Since this method does not require following the common support assumption, these calculations include the whole effective sample. Boys have the highest negative coefficients consistently in mathematics starting in round 3. While their negative associations are -0.16 and -0.25 SD in rounds 3 and 4, respectively, girls show much smaller coefficients: -0.01

²⁹ The EB results using all the effective sample are similar can be seen in columns 4-6 of [Results of the sensitivity analysis Table D1](#) and [Table D2](#).

and -0.06 SD. Then, in round 5, boys show a difference of -0.21 SD, which doubles the girls' coefficient. The differences between both sexes are in no round statistically significant. However, this could be related to the lack of statistical power that results from splitting our sample in almost half in each case, especially considering that the treatment groups represent roughly between 13 and 15% of both subsamples. We further discuss this issue in Section 6. Despite this lack of significance, the magnitudes of both estimates somewhat suggest that boys' may have had more difficulties in mathematics associated with early father absence. In contrast, both boys and girls show relatively high negative associations in reading skills in round 5. Although only the boys' coefficient (-0.26 SD) is statistically significant in this case, it seems that both may have experienced challenges in reading skills related to early father absence.

Other results arise in non-cognitive outcomes. There are large positive associations between boys' self-efficacy in round 4 (0.22 SD) and agency in round 5 (0.32 SD). Instead, girls show high negative differences in self-esteem (-0.21 SD) and self-efficacy (-0.37 SD) in round 4 and self-esteem (-0.33 SD) in round 5. These between-sex differences are relatively large but, like in cognitive outcomes, not statistically significant, potentially due to the same issues with statistical power.

5.2. Mechanisms

According to the model of skills formation, the influence of early father absence on these outcomes could be explained by lower investments in child skills. The YLS does not measure directly either the magnitude of the monetary expenses on children nor the quantity or quality of the time mothers and other caregivers spent with them. Taking this into account, we used proxy variables to explore some of these mechanisms in [Table 9](#). The table differentiates between three groups of variables: measures related to household income as proxies of monetary investments, variables associated with the availability of time, and approximations to whether father absence may have influenced mothers. Each column represents the adjusted differences between the treatment and control groups for a given round.

There are two main elements about the income variables. Firstly, households that experienced early father absence had a much higher likelihood of having received monetary remittances. On one side, it may be that some fathers kept being involved

in their children's development through monetary transfers despite their physical absence. On the other side, it could also be that mothers received consistently financial support from relatives living outside the household. In the second place, despite the higher monetary transfers, the other variables suggest that father absence is still associated with lower income. For instance, we see constant drops in households' per capita expenditure and worse perceptions of their socioeconomic status. Also, we find a reduction in the wealth index in round 4, a variable related to structural aspects of households' financial situation.

The variables associated with caregivers' time availability are also consistent with reductions in household income. Mothers belonging to households with early father absence showed a higher propensity permanently to participate in the labour market, as well as a higher number of jobs during the first two rounds of the survey. Both could be potential consequences of the adverse associations between household income and early father absence. In this sense, mothers had likely less time to dedicate to their children throughout their childhood, especially during their first five years of life. The third variable measures the absence of at least another potential caregiver in the household apart from mothers: fathers, stepfathers, grandparents, uncles, or aunts. Even though our treatment group has, by definition, a lesser presence of fathers, this variable should not show positive differences if any of the other family members substitutes their place. The large and consistent positive estimates show that this was not necessarily the case, as mothers in the treatment group were more likely not to have cohabitated with any other source of support. This variable is still an imperfect measure since it says nothing about how much support mothers have obtained from people outside their household. Nevertheless, it still suggests that mothers may have experienced more difficulties since sharing the household with another caregiver could have made it easier to arrange for childcaring whenever needed to work.

Finally, the last set of variables is related to potential effects on mothers' mental health, perceptions, and parenting skills. Even though both groups showed similar levels of subjective well-being throughout the rounds, agency was consistently lower in the treatment group, with negative associations that were statistically significant in the rounds 2,3, and 4. These differences are relevant because this variable measured caregivers' perceptions about their capabilities to help their children in aspects related

to their health and education.³⁰ Additionally, we observed a negative association during round 2 with self-esteem, measured as how proud they were regarding several aspects of their lives. Likewise, we find some evidence suggesting that caregivers who experienced father absence may have had, on average, higher expectations of their children, perhaps due to the financial complications still experienced. For instance, compared with caregivers in the control group, during round 4, they were expecting their children to start earning money at an earlier age. Additionally, during round 5, they expected their children to move out of the household at a younger age.

Although we cannot tell straight from the YLS survey whether all these aspects could have affected caregivers' parenting styles and behaviours, the last two rounds measured children's perceptions about their relationship with their parents.³¹ In both rounds, children from the treatment group had, on average, a much worse view of this relationship. Nevertheless, this association should be interpreted cautiously because it is unclear whether treated children thought only about their mothers' attitudes and behaviours when answering or if they also considered their fathers' absence, and, where applicable, also their relationships with their stepfathers.

Finally, to make more sense of the suggestive evidence regarding potentially greater difficulties for boys in math skills and girls in some non-cognitive outcomes, we also explored between-sex differences in children's use of time. For instance, boys' math skills could have been more negatively affected than girls if they had a higher reduction of their hours dedicated to studying and doing school assignments. This may have happened if they had to increase their participation in income-generating or household responsibilities (such as household chores or caring for other household members) that in the presence of fathers may have been allocated using more traditional gender-based criteria. Meanwhile, girls' non-cognitive outcomes could have also been more adversely affected if they had to dedicate more time to helping their mothers or replacing them in certain activities they could no longer keep doing due to their work obligations.

³⁰ For instance, items included "If my child gets sick, I can do little to help him/her get better" and "I can do little to help my child do well in school, no matter how hard I try"

³¹ This variable was measured using items from the SDQ. Examples of the items include: "I like my parents", "my parents like me", "I get along with my parents", "my parents understand me", "my parents are easy to talk to".

[Table 10](#) shows the results of this analysis, which do not support our hypotheses. Boys neither spent consistently more time on income-generating nor did they dedicate more time to household-related activities. Besides, their study time was not affected. On the contrary, even though girls did not spend more time in household chores, they did increase in rounds 2, 3, and 4 their participation in other activities related to income generation, such as doing tasks on family farms or family businesses. This rise was particularly high during round 4 (0.40 SD), when they were old enough to contribute more while still not having the higher levels of freedom and independence that come in adolescence. Even though these between-sex differences were only statistically significant in round 4, the coefficients of rounds 2 and 3 are close enough, especially considering the issues of statistical power.

Table 5. ATT of early father absence on cognitive outcomes (complete sample)

	PPVT			EGRA/Reading			CDA/Math		
	(1) OLS	(2) EB	(3) KPS	(4) OLS	(5) EB	(6) KPS	(7) OLS	(8) EB	(9) KPS
Round 2 (4-6 years)	-0.063 (0.1141)	-0.044 (0.0908)	-0.063 (0.0898)	-	-	-	-0.11 (0.1082)	-0.13 (0.1073)	-0.16 (0.1187)
Round 3 (7-8 years)	-0.044 (0.1042)	-0.074 (0.0779)	-0.075 (0.0907)	0.061 (0.0990)	-0.017 (0.0873)	-0.043 (0.0858)	0.038 (0.0756)	-0.087 (0.0648)	-0.10 (0.0702)
Round 4 (11-12 years)	-0.031 (0.1067)	-0.038 (0.0816)	-0.034 (0.0829)	0.033 (0.0911)	0.011 (0.0781)	-0.001 (0.0817)	-0.068 (0.0918)	-0.15** (0.0637)	-0.14** (0.0660)
Round 5 (14-15 years)	0.005 (0.0961)	-0.031 (0.0714)	-0.060 (0.0721)	-0.086 (0.0895)	-0.18** (0.0843)	-0.17* (0.0836)	-0.12 (0.0695)	-0.14** (0.0506)	-0.11** (0.0509)
Observations	1,296	1,296	1,296	1,296	1,296	1,296	1,296	1,296	1,296

Source: own elaboration. Clustered standard errors (by sentinel site) in parentheses * $p < 0.1$, ** $p < .05$, *** $p < 0.01$. All estimates are expressed in effect sizes based on the control group's mean and standard deviation. Empty cells ("-") indicate that the outcome was not measured at that round. EB stands for "entropy balancing" and KPS for "kernel propensity score".

Table 6. ATT of early father absence on non-cognitive outcomes (complete sample)

	Agency			Self-esteem			Self-efficacy		
	(1) OLS	(2) EB	(3) KPS	(4) OLS	(5) EB	(6) KPS	(7) OLS	(8) EB	(9) KPS
Round 3 (7-8 years)	0.037 (0.0675)	-0.007 (0.0592)	-0.003 (0.0613)	-	-	-	-	-	-
Round 4 (11-12 years)	-0.056 (0.0858)	0.012 (0.1102)	0.047 (0.1261)	-0.090 (0.0606)	-0.091 (0.0856)	-0.11 (0.0894)	0.004 (0.0712)	-0.067 (0.1020)	-0.10 (0.1302)
Round 5 (14-15 years)	0.12* (0.0662)	0.087 (0.0752)	0.15* (0.0794)	-0.12* (0.0658)	-0.23** (0.1053)	-0.17* (0.0833)	0.018 (0.0825)	-0.018 (0.1099)	0.057 (0.0967)
Observations	1,296	1,296	1,296	1,296	1,296	1,296	1,296	1,296	1,296

Source: own elaboration. Clustered standard errors (by sentinel site) in parentheses * $p < 0.1$, ** $p < .05$, *** $p < 0.01$. All estimates are expressed in effect sizes based on the control group's mean and standard deviation. Empty cells ("-") indicate that the outcome was

not measured at that round. EB stands for “entropy balancing” and IPW for “inverse probability weighting

Table 7. ATT of early father absence on cognitive outcomes (differences by sex)

		(1) PPVT	(2) EGRA/ Reading	(3) CDA/ Math
Round 2 (4-6 years)	Father absence-Male	-0.006 (0.1339)	-	-0.125 (0.1331)
	Father absence-Female	-0.066 (0.1294)	-	-0.149 (0.1399)
	Difference (Female-Male)	-0.059 (0.1683)	-	-0.025 (0.1455)
Round 3 (7-8 years)	Father absence-Male	-0.156 (0.1229)	-0.163 (0.1228)	-0.157 (0.1228)
	Father absence-Female	0.007 (0.1197)	0.035 (0.1352)	-0.012 (0.1216)
	Difference (Female-Male)	0.162 (0.1171)	0.198 (0.1423)	0.145 (0.1351)
Round 4 (11-12 years)	Father absence-Male	-0.066 (0.1219)	-0.017 (0.1301)	-0.248** (0.1166)
	Father absence-Female	0.034 (0.1181)	0.091 (0.1285)	-0.056 (0.1246)
	Difference (Female-Male)	0.099 (0.1206)	0.108 (0.1903)	0.192 (0.1843)
Round 5 (14-15 years)	Father absence-Male	-0.068 (0.1264)	-0.259** (0.1307)	-0.208 (0.1359)
	Father absence-Female	0.039 (0.1123)	-0.186 (0.1207)	-0.100 (0.1218)
	Difference (Female-Male)	0.107 (0.1275)	0.073 (0.1745)	0.108 (0.1925)
Observations		1,360	1,360	1,360

Source: own elaboration. Clustered standard errors (by sentinel site) in parentheses
^{*} $p < 0.1$, ^{**} $p < .05$, ^{***} $p < 0.01$. Estimates correspond to calculations using entropy balancing. All coefficients are expressed in effect sizes based on the control group's mean and standard deviation. Empty cells ("-") indicate that the outcome was not measured at that round. Male sample size was 685 observations (92 treated and 593 controls) due to a. Female sample size was 675 observations (104 treated and 571 controls).

Table 8. ATT of early father absence on non-cognitive outcomes (differences by sex)

		(1) Agency	(2) Self- esteem	(3) Self- efficacy
Round 3 (7-8 years)	Father absence -Male	0.103 (0.1247)	-	-
	Father absence-Female	-0.112 (0.1246)	-	-
	Difference	-0.215 (0.1666)	-	-
Round 4 (11-12 years)	Father absence -Male	0.017 (0.1438)	-0.003 (0.1075)	0.218* (0.1161)
	Father absence-Female	-0.013 (0.1673)	-0.209 (0.1332)	-0.371*** (0.1301)
	Difference	-0.030 (0.2455)	-0.206 (0.1250)	-0.589*** (0.1829)
Round 5 (14-15 years)	Father absence -Male	0.323*** (0.1204)	-0.126 (0.1247)	0.044 (0.1237)
	Father absence-Female	-0.092 (0.1384)	-0.334** (0.1610)	-0.095 (0.1590)
	Difference	-0.416 (0.2803)	-0.208 (0.2622)	-0.139 (0.2481)
Observations		1,360	1,360	1,360

Source: own elaboration. Clustered standard errors (by sentinel site) in parentheses
 * $p < 0.1$, ** $p < .05$, *** $p < 0.01$. Estimates correspond to calculations using entropy balancing. All coefficients are expressed in effect sizes based on the control group's mean and standard deviation. Empty cells ("-") indicate that the outcome was not measured at that round. The male sample size was 685 observations (92 treated and 593 controls) due to a. The female sample size was 675 observations (104 treated and 571 controls).

Table 9. Potential mechanisms to explain the effects of early father absence

		(1) Round 1	(2) Round 2	(3) Round 3	(4) Round 4	(5) Round 5
Income	Remittances (# transfers)	0.749*** (0.1632)	0.686*** (0.1452)	0.476*** (0.0960)	0.148 (0.1025)	-
	Wealth index	0.027 (0.1009)	0.005 (0.0930)	-0.089 (0.0910)	-0.136** (0.0583)	-0.037 (0.0789)
	Per capita expenditure (log)	-	-0.177** (0.0761)	-0.166** (0.0713)	-0.104 (0.0805)	-0.163** (0.0745)
	Perceived SES	-	-0.112 (0.0921)	-0.155 (0.1183)	-0.126** (0.0495)	-0.180* (0.0874)
Time	Mother works	0.250*** (0.0731)	0.272** (0.1014)	0.515*** (0.1333)	0.380*** (0.0783)	-
	Mother # jobs	0.464*** (0.1352)	0.441*** (0.1495)	-0.092 (0.1607)	0.087 (0.0988)	-
	No other main caregiver	1.881*** (0.2974)	3.063*** (0.4505)	0.749*** (0.1384)	0.637*** (0.1245)	0.607*** (0.1535)
Mothers	Subjective Wellbeing	-	0.027 (0.1116)	-0.068 (0.0948)	0.130 (0.0906)	-0.060 (0.0890)
	Agency	-	-0.209** (0.0843)	-0.321*** (0.0838)	-0.198** (0.0813)	-0.074 (0.1033)
	Self-esteem	-	-0.221* (0.1167)	-0.108 (0.0956)	-0.045 (0.0512)	-0.111 (0.0716)
	Expected age (start earning)	-	-	-	-0.234** (0.0975)	-0.115 (0.1232)
	Expected age (leave edu.)	-	-	-	-0.131 (0.0960)	-0.127 (0.1472)
	Expected age (leave HH)	-	-	-	-	-0.259*** (0.0885)
	Relationship with child	-	-	-	-0.309*** (0.0887)	-0.378*** (0.0910)
Observations (range)		1,356 1,360 ^a	1,322 1,360 ^b	1,308 1,360 ^c	1,271 1,360 ^d	1,285 1,360 ^e
Observations (mean) ^f		1,359	1,355	1,352	1,341	1,334

Source: own elaboration. Clustered standard errors (by sentinel site) in parentheses * $p < 0.1$, ** $p < .05$, *** $p < 0.01$. All estimates are expressed in effect sizes based on the control group's mean and standard deviation. Empty cells ("-") indicate that the outcome was not measured at that round. Sample sizes are shown in ranges and means because they somewhat varied depending on the covariate. ^{a,b,c,d,e} The sample sizes of the variables related to mothers' employment were 784 (round 1), 721 (round 2), 382 (round 3), and 802 (round 4) because the number of mothers living in the household dropped continuously between rounds. ^f Calculations do not include the sample sizes of "mothers' number of jobs" because mothers who were not employed were taken as missing values.

Table 10. ATT of father absence on children's time use (differences by sex)

	(1) Round 2	(2) Round 3	(3) Round 4	(4) Round 5
<i>Doing HH duties</i>				
Male	0.059 (0.1617)	-0.156 (0.1324)	-0.230* (0.1234)	-0.064 (0.1225)
Female	-0.036 (0.1198)	-0.051 (0.1375)	-0.020 (0.1394)	0.042 (0.1563)
Difference	-0.095 (0.1578)	0.105 (0.1943)	0.210 (0.2029)	0.106 (0.2283)
<i>Working</i>				
Male	0.072 (0.0764)	-0.116 (0.1473)	-0.162 (0.1252)	0.159 (0.1864)
Female	0.229 (0.1575)	0.141 (0.1022)	0.238** (0.1027)	-0.014 (0.0851)
Difference	0.157 (0.0969)	0.257 (0.1680)	0.400** (0.1553)	-0.173 (0.1392)
<i>Doing homework/studying</i>				
Male	-0.100 (0.1450)	-0.012 (0.1182)	0.178 (0.1392)	0.017 (0.1425)
Female	-0.064 (0.1232)	0.044 (0.1170)	-0.126 (0.1263)	-0.195 (0.1241)
Difference	0.037 (0.1640)	0.056 (0.1414)	-0.303 (0.2378)	-0.212 (0.1831)
<i>Playing/leisure</i>				
Male	0.006 (0.1497)	0.185 (0.1385)	0.015 (0.1098)	0.253 (0.1547)
Female	-0.063 (0.1385)	-0.193* (0.1127)	-0.090 (0.1508)	0.099 (0.1285)
Difference	-0.069 (0.1886)	-0.378** (0.1567)	-0.105 (0.1794)	-0.155 (0.1513)
Observations	1,359	1,359	1,360	1,360

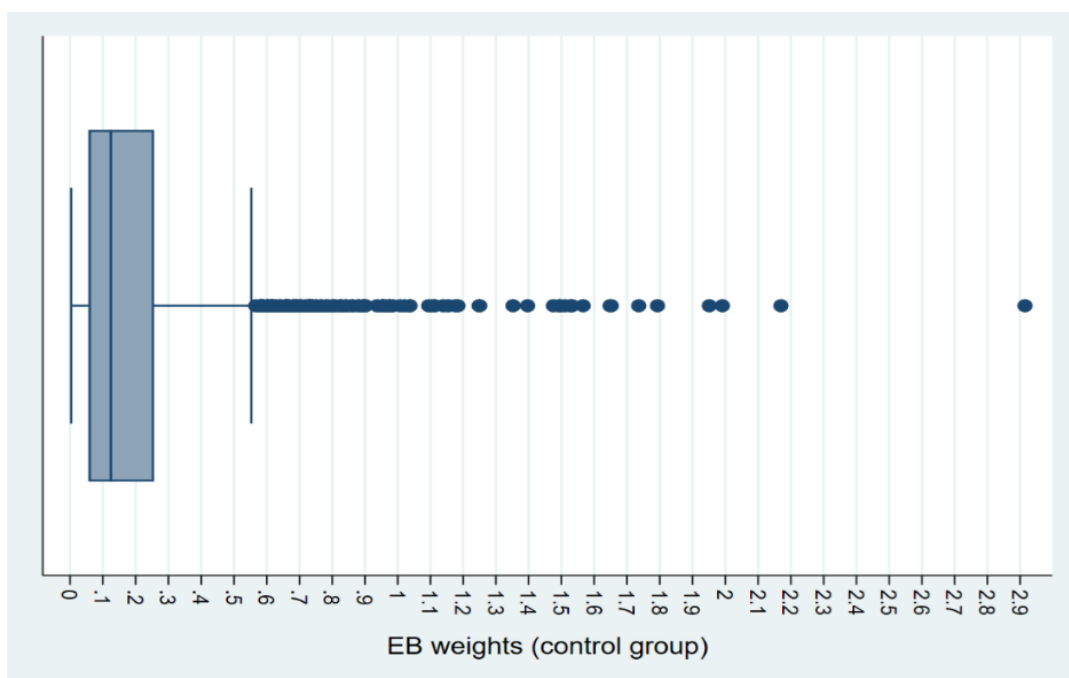
Source: own elaboration. Clustered standard errors (by sentinel site) in parentheses
 *p < 0.1, ** p < .05, *** p < 0.01. All estimates are expressed in effect sizes based on the control group's mean and standard deviation.

5.3. Sensitivity analysis

[Figure 3](#) shows that EB is giving disproportionately high weights to very few observations. More than three-fourths of the control units have weights smaller than 0.3, most of the last fourth lies between 0.3 and 0.6, and just a few have weights as large as 2.9. Even though this last group is composed of just a small number of children, their weights are extraordinarily high. For instance, the highest weight (right end of the figure) is equivalent to the sum of the 238 smallest weights (which represents 20.44% of the sample). If a very small number of controls drives our results, we would be falling too deep into model dependency. Hainmueller (2012) already warned about this risk when using EB and suggested using an iterative weight refinement strategy as a sensitivity test.

In our case, we opted to rerun our analysis excluding the observations with extreme values, which we defined as having a weight higher than 1. Even though this meant dropping only 25 observations (2% of the sample), their weights combined were 19% of the sum of weights. The results of this exercise are reported in [Appendix D](#). None of our results faded after the exclusion. On the contrary, we now observe larger effects, more being now statistically significant at conventional levels. Therefore, our results are not sensitive to the overrepresentation of this tiny group of children.

Figure 3. Comparison of control group weights (EB and IPW)



Source: own elaboration

6. Discussion and conclusion

Consistent with the framework of Cunha and Heckman (2007) regarding human skills formation, our findings suggest negative long-term associations between early father absence and children's cognitive (math and reading skills) and non-cognitive (self-esteem) outcomes. These relationships can be potentially a result of a lower accumulation of monetary and non-monetary investments in child skills throughout their childhood and adolescence. Thus, households without fathers not only lose a person who may have been the primary source of income but also someone who could have dedicated time for childcaring and supporting children in developing their skills.

Additionally, their absence puts higher pressures and burdens on mothers, who must face the challenging task of raising a child in more complicated conditions. In our case, we observed that mothers in this situation may have gone through more financial complications, and that this could have probably obliged many of them to enter the labour market. This may have impeded them to dedicate more time to their children since early ages. Besides, this situation may have also affected them non-cognitively in aspects such as agency. Given that financial hardship and mental health have been widely associated in the academic literature, this possibility seems logical.³² In this sense, it is also not surprising that they may have had higher expectations towards their children regarding their contributions to the household despite their young ages. Similarly, it is also no surprise that these children had a worse perception of their relationships with their parents, even though some could be attributable directly to fathers' absence and not to mothers' attitudes and behaviours. Besides, the greatest differences occurred in adolescence, which may be related to the particularities of this developmental stage, characterised by a series of physical and social changes and transitions (Harris et al., 2018).

Furthermore, the results of our work indicate that some associations do not seem to be gender-neutral. However, issues with statistical power do not allow us to have more certainty over this matter. Regarding boys, we found some suggestive evidence about more difficulties in mathematics and higher levels of agency (round 5) and self-efficacy (round 4). Meanwhile, girls showed higher negative associations in self-efficacy (round

³² See, for example, Broussard (2010); Crosier et al. (2007); Hope et al. (1999); Rousou et al. (2013); Stack & Meredith (2018)

4) and self-esteem (rounds 4 and 5). Additionally, we did not observe an increase in boys' dedication to household or income-generating activities, but we did find that girls seem to have augmented their participation in income-related activities such as helping in family farms or family businesses.

Considering that we did not find relevant differences in boys' use of time, their potential higher struggles in math that could be associated with father absence might be related to differences in the way boys and girls learn, making girls less susceptible to needing parental support. Even though some academics have argued that sex-based learning differences exist and are related to biological differences in brain development throughout childhood, research is still far from being conclusive in this area (Bonomo, 2017). However, if this were the case, then we should also wonder why we found, in contrast, similar negative associations in reading skills between boys and girls. One possibility is that since learning math and reading require different sets of skills, perhaps the importance of fathers' contributions differs in both cases. Further research using other samples could help to clarify these interrogations.

The positive associations of early father absence with agency (round 5) and self-efficacy (round 4) in boys, let us hypothesise that they could have been starting to assume more responsibilities and become more involved in the households' decision-making process. However, the fact that we did not find similar associations in self-esteem and self-efficacy (during round 5) makes it more difficult to determine whether those potential gains could be translated into better future outcomes for boys.

On the other side, the potential higher associations of girls' non-cognitive outcomes and early father absence are consistent with findings from Keizer et al. (2019) and Liu (2008) regarding the importance of fathers in girls' development of self-worth during adolescence. Keizer et al. (2019) argue that boys are generally more encouraged to be independent and confident from young ages, while girls are not, which is why they can find higher gains from fathers' independence-encouraging behaviours. Alternatively, these associations could be related to girls' increased participation in income-generating activities, which may have shaped in some way the interactions with their mothers. Nevertheless, more evidence regarding the effects of father absence on girls is also needed, especially in developing countries since traditional gender-based family dynamics usually are still present in these settings.

Our work leads to some policy implications. In the first place, the potentially negative associations of early father absence and children's outcomes make it logically desirable to avoid this situation. However, going further than implementing information campaigns with the consequences of this event targeted at populations with a higher propensity of experiencing this situation might be neither realistic nor ethical. Therefore, it might be more advisable to focus efforts on reactive policies towards early father absence. In this sense, supporting economically vulnerable mothers throughout their path of raising their children seems fundamental, especially during early childhood. This does not mean lecturing them about the "right" ways of childcaring and parenting, but instead understanding their obstacles and helping them to overcome those hurdles. In this sense, if financial hardship is the principal cause of their difficulties, then policies should focus on alleviating such burdens. Cash transfers are a seemingly straightforward tool and are already being implemented in some countries. Additionally, mothers should have access to support systems when participating in the labour market. For instance, if their children are not yet in school age, they should have access to day-care centres; and when their kids reach school age, they should receive support with their transportation before and after school. Also, mothers should have the possibility to participate in job schemes that give them adequate flexibility to combine their work with their responsibilities as single mothers.

Children at risk should also be the target of remedial policies, and early father absence should be incorporated as another criterium when assessing and monitoring child vulnerability. Building a school-based support system for them could be advisable. Schools could offer them reinforcement programmes after classes while also closely monitoring their evolution and establish a closer communication with their caregivers.

Finally, this research has several limitations that are important to consider. The most important one is related to the extent whereby our results reflect the causal effects of early father absence accurately. Even though our methodology may be better calculating unbiased ATTs under strong ignorability than efforts based on OLS regressions, it still relies on the inexistence of unobservable differences. We have shown that the covariates we used to perform the matching and weighting strategies are predictors of early father absence. However, it is difficult to believe that there are no other relevant elements that we are not incorporating, especially considering that

our definition of absence is based on retrospective information. For instance, as mentioned earlier, mothers' participation in the labour market is a relevant predictor that we could not include. In this sense, future studies on this topic should try using datasets with richer baseline information *before* the occurrence of father absence.

Although generating such a database would probably imply great challenges, especially for measuring early absence, some relatively simple details might already add important information. For instance, YLS is missing basic demographic information (age, education, religion, employment, among others) of absent fathers during round 1. Another useful aspect is collecting more information to have a better idea of family structure stability like, for example, the quality of parents' relationship, whether they were cohabitating and how long before the pregnancy and their marital status. Furthermore, undertaking cohort studies that start collecting information from mothers during their early days of pregnancy might be also a helpful way for generating better samples for studying this topic.

Secondly, our definition of father absence has a relevant margin of improvement. On one side, it is retrospective and only covers the six months previous to each round. Therefore, we could not tell what happened in the rest of the period between both rounds. On the other side, it only measures the frequency of father-child interactions in a very general way. Further research should try to use more accurate measurements that take into account the *quality* of those interactions. This is relevant because even when fathers could see their children every day, this does not mean that they engaged with them in ways that would help them to develop better their skills. Additionally, our measures of potential mechanisms could also be vastly improved. For instance, mothers' mental health is a relevant issue that this work could not explore adequately.

In the third place, we experienced issues with statistical power. Keeping everything else equal, the statistical power in continuous outcomes is maximised when the share of treatment and control groups is the same, as this would allow calculating the estimates of both groups with a similar precision (Glennerster & Takavarasha, 2013). However, due to the relatively low occurrence of early father absence, our treatment group represents roughly 15% of our sample. Additionally, statistical power is positively related with sample size. In our case, attrition and incomplete information of

outcomes caused a reduction of 15% of our sample (see [Table 2](#)). These matters have limited our capability of producing more trustworthy associations, especially when we split the sample into halves to explore heterogeneous effects based on children's sex. In these cases, we observed some relatively high between-sex differences that we could not indicate as being strictly different from zero based on the traditional statistical confidence levels. Thus, studying early father absence also requires efforts to define samples that over-represent children going through this situation.

Finally, this work can have limitations for extrapolating its results to the YLS whole sample of interest due to missing values in outcomes and attrition in 16% of the sample of interest. External validity problems are more relevant if children or their mothers self-select to participate or not in the survey. [Table E1](#) shows the results of comparing those belonging to the sample of interest who were included and excluded. Even though there are no relevant differences in most cases, differences in some variables related to the pregnancies make us cautious. Mothers who were excluded were less likely to have participated in at least one antenatal visit and were more likely to have given birth at home and without the presence of any health professional. These aspects might be related to their higher probability of living in a rural setting. Moreover, they also have more difficulties reading in Spanish and come from a background wherein it may not have been the first language. All these elements give some signs of their potential higher economic vulnerabilities. Therefore, the effects on early father absence may have been different (perhaps worse) in this subgroup.

References

- Amato, P. R., & Anthony, C. J. (2014). Estimating the Effects of Parental Divorce and Death With Fixed Effects Models. *Journal of Marriage and Family*, 76(2), 370-386. <https://doi.org/10.1111/jomf.12100>
- Amato, P. R., & Patterson, S. E. (2017). The Intergenerational Transmission of Union Instability in Early Adulthood. *Journal of Marriage and Family*, 79(3), 723-738. <https://doi.org/10.1111/jomf.12384>
- American Psychological Association. Receptive vocabulary. In *APA Dictionary of Psychology*. <https://dictionary.apa.org/receptive-vocabulary>
- American Psychological Association. Self-efficacy. In *APA Dictionary of Psychology*. <https://dictionary.apa.org/self-efficacy>

- American Psychological Association. Self-esteem. In *APA Dictionary of Psychology*. <https://dictionary.apa.org/self-esteem>
- Antman, F. M. (2012). Gender, educational attainment, and the impact of parental migration on children left behind. *Journal of Population Economics*, 25(4), 1187-1214. <https://doi.org/10.1007/s00148-012-0423-y>
- Augurzy, B., & Kluve, J. (2007). Assessing the performance of matching algorithms when selection into treatment is strong. *Journal of Applied Econometrics*, 22(3), 533-557. <https://doi.org/10.1002/jae.919>
- Austin, P. C. (2011). An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. *Multivariate Behavioral Research*, 46(3), 399-424. <https://doi.org/10.1080/00273171.2011.568786>
- Austin, P. C. (2014). A comparison of 12 algorithms for matching on the propensity score. *Statistics in Medicine*, 33(6), 1057-1069. <https://doi.org/10.1002/sim.6004>
- Bai, H. (2011). A comparison of propensity score matching methods for reducing selection bias. *International Journal of Research & Method in Education*, 34(1), 81-107. <https://doi.org/10.1080/1743727X.2011.552338>
- Bai, Y., Zhang, L., Liu, C., Shi, Y., Mo, D., & Rozelle, S. (2018). Effect of Parental Migration on the Academic Performance of Left Behind Children in North Western China. *The Journal of Development Studies*, 54(7), 1154-1170. <https://doi.org/10.1080/00220388.2017.1333108>
- Bandura, A. (1993). Perceived Self-Efficacy in Cognitive Development and Functioning. *Educational Psychologist*, 28(2), 117-148. https://doi.org/10.1207/s15326985ep2802_3
- Beck, A. N., Cooper, C. E., McLanahan, S., & Brooks-Gunn, J. (2010). Partnership Transitions and Maternal Parenting. *Journal of Marriage and Family*, 72(2), 219-233. <https://doi.org/10.1111/j.1741-3737.2010.00695.x>
- Bernal, R., & Keane, M. P. (2011). Child Care Choices and Children's Cognitive Achievement: The Case of Single Mothers. *Journal of Labor Economics*, 29(3), 459-512. <https://doi.org/10.1086/659343>
- Bick, J., & Nelson, C. A. (2016). Early Adverse Experiences and the Developing Brain. *Neuropsychopharmacology*, 41(1), 177-196. <https://doi.org/10.1038/npp.2015.252>
- Bonomo, V. (2017). Brain-based learning theory. *Journal of Education and Human Development*, 6(1), 27-43. <https://doi.org/10.15640/jehd.v6n1a3>
- Boyden, J. (2018a). *Young Lives: an International Study of Childhood Poverty: Round 2, 2006* (6852; Version 3rd Edition). <https://doi.org/10.5255/UKDA-SN-6852-3>

- Boyden, J. (2018b). *Young Lives: an International Study of Childhood Poverty: Round 3, 2009* (6853; Version 3rd Edition) UK Data Service. <https://doi.org/10.5255/UKDA-SN-6853-3>
- Britto, P. R., & Pérez-Escamilla, R. (2013). No second chances? Early critical periods in human development. *Social Science & Medicine*, 97, 238-240. <https://doi.org/10.1016/j.socscimed.2013.09.001>
- Broussard, C. A. (2010). Research Regarding Low-Income Single Mothers' Mental and Physical Health: A Decade in Review. *Journal of Poverty*, 14(4), 443-451. <https://doi.org/10.1080/10875549.2010.518003>
- Brown, J. D. (2010). High self-esteem buffers negative feedback: Once more with feeling. *Cognition and Emotion*, 24(8), 1389-1404. <https://doi.org/10.1080/02699930903504405>
- Bucheli, M., & Vigorito, A. (2021). *Short-and Medium-term effects of parental separation on children's well-being. Evidence from Uruguay*. Universidad de La República <https://www.colibri.udelar.edu.uy/jspui/bitstream/20.500.12008/28797/1/DT%20E%202021%2007.pdf>
- Burger, A., & Chong, I. (2011). Receptive Vocabulary. In S. Goldstein & J. A. Naglieri (Eds.), *Encyclopedia of Child Behavior and Development* (pp. 1231-1231). Springer US. https://doi.org/10.1007/978-0-387-79061-9_2359
- Busso, M., DiNardo, J., & McCrary, J. (2014). New Evidence on the Finite Sample Properties of Propensity Score Reweighting and Matching Estimators. *The Review of Economics and Statistics*, 96(5), 885-897. https://doi.org/10.1162/REST_a_00431
- Caliendo, M., & Kopeinig, S. (2008). Some practical guidance for the implementation of propensity score matching. *Journal of economic surveys*, 22(1), 31-72.
- Caplan, L. J., & Schooler, C. (2007). Socioeconomic status and financial coping strategies: The mediating role of perceived control. *Social psychology quarterly*, 70(1), 43-58. <https://doi.org/10.1177/019027250707000106>
- Chamie, J. (2016). 320 Million Children in Single-Parent Families. <http://www.ipsnews.net/2016/10/320-million-children-in-single-parent-families/>
- Cooper, C. E., Osborne, C. A., Beck, A. N., & McLanahan, S. S. (2011). Partnership Instability, School Readiness, and Gender Disparities. *Sociology of education*, 84(3), 246-259. <https://doi.org/10.1177/0038040711402361>
- Crabtree, S., & Kluch, S. (2020, 5 March 2020). *How Many Women Worldwide Are Single Moms?* Gallup. <https://news.gallup.com/poll/286433/women-worldwide-single-moms.aspx>

- Crawford, C., & Cribb, J. (2013). *Reading and maths skills at age 10 and earnings in later life: a brief analysis using the British Cohort Study* (Research Report REP03, Issue.
- Crawford, C., & Cribb, J. (2015). *The link between childhood reading skills and adult outcomes: analysis of a cohort of British children* (IFS Briefing Note BN169, Issue.
- Crosier, T., Butterworth, P., & Rodgers, B. (2007). Mental health problems among single and partnered mothers. *Social Psychiatry and Psychiatric Epidemiology*, 42(1), 6-13. <https://doi.org/10.1007/s00127-006-0125-4>
- Cueto, S., & Leon, J. (2012). *Psychometric Characteristics of Cognitive Development and Achievement Instruments in Round 3 of Young Lives* (Technical Note 25, Issue.
- Cueto, S., Leon, J., Guerrero, G., & Muñoz, I. (2009). *Psychometric characteristics of cognitive development and achievement instruments in Round 2 of Young Lives* (Technical Note 15, Issue.
- Cunha, F., & Heckman, J. (2007). The Technology of Skill Formation. *American Economic Review*, 97(2), 31-47. <https://doi.org/10.1257/aer.97.2.31>
- Cunha, F., & Heckman, J. J. (2008). Formulating, Identifying and Estimating the Technology of Cognitive and Noncognitive Skill Formation. *The Journal of Human Resources*, 43(4), 738-782. <https://doi.org/10.3368/jhr.43.4.738>
- Cunha, F., Heckman, J. J., Lochner, L., & Masterov, D. V. (2006). Interpreting the Evidence on Life Cycle Skill Formation. In E. Hanushek & F. Welch (Eds.), *Handbook of the Economics of Education* (Vol. 1, pp. 697-812). Elsevier. [https://doi.org/https://doi.org/10.1016/S1574-0692\(06\)01012-9](https://doi.org/https://doi.org/10.1016/S1574-0692(06)01012-9)
- Cunha, F., Heckman, J. J., & Schennach, S. M. (2010, 2010/05/01). Estimating the Technology of Cognitive and Noncognitive Skill Formation [<https://doi.org/10.3982/ECTA6551>]. *Econometrica*, 78(3), 883-931. <https://doi.org/https://doi.org/10.3982/ECTA6551>
- Dawes, A. (2020). *Measuring the Development of Cognitive Skills Across Time and Context: Reflections from Young Lives* (Young Lives Insight Reports, Issue.
- de Vaus, D., Gray, M., Qu, L., & Stanton, D. (2017). The economic consequences of divorce in six OECD countries. *Australian Journal of Social Issues*, 52(2), 180-199. <https://doi.org/10.1002/ajs4.13>
- Dearden, K., Crookston, B., Madanat, H., West, J., Penny, M., & Cueto, S. (2013, 2013/01/01). What difference can fathers make? Early paternal absence compromises Peruvian children's growth [<https://doi.org/10.1111/j.1740-8709.2011.00347.x>]. *Maternal & Child Nutrition*, 9(1), 143-154. <https://doi.org/https://doi.org/10.1111/j.1740-8709.2011.00347.x>

- DiNardo, J., & Tobias, J. L. (2001). Nonparametric Density and Regression Estimation. *The Journal of Economic Perspectives*, 15(4), 11-28. <http://www.jstor.org/stable/2696513>
- Dong, N., Stuart, E. A., Lenis, D., & Quynh Nguyen, T. (2020). Using Propensity Score Analysis of Survey Data to Estimate Population Average Treatment Effects: A Case Study Comparing Different Methods. *Evaluation Review*, 44(1), 84-108. <https://doi.org/10.1177/0193841x20938497>
- Dunn, L., Padilla, E., Lugo, D., & Dunn, L. (1986). *Manual del Examinador para el Test de Vocabulario en Imágenes Peabody–Adaptación Hispanoamericana [Examiner’s Manual for the Peabody Picture Vocabulary Test Hispanic–American Adaptation]*.
- Escobal, J., Lanata, C., Madrid, S., Penny, M., Saavedra, J., Suárez, P., Verastegui, H., Villar, E., & Huttly, S. (2003). *Young Lives preliminary country report: Peru*.
- Foster, E. M., & Kalil, A. (2007). Living Arrangements and Children’s Development in Low-Income White, Black, and Latino Families. *Child Development*, 78(6), 1657-1674. <https://doi.org/10.1111/j.1467-8624.2007.01091.x>
- Fox, S. E., Levitt, P., & Nelson, C. A., 3rd. (2010). How the timing and quality of early experiences influence the development of brain architecture. *Child Development*, 81(1), 28-40. <https://doi.org/10.1111/j.1467-8624.2009.01380.x>
- Galdo, J. C., Smith, J., & Black, D. (2008). Bandwidth Selection and the Estimation of Treatment Effects with Unbalanced Data. *Annales D’Économie et de Statistique*(91/92), 189-216. <https://doi.org/10.2307/27917245>
- Georgieff, M. K., Brunette, K. E., & Tran, P. V. (2015). Early life nutrition and neural plasticity. *Development and psychopathology*, 27(2), 411-423. <https://doi.org/10.1017/S0954579415000061>
- Gertler, P. J., Martinez, S., Premand, P., Rawlings, L. B., & Vermeersch, C. M. J. (2016). Matching. In *Impact Evaluation in Practice, Second Edition* (Second ed., pp. 143-158). https://doi.org/10.1596/978-1-4648-0779-4_ch8
- Gilbert, S. J., & Burgess, P. W. (2008). Executive function. *Current Biology*, 18(3), R110-R114. <https://doi.org/10.1016/j.cub.2007.12.014>
- Glennerster, R., & Takavarasha, K. (2013). Statistical Power. In *Running Randomized Evaluations* (pp. 241-297). Princeton University Press. <https://doi.org/10.2307/j.ctt4cqd52.10>
- Gove, A. K., & Wetterberg, A. (2011). The Early Grade Reading Assessment: An Introduction. In A. K. Gove & A. Wetterberg (Eds.), *The early grade reading assessment: Applications and interventions to improve basic literacy* (Vol. BK-0007-1109). RTI Press. <https://doi.org/10.3768/rtipress.2011.bk.0007.1109>

- Gu, X. S., & Rosenbaum, P. R. (1993). Comparison of Multivariate Matching Methods: Structures, Distances, and Algorithms. *Journal of Computational and Graphical Statistics*, 2(4), 405-420. <https://doi.org/10.2307/1390693>
- Hainmueller, J. (2012). Entropy Balancing for Causal Effects: A Multivariate Reweighting Method to Produce Balanced Samples in Observational Studies. *Political Analysis*, 20(1), 25-46. <https://doi.org/10.1093/pan/mpr025>
- Harris, M. A., Wetzell, E., Robins, R. W., Donnellan, M. B., & Trzesniewski, K. H. (2018). The development of global and domain self-esteem from ages 10 to 16 for Mexican-origin youth. *International Journal of Behavioral Development*, 42(1), 4-16. <https://doi.org/10.1177/0165025416679744>
- Heckman, J. J. (2011). The economics of inequality: The value of early childhood education. *American Educator*, 35(1), 31.
- Hemphill, L., & Tivnan, T. (2008). The Importance of Early Vocabulary for Literacy Achievement in High-Poverty Schools. *Journal of Education for Students Placed at Risk (JESPAR)*, 13(4), 426-451. <https://doi.org/10.1080/10824660802427710>
- Herbst, C. M. (2013). The impact of non-parental child care on child development: Evidence from the summer participation "dip". *Journal of Public Economics*, 105, 86-105. <https://doi.org/10.1016/j.jpubeco.2013.06.003>
- Himaz, R. (2013). Impact of Parental Death in Middle Childhood and Adolescence on Child Outcomes. *Journal of African Economies*, 22(3), 463-490. <https://doi.org/10.1093/jae/ejt001>
- Himaz, R. (2020, 2020/08/01/). Sweet are the fruit of adversity? The impact of fathers' death on child non-cognitive outcomes in Ethiopia. *Economics & Human Biology*, 38, 100890. <https://doi.org/https://doi.org/10.1016/j.ehb.2020.100890>
- Hodel, A. S. (2018). Rapid infant prefrontal cortex development and sensitivity to early environmental experience. *Developmental Review*, 48, 113-144. <https://doi.org/10.1016/j.dr.2018.02.003>
- Holland, P. W. (1986). Statistics and causal inference. *Journal of the American statistical Association*, 81(396), 945-960.
- Hope, S., Power, C., & Rodgers, B. (1999). Does financial hardship account for elevated psychological distress in lone mothers? *Social Science & Medicine*, 49(12), 1637-1649. [https://doi.org/10.1016/S0277-9536\(99\)00251-8](https://doi.org/10.1016/S0277-9536(99)00251-8)
- Huber, M., Lechner, M., & Wunsch, C. (2013). The performance of estimators based on the propensity score. *Journal of Econometrics*, 175(1), 1-21. <https://doi.org/10.1016/j.jeconom.2012.11.006>

- Imbens, G. W. (2004). Nonparametric Estimation of Average Treatment Effects Under Exogeneity: A Review. *The Review of Economics and Statistics*, 86(1), 4-29. <https://doi.org/10.1162/003465304323023651>
- Imbens, G. W. (2015). Matching methods in practice: Three examples. *Journal of Human Resources*, 50(2), 373-419. <https://doi.org/10.3368/jhr.50.2.373>
- Intemann, Z., & Katz, E. (2014). Migration and children's schooling and time allocation: Evidence from El Salvador. *International Journal of Educational Development*, 39, 264-274. <https://doi.org/10.1016/j.ijedudev.2014.03.004>
- Jamil, N. (2017). The Impact of Remittances Versus Parental Absence on Children's Wellbeing: Evidence from Rural Punjab. *The Lahore Journal of Economics*, 22(2), 1-37.
- Jann, B. (2017). *kmatch: Stata module for multivariate-distance and propensity-score matching, including entropy balancing, inverse probability weighting, (coarsened) exact matching, and regression adjustment*. In <https://ideas.repec.org/c/boc/bocode/s458346.html>.
- Joffe, M. M., & Rosenbaum, P. R. (1999). Invited commentary: propensity scores. *American journal of epidemiology*, 150(4), 327-333.
- Jones, N., & Huttly, S. (2018). *Young Lives: an International Study of Childhood Poverty: Round 1, 2002*. (5307; Version 6th Edition) UK Data Service. <https://doi.org/10.5255/UKDA-SN-5307-3>
- Keizer, R., Helmerhorst, K. O. W., & van Rijn-van Gelderen, L. (2019). Perceived Quality of the Mother–Adolescent and Father–Adolescent Attachment Relationship and Adolescents' Self-Esteem. *Journal of Youth and Adolescence*, 48(6), 1203-1217. <https://doi.org/10.1007/s10964-019-01007-0>
- King, G., & Nielsen, R. (2019). Why Propensity Scores Should Not Be Used for Matching. *Political Analysis*, 27(4), 435-454. <https://doi.org/10.1017/pan.2019.11>
- Knudsen, E. I. (2004). Sensitive Periods in the Development of the Brain and Behavior. *Journal of Cognitive Neuroscience*, 16(8), 1412-1425. <https://doi.org/10.1162/0898929042304796>
- Knudsen, E. I., Heckman, J. J., Cameron, J. L., & Shonkoff, J. P. (2006). Economic, neurobiological, and behavioral perspectives on building America's future workforce. *Proceedings of the National Academy of Sciences*, 103(27), 10155. <https://doi.org/10.1073/pnas.0600888103>
- Krauss, S., Orth, U., & Robins, R. W. (2020). Family environment and self-esteem development: A longitudinal study from age 10 to 16. *Journal of personality and social psychology*, 119(2), 457. <https://doi.org/10.1037/pspp0000263>

- Lamb, M. E., Pleck, J. H., Charnov, E. L., & Levine, J. A. (1985). Paternal Behavior in Humans. *American Zoologist*, 25(3), 883-894. <http://www.jstor.org/stable/3883043>
- Liu, Y.-L. (2008). An Examination of Three Models of the Relationships between Parental Attachments and Adolescents' Social Functioning and Depressive Symptoms. *Journal of Youth and Adolescence*, 37(8), 941-952. <https://doi.org/10.1007/s10964-006-9147-1>
- Luo, J., Wang, L.-G., & Gao, W.-B. (2012). The influence of the absence of fathers and the timing of separation on anxiety and self-esteem of adolescents: a cross-sectional survey. *Child: Care, Health and Development*, 38(5), 723-731. <https://doi.org/10.1111/j.1365-2214.2011.01304.x>
- Marsh, H. W. (1990). *Self description questionnaire I : SDQ I : manual*. University of Western Sydney, Macarthur.
- Meng, X., & Yamauchi, C. (2017). Children of Migrants: The Cumulative Impact of Parental Migration on Children's Education and Health Outcomes in China. *Demography*, 54(5), 1677-1714. <https://doi.org/10.1007/s13524-017-0613-z>
- Mirowsky, J., & Ross, Catherine E. R. (2007). Life Course Trajectories of Perceived Control and Their Relationship to Education. *American Journal of Sociology*, 112(5), 1339-1382. <https://doi.org/10.1086/511800>
- Mitchell, C., McLanahan, S., Hobcraft, J., Brooks-Gunn, J., Garfinkel, I., & Notterman, D. (2015). Family Structure Instability, Genetic Sensitivity, and Child Well-Being. *American Journal of Sociology*, 120(4), 1195-1225. <https://doi.org/10.1086/680681>
- Mooney, A., Oliver, C., & Smith, M. (2009). *Impact of family breakdown on children's well-being: Evidence review*. <https://core.ac.uk/download/pdf/4160412.pdf>
- Mortelmans, D. (2020). Economic Consequences of Divorce: A Review. In M. Kreyenfeld & H. Trappe (Eds.), *Parental Life Courses after Separation and Divorce in Europe* (Vol. 12, pp. 23-41). Springer Nature. <https://doi.org/10.1007/978-3-030-44575-1>
- Mortelmans, D., & Jansen, M. (2010). Income dynamics after divorce: A latent growth model of income change on the European Community Household Panel. *Brussels economic review*, 53(1), 85-107.
- Muter, V., Hulme, C., Snowling, M. J., & Stevenson, J. (2004). Phonemes, rimes, vocabulary, and grammatical skills as foundations of early reading development: evidence from a longitudinal study. *Developmental psychology*, 40(5), 665-681. <https://doi.org/10.1037/0012-1649.40.5.665>
- National Research Council. (2000). The Developing Brain. In *From Neurons to Neighborhoods: The Science of Early Childhood Development* (pp. 182-218). The National Academies Press. <https://doi.org/10.17226/9824>

- National Scientific Council on the Developing Child. (2007). *The Timing and Quality of Early Experiences Combine to Shape Brain Architecture*. www.developingchild.harvard.edu
- Nguyen, C. V. (2016, 2016/03/01/). Does parental migration really benefit left-behind children? Comparative evidence from Ethiopia, India, Peru and Vietnam. *Social Science & Medicine*, 153, 230-239. <https://doi.org/https://doi.org/10.1016/j.socscimed.2016.02.021>
- Nguyen, C. V., & Vu, L. (2016). Should Parents Work Away from or Close to Home? The Effect of Parental Absence on Children's Time Use in Vietnam. *Journal of Human Development and Capabilities*, 17(1), 110-124. <https://doi.org/10.1080/19452829.2015.1103711>
- Oakhill, J. V., & Cain, K. (2012). The Precursors of Reading Ability in Young Readers: Evidence From a Four-Year Longitudinal Study. *Scientific Studies of Reading*, 16(2), 91-121. <https://doi.org/10.1080/10888438.2010.529219>
- OECD. (2016). *OECD Family Database-SF1.1: Family size and household composition*. OECD - Social Policy Division - Directorate of Employment, Labour and Social Affairs. Retrieved April 12 2021 from <https://www.oecd.org/els/family/database.htm>
- OECD. (2018). *OECD Family Database-SF1.3: Further information on the living arrangements of children*. OECD - Social Policy Division - Directorate of Employment, Labour and Social Affairs. Retrieved April 12 2021 from <https://www.oecd.org/els/family/database.htm>
- Ogando, M. J., & Yorke, L. (2018). *Psychosocial scales in the Young Lives round 4 survey: Selection, adaptation and validation* (Technical Note 45, Issue).
- Orth, U. (2018). The family environment in early childhood has a long-term effect on self-esteem: A longitudinal study from birth to age 27 years. *Journal of personality and social psychology*, 114(4), 637-655. <https://doi.org/10.1037/pspp0000143>
- Orth, U., & Robins, R. W. (2014). The Development of Self-Esteem. *Current Directions in Psychological Science*, 23(5), 381-387. <https://doi.org/10.1177/0963721414547414>
- Park, L. E., & Crocker, J. (2008). Contingencies of self-worth and responses to negative interpersonal feedback. *Self and Identity*, 7(2), 184-203. <https://doi.org/10.1080/15298860701398808>
- Peruvian National Institute of Statistics. (2019). *Peru: Household Types and Life Cycles*.

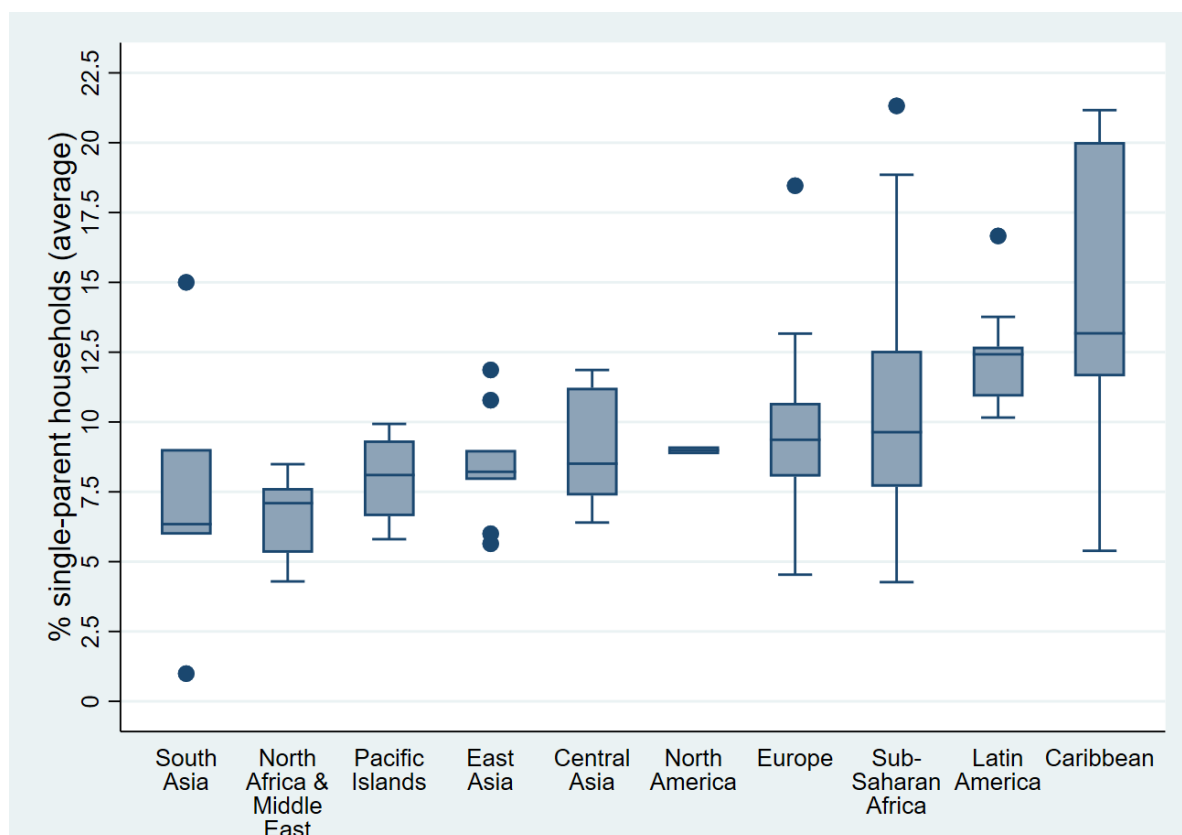
- Pleck, J. H. (2010). Paternal involvement: revised conceptualization and theoretical linkages with child outcomes. In M. E. Lamb (Ed.), *The Role of the Father in Child Development* (Fifth ed., pp. 58-93). John Wiley & Sons, Inc.
- Reynolds, S., Fernald, L., Deardorff, J., & Behrman, J. (2018). Family structure and child development in Chile: A longitudinal analysis of household transitions involving fathers and grandparents. *Demographic Research*, 38(58), 1777-1814. <https://doi.org/10.4054/DemRes.2018.38.58>
- Ribar, D. (2004). *What Do Social Scientists Know About the Benefits of Marriage? A Review of Quantitative Methodologies* (IZA Discussion Paper No. 998, Issue. <https://ssrn.com/abstract=500887>
- Ripollone, J. E., Huybrechts, K. F., Rothman, K. J., Ferguson, R. E., & Franklin, J. M. (2018). Implications of the Propensity Score Matching Paradox in Pharmacoepidemiology. *American journal of epidemiology*, 187(9), 1951-1961. <https://doi.org/10.1093/aje/kwy078>
- Ritchie, S. J., & Bates, T. C. (2013). Enduring Links From Childhood Mathematics and Reading Achievement to Adult Socioeconomic Status. *Psychological Science*, 24(7), 1301-1308. <https://doi.org/10.1177/0956797612466268>
- Rosenbaum, P. R., & Rubin, D. B. (1983). The Central Role of the Propensity Score in Observational Studies for Causal Effects. *Biometrika*, 70(1), 41-55. <https://doi.org/10.2307/2335942>
- Rotter, J. B. (1966). Generalized expectancies for internal versus external control of reinforcement. *Psychological monographs: General and applied*, 80(1), 1-28.
- Rousou, E., Kouta, C., Middleton, N., & Karanikola, M. (2013). Single mothers' self-assessment of health: a systematic exploration of the literature. *International Nursing Review*, 60(4), 425-434. <https://doi.org/10.1111/inr.12044>
- Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of educational Psychology*, 66(5), 688.
- Rubin, D. B. (2008). For objective causal inference, design trumps analysis. *The Annals of Applied Statistics*, 2(3), 808-840, 833. <https://doi.org/10.1214/08-AOAS187>
- Sanchez, A., Penny, M., Woldehanna, T., Galab, S., Boyden, J., & Duc, L. T. (2018). *Young Lives: an International Study of Childhood Poverty: Round 5, 2016* (8357; UK Data Service. <https://doi.org/10.5255/UKDA-SN-8357-1>
- Sarma, V. J., & Parinduri, R. A. (2016). What happens to children's education when their parents emigrate? Evidence from Sri Lanka. *International Journal of Educational Development*, 46, 94-102. <https://doi.org/10.1016/j.ijedudev.2015.11.007>

- Scarborough, H. S. (2009). Connecting early language and literacy to later reading (dis) abilities: Evidence, theory, and practice. In F. Fletcher-Campbell, J. Soler, & G. Reid (Eds.), *Approaching difficulties in literacy development: Assessment, pedagogy and programmes* (Vol. 10, pp. 23-38). SAGE.
- Schwarzenberg, S. J., Georgieff, M. K., & Committee on Nutrition. (2018). Advocacy for Improving Nutrition in the First 1000 Days to Support Childhood Development and Adult Health. *Pediatrics*, 141(2), e20173716. <https://doi.org/10.1542/peds.2017-3716>
- Schwarzer, R., & Jerusalem, M. (1995). Generalized self-efficacy scale. *Measures in health psychology: A user's portfolio. Causal and control beliefs*, 1(1), 35-37.
- Sigle-Rushton, W., & McLanahan, S. (2004). Father absence and child well-being: A critical review. In D. P. Moynihan, T. M. Smeeding, & L. Rainwater (Eds.), *The future of the family* (pp. 116-158). The Russell Sage Foundation.
- Silva, M., & Cain, K. (2015). The relations between lower and higher level comprehension skills and their role in prediction of early reading comprehension. *Journal of educational Psychology*, 107(2), 321-331. <https://doi.org/10.1037/a0037769>
- Smith, J. A., & Todd, P. E. (2005). Does matching overcome LaLonde's critique of nonexperimental estimators? *Journal of Econometrics*, 125(1), 305-353. <https://doi.org/10.1016/j.jeconom.2004.04.011>
- Stack, R. J., & Meredith, A. (2018). The Impact of Financial Hardship on Single Parents: An Exploration of the Journey From Social Distress to Seeking Help. *Journal of Family and Economic Issues*, 39(2), 233-242. <https://doi.org/10.1007/s10834-017-9551-6>
- Suggate, S., Schaughency, E., McAnally, H., & Reese, E. (2018). From infancy to adolescence: The longitudinal links between vocabulary, early literacy skills, oral narrative, and reading comprehension. *Cognitive Development*, 47, 82-95. <https://doi.org/10.1016/j.cogdev.2018.04.005>
- Thompson, R. A., & Nelson, C. A. (2001). Developmental science and the media: Early brain development. *American Psychologist*, 56(1), 5.
- UNICEF. (2018). *Early Childhood Development in the UNICEF Strategic Plan 2018-2021*. <https://www.unicef.org/sites/default/files/2019-05/Early%20Childhood%20Development%20in%20the%20UNICEF%20Strategic%20Plan%202018-2021.pdf>
- United Nations. (2019). *Database on Household Size and Composition 2019*.
- Vera Rueda, J. (2021). *When fathers are gone: the consequences of paternal absence during the early years (2020-67)*. Paris School of Economics. <https://halshs.archives-ouvertes.fr/halshs-02978563>

- Verhoeven, L., & van Leeuwe, J. (2008). Prediction of the development of reading comprehension: a longitudinal study. *Applied Cognitive Psychology*, 22(3), 407-423. <https://doi.org/10.1002/acp.1414>
- Vikram, K. (2021). Fathers' Migration and Academic Achievement among Left-behind Children in India: Evidence of Continuity and Change in Gender Preferences. *International Migration Review*, 0197918321989279. <https://doi.org/10.1177/0197918321989279>
- Woldehanna, T., Galab, S., Sanchez, A., Penny, M., Duc, L. T., & Boyden, J. (2018). *Young Lives: an International Study of Childhood Poverty: Round 4, 2013-2014* (7931; Version 2nd Edition) UK Data Service. <https://doi.org/10.5255/UKDA-SN-7931-2>
- Yang, Z., & Schaninger, C. M. (2010). The Impact of Parenting Strategies on Child Smoking Behavior: The Role of Child Self-Esteem Trajectory. *Journal of Public Policy & Marketing*, 29(2), 232-247. <https://doi.org/10.1509/jppm.29.2.232>
- Zeigler-Hill, V. (2013). The importance of self-esteem. In V. Zeigler-Hill (Ed.), *Self-esteem* (pp. 1-20). Psychology Press.
- Zhang, J., Appleton, S., Song, L., & Liu, B. (2021). Who Looks after the Kids? The Effects of Childcare Choice on Early Childhood Development in China*. *Oxford Bulletin of Economics and Statistics*, 83(3), 619-640. <https://doi.org/10.1111/obes.12410>
- Zhao, Z. (2004). Using Matching to Estimate Treatment Effects: Data Requirements, Matching Metrics, and Monte Carlo Evidence. *The Review of Economics and Statistics*, 86(1), 91-107. <http://www.jstor.org/stable/3211662>
- Zhao, Z. (2008). Sensitivity of propensity score methods to the specifications. *Economics Letters*, 98(3), 309-319. <https://doi.org/10.1016/j.econlet.2007.05.010>

Appendix A

Figure A1. Distribution of the ratio of single-parent and biparental households by regions



Source: own elaboration using data from United Nations (2019). This plot shows the ratios of single-parent and biparental households instead of the shares of single-parent households to consider that households in every country do not have the same propensity of having children.

Appendix B

Table B1. Non-cognitive outcomes, instruments, and items

Outcome	Instrument	Items
Self-esteem	SDQ-I	1. I do lots of important things
		2. In general, I like being the way I am
		3. Overall, I have a lot to be proud of
		4. I can do things as well as most people
		5. Other people think I am a good person
		6. A lot of things about me are good
		7. I'm as good as most other people
		8. When I do something, I do it well
Self-efficacy	GSES	1. I can always manage to solve difficult problems if I try hard enough.
		2. If someone opposes me, I can find the means and ways to get what I want.
		3. It is easy for me to stick to my aims and accomplish my goals.
		4. I am confident that I could deal efficiently with unexpected events.
		5. Thanks to my resourcefulness, I know how to handle unforeseen situations.
		6. I can solve most problems if I invest the necessary effort.
		7. I can remain calm when facing difficulties because I can rely on my coping abilities.
		8. When I am confronted with a problem, I can usually find several solutions.
		9. If I am in trouble, I can usually think of a solution.
		10. I can usually handle whatever comes my way.
Agency		1. I have no choice about the work I do
		2. If I study hard, I will be rewarded with a better job in the future
		3. I like to make plans for my future studies and work
		4. Other people in my family make all the decisions about how I spend my time
		5. If I try hard, I can improve my situation in life

Source: own elaboration based on the YLS surveys

Appendix C

Matching and weighting methods

1. Description of the alternatives:

Implementing matching requires taking three main decisions: choosing the metric, algorithm, and its parameters (Caliendo & Kopeinig, 2008). Similarly, there are several ways of weighting.

Matching metrics:

Selection into treatment is usually affected by several factors in non-randomised evaluations. Depending on the depth and quality of the available data, some might be observable, and some might not. Considering the unconfoundedness assumption, we would like to include in the matching strategy as many relevant elements as possible. However, the larger the number of covariates, the more difficult it becomes to find observations that share the same values, which is a frequent issue called the "curse of dimensionality" (Gertler et al., 2016). To overcome this problem, researchers perform matching regularly using measures that allow combining all relevant covariates in a single variable.

In this case, we tested the two aggregation metrics that have been most widely utilised in the academic literature:

- a. Propensity score (PS): $PS_i = P(W_i = 1|X_i)$

The PS is the individual's probability of being selected into the treatment group ($W_i = 1$) based on their characteristics (X_i). Rosenbaum and Rubin (1983) proposed this method as a way to solve the problem of dimensionality, proving that if the set of covariates that affect this probability allow keeping the unconfoundedness assumption, then the PS should also preserve this condition. Matching using this probability is referred to as *propensity score matching* (PSM).

Despite its wide acceptability, some aspects about this metric need to be taken into account. Even though the *true* PS should theoretically allow achieving strong ignorability (Austin, 2011), it is seldom known in observational studies because

researchers usually ignore the assignment mechanism. In our case, we ignore why some children ended experiencing father absence while others did not. Thus, this probability needs to be *estimated* using non-linear regression methods such as logit or probit. This can lead results based on PSM to be sensitive to misspecifications of the PS (Smith & Todd, 2005; Zhao, 2008). Additionally, even in cases when the PS is correctly specified, it might not do much to eliminate the bias due to unobservable characteristics (Joffe & Rosenbaum, 1999). However, this is something shared by every matching metric.

Besides, King and Nielsen (2019) argue, based on simulations and exercises with real world data, that PSM leads to suboptimal results in terms of efficiency and bias reduction and is more prone to model dependence in comparison to matching using metrics not estimated through regression methods. According to these authors, the loss of information that comes with the reduction of dimensionality achieved by the PS estimation impedes generating better matches. As a result, while PSM approximates a *completely randomised experimental design*, using other metrics allow getting closer to a *fully blocked randomised experimental design*. The results of this work showed that despite PSM seeming to perform the task of pruning the observations with greater dissimilarities adequately, it also appears to fail on keeping the best matches if the researcher goes further with the trimming aiming for a higher bias reduction. The authors called this issue "the propensity score paradox".

The results of this work warn against using the PSM in the traditional ways academics have been employing it and advocate instead for using other metrics. Nevertheless, much more empirical evidence using real data is needed to establish any conclusion. In fact, the findings from Ripollone et al. (2018) based on pharmacoepidemiologic data sets indicate that this paradox might be less problematic in some applications.

b. Mahalanobis distance (MD):
$$D(X_t, X_c) = \sqrt{(X_t - X_c)^T * S^{-1} * (X_t - X_c)}$$

It is a similarity metric constructed using the multivariate *distance* between two observations, a treated unit t and a control unit c , based on the values of covariates

represented by the matrices X and the *correlations* based on the sample covariance matrix S .

In contrast to the PS, the MD is a metric calculated directly from the covariates without using regression methods. According to King and Nielsen (2019), its calculation loses less information from the covariates, so using it for matching permits to retrieve a fully blocked randomised experimental design from observational data. Thus, it should yield better results than PSM. This is also supported by the results of Zhao (2004) using small samples. Nevertheless, other studies have pointed out that the MD performs worse when facing a large set of covariates (Gu & Rosenbaum, 1993).

Matching algorithms

Once the metric has been defined and calculated, there are several different ways of performing the matching. This work tested the algorithms more frequently used in the academic literature, namely those described in Caliendo and Kopeinig (2008). Some other techniques do not even a predefined matching metric but instead use the information of the relevant covariates. We also evaluated one of these alternatives: entropy balancing (Hainmueller, 2012).

Now, we will describe the algorithms or methods that were tested, beginning with those based on matching metrics. As well, when applicable, we will also describe the relevant parameters briefly within each approach.

a. Nearest neighbours (NN):

Perhaps the most popular matching algorithm, NN consists in matching each treated units with the k closest control units based on the selected metric, where k is a number between 1 and N_c (size of the control group), and discarding the rest of the controls that are not matched. There are two relevant decisions regarding this method: choosing the number of nearest controls units (parameter k) and allowing each control unit the possibility of being matched to more than one treatment unit (*NN with replacement*) or not (*NN without replacement*). Regarding this second decision, assuming a one-to-one matching (k equals 1), NN without replacement will generally work the following way: it would start with a given treated

unit t_1 and match it with its closest neighbour, a control unit c_1 . Afterwards, this control unit would be dropped from the pool of potential matches for other treated units. Then, the algorithm will continue with another treated unit t_2 and so on. However, it may come to a point where the closest match within the sample for another treated unit, t_n , will not be available because it was already matched (for instance c_1). Instead, t_n will be matched to another control unit c_n as the nearest available neighbour. On the contrary, in NN with replacement, c_1 would continue to be available as an option even after having been matched to t_1 , and thus, would also be the match of t_n .

Following Caliendo and Kopeinig (2008), both decisions pose a trade-off between bias and variance. Regarding the number of nearest controls units, matching each treated unit to more than one control ($k > 1$) would allow the use of more information and potentially include more control units in the analysis, reducing the variance of the estimations. However, this will also increase the bias because each treated unit will not only be compared with the control unit that is, at least in theory, more similar, but also with others that share less similarities. Moving to the decision of allowing or not replacements, NN without replacement increases the bias by not guaranteeing each treated unit to be matched with its nearest neighbour within the sample (sometimes they would be compared with the closest available neighbour). Nevertheless, it also decreases variance by making it possible to include some control units that would have been dropped otherwise for not being the nearest neighbours of any treated unit. Following this logic, NN with replacement allows, contrarily, to decrease bias but increase variance.

b. Kernel (KM):

This approach compares each treated unit with several controls, which are reweighted based on their distance so that those who are closer to the treated unit end with a higher weight and those who are further with a lower weight. Therefore, each control unit will have several different weights depending on the treated unit it is being compared to. In comparison to NN, KM may usually increase the bias of an estimator but decrease its variance (Caliendo & Kopeinig, 2008).

However, it can also allow avoiding higher bias levels by imposing a maximum acceptable distance between a treated and a control unit (commonly referred to as a *bandwidth*). The choice of the bandwidth can play a crucial role on the quality of the matches. Following the academic practices that seek to prevent both selection and researcher biases (Rubin, 2008), we will follow the cross-validation procedure defined by Jann (2017), which seeks for a value that locally minimises the mean squared error of the PS or the means of the covariates in the case of MD and includes a penalty if this optimal caliper is larger is excessively large (greater than the standard deviation of the PS or the square-root of the number of covariates). Another parameter related to this method is the *kernel function*, which will determine how weights are assigned to the control units. DiNardo and Tobias (2001) suggest that there are no relevant differences among the variety of functions that calculate weights based on the distance (which are called "non-rectangular functions"), which is why this decision does not seem to be crucial.

Weighting methods

c. Inverse probability weighting (IPW):

This alternative consists in weighting each observation based on its PS, so that treated units with low PS and control unit with high PS receive a greater weight in the estimations (Imbens, 2004). Even though IPW is not strictly speaking a matching method, it seems to perform similarly in terms of bias reduction (Busso et al., 2014).

d. Entropy balancing (EB):

This technique was introduced in Hainmueller (2012). It consists in assigning weights to each observation to eliminate any differences in the distributions of each covariate between the treatment and control groups. This is achieved by choosing by minimising the entropy distance metric (EDM) of the weights while also considering a series of balancing and normalising constraints. Formally, the EDM is calculated using the following formula:

$$H(w) = \sum_{\{i|D=0\}} w_i * \log (w_i/q_i)$$

Where w_i represents the final weights and q_i the base weights. Thus, EDM represents a measure of the distance between the final and base weights, being these second values those that give every observation the same weight. In other words: $q_i = \frac{1}{N}$, where N is the sample size. Minimising this metric is relevant so that the final weights end up being as close as possible to the base weights, thereby reducing the loss of information.

A relevant aspect of this approach is its flexibility to set the constraints wherein the EDM minimisation occurs, especially those related to balancing the covariates. In this sense, apart from calculating weights that allow balancing between-group differences in the mean values of each covariate, EB goes further and also permits to find weights that also balance differences in variance and skewness, and also the covariances between the covariates. Therefore, the weights generated by EB should outperform normally the levels of balance achieved by techniques based on the PS or MD. Nevertheless, depending on amount and complexity of the defined constraints and the amount of data, sometimes the algorithm might not find a set of weights that satisfy all the restrictions.

2. Propensity score calculation:

Table C1. Propensity score calculation

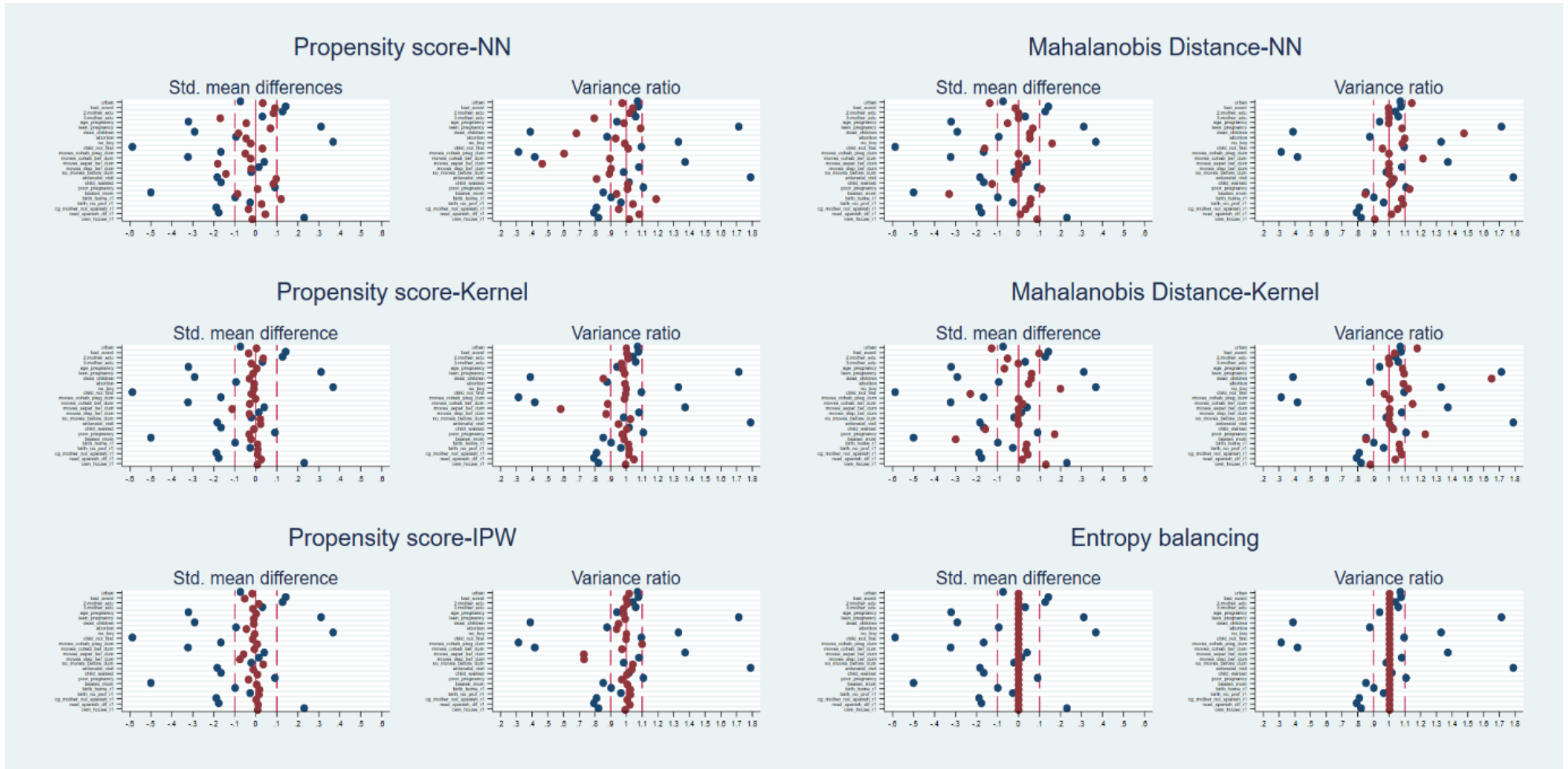
	(1) Early father absence
Urban setting	0.482*** (0.1243)
Negative economic shocks	1.340 (0.2437)
Mother education (<i>ref. less than secondary</i>)	
Secondary completed	1.105 (0.2753)
At least some college	0.865 (0.2803)
YLS pregnancy: mum age	1.017 (0.0200)
YLS-pregnancy: teen	1.328 (0.3564)
Experienced death of a child	0.689 (0.2673)
Experienced an abortion	0.742 (0.1762)
YLS-parents have no boys	1.265 (0.2496)
YLS-not first child	0.304*** (0.0764)
Cohabitation during YLS-pregnancy	0.0450*** (0.0476)
Cohabitation before YLS-pregnancy	0.304*** (0.1140)
Separation before YLS-pregnancy	0.468 (0.3942)
Family dispute before YLS-pregnancy	1.573 (0.8518)
Never moved before YLS-pregnancy	0.655** (0.1391)
Antenatal visit during YLS-pregnancy	0.455** (0.1507)
YLS-pregnancy was wanted	0.565*** (0.0995)

YLS-pregnancy was difficult	1.279 (0.2633)
Mum beaten during childhood	0.280*** (0.0539)
Gave birth at home	0.625 (0.2429)
No health professionals in birth	0.961 (0.3782)
Grandmother's first language not Spanish	0.646 (0.2191)
Mother has difficulties reading in Spanish	0.807 (0.2185)
Own house	1.677*** (0.3334)
Constant	3.919* (2.7416)
Observations	1,360
Pseudo R^2	0.20

Source: own elaboration. Coefficients expressed in odds-ratios;
Standard errors in parentheses. * $p < 0.1$, ** $p < .05$, *** $p < 0.01$.
This regression included sentinel site fixed effects

3. Complementary plots:

Figure C1. Comparison of matching and weighting alternatives



Source: own elaboration with data from YLS. Blue dots correspond to raw (before matching or weighting) values and red dots correspond to adjusted (after matching or weighting) values.

Results: complementary analysis

1. Results of the sensitivity analysis

Table D1. Cognitive outcomes

	Excluding large weights ^A			All weights		
	(1) PPVT	(2) EGRA/ Reading	(3) CDA/ Math	(4) PPVT	(5) EGRA/ Reading	(6) CDA/ Math
Round 2 (4-6 years)	-0.099 (0.0731)		-0.20* (0.1023)	-0.043 (0.0891)	-	-0.13 (0.1076)
Round 3 (7-8 years)	-0.13** (0.0639)	0.026 (0.0965)	-0.076 (0.0796)	-0.070 (0.0772)	-0.028 (0.0856)	-0.089 (0.0648)
Round 4 (11-12 years)	-0.11 (0.0728)	0.0018 (0.0944)	-0.18** (0.0716)	-0.031 (0.0811)	0.013 (0.0770)	-0.15** (0.0623)
Round 5 (14-15 years)	-0.066 (0.0735)	-0.18* (0.0879)	-0.15** (0.0570)	-0.024 (0.0725)	-0.19** (0.0800)	-0.13** (0.0513)
Observations	1,335	1,335	1,335	1,360	1,360	1,360

Source: own elaboration. Clustered standard errors (by sentinel site) in parentheses * $p < 0.1$, ** $p < .05$, *** $p < 0.01$. All estimates are expressed in effect sizes based on the control group's mean and standard deviation. Empty cells ("–") indicate that the outcome was not measured at that round.

^A Regressions exclude all control units with weights larger than 1.

Table D2. Non-cognitive outcomes

	Excluding large weights ^A			All weights		
	(1) Agency	(2) Self- esteem	(3) Self- efficacy	(4) Agency	(5) Self- esteem	(6) Self- efficacy
Round 3 (7-8 years)	0.049 (0.0510)			-0.004 (0.0590)	-	-
Round 4 (11-12 years)	-0.056 (0.0929)	-0.19* (0.0944)	-0.047 (0.0922)	0.020 (0.1112)	-0.100 (0.0880)	-0.085 (0.1053)
Round 5 (14-15 years)	0.061 (0.0797)	-0.25** (0.0864)	-0.011 (0.0934)	0.071 (0.0808)	-0.24** (0.1106)	-0.031 (0.1120)
Observations	1,335	1,335	1,335	1,360	1,360	1,360

Source: own elaboration. Clustered standard errors (by sentinel site) in parentheses * $p < 0.1$, ** $p < .05$, *** $p < 0.01$. All estimates are expressed in effect sizes based on the control group's mean and standard deviation. Empty cells ("–") indicate that the outcome was not measured at that round.

^A Regressions exclude all control units with weights larger than 1.

Appendix E

Limitations

Table E1. Sample of interest: comparing complete and missing observations

Covariates	Included (I)	Excluded (E)	Difference (I-E)
Household setting: urban	0.00	-0.21	0.21***
External bad event index	0.00	-0.07	0.07
Mum education level	0.00	-0.14	0.14**
YLS pregnancy: mum age	0.00	0.01	-0.01
YLS-pregnancy: teen	0.00	-0.04	0.04
Experienced death of a child	0.00	0.11	-0.11
Experienced abortion/miscarriage	0.00	-0.11	0.11
YLS-parents have no boys	0.00	-0.01	0.01
YLS-child not first child	0.00	0.07	-0.07
YLS-pregnancy was wanted	0.00	-0.02	0.02
Cohabitation during YLS-pregnancy	0.00	0.08	-0.08
Cohabitation before YLS-pregnancy	0.00	-0.05	0.05
Separation before YLS-pregnancy	0.00	0.03	-0.03
Family dispute before YLS-pregnancy	0.00	-0.02	0.02
Never moved before YLS-pregnancy	0.00	-0.16	0.16**
Some antenatal visit	0.00	-0.03	0.03
YLS-pregnancy was difficult	0.00	0.06	-0.06
Mum beaten during childhood	0.00	-0.10	0.10
Gave birth at home	0.00	0.23	-0.23***
No health professionals at birth	0.00	0.27	-0.27***
Grandmother's first language is not Spanish	0.00	0.21	-0.21***
Mother has difficulties reading in Spanish	0.00	0.16	-0.16**
Own house	-0.00	-0.00	0.00
Observations	1,360	257	1,617

Source: own elaboration. * $p < 0.1$, ** $p < .05$, *** $p < 0.01$. All estimates are expressed in effect sizes based on the excluded subgroup's mean and standard deviation.

Appendix F

Ethics approval

Anyone conducting research under the auspices of the Institute (staff, students or visitors) where the research involves human participants or the use of data collected from human participants, is required to gain ethical approval before starting. This includes preliminary and pilot studies. Please answer all relevant questions in terms that can be understood by a lay person and note that your form may be returned if incomplete.

For further support and guidance please see accompanying guidelines and the Ethics Review Procedures for Student Research or contact your supervisor.

Before completing this form you will need to discuss your proposal fully with your supervisor(s).

Please attach all supporting documents and letters.

For all Psychology students, this form should be completed with reference to the British Psychological Society (BPS) Code of Human Research Ethics and Code of Ethics and Conduct.

Section 1: Project Details

- a) Project title: **Does being there at the beginning matter? Studying the effects of father absence during the first year of life on outcomes during childhood and adolescence**
- b) Student name: **Jhony Pulido**
- c) Supervisor/Personal Tutor: **Rafael Novella**
- d) Department: **Social Science**
- e) Course category (tick one):
 - PhD/MPhil ☐
 - EdD ☐
 - MRes ☐
 - DEdPsy ☐
 - MTeach ☐
 - MA/MSc ☒
 - ITE ☐
 - Diploma (state which) ☐
 - Other (state which) ☐
- f) Course/module title: **Dissertation**
- g) If applicable, state who the funder is and if funding has been confirmed:

- h) Intended research start date: [31-01-2021](#)
i) Intended research end date: [03-09-2021](#)
j) Country fieldwork will be conducted in: [No fieldwork is required](#)

If research to be conducted abroad please check the [Foreign and Commonwealth Office \(FCO\)](#) and submit a completed travel risk assessment form (see guidelines). If the FCO advice is against travel this will be required before ethical approval can be granted: [UCL travel advice webpage](#)

- k) Has this project been considered by another (external) Research Ethics Committee?

Yes ☐

External Committee Name:

Date of Approval:

No ☒ **go to Section 2**

If yes:

- Submit a copy of the approval letter with this application.
- Proceed to Section 10 Attachments.

Note: Ensure that you check the guidelines carefully as research with some participants will require ethical approval from a different ethics committee such as the [National Research Ethics Service](#) (NRES) or [Social Care Research Ethics Committee](#) (SCREC). In addition, if your research is based in another institution then you may be required to apply to their research ethics committee.

Section 2 - Research methods summary (tick all that apply)

- ☐ Interviews
- ☐ Focus Groups
- ☐ Questionnaires
- ☐ Action Research
- ☐ Observation
- ☐ Literature Review
- ☐ Controlled trial/other intervention study
- ☐ Use of personal records
- ☐ Systematic review – **if only method used go to Section 5**
- ☒ Secondary data analysis – **if secondary analysis used go to Section 6**
- ☐ Advisory/consultation/collaborative groups
- ☐ Other, give details:

Please provide an overview of the project, focusing on your methodology. This should include some or all of the following: purpose of the research, aims, main research questions, research design, participants, sampling, data collection (including justifications for methods chosen and description of topics/questions to be asked), reporting and dissemination. Please focus on your methodology; the theory, policy, or literary background of your work can be provided in an attached document (i.e. a full research proposal or case for support document). *Minimum 150 words required.*

This information is all detailed in the research proposal that will also be attached with this form

Section 3 – research Participants (tick all that apply)

a) Will your research involve human participants?

Yes ☐

No ☐ (if 'No', go to Section 4)

b) Who are the participants for this project (i.e. what sorts of people will be involved)? Tick all that apply

☐ Early years/pre-school

☐ Ages 5-11

☐ Ages 12-16

☐ Young people aged 17-18

☐ Adults please specify below

☐ Unknown – specify below

☐ No participants

Enter text

Note: Ensure that you check the guidelines carefully as research with some participants will require ethical approval from a different ethics committee such as the [National Research Ethics Service](#) (NRES) or [Social Care Research Ethics Committee](#) (SCREC).

c) If participants are under the responsibility of others (such as parents, teachers or medical staff) how do you intend to obtain permission to approach the participants to take part in the study? (*Please attach approach letters or details of permission procedures - see Section 9*)

Enter text

d) How will participants be recruited (identified and approached)?

Enter text

e) Describe the process you will use to inform participants about what you are doing

Enter text

- f) How will you obtain the consent of participants? Will this be written? How will it be made clear to participants that they may withdraw consent to participate at any time?

Enter text

- g) **Studies involving questionnaires:** will participants be given the option of omitting questions they do not wish to answer?

Yes ☐

No* ☐

***If no**, please explain why below, and ensure that you cover any ethical issues arising from this in Section 8

Enter text

- h) **Studies involving observation:** please confirm whether participants will be asked for their informed consent to be observed

Yes ☐

No* ☐

***If no**, read the guidelines (Ethical Issues section) and explain why below and ensure that you cover any ethical issues arising from this in section 8

Enter text

- i) Might participants experience anxiety, discomfort, or embarrassment as a result of your study?

Yes* ☐

No* ☐

***If yes**, what steps will you take to explain and minimise this?

Enter text

***If no**, explain how you can be sure that no discomfort or embarrassment will arise?

Enter text

- j) Will your project involve deliberately misleading participants (deception) in any way?

Yes* ☐

No ☐

If **yes**, please provide further details below and ensure that you cover any ethical issues arising from this in Section 8

Enter text

- k) Will you debrief participants at the end of their participation (i.e. give them a brief explanation of the study)?

Yes ☐

No* ☐

*If **no**, please explain why below and ensure that you cover any ethical issues arising from this in section 8

Enter text

- l) Will participants be given information about the findings of your study? (This could be a brief summary of your findings in general; it is not the same as an individual debriefing)

Yes ☐

No* ☐

If **no**, why not?

Enter text

Section 4 - Security-sensitive material (only complete if applicable)

Security sensitive research includes: commissioned by the military; commissioned under an EU security call; involves the acquisition of security clearances; concerns terrorist or extreme groups.

- a. Will your project consider or encounter security-sensitive material?

Yes* ☐ No ☐

- b. Will you be visiting websites associated with extreme or terrorist organisations?

Yes* ☐ No ☐

- c. Will you be storing or transmitting any materials that could be interpreted as promoting or endorsing terrorist acts?

Yes* ☐ No ☐

* Give further details in **Section 8 Ethical Issues**

Section 5–Systematic reviews of research (only complete if applicable)

- a. Will you be collecting any new data from participants?

Yes* ☐ No ☐

- b. Will you be analysing any secondary data? Yes* ☐ No ☐

** Give further details in **Section 8 Ethical Issues***

*If your methods do not involve engagement with participants (e.g. systematic review, literature review) **and** if you have answered **No** to both questions, please go to **Section 10 Attachments**.*

Section 6 - Secondary data analysis (only complete if applicable)

- a. Name of dataset/s: [Young Lives: an International Study of Childhood Poverty: Rounds 1-5](#)

- b. Owner of dataset/s: [UK Data Service](#)

- c. Are the data in the public domain? Yes ☒ No ☐

*If **no**, do you have the owner's permission/license?*

Yes ☐ No* ☐

- d. Are the data special category personal data (i.e. personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation)?

Yes* ☐ No ☒

- e. Will you be conducting analysis within the remit it was originally collected for?

Yes ☒ No* ☐

- f. **If no**, was consent gained from participants for subsequent/future analysis?

Yes ☐ No* ☐

- g. **If no**, was data collected prior to ethics approval process? Yes ☐ No* ☐

** Give further details in **Section 8 Ethical Issues***

*If secondary analysis is only method used **and** no answers with asterisks are ticked, go to **Section 9 Attachments**.*

Section 7 – Data storage and security

- a) Confirm that all personal data will be stored and processed in compliance with the General Data Protection Registration (GDPR). (See the [Guidelines](#) and the Institute's Data Protection & Records Management Policy for more detail.)

Yes ☐

- b) Will personal data be processed or be sent outside of the European Economic Area (EEA)? Yes*

☐

No ☐

***If yes**, please confirm that there are adequate levels of protections in compliance with the DPA 1998 and state what these arrangements are below

Enter text

- c) Who will have access to the data and personal information, including advisory/consultation groups, and during transcription?

Enter text

During the research

- d) Where will the data be stored?

Enter text

- e) Will mobile devices such as USB storage and laptops be used? Yes ☐

No ☐

***If yes**, state what mobile devices will be used

Enter text

After the research

- f) Where will the data be stored?

Enter text

- g) How long will the data and records be kept for, and in what format?

Enter text

- h) Will the data be archived for use by other researchers? Yes

☐

No ☐

***If yes**, please provide details

Enter text

Section 8 – Ethical Issues

Are there particular features of the proposed work which may raise ethical concerns or add to the complexity of ethical decision making? If so, please outline how you will deal with these below.

It is important that you demonstrate your awareness of potential risks or harm that may arise as a result of your research. You should then demonstrate that you have considered ways to minimise the likelihood and impact of each potential harm that you have identified. Please be as specific as possible in describing the ethical issues you will have to address. Please consider/address ALL issues that may apply from the below.

Ethical concerns may include, but not be limited to, the following areas:

- Methods
- Sampling
- Recruitment
- Gatekeepers
- Informed consent
- Potentially vulnerable participants
- Safeguarding/child protection
- Sensitive topics
- International research
- Risks to participants and/or researchers
- Confidentiality/Anonymity
- Disclosures/limits to confidentiality
- Data storage and security both during and after the research (including transfer, sharing, encryption, protection)
- Reporting
- Dissemination and use of findings

There are no potential ethical issues because the data that I will use is already anonymised and it is publicly available.

Section 9 – Further information

Outline any other information you feel is relevant to this submission, using a separate sheet or attachments if necessary

There is no further relevant information.

Section 10 – Attachments

Please attach the following items to this form, or explain if not attached:

- a) Information sheets and other materials to be used to inform potential participants about the research, including approach letters
Yes ☐ No ☐ N/A ☒
- b) Consent form
Yes ☐ No ☐ N/A ☒
- c) The proposal for the project
Yes ☒ No ☐ N/A ☐
- d) Approval Letter from external Research Ethics Committee Yes
☐ No ☐ N/A ☒
- e) Full risk assessment
Yes ☐ No ☐ N/A ☒

Section 11 – Declaration

I have read, understood, and will abide by the following set of guidelines:

Yes ☒ No ☐

Please select guidelines this project will be abiding by:

BPS ☐

BERA ☒

BSA ☐

Other (please state) ☐ Enter text

I have discussed the ethics issues relation to my research with my supervisor

Yes ☒ No ☐

I have attended the appropriate ethics training provided by my course

Yes ☒ No ☐

I confirm that to the best of my knowledge:

The above information is correct and this is a full description of the ethics issues that may arise in the course of this project

Name: Jhony Pulido

Date: 15-04-2021

Departmental use

If a project raises particularly challenging ethics issues, or a more detailed review would be appropriate, the supervisor must refer the application to the Research Development Administrator via email so that it can be submitted to the IOE Research Ethics Committee for consideration. A departmental research ethics coordinator or representative can advise you, either to support your review process, or help decide whether an application should be referred to the REC. If unsure please refer to the guidelines explaining when to refer the ethics application to the IOE Research Ethics Committee, posted on the committee's website.

Student name: **Jhony Pulido**

Course: **MSc. Social Policy and Social Research**

Project Title: **Does being there at the beginning matter? Studying the effects of father absence during the first year of life on outcomes during childhood and adolescence**

Reviewer 1

Supervisor name: **Rafael**

Novella Supervisor

comments:

The ethics application is correct and does not represent any particularly challenging ethics issue.

Supervisor/first reviewer signature:



Date: **22 April 2021**

Reviewer 2

Advisory committee/course team member name:

Enter text Advisory committee/course team member comments:



Enter text

Advisory committee/course team member signature:

Enter text Date: **Enter text**

Decision on behalf of reviewers

Approved

Referred back to applicant and supervisor

Referred to the REC for review



Recording

Recorded in the student information system

Once completed and approved, please send this form and associated documents to the relevant programme administrator to record on the student information system and to securely store.

Further guidance on ethical issues can be found on the UCL Institute of Education Research Ethics Committee website
