Effects of Family Disruption on Child Development: The Moderating Role of Residential Relocation*

Lucienne N.Y. Disch†
University of Pennsylvania

April 13, 2025

Abstract

This paper studies the consequences of family disruption and associated change of residence for human capital formation. I estimate the impact of family disruption on children's test scores by exploiting variations in family composition and their effects on children's academic performance, due to the father's initial presence and subsequent absence from the household of the mother and the child. To address parents' selection into disruption, I use a dynamic within-child difference-in-differences approach, comparing longitudinal test scores of children who experience family disruption to those of children whose families have not yet separated. I find that, on average, family disruption leads to moderate but significant declines in test scores. However, I highlight that residential relocation emerges as a key factor in the context of family disruption. In the United States, more than one-third of children whose families disrupt have to relocate, and the majority of those move more than a mile away. Using confidential geocoded data, I demonstrate that, on average, children who relocate to a new residence due to family disruption experience significant declines in school performance, particularly those who move more than a mile from their original home. In contrast, children who remain in their current residence or relocate within the same neighborhood exhibit less pronounced declines following family disruption. These findings indicate that the act of relocating, rather than family disruption itself, is the primary factor contributing to the observed test score gap. Consequently, targeted policies – such as providing support to newly single mothers and their children to help them remain in their familiar residential areas after separation – could alleviate the negative impact of long-distance moves on children's academic performance.

*I am indebted to Francesco Agostinelli, Jere Behrman, Margaux Luflade, and Petra Todd for many fruitful discussions and suggestions. For helpful comments, I thank my discussant Susan Clampet-Lundquist, as well as seminar and conference participants at the University of Pennsylvania, the SEA 2023 Conference at the University of Zurich, the 2023-2024 Penn Urban Studies Graduate Colloquium, and the 2024 Annual Congress of the Swiss Society of Economics and Statistics.

†University of Pennsylvania, Department of Economics. Email: brunnerl@sas.upenn.edu.

1 Introduction

Both the family and the neighborhood in which a child is raised are crucial to their development (Agostinelli, Doepke, Sorrenti, and Zilibotti, 2024). Social disruptions, such as parental separation or a move, can have an adverse impact on human capital formation. Such disruptions also often coincide – I find that 38% of children who experience family disruption have to residentially relocate, and 82% of those move more than a mile away.[1] Father gone, friends gone? Little is known about the effects of family disruption on child development and what role residential relocation plays. What are the causal effects of a change in family composition on test scores, and does it matter whether and how far away children have to move as a consequence?

To answer these questions, I examine the impact of family disruption on children's test scores by exploiting variations in family stability and their effects on children's academic performance, due to the father's initial presence and subsequent absence from the household. To overcome parent's selection into separation, I use a dynamic within-child difference-in-differences approach, comparing longitudinal test scores of children who experience family disruption to those of children whose parents have not yet separated.

Consistent with previous research – which often focuses only on married couples and uses the legal date of divorce as the point of separation – I find that family disruption has moderate but significant adverse effects on test scores. I highlight an important finding in the context of family disruption, which is *residential relocation* acting as a moderating variable. I suggest that the relationship between family disruption and test score performance depends on whether there is an associated move. On average, the phenomenon of moving contributes to the test score gap, rather than family disruption by itself. More specifically, the main transmission channel appears to work through the *distance* of the move. If they move not more than a mile away, test scores are less

[1] Calculations are based on a sample of school-aged children in the NLSY Geocode Data, the National Longitudinal Survey of Youth 1979 (NLSY79) confidential spatial data files from the U.S. Bureau of Labor Statistics (BLS).

affected. In fact, staying in the same area mitigates – if not offsets – the adverse effects of family disruption.

From a policy perspective, my results – which show that the negative impact on test scores is concentrated among children who relocate to a new area – are relatively encouraging. They suggest that mitigating the adverse effects of family disruption may be more achievable by preventing moves, rather than intervening directly within the family, which could be more difficult and potentially counterproductive. Targeted policies, such as helping newly single mothers and their children stay in their familiar residential areas after separation, could help reduce the negative consequences of long-distance moves on children's academic performance.

This paper is related to a growing literature studying the effects of family disruption, such as parental relationship dissolution, death, or incarceration. While there is mounting evidence on the importance of the family background[2] and the power of place[3] for child development, less is known about the disruptive impact of a change in the social environment.[4] A key insight in this literature is that disruption occurrences are not random (Marinescu (2016); Greenwood et al. (2016); Nielsen et al. (2025)). For example, those families experiencing family disruption have different characteristics than more stable families. Therefore, it is important to account for potential selection into "treatment" of family disruption. Because, when measuring the relationship between family disruption and outcome variables such as test scores, models can either overstate or understate the effect if there is an omitted variable that is both correlated with parental separation and is a determinant of test scores. In the last years, the literature has made substantial progress in measuring the causal effect of family disruption on the child's outcomes in the presence of sorting, identified with the help of

[2]E.g., Cunha and Heckman (2007); Cunha and Heckman (2008); Agostinelli et al. (2024); Agostinelli and Wiswall (2024), among many others.

[3]E.g., Chetty et al. (2016); Chyn (2018), among many others.

[4]E.g., Tartari (2015); Selya et al. (2016); Chetty and Hendren (2018a); Cordes et al. (2019); Chan and Liu (2024).

changes in marital status and dynamic structural modeling (Tartari, 2015), survey data and a control function approach (Chan and Liu, 2024), or administrative data and an event study analysis (Holm et al., 2023). A difficult problem in this context is the identification of the date of separation. Tartari (2015) addressed this issue by using the legal date of divorce. Chan and Liu (2024) and Holm et al. (2023) focus on changes in family stability based on survey and register data, respectively. The idea is that variations in family composition should be reflected in the living situations of both parents. Family disruption typically implies the transformation of a two-parent household to a single-mother household.[5] These circumstances can lead to reduced financial and time investments in the child, which may negatively affect the child in ways that extend beyond changes within the family environment (Cunha and Heckman (2007)). A number of studies investigate changes in financial resources available to the child and mother, as well as engagement of the father with the child, after separating from the family (Kalil et al. (2011); Del Boca et al. (2014); Chan and Liu (2024)). The effects of family disruption on children's outcomes, particularly test scores, may be heterogeneous. These outcomes can vary significantly depending on several factors, such as changes in social support, the emotional climate at home, or the child's new living situation. Family disruption often involves changes in living arrangements, which can lead to shifts in a child's social environment—such as moving to a new neighborhood and attending a new school, forming new peer relationships, or being closer to supportive grandparents. Additionally, the level of conflict within a household can have a more severe negative impact on a child's well-being than the physical separation of parents. In some cases, children may even show improved test scores following family disruption, especially if the separation reduces household tension. This complexity underscores the difficulty

[5]Most datasets have the shortcoming of not capturing a significant number of cases where children reside with their fathers after parental separation. The NLSY, which I use, is no exception, as it treats mothers and children as paired units while handling fathers separately. Furthermore, family disruption typically involves a transition from a shared household to two separate ones—implying that the father remains associated with a household but no longer lives with the rest of the family. However, in some cases, family disruption results from more involuntary events, such as the father's death or incarceration. My baseline analysis includes a pooled specification covering all types of family disruption, alongside separate estimates for parental breakups.

of isolating the specific causes and consequences of family disruption on children's outcomes, illustrating the challenge in causally identifying how such disruptions influence academic performance.

This paper contributes to the literature by introducing a new source of information and an identification strategy that illuminate how residential mobility linked to family disruption affects children. Survey data on household composition changes are used to identify family disruption events during periods when school-aged children take standardized tests biennially. Geospatial microdata is further used to analyze the geographic dimensions of a shared household splitting into two separate ones. While I do not explicitly account for exact latitude and longitude, I demonstrate that proximity between the old and new residences provides valuable information for identifying the distance of moves associated with family disruption.[6] The geographic distances between the old and new household locations turn out to be strong mediators for the effect of family disruption on test scores. This is relevant, as, for example, Holm et al. (2023), the study most closely related to mine, rely solely on register data and do not account for the geographic dimension of changing living situations. In addition, other proxies for family disruption, including Tartari (2015)'s approach of taking the legal date of divorce, have been found to not identify the exact date of physical separation and exclude cohabitants. Moreover, the literature generally does not account for other types of family disruption beyond parental relationship dissolution, such as parental incarceration or death. In contrast, I distinguish between different types of disruption, allowing for a more comprehensive understanding of how children's outcomes are affected by a range of family changes.

From a methodological perspective, my approach is closely related to event study strategies aiming to evaluate the impact of interventions or events, such as the introduction of a policy. The idea is to track individuals over time, measuring their outcomes both before and after a significant change, and comparing them to a control group of

[6] The data allows me to identify their domiciles at the county level. While I also know whether they changed the ZIP codes when moving and the exact distance they moved, I do not know their actual ZIP codes. Plus, I cannot infer whether that includes a change in school zones or schools.

individuals who are identical in all relevant aspects except for the treatment. I use a dynamic within-child difference-in-differences approach with multiple time periods that accounts for staggered treatment and heterogeneous treatment effects among children to quantify the causal impact of family disruption – and any associated changes in residence – on test scores. Recent econometric literature has identified that the traditional two-way fixed effects estimator can be biased when treatment is staggered, treatment effects are heterogeneous, and there are dynamic treatment effects over time, and therefore offers alternative methods (Sun and Abraham (2001); de Chaisemartin and D'Haultfœuille (2020); Goodman-Bacon (2021); Baker et al. (2022); Dube et al. (2025)). For an overview of recent advances in the difference-in-differences literature, see Roth et al. (2023). To address these concerns, I follow the approach by Callaway and Sant'Anna (2021) for my baseline specification. This method provides a dynamic roll-out framework that accounts for variations in the timing and impact of family disruption, recognizing that children experience it at different ages, in different years, and with potentially different effects. I also conduct the analysis using the method proposed by Wooldridge (2021) as robustness check. An important challenge in measuring the effect of family disruption is defining the appropriate counterfactual. The issue is that stable two-parent households, which are unlikely to experience disruption, may not serve as an appropriate control group. Because, they may fundamentally differ in their characteristics and do not face the same levels of tension or struggle, as households that are on the verge of disruption. However, to accurately estimate the causal effect of the transition from a two-parent to a single-parent household, it is essential that the counterfactual is realistic. Another crucial advantage of the method by Callaway and Sant'Anna (2021) is that it helps address concerns about parents' selection into separation by using not-yet-separated families, rather than never-separated families, as the control group to estimate the impact of separation. Thus, I compare the longitudinal test scores of children who experience family disruption – along with any potential associated change of residence – to those of children who have not yet been affected by the time of comparison but may experience it later. To further ensure robustness, I am reducing my sample to a balanced panel, so that I am tracing the exact same individuals over time with no missing information for any year.

The key identification assumption for causal estimates in the difference-in-differences framework posits that, after controlling for family and cohort fixed effects, the test scores of children who experience disruption and those who do not would follow parallel trends in the absence of the disruption. This guarantees that the control group provides a valid counterfactual for the treated group. I support this assumption by showing that there are no differential pre-trends in the outcomes of interest. Thus, since parallel pre-trends hold, any difference in post-treatment outcomes can be attributed to the causal effect of the family disruption, rather than underlying differences between groups. The method proposed by Callaway and Sant'Anna (2021) allows to include anticipation of family disruption. However, by testing for significant deviations in outcomes prior to treatment, I verify that anticipation effects are not a concern. Pre-treatment estimates are close to zero and statistically insignificant, indicating no systematic behavioral response before the treatment starts.

Instead of relying on the legal date of divorce, which often overlooks cohabiting parents who split and typically occurs months or even years after the actual physical and emotional separation – or, in some cases, never at all – my spatial identification strategy provides a more nuanced perspective. It allows me to examine children who transition from a two-parent household to a one-parent household, taking into account the geographical dimensions of this transition. While this paper is not the first to apply a difference-in-differences approach to investigate the effects of parents splitting households on children's test scores (Holm et al. (2023)), to the best of my knowledge, it is the first to move beyond register data and focus not only on the parental separation event per se, but also on other types of family disruption and the spatial dimensions of the living situations in the context of family disruption.

The results show that family disruption significantly decreases test scores. On average, the test scores of children experiencing separation fall by 1 point, corresponding to a 3% decrease. A longer-term analysis using a subsample with a more extended panel suggests that this gap persists for 3 to 5 years after the disruption event, though the estimates are no longer statistically significant, potentially due to reduced statistical power in the smaller subsample. The study further reveals that the adverse effects on test scores are primarily concentrated among children who move in association with family disruption. While affected children who stay show no significant decline in test scores, those who move experience significantly worse outcomes of almost -2 raw points. The distance of the move also matters: long-distance moves are associated with particularly large significant negative effects of more than -2 raw points, corresponding to a 6% decrease, while short-distance moves have smaller, statistically insignificant impacts. These findings suggest that much of the observed academic harm stems not from family disruption itself, but from the residential instability that often accompanies it. Robustness checks focusing on parental breakup alone yield qualitatively similar patterns, although the overall negative effect on test scores is somewhat smaller in magnitude.

Overall, my results suggest that whether children relocate due to family disruption—and the distance of their new home from the previous one—can significantly influence their test scores. By considering family disruption and residential relocation together, I provide new evidence of a key mechanism behind the negative effects of parental separation on academic performance: the concurrent occurrence of residential relocation, which acts as a moderating factor. From a policy perspective, these findings emphasize the importance of accounting for residential relocation when analyzing family disruption. Solely focusing on the general effects of family disruption can lead to an overestimation of its true impact. Interestingly, while children who move in association with family disruption experience adverse outcomes, those from stable families tend to perform better after a move than before. This suggests that the effects of relocation depend on the reasons behind the move, highlighting the endogenous nature of mobility. In this way, I also contribute to the literature on the role of residential relocation in shaping children's outcomes (Chetty and Hendren (2018a); Chetty and Hendren (2018b); Chetty et al. (2016)).

2 Identification Strategy

Measuring family disruption is inherently challenging. While the legal date of divorce is a common metric, it is insufficient for at least two reasons. First, divorce may occur years after parents have emotionally and physically separated – or it may not happen at all. Second, focusing solely on divorce dates excludes cohabiting families who were never married and overlooks more "involuntary" family disruptions caused by incarceration or the death of a parent. I propose a geospatial identification strategy. I follow school-aged children who take standardized tests biennially, while simultaneously tracking their living situations – accounting for both family composition and residential mobility. This approach allows me to identify family disruptions, residential relocation events and distances of moves, which can serve as indicators of shocks in a child's life that may impact their test score performance. In this chapter, I present the data sources and describe the key variables, such as the standardized test scores that serve as the outcome variables, in more detail. I also detail the measurement of family disruption, the tracking of children's living situations, and the selection of the control group that underpins my event study framework.

2.1 Data Sources and Key Variables

The main data source for this project consists of panel data from the National Longitudinal Survey of Youth (NLSY) by the U.S. Bureau of Labor Statistics (BLS).[7] In particular, I base my analysis on the NLSY cohort of Americans born between $1957 - 1964$ (NLSY79). The cohort originally included $12,686$ respondents ages $14 - 22$ when first interviewed in 1979 and data are now available from Round 1 (1979 survey year) to Round 29 (2020 survey year). I link the women in the NLSY79 with their biological children in the NLSY79 Children and Young Adults cohort (NLSY79 Child/YA). To date, a total of $11,545$ children have been identified as born to interviewed NLSY79

[7]For more information and details, consult the NLSY website of the U.S. Bureau of Labor Statistics: https://www.nlsinfo.org/content/cohorts, accessed November 2024.

mothers. Data are now available from 1986 to 2018, representing 17 survey rounds for the child sample and 13 for young adults in that time span. I further match these linked files with the corresponding confidential NLSY79 and NLSY79 Child/YA Geocode data, which contain information on respondents' geographic locations.

These geocoded files reveal information such as from which county to which county a person moved. They also provide information regarding the geographical distance of the places of residence between two consecutive interview rounds at a more granular level. While the confidential Geocode data files do not reveal the identified ZIP codes where people live, they do provide information regarding whether they changed the ZIP or not between two survey years. On top of that, the distance between two places of residence between two interview rounds is reported more specifically (0 miles (non-mover), $0-999$ feet, 1000 feet-1 mile, $1-5$ miles, ..., 500+ miles). The data also allows me to infer the identification of those moving abroad. While latitudes and longitudes were generated from exact address matches in most cases, migration distances were calculated using ZIP code centroids when exact addresses were unavailable.

The combination of matched survey and geospatial data from the NLSY enables me to capture a range of characteristics related to a school-aged child's living situation every other year, across at least three and up to five observation periods in my panel. Specifically, for each observation period, I have information on whether the father resides in the same household, the child's location at the county level, whether the child moved to a new residence compared to the previous observation period, and – conditional on moving – the exact geographical distance between the old and new residences. Simultaneously, I observe children's standardized test scores over these three to five consecutive periods. These test scores serve as the primary outcome variables to measure the effects of changes in family composition and related moves on children's academic performance.

Key measurement variables in this paper are the raw Peabody Individual Achievement Test (PIAT) scores. The PIAT is one of the most popular used assessments of academic achievements in the United States, primarily thanks to its high test-retest reliability and concurrent validity. Children take these standardized tests every other year during K-12 (between five and 22 years of age). I focus on the three most widely used PIAT sub-assessments available in the NLSY data from 1986 to 2020: Mathematics, Reading Recognition, and Reading Comprehension, with the latter serving as my baseline outcome measure. All children take the same tests in a particular year. Relatively older children are expected to score higher than younger ones, although there are some adjustments in the difficulty depending on which grade the child is enrolled in, which I will highlight in more detail later. However, given that the baseline test is the same irrespective of the age of the child, this also implies that children typically improve their performance in the tests as they get older throughout their school years. A detailed explanation of what these assessments contain in evaluating a child's skills in both reading and mathematics can be found in *Appendix A*, as well as on the NLSY website of the U.S. Bureau of Labor Statistics.[8].

The PIAT consists of many institutional features that are very advantageous and appealing when comparative studies of academic performance among children are desirable. Given that all children take the same baseline standardized test every other year during K-12, irrespective of their age, school attended, and neighborhood they live in, it is very comparable, for example, across different children of similar age. This test is not only comparable from a cross-sectional viewpoint, but also from a longitudinal one, as it enables the comparison of trends in raw test scores over time across different children. Children take these tests up to five times in their life, so I observe at most five consecutive biennial test score performances for the same child in my sample. Thus, the longitudinal sample period does not exceed ten years.

To simplify the analysis, I transform the years in which I observe test scores into consecutive periods $(1, 2, 3$, or $1, 2, 3, 4$, or $1, 2, 3, 4, 5)$ for each child. This transformation

[8]https://www.nlsinfo.org/content/cohorts/nlsy79-children/topical-guide/assessments/piat-reading-reading-recognitionreading, accessed November, 2024.

does not affect the validity of the technique or results, but it enhances statistical power and precision when comparing children who experience family disruption in a given year with those who remain in a two-parent household during the same time (but may separate later). I believe that the specific year of disruption is not critical to the analysis; what matters more is the child's age at the time of disruption and the distance between the event and the measured outcome. Given the study period (1986–2014), it is important to have a sufficiently large number of families experiencing disruption each year to ensure that the treated group is adequately sized for comparison to the larger control group. Thus, transforming the time domain reduces the number of parameters and enhances power without any loss of generality. Additionally, this approach makes the panel data more concise and well-structured by consistently defining each child's time period as "Year 1" through "Year 5".

2.2 Measuring Family Disruption

To identify family disruption events, I employ a panel data approach by tracking school-aged children over ten years, for whom I have biennial PIAT scores. I include only those children who are living in two-parent households at the time of their initial entry into the panel, which is also when they take their first of up to five tests. To track children who transition from a two-parent household to a single-parent household over the ten-year period, I examine changes in family composition. Based on this analysis, I create two additional indicator variables, defined below.

The first variable, $FamDisrup_i$, indicates whether and when child i experienced a family disruption event. For children who experienced disruption, $FamDisrup_i$ takes on the values 2, 3, 4, or 5, corresponding to the year of disruption. For those who did not experience disruption, $FamDisrup_i = 0$. "Year 1" serves as the reference or baseline period where all children are living in two-parent households, so no disruptions occur then (i.e., $Fam.Disrup_i \neq 1$). This setup provides four consecutive years (Years

2 to 5) during which a disruption event can occur. I group all children who experience disruption in the same year into the same group (or "cohort").

The second indicator variable, $Treat.Fam_i$, is a binary variable indicating whether child i experienced family disruption at any point during the ten-year period. Specifically, $Treat.Fam_i = 1$ if the child experienced a disruption, and $Treat.Fam_i = 0$ if the child remained in a two-parent household throughout the entire period.

Definition 1 (Family Disruption Start):

$$Fam.Disrup_i = \begin{cases} 2 & \text{if} \quad (i_{t=1} \in \mathrm{T}) \cap (i_{t=2} \in \mathrm{S}) \\ 3 & \text{if} \quad (i_{t=1} \in \mathrm{T}) \cap (i_{t=2} \in \mathrm{T}) \cap (i_{t=3} \in \mathrm{S}) \\ 4 & \text{if} \quad (i_{t=1} \in \mathrm{T}) \cap (i_{t=2} \in \mathrm{T}) \cap (i_{t=3} \in \mathrm{T}) \cap (i_{t=4} \in \mathrm{S}) \\ 5 & \text{if} \quad (i_{t=1} \in \mathrm{T}) \cap (i_{t=2} \in \mathrm{T}) \cap (i_{t=3} \in \mathrm{T}) \cap (i_{t=4} \in \mathrm{T}) \cap (i_{t=5} \in \mathrm{S}) \\ 0 & \text{if} \quad (i_{t=1} \in \mathrm{T}) \cap (i_{t=2} \in \mathrm{T}) \cap (i_{t=3} \in \mathrm{T}) \cap (i_{t=4} \in \mathrm{T}) \cap (i_{t=5} \in \mathrm{T}), \end{cases}$$

where T stands for "Two-Parent-Household" and S stands for "Single-Parent-Household".

Definition 2 (Family Disruption Event):

$$Treat.Fam_i = \begin{cases} 1 & \text{if} \quad Fam.Disrup_i \in [2, 3, 4, 5] \\ 0 & \text{otherwise.} \end{cases}$$

I can then construct a panel data frame as shown in Table 1, that captures all the relevant data for my analysis. Importantly, I have one row per child ($Child.ID$) per time period ($Year$), along with the child's PIAT scores ($Test.Scores$) for Year 1 up to Year 5. Note that, even though I call them "Year 1", "Year 2", etc., there is a two-year gap between two such time periods, because these tests are only taken every other year. I further have my two indicator variables that inform whether there was a potential family disruption event ($Treat.Fam$) and in what year ($Fam.Disrup$). Lastly, I control for a set of time-invariant covariates, including the child's age and race, the mother's AFQT score and years of education, and an indicator for whether the child has siblings.

To ensure consistency in data and facilitate meaningful comparisons, I restrict the analysis to children who consistently live with their biological mother. Additionally, the sample is further narrowed to children who live with their biological father during the first period, excluding cases where a stepfather or another individual is present in the household. This focus is partly due to data limitations, as the dataset does not provide sufficient observations of single-parent fathers or step-parents, among other scenarios. In this setup, any transition to a "single-parent" household is defined as either the biological father exiting the household or the mother and child relocating to establish a new household elsewhere.

Table 1: Illustration of the key variables for the analysis at hand

Child.ID	Year	Test.Scores	Fam.Disrup	Treat.Fam
244	1	66	2	1
244	2	45	2	1
244	3	41	2	1
244	4	38	2	1
244	5	62	2	1
817	1	88	0	0
817	2	93	0	0
817	3	79	0	0
817	4	98	0	0
817	5	99	0	0
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.

2.3 Measuring Residential Relocation

To investigate the effects of family disruption on test scores, I leverage up to five consecutive years of test score observations, which allows for a detailed view of how academic performance evolves before and after the disruption. However, to increase statistical power and ensure a larger, more stable sample, my baseline event study includes all children with at least three consecutive years of test score data. I adopt this three-year minimum to align the sample more closely with analyses that examine the interaction between family disruption and residential relocation. This restriction is particularly important because incorporating residential moves substantially reduces the number of treated children, and a lower threshold for inclusion helps preserve sufficient statistical power.

The identification and definition of residential relocation follow a similar approach to that used for family disruption: I construct a variable $Move_i \in [2, 3, 0]$ and an additional variable $Treat.Move_i \in [1, 0]$. To capture the combined experience of family disruption and relocation, I create unified measures. These include $Treat_i \in [2, 3]$ and $First.Treat_i = 1$, identifying children who experience both events. Conversely, children who remain in two-parent households and the same residence throughout the sample period are characterized by $Treat_i = 0$ and $First.Treat_i = 0$.

2.4 Diagnostics of the Counterfactual (Control Group)

A critical aspect of analyzing the effects of disruption events on children's outcomes is selecting an appropriate counterfactual or comparison group, as this decision fundamentally shapes the interpretation of the causal effect. While the literature typically compares families undergoing disruption to very stable two-parent households, this can lead to biased estimates of the causal effect. Families experiencing disruption may face pre-existing tension in the years leading up to the event. However, comparable levels of tension may also exist in families that ultimately do not separate. Overlooking this nuance risks conflating the effects of family tension with the causal impact of the disruption itself. Another key challenge is sorting—families with certain underlying characteristics may be more prone to experiencing disruption, leading to selection into treatment. If not addressed, this selection process can distort the validity of comparisons. To mitigate these concerns, my baseline analysis uses a control group consisting of individuals who have not yet experienced family disruption. This "not-yet-treated" control group includes children who never experience disruption during the observed time period, as well as those who have not yet experienced disruption at a given point in time but eventually do. This group is at least as large as the never-treated group, though it changes across time periods. This approach ensures greater similarity in pre-treatment characteristics and family dynamics between treated and control groups, providing a more robust basis for causal inference.

However, before turning to my main analysis, which allows for the inclusion of "not-yet-treated" children as a control group, I first present simpler analyses that rely on the never-treated group. The methods used in the next section cannot accommodate not-yet-treated controls, but they provide an initial look at the relationship between social disruption and test scores. This preliminary analysis offers a useful point of contrast to the more advanced approach introduced later.

3 Descriptive Evidence

In this chapter, I begin by presenting summary statistics to describe the data. I then conduct an initial analysis of the relationship between social disruption events and test scores using Ordinary Least Squares (OLS) regressions and static two-way fixed effects (TWFE) models.

3.1 Summary Statistics

Outcome Variables. Table 2 presents detailed summary statistics on the evolution of test score performance in reading comprehension and mathematics for children with at least three years of observations. The sample includes all children in the dataset, encompassing both those who experienced a disruption during the observed period and those who did not.

Table 2: Summary Statistics

Obs	Min	Max	Avg	Std
Year 1	0	67	21.8	9.6
Year 2	0	78	36.3	11.1
Year 3	0	80	46.1	10.7

Reading Comprehension

Obs	Min	Max	Avg	Std
Year 1	0	64	21.1	10.3
Year 2	6	84	37.3	11.9
Year 3	0	84	49.4	10.5

Mathematics

Family Disruption. In my sample, the probability that a child experiences family disruption in a given year is 12%. This probability is notably higher for Black children (21%) than for non-Black children (11%). However, because non-Black children outnumber Black children by nearly five to one in the sample, less than a third (29%) of all children who experience family disruption are Black. Children experiencing family disruption are, on average, 6.7 years old at the start of the three observation periods, which span a total of six years, with observations taken every other year. Furthermore, among those families experiencing family disruption, the mother has, on average, 14 years of education and an average AFQT score of 42694.

The numbers provided above refer to all types of family disruption, including parental breakups and more involuntary disruptions, such as paternal death or incarceration. When focusing solely on parental breakups, where the father is still "around" but no longer resides with the child and mother, the numbers are similar but lower. For this subsample of family disruption types, the probability of experiencing a parental breakup is 9% overall, 18% for Black children, and 7% for non-Black children. In addition, when focusing solely on parental breakups, on average, mothers have 13.4 years of education and an AFQT score of 37673.

Residential Relocation. Conditional on experiencing family disruption (parental breakup), 33% (37%) of the children have to move to a new residence with their mother. Among these movers due to family disruption (parental breakup), 37% (39%) are Black, and, on average, their mothers have 13.9 (13.7) years of education and an AFQT score of 36331 (32302).

Among stayers who experience family disruption (parental breakup), 25% (34%) of children are Black. Plus, on average, their mother's years of education amounts to 13.9 (13.2) with an AFQT score of 45773 (40860). Average age of the children and probability for siblings is similar among all groups discussed.

3.2 OLS Regression Models

As an initial analysis of the relationship between social disruption and test scores, I estimate a series of OLS regressions. This provides a first description at how academic outcomes – measured via raw PIAT scores – differ between children who experience family disruption and residential relocation and those who do not. My baseline outcome measure consists of the PIAT reading comprehension score.

3.2.1 OLS Regression Model (FamDisrup)

I begin by examining the relationship between family disruption and test scores using a simple OLS regression model that abstracts from whether the disruption involves residential relocation. The timing of observations is structured relative to the disruption event: I normalize the period in which family disruption occurs to $t = 0$, and observe test scores at five time points relative to this event – specifically, $t \in \{-3, -1, +1, +3, +5\}$. I restrict the sample to children who live with both biological parents in $t = -3$ and

$t = -1$, which is prior to the potential disruption event. In addition, I restrict that those children who experience family disruption in $t = 0$ remain in a single-parent household with their mother from $t = +1$ through $t = +5$. This selection ensures a clearly defined timing of family disruption between the second and third PIAT observations and allows for a cleaner interpretation of how test scores evolve around the disruption event. To estimate the association between family disruption and test performance, I run regressions of the following form:

$$y_{it} = \beta_0 + \beta_1 FamDisrup_{i0} + \mathbf{X}_i'\gamma + \varepsilon_{it},$$

where y_{it} is the raw PIAT reading comprehension score of child i at time t, and $FamDisrup_{i0}$ is a binary variable equal to 1 if the child experiences family disruption at $t = 0$, and 0 otherwise. The vector $\mathbf{X}_i$ includes the following time-invariant child characteristics: race, age at first observation, presence of siblings, mother's years of education, and her AFQT score. These variables are included to account for differences across children that may influence both their likelihood of experiencing family disruption and their academic performance, thereby helping to isolate the relationship between family disruption and test scores. Importantly, I do not control for household income in this analysis. Income may both influence the likelihood of family disruption and be affected by it, introducing post-treatment bias if included as a covariate. Moreover, income is strongly correlated with maternal education and cognitive ability, which are already captured through years of education and AFQT score. Lastly, the error term ε_{it} captures all other unobserved influences on test scores at time t that are not accounted for by observed covariates or family disruption status.

In this simple model, identification comes from changes in family composition: $\mathbb{E}\left[\varepsilon_{it} | FamDisrup_{i0}, \mathbf{X}_i\right] = 0$. The assumption that, conditional on observed child and family characteristics, the occurrence of family disruption is exogenous from the child's perspective, is a strong one. However, it is important to note that this is a more descriptive, correlational analysis and a more rigorous empirical strategy will be introduced in a later section. The main study then includes a difference-in-differences approach with a control group of children not yet treated and parallel pre-treatment trends between the treated and control groups. This approach will better address potential sorting into separation and more accurately estimate the causal effect, while tackling endogeneity concerns.

The above regression model is estimated separately for each of the five observed test score periods. This allows me to examine how the association between family disruption and academic performance evolves over time, from several years before to several years after the event. This is important because the magnitude of the association may differ depending on the timing – for example, whether test scores are measured one period before or three periods after the disruption.

Table 4 presents estimates from the regression model, showing the impact of family disruption at $t = 0$ on reading comprehension scores across different time spans between the disruption and the timing of test score measurement. The model includes a dummy variable for family disruption, equal to 1 if disruption occurred and 0 otherwise. All estimates are conditional on the inclusion of control variables. When controls are omitted, the results are in line with those presented, and the effects are even stronger.

Table 4: OLS Regression (Family Disruption)

	Test Scores Measured Around Family Disruption Event in t				
Time	**t − 3**	**t − 1**	**t + 1**	**t + 3**	**t + 5**
Family Disruption (Dummy)	**-0.70****	**-2.03*****	**-3.21*****	**-5.03*****	**-5.34*****
	(0.35)	(0.64)	(0.64)	(0.70)	(0.76)
Observations (balanced panel)	3,435	3,435	3,435	3,435	3,435
Experiencing Fam Disruption	6.8%	6.8%	6.8%	6.8%	6.8%
R^2	0.16	0.12	0.07	0.11	0.11
Note:					*p<0.1; **p<0.05; ***p<0.01

Note: Impact of family disruption in $t = 0$ on test scores as a function of time span between family disruption and the timing of test scores measurement.

Table 4, presenting estimates of OLS regressions of family disruption on test scores, shows significant negative estimates. For example, the estimated effect is a decrease of 5.34 raw points in test scores five years after experiencing family disruption. This appears to be a high adverse effect on test scores given that average test scores for reading comprehension are around 34.7.

This descriptive evidence shows that, at any point in time, there is a gap between the affected group and those not affected. However, it is not clear whether that gap is increasing or not since family disruption. While the values do increase over time, the trends do not necessarily need to do so. It could be that part of the trend is just continuing. And if such a trend is continuing, then this implies extra losses around 1.5 every other year, which are not because of what happened in between. However, a difference-in-differences analysis in the next chapter will test that. Before turning to the main analysis, I conduct an additional descriptive analysis examining the interaction effects of family disruption and residential relocation on children's test scores using an OLS regression model.

3.2.2 OLS Regression Model (FamDisrup×Move)

To undertake a preliminary examination of whether the impact of family disruption on test scores depends on a concurrent residential move, I focus on children for whom I observe at least three consecutive test scores, and estimate regression models that include indicators for family disruption, moving, and their interaction:

$$y_{it} = \beta_0 + \beta_1 FamDisrup_{i0} + \beta_2 Move_{i0} + \beta_3 (FamDisrup_{i0} \times Move_{i0}) + \mathbf{X}_i' \gamma + \varepsilon_{it},$$

where $\mathbb{E}\left[\varepsilon_{it} | FamDisrup_{i0}, Move_{i0}, FamDisrup_{i0} \times Move_{i0}, \mathbf{X}_i\right] = 0$. The binary variable $Move_{i0}$ is equal to 1 if the child has to move in period 0 and equal to 0 if the child stays in the same place throughout the years during which I observe test score outcomes. Table 5 presents OLS estimates from the above regression model.

Table 5: OLS Regression (Family Disruption X Move)

	Test Scores Around Disruption Event in t		
Time	**t − 1**	**t + 1**	**t + 3**
FamDisrup (Dummy)	1.71***	1.35***	1.01**
	(0.35)	(0.52)	(0.53)
Move (Dummy)	-0.34	0.47	1.28***
	(0.23)	(0.34)	(0.35)
FamDisrup x Move	-2.14***	-4.71***	-4.39***
	(0.59)	(0.87)	(0.89)
Observations	4,410	4,410	4,410
R^2	0.54	0.28	0.19
Note:	*p<0.1; **p<0.05; ***p<0.01		

Note: Impact of *Family Disruption* × *Move* in $t = 0$ on test scores as a function of time span between simultaneous disruption events and the timing of test scores measurement.

Table 5 shows the resulting impact of family disruption in interaction with residential relocation on test scores. For example, three years after family disruption, for stayers, the estimated effect is an *increase* of 1.01 raw points, while for children who had to move, the predicted *decrease* in test scores amounts to $1.01 - 4.39 = -3.38$ raw points. OLS estimates are statistically significant, and their magnitudes increase in models that omit covariates (not shown). The results from this preliminary descriptive analysis highlight that the negative impact of family disruption on academic performance appears to be driven primarily by children who experience a simultaneous residential move. Next, I further investigate whether the distance of the move appears to matter in terms of its impact on academic outcomes.

3.2.3 OLS Regression Model (FamDisrup×DistMove)

I extend the analysis by accounting for the distance of the move, rather than simply whether a child moved. To do so, I again estimate regression models interacting family disruption with residential relocation, but this time conducted separately for two subsamples: children who moved at most one mile and those who moved more than one mile. Table 6 presents the results.

Table 6: OLS Regression (Family Disruption X DistMove)

	Test Scores Around Disruption Event in t		
MOVE ≤ 1 MILE	**t − 1**	**t + 1**	**t + 3**
FamDisrup (Dummy)	1.70***	1.43***	1.02**
	(0.35)	(0.52)	(0.54)
Move (Dummy)	-0.95***	0.39	1.00
	(0.48)	(0.71)	(0.74)
FamDisrup x Move	1.43	-0.57	-0.55
	(1.14)	(1.69)	(1.75)
MOVE > 1 MILE	**t − 1**	**t + 1**	**t + 3**
FamDisrup (Dummy)	1.71***	1.33***	1.03**
	(0.35)	(0.52)	(0.53)
Move (Dummy)	-0.21	0.51	1.34
	(0.22)	(0.37)	(0.37)
FamDisrup x Move	-2.94***	-5.69***	-5.26***
	(0.63)	(0.93)	(0.95)
Note:	*p<0.1; **p<0.05; ***p<0.01		

Note: Impact of *Family Disruption* × *Distance of Move* in $t = 0$ on test scores as a function of time span between simultaneous disruption events and the timing of test score measurement.

Table 6 provides descriptive evidence demonstrating that children who move more than a mile away, in connection with family disruption, experience significantly greater adverse effects on their test scores compared to those who move within a mile.

The initial analysis using OLS with time-invariant controls provides useful descriptive and correlational insights into the relationship between family disruption, residential relocation, and test scores. However, this approach does not account for potential unobserved child-specific characteristics that could be correlated with both the independent variables (family disruption, moving) and the outcome (test scores). These unobserved factors could lead to biased estimates, potentially overstating or understating the effect of family disruption and moving on test scores. To address this limitation and provide more robust estimates, I now turn to a Two-Way Fixed Effects (TWFE) model.

3.3 Static Two-Way Fixed Effects Model

To analyze the relationship between social disruption and test scores, I now estimate a series of static TWFE models to provide alternative estimates. This provides a more sophisticated look at how academic outcomes – measured via raw PIAT scores – differ between children who experience family disruption and residential relocation and those who do not. Again, my baseline outcome measure consists of the PIAT reading comprehension score. I focus on a specific simple difference-in-differences case of the staggered version, considering a group of individuals who experience disruption at the same specific point in time. Since in this case, I do not have a roll-out design anymore, a TWFE model is valid. Disruption effects can still be heterogeneous through time, so that depending on the distance to the disruption event, the strength of treatment can vary. This approach aims to improve causal inference by isolating the effect of family disruption and relocation on test scores, while accounting for unobserved heterogeneity that may otherwise confound the results.

3.3.1 TWFE Model (FamDisrup)

To investigate the relationship between social disruption and test scores, I first focus on all children for whom I have five observation periods and who either experience family disruption in $t = 0$ – i.e., get treated between periods $t = -1$ and $t = +1$ – which means between the second and third of their five observed time periods, or never. Given that this constitutes a static setup, a TWFE model is appropriate. Nonetheless, disruption effects may remain heterogeneous over time, with treatment intensity potentially varying based on the distance from the disruption event. I estimate the following panel data model with child and time fixed effects.

$$y_{it} = \alpha_i + \delta_{-1} f_t^{-1} + \delta_{+1} f_t^{+1} + \delta_{+3} f_t^{+3} + \delta_{+5} f_t^{+5} + \omega_{-1} d_s f_t^{-1} + \tau_{+1} d_s f_t^{+1} + \tau_{+3} d_s f_t^{+3} + \tau_{+5} d_s f_t^{+5} + \varepsilon_{it},$$

for $t \in \{-3, -1, +1, +3, +5\}$. In the equation, y_{it} stands for raw test scores for child i at time t, α_i is a child fixed effect, and δ_t represents time fixed effects. In addition, d_s is a dummy variable equal to 1 if the child ever experiences family disruption during the observation period. Furthermore, $f_t^{-1}, f_t^{+1}, f_t^{+3}, f_t^{+5}$ are year-dummies for the 2nd, 3rd, 4th, and 5th observation period, respectively, measured in $t = -1, t = +1, t = +3$, and $t = +5$, respectively. Lastly, ε_{it} is an error term. Since the effect of family disruption can be heterogeneous through time, I estimate four event effects: ω_{-1}, τ_{+1}, τ_{+3}, τ_{+5}. Results are presented in Table 7.

Table 7: TWFE (Family Disruption, Long-Term)

	Coefficient compared to t–3				
Time	**t − 3**	**t − 1**	**t + 1**	**t + 3**	**t + 5**
DD Estimate	**0**	**-0.76**	**-2.19****	**-3.64*****	**-3.48*****
(FamDisrup-TwoParent & versus t-3)		(0.83)	(0.97)	(1.12)	(1.11)
Obs (balanced panel): 5,275					
Experiencing Fam Disrup: 12.7%					
R^2: 0.88; Adj. R^2: 0.85					

Note: *p<0.1; **p<0.05; ***p<0.01

The results in Table 7 indicate significant negative effects of a family disruption event at time $t = 0$ on children's test scores, relative to the baseline period three years prior to the event at $t - 3$. The difference-in-differences (DD) estimates are especially large and statistically significant at the three-year follow-up ($t+3$), where the effect of family disruption on test scores is a decrease of 3.64 raw points. In contrast, the coefficient for the year before the disruption ($t - 1$) is statistically indistinguishable from zero, suggesting that, in the absence of the disruption, the treated and control groups were likely following similar trends over time.

Next, I replicate the analysis on a larger sample comprising all children for whom at least three consecutive test score observations are available. I use only three observation periods, without yet accounting for residential moves, to assess how the estimates compare when the sample is restricted in this way. This serves as a baseline for the next step of the analysis, in which I distinguish between movers and stayers within the same three-period sample. Requiring five consecutive test scores would substantially reduce the sample size, as most children take these tests only three times in a row. Consequently, imposing a five-year observation window would then also decrease the number of children who experience both a family disruption and a residential move. Limiting the sample to three periods therefore helps preserve statistical power. Moreover, as shown in Table 7, there is no evidence of significant pre-disruption effects, so I use the year prior to the event ($t-1$) as the baseline period in this shorter panel. Results are presented in Table 8.

Table 8: TWFE (Family Disruption, Short-Term)

	Coefficient compared to t-1		
Time	$\mathbf{t-1}$	$\mathbf{t+1}$	$\mathbf{t+3}$
DD Estimate	**0**	**-1.44****	**-2.33*****
(FamDisrup-TwoParent & versus t-1)		(0.56)	(0.70)
Obs (balanced panel): 9,042			
Experiencing Fam Disrup: 13.2%			

Note: *p<0.1; **p<0.05; ***p<0.01

The estimates in Table 8 are slightly smaller in magnitude but otherwise consistent with those in Table 7. I also conducted this type of analysis with an intermediate-sized subsample consisting of all children for whom I have at least four consecutive test score observations available. Results are reported in Table 24 in *Appendix B* and closely align with the estimates presented in Table 8, both in magnitude and direction.

3.3.2 TWFE Model (FamDisrup×Move)

Next, I estimate the panel data model with child and time fixed effects separately for two subsamples: one consisting of movers and the other of stayers. The difference between these two sets of estimates forms the basis of the triple difference-in-differences (DDD) estimate. Results are presented in Table 9.

Table 9: TWFE (Family Disruption X Move, Short-Term)

	Coefficient compared to t-1	
MOVERS	**t + 1**	**t + 3**
DD Estimate (FamDisrup-TwoParent & versus t-1)	-3.19**	-3.24*
	(1.30)	(1.75)
Obs: 1,290		
Experiencing Fam Disrup: 17.4%		
STAYERS	**t + 1**	**t + 3**
DD Estimate (FamDisrup-TwoParent & versus t-1)	-0.68	-1.04
	(1.03)	(1.21)
Obs: 3,120		
Experiencing Fam Disrup: 11.9%		
Triple DD (DDD) Estimate:	**-2.51**	**-2.20**

Note: *p<0.1; **p<0.05; ***p<0.01

The results in Table 9 suggest that the impact of family disruption on test scores varies substantially by residential mobility. The DDD estimate is -2.51 in the year following the disruption and -2.20 three years after. While this "difference-in-differences-in-differences" value is difficult to interpret on its own, the table reflects a meaningful gap between movers and stayers: the effect of family disruption on test scores is more than three times larger – and statistically significant – for children who moved compared to those who remained in place.

3.3.3 TWFE Model (FamDisrup×DistMove)

Similar as for the study based on the OLS regressions, I also conduct an analysis that accounts for the distance of the move, rather than simply whether the child moved. Preliminary results indicate that the effects of family disruption on academic outcomes are more detrimental for moves greater than one mile and less pronounced for moves of one mile or less, suggesting that the distance of the move may amplify its disruptive impact. However, given the limited number of observations and resulting concerns about statistical power, I do not formally report these estimates here. But, I do a separate related analysis, where I distinguish between children who change ZIP codes following a family disruption and those who remain within the same ZIP code, a category that includes both non-movers and short-distance movers. These estimates are reported in Table 10 and suggest that the negative effect of family disruption on test scores is approximately twice as large for children who move to a different ZIP code compared to those who stay within the same ZIP code.

Table 10: TWFE (FamDisrup X ChangeZIP), 3 Observation Periods

	Coefficient compared to t-1	
CHANGE ZIP	**t + 1**	**t + 3**
DD Estimate (FamDisrup-TwoParent & versus t-1)	-3.37**	-3.80**
	(1.42)	(1.87)
Obs: 1,005		
Experiencing Fam Disrup: 18.2%		
SAME ZIP	**t + 1**	**t + 3**
DD Estimate (FamDisrup-TwoParent & versus t-1)	-1.57*	-1.99*
	(0.85)	(1.07)
Obs: 4,179		
Experiencing Fam Disrup: 12.6%		
Triple DD (DDD) Estimate:	**-1.80**	**-1.81**
Note:	*p<0.1; **p<0.05; ***p<0.01	

Note: Effect of family disruption on test scores is $\sim 2x$ higher for those changing ZIP's compared to those staying within the same ZIP. Note that staying within the same ZIP includes both stayers and short-distance movers who move within their ZIP.

3.4 Diagnostics of the Reduced Form Analysis

The descriptive regressions presented in this section – including both OLS and TWFE models – provide useful insights into the relationship between social disruptions and children's test scores. However, these estimates are subject to several limitations. First, they may suffer from selection bias, as parental separation is not randomly assigned. As shown in the summary statistics, families experiencing a disruption differ systematically from more stable families, raising concerns that pre-existing differences—not the disruption itself—may drive the observed outcomes.

Moreover, these models may be affected by omitted variable bias. If unobserved factors influence both the likelihood of separation and children's academic performance, the estimated effects of family disruption may be either overstated or understated. While TWFE models help control for time-invariant unobserved heterogeneity and common shocks, they do not address time-varying confounders or heterogeneous treatment effects. Additionally, the current specifications focus on a narrow comparison group—children experiencing a disruption between periods 2 and 3—rather than leveraging the full sample. A staggered event study framework can better utilize variation in the timing of treatment across individuals.

The choice of the control group further complicates interpretation. In these models, the control group consists of children in two-parent households who remain intact throughout the observation period. Yet this may not represent the appropriate counterfactual, as some of these families may face internal tensions or may eventually separate outside the observation window. A more appropriate comparison involves children who are not yet treated, allowing for a cleaner identification of the causal effect of transitioning from a two-parent to a single-parent household.

To address these challenges and move toward causal inference, I adopt a modern event-study design in the next section. Specifically, I implement the staggered difference-in-differences estimator developed by Callaway and Sant'Anna (2021), which accommodates variation in treatment timing and enables the use of not-yet-treated units as controls. This framework allows for dynamic treatment effects and pre-trend diagnostics, and is well-suited to studying disruptions that occur at different times and ages across children in the sample.

4 Econometric Approach

In this chapter, I use a dynamic within-child difference-in-differences approach à la Callaway and Sant'Anna (2021) to compare longitudinal test scores of children who experience family disruption, and potential associated change of residence, to those of children who have not yet been affected. The estimator allows for staggered treatment and heterogeneous treatment effects and will quantify the causal effects of family disruption, and potential associated change in the place of residence, on test scores. In what follows, I provide the framework of analysis.

Balanced Panel. The sample I consider has no missing values for any child or year and can be described as $\{(Y_{i1}, Y_{i2}, ..., Y_{iT}, D_{i1}, D_{i2}, ..., D_{iT}, \mathbf{X}_i)\}_{i=1}^N$, where Y_{it} represents PIAT scores for child i at time t. Since children take these biennial tests at most five times, I can follow a child's performance over a period for up to ten years. In my initial analysis, I exploit this longest panel possible, $T = 5$, and eliminate all children with fewer observations. The sample consists of $i = 1, ..., N$ children and includes the following time-invariant covariates, $\mathbf{X}_i$, which must be measured before treatment takes place: race, age at first observation, presence of siblings, mother's years of education, and mother's AFQT score. Lastly, D_{it} is a dummy variable taking the value 1 if child i experienced family disruption, and 0 otherwise. I design it so that once a child is treated, it stays so throughout the rest of the time.

Parameter of Interest. The main parameter that I try to recover in this paper is the "average treatment effect on the treated", ATT, which is the average effect on test scores for those children who experienced family disruption. In particular, I am interested in the average treatment effect at time period t for children that experienced the family disruption event in period g, $ATT(g, t)$. Children in my first subsample can either be affected in period 2, 3, 4, or 5. I combine all children who experienced family disruption in the same period, resulting in four groups (or "cohorts"). Each of these groups is going to have their own $ATT(g, t)$. The main building block of the paper can be summarized with the following equation (Callaway and Sant'Anna, 2021):

$$ATT(g, t) = \mathbb{E}\left[Y_t(g) - Y_t(0) | G_g = 1\right], \quad \text{for t} \geq g,$$

which is the *group-time average treatment effect*, i.e., the ATT for children who are members of a particular "treatment" group g at a particular time t, $ATT(g, t)$. So, the ATT is allowed to vary across groups g and calendar time t, $ATT(g, t)$, and estimates will show whether and by what magnitude. Since the maximum amount of biennial tests that one can take is five, all children in my baseline sample are roughly of same age when entering the panel in period 1 (roughly six years old) and taking their first PIAT tests as well as when taking their fifth PIAT tests and subsequently leaving the panel. Thus, the four cohorts (groups), which are created according to the time of treatment, represent children of different ages. For instance, children who experienced family disruption in period 2, and therefore belong to group 2, are six years younger than children who experienced family disruption in period 5, and therefore belong to group 5. On top of potential variation across groups who are of different average age, the $ATT(g,t)'s$ can also vary across time to treatment, i.e., depend on the distance to treatment. For example, the instantaneous effects of family disruption on test scores can be stronger than the one five years after the event.

However, in order to get the overall ATT, I then also have to aggregate the $ATT(g,t)'s$ across groups and time. In the $ATT(g,t)$ equation above, I fix a group g, $G_g = 1$, and I then have the treatment effect at time period t, $Y_t(g)$. While the expected value of Y_t among treated groups g, $\mathbb{E}[Y_t(g)|G_g = 1]$, is observed from the data, the challenge is to recover $\mathbb{E}[Y_t(0)|G_g = 1]$. To generate this missing counterfactual measure in a difference-in-differences setup, two key assumptions must generally hold, although the first one can be relaxed (Callaway and Sant'Anna, 2021). Identification of the ATT is achieved via the following two main identification assumptions.

Assumption 1 (No Treatment Anticipation).
$\mathbb{E}[Y_t(g)|X, G_g = 1] = \mathbb{E}[Y_t(0)|X, G_g = 1]$ *almost surely* $\forall g \in \mathcal{G}, t \in T$, *such that* $t < g$.

This assumption states that, before treatment takes place ($t < g$), there is no treatment effect. And if no treatment, then the treatment effect for this group G is also zero. For example, test score performance outcome in period 1 is not affected by treatment status in period 2 (whether the child experiences family disruption or not). This assumption guarantees that the difference-in-differences estimand can be interpreted as a causal effect in period 2.

Even though this no treatment anticipation is a common requirement in empirical research, Callaway and Sant'Anna (2021)'s method allows to deal with cases where there is treatment anticipation. And in my context, it seems reasonable that some of these school-aged children were already informed or experienced tension between parents and anticipated a parental breakup before the actual event, which could have impacted their school performance due to the announcement or anticipation. However, there is a two-year gap between the immediate test scores taken before and after such a family disruption, and it is probably less likely that parents already announced their breakup two years before the event. In addition, Holm et al. (2023) find that periods of anticipation lasting more than one year before the event are not significant. Yet, I cannot be certain. Therefore, I will test this in the next section. First, I also have to impose a second assumption (Callaway and Sant'Anna, 2021).

Assumption 2 (Parallel Trends).

For each $(s, t) \in \{2, ..., T\} \times \{2, ..., T\}, g \in \mathcal{G}$ such that $t \geq g, s \geq t$,

$\mathbb{E}[Y_t(0) - Y_{t-1}(0)|X, G_g = 1] = \mathbb{E}[Y_t(0) - Y_{t-1}(0)|X, D_s = 0, G_g = 0]$ *almost surely.*

This is a parallel trends assumption conditional on observables and comparing the affected children to a control group consisting of children who are not yet affected at the time of the comparison, meaning that they could either be affected later or never during the sample observed. To be more specific, *Assumption 2* states that, conditional on a vector of covariates X, the evolution of the outcome in the absence of treatment, $Y_t(0) - Y_{t-1}(0)$, among this group g, is the same as among the not-yet-treated children on average, where $D_s = 0$ means not-yet-treated by time s. The notation $(t \geq g, s \geq t)$ restricts parallel pre-trends, as required when using not yet treated children as control group (Callaway and Sant'Anna, 2021).

Nonparametric Identification. Given both restrictions, *Assumption 1* and *Assumption 2*, hold, I can then nonparametrically point-identify the group-time average treatment effects, $ATT(g,t)'s$, via the following representation (Callaway and Sant'Anna, 2021):

$$ATT(g,t) = \mathbb{E}\left[\left(\frac{G_g}{\mathbb{E}[G_g]} - \frac{\frac{p_{g,t}(X)(1-D_t))}{1-p_g(X)}}{\mathbb{E}\left[\frac{p_{g,t}(X)(1-D_t))}{1-p_{g,t}(X)}\right]}\right)(y_t - y_{g-1} - \mathbb{E}[y_t - y_{g-1}|X, D_t = 0, G_g = 0])\right]$$

Even though I have multiple groups of children affected by family disruption at different times and multiple periods of test score observations around such a potential disruption event in my sample, the subsetting of the data into $ATT(g,t)'s$ is letting me regard each $ATT(g,t)$ as a two-group and two-period case. So, the comparison group changes over time. Thus, I have two groups (or "cohorts"): G_g representing children from affected group g and the group not yet affected, D. Plus, I have two time periods (t and $g-1$), as well as stabilizing weights, $Y_t - Y_g - m(X)$. Also, $p_{g,t}(X)$ is the time-varying propensity score of being in group g as a function of the vector of additional control variables, X. I am going to model both the probability of belonging to the treatment group, denoted by the propensity score $p_g(X)$, and I am going to put more weights on children who are more similar to the group's g, G_g. So, $p_g(X)$ reweights the comparison groups (D) such that they look similar to the cohort g, G_g. Further, it includes an outcome regression model, $E(y_t - y_{g-1}|X, D_t = 0, G_g = 0)$, which models the outcome evolution of the not-yet affected children. Implicitly, $y_t - y_{g-1} - E(y_t - y_{g-1}|X, D_t = 0, G_g = 0)$ is a difference-in-difference with recentered covariates ("residuals"): $y_t - y_{g-1} - E(y_t - y_{g-1}|X, D_t = 0, G_g = 0)$, where $E(y_t - y_{g-1}|X, D_t = 0, G_g = 0)$ is the outcome regression. And then I am reweighting the residuals with the large bracket on the left-hand side, so that I am balanced. See Callaway and Sant'Anna (2021) for more details. The idea is that, as long as my model for the propensity score, $p_g(X)$, *or* my model for the outcome evolution, $E(y_t - y_{g-1}|X, D_t = 0, G_g = 0)$, are correctly specified, this is going to recover the ATT. Since this is giving me two "chances" to estimate the ATT, this estimator is also called the "doubly-robust" estimator. This estimator by Callaway and Sant'Anna (2021) extends Heckman et al. (1998); Abadie (2005); Sant'Anna and

Zhao (2000). Without covariates, Callaway and Sant'Anna (2021)'s procedure would be similar to Sun and Abraham (2001) and de Chaisemartin and D'Haultfœuille (2020).

Aggregation. To enhance precision and regain statistical power, I can aggregate the subsetted data to construct a unified overall $ATT(g,t)$. The idea is to calculate a weighted average by summarizing the $ATT(g,t)$ components for every group g and every time period t. In particular, Callaway and Sant'Anna (2021) propose taking the weighted averages of the $ATT(g,t)'s$ of the following form:

$$\sum_{g=2}^{T}\sum_{t=2}^{T}\mathbb{1}\{g \leq t\} w_{gt} ATT(g,t),$$

where the indicator function guarantees that I have zero weights before treatment takes place and non-zero weights afterwards – because the goal is to average whatever happens post-treatment (and not pre-treatment). Aggregation then follows by computing weighted averages of the $ATT(g,t)'s$ with given weights w_{gt} that can vary across groups g and across time t.

While there are different types of weights possible, for my baseline analysis, I focus on event-study type weights (or dynamic treatment effect weights). This weighting method seems most feasible, because the effect of family disruption on test scores may depend on the length of exposure to a single-parent household, so I want to emphasize how treatment effects may vary with elapsed treatment. However, I also check alternative weighting methods as robustness. Average effect of participating in the treatment for the group of children that have been exposed to the treatment for exactly e time periods:

$$\theta_D(e) = \sum_{g=2}^{T}\mathbb{1}\{g + e \leq T\} ATT(g, g+e) P(G = g | G + e \leq T, C \neq 1)$$

The time period is now reparameterized to $t = g + e$, where e stands for time since treatment took place. So, $e = 0$ would be the instantaneous treatment effect, $e = 1$

would be the treatment effect 1 period after treatment, etc. Since there is a two-year gap between any two consecutive observations, $e = 0$ can be interpreted as one year after family disruption or two years after the previous observations (where they were still living in a two-parent household). I then average the $ATT(g,t)'s$ for a fixed e across different groups $G = g$. I will also report robustness checks based on two alternative weighting methods.

First, I undertake an aggregation that takes the time average of these $ATT(g,t)'s$ for each group. This gives me the average effect of "participating" in the treatment of family disruption that children in group g experienced, which allows me to highlight heterogeneity across groups. Second, I fix time period t and average across groups. This allows me to highlight heterogeneity across time periods (calendar-time heterogeneity). The result is an average effect of participating in the treatment in time period t for groups that have participated in the treatment by time period t.[9] For estimation, Callaway and Sant'Anna (2021) allow to choose between the outcome regression, the inverse probability weighting, and the doubly-robust one.

5 Results

In this section, I present the main results. I begin by examining the effects of family disruption on children's test scores, focusing on how these effects vary depending on the timing of test score observations relative to the disruption event. Since children take these tests at most five times, and I restrict the sample to those observed in all five periods for the initial analysis, the children in this analysis are approximately the same age across cohorts. However, I account for variation in the timing of the disruption—children may experience family disruption at different points across the five observation periods, meaning the age at which they are affected can differ.

[9] An alternative way of combining $ATT(g,t)'s$ across g and t could just be to take a simple weighted average (give weight 1 to everybody, i.e., everybody has the same weight such that these weights sum up to 1). However, this is not very appealing as some groups are much larger than others (group g could be $10x$ bigger than group g', so cohort g deserves to be more "informative"). An alternative would be to attach weights on the size of group g. However, earlier groups get larger weights.

I also conduct this baseline analysis using a broader sample of children observed in at least three consecutive waves. Most children in the dataset take tests only three times in a row, and using this sample increases statistical power and ensures comparability with later analyses. This is especially important because I later examine the interaction between two potential forms of social disruption: family disruption and residential relocation. Requiring five consecutive test score observations would result in too few children who experience both events, limiting the analysis. While the five-period sample offers a longer-term view of how family disruption affects test scores before and after the event, the three-period setup provides a more robust foundation for identifying differences between children who move and those who stay following a family disruption. In subsequent sections, I further refine the treatment groups by incorporating information on the distance of relocation.

5.1 Effects of Family Disruption on Test Scores

I estimate the effects of family disruption on test scores using the staggered difference-in-differences framework proposed by Callaway and Sant'Anna (2021). In order to provide evidence of the aggregate effects of family disruption on test scores, the idea is to aggregate group-time average treatment effects, $ATT(g, t)$'s. This group-time treatment effect framework fits well with my setting because children experience family disruption at different time periods (ages), and the impact on test scores in t may depend on how close the test observation is to the disruption event in g. So, on top of an overall average treatment effect, which averages the treatment effects across all lengths of exposure to the family disruption event, this analysis sheds light on how average treatment effects evolve over time, capturing variation in test scores both before and after the disruption. Within the 10-year observation period, 39% of the children experience family disruption. The results of the effects of family disruption on test scores, both overall and by length of exposure, are reported in Table 11.

Table 11: Dynamic Aggregation of Group-Time Average Treatment Effects

Event Time	Family Disruption	Family Disruption
-3	0.29	0.34
	(0.64)	(0.68)
-2	-0.36	-0.30
	(0.58)	(0.57)
-1	-0.25	-1.14
	(0.43)	(0.42)
0	-0.96	-0.74
	(0.40)	(0.38)
+1	-1.70^{*}	-1.15
	(0.51)	(0.51)
+2	-2.79^{*}	-1.50
	(0.62)	(0.62)
+3	-1.97	0.08
	(0.93)	(0.81)
Overall ATT:	-1.85^{*}	-0.83
	(0.48)	(0.46)
Controls		✓
# Obs	7,345/5 = 1,469 kids	
Fam Disrup	39%	

Table 11 reports the dynamic effects of family disruption on raw PIAT reading comprehension scores across all available time distances to the disruption event, as well as the overall average treatment effect ("Overall ATT"). Results are shown for an analysis without and one with controls. Estimates suggest that family disruption lowers children's test scores. However, when controls are included, the estimates are no longer statistically significant. Importantly, this holds for the ATT: when controls are included, the average effect remains a reduction of 0.83 raw points, but the estimate is no longer statistically significant.

To verify that I identify unbiased treatment effects, I require the classic parallel trends assumption to hold, that in the absence of family disruption, test scores for affected and unaffected children would have evolved in a similar fashion. The parallel trends assumption cannot be directly tested because I do not observe the counterfactual outcome for treated children. But I can test whether groups of observations have a different trend with respect to the amount of time left until the disruption event in the pre-treatment period. Figure 1 aggregates group-time effects into an event study plot to illustrate the average effect by length of exposure to the family disruption event for the baseline case including control variables. The graph confirms that there are no differential trends in the pre-treatment period between groups that are periods away from being disrupted and other groups that are not yet disrupted. This provides evidence in support of the parallel trends assumption. Since affected children do not trend differently from not yet affected children in the periods leading up to the disruption, we should not expect them to have trended differently in the absence of treatment. The fact that treated and untreated children only diverge once the treatment occurs provides strong evidence that family disruption is causing the change. The figures also show no evidence of anticipation. No evidence of anticipatory effects prior to the implementation date of the separations could be due to lack of awareness. It is plausible that most potential children were unaware of the disruption until after they were affected. I can also examine the pre-treatment coefficients in Table 11. Since these pre-treatment estimates are small and insignificant, I can conclude that anticipation does not appear to be a concern. I have thus justified that my difference-in-differences specification does not need to explicitly account for anticipation effects.

It is important to note that in the analysis based on the sample with five observation periods—i.e., an observation every other year over a 10-year span—the staggered difference-in-differences approach enables a dynamic aggregation of group-time average treatment effects, ranging from five years before to seven years after the disruption event. While this setup offers a valuable long-term perspective on the effects of family disruption on test scores, the analysis becomes more robust when I restrict the sample to children for whom I observe at least three consecutive test scores instead of five. This broader sample increases statistical power and forms the basis for subsequent refinements, where I distinguish between movers and stayers in relation to family disruption.

I therefore now turn to estimates of the effects of family disruption on test scores for all children observed in at least three consecutive waves. Table 12 reports the results.

Table 12: Dynamic Aggregation of Group-Time Average Treatment Effects

Event Time	Family Disruption
-1	-0.16
	(0.49)
0	-0.75^{*}
	(0.34)
$+1$	-1.22
	(0.55)
Overall ATT	-0.99^{*}
	(0.38)
Controls	✓
# obs	9,762/3=3,254
Treated	22%

Note: Test scores and family composition are observed every other year and family disruption happens in year t, which is between two such observed test score and family stability measures. Thus, event time -1 translates into $t-1$, event time 0 translates into $t+1$, and event time $+1$ translates into $t+3$.

While the estimates for the three-observation-period sample are similar in magnitude to those from the five-observation-period sample, they are slightly stronger and statistically significant. This suggests that family disruption has immediate adverse effects on test scores. Furthermore, if I restrict my sample to those children for whom I observe at least four consecutive test scores, the *Overall ATT* is in line and significant (see Table 25 in *Appendix B*). I will now refine the analysis by distinguishing between movers and stayers among those who experience family disruption.

5.2 Effects of (Family Disruption × Move) on Test Scores

In the previous section, I presented the difference-in-differences estimates for the unconditional sample, irrespective of whether children moved or not, to calculate the general effects of family disruption on test scores. In this section, I examine the effects of family disruption and simultaneous residential relocation on children's test scores. To better understand the interaction between these two factors, I focus on a triple difference-in-differences framework. Specifically, I aim to compare the effects of family disruption on test scores in two distinct samples: movers and stayers.

First, among movers, I compare children who experience family disruption with those from stable families who also relocate. This comparison isolates the effect of family disruption among children who move. Second, among stayers, I estimate the effect of family disruption on test scores by comparing children who experience disruption with those who do not. This allows me to isolate the effect of family disruption for children who remain in their original residence. Finally, by comparing the results from the two samples—movers and stayers—I can assess the interaction effects of family disruption and residential relocation on children's test scores, effectively providing insight into how these two factors jointly influence outcomes. This analysis follows the structure of a triple difference-in-differences approach. Table 13 presents the results of the intermediate step, where, within the sample of movers, I estimate the effects of family disruption (disrup(cond)) compared to stable families who also relocate but remain in a two-parent household (stable(cond)). The differences of these estimates gives me the effects of family disruption on test scores for movers compared to stayers (disrupXmove).

Table 13: Effects of (Family Disruption × Move) on Test Scores

Event Time	(2) disrup(cond)	(3) stable(cond)	(4) disrupXmove
t−1	−0.89	−0.17	−0.72
	(1.34)	(0.67)	(0.67)
t+1	−0.58	0.95	−1.53
	(0.84)	(0.45)	(0.39)
t+3	−0.47	1.77*	−2.24*
	(1.32)	(0.56)	(0.76)
Overall ATT	−0.53	1.36*	−1.89*
	(0.94)	(0.41)	(0.53)
Controls	✓	✓	✓
# obs	3,015/3=1,005	4,296/3=1,432	
Treated	11%	38%	

The *Overall ATT* in Table 13 suggests a decrease of 1.89 raw points ($-0.53 - 1.36 = -1.89$), which is significant. This implies that, among movers, children experiencing family disruption experience a drop of almost 2 raw test score points compared to children in stable two-parent households. Interestingly, while children who move in association with family disruption experience adverse outcomes, those from stable families tend to perform better after a move than before. This suggests that the effects of relocation depend on the reasons behind the move, highlighting the endogenous nature of mobility.

I do a similar analysis based on the sample of stayers only and compare how test scores of children experiencing family disruption are affected compared to stable two-parent households (disrupXstay). The estimates of this investigation are reported in Table 14, together with the unconditional effects of family disruption on test scores (disrupXall) as well as those for the sample of movers (disrupXmove).

Table 14: Effects of (Family Disruption × Move) on Test Scores

Event Time	(1) disrupXall	(4) disrupXmove	(5) disrupXstay
t–1	−0.16	−0.72	0.89
	(0.49)	(0.67)	(0.78)
t+1	−0.75*	−1.53	−0.96
	(0.34)	(0.39)	(0.61)
t+3	−1.22	−2.24*	−0.08
	(0.55)	(0.76)	(0.94)
Overall ATT	−0.99*	−1.89*	−0.52
	(0.38)	(0.53)	(0.60)
Controls	✓	✓	✓
# obs	9,762/3=3,254		3,456/3=1,152
Treated	22%		22%

Table 14 now disentangles the effects of family disruption on test scores (disrupXall) and gives a full picture of how children experiencing family disruption in association with a move (disrupXmove) fare compared to children experiencing family disruption staying (disrupXstay). The difference of the *Overall ATT*'s in columns (4) and (5) results in the triple difference-in-differences, the *Triple Overall ATT*.

The unrestricted effect of family disruption on children's test scores for the considered sample with three observation periods is a decrease of 1 raw test score point, as reflected by the *Overall ATT* in the first column (1). When family disruption is associated with a move (middle column 4), then that implies a decrease of 1.9 raw test score points. Both are significant. However, for stayers experiencing family disruption (right column 5), the decrease in raw test score points is 0.5 only and not significant. In a triple difference-in-differences interpretation fashion, among children who experience family disruption, those who move suffer a significantly larger negative effect (−1.37 points worse) than those who stay. Overall, this analysis indicates that while family disruption negatively impacts children's test scores, the adverse effect is primarily concentrated among those who move. This suggests that residential relocation acts as a moderating variable, amplifying the negative impact of family disruption on children's test scores. For stayers, the effects of family disruption are not significant, highlighting that it is the combination of family disruption and residential relocation that most strongly influences children's test scores. Next, I further distinguish between different distances of the move.

5.3 Effects of (Family Disruption $\times$ Distance of Move) on Test Scores

In addition to examining whether a move is associated with family disruption, I further account for the distance of the move. Specifically, I distinguish between two mover samples: Short-Distance Moves and Long-Distance Moves. The Long-Distance sample includes movers whose new residence is located more than one mile from their previous home, while the Short-Distance sample includes those who moved no more than one mile. Table 15 presents the results separately for (a) Short-Distance Moves and (b) Long-Distance Moves. Column (1) presents the estimates for the effects of family disruption on test scores for the entire sample, irrespective of whether there is an associated move or not. The differences between estimates in column (2) and (3) yield those in column (4), which represent the effects of experiencing family disruption among movers. Lastly, column (5) shows the effects of family disruption on test scores for the sample of stayers. The differences of the *Overall* ATT's in columns (4) and (5) result in triple difference-in-differences, the *Triple Overall* ATT's.

Table 15: Distance of Moves: Short vs. Long

(a) Short-Distance Moves

Event Time	(1) disrupXall	(2) disrup(cond)	(3) stable(cond)	(4) disrupXmove	(5) disrupXstay
t–1	-0.16	1.39	-0.04	1.43	0.89
	(0.49)	(2.48)	(1.04)	(1.44)	(0.78)
t+1	-0.75*	-0.11	1.06	-1.17	-0.96
	(0.32)	(2.21)	(1.03)	(1.18)	(0.55)
t+3	-1.22	0.82	1.76	-0.94	-0.08
	(0.52)	(2.74)	(1.22)	(1.52)	(0.92)
Overall ATT	-0.99*	0.35	1.41	-1.06	-0.52
	(0.39)	(2.04)	(0.93)	(1.11)	(0.63)
Controls	✓	✓	✓	✓	✓
# obs	9,762/3=3,254	2,742/3=914	3,003/3=1,001		3,456/3=1,152
Treated	22%	2%	11%		22%

(b) Long-Distance Moves

Event Time	(1) disrupXall	(2) disrup(cond)	(3) stable(cond)	(4) disrupXmove	(5) disrupXstay
t–1	-0.16	-1.25	-0.18	-1.07	0.89
	(0.50)	(1.44)	(0.76)	(0.68)	(0.76)
t+1	-0.75*	-0.67	0.93	-1.60	-0.96
	(0.32)	(0.83)	(0.47)	(0.36)	(0.57)
t+3	-1.22	-0.72	1.77*	-2.49*	-0.08
	(0.56)	(1.57)	(0.55)	(1.02)	(0.88)
Overall ATT	-0.99*	-0.69	1.35*	-2.04*	-0.52
	(0.36)	(1.07)	(0.42)	(0.65)	(0.62)
Controls	✓	✓	✓	✓	✓
# obs	9,762/3=3,254	2,958/3=986	3,978/3=1,326		3,456/3=1,152
Treated	22%	9%	33%		22%

Table 15 presents results by distance of the move, distinguishing between short- and long-distance relocations. The findings indicate that children who move more than one mile away in connection with family disruption experience a substantially larger and statistically significant decline in test scores (a decrease of 2.04 raw points). In contrast, those who move within a one-mile radius exhibit a smaller decrease (1.06 raw points), which is not statistically significant. For stayers, the decrease amounts to a decrease of 0.52. Consequently, the *Triple Overall ATT*'s differ by pretty much 1 raw point. This suggests that the disruptive impact of moving on children's academic performance may be more pronounced when the move involves a greater geographic distance. This might be due to accompanying changes such as switching schools, losing access to familiar peer networks, or transitioning to a different neighborhood environment.[10]

6 Sensitivity Analysis

In this section, I conduct a series of robustness checks. First, I present additional estimates for a subsample of children who experienced family disruption through parental separation. Next, I report results using an alternative method to aggregate the $ATT(g,t)$ estimates. Finally, I provide estimates based on Wooldridge (2021)'s approach, which proposes a Difference-in-Differences specification for staggered adoption similar to the one employed in my main analysis.

6.1 Family Disruption Consisting of Parental Separation Only

I now restrict the treatment group to cases where the parents separate, excluding other forms of family disruption such as paternal death or incarceration. This narrower definition focuses specifically on situations in which the father remains present but no longer resides in the same household. In other words, I capture instances where a two-parent household splits into two separate households, rather than broader disruptions that also result in the mother and child living together. As a result of this more selective definition, the number of treated children in the sample decreases. Within the 10-year

[10]In fact, analyzing a smaller subsample supports this finding: the further children must relocate due to family disruption, the more adversely they are affected. However, beyond a certain threshold, the distance no longer appears to have an impact. Given the small size of this subsample and the potential limitations in statistical power, I refrain from including a more granular distinction based on the distance of the move in the current paper.

observation period, 28% of the children experience a parental relationship dissolution. Due to the smaller sample size, I no longer distinguish moves by distance. However, I do continue to account for whether a move occurs in connection with the parental separation, in addition to the separation itself. Table 16 presents the effects of parental separation on children's test scores for the sample of children with five consecutive test score observations. Table 17 reports the same effects for children with at least three consecutive test score observations and additionally includes estimates of the interaction between parental separation and a residential move.

Table 16: Group-Time Average Breakup Effects (Breakup)

Event Time	Breakup	Breakup
−3	−0.06	−0.08
	(0.84)	(0.83)
−2	−0.55	−0.29
	(0.75)	(0.78)
−1	−0.32	−0.10
	(0.59)	(0.59)
0	−1.46*	−1.05
	(0.48)	(0.48)
+1	−1.99*	−1.14
	(0.70)	(0.64)
+2	−3.02*	−1.41
	(0.86)	(0.83)
+3	−2.61	−0.29
	(1.12)	(1.07)
Overall ATT:	−2.27*	−0.97
	(0.58)	(0.55)
Controls		✓
# Obs	6,240/5 = 1,248 kids	
Breakup	28%	

Table 17: Group-Time Average Breakup Effects (Breakup × Move)

Event Time	(1) breakupXall	(2) breakup(cond)	(3) stable(cond)	(4) breakupXmove	(5) breakupXstay
t–1	-0.38	0.73	-0.04	0.77	0.34
	(0.61)	(1.56)	(0.64)	(0.92)	(1.13)
t+1	-0.24	-0.44	0.94	-1.38	0.16
	(0.39)	(0.99)	(0.42)	(0.57)	(0.74)
t+3	-0.63	0.86	1.73*	-0.87*	0.47
	(0.59)	(1.48)	(0.55)	(0.93)	(1.20)
Overall ATT	-0.43	0.21	1.34*	-1.13*	0.31
	(0.41)	(1.07)	(0.42)	(0.65)	(0.83)
Controls	✓	✓	✓	✓	✓
# obs	9,093/3=3,031	2,904/3=968	4,266/3=1,422		3,060/3=1,020
Treated	16%	8%	37%		13%
R^2	0.67	0.68	0.68		0.68

Qualitatively, the results accounting for breakups only, as shown in Tables 16 and 17, are similar to those in the main analysis, which includes all types of family disruption cases. But for the case with three observation periods, the results reported here are slightly smaller in magnitude and are estimated with less precision. To be specific, for breakups only, the *Overall ATT*, which averages the treatment effects across all lengths of exposure to the family disruption event, is −0.43 and not significant, while for all types of family disruption cases, the *Overall ATT* was −0.99 and significant. On the other hand, for all children for whom I have five observation periods, the *Overall ATT* for breakups is slightly larger in magnitude (−0.97) compared to the analysis that was based on all family disruption types (−0.83), but not significant in both cases. Furthermore, if I restrict my sample to those children for whom I observe at least four consecutive test scores, the *Overall ATT* measuring the effect of a breakup on test scores is also in line (see Table 25 in *Appendix B*).

The estimates of the effect of a parental breakup on children's test scores also rely on the parallel trends assumption. I support this assumption by finding that, even though families that experience a breakup are observably different, pre-treatment trends are similar. Figure 2 aggregates group-time effects into an event study plot to visualize the average effect by length of exposure to the breakup event, including control variables. This graph confirms that there are no differential trends in the pre-treatment period between groups that are periods away from being separated and other groups that are not yet separated.

6.2 Alternative Aggregation of the ATT(g,t)'s

In this subsection, I perform an additional aggregation of the $ATT(g,t)$'s to derive the *Overall ATT*. I take the average effect by time period to estimate the effects of family disruptions and breakups, respectively, on children's test scores. To be more specific, I calculate the average effect of participating in the treatment in a particular time period for all groups that participated in the treatment in that time period. Details are in Table 18.

Table 18: Average Effect by Time Period (Family Disruption; Breakup)

Time Period	(1) Family Disruption	(2) Family Disruption	(3) Breakup	(4) Breakup
2	-1.57^{*}	-0.79	-2.34^{*}	-1.43
	(0.59)	(0.62)	(0.72)	(0.65)
3	-1.83^{*}	-1.18	-2.05^{*}	-1.20
	(0.58)	(0.59)	(0.80)	(0.77)
4	-1.91^{*}	-1.02	-2.22^{*}	-0.99
	(0.55)	(0.50)	(0.63)	(0.64)
5	-1.40^{*}	-0.75	-1.90^{*}	-0.92
	(0.51)	(0.50)	(0.64)	(0.57)
Overall ATT	-1.68^{*}	-0.94^{*}	-2.13^{*}	-1.14^{*}
	(0.40)	(0.38)	(0.52)	(0.50)
Controls		✓		✓
# obs	7,345/5 = 1,469 kids		6,240/5 = 1,248 kids	
Experiencing Fam Disrup	39%		28%	

Table 18 reports the average effects by time period. For example, for time period 2, estimates show the average effect among those who experienced disruption in the second time period, which is group 2 only. For time period 3, it then shows the average effect among those who experienced disruption in time periods 2 or 3, which is the average of groups 2 and 3. Consequently, for time period 4, it shows the average for groups 2, 3, and 4. Lastly, for time period 5, it shows the average for groups $2-5$. The *Overall ATT*'s in Table 18 are all negative and significant, indicating that family disruption (parental breakup) decreases test scores by 0.94 (1.14) raw points.

Now, I turn to the larger sample consisting of all children for whom I observe at least three test scores in a row. Again, I calculate the average effect of experiencing family disruption or parental breakup only, respectively, in a particular time period, for all groups that experienced it by that time period. In addition, I compare them to the average effects of experiencing both family disruption (breakup) and residential relocation. These estimates are reported in Tables 19 for family disruption and 20 for breakups, respectively, and are in line with previous results.

Table 19: Average Effect by Time Period (Family Disruption × Move)

Time Period	(1) disrupXall	(2) disrup(cond)	(3) stable(cond)	(4) disrupXmove	(5) disrupXstay
2	-0.80	-1.24	1.03	-2.27	-0.25
	(0.43)	(1.01)	(0.50)		(0.85)
3	-0.99*	-0.06	1.43*	-1.49	-0.88
	(0.38)	(0.94)	(0.48)		(0.61)
Overall ATT	-0.89*	-0.65	1.23*	-1.88	-0.57
	(0.36)	(0.84)	(0.41)		(0.58)
Controls	✓	✓	✓		✓
# obs	9,762/3=3,254	3,015/3=1,005	4,296/3=1,432		3,456/3=1,152
Treated	22%	11%	38%		22%

Table 20: Average Effect by Time Period (Breakup × Move)

Time Period	(1) disrupXall	(2) disrup(cond)	(3) stable(cond)	(4) disrupXmove	(5) disrupXstay
2	-0.25	-0.30	1.11	-1.41	0.98
	(0.53)	(1.18)	(0.53)		(1.03)
3	-0.45	0.38	1.35*	-0.97	-0.09
	(0.45)	(1.16)	(0.50)		(0.78)
Overall ATT	-0.35	0.04	1.23*		0.44
	(0.42)	(0.99)	(0.42)	-1.19	(0.77)
Controls	✓	✓	✓		✓
# obs	9,762/3=3,254	3,015/3=1,005	4,296/3=1,432		3,456/3=1,152
Treated	22%	11%	38%		22%

Overall, estimates from this alternative aggregation method confirm the findings from the baseline analysis. Family disruption has significant adverse effects on children's test scores, primarily driven by children who move following the disruption. In contrast, the effects for children who stay in the same location are small and statistically insignificant.

6.3 Alternative Difference-in-Differences Specification for Staggered Adoption

In this subsection, I also conduct a series of sensitivity analyses using the alternative Difference-in-Differences specification for staggered adoption proposed by Wooldridge (2021). For all children with five consecutive test scores, the average effects by time period are reported in Table 21, separately for general family disruption and breakup cases. The same analysis based on a larger sample of children with at least three consecutive test scores is presented in Tables 22 and 23, which also report interaction effects with residential relocation. Finally, Figure 3 in *Appendix B* also explores heterogeneity in the effects of breakups by displaying ATT's by sex and race.

Table 21: W: Average Effect by Time Period (Family Disruption; Breakup)

Time Period	(1) Family Disruption	(2) Family Disruption	(3) Breakup	(4) Breakup
2	-1.11	-0.83	-1.66	-1.12
	(0.35)	(0.34)	(0.45)	(0.43)
3	-1.95	-1.39	-2.33	-1.35
	(0.50)	(0.48)	(0.62)	(0.58)
4	-2.83	-1.47	-3.13	-1.42
	(0.64)	(0.60)	(0.80)	(0.74)
5	-1.88	0.04	-2.49	-0.35
	(0.91)	(0.82)	(1.10)	(0.98)
Controls		✓		✓
# obs	7,345/5 = 1,469 kids		6,240/5 = 1,248 kids	
Treated	39%		28%	
R^2	0.69	0.76	0.69	0.76

Table 22: W: Average Effect by Time Period (Family Disruption × Move)

Time Period	(1) disrupXall	(2) disrup(cond)	(3) stable(cond)	(4) disrupXmove	(5) disrupXstay
2	-0.79	-0.73	0.93	-1.66	-0.78
	(0.32)	(0.75)	(0.42)		(0.53)
3	-1.23	-0.62	1.81	-2.43	-0.17
	(0.52)	(1.19)	(0.54)		(0.91)
Controls	✓	✓	✓		✓
# obs	9,762/3=3,254	3,015/3=1,005	4,296/3=1,432		3,456/3=1,152
Treated	22%	11%	38%		22%
R^2	0.67	0.68	0.68		0.67

Table 23: W: Average Effect by Time Period (Breakup × Move)

Time Period	(1) disrupXall	(2) disrup(cond)	(3) stable(cond)	(4) disrupXmove	(5) disrupXstay
2	-0.33	-0.29	0.94	-1.23	0.19
	(0.38)	(0.87)	(0.42)		(0.74)
3	-0.63	0.82	1.77	-0.95	0.39
	(0.60)	(1.38)	(0.54)		(1.18)
Controls	✓	✓	✓		✓
# obs	9,093/3=3,031	2,904/3=968	4,266/3=1,422		3,060/3=1,020
Treated	16%	8%	37%		13%
R^2	0.67	0.68	0.68		0.68

The results from these sensitivity analyses are broadly consistent with the findings from the main analysis. Family disruption is associated with adverse effects on children's test scores, and these effects appear slightly less pronounced in the case of parental separation. In line with the main results, the most substantial negative effects are concentrated among children who experience family disruption and who subsequently relocate, suggesting that residential moves may amplify the academic consequences of family disruption.

7 Conclusion

This paper examines the academic consequences of family disruption—including parental breakup, incarceration, or death—among school-aged children in the United States. I use a staggered difference-in-differences design that compares children who experience family disruption to those not yet treated, ensuring a relevant counterfactual. With parallel pre-trends holding, the method allows for causal interpretation. Across all specifications, I find that family disruption has a negative effect on children's standardized test scores, although the magnitude and statistical significance of the effect depend on the sample at hand. In the sample of children with at least three observation periods, the estimated effect is relatively large and statistically significant. In the smaller sample of children with five observation periods, the estimated effect is still negative but slightly smaller in magnitude and not statistically significant. Although this allows for observing outcomes over a broader time window before and after the disruption, it likely reflects reduced statistical power in the longer panel.

The analysis further shows that the adverse academic effects are primarily concentrated among children who change their place of residence in association with family disruption. Children who remain in place show no significant decline in test scores, whereas those who move experience significantly worse outcomes. The distance of the move also matters: long-distance moves are associated with particularly large negative effects, while short-distance moves have smaller, statistically insignificant impacts. These findings suggest that much of the observed academic harm stems not from family disruption itself, but from the residential instability that often accompanies it.

Robustness checks focusing on parental breakup alone yield qualitatively similar patterns, although the overall negative effect on test scores is somewhat smaller in magnitude. Children who move following a breakup still experience significant declines, while those who remain in place do not. Taken together, these results underscore the role of residential mobility as a key mechanism through which family disruption affects children's academic performance.

This analysis opens several avenues for future research. First, my findings highlight the role of residential mobility in shaping children's academic outcomes after family disruption. Future research should investigate the mechanisms driving the adverse effects of long-distance moves. Are these effects primarily due to factors such as changing schools, adjusting to a new neighborhood, reduced time with the mother, or less father involvement? Additionally, moving to a potentially supportive environment, like living with grandparents, might mitigate some of these negative impacts. Further exploration of heterogeneity in how different types of moves affect children would also be illuminating. Second, while this analysis does not capture spillover effects, examining general equilibrium effects in future studies could provide insights into how disruptions in one family might affect their neighborhood or broader social network.

References

Abadie, A. (2005). Semiparametric difference-in-differences estimators. *The Review of Economic Studies 72*(1), 1–19.

Agostinelli, F., M. Doepke, G. Sorrenti, and F. Zilibotti (2024). It takes a village: The economics of parenting with neighborhood and peer effects. *Journal of Political Economy, forthcoming*.

Agostinelli, F. and M. Wiswall (2024). Estimating the technology of children's skill formation. *Journal of Political Economy forthcoming*.

Baker, A. C., D. F. Larcker, and C. C. Wang (2022). How much should we trust staggered difference-in-differences estimates? *Journal of Financial Economics 144*(2), 370–395.

Callaway, B. and P. H. Sant'Anna (2021). Difference-in-differences with multiple time periods. *Journal of Econometrics 225*(2), 200–230.

Chan, M. K. and K. Liu (2024). Changing families: Family relationships, parental decisions and child development. *Journal of Labor Economics, Forthcoming*.

Chetty, R. and N. Hendren (2018a). The impacts of neighborhoods on intergenerational mobility i: Childhood exposure effects. *The Quarterly Journal of Economics 133*(3), 1107–1162.

Chetty, R. and N. Hendren (2018b). The impacts of neighborhoods on intergenerational mobility ii: County-level estimates. *The Quarterly Journal of Economics 133*(3), 1163–1228.

Chetty, R., N. Hendren, and L. F. Katz (2016). The effects of exposure to better neighborhoods on children: New evidence from the moving to opportunity experiment. *American Economic Review 106*(4), 855–902.

Chyn, E. (2018). Moved to opportunity: The long-run effects of public housing demolition on children. *American Economic Review 108*(10), 3028–56.

Cordes, S. A., A. E. Schwartz, and L. Stiefel (2019). The effect of residential mobility on student performance: Evidence from new york city. *American Educational Research Journal 56*(4), 1380–1411.

Cunha, F. and J. J. Heckman (2007). The technology of skill formation. *American Economic Review 97*(2), 31–47.

Cunha, F. and J. J. Heckman (2008). Formulating, identifying and estimating the technology of cognitive and noncognitive skill formation. *The Journal of Human Resources 43*(4), 738–782.

de Chaisemartin, C. and X. D'Haultfœuille (2020). Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review 110*(9), 2964—-96.

Del Boca, D., C. Flinn, and M. Wiswall (2014). Household choices and child development. *The Review of Economic Studies 81*(1), 137–185.

Dube, A., D. Girardi, Jordà, and A. M. Taylor (2025). A local projections approach to difference-in-differences. *NBER Working Paper No. 31184*.

Dunn, L. and F. Markwardt (1970). Manual for the peabody individual achievement test. *Circle Pines, MN: American Guidance Service*.

Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics 225*(2), 254–277.

Greenwood, J., N. Guner, G. Kocharkov, and C. Santos (2016). Technology and the changing family: A unified model of marriage, divorce, educational attainment, and married female labor-force participation. *American Economic Journal: Macroeconomics 8*(1), 1–41.

Heckman, J. J., H. Ichimura, and P. Todd (1998). Matching as an econometric evaluation estimator. *The Review of Economic Studies 65*(2), 261–294.

Holm, M. L., P. Fallesen, and E. Heinesen (2023). The effects of parental union dissolution on children's test scores. *The ROCKWOOL Foundation Research Unit* (Study Paper No. 185), 1–60.

Kalil, A., M. Mogstad, M. Rege, and M. Votruba (2011). Divorced fathers' proximity and children's long-run outcomes: Evidence from norwegian registry data. *Demography 48*(3), 1005–1027.

Marinescu, I. (2016). Divorce: What does learning have to do with it? *Labour Economics 38*(2), 90–105.

Nielsen, M. M., P. Fallesen, and M. Gähler (2025). Parental union dissolution and children's emotional and behavioral problems: Addressing selection and considering the role of post-dissolution living arrangements. *Social Forces*, 1–22.

Roth, J., P. Sant'Anna, A. Bilinski, and J. Poe (2023). What's trending in difference-in-differences? a synthesis of the recent econometrics literature. *Journal of Econometrics 235*(2), 2218–2244.

Sant'Anna, P. and J. Zhao (2000). Doubly robust difference-in-differences estimators. *Journal of Econometrics 219*(1), 101–122.

Selya, A., E. Engel-Rebitzer, L. Dierker, E. Stephen, J. Rose, D. Coffman, and M. Otis (2016). The causal effect of student mobility on standardized test performance: A case study with possible implications for accountability mandates within the elementary and secondary education act. *Frontiers in Psychology 7*(1096), 1–10.

Sun, L. and S. Abraham (2001). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics 225*(2), 175–199.

Tartari, M. (2015). Divorce and the cognitive achievement of children. *International Economic Review 56*(2), 597–645.

Wooldridge, J. (2021). Two-way fixed effects, the two-way mundlak regression, and difference-in-differences estimators. *Working Paper*.

Appendices

Appendix A. Details on PIAT

Key outcome and measurement variables in this paper are PIAT scores. To assess a child's reading skills, the PIAT encompasses two sub-tests: Reading Comprehension and Reading Recognition. The PIAT Reading Comprehension test measures the child's ability to understand and interpret a sentence. This sub-test consists of 66 items of increasing difficulty, where the child has to select one out of four pictures that best describes what is written. The PIAT Reading Recognition test, on the other hand, measures word recognition and pronunciation ability, which are essential components of reading achievements. It contains 84 items, each with four options, which increase in difficulty from preschool to high school levels. Skills assessed include matching letters, naming names, and reading single words aloud. Adverse events, such as social disruptions, could impact the broader, compounding development of the child. Furthermore, Dunn and Markwardt (1970) emphasize that reading aloud is a useful skill throughout life in a wide range of situations also outside of school. The authors also stress that, as the child gets older, performance becomes more and more confounded with the acculturation factors. The third main component of the PIAT assesses skills in Mathematics. It consists of 84 multiple-choice items of increasing difficulty. It begins with recognizing numerals and progresses to measuring advanced concepts in geometry and trigonometry. The child looks at each problem on a flipboard and then chooses an answer by pointing to or naming one of four answer options. More information regarding the PIAT can be found on the NLSY website of the U.S. Bureau of Labor Statistics.[11]

[11] https://www.nlsinfo.org/content/cohorts/nlsy79-children/topical-guide/assessments/piat-reading-reading-recognitionreading, accessed November, 2024.

Appendix B. Additional Tables and Figures

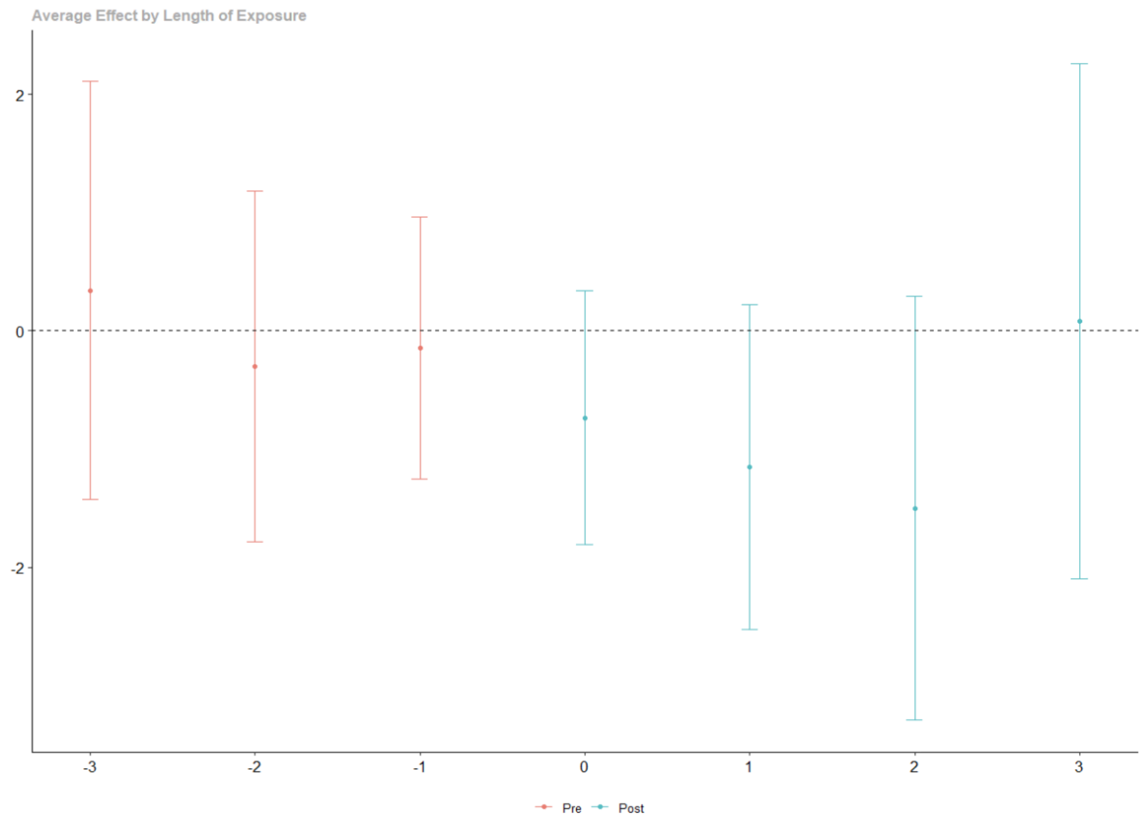
Table 24: TWFE (FamDisrup), 4 Observation Periods

	Coefficient compared to t-3		
Time	**t − 1**	**t + 1**	**t + 3**
DD Estimate (FamDisrup-Stable & versus t-3)	-0.45	-1.39**	-2.42***
	(0.61)	(0.69)	(0.79)
Obs (balanced panel): 9,088			
Experiencing Fam Disrup: 11.9%			
R^2: 0.87; Adj. R^2: 0.83			
Note:	*p<0.1; **p<0.05; ***p<0.01		

Table 25: Main DD Model (FamDisrup), 4 Observation Periods

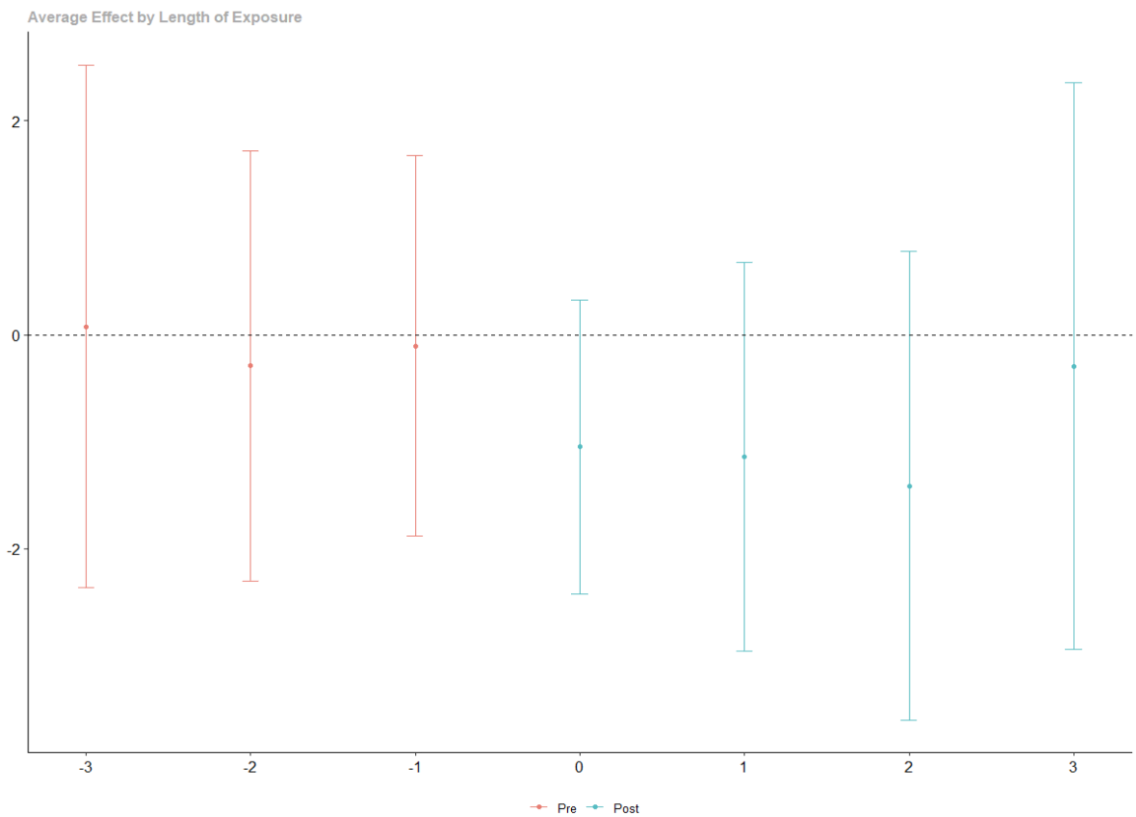
Event Time	(1) Family Disruption	(2) Family Disruption	(3) Breakup	(4) Breakup
−2	−0.42	−0.44	−0.60	−0.21
	(0.58)	(0.55)	(0.70)	(0.70)
−1	0.03	0.15	−0.28	0.04
	(0.39)	(0.40)	(0.49)	(0.47)
0	−0.89*	−0.61	−1.03*	−0.55
	(0.31)	(0.30)	(0.36)	(0.38)
1	−2.03*	−1.30*	−1.98*	−0.90
	(0.44)	(0.41)	(0.53)	(0.52)
2	−2.64	−1.06	−2.50*	−0.40
	(0.65)	(0.59)	(0.78)	(0.65)
Overall ATT:	−1.85*	−0.99*	−1.84*	−0.61
	(0.37)	(0.37)	(0.43)	(0.44)
Controls		✓		✓
# Obs	11,240/4 = 2,810 kids		10,172/4 = 2,543 kids	
Fam Disrup	30%		22%	

Figure 1: Dynamic Aggregation of Group-Time Average Disruption Effects



Note: Graph shows aggregated average treatment effects on treated by length of exposure. The dynamic aggregation of group-time average treatment effects includes all types of family disruption. Whisker plots represent 95% confidence interval constructed from bootstrapped standard errors with 1000 iterations.

Figure 2: Dynamic Aggregation of Group-Time Average Breakup Effects



Note: Graph shows aggregated average treatment effects on treated by length of exposure. The dynamic aggregation of group-time average treatment effects includes breakups only. Whisker plots represent 95% confidence interval constructed from bootstrapped standard errors with 1000 iterations.

Figure 3: W: ATT Heterogeneity by Sex and Race

Comparing the ATT on Girls (=1) and Boys (=0)

	Girls	
	0	1
ATT	-1.176	-1.123
	(0.479)	(0.506)

Comparing the ATT on Black (=1) and Non-Black (=0)

	Black	
	0	1
ATT	-0.246	-0.135
	(1.508)	(1.768)