I Z A Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 17057

Financial Incentives for Adoption and Kin Guardianship Improve Achievement for Foster Children

David Simon
Aaron Sojourner
Jon Pedersen
Heidi Ombisa Skallet

JUNE 2024

I Z A Institute
of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 17057

Financial Incentives for Adoption and Kin Guardianship Improve Achievement for Foster Children

David Simon
University of Connecticut and NBER

Aaron Sojourner
W.E. Upjohn Institute and IZA

Jon Pedersen
Minnesota Department of Human Services

Heidi Ombisa Skallet
Minnesota Department of Human Services

JUNE 2024

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Financial Incentives for Adoption and Kin Guardianship Improve Achievement for Foster Children*

Difference-in-differences analysis of linked administrative data estimates effects of a Minnesota policy change that increased the payments to adoptive parents and kin guardians for children ages 6 and older, making them equal to what foster care payments were, but didn't for younger children. Equalizing payments raised average academic achievement by 31 percent of a standard deviation three years after foster children's cases started, raised the value of payments by about $2,000 per child during this period, raised the monthly chance of moving from foster care to adoption or kin guardianship by 29 percent, improved school stability, and reduced school suspensions.

JEL Classification: H75, I38, D19

Keywords: child welfare, education, human capital

Corresponding author:
Aaron Sojourner
W.E. Upjohn Institute for Employment Research
300 S. Westnedge Ave
Kalamazoo, MI 49007
USA
E-mail: sojourner@upjohn.org

* Simon acknowledges funding from the University of Connecticut Health Economics and Policy Education Lab. Sojourner acknowledges funding from Casey Family Programs. We would like to thank Kasey Buckles, Joe Doyle, Hilary Hoynes, workshop participants at the NBER Children's Program Summer Institute, Society of Labor Economists annual meeting, SOLE annual meetings, the 2023 Atlanta Workshop on Public Policy and Child Well-Being, Maggie Brehm along with the Oberlin economics seminar participants, and the University of Pittsburgh economics seminar participants for their valuable comments. We would like to thank Leonardo Maldanado, Celeste Siameh, Yanxu Long, Keyongki Park, and Jing Cai for their fantastic research assistance.

1 Introduction

What role can financial incentives play in encouraging adoption and kin guardianship and improving the outcomes of affected children? Such financial incentives can be large and are a potentially important tool available to state agencies to improve the outcomes of children in foster care. Earlier work suggests that families respond to these incentives by taking more children into permanent adoptive or kin-guardian arrangements (Buckles (2013), Brehm (2021)). In spite of this, we know little about impacts on the ultimate goal of such incentives: do they improve the longer term outcomes of children? In this paper we use a Minnesota policy reform combined with linked administrative data and a difference-in-differences identification strategy to generate evidence on this question.

While many children in foster care reunite with their family of origin, others remain in foster care for an extended time. Nationally, 10 to 25 percent age out of the system without being reunified or adopted. Typically, states provide financial support to children in foster care, but this support often ends when a child is adopted or placed in a kin guardianship, leaving the new family to fully support the child financially. Meanwhile, states face challenges in finding safe, supportive adoptive homes for foster children. However, the effect of increasing payments in permanency on children's later human capital outcomes could be positive, null, or negative.

On the positive side, financial resources coming into the family may directly improve the foster child's outcomes.[1] Prolonged exposure to foster care and aging out of foster care is itself correlated with poor transitions into adulthood, including homelessness (Congressional Research Service, 2019). Financial incentives may mitigate this through accelerating the move to "permenency" (adoption or kinship care) sooner, and may lead adoptive parents to invest more resources in the child sooner. For the remainder of this paper we will use the term "permanency" to refer to the two types of post-foster alternative arrangements directly affected by the payment reform: adoption (a permanent and lifelong legal arrangement) or kin guardianship (an arrangement that ends when the child turns 18, typically used by the child's extended family).[2] Finally, payments could enable low-income families that

[1]In general, more resources for children's households evidently improve long-run outcomes (Institute of Medicine and National Research Council, 2000). Evidence points to beneficial effects of family income, especially among children in more-disadvantaged families (Duncan et al., 2010, 2011; Dahl and Lochner, 2012; Løken et al., 2012; Aizer et al., 2016), and of food stamps (Hoynes et al., 2016). Among children in low-income families, increased access to health insurance raises educational attainment, employment, and hourly earnings, and reduces disability in adulthood (Brown et al., 2020; Cohodes et al., 2016; Goodman-Bacon, 2021). Evidence shows increased access to expensive high-quality care and education services improve children's short- and long-run outcomes (Duncan and Magnuson, 2013; Elango et al., 2015; Hendren and Sprung-Keyser, 2020).

[2]Both adoption and kin guardianship are forms of post-foster permanency that are alternatives to reuni-

have important non-financial resources and the child's interests at heart to take on the responsibility of care they otherwise would not be able to afford.

For null effects, a child's foster family is by far the one most likely to adopt a child, and it is often required that prospective families spend some time fostering the child they wish to adopt. So, the child will often be in the same family setting either way, just changing the label on the time (foster or permanency) but not the family setting itself. Also, because parents and guardians have control over the payment resources, they may direct them to purposes other than the child's development.

For negative effects, the highly personal nature of child rearing means that matching between prospective adoptive parents and children is important. There's little prior evidence about how financial incentives affect match quality. On the margin, the policy's boost to pecuniary incentives could crowd out altruistic incentives and worsen match quality, with negative consequences for children's development (Bowles, 2016). Ultimately, we have little knowledge of the long term effects of financial incentives to move children into permanency, and it is difficult to make a prediction of even of the *sign* of the impact of such a policy on children's outcomes.

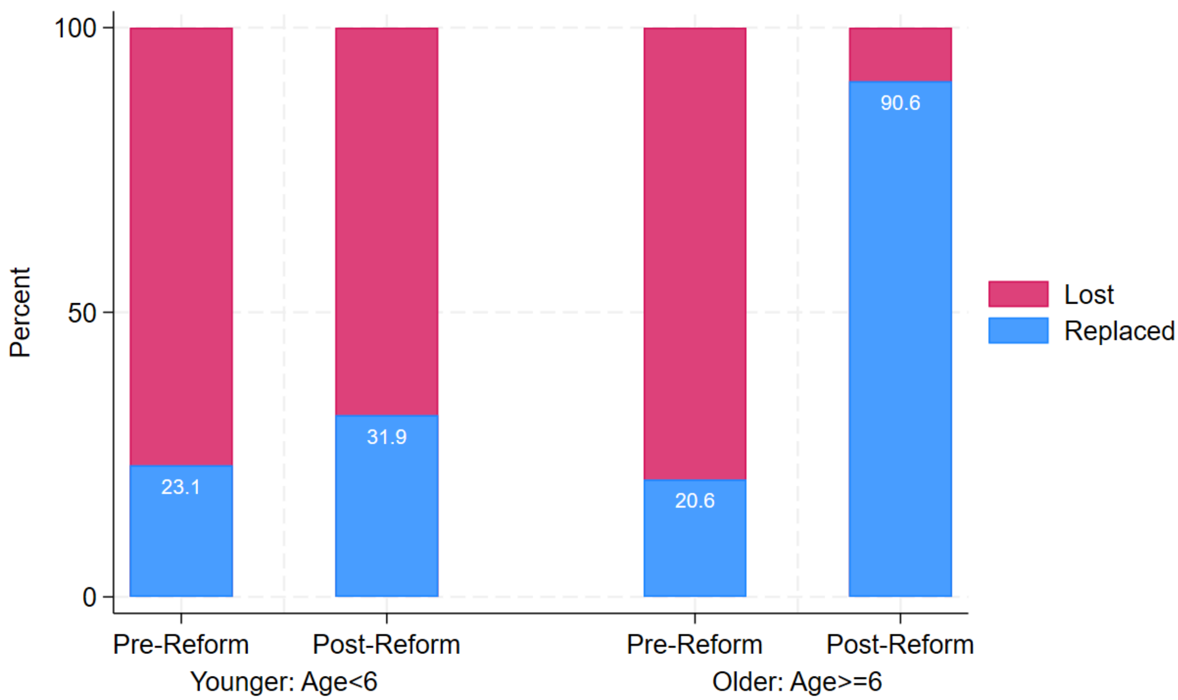
To understand the effects of financial incentives, we leverage a payment reform that was part of Minnesota's January 2015 Northstar child welfare policy reform. Minnesota policymakers were dissatisfied that older foster children (ages 6 years and up) spent too long in foster care rather than being reunited with their origin family or finding an alternative permanent home. In contrast, younger foster children who were not reunified with their origin family moved to an alternative permanency arrangement more quickly. Policymakers worried that the termination of payments that occurred with permanency created unintended disincentives against it. For children exiting foster care into permanency, the state's monthly payment rate fell by an average of 80 percent for older children and by 77 percent for younger children (Figure 1).[3] Following the reform, among children ages 6 and up, the within-child ratio of monthly payment in permanency to foster care rose from 20 percent to 90 percent (Figure 1). In contrast, the ratio rose only from 23 percent to only 32 percent for younger

fication with the origin family. We use permanency as an umbrella term that refers to both adoption and kin guardianship so as not to privilege one of these arrangements over the other. There is a debate among child protection and social work researchers about whether kin guardianship is less beneficial for children than adoption. Our findings apply to both arrangements.

[3]Northstar's implementation was preceded by a demonstration study with experimental and non-experimental designs that looked only at changes in length of time the child spent in foster care. The demonstration project aimed "to determine whether a continuous (or single) benefit program would increase permanency rates and shorten foster care stays among children who have been in foster care for an extended period of time" (Institute of Applied Research, 2011). The study suffered from a variety of implementation challenges, making the findings unclear. An experimental part of the study that ran in only two counties found some imprecise, suggestive evidence of accelerating children's moves from foster care into permanency.

children exiting foster care to permanency. Consequently, the policy had a much larger effect on caregiver financial incentives for older than for younger children.

Figure 1: Replacement Rates of Permanency Stipends, by Child Age and Policy Regime



Source: Minnesota Department of Human Services data on payments. See the text in section 5 and Appendix D for details.
Notes: This figure focuses on children who exited from foster care to adoption or kin guardianship, referred to jointly as permanency. For each child, we compute the ratio of their monthly post-foster stipend in permanency to their monthly stipend in foster care. We partition children into four groups by age group and policy regime at exit from foster care. Across children within group, the bottom bar reflects the average percent of their foster stipend replaced by their stipend in permanency and the top bar reflects the complementary percent of foster stipend lost in permanency.

We analyze the effects of payment equalization on children's academic achievement scores three years after the start of their foster case via linked administrative data. This is the first paper to use quasi-experimental methods to investigate the impact of payment policies in permanency on children's later human capital accumulation. Due to children changing their name in adoptive families, linking children in foster care out of the child welfare system into other administrative records is challenging. We address this difficulty with new linkages within and across public agencies administered by the University of Minnesota's Center for Advanced Studies in Child Welfare via its Minn-LInK project discussed below.

To identify effects of the payment equalization policy, we employ a variety of difference-in-differences (DiD) specifications. We compare outcome changes experienced by older children in foster care relative to the changes of younger children with similar observable characteristics. To understand impacts of the policy on time spent in foster care, we use duration models with calendar year-month fixed effects and foster child age fixed effects. Linking Child Protective Services administrative data to schooling and academic records allows us to explore the impact of the permanency subsidies on longer-run outcomes such as standardized academic achievement scores, using child age and date at the start of the foster placement as exogenous variation in exposure to the policy of payment equalization. We augment this identification strategy with random forest techniques to predict which children at foster care entry are most likely to be impacted by the policy.

We find large medium-term gains to child human capital. For children exposed to the policy, achievement test scores increased by a third of a standard deviation three years after case start. To understand these gains, the rich data allows study of a variety of mechanisms. The policy increased the net present value of the total stream of payments paid in the years between case start and achievement testing by an average of $1,964. This implies a cost benefit ratio around 16, given how achievement gains lift lifetime earnings (Hanushek, 2011). The achievement effects are 3–5 times larger than would be predicted based on that payment difference and prior research on the causal effect of increased family income on test scores, though prior work was outside of the foster care setting. Therefore, while more money going to families is likely part of the story driving test score gains, improved stability in the child's life likely also played a role. We further see substantial reductions in children's behavioral problems, as proxied by a decline in the likelihood of school suspensions, and a reduction in school instability, proxied by their average number of schools attended per year.

Closest to this study is a literature that evaluates the impact of monetary incentives on the occurrence of adoption and length of time spent in foster care. Such studies find that increasing monetary payments to adoptive families increases the number of adoptions in a given period of time.[4] For example, Buckles (2013) looks at federal funds for adoption subsidies provided through the 1980 Adoption Assistance and Child Welfare Act, showing that such funds increased the number of adoptions. Similarly, Brehm (2021) uses a bunching analysis to show that making the federal adoption tax credit refundable resulted in 2,400

[4]Historically, the quantitative analysis of foster care policies is likely subject to selection bias (Cuddeback, 2004; Buckles, 2013). More recent studies trying to estimate causal impacts focus on whether home removal is good for child well-being. Two influential papers use random assignment to caseworkers and variation in relative caseworker leniency (Doyle Jr., 2007, 2008; Bald et al., 2019). Doyle (2007, 2008) compare children who were assigned to caseworkers with a predisposition for home removal versus children assigned to a more lenient caseworker. Applying this design to Illinois data, Doyle finds worse outcomes for children (particularly older children) who were removed from the home. More recently, Roberts (2019), Bald et al. (2019), and Gross and Baron (2022) use data from other states with similar research designs and find improvements in educational and/or safety outcomes for children. Warburton et al. (2014) document a sudden increase in home removals following a high profile child death due to a failure in the foster care system. Such removals are associated with worse educational and economic self-sufficiency outcomes for older boys.

more adoptions in the year following the policy change. Argys and Duncan (2013) look at state-level policy variation in the adoption subsidy; using a state-year panel they also find such subsidies increase the number of adoptions.[5] Unlike our work, most of their cross-state policy variation decreases but does not eliminate the financial "penalty" associated with moving from foster care to adoption.[6]

We confirm and extend this earlier literature. In a related paper, Perales (2024) looks at the impact of a retrospective increase in payments to kin-guardians and finds that this increase in payments decreases the long term likelihood of maltreatment of children in California. We differ from Perales (2024) in focusing on academic outcomes which are directly related to child human capital accumulation. Partnering with the government of the state of Minnesota allows access to these richer data to look at longer-term outcomes not previously investigated. Further, unlike Perales (2024), our identification strategy uses an increase in the amount of payments to permanent families *relative* to foster payments: such that our treatment captures an incentive to adopt children into kinship arrangements. Finally, we also capture effects on both kinship and adoptive arrangements.

This is also the first paper we know of in the economics literature to specifically focus on the elimination of the disparity in payments between foster care and permanency. Equalizing the stipend offered in foster care and permanency potentially results in larger impacts than found in earlier work. Further, state policy changes rarely happen as an isolated event but tend to be associated with a number of additional reforms. Focusing on a single state allows us to look closer at the institutional details associated with a policy change, and to incorporate this into our model. Because we isolate a single large policy shock, we can also transparently test for pre-trends leading up to the reform. While Northstar did change other aspects to the policy, the others did not have age-specific thresholds, and so the first differences absorbs their effects.

In addition, we check for anticipation effects, such as potential adoptive parents delaying

[5]Rodgers and Wallace (2020) and Hansen (2007) also find that financial incentives encourage adoption.

[6]A related literature looks specifically at how subsidies to foster families impact foster outcomes while the child remains in foster care. Critically, such payments could decrease the incentive to adopt a child if the payments made when child is adopted don't also increase. Testa and Slack (2002) and Doyle (2007) investigate this in the context of a policy in Illinois that decreased payments to foster parents. These papers find that as payments to foster care families decline both the amount and quality of care declines, particularly for "costly" child cases such as children with diagnosed mental health problems. Likewise, Testa and Rolock (1999) and Doyle and Peters (2007) show that higher payments to foster care families increase the supply of available foster families. These papers also shows improvements in foster placement stability with increased payment levels. More recently, Perales (2024) finds that increasing payments to county foster care providers decreased the maltreatment of children in private agency homes as the private providers felt pressure to provide a higher quality of service in order to remain competitive with county care providers. On the other hand, Chorniy and Mills (2022) find no evidence that such increases in payments to foster families improve child health as measured by emergency room visits.

permanency until their foster child ages into the higher, post-reform adoptive payment. We find evidence of some children younger than 6 having their permanency delayed until they are older due to the law, but the delays are short and isolated to those children who were directly below the age 6 threshold. Accounting for such red-flagging or "strategic delays" of permanency has only minimal impacts on our estimated average treatment effects.

Section 2 goes into details on policy context along with a description of the data. Section 3 describes the basics of the identification strategy. We then document the large test score gains, as well as impacts on other human capital outcomes from the policy in Section 4. We place these findings in context by using payment data in Section 5. Section 6 shows that time to exit from foster care also greatly accelerates. Section 7 places the various estimates throughout the paper into context while discussing mechanisms and implications, and Section 8 concludes.

2 Context and Data

Our analysis focuses on comparing changes in the outcomes of foster cases before versus after the reform with attention to differences in changes for children above and below the age 6 threshold. Payments in adoption increased much more for older than for younger children (Figure 1). This differential increase in the permanency-to-foster payment ratio is the basis of our research design. The differential increase in payments is not just for children ages 6 or higher at the time of the reform but for all children who age into 6 at any time after the reform is implemented.[7]

Our main results are on outcomes three years after foster care start. To explore how permanency payments to families changed with Northstar, we use payment data from Minnesota DHS to construct the total Net Present Value (NPV) of payments from foster care entry through three years after case start (including both foster care and permanency payments), chosen to parallel the period between case start and the time of the standardized achievement score measurement for the test-score subsample. The average permanency placements pre-Northstar took 16 months from case start: so the policy on average reflects 1–2 years of additional monthly permanency payments.[8]

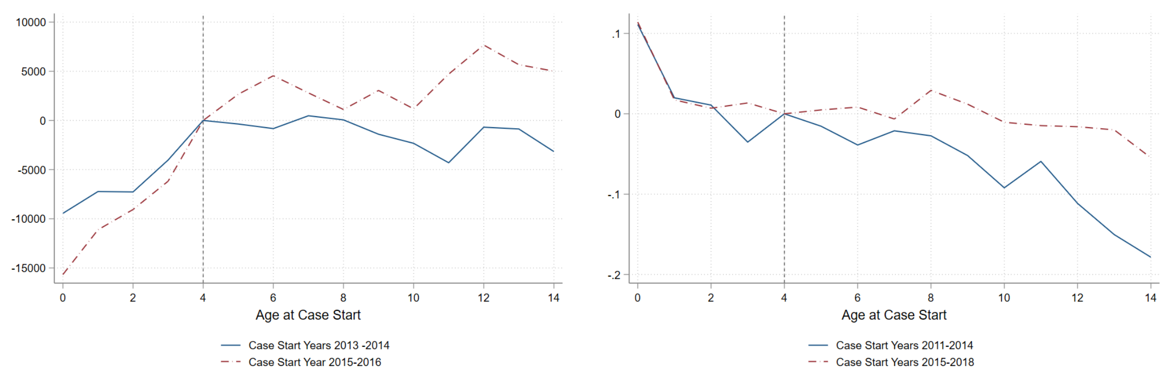
Around the policy change, the NPV of payments is relatively stable for children whose cases start at ages that likely make them too young to qualify for payment equalization but increases substantially for older children, those likely subject to the payment equalization

[7]The median post-reform ratios are 50 percent for older adopted children and 100 percent for younger adopted children, though there are many particular circumstances where policy does not extend payments post permanency or that adjust the ratio somewhat.

[8]For more details on how we construct net present value from the payment data see Appendix D

reform. Panel A of Figure 2 shows raw mean trends in NPV for all children who get placed in permanency by their age at case start for both the pre-equalization-policy years of 2013–2014 and the post-policy years of 2015–2016. To ease comparison, both series are normalized to be zero for those age 4 at case start. Given the average of 16 months in foster care before permanency, children whose cases start at ages older than 4 in the later years would be expected to remain in foster care long enough for them to qualify for equalized payments; those 4 and younger are expected not to stay in long enough to qualify in any year, though some individuals do of course. In the pre-policy years, NPV increases steadily with age until leveling out at around age 4. This increase is due to the fact that younger children exit to permanency much more quickly than older children in foster care and, in permanency, received few to no payments from the state. For the post-policy years NPV also increases with age but increases past age four before leveling out with higher payments for children ages six and older: the age at which the policy provides higher payments. This differential increase in payments before versus after the policy reflects the increased generosity of payments in permanency to older children. Difference in the raw data in payments between the two series ranges from \$5,000 to \$7,000.

Figure 2: Relative Net Present Value of Payments and Share Placed in Permanency Within Two Years by Age at Case Start



(a) Net Present Value (b) Share Adopted in Two Years

Source: Minnesota Department of Human Services data on payments. See the text in section 5 and Appendix D for details.
Notes: In both figures each series is normalized to be relative to their age 4 values. Panel A shows the average Net Present Value of payments made to children who enter permanency (2022\$), where the horizontal-axis is age of child in years at case start. Series are shown for two sets of foster case start years, immediately before the payment equalization reform (2013–2014) and immediately after (2015–2016). Panel B shows the percent of cases adopted within two years by child age at case start. We observe more years of data on foster care cases than on payments, so Panel B shows the series over longer periods.

Families responded to these incentives by adopting more of the older children whose NPVs increased more (Figure 2: Panel B). The vertical-axis is the normalized share of cases that exited to permanency within two years of case start by age at case start, for the pre-policy years of 2011–2014 and post policy years of 2015–2018.[9] Starting around birth the likelihood of a permanency placement within two years decreases with age for both series. In the pre-policy years, adoption shares continue to decline for older children past age 4. In contrast, for those (older) children exposed to the policy, the decline in adoption shares flatten out: reflecting a relatively larger likelihood of adoption for older children post-reform. Our identification strategy builds off these types of comparisons apparent in the raw data, using regression analysis to hold the composition of case types constant over time.[10]

Recent data quality improvements that improve linking of a child's records within the Minnesota child welfare administrative data over time and across other administrative data sets reduces measurement error here. Preivously, child name changes at permanency made it challenging to link a child's pre-permanency records with any post-permanency record, even within child welfare data. Linking to K12 or other data systems adds more challenge. This study benefited from Minn-LInK's recent investment in following children across name changes and their long-running ability to link a child across state administrative data systems. Minn-LInk analysts linked children across adoption name changes using algorithmic linking backed by substantial human verification. Researchers access de-identified, linked data via a secure server. Leveraging identifiers in state data, it is now more possible to reliably identify a child over time and across state agencies. We observe both time spent in foster care and payments from the state to both foster and adoptive or kin guardian families. We trace the outcomes of these children in both the Department of Human Services (DHS) records, Department of Education data on school attendance, out-of-school suspensions, and academic achievement on standardized tests Department of Health records on Medicaid-funded mental health service use.

Additional reforms occurred around the same time as Northstar's permanency payment reforms but only payment reform differentiated at the age 6 threshold.[11] The other main reform was that Minnesota began using a new system for assessing children and caregivers in order to determine payment levels. Specifically, the state moved away from using different rubrics for setting payments in foster, kin guardianship, and adoption. In the new system, all children in foster care were assessed in a single rubric that determined payments in all arrangements. Because it affected all children in foster care rather than differentiating by

[9] We have more years of data on foster care cases then on payments.

[10] Our access to the academic achievement data is dormant until after we get an R&R. We devised the design for these figures after our access went dormant, so we do not have an analogous figure for student achievement but plan to include one in the next iteration.

[11] Northstar is not unique in this respect. States that reform their permanency subsidy payments typically implement a number of additional reforms to qualification requirements and administration.

age, the DiD identification strategy should pull the effects of this general change out as a time effect. We also observe the assessment rubric and score, and controlling for assessment score directly does not affect our results. We further discuss this test and the changes in assessment in Section 4.2 below.[12]

2.1 Foster Care Data and Descriptive Statistics

We analyze Minnesota DHS data on all foster care cases starting from January 2011 through July 2019. A case starts when the state removes a child from their family of origin and places them into foster care.[13] For each case, we observe start date, child demographics (age at start, race, ethnicity, and gender) and reasons for removal from the family of origin. Panel A of Table 1 shows summary statistics across cases. The first column summarizes case characteristics across all foster care cases. The second column summarizes the subsample of records linked to K12 records 36 months to 47 months (three years) after case starts: this is our main sample for educational outcomes besides test scores. The third column summarizes variables for the subsample with a state-mandated standardized achievement test score at three years after case start. This subsample is smaller. Because the state mandates testing only in grades from three to eleven, this subsample excludes children who are the youngest and oldest at case start.

Case characteristics are similar across samples. Children average a bit older than age 8 at the start of their foster care placement. Twenty percent of cases involved African American children and 15 percent involved Native American children, far higher than these groups' shares in Minnesota's child population. Panel B of Table 1 shows the reason for exit from foster care for the 88 percent of cases that end by the observation window's close in July

[12] Additionally, directly following the Northstar reform, the state also began requiring a licensure requirement for kin guardians. Specifically, kin guardianship could only occur after the prospective guardian had cared for the child for six months as a licensed foster parent. Prospective parents already were required to foster before adoption. The state wanted to reduce the share of guardianships that later dissolved, returning the child to foster care. By ensuring guardians met foster parent licensing requirements and succeeded through a six-month trial period, this reform aimed to promote post-foster placement stability. Requiring the family to have already spent some time with the child ensured that the child and their permanent caregiver had a strong attachment to one another. To account for this, we directly model this requirement when looking at kin guardianship by controlling for an indicator variable when in the first six months of foster care out of the risk set for guardianship after the policy came into effect. Finally, Medicaid was extended to some additional families following kin guardianship, though not adoption. This policy impacted all children and did not differentiate by child age. Further, the extent of take-up is unclear as Medicaid was sometimes available to these families through other sources. It is possible that older children benefit more from a Medicaid expansion, but the existing literature suggests that the long-term health benefits of Medicaid are focused on children exposed in the prenatal period and those younger than age 5, which would (if anything) bias our estimates downward (Miller and Wherry, 2019).

[13] We exclude the few cases where parents voluntarily placed their children in foster care because these are nearly guaranteed to end in reunification.

Table 1: Summary Statistics of Foster Care Cases

	(1)	(2)	(3)
		Subsample linked to:	
Sample of cases:	All	K12 Records	Test Scores
Panel A: Case Characteristics at Start			
Age, years	8.34	7.27	8.57
Average number of cases per child	1.37	1.37	1.28
White	37%	37%	41%
African American	20%	20%	18%
American Indian	15%	16%	16%
Hispanic	10%	10%	10%
Removed for neglect	26%	30%	32%
Removed for physical abuse	10%	12%	14%
Removed for caretaker drug use	24%	22%	23%
Removed due to child behaviors	19%	15%	8%
Panel B: Case Outcomes If Closed			
Average case length, months	11.42	13.16	13.38
Exit to family reunification	58%	62%	62%
Exit to any permanency	19%	26%	27%
Exit to adoption	11%	15%	13%
Exit to kin guardianship	8%	11%	14%
Panel C: Test Scores 3–4 Years after Case Start			
Math z-score	—	—	-0.81
Reading z-score	—	—	-0.70
Average z-score	—	—	-0.77
Panel D: School Outcomes 3–5 Years after Case Start			
Any out-of-school suspensions	—	13.5%	22%
Average School attendance rate	—	88.5%	91%
Average number of schools attended per year	—	1.49	1.58
Any use of mental health services	—	10%	7.5%
Number of cases	52,344	20,407	6,908

Source: Minnesota Department of Human Services foster case data, Minnesota Department of Education K12 school data, and Minnesota Department of Health Medicaid data for 2011–2019.

Notes: This table shows summary statistics at the foster case level linked with educational and health records. The first column shows statistics on all foster cases starting during this time. The second describes the subsample of cases that link with educational records. The third describes the further subsample for which achievement test scores 3 years (36-47 months) after case start are observable.

2019. Such dispositions include reunification with the family of origin (58 percent), any other type of permanency (19 percent), including adoption (11 percent) and kin guardianship (8 percent). Of the remaining cases (33 percent) either did not exit by the end of the observed sample time frame or exited for a variety of less common reasons such as aging out of foster care.[14] For children with cases starting in this time range, any prior foster cases are largely observed as well.[15] Fully observed cases average 11.4 months over the entire period.

2.2 Outcomes

We first look at the impact of payment equalization on children's academic achievement, measured by standardized tests taken after three and before four years from a foster case's start. Test scores are transformed into standardized z-scores, normalized within the distribution of scores among all Minnesota children in the same grade, subject, and school year.

To understand potential mechanisms, we study effects on related academic outcomes during these three years between case start and the achievement test, and separately for a few years after. These related outcomes are whether the child experienced any out-of-school suspensions, school attendance rates (percent of days attended per year), school instability (number of schools attended per year), and mental health service use. Data on academic outcomes is typically observed annually and, in the case of test scores, only in certain grades. Child welfare system outcomes, such as whether the child exits from foster care to adoption, are observed monthly. We also look at outcomes in the child welfare system, such as the rate of exit from foster care into permanency and the stream of payments to these children after the start of their case.

An outcome of the child in case-i measured at calendar time-t is normalized according to the number of years elapsed since the start of their case (E_{it}). Outcomes occurring within 0 to 12 months of the start of a case occur before a year has passed have $E = 0$. Outcomes occurring 12 to 24 months after the case start occur in elapsed year $E = 1$, and so on. Each achievement test score (as well as other academic outcomes) is allocated to an elapsed year based on time since case start and coverage of the school year leading up to the spring the test was taken.[16]

[14] Other examples include when children run away or child death.

[15] All past child welfare records are preserved for 10 years after the child's most recent case ends. So we only would miss past cases for older children who had no case in the decade prior but did have an earlier case.

[16] Consider a child who takes a standardized test at the end of April 2014. The year leading up to it runs May 2013 to April 2014. The first half of that year ends in October 2013. If the child's foster case started in October 2012, then $E = 1$ starts October 2013 and runs through September 2014, including the majority of the year ending in the month of the test. So the April 2014 test would be coded as $E = 1$. If instead a

For academic test scores, we focus on test scores three years after case start. We estimate the effect at a fixed elapse since case start in order to hold the amount of time children are exposed to policy and their maximum length in foster care fixed.[17] Three constraints limit the window over which we can observe the test score sample. First, our sample must include children under the age of 6 at case start (control group), as well as older children. Second, standardized tests are administered only starting at the end of third grade. Third, there are less than 5 years between the policy change in January 2015 and the last wave of test scores observed in our data in spring 2019. The sample of children 1) under age 6 at case start, 2) exposed to the policy reform, and 3) with observed test scores, has the most mass at $E = 3$.[18] Using $E > 3$ would reduce the already limited number of post-policy years we could look at in an event study. Using $E < 3$ would dramatically shrink the number of young at case start children we could observe with test scores. That said, some test scores for control group children with $E = 2$ are observed, and they are included in some models to assess robustness.

Overall, our linked sample contains 6,908 cases with at least one $E = 3$ test score (Table 1, Column 3). Restricting the sample in this way limits the age profile of the children in our test score sample to those who start foster care between the ages of 4 and 14. Younger and older children are not observed in a tested grade at $E = 3$.

The elapsed time structure of the test score analysis informs our analysis of attendance rates, out-of-school suspensions, school stability, and mental health service use. These outcomes are all observed at more ages than test scores are, so the linked sample is larger. We pool together outcomes in earlier years ($0 \leq E < 3$) leading up to when test scores are measured, and those observed in the subsequent two years ($3 \leq E \leq 5$).[19] The linked sample for outcomes ($3 \leq E \leq 5$) years after case start contains 20,407 cases (Table 1, Column 2).[20]

child's foster case started November 2012, then the majority of the year ending the month of the test is in $E = 0$, so the test would be coded as $E = 0$.

[17] In our design, since we compare older to younger children, either elapse or grade could be held fixed but not both. Looking instead at effects on test scores in a fixed grade (e.g., third grade) would require comparing children with greatly differing lengths of potential policy exposure. For example, children with a case starting at age five might be exposed to policy changes for three years before third grade scores are measured, while those starting at age seven might be exposed to policy changes for less than half as long. Because all scores are normalized within grade-subject-year in the full Minnesota student population, DiD comparisons of changes between different age groups at the same years-elapsed since case start seem more informative than comparisons of changes between different years-elapsed groups at the same age.

[18] For instance, a four-year-old whose case starts in January 2015 ends third grade in April 2019, yielding an observable third grade test score with $E = 3$.

[19] We still are limited with data through 2019, so outcomes 4-5 years after case start have one fewer year of post policy exposure.

[20] Mean school outcomes for years zero to three are shown in Table 5 and Table A-2. We show school outcomes three to five years after case start in Panel D so that the numbers in Table 1 are the most comparable to our main test score outcomes three to four years after case start, given limited table space.

Children who experienced foster care are highly disadvantaged in terms of achievement, with average test scores 0.77 standard deviations below the statewide mean (Table 1, Panel C). Likewise, they have relatively low average school attendance rates (88.5 percent), a high likelihood of having had any out-of-school suspension three to five years following foster case start (13.5 percent), and substantial school instability averaging 1.49 schools per year. Approximately 10 percent of them use state-provided mental health services at some point during these years (Panel D).

To investigate the impact of the policy on time to permanency, we use hazard models where the outcome is whether foster-care exit is observed in a given month. Given that we use policy variation at the monthly level, we transform the data so each observation represents a case-month, allowing age and calendar time to vary within each case episode so that we measure exactly when children move into different policy regimes. We show summary statistics at the case-year level for these regressions in Appendix Table A-1. Figure 3 visually represents exit to permanency or reunification at the case-month level by showing a survival curve. As discussed above, younger children exit from foster care faster than older children, though many children of both age groups remain in foster care for multiple years.

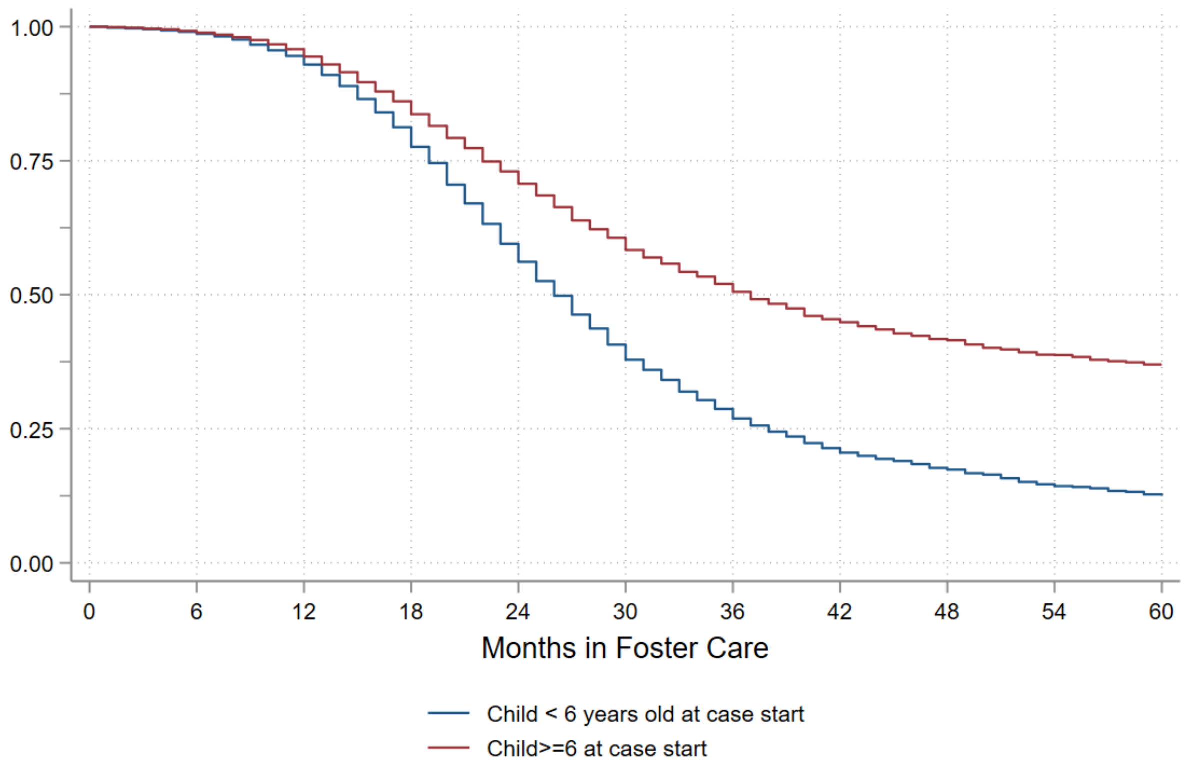
3 Identification and Estimation

We first outline the thought experiment behind our DiD approach, and then describe how we adapt this approach to enable analysis of academic, schooling, and payment outcomes. For analyzing the impacts on time spent in foster care the setup is somewhat different and we discuss that strategy in detail in Section 6. Generally, policy variation comes from the differential post-reform change in payments for older versus younger foster children based on age at case start (Figure 2), and identification of the treatment effect comes from the change in average outcomes from before the reform to after the reform for children age 6 and older (treatment group) relative to children under 6 (control group). For child-i observed at time t when they are a years of age, a standard difference-in-differences model would be:

$$Y_{ita} = \alpha_1 \mathbf{1}(t \geq 2015)\mathbf{1}(a \geq 6) + \alpha_2 X_i + \gamma_a + \delta_t + \epsilon_{ita} \tag{1}$$

Y_{ita} expresses an outcome for case i observed at time t, when i is age a. $\mathbf{1}(t \geq 2015)$ indicates t being post the January 2015 reform. $\mathbf{1}(a \geq 6)$ indicates child i is 6 or older at time t. Interest centers on the DiD parameter α_1, expressing the change in average outcomes following the policy among older children less the change among younger children. X is a vector of case-specific control variables fixed at case start. We control for child race,

Figure 3: Foster Care Case Survival Curves, by Child Age Group at Case Start



Source: Minnesota Department of Human Services data on foster cases. See Section 2 for more details.
Notes: Kaplan-Meier survival curves for the likelihood of not yet exiting foster care as a function of elapsed months since case start for the full sample of cases. The top line shows the survival probabilities for children 6 or older at case start. The bottom shows survival for kids younger than 6 at case start. Over the whole period, older children remain in foster care longer. While children may age into treatment post reform, age at case start provides one proxy for which children are most likely exposed to the reform.

ethnicity, gender, and reasons for the child's removal to foster care.[21] The model absorbs fixed additive differences by age using years-of-age fixed effects (γ_a) and over time using calendar year-month fixed effects (δ_t). This is a non-staggered DiD design estimating the effect of a single policy change where the age and time fixed effects subsume indicators for "treatment" and "post," respectively.[22] Standard errors are clustered at the child level.

[21] We show results with and without controls in our main models. Race and ethnicity dummies are White non-Hispanic, Black non-Hispanic, Hispanic, Native American non-Hispanic, and other. Reasons for removal dummies are neglect, physical abuse, caretaker drug use, behavioral problems, and other. In the raw data, reasons for removal are not mutually exclusive, so we control for what is reported as the primary reason for removal, even though several reasons can be listed. We also include a female dummy to control for gender. The analysis conditions only on variables fixed at case start because policy can affect later time-varying case characteristics.

[22] Recent critiques of two-way fixed effects models focus on staggered-adoption designs, with multiple groups treated at different times (Goodman-Bacon, 2018; Baker et al., 2021). Given that treatment began at the same time for all treated children, this is not an issue.

For our main results, we adapt Equation (1) to estimate the reform's effect on children's academic achievement several years after case start. Treated foster care cases are children who spend some or all of their time older than age 6 when in foster care and some or all of the case occurs post January 2015. We do not assign treatment based on the observed length of the foster care case because length of time in foster care is endogenous to the policy, as we document in Section 6 below. Instead, we link academic records to the foster care case level but use only variation fixed at the *start* of each foster case, the year-month of case start and child age at case start, along with a predicted value for "expected" length of foster care for case-i (L_i) conditional only on characteristics at case start. L_i is a prediction of the time over which the child will be in foster care and therefore exogenously exposed to the policy. In our main models we set L_i equal to 16 months—the average length of foster cases that started and ended within the pre-Northstar period, excluding cases that ended in reunification with the child's origin family.[23]

We replace the DiD treatment variable in Equation (1) ($\mathbf{1}(t \geq 2015)\mathbf{1}(a(it) \geq 6)$) with the share of months each child is expected to spend treated ($ShareTreated_i$) based on this 16-month window of expected case length, given the year-month and age at which the foster case starts. For example, if a child begins their foster care case when they are older than 6 and past January 2015, then their foster care case would be unambiguously treated for its whole expected duration and $ShareTreated_i = 1$. If, based on the child's age and the calendar time at case start, it would be longer than 16 months before the child turned 6 *or* the post-reform period of January 2015 is more than 16 months away, then $ShareTreated_i = 0$. This will be the case for any foster care cases that began before midway through 2013 or for any case that started when the child is younger than 56 months old (approximately 4.66 years old). For intermediate cases, $ShareTreated_i$ ranges between 0 and 1 based on the percent of months that the child is both older than 6 and in the post-2014 reform period.[24] An advantage of this approach is that we can easily modify the value of L_i to be over different expected policy horizons to test different assumptions about timing.

Panel A of Figure A-1 shows variation in $ShareTreated_i$ by year-month of case start (in 6-month intervals) for three different ages at case start: 4, 5, and 6; using the 16-month

[23]The included cases—children who would be ultimately adopted, enter kin guardianship, or whom we don't observe an exit for and may age out of foster care—are the most likely to be affected by the Northstar payment reform's change in financial incentives.

[24]Formally, let t_i^0 be the calendar year-month at case start and a_i^0 be age in months at case start. A case's policy exposure window runs from t_i^0 to $t_i^0 + L_i$. The number of months from the start of the case until the child is at least 6 years old and the policy reform occurred is $W_i \equiv \max\{(t_{\text{Jan2015}} - t_i^0), (a_6 - a_i^0)\}$. The share of the exposure window the child is treated is then $ShareTreated_i = (L_i - W_i)/L_i$ if $W_i \leq L_i$ & $W_i \geq 0$. If $W_i > L$ then $ShareTreated_i = 0$. If $W_i \leq 0$ that implies $t_{0i} \geq t_{\text{Jan2015}}$ and $a_i^0 \geq a_6$, in which case we set $ShareTreated = 1$.

expected treatment window. The pre-period treatment throughout is zero for all three ages. Starting midway through 2013, some children will spend part of their expected case length treated. We see those who were age 6 at case start gradually increasing their treatment exposure starting in mid 2013 until they spend their entire case in the post period (cases that start January 2015 or later). Those who are age 4 at case start never spend any of their expected case treated because even after 16 months they are still younger than 6. Those who are age 5 at case start gradually increase treatment midway through 2013 as their case is predicted to overlap with the post-reform period. However, because only the last four months of their case is predicted to be spent above age 6, treatment only ever increases to 25 percent. Panel B shows the same variation more richly in a heat map with time of case start on the x-axis and age of case start on the y-axis (both in months).

With exposure to treatment measured for each case, we estimate a linear DiD model adapting Equation (1):

$$Y_{(iat)} = \beta_1 ShareTreated_i + \beta_2 X_i + \gamma_{a_i^0} + \delta_{t_i^0} + \epsilon_{iat}. \tag{2}$$

Because an observation here is at the case level and test scores are measured at an elapse of $E = 3$ from case start, age (a) only varies by age at case start (a_i^0) and time (t) only varies by year-month at case start (t_i^0). The model therefore uses age-at-case-start fixed effects $\gamma_{a_i^0}$ and year-month of case start fixed effects ($\delta_{t_i^0}$). Y_{iat} denotes average test z-score for child in case i whose case starts at age a_i^0 and whose test scores are observed $E = 3$ years elapsed (36 through 47 months elapsed) after case start in time t_i^0. For our main test score outcome we average together reading and math z-scores. While we prefer using this treatment intensity approach, we can alternatively employ a dichotomous treatment indicator equal to one for children who are "fully" exposed to the policy, i.e., older than 6 and post 2015 at case start, and zero for all other children. This is identical to assuming a treatment exposure window of $L_i = 0$. This case reduces to a traditional DiD design and we show that results are robust to such an approach.

A common challenge in papers on adoption out of foster care is that our sample is a mixture of two latent types of cases. Most children in foster care reunify quickly with their parents without ever being "at risk" of remaining in foster care indefinitely or until permanency. Alternatively, there are cases that will remain in foster care for a longer period, eventually having parental rights dissolved. When looking at the effects of a policy that impacts payments in permanency, likely-to-reunify cases are unlikely to be affected. In such cases, Child Protective Services reunifies the family if they judge it is low risk and parents pass concrete milestones that can often be achieved quickly with a combination of counseling and additional resources. Reunification occurs in 58 percent of cases in our sample. On average, reunification cases are short compared to the rest of the sample, many lasting less than two months. Though reunification is observable in administrative data *ex post*, it is unobservable *ex ante*, near the beginning of the case.[25] Because of this and because financial incentives to adopt could conceivably decrease the likelihood of reunification if permanency happens in its place, it has been common in the literature to include such cases in the sample (Buckles, 2013; Argys and Duncan, 2013).

We can focus our analysis on the type of cases where the policy is liable to have more substantial effects (cases that would tend to end in adoption, kin guardianship, or aging out) and exclude others (those predicted as likely to reunify). We do not want to condition on realized reunification, which could be endogenous to the policy. Instead, we use random forest models to predict likelihood of reunification and then classify cases as either highly likely or less likely to reunify. The random forest is first trained using a random subset of cases in the early pre-reform period and predicted off of characteristics fixed at the time of case start. We follow conventional machine learning methods to guard against over-fitting and to test the quality of our predictions. Then, we score all cases and partition them between those predicted to have an 80 percent or greater likelihood of reunification versus other cases. We restrict the sample to less-likely-to-reunify when analyzing outcomes beyond the foster care system. We classify observations as less-likely-to-reunify if they are predicted by the random forest to have less than an 80 percent chance of reunification. Appendix C discusses details of the random forest model.[26]

Our key identifying assumption is that there are no confounding differential unobservables changing between older and younger children concurrent with the reform. To test this, we adapt our model to an event study by replacing $ShareTreated_i$ with an interaction between indicators for case start time aggregated into six month bins and the fraction of L_i for which the child is predicted to be "treated" (age 6 or older).[27] Given the 16-month exposure window, event studies are normalized such that the first half of 2013 is the excluded period as that's the last fully unexposed half year; cases starting then have their exposure window end before Northstar begins. Cases beginning in the second half of 2013 are partially treated.

[25] We discussed the possibility of identifying such cases *ex ante* with administrators at Minnesota DHS. Placements that seem straightforward at case start can end up having deeper problems that ultimately prevent returning the child to a family of origin. Parental rights are sometimes dissolved late in the process, after the child has been living with their prospective adoptive parents in foster care for some time.

[26] The 16 month window of expected case length is similar for the predicted to not reunify sample and the whole sample by construction: 16 months is the average length of foster cases in the pre-Northstar period *excluding* reunifications.

[27] Specifically ,we modify Equation 2 to be: $Y_{(iat)} = \sum_{t=2011}^{2016} ShareOld_i \delta_{t(0_i)} + \beta_2 X_i + \gamma_{a_i^0} + \epsilon_{iat}$. Where $ShareOld_i$ is the share of the expected case spent over the age of 6 based on age at case start. $\delta_{t(0_i)}$ is a set of dummy variables for calendar time of case start.

Treatment over time is illustrated in Panel A of Appendix Figure A-1, which graphically shows the variation in treatment exposure generated by this design. In addition to six month bins, we also show results with more-aggregated one-year bin time dummies, to improve statistical power. To clearly identify treatment effects, we focus the event study on the sample less likely to reunify.

Another concern is bias that arises from systemic unobserved differences between old and young children. Narrowing the sample's age range closer to the age 6 threshold should reduce this bias, making those on either side more similar. However, excluding ages reduces the sample size, a bias-variance trade-off. We address this through analyzing different samples: children of all ages (-6 to +12 years around sixth birthday), those aged 2–9 only (± 4 years), those aged 4–8 only (± 3), and those aged 3–7 only (± 2). Results are largely similar across bandwidths. Finally, we assess balance on observable characteristics by using case demographics as outcomes in Equation (1) and showing that changes in these demographics aren't predicted by the policy.

4 Child Human Capital Outcomes: Results

4.1 Results: Academic Achievement

Exposure to the payment reform substantially boosts academic achievement three to four years after case start. Analysis of the full sample in Table 2, only controlling for age and year fixed effects, shows a 0.32 standard deviation improvement in text scores (Col. 1). The coefficient stays nearly the same after adding controls for demographics and reason for removal (Col. 2). Impacts are generally similar for reading and math scores though with slightly larger and more precise estimates on math (Panels B and C). Focusing only on the less likely to reunify subsample (Col. 3), effects are larger at 0.46 of a standard deviation.

One-third to one-half of a standard deviation increase in test scores over three years is a large learning gain. These children otherwise face tremendous disadvantage, as reflected in a pre-treatment average test score 77 percent of a standard deviation below the all-student average. Section 5 will put these results into the context of the policy's estimated effect on the amount of money coming into the household from the state. Overall, the implied returns—in terms of academic achievement relative to extra money spent—are high.

4.2 Additional Validation Tests

Bias could come from unobserved confounding differences between older and younger children. To first address this, we assess robustness to narrowing the sample's bandwidth around

Table 2: The Impacts of Subsidizing Permanency on Academic Achievement Test Scores Three Years After Case Start

	(1)	(2)	(3)
Panel A: Child's Average Achievement z-Scores			
Policy Exposure	0.319 *	0.311 **	0.455 **
	(0.170)	(0.155)	(0.214)
Pre-policy Mean	-0.78	-0.78	-0.71
# of cases	6908	6908	3155
Panel B: Child's Math z-Score			
Policy Exposure	0.334 *	0.308 *	0.411 *
	(0.177)	(0.165)	(0.225)
Pre-policy Mean	-0.81	-0.81	-0.74
# of cases	6142	6142	2838
Panel C: Child's Reading z-Score			
Policy Exposure	0.250	0.275	0.468 **
	(0.186)	(0.169)	(0.235)
Pre-policy Mean	-0.72	-0.72	-0.67
# of cases	6160	6160	2902
Controls	No	Yes	Yes
Sample	Full	Full	not reunify

Source: Minnesota Dept, of Human Services foster case data and Minn. Dept, of Education K12 data.
Notes: This table shows results from estimating Equation 2 where each observation is a foster care case. Each coefficient above is from a different regression. Panel A shows effects on the outcome of the average of both math and reading z-scores. Panel B shows effects just on math, Panel C just on reading. Columns 2 and 3 include controls for child race, ethnicity, gender and reasons for the child's removal to foster care. Race and ethnicity categories are White non-Hispanic, Black non-Hispanic, Hispanic, Native American non-Hispanic, and other. Reasons for removal categories are neglect, physical abuse, care taker drug use, behavioral problems, and other. The first two columns are estimated on the full sample while the third column is estimated on the subsample of cases that were predicted to not reunify based on a random forest algorithm. See the text for more details. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

the sixth birthday policy threshold (Table 3, Panel A). Cases are included if the child's age at case end is within the relevant age bandwidth. The baseline specification on the full sample, including controls, is reproduced to ease comparison (Column 1). First, we drop cases of children who spend their policy exposure window outside ±4 years of the sixth birthday threshold; that is, we only include children ages 4 to 9 when looking at test scores (Column 2) or 2–9 when looking at time to permanency discussed below.[28] Next, the sample is narrowed only to cases of children within ±3 and ±2 years of the age threshold (Columns 3 and 4). Estimated effects are largely robust to changes in bandwidth. Though the coefficients decrease slightly for the ±4 and ±3 age bandwidths, the results are qualitatively similar. Further, a key goal of the payment reform policy was to increase older children's timeliness to permanency: excluding the older children excludes many of the policy's intended beneficiaries. The estimates lose statistical significance in the narrowest ±2 bandwidth sample, but the point estimate is similar to that of the 3–8 age bandwidth. We keep all children in the sample both for power and to broaden the external validity of our main estimates.

Estimating the test score model requires a number of specification decisions against which we show robustness. As an alternative to the assumed 16-month policy exposure window set equal to the average length of non-reunification, pre-reform cases, we assess the impact of assuming a policy exposure window that is instead: 1) the average foster care case length observed entirely in the pre-reform period including reunification, which is seven months; 2) individualized predicted case length based on a linear regression model estimated on pre-reform observations and using covariates fixed at case start; and 3) the exposure window is only the case's first month (Treated = Post × Old at case start), implying a case is treated if and only if the child is exposed for their entire foster case. We also assess robustness to which test scores are included (reading, math, or their average), whether any available $E = 2$ test scores are averaged together with $E = 3$, and whether to include covariates in the model. Finally, we show robustness to the sample used: the full sample, the sample predicted less likely to reunify using linear regression, or the sample predicted less likely to reunify using our random forest model.

We illustrate the robustness of the result to these decisions across this wide range of decisions in a specification curve figure (Figure 4). Each point on the figure is a coefficient estimate, with 95 percent (and 90 percent) confidence intervals shown in light (dark) grey bars, and the particular specification detailed below. A blue diamond marks our preferred specification (matching Table 2, Column 2, Row 1). Several key patterns are evident. First, results are robust across many specifications. Estimates from all specifications are positive

[28]The difference is because age is already restricted to be no younger than 4 in the test score sample, so the Column 2 bandwidth is binding only for older children.

Table 3: Robustness to Varying Age Bandwidths: Effects on Average Test Scores Three Years after Case Start and Foster Care Exit Rates

	(1)	(2)	(3)	(4)
Sample:	All Ages	Ages 2-9	Ages 3-8	Ages 4-7
Bandwidth from 6th Birthday:	[-6,12]	[±4]	[±3]	[±2]
Panel A: Average Achievement z-Scores†				
Policy Exposure	0.31**	0.28*	0.27*	0.25
	(0.16)	(0.16)	(0.16)	(0.16)
Pre-policy mean	-0.78	-0.73	-0.73	-0.69
# of cases	6,908	4,597	3,772	2,908
Panel B: Exit to Any Permanency				
(Age 6+) x (Post 2014)	0.21***	0.11*	0.16**	0.24**
	(0.06)	(0.07)	(0.08)	(0.11)
Pre-policy mean				
# of cases	52,334	23,845	18,220	13,049
Case-month observations	667,992	295,885	218,864	144,981

Source: Minnesota Dept. of Human Services foster case data linked to Minn. Dept. of Education K12 data. *Notes:* Each coefficient in the above table is from a different regression. Panel A shows estimates of Equation 2 on average within-child math and reading z-scores where each foster care case is a separate observation, and limiting the regression to the test score linked sample. Panel B shows coefficients from the estimation of Equation 3 on the likelihood of exiting foster care to permanency in a given month, where each observation is a foster care case-month. In both panels, the first column includes all ages. Subsequent columns tighten the age band around the age 6 threshold. All specifications control for child race, ethnicity, gender and reasons for the child's removal to foster care. Race and ethnicity categories are White non-Hispanic, Black non-Hispanic, Hispanic, Native American non-Hispanic, and other. Reasons for removal categories are neglect, physical abuse, caretaker drug use, behavioral problems, and other. See the text for more details.
† *Caveat*: given the constraints of observability window and testing starting at the end of third-grade, the linked test score sample does not include any children who started their case younger than age 4; in Panel A, only the top of the age range changes across columns.

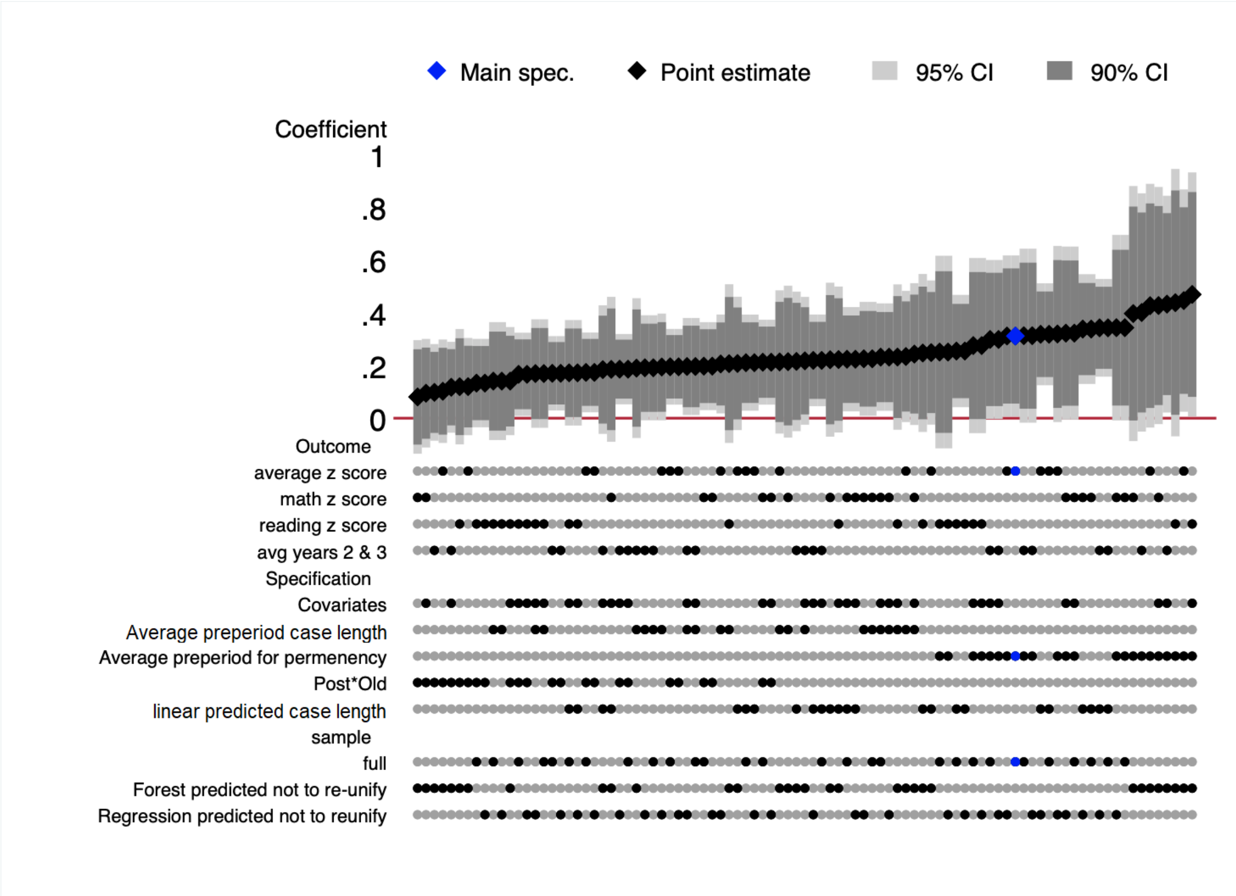
and the majority are statistically significant. Our preferred specification is not one that has a particularly large coefficient nor a particularly small confidence interval. Results are robust regardless of which test score measure we use, whether test scores two years after case start are averaged to the outcome, and the timing on how policy exposure is assigned, though coefficient estimates tend to be smaller for reading z-scores.

While coefficients range from approximately 0.15 to 0.60, estimates most often lose significance when in specifications that combine two assumptions: a case is treated only if it starts post 2015 at age 6 and older (i.e. $L_i = 0$) *and* including only the random forest's less likely to reunify subsample. The less likely to reunify sample has the longest case lengths (whether reunified or not) and thus the longest actual exposure window, whereas assigning treatment based on Post $\times$ Old is equivalent to assuming the exposure window is only the case's first month. Therefore, many cases that are in fact treated whose case started over a year before January 2015 (due to the long de facto exposure window of cases predicted not to reunify) are assigned as not being treated. This measurement error in turn attenuates the estimates. We generally see the largest coefficients for the less likely to reunify subsample when one of the longer treatment exposure window assumptions are made.

Next, we turn to an event study to illuminate trends in the difference between treatment and control groups. In Figure 5 the outcome is the average of reading and math scores. We plot coefficients in six-month bins as the light grey circles without confidence intervals starting in 2011. Overlaid with this, we also plot yearly bins for more statistical power as the black circles with confidence intervals starting midway through 2011 (11-12). Both sets of estimates include controls. Differential trends in test scores between younger and older exposed foster children are fairly flat between 2011 and the first half of 2013, with no pre-trend evident. The first cases expected to be treated begin halfway through 2013 (16 months before January 2015). For these cases that begin in late 2013, which are predicted to last until after the policy change, we see lasting and increasing relative improvements in achievement tests for older children. The visual magnitude of the effect in Figure 5 is comparable to our estimates in Table 2.

Another identification concern is that unobservable case characteristics could be changing over time in a way that is correlated with the policy. To test this, we perform a balancing test where we put observable case characteristics on the left-hand side of our estimating equation. This test is meaningful under the assumption that, if observable characteristics are not changing systematically with the policy, then unobservable characteristics are unlikely to do so. Dropping all categorically related controls from the model (e.g., dropping all race and ethnicity dummies when regressing on Black), there is no significant relationship between our policy variable and indicators for gender, race, and ethnicity (Black, White, Native,

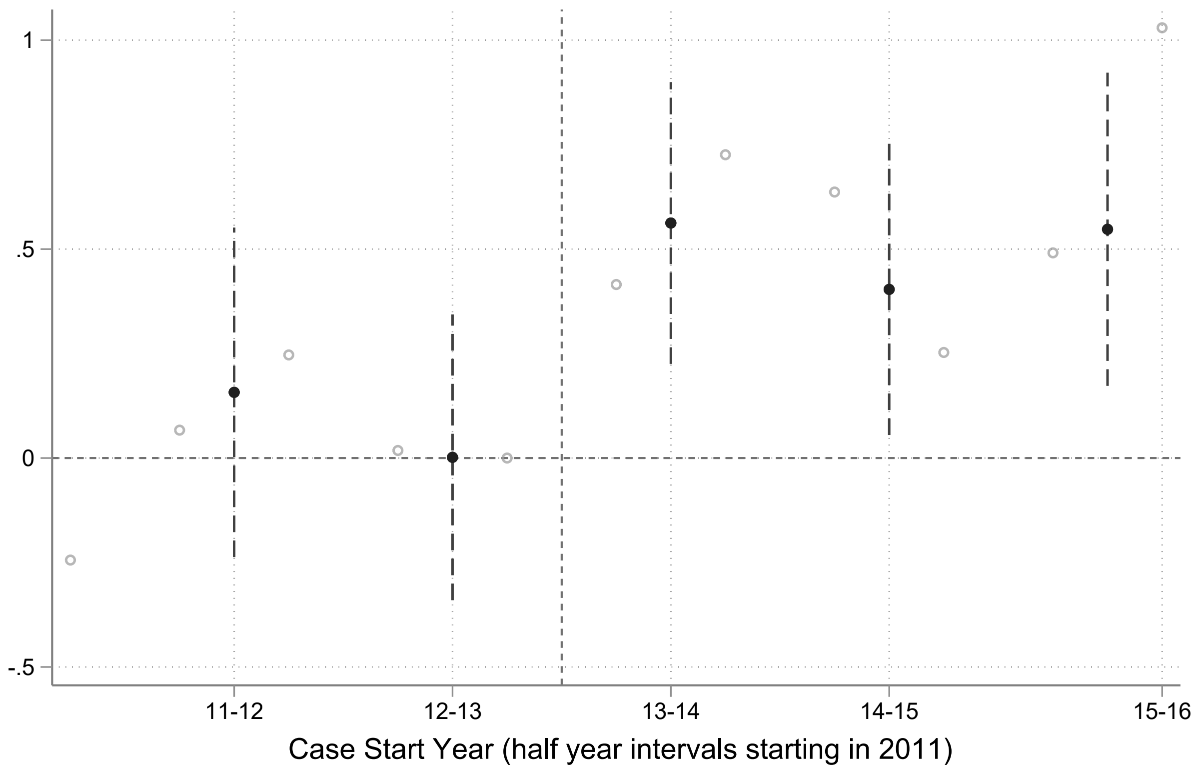
Figure 4: Specification Curve



Source: Minnesota Dept. of Human Services foster case data and Minn. Dept. of Education K12 school data. See Equation 2 in the text.

Notes: This figure shows estimates of the effect of reform on academic achievement across a range of models with different combinations of outcomes, covariates, policy exposure windows, and samples. Each specification uses a difference-in-difference model of a test score measure. Each small black square on the figure is a different coefficient estimate, with its 90% confidence interval in dark grey and 95% interval in light gray, for a different specification. The main specification is highlighted as the blue diamond. Each row below the graph identifies particulars of the specification. The first four rows specify the outcome used, either the average math and reading z-scores from 3–4 years after case start, these math and reading z-scores separately, or average math and reading z-scores averaged over both 2–3 and 3–4 years post case start. The second five rows specify whether covariates fixed at case start are included and how predicted length of case exposure to the policy (L_i) is defined. Exposure variations are average pre-period case length, average pre-period case length excluding reunification, Post×Old (i.e., only treated if the case starts post Northstar and when the child is 6 or older), or using linear regression on pre-reform cases to predict case length with the same case covariates used as predictors. The final three rows specify three different samples: the full sample, the subsample predicted as less than 80% likely to reunify using a random forest model, and the subsample predicted as less than 80% likely to reunify using linear regression (again with predictions based on the case covariates).

Figure 5: Event Study on Standardized Achievement Test Z-scores: Three Years after Case Start by Calendar Time of Start



Source: Minnesota Dept. of Human Services foster case data and Minn. Dept. of Education K12 data.
Notes: This figure shows how differences by age group in the average of math and reading test scores 3 years after case start change over calendar time as different cohorts become exposed to the policy reform. Specifically we modify Equation 2 to be: $Y_{(iat)} = \sum_{t=2011}^{2016} ShareOld_i \delta_{t(0_i)} + \beta_2 X_i + \gamma_{a_i^0} + \epsilon_{iat}$. Where $ShareOld_i$ is the share of the expected case spent over the age of 6 based on age at case start. $\delta_{t(0_i)}$ is a set of dummy variables for calendar time of case start. Small gray dots show coefficient point estimates from a model that aggregates $\delta_{t(0_i)}$ by half year cohorts starting in 2011. The solid black dots show coefficient estimates from a separate model aggregating $\delta_{t(0_i)}$ into yearly bins starting midway through 2011 (11-12). In both cases the outcome is the average of math and reading z-scores. In both models, the graph is normalized so that the coefficient representing the first six months of 2013 equals 0 for ease of comparison with prior and subsequent years. For the cohort well before the reform (2011-12), the coefficient expressing the older-group gradient in test scores is similar to that of the last cohort fully not exposed to the reform (2012-13). For subsequent cohorts, which are exposed to the reform, the older-group gradient in test scores increases, showing that the older-child differences over time are larger than younger-child differences over time.

Hispanic, or Other), nor reasons why the child was placed in foster care (Table 4). This holds for both the full sample (Panel A) and the subsample predicted less likely to reunify (Panel B). Further, in addition to the coefficients being insignificant, the coefficients' signs do not consistently suggest that advantage is increasing or decreasing with the sample.[29]

[29]For example, the sign on neglect cases is positive and children from neglect cases are relatively more advantaged than cases involving abuse; however, the sign on female cases is negative and girls tend to perform better in academics than boys.

Table 4: Balancing Test: Northstar Policy Exposure Regressed on Case Characteristics

	Female	Black	Native	White	Hispanic	Neglect	Abuse	Drug Use
				Sample: All				
Policy Exposure	-0.095	0.019	-0.045	0.019	0.019	0.025	-0.020	0.030
	(0.091)	(0.069)	(0.068)	(0.089)	(0.052)	(0.084)	(0.056)	(0.080)
Pre-policy mean	0.48	0.19	0.16	0.40	0.09	0.33	0.14	0.20
Obs	6908	6908	6908	6908	6908	6908	6908	6908
				Sample: Predicted to Not-Reunify				
Policy Exposure	-0.047	0.066	-0.087	-0.061	0.024	-0.033	-0.038	0.027
	(0.125)	(0.070)	(0.102)	(0.121)	(0.084)	(0.115)	(0.057)	(0.117)
Pre-policy mean	0.49	0.12	0.22	0.43	0.09	0.35	0.08	0.31
Obs	3073	3073	3073	3073	3073	3073	3073	3073

Source: Minnesota Department of Human Services foster case data and Minnesota Department of Education K12 school data.
Notes: This table shows results from a "balancing" test where we put indicators for various case characteristics that are fixed at the start of the case as outcomes of Equation 2 and test if the policy change predicts differential changes in these case characteristics by age group. Each number and standard error reflects the coefficient on the Northstar policy variable from a different regression. Case characteristic indicator outcomes are listed in the column headings. Models include covariates leaving out those sets cateogrigically related to the outcome of interest: ie. When testing the effect of the policy on indicators for race we only control for reason for removal indicators and gender.

Lastly, we check whether results are driven by the change in the assessment rubric used to set the amount of the foster care payment that occurred with the Northstar reform. In both policy regimes, there was a basic monthly foster payment determined only by child age. There could also be supplemental payments that rise with child and parenting needs.[30] Including assessment level fixed effects does not qualitatively change our results: the coefficient on $ShareTreated$ stays at 0.319 while the standard error declines to (0.156). This is as expected, since the assessment did not change around the age 6 threshold used to identify our policy.

4.3 Other Educational and Mental Health Outcomes

Next we study effects on additional education and health care outcomes, an indicator for any out-of-school suspensions, average school attendance rate, and average number of schools attended per academic year (school instability), as well as an indicator for the use of any mental health services. This analysis focuses on the sample of children observed in any Minnesota K12 school. Unlike test scores, these outcomes are observed every academic year. To provide context for the test score results, we partition observations by elapsed years since case start into two groups: those in years leading up to the analyzed test scores (in the first three years after case start $= 0 \leq E \leq 2$) and outcomes in the three years after that ($3 \leq E \leq 5$). In robustness tests, we also show results separately by each elapsed year.

Estimates imply a substantial, persistent decline in child behavioral problems. The policy reduces the probability of having any out-of-school suspension between case start and test score measurement by 2.1 percentage points when including controls. The pre-reform proportion of children with any suspension is 19 percent, implying an 11 percent reduction relative to the base rate (Table 5, Panel A, Column 2). In years three through five, we see a long-term effect of a 3 percentage point decline in having had a suspension or 21 percent of the mean (Col. 4). In both cases the magnitudes on the coefficients decline only slightly after adding controls.

Counter to these positive results, we see evidence that the policy change caused a small decline in school attendance rates three to four years after case start, averaging 1.2 percent fewer days in school annually, which is about two days a year (Panel C, Columns 1 and 2). If

[30] Before Northstar, DHS assessed every child in foster care using the Difficulty of Care (DOC) rubric. After Northstar, it shifted to using the Minnesota Assessment of Parenting for Children and Youth (MAPCY) rubric. We observe each case's points value on its contemporaneous rubric. DHS developed a crosswalk to approximate the equivalence between DOC point ranges and MAPCY point ranges (Figure A-2). Each row corresponds to a payment category with relevant DOC and MAPCY point ranges corresponding. Need and payments rise moving down the table. This allows us to break cases into more-homogeneous groups that are stable across the whole time period. We define fixed effects for the rows and assign these to cases based on each case's DOC or MAPCY score. This brings in more refined measure of child types, capturing if some children are systematically more likely to get higher letters in a way that is correlated with the treatment.

Table 5: Estimated Impacts of Payment Reform On Other Educational and Mental Health Care Outcomes

	(1)	(2)	(3)	(4)
Elapsed years since case start	0-2	0-2	3-5	3-5
	Panel A: School Suspensions			
Policy Exposure	-0.035***	-0.021***	-0.038***	-0.030***
	(0.008)	(0.007)	(0.010)	(0.010)
Pre-Reform Mean	0.19	0.19	0.14	0.14
Obs	33,824	33,824	20,407	20,407
	Panel B: Attendance			
Policy Exposure	-0.012**	-0.014***	0.0001	-0.002
	(0.005)	(0.005)	(0.005)	(0.005)
Pre-Reform Mean	0.88	0.88	0.88	0.88
Obs	27,393	27,393	17,204	17,204
	Panel C: Average # of Schools per Year			
Policy Exposure	-0.082***	-0.041**	-0.005	0.009
	(0.019)	(0.018)	(0.023)	(0.023)
Pre-Reform Mean	1.49	1.49	1.41	1.41
Obs	33,824	33,824	20,407	20,407
	Panel D: Mental Health Services			
Policy Exposure	-0.016***	-0.007	-0.007	-0.004
	(0.006)	(0.006)	(0.007)	(0.007)
Outcome Mean	0.10	0.10	0.06	0.06
Obs	33,824	33,824	20,407	20,407
Controls	No	Yes	No	Yes
Sample	Full	Full	Full	Full

Source: Minnesota Department of Human Services foster case data, Minnesota Department of Education K12 school data, and Minnesota Department of Health Medicaid data.
Notes: This table shows results from estimating Equation 2 on outcomes other than test scores. Each row and column shows a coefficient from a different regression. The outcome shown in Panel A is any out of school suspension over the defined elapsed time period. Panel B shows average attendance over the defined elapsed time period. Panel C shows average number of schools attended per year averaged over the defined elapsed time period. Panel D shows any mental health service use defined over the elapsed time period. Columns 1 and 2 define elapsed time as 0-2 years after case start. Columns 3 and 4 defined elapsed time as 3-4 years after case start. Columns 2 and 4 add controls for race, ethnicity, gender and removal reason.

finalizing a permanency placement requires legal obligations that remove the child from the classroom, this could explain the declines in attendance. Regardless, improved attendance rates do not seem to be a mechanism for raising achievement. Attendance rate effects fade out in the longer-term period of three to five years after case start.

We next estimate effects on school instability. Children experiencing foster care struggle to maintain continuity in the same school, averaging 1.5 schools per year in our sample in the three years after case start pre-policy reform. Payment equalization causes a small reduction in instability. Students attend 0.04 fewer schools per year, a 2 percent decline, in the first three years after case start. While small, this increased stability potentially contributes to improvements in test scores. Effects on school instability fade out four to five years after case start.

We observe an indicator for any mental health service use in matched Medicaid records and look at effects in the given time frames. We see declines in the probability of mental health service use in the first three years without controls (Panel B, Column 1), though the estimate is not robust to controls and attenuates (remaining negative) during four to five years after case start. It is not clear if a decline represents improvements in mental health or loss of access to care. For most of these outcomes we see a similar (albeit noisier) pattern when examining individual elapsed years (Appendix Table A-2).

4.4 Heterogeneity

Next, we estimate heterogeneity of effects across different subgroups. Doing so is of interest to understand which groups experience the largest marginal returns from the policy. We look at differences by child demographics (race, ethnicity, and gender) as well as reason for removal (neglect, physical abuse, caretaker drug use, or one of the many other smaller categories of abuse). It is not obvious *ex ante* if the additional incentives to adopt and subsequent improvements in achievement are likely to be larger for more or less disadvantaged groups of children. For example, children removed for neglect instead of physical abuse may be more likely to benefit from larger payments due to there being more candidate adoptive families willing to take on a child with less severe problems. On the other hand, it may be less beneficial because the additional payments may have the highest marginal benefit in cases where the child is the most in need. Overall, we see larger test score effects among male and Native American children as well as those who are removed due to parental drug abuse (Appendix Table A-3). Strikingly, the test score effects become small, not significant, and negative for Black children, though the standard error is large. Effects are fairly similar across the other different reasons for removal.

5 Costs and Payments

We now estimate how the payment reform policy changed the stream of payments going from the state of Minnesota to impacted households. This analysis has two main purposes: 1) to estimate the fiscal cost of the policy to Minnesota, and 2) to estimate the magnitude of the increase in family income as a potential mechanism causing the improvement in test scores and other child outcomes. To line up the timing of the main benefit (test scores) with the cost (payment stream), we focus on measuring payment streams between case start and the time of the $E = 3$ test score.

DHS payment data shows each monthly foster care payment associated with each foster case and, if a child moves into permanency, the first payment made to the adoptive family. We calculate the total payment stream that goes to each foster child's foster and permanent families as 1) the sum of any foster payments observed in foster care added to 2) the observed adoptive payment for cases that end in permanency, repeated monthly from when the payment is first observed at case end until the standardized testing year and month. Because the time path of payments affects public cost, we also compute a net present value of each monthly payment stream by discounting monthly payments using the yield rate on the State of Minnesota's taxable General Obligation Bonds, 4.6 percent in August 2023. We inflation adjust all payments to 2022 dollars. The payment data is more limited than our main data, only covering 2013 to 2019. Appendix D gives more details about the data on payments and how we construct the payment outcomes.

We estimate the impact on payments using the same strategy used to estimate test score effects (see Equation (2)). The estimated policy effect on average total payments from case start through month of the test is \$2,141 or \$1,964 in net present value, 9 percent of the pre-reform mean (Table 6, Columns 1 and 2). Average monthly payments of all kinds increased by \$52 a month (Col. 3). Average payments here include foster and permanency payments for the entire sample, including those who do not exit to permanency. Since the policy only increased payments in permanency but not foster care, we estimate the policy effect on average foster payments as a placebo. As expected, the estimated effect is small and not significant. Next, we estimate effects on monthly payments only for those whose case ended in a permanency placement: up \$128 (17 percent) among those in adoption and about \$362 (56 percent) among those in kin guardianship. Since the reform equalized payments in adoption and kin guardianship, we would expect a larger increase for kin guardianship as these tended to receive a lower pre-reform payment level relative to adoption.

Table 6: Effects on Payments

	(1)	(2)	(3)	(4)	(5)	(6)
Payment Outcome:	Total	NPV	All Monthly	Foster	Adoption	Kinship
Policy Exposure	$ 2,141**	$1,964 **	$ 48 **	$ -9	$ 128 ***	$ 362 ***
	(958)	(888)	(18)	(17)	(27)	(29)
pre-policy mean	$22,965	$21,702	$448	$ 1,051	$ 736	$648
Demographic Covariates	Yes	Yes	Yes	Yes	Yes	Yes
# of Foster care spells	19,117	19,117	19,117	19,117	3051	3331

Source: Minnesota Dept. of Human Services data on payments. See Section 5 and Appendix D for details.
Notes: This table shows results from estimating Equation 2 on different measures of payments to families from the state of Minnesota for fostering or taking a child into permanency. Data come from the Minnesota Department of Human Resources payment records between January 2013 and December 2016, inflation adjusted to be in 2022 dollars. Each observation is a different foster care case, and each row and column shows a coefficient on the policy exposure variable from a different regression. Total payments (the first column) are calculated as the sum of all payments made in foster care and in permanency (if applicable) paid to the family from the start of foster care until the month before the standardized achievement test taken three years after foster care start. Net Present Value (NPV – the second column) discounts the total payments by using the discount rate implied by the State of Minnesota borrowing cost in August 2023: 4.6% in August 2023. The third column shows average monthly payments for any type of payment (foster payments, or payments made to adopted or kinship care families). The fourth column shows results on only the average foster care payments made to the child (excludes payments made in permanency). The fifth column only includes those children who exited to adoption and shows results on the average adoption payments made to those children. The sixth column does the same thing but for kinship care. All specifications control for child race, ethnicity, gender and reasons for the child's removal to foster care.

6 Time in Foster Care: Estimation and Results

We examine how payment reform impacted time in foster care and the likelihood of exit into permanency. Prior descriptive evidence suggests that both prolonged stays in foster care are harmful to children and that children do best in a safe and supportive permanent home.

Here, we use a survival analysis where for each child in foster care we can observe whether or not they exit each month, so each observation is now at the foster care case by month level. We adapt Equation (1) so that Y_{iat} is an indicator variable equal to one if the child exited foster care in month t and zero otherwise. Our main analysis focuses on exit to any type of permanency. In supplementary results, we consider exits to adoption and to kin guardianship separately.

The probability of exit in a month is a function of the duration of the foster care episode to that point. Without accounting for duration dependence, estimates of the DiD could potentially be biased. We account for this using survival models, beginning with the standard Cox proportional hazard, which the earlier literature has used in DiD models on foster care cases (Buckles, 2013). Consider survival time modeled as:

$$h_{iat,p|\mathbf{x},\beta} = h_0(p)e^{\mathbf{x}'\beta} \tag{3}$$

where $h_{iat,p|\mathbf{x}}$ is the hazard of exit to permanency for child i, who is age a in calendar time t, while p signifies that the child has remained in foster care for p periods so far (survival duration).

Using the Cox model as a functional form for the duration dependence, we adapt the research design discussed in Section 3 by substituting the right side of the linear model in Equation 1 into $\mathbf{x}'\beta$. For ease of reading, we take logs of both sides of Equation (3):

$$\ln(h_{iat,p}) = \lambda(p) + \beta_1(t \geq 2015)\mathbf{1}(a \geq 6) + \beta_2 X_i + \gamma_a + \delta_t + \epsilon_{iat} \tag{4}$$

where β_1 is the DiD effect of payment equalization on the hazard. $\lambda(p)$ is the baseline hazard, which captures the likelihood of exit as a function of duration of the case (note that $\lambda(p) = ln(h_0(p))$. The baseline hazard is directly estimated in the Cox case without assuming a functional form. Standard errors are clustered at the child level.

Cox proportional hazard models are easy to interpret, commonly used, and avoid making assumptions about the form of the baseline hazard; however, they do assume that the hazard functions across treatment and control groups are proportional over the duration of the foster care episode.[31] To make sure the results are not sensitive to this assumption, we also estimate

[31]Such an assumption could be violated if, within a foster care case, exit probabilities change in the

discrete time hazard and linear probability models, neither of which make the proportionality assumption. Results prove similar across all models.

To test for differential pre-trends, we reformulate Equation (4) as an event study by dropping the single interaction term $\mathbf{1}(t \geq 2015)\mathbf{1}(a \geq 6)$ and replacing it with a vector of interactions between $\mathbf{1}(\mathrm{a} \geq 6)$ and each calendar year. Normalizing against the difference in the immediate pre-reform year, the trend in these calendar-year specific group-difference estimates reveals any trends in the difference leading up to the policy change.

6.1 Results: Time to Permanency

Here we estimate the impact of Northstar's payment reform on the monthly probability of exit from foster care into permanency and find that, across models, the policy is estimated to substantially raise this likelihood (Table 7). Results are similar across Cox models that include only covariates needed for identification—year-month fixed effects and age fixed effects (Column 1)—and our preferred specification that adds case covariates (Column 2). Positive coefficients express a percent increase in exit probabilities for the older relative to younger children and negative coefficients a percent decrease.[32] Our preferred specification that uses the Cox model and includes case characteristic controls finds that payment equalization led to a 29 percent (not percentage point) increase in the probability of exit from foster care to permanency in a given month. As a first alternative to the Cox model, we estimate a linear probability model predicting an indicator of exit each month presented in percentage points. The result implies a 0.5 percentage point higher likelihood of exit off a pre-policy mean of 1.6 (Column 3), implying a 31 percent increase, similar to the Cox model estimates. A discrete-time survival model, another alternative (Column 4), yields a similar estimate: a 23 percent increase in monthly exit probability.

Panel B of Table 3 presents results for exit to permanency when narrowing the sample's age bandwidth around the sixth birthday policy threshold. Coefficients decrease somewhat for the age 2–9 and 3–8 bandwidths, though estimates return to a 22 percent increase in exit probabilities in the narrowest 4–7 years bandwidth. This is quite similar to estimates from the full sample, though the standard errors are almost double. As with test scores, shown in Panel A and discussed in Section 4 above, this provides evidence that results are not driven by differences in unobserved confounding factors between children of different ages on either side of the age threshold.

treatment relative to control group conditional on the same duration p spent in the episode. A necessary condition for this to hold is that the survival functions of the two groups do not cross; they do not (Figure 3). See `https://bookdown.org/sestelo/sa_financial/how-to-evaluate-the-ph-assumption.html`

[32]This is derived simply by subtracting one from our estimates of the hazard ratio.

Estimated hazards give the relative decrease in time to permanency for treated children. We translate this into decreased expected time in foster care by converting the hazard model estimates into survival probabilities relative to a counterfactual case where the policy did not happen. Doing so yields an average decline of time spent in foster care of 5.2 months.[33] We compare this decreased time in foster care with the increased money and other potential mechanisms to shed light on our child human capital results in Section 4.

Table 7: Effects of Northstar Payment Equalization on Foster Care Exits

	(1)	(2)	(3)	(4)
Exit to Permanency: Adoption or Kin Guardianship				
(Age 6+) x (Post 2014)	0.28 ***	0.21 ***	0.005 ***	0.23 ***
	(0.06)	(0.06)	(0.0007)	(0.06)
pre-policy mean			.016	
% impact	28%	21%	31%	23%
Model	Cox	Cox	LPM	Discrete time log-log hazard
Controls	No	Yes	Yes	Yes
# of Foster care spells	52,334	52,334	52,334	52,214
Observations	667,992	667,992	667,992	662,906

Source: Minnesota Dept. of Human Services foster case data.
Notes: An observation is a year-month that a child is observed in a foster care case. Each column and row are from a separate DD regression of the interaction between being age 6+ and in the post Northstar period (2015+) with age and year-month fixed effects regressed on whether there was an exit to permanency in that month (see equation 4 in the text). Columns 1-2 show results from cox proportional hazard models without and with controls. Column 3 shows results from a linear probability model. Column 4 shows results from a discrete time hazard model. For the hazard models we report relative hazard ratios after subtracting one from them. We subtract 1 from the ratios so that they reflect the effect of treatment on the % change in the outcome relative to the comparison group. For the linear probability model we multiply the coefficients by 100 so they reflect a percentage point likelihood of exit in a given year-month, and we then calculate the % change relative to the pre-policy mean for comparison with the hazard models. Models with controls control for child race, ethnicity, gender and reasons for the child's removal to foster care.

[33] To do this, we derive one set of survival probabilities based on the estimated parameters and observed cases, and another under the counterfactual that the policy was never implemented; i.e., the 1(Age 6+)*1(post 2014) term is set to zero for all children. For each of the two sets of estimates, we then take the average of the survival probabilities over all observations for every month of foster care duration, giving us two estimated predicted average survival curves. The difference in the areas under the two survival curves expresses the estimated effect of the payment reform on time in foster care.

To give evidence on possible differential pre-trends, we estimate an event study using the Cox specification for exit to any type of permanency from Column 2 of Table 7. One series includes only age and time fixed effects (triangle: corresponding to Column 1 above). Overall we see a slight downward trend in the pre-period, followed by a clear increase in estimated exit probabilities following the policy reform (Figure 6). The second (box-shaped) series directly connects these results to our estimated effects on children's later academic achievement outcomes by narrowing our sample to the same age ranges used in the analysis of academic achievement (For academic outcomes, only children ages 4–14 are observed in foster care; see the discussion on academic outcomes in Section 2.2) and limiting to the sample of those who are predicted as unlikely to reunify with their family of origin. For this subsample, the estimated policy impact is substantially larger.[34]

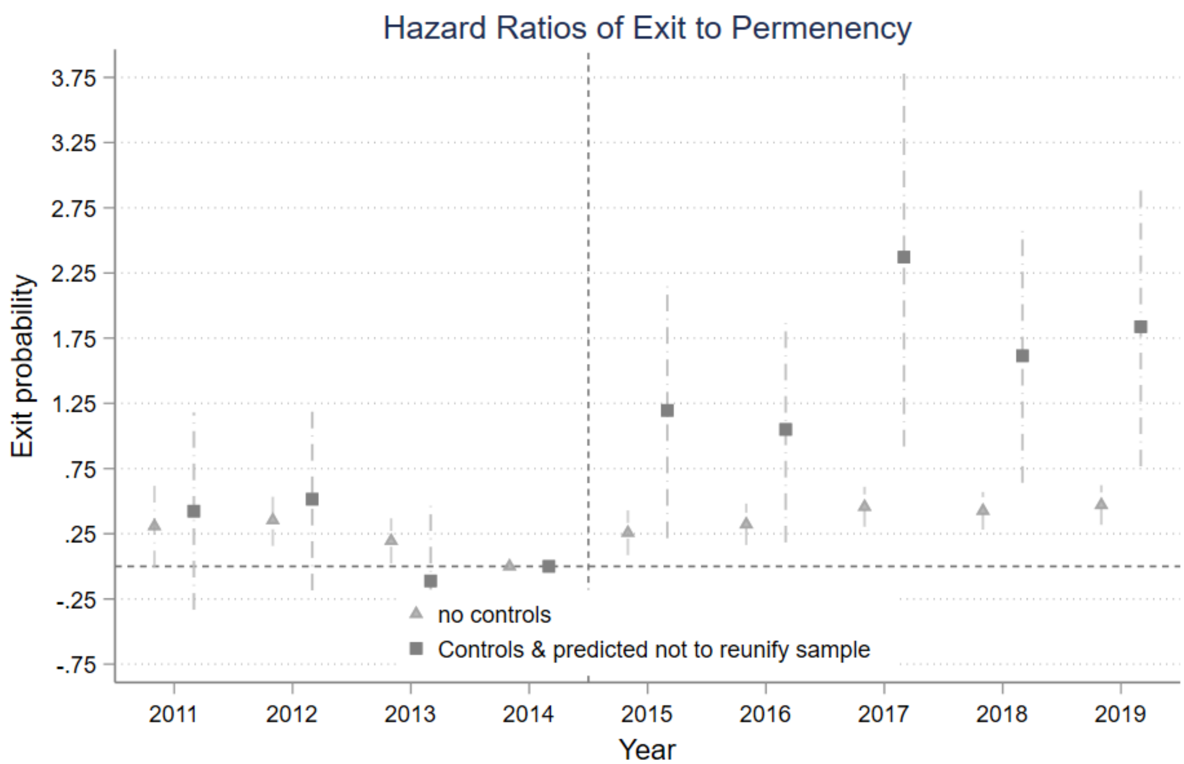
We continue to estimate substantial and significant increases in hazards in separate models for exits to kin guardianship and to adoption (Table A-4). The estimates show a reduction in time to permanency for both sub-types of exits and are similar to each other.[35] We show the event study for exit to adoption and to kin guardianship in the panels of Figure A-3. For kin guardianship, the estimated impact in the years leading up to the policy is relatively flat, with a discrete and persistent increase in kin guardianship in the years following the policy.

The final column of Table A-4 tests to see if there is any impact of the increased permanency payments on reunification with the family of origin. This checks against the perverse incentive of adoption occurring at the expense of reunification. Further, Minnesota DHS believed it was possible that the policy could delay reunification by improving the bargaining power of kin who wish to see improvements in the origin family's home and also who might serve as alternative guardians should reunification fall through. Results in Column 4 show that effects on reunification are negative but not quite significant at the 10% level, and almost an order of magnitude smaller than our main results (3.8 percent versus 21 percent for any exit to permanency).

[34] The event study places these effects of an increase in the hazard of 100–125 percent beyond the control group. Equivalently estimated using a Cox proportional hazard model with controls (equivalent to the model ran to estimate Column 2 of Table 7).

[35] Another robustness test is shown for kin guardianship in Table A-4. As part of the reform, new federal incentives led to a licensure requirement: to become a child's permanent kin guardian, kin needed to spend the prior six months as the child's licensed foster parent (also discussed in Section 2 footnote 9). We show the sensitivity of results to controlling directly for having met the licensure requirement (Column 2) by adding an indicator variable for being post 2015 and having had at least six months as the child's foster parent. Similar requirements already existed for adoptions, so we do not include this control when the outcome is permanency overall or adoption. The difference between Column 1 and Column 2 shows that accounting for whether a child-month is ineligible for kin guardianship under this policy increases the effects somewhat but gives qualitatively similar results.

Figure 6: Event Study for Effect of Policy on Exit from Foster Care to Permanency



Source: Minnesota Dept. of Human Services foster case data. *Notes:* This figure shows estimated hazard ratios from the event study version of Equation 1. Each coefficient estimate represents the proportional increase in exits of older (age 6+) children relative to younger children, in a given year, normalized so that the last pre-reform year (2014) equals 0. The triangles show baseline estimates without controls beyond fixed effects. The boxes show our preferred specification including observable controls (such as child years of age, gender, race/ethnicity and reason for child removal into foster care) and limiting the event study to the predicted unlikely-to-reunify sample. In the preferred specification, the age-group difference in foster care exit hazards is similar in the years leading up to the policy change. After the reform, older children's exit hazards fall less than younger children's yielding a positive treatment effect estimate.

We do additional robustness exercises. To further test for confounding contemporaneous shocks, we perform a placebo test limiting the sample to children younger than age 6, estimate the DiD models keeping the calendar timing of the policy the same (post Jan 2015), but assign a fictitious policy threshold of different birthdays than the sixth. Table A-5 reports results. In Column 1, the threshold is wrongly assumed to be the second birthday, and children ages 0–1 are "control" and those ages 2–5 are considered as placebo treated. Column 2 uses the third birthday and Column 3 the fourth. None of the estimated effects are significant, and magnitudes are all substantially smaller than our main estimates.[36] Finally, as with long-term outcomes above, we include assessment score fixed effects in the model and find no qualitative difference from our main DiD estimates.

6.2 Evaluating Strategic Delays of Permanency Placements

A concern for our identification strategy is possible strategic delay. In the period leading up to the policy reform date, potential adoptive parents or kin guardians serving as foster parents of young children might anticipate that their families would receive a larger stream of payments in permanency if they were to delay the child's exit until the child's sixth birthday. If such delaying of permanency is widespread and remaining in foster care is harmful, then this strategic delay could bias our human capital results upward. This would occur if increased time in foster care causes younger children to do worse, making it appear that older children have relatively improved achievement. Fortunately, the same data setup we use for our time-in-foster-care estimation allows us to investigate this potential "red-flagging" of 5-year-old children.[37]

Such strategic behavior should be observed as a shift in the mass of exits from before to directly after the age 6 threshold. We plot a separate histogram of the number of exits to permanency by age in months for children ages 2–10 separately before and after the reform. Pre-reform, the number of exits declines as children age, consistent with younger children exiting to permanency more quickly, but there is no clear bunching of exits after the age 6 (72 month) threshold (Figure A-4). Post reform, exits also decline as children age, though we can see this trend is offset somewhat by the policy as the decline in exits becomes less pronounced for children over the age 6, consistent with the goals of the reform (Figure A-5). Further, directly after age 6 there is evidence of some bunching in exits between the ages of

[36] We cannot do this test on our long-term outcomes because we are limited to test scores on children who are at least age 4.

[37] Another concern is that foster parents of older children immediately before the reform anticipate the possibility of higher payments if they were to delay until post reform. This would present in the event study of exit rates as a dip immediately pre-reform and an extra bump up immediately post reform. This is not observed (Figure 6).

73 months and 80 months. Likewise, there appears to be a small (off-trend) notch in the number of exits several months leading up to the age 6 threshold. Visually, the shift in mass appears to be relatively small.

To quantify the amount of sorting over the threshold, we draw inspiration from the literature on bunching and notches (Kleven, 2016). These tools from the bunching literature allow us to estimate the amount of displaced mass and translate this into child-months of delayed exit from foster care. Details of this exercise are provided in Appendix E. We find that only 81 cases are delayed, and the average delay of these 81 cases seems to be about 6.4 months. This is likely an overestimate and represents a small portion of our sample that is unlikely to be driving our main regression estimates. At the end of subsection 6.1, we calculated that the policy decreased time in foster care by 5.2 months on average for the 3,860 older children who were adopted or taken into kin guardianship after the policy was enacted in 2015. If 81 of those exits were actually *delayed* by 6.4 months due to red-flagging, we can re-weight the average treatment effect as $ATE = \frac{-5.2*3779+6.4*81}{3860} = -4.95$ months. That is, this anticipatory delay factor would lead to an overstatement of the magnitude of the time-to-permanency effect of 0.2 months.

Alternatively, we directly check this concern by estimating a "donut hole" type regression that removes all child-month observations within a year of age of the sixth birthday threshold (ages 5 and 6) to eliminate cases most likely to be affected by anticipation (Lindo and Waddell, 2011). Estimating our models on this sample produces results that are either unchanged or stronger. Appendix Table A-6 shows results on permanency with and without controls and shows that results are similar when the children most at risk of strategic delay are excluded.

7 Mechanisms

In this section we discuss the magnitudes of our test score impacts and consider mechanisms more closely. We find that a net present value investment of $1,964 per child over three years leads to a test score gain of 30 percent of a standard deviation. Turning to an earlier literature on household income and child achievement, Dahl and Lochner (2017) find that a $2,000 increase in income from the Earned Income Tax Credit increased test scores by 0.06 of a standard deviation.[38] Duncan et al. (2011) find similar effects using experiments from cash welfare. Closer to our findings, Black et al. (2014) find impacts roughly twice the

[38]We took their estimate of 4 percent increase for $1,000 in 2010 dollars and inflation adjusted to be in 2022 dollars, and scaled to a $2,000 increase to be comparable to our own intervention size. We similarly scaled the findings of other papers we cite in this section to make them comparable.

size of earlier work, an increase of 0.12 of a standard deviation, from a child care subsidy of $2,000. Milligan and Stabile (2011) find a somewhat larger increase of 9 percent of a standard deviation in test scores from $2,000 (in U.S.) dollars of income from a Canadian tax credit. Our results therefore are 3–5 times larger than the earlier literature and potentially reflect a "good deal" relative to other investments in child achievement.[39]

An important caveat is that the above discussion assumes our findings are purely driven by increased income for adoptive and kin guardian families. However, there are some key differences between our context and the earlier literature on family income and test scores. First, the above papers estimate the impacts of a single transfer within a year on test scores taken that year, whereas we estimate the effect of an increased payment stream over 3–4 years on test scores at the end. Regular monthly payments may be more valuable to accumulating child achievement over time than a transfer within a single year.

Further, unlike most other transfer programs, permanency payments are guaranteed by the state until the child turns 18. If adoptive parents base their investments in a child on the total expected flow of income from the state, then this could help explain the larger impacts. For example, parents may base their perceived likelihood of a child going to college on the amount of money that is predicted to go into the household up until age 18 and be more likely to prepare a child academically for college if this amount is higher. To get a sense of the magnitude of expected payments through age 18, we extrapolate the results on increased average monthly payments in Table 6. We first calculate the average age of exit in months to permanency for older children in the post period. This allows us to back out an estimate of the expected total monthly payments through childhood (94 months in our sample). We then calculate the effect on the total payment amount between case start and age 18 by multiplying the increase in average monthly payments from Table 6 by the number of months until age 18: $12,048 for adoption and $28,452 for kin guardianship. In net present values, these are $10,121 and $24,796, respectively. This represents a substantial increase in expected income that is more in line with the earlier literature's findings on test score gains. However, it is important to note that the test score gain happens before most of those payments are made.

[39]Rather than benchmarking against income's effects on test scores, one can also benchmark against evidence on educational interventions' costs and effects on test scores. There are a few interventions that show comparable impacts at similar cost. According to the U.S. Department of Education's What Works Clearinghouse, two middle school interventions have cost estimates and strong evidence of positive effects. We translate effects to be per $2,000 (2022$) for comparability. The READ 180 literacy program is estimated to raise scores by 36 percent of a standard deviation (15 percent for $717 in 2017$) and the Knowledge is Power Program 47 percent (24 percent average effect size across math and English for the cost of $771 per student year in 2008$). However, these interventions cannot really be scaled up linearly to produce as large gains as Northstar evidently did.

Further, the impact of additional money on test scores may be larger for foster children than for other children. This could be the case if there are decreasing marginal returns of money on achievement, implying a higher return for this very disadvantaged group of children. As shown in Table A-1, foster children have low scores of -0.77 standard deviations below the state average; placing their average around the 22nd percentile of all students.

An important part of understanding these results is that the reform doesn't just increase family income but is tied directly to promoting permanency, increasing the stability of adoptive families and shortening time spent in foster care. Estimates in Section 6 show a shortening of foster care spells by around 5.2 months. Some of this shortening of foster care time represents permanency payments that otherwise would not have happened. To get a sense of this, we can calculate the number of additional permanency placements within the period of the treatment due to the policy. This comes to there being 474 more placements than in the counterfactual case where the policy was not implemented.

The reduction of 5.2 months of time in foster care could itself be valuable, to the degree that this is less time in a home with foster parents who may not be investing as much in the long-term future of a child, and there are reasons why this could matter. In the common case where the adoptive parent was also the foster parent (or fostering kin guardian), permanency reflects both an emotional commitment and legal obligation on the part of the otherwise fostering parent. Due to the legal liability for the child, parents face more incentives to invest in the child. We believe this speaks to potentially large benefits of incentivizing permanency for children in foster care who otherwise would not return to their families of origin. As one illustration of the change in legal power, kin foster parents often need to bargain with the origin parents over decisions related to care for a foster child. Given that origin parents still have many parental rights over foster children, they could pressure changes in how the child is handled. Full parental rights being transferred to the kin guardian could, in some cases, benefit the former foster child through increasing the kin's bargaining power relative to the origin parent.

In sum, we believe the combination of money and entrance into an adoptive home could be complementary in producing the sizable test score impacts. The increase in financial resources from expected payments is intuitively most valuable in tandem with a commitment to investing these additional resources in the child. Such an attachment is less likely to come from a caregiver who lacks full social, legal, and emotional commitment to the child. Supporting this, Chorniy and Mills (2022) find little impact of paying an increased stipend in foster care on child well-being, contrasting with the Northstar payments that only go to families where the child is taken into permanency. A key empirical question remains: Is the "marginal" permanency placement encouraged by financial incentives more or less stable than average? This would not be the case if financial motives crowd our altruistic ones, but possibly would be the case if the slackening of financial constraints allows more potential matches between adoptive parents and foster children, in turn resulting in better matches. In the next section, we estimate impacts on the likelihood that, after being placed in adoption or kinship care, the child re-enters foster care.

7.1 Stability of Permanency

A final mechanism considered is how payments impact the quality of the permanency matches formed. The quality of a match is difficult to measure because direct parental inputs into children are not measured. As a proxy for quality, we look at how the policy affects the likelihood that a child re-enters foster care after exiting into an intended-as-permanent adoption or kin guardianship arrangement. This analysis benefits from the extra work Minn-LInK did to match cases after name changes into re-entry into foster care. We estimate DiD hazard models similar to Equation 3 but with re-entry into foster care as an outcome. Each case of interest starts with a former foster child's transition into adoption or kin guardianship. Whether or not the child re-enters foster care each month is put on the left-hand side. Treatment is assigned based on having been age 6 or older when they entered permanency and having entered permanency after 2014. The results show insignificant effects on hazard ratios in models both with and without controls (Table 8). In our preferred model with controls, the point estimate on treatment suggests a large 45 percent *decline* in the likelihood of re-entry for treated children. However, the estimates are not statistically significant (t-stat of 1.41). This is likely due to re-entry being (fortunately) relatively uncommon: out of the 10,032 children who enter permanency during our sample, only 87 re-enter foster care to be placed with a new foster family. This in turn limits the amount of statistical power we have to identify effects. Regardless, our results at least show no evidence that the increased rates of exit caused by the payment equalization policy come at the cost of more fragile matches, and the sign on the coefficient suggests that such matches may be *more* stable.

8 Conclusion

In this paper we explored in detail how increasing payments to adoptive parents improves child outcomes. Not only did the Northstar policy increase the flow of money to these families, but equalization of payments between foster care and permanency substantially accelerated exit from foster care into permanency. Our analysis of mechanisms showed no evidence that these adoptive arrangements were less stable. Indeed, the family and emotional

Table 8: Impact of Policy on Foster Care Re-Entry

	(1)	(2)
(Age 6+) x (Post 2014)	-0.45	-0.45
	(0.32)	(0.32)
# of Permanency Spells	10,032	10,032
# of Re-entries	87	87
Model	Cox	Cox
Controls	No	Yes

source: Minnesota Department of Human Services foster case data.
Notes: An observation is a year-month that a child is observed in a permanency arrangement after leaving foster care. Results are from a DiD regression on the interaction between being age 6+ in the post Northstar period (2015+) with age and year-month fixed effects on the likelihood of being placed *back* into foster care after permanency (see Section 7 in the text for details). We estimate these models using a cox-proportional hazard model. We report relative hazard ratios after subtracting one from them. We subtract 1 from the ratios so that they reflect the effect of treatment on the % change in the outcome relative to the comparison group. Column 2 includes controls for child race, ethnicity, gender and reasons for the child's removal to foster care.

life of the child seems to improve, with declines in suspensions and increased school stability.

In turn, these improvements are shown to translate into achievement gains. Three to four years following the start of a foster care case, children experience large increases in academic achievement. Our estimate is that $1,964 in increased net present value of payments results in a 0.3 standard deviation increase in test scores. This is a high return per dollar relative to prior literature on income's effects on child academic achievement. Further, this 0.3 rise in test scores is worth about $32,000 in boosted expected net present value of child's lifelong earnings, an implied benefit to cost ratio of 16.[40] Knowing that foster children are severely disadvantaged academically and may have high returns to spending means they are a relatively easy group to target for financial investment. Such a policy is potentially "low hanging fruit" for states and policymakers who wish to decrease achievement gaps. In part this could be relatively easy to implement because the institutional mechanisms for providing foster families with payments already exist, and simply need to be extended into adoptive or kin guardian homes and/or increased as the Northstar policy did.

Future research can attempt to better understand why the returns to investing in foster children in this context is large, as well as document longer-term gains to adoption and kin guardianship. It would be particularly interesting to follow adoptive and foster children over a longer period of time so that the counterfactual children not exposed to the policy can be

[40]Hanushek (2011) finds +0.13 standard deviation in test scores is worth an average $10,600 in net present value lifetime earnings in 2011 dollars. Inflating to 2022 dollars multiplies this by 1.3. The BC ratio of 16 may be overstated given that foster student earnings might be lower on average and that the $2,141 cost only includes policy costs (and benefits) to the time of the test score.

observed as they transition out of foster care. Given the large negative associations between aging out of foster care and life outcomes, documenting the longer-term benefits of financial incentives to adopt could be particularly valuable.

References

Aizer, A., S. Eli, J. Ferrie, and A. Lleras-Muney (2016). The long-run impact of cash transfers to poor families. *American Economic Review 106*(4), 935–71.

Argys, L. and B. Duncan (2013). Economic incentives and foster child adoption. *Demography 50*(3), 933–954.

Baker, A., D. F. Larcker, and C. C. Wang (2021). How much should we trust staggered difference-in-differences estimates? *Available at SSRN 3794018*.

Bald, A., E. Chyn, J. S. Hastings, and M. Machelett (2019). The causal impact of removing children from abusive and neglectful homes. Technical report, National Bureau of Economic Research.

Black, S. E., P. J. Devereux, K. V. Løken, and K. G. Salvanes (2014). Care or cash? the effect of child care subsidies on student performance. *Review of Economics and Statistics 96*(5), 824–837.

Bowles, S. (2016). *The moral economy: Why good incentives are no substitute for good citizens*. Yale University Press.

Brehm, M. E. (2021). Taxes and adoptions from foster care evidence from the federal adoption tax credit. *Journal of Human Resources 56*(4), 1031–1072.

Brown, D. W., A. E. Kowalski, and I. Z. Lurie (2020). Long-term impacts of childhood medicaid expansions on outcomes in adulthood. *The Review of Economic Studies 87*(2), 792–821.

Buckles, K. S. (2013). Adoption subsidies and placement outcomes for children in foster care. *Journal of Human Resources 48*(3), 596–627.

Chorniy, A. and C. Mills (2022). More money, fewer problems? the effect of foster care payments on children's quality of care. In *2021 APPAM Fall Research Conference*. APPAM.

Cohodes, S. R., D. S. Grossman, S. A. Kleiner, and M. F. Lovenheim (2016). The effect of child health insurance access on schooling: Evidence from public insurance expansions. *Journal of Human Resources 51*(3), 727–759.

Congressional Research Service (2019). Youth transitioning from foster care: Background and federal programs.

Cuddeback, G. S. (2004). Kinship family foster care: A methodological and substantive synthesis of research. *Children and youth services review 26*(7), 623–639.

Dahl, G. B. and L. Lochner (2012). The impact of family income on child achievement: Evidence from the earned income tax credit. *American Economic Review 102*(5), 1927–56.

Dahl, G. B. and L. Lochner (2017). The impact of family income on child achievement: Evidence from the earned income tax credit: Reply. *American Economic Review 107*(2), 629–631.

Doyle, J. J. (2007). Can't buy me love? subsidizing the care of related children. *Journal of Public Economics 91*(1-2), 281–304.

Doyle, J. J. and H. E. Peters (2007). The market for foster care: an empirical study of the impact of foster care subsidies. *Review of Economics of the Household 5*(4), 329.

Doyle Jr., J. J. (2007). Child protection and child outcomes: Measuring the effects of foster care. *American Economic Review 97*(5), 1583–1610.

Doyle Jr., J. J. (2008). Child protection and adult crime: Using investigator assignment to estimate causal effects of foster care. *Journal of political Economy 116*(4), 746–770.

Duncan, G. J. and K. Magnuson (2013). Investing in preschool programs. *Journal of Economic Perspectives 27*(2), 109–32.

Duncan, G. J., P. A. Morris, and C. Rodrigues (2011). Does money really matter? estimating impacts of family income on young children's achievement with data from random-assignment experiments. *Developmental psychology 47*(5), 1263.

Duncan, G. J., K. M. Ziol-Guest, and A. Kalil (2010). Early-childhood poverty and adult attainment, behavior, and health. *Child Development 81*(1), 306–325.

Elango, S., J. L. García, J. J. Heckman, and A. Hojman (2015). Early childhood education. In R. A. Moffitt (Ed.), *Economics of Means-Tested Transfer Programs in the United States, Volume 2*, pp. 235–297. University of Chicago Press.

Goodman-Bacon, A. (2018). Difference-in-differences with variation in treatment timing. Technical report, National Bureau of Economic Research.

Goodman-Bacon, A. (2021). The long-run effects of childhood insurance coverage: Medicaid implementation, adult health, and labor market outcomes. *American Economic Review 111*(8), 2550–93.

Gross, M. and E. J. Baron (2022). Temporary stays and persistent gains: The causal effects of foster care. *American Economic Journal: Applied Economics 14*(2), 170–199.

Hansen, M. E. (2007). Using subsidies to promote the adoption of children from foster care. *Journal of Family and Economic Issues 28*(3), 377–393.

Hanushek, E. A. (2011). The economic value of higher teacher quality. *Economics of Education Review 30*(3), 466–479.

Hendren, N. and B. Sprung-Keyser (2020). A unified welfare analysis of government policies. *The Quarterly Journal of Economics 135*(3), 1209–1318.

Hoynes, H., D. W. Schanzenbach, and D. Almond (2016). Long-run impacts of childhood access to the safety net. *American Economic Review 106*(4), 903–34.

Institute of Applied Research (2011). Minnesota permanency demonstration: Final evaluation. Technical report, Minnesota Department of Human Services.

Institute of Medicine and National Research Council (2000). *From neurons to neighborhoods: The science of early childhood development*. National Academies Press.

Kleven, H. J. (2016). Bunching. *Annual Review of Economics 8*, 435–464.

Lindo, J. M. and G. R. Waddell (2011). Saving babies? revisiting the effect of very low birth weight classification. *The Quarterly Journal of Economics 126*, 2117–2123.

Løken, K. V., M. Mogstad, and M. Wiswall (2012). What linear estimators miss: The effects of family income on child outcomes. *American Economic Journal: Applied Economics 4*(2), 1–35.

Miller, S. and L. R. Wherry (2019). The long-term effects of early life medicaid coverage. *Journal of Human Resources 54*(3), 785–824.

Milligan, K. and M. Stabile (2011). Do child tax benefits affect the well-being of children? evidence from canadian child benefit expansions. *American Economic Journal: Economic Policy 3*(3), 175–205.

Perales, N. (2024). Do higher wages improve service quality? evidence from foster care in california using a regression discontinuity design. Unpublished working paper.

Roberts, K. V. (2019). Foster care and child welfare.

Rodgers, L. P. and C. T. Wallace (2020). Who responds to changes to the federal adoption tax credit? evidence from florida. *Southern Economic Journal 87*(2), 483–516.

Schonlau, M. and R. Y. Zou (2020). The random forest algorithm for statistical learning. *The Stata Journal 20*(1), 3–29.

Testa, M. F. and N. Rolock (1999). Professional foster care: A future worth pursuing? *Child Welfare 8*, l.

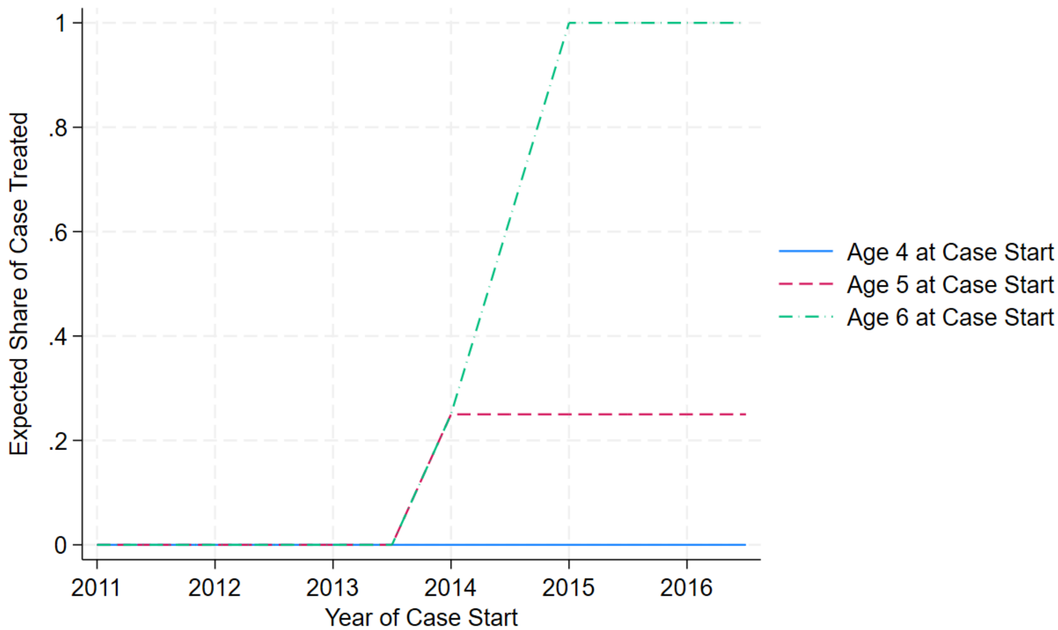
Testa, M. F. and K. S. Slack (2002). The gift of kinship foster care. *Children and Youth Services Review 24*(1-2), 79–108.

Warburton, W. P., R. N. Warburton, A. Sweetman, and C. Hertzman (2014). The impact of placing adolescent males into foster care on education, income assistance, and convictions. *Canadian Journal of Economics/Revue canadienne d'économique 47*(1), 35–69.

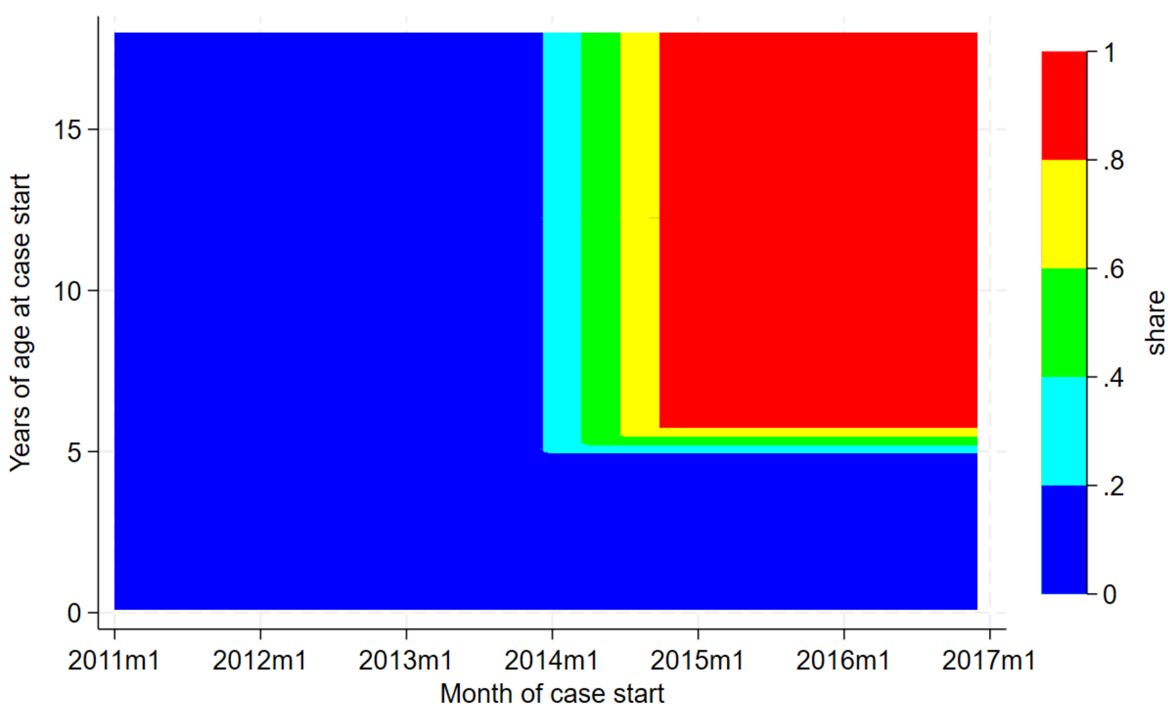
A Appendix Figures

Figure A-1: Variation in Assigned Treatment by Age and Time of Case Start

(a) Treatment Intensity Over Time



(b) Treatment Intensity by Age and Year of Case Start in Months



Notes: This figure illustrates the variation used to identify the % months the case is expected to spend treated ($shartreated_i$) in Equation (2), assuming a 16-month expected foster case length.

Figure A-2: Pre- to Post-Northstar Case Assessment Scale Crosswalk

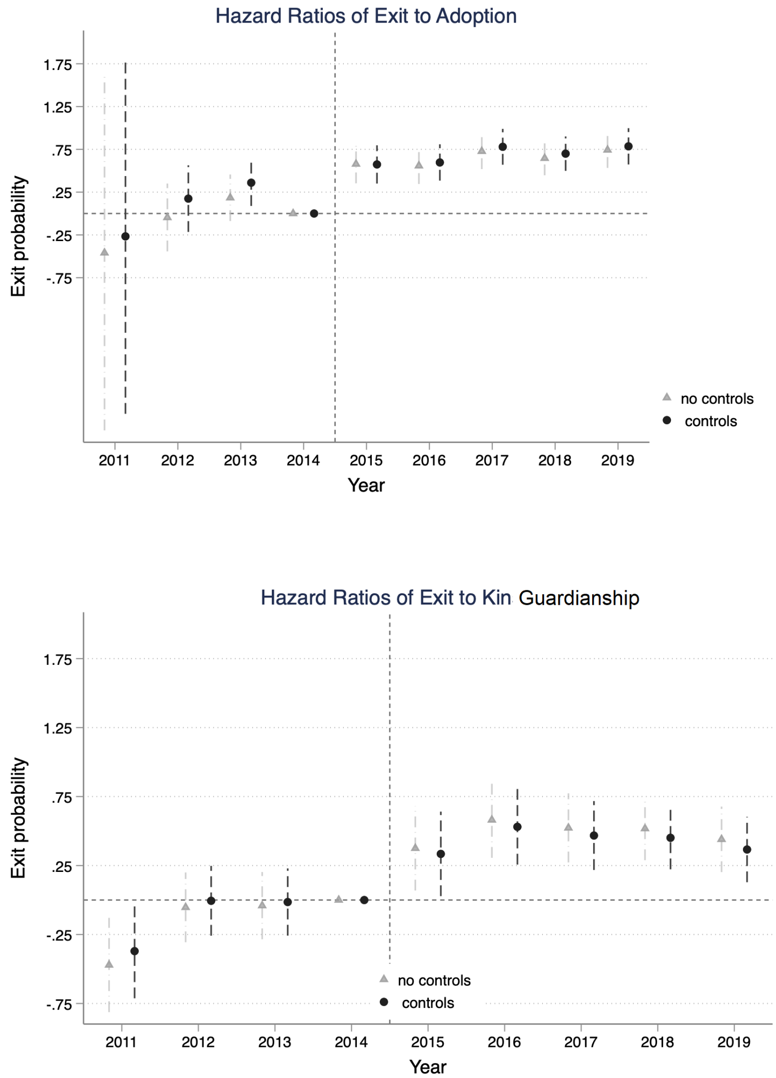
Translation Table

MAPCY Levels of Points to Existing Programs during the Phase 3 MAPCY Field Test in 2014 - January 15, 2014

MAPCY Benefit Level	Foster Care DOC Points	RCA Supplemental Level	MAPCY Total Points	AA Supplemental Level
B	0	Basic Only	0-17	Basic Only
C	18	Basic Only	18-25	Basic Only
D	36	I	26-32	I
E	54	I	33-39	I
F	72	II	40-46	II
G	90	II	47-53	II
H	108	III	54-61	III
I	126	III	62-68	III
J	144	III	69-74	III
K	160	IV	75-81	IV
L	175	IV	82+	IV
M	190	IV	DHS will specify	IV
N	202	IV	DHS will specify	IV
O	210	IV	DHS will specify	IV
P	218	IV	DHS will specify	IV
Q	225	IV	DHS will specify	IV

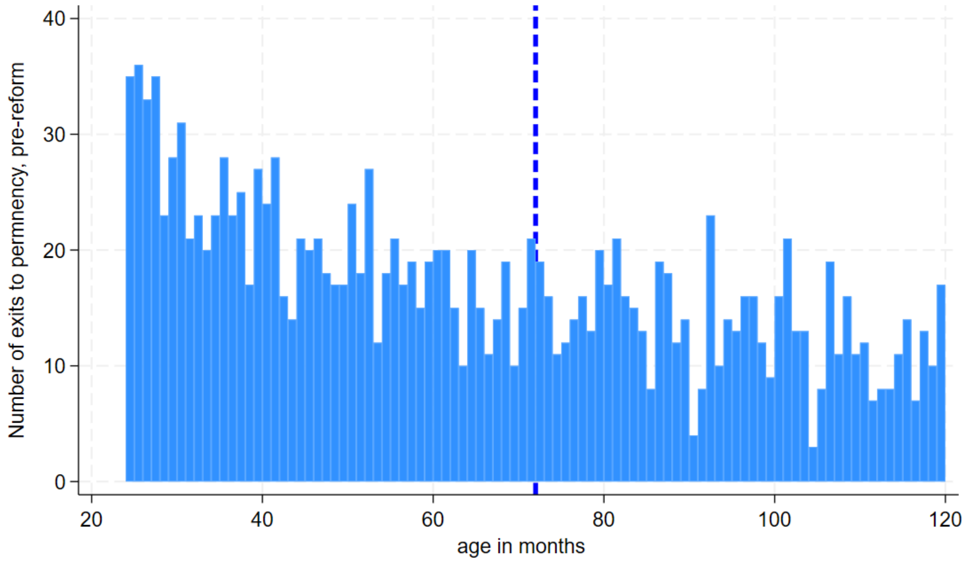
INSTRUCTIONS: 1) Find the MAPCY Level or Total Points on the MAPCY grid or Rating Sheet in SSIS.
2) For Foster Care, the MAPCY Level is translated to determine the DOC points. The translated DOC points are then entered in SSIS, which calculates the payments as usual.
3) For RCA, the MAPCY Level is translated into a Supplemental Level (Basic only, I, II, III, or IV). This level is then used in the RCA Payment Worksheet as usual.
4) For Adoption Assistance, the MAPCY Total Points (not the level) is translated into a Supplemental Level (Basic only, I, II, III, or IV). This level is then submitted to DHS along with the MAPCY Rating Sheet. For legacy Adoption Assistance, it is necessary to go to MAPCY Total Points to translate to legacy Adoption Assistance, because child care is a separate benefit available under that program.
5) For an approved Extraordinary Levels Request, DHS will calculate the additional levels and include the translation in the approval e-mail.

Figure A-3: Event Study of Policy Effects on Hazard Ratios of Exit



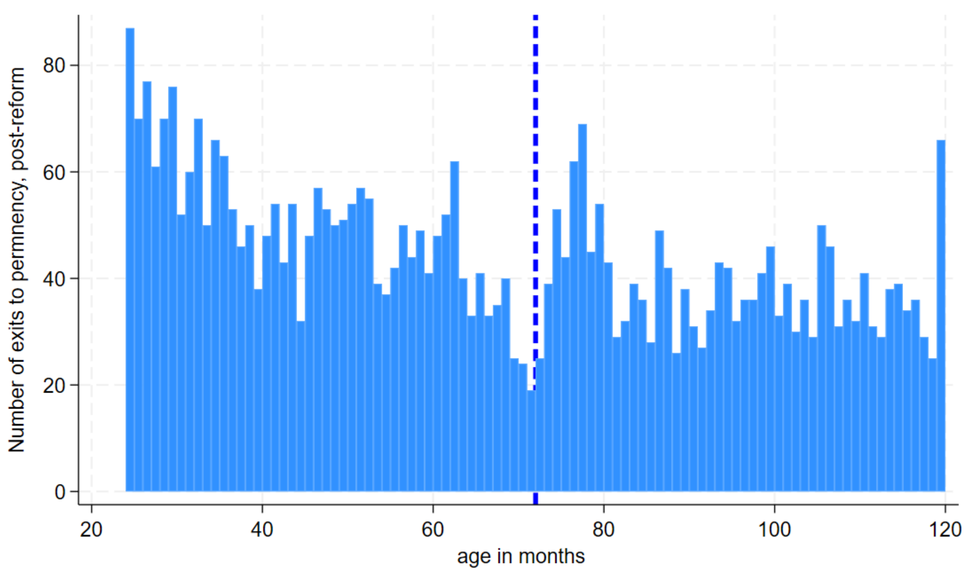
Source: Minnesota Department of Human Services foster case data.
Notes: This figure shows hazard ratios (with one subtracted from them) from estimating the event study version of Equation 4. Each coefficient estimate represents the proportional increase in exits of older (6+) children relative to younger children, in a given year, normalized so that 2014=0.

Figure A-4: Exits by Months in Age, before Northstar Reform



Source: Minnesota Department of Human Services foster case data.
Notes: This figure shows counts of exits to permanency (the y-axis) by child age in months (x-axis) for years 2011-2014. The dotted blue line reflects the age 6 cutoff of 72 months: when the child would qualify for a higher permanency stipend under Northstar.

Figure A-5: Exits by Months in Age, after Northstar Reform



Notes: This figure shows counts of exits to permanency (the y-axis) by child age in months (x-axis) for years 2015-2019. The dotted blue line reflects the age 6 cutoff of 72 months: when the child would qualify for a higher permanency stipend under Northstar.

B Appendix Tables

Table A-1: Summary Statistics of Foster Care Cases

	Mean	*Standard Deviation*
# of case-month observations	699,413	NA
Share post Northstar	0.72	0.45
Share 6 or older	0.59	0.49
Exit to reunification this month	0.0450	0.207
Exit to any permanency this month	0.015	0.122
Exit to adoption this month	0.009	0.093
Exit to kin guardianship this month	0.006	0.078

Source: Minnesota Department of Human Services foster case data.
Notes: This table shows summary statistics where each observation is a foster case-month for years 2011–2019.

Table A-2: The Impacts of Subsidizing Permanency on Ancillary Outcomes, Broken up Yearly

Years after CPE start	1-2	2-3	3-4	4-5
	Panel A: School Suspensions			
Policy Exposure	-0.002	-0.007	-0.003	0.006
	(0.007)	(0.008)	(0.009)	(0.013)
Pre-Policy Mean	0.13	0.12	0.10	0.10
Obs	31,376	33,824	20,407	20,407
	Panel B: Mental Health Services			
Policy Exposure	-0.011*	-0.004	-0.004	0.003
	(0.006)	(0.006)	(0.007)	(0.009)
Outcome Mean	0.08	0.07	0.05	0.04
Obs	31,376	24,932	18,966	13,502
	Panel C: Attendance			
Policy Exposure	-0.014**	-0.016**	-0.005	-0.007
	(0.006)	(0.005)	(0.005)	(0.007)
Outcome Mean	(0.88)	(0.89)	(0.89)	(0.89)
Obs	24,851	19,940	15,595	11,508
	Panel D: Average # of Schools per Year			
Policy Exposure	-0.063***	-0.042*	-0.021	-0.095
	(0.022)	(0.023)	(0.026)	(0.036)
Outcome Mean	1.75	1.61	1.53	1.48
Obs	31376	24,932	18,966	13,502
Controls	Yes	Yes	Yes	Yes

Source: Minnesota Department of Human Services foster case data, Minnesota Department of Education K12 school data, and Minnesota Department of Health Medicaid data. *Notes:* This table shows results from estimating Equation 2 on outcomes other than test scores. Each row and column shows a coefficient from a different regression. The outcome shown in Panel A is any out of school suspension, Panel B shows average yearly attendance. Panel C shows average number of schools attended per year. Panel D shows any mental health service use defined over a year. In each case this looks over a yearly period. Column 1 averages over years 1–2 after case start. Column 2 averages over years 2–3; columns 3 averages over years 3–4; and column 4 averages over years 4–5. All specifications include controls for child race, ethnicity, gender and reasons for the child's removal to foster care.

Table A-3: Heterogeneous Effects by Child Demographic and Reason for Removal Subgroup

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Sample:	All	Female	Male	White	Black	Hispanic	Native
Policy Exposure	0.311 **	0.084	0.618 ***	0.273	-0.001	0.262	0.620 *
	(0.155)	(0.231)	(0.228)	(0.286)	(0.324)	(0.471)	(0.324)
Pre-Policy Mean	-0.78	-0.69	-0.87	-0.54	-1.20	-0.87	-0.84
% impact	48.7 %	12.7 %	71.03 %	50.56 %	0.08 %	30.11 %	73.81 %
# Cases	6908	3399	3509	2806	1221	703	1071
Sample:	All	Neglect/Behavior	Abuse	Drug Use	Other		
Policy Exposure	0.311 **	0.34	0.327	0.403	0.044		
	(0.155)	(0.26)	(0.506)	(0.287)	(0.308)		
Pre-policy Mean	-0.78	-0.86	-0.95	-0.57	-0.72		
% impact	39.74%	39.53 %	34.42 %	75.44%	6.11%		
	(0.008)	(0.011)	(0.024)	(0.014)	(0.015)		
Pre policy Mean	0.19	0.22	0.23	0.07	0.16		
% impact	13.68%	12.27%	0.87%	23.33%	14.38%		
# Cases	6908	2266	940	1563	2139		

Source: Minnesota Department of Human Services foster case data and Minnesota Department of Education K12 school data.
Notes: This table shows results from estimating Equation 2 on the average of reading and math z-scores, where each observation is a separate foster care case. Each column limits the sample to a different subgroup listed at the top of the column. Each coefficient in the above table is from a different regression. Each foster case is a separate observation. When looking at a demographic subsample (female, White, etc.) we control for reason for removal dummies. When looking at a reason for removal subsample we control for demographic indicators.

Table A-4: Effects of Northstar Payment Equalization on Foster Care Exits for Different Exit Types

Exit Type:	Kin Guardianship	Kin Guardianship adding licensure controls	Adoption	Reunification
(Age 6+) x (Post 2014)	0.49 ***	0.61 ***	0.76 ***	-0.038
	(0.10)	(0.12)	(0.13)	(0.024)
Pre-reform mean	0.009	0.009	0.006	0.069
% impact	49%	61%	76%	3.8%
Model	Cox	Cox	Cox	Cox
# of Foster care spells	52,334	52,334	52,334	52,334
Observations	667,992	667,992	667,992	667,992

Source: Minnesota Department of Human Services foster case data.
Notes: An observation is a year-month that a child is observed in a foster care case. Each column and row report results from a separate DiD regression of the interaction between being age 6+ and in the post Northstar period (2015+) with age and year-month fixed effects regressed on whether there was an exit to different types of permanency arrangements in that month (see equation 4 and our discussion in subsection 6.1). All models include indicator controls for race, gender, and reason for removal. Column 1 regresses on the outcome of exit to Kin-Guardianship, column 2 adds controls for the licensure requirements mandated for kin guardian families to this regression, column 3 changes the outcome to exit to adoption (without kin gaurdian licensure controls), while column 4 changes the outcome to exit to reunification. For the hazard models we report relative hazard ratios after subtracting one from them. We subtract 1 from the ratios so that they reflect the effect of treatment on the % change in the outcome relative to the comparison group.

Table A-5: Timeliness to Permanency: Placebo Treatment on Children Younger than 6

	(1)	(2)	(3)
	Exit to Permanency		
(Age 2-5) x (Post 2014)	0.02		
	(0.07)		
(Age 3-5) x (Post 2014)		0.01	
		(0.06)	
(Age 4-5) x (Post 2014)			0.03
			(0.07)
Model	Cox	Cox	Cox
Controls	yes	Yes	yes
# of Foster care spells	21,828	21,828	21,828
Observations	275,878	275,878	275,878

Source: Minnesota Department of Human Services foster case data.
Notes: The full sample is limited to children younger than 6 in order to demonstrate placebo effects. Each column and row are from a separate DiD regression of the interaction between being in one of the placebo age ranges and in the post Northstar period (2015+) with age and year-month fixed effects regressed on whether there was an exit to permanency in that month (see equation 4 in the text). The first column defines the placebo "treated" ages as ages 2-5, the second as 3-5, and the third as 4-5. For the hazard models we report relative hazard ratios after subtracting one from them. We subtract 1 from the ratios so that they reflect the effect of treatment on the % change in the outcome relative to the comparison group. Models with controls control for child race, ethnicity, gender and reasons for the child's removal to foster care.

Table A-6: Timeliness to Permanency: Robustness to Donut Hole Regressions Removing Ages 5-6

	(1)	(2)
	Exit to Permanency	
(Age 6+) x (Post 2014)	0.28 ***	0.20 ***
	(0.07)	(0.06)
Model	Cox	Cox
Controls	No	Yes
# of Foster care spells	49,667	49,667
Observations	596,183	596,183

Source: Minnesota Department of Human Services foster case data.
Notes: This table shows our estimates on time to permanency removing ages over which there could potentially be bias due to strategic delays (ages 5-6). An observation is a year-month that a child is observed in a foster care spell. Each column and row are from a separate DiD regression of the interaction between being age 6+ and in the post Northstar period (2015+) with age and year-month fixed effects regressed on whether there was an exit to permanency in that month (see equation 4 in the text). For the hazard models we report relative hazard ratios after subtracting one from them. We subtract 1 from the ratios so that they reflect the effect of treatment on the % change in the outcome relative to the comparison group. Models with controls include controls for child race, ethnicity, gender and reasons for the child's removal to foster care.

C Appendix: Predicting Reunification

At times we run models estimated on the subsample of CPEs that are predicted not to reunify. Cases where the child is reunified with their parent represent 58 percent of foster care episodes. A reunification is typically the end result of cases where the underlying household issues are less severe and can be overcome with additional counseling or resources provided by a social worker. It is common in these less severe cases that, in a relatively short amount of time, Child Protective Services deems it is safe for the child to return to their origin home. Such foster care episodes can only last a few months and reflect cases where a child is unlikely to be adopted, because parental rights will never be dissolved.

Case characteristics that lead to reunification are predictive of cases less likely to be directly impacted by the policy, though it is possible that permanency placements displace reunification on the margin. If there are no or small effects on children likely to reunify, then including such cases could mute the estimated average treatment effects. Similarly, if these cases are fundamentally different from the sample of children who are the main target of the policy, then including them could introduce confounding trends in our DiD analysis (for example, this could happen if the composition of such children is changing over time or across the age distribution).

Speaking with adoption experts at Allegheny county, they made it clear that, *ex ante*, it is not obvious which cases will likely end in reunification versus which children will ultimately stay in foster care and be candidates for adoption or kin guardianship. Cases that begin with relatively simple barriers to reunification can grow in complexity over time. Further, when we began this project we did not rule out the possibility that incentives to adopt may lower the likelihood of reunification for cases on margin of being reunited, making the assignment of reunification endogenous to the policy.[41] We find that these issues remain particularly salient when looking at long-term outcomes, where the sample is smaller and downstream impacts of the policy are harder to detect.

To address these concerns, we supplement our analysis by using machine learning methods to predict which cases are most likely to reunify based on a rich set of case characteristics. We apply a random forest to all cases that ended before the start of the Northstar Policy.

The basic idea of the random forest is to construct "a tree" or a series of partitions based on different covariate values and then calculate the mean within each partition. Predictions can then be made by assigning the within-partition means to any observation where you observe the same covariates used in the partitioning. The algorithm chooses the covariate

[41] While we estimate non-significant effects of the policy on reunification, the sign is negative and close to being significant at the 10 percent level.

values that define the partition, as well as the number of partitions, in order to maximize the amount of information conveyed in the model. Such trees are subject to over fitting, which is guarded against by repeatedly bootstrapping the samples to construct many trees (a forest) in a processes called "bagging." The algorithm then makes its prediction for an observation by averaging the predicted values in each tree over the entire forest. In our classification problem, the covariates predict a dichotomous variable: whether the case ended in reunification or not.

To estimate a random forest:

1. Starting with all cases that ended before the start of Northstar, we randomly select 60 percent of this sample (10,000 cases) to use as a training data set. The remainder are set aside for validation.

2. We next need to calibrate hyper-parameters for the maximum number of nodes for each tree (the number sub-trees or times at which a tree gets partitioned) and the number of explanatory variables that are randomly chosen to determine at whose values there should be partitions in the tree.

3. Since individual trees are built on a bootstrapped sample, we calculate the error rates based on how the algorithm compares with the out of bootstrap sample (or out of bag error) as well as to the validation data set ("validation error"). Our hyper parameters are then chosen when we see these errors begin to converge to a minimum level. In our case at 300 maximum iterations with randomly selecting 9 variables to partition at each node.

4. After running the forest we then can predict cases that are likely to reunify. We exclude cases that are more than 80 percent likely to reunify from our "unlikely to reunify" data set. These cases are ones that the forest predicts will likely end in another way: ages out of foster care, gets adopted or put in kinship care, a tribal adoption, or one of a few other unlikely scenarios, such as child death. These are the cases we believe the policy is likely the largest positive effects on.

See Schonlau and Zou (2020) for more details.

D Appendix: Measuring Payment Streams

We construct the payment outcomes from 20,811 foster care cases in the state of Minnesota. The data on payments available from the state of Minnesota is more limited than our main sample, with information on payments limited to those who entered foster care between January 2013 and December 2016. We inflation adjust all payments to 2022 dollars.

For our main payment results, we compare the costs to the state of paying the permanency stipend with the social benefit of the stipend, as measured by child achievement gains observed $E = 3$ years after foster care start. To this end, we either observe or impute (in cases of permanency) a monthly payment amount from each episode's start time until the month three to four years later that the child would take the achievement test. The exact number of months that occur in this time frame is chosen for each observed case based on calendar month and child age at time of case start. Achievement tests are taken in April of third to eighth grade as well as tenth, and eleventh grade. The assigned number of months/payments before the test therefore ranges from 36 to 47, depending on when the foster care episode started.

For children who are exclusively in foster care during this time, their monthly payments are observed in our data as foster care payments.[42] For children who exit to permanency, we only have data on the first permanency payment as such payments rarely change in the medium run. We extrapolate out from the first payment by assigning the same permanency payment up until the month before the test. In cases where no first permanency payment is observed, we interpret this as no payments in permanency, which commonly occurred before the Northstar reform.

Child care subsidies were offered in permanency before Northstar to some children and discontinued after Northstar, as such subsidies were supposed to be more than made up for by the increased stipend amount. To measure the removal of the child care subsidy as part of the change in payments, we calculated the average child care subsidy that were paid out to adopted children.[43] We then assigned this average to all adopted children in our sample who exited to adoption before Northstar and set the subsidy equal to zero for all children who exited after Northstar. To test the robustness of the payment results to how we handled imputing this child care subsidy, we also calculated alternative versions of the payment stream that either ignored the child care subsidy altogether or imputed the maximum subsidy amount allowed under Minnesota policy. Regardless of how we assigned the value of the subsidy, it made little difference to the payment results shown in Table 6, suggesting that the change in child care policy was not a major factor in changing the amount of the stipend.

For each foster case, we measure the total payment stream as each foster payment plus

[42] In a few cases, there are missing values between monthly payments, which we extrapolate based on the most recent observed monthly payment.

[43] While some children who exited to kin guardianship also got a child care stipend before northstar this typically affected fewer children and good data on the extent of these payments from DHS was not available.

any permanency payment (including payments from child care), ranging from case start until the standardized testing year and month. The remaining children exited foster care via other paths (e.g., reunification with family of origin). They receive monthly foster care payment until their foster case ended and zero payments after that. Adding up this stream of payments yields total payments from start of foster care until time of the test. We additionally calculate the net present value of payments using the discount implied by the state of Minnesota borrowing cost in August 2023: 4.6 percent in August 2023.

We also construct as case-specific outcomes average monthly payment of all types, average monthly payments in foster care, average monthly payments in adoption if applicable, and average monthly payment in kin guardianship care if applicable. Every child in foster care has a foster payment, so average total payments and average foster care payments can be calculated for the entire sample in the payment data. Average adoption and kin guardianship payments are computed only for those children who are observed to exit into those paths, respectively.

E Appendix: Estimating Bunching over the Age 6 Threshold

In this appendix we discuss how we estimate the extent that Northstar might incentivize "strategic delay" of the permanency placement of children whose age places them near the age 6 cutoff so that their parents can take advantage of the higher permanency payments for older children. Figure A-4 and Figure A-5 plot the number of exits to permanency at every child age (in month) for years before and after the policy change, respectively. The vertical blue line marks when the child turns 6 (72 months). Both graphs show a long-term trend in declining number exits as children get older. Visually, in the years before Northstar we see no clear displaced mass of exits before the age 6 cutoff. Post Northstar it appears there may be a small notch of "missing" exits starting around an age of 69 months, with the displaced mass bunching beginning at age 73. The fact that there are still a substantial number of exits for children in the month or two before their sixth birthday implies either that there is a lack of transparency about how payments work or administrative burdens keep prospective adopters from precisely timing the move to permanency, or both.

To quantify the number of displaced exits, and the average length they were delayed, we employ techniques from the bunching literature often used around notches in the tax schedule (see Kleven (2016) for a review). We begin with the simplest approach, which is to look within a small window leading up to the notch (age in months 55–68), and calculate the average number of exits per month within this window: $\widehat{y}$. This average is then projected over the notch ages of 69–72 (y_a) months. The estimated "missing" exits per month in the notch is estimated as $\sum_{a=69}^{72} \widehat{y} - y_a$, the difference between the mean exits leading up to the notch and the observed number of exits in the range of the notch.

To quantify the average time cases in this notch area are delayed, we start with the first month of the notch ($a = 69$) and assume those exits happen during the first month of displaced "bunching" ($a = 73$); once all the displaced mass at age 69 is accounted for, we continue to age 70 and so on (assuming the mass at each age is systematically displaced up through at most age 80). We then present the average time these cases are delayed which can be used to bound the extent of the bias in our time to permanency hazard regressions. Note that the extent of estimated bunching is determined the same way the notch is: $\sum_{a=73}^{80} \widehat{y} - y_a$. Across specifications the estimated bunching mass is roughly equivalent to estimated displaced mass in the notch, further reassuring us that our visual inspection of the start of the notch/bunching was correct.[44]

[44]There were at most five extra estimated cases in the notch that we assigned an exit of $a = 81$

The above approach assumes the counterfactual number of exits in the absence of any strategic delays can be estimated as the average number of exits between ages 55–68 months. We estimate alternative counterfactuals under different assumptions. To do this, first we collapse the sample so that each observation is an age-month cell with our outcome of interest being the counts of exits to permanency at each age. We limit the sample to range from the youngest being ages 20–55 in months and the oldest being 81 months (when bunching is last clearly observed). We next estimate the following regression model:

$$Y_a = f_n(a) + \sum_{a=69}^{81} \gamma_a \tag{E-1}$$

where a indexes age in months, Y_a is the number of exits at age a, and $f_n(a)$ is an an n degree polynomial chosen to capture the trend between age in months and and number of exits. γ_a is a vector of dummies for each age in month over the observed density of the notch and bunching. These dummies are meant to capture the potential area of displaced mass from sorting below to over the 72 months threshold. Successfully estimating the coefficients on the polynomial $f(a)$ requires looking over a horizon of ages leading up to the notch. Typically, we allow a more flexible polynomial by estimating the polynomial over a longer window. Specifically, we estimate the window from months 55–68 in the uniform case $n = 0$, 30–68 months in the linear case $n = 1$ and 20–68 months in the quadratic case $n = 2$.

After estimating the coefficients on $f(a)$ from Equation E-1, we predict the counterfactual number of exits in the absence of bunching ($\widehat{Y_a}$ for $min <= a <= 72$). We then estimate the extent of delayed exits as $\sum_{a=69}^{72} \widehat{Y_a} - Y_a$ and otherwise estimate the time each exit was delayed analogously to the simple uniform case.

Table E-1 below shows the results of this exercise. We report the number of estimated delayed exits and the average months these exits were delayed. A roughly similar number of delayed exits are found using uniform ($n = 0$) and linear ($n = 1$) terms, while using a quadratic polynomial estimates somewhat fewer exits and a shorter average time to exit. Overall, the three models give similar results with a range of 59–81 delayed exits with an average delay of 4.85 to 6.54 months.

Given that we see a fairly linear decline in exits over age in months in the pre-period, our preferred specification is the linear polynomial, which suggests that there were 81 exits, each delayed an average of 6.37 months. At the end of subsection 6.1 we calculated the policy on average decreased time in foster care by 5.2 months for the 3,860 older children who were adopted or taken into kinship care after 2014. If 81 of those exits were actually *delayed* by six months due to red-flagging, we can re-weight the average treatment effect as $ATE = \frac{-5.2*3779+6.4*81}{3860} = -4.95$ months. That is, our treatment effects of time to permanency are at most a fifth of a month overstated due to red-flagging.

Table E-1: Bunching Analysis of Strategic Delay Times

polynomial	# of Delayed Exits ($\sum_{a=69}^{72} \widehat{y} - y_a$)	Average Length of Delay
Uniform ($n = 0$)		
Estimates	-79	6.54 months
Estimation window	55-81	55-81
Linear ($n = 1$)		
Estimates	-81.2	6.37 months
Estimation window	30-81	30-81
Quadratic ($n = 2$)		
Estimates	-59	4.85 months
Estimation window	20-81	20-81

Source: Minnesota Department of Human Services foster case data.
Notes: This table presents results from our estimates of Equation E-1 for # of delayed exits and the average length of case delay assuming different ways of modeling $f_n(a)$ as presented in equation Equation E-1. Across the three panels we assume the polynomial ($f_n(a)$) is of different degrees (uniform, linear, or quadratic). We also limit the sample to different ages over which strategic delay is estimated to match the potential amount of displaced mass observed in Figure A-5. See Appendix E for details.