Preserving Families Through Infant-Toddler Court Teams

An Evaluation of New York State's Strong Starts Court Initiative

By Jordan Conan and Jeffrey Sharlein

Center
for
Justice
Innovation

Preserving Families Through Infant-Toddler Court Teams: An Evaluation of the Strong Starts Court Initiative

By Jordan Conan and Jeffrey Sharlein

© April 2024

Acknowledgements

Table of Contents

Executive Summary

Infants and toddlers (aged 0-3) are overrepresented in the child welfare system and are more likely than children of other ages to be removed from their original caregiver and placed in out-of-home care. Such early childhood attachment disruptions have been linked to enduring impacts on emotional regulation, psychopathology, and social cognition, and the often-extended stays in the child welfare system are associated with an increased likelihood of additional attachment disruptions and poor physical and mental health (Afifi et al. 2020; Casey Family Programs 2023; Coan 2016; Shonkoff and Phillips 2000; Sroufe 2005; Thompson 2016).

Specialized Infant-Toddler Court Team (ITCT) programs, often supported by the national ZERO TO THREE network, have been implemented to better support and address the needs of these young children and their families, typically following a child's removal. The Strong Starts Court Initiative (hereafter, Strong Starts) in New York State is one such program. Strong Starts pairs a judge with specialized training and a clinical coordinator with expertise in early childhood mental health (ECMH) to promote a more collaborative court process and provide ongoing ECMH training for the court practitioner community. Somewhat unique among ITCTs, both families of children under court-ordered supervision and those that have experienced a removal are eligible for programming. Strong Starts' theory of impact posits that the program will lead to a decrease in attachment disruptions, strengthened and/or repaired attachments between children and their original caregivers, and increased stability for those caregivers. Since its 2015 inception in Bronx County, Strong Starts has grown to incorporate sites in all New York City counties, and in Erie and Westchester counties.

This evaluation of Strong Starts addresses the following research questions:

- Does Strong Starts implementation lead to a decrease in rates of removal to out-of-home care for program-eligible children in the Strong Starts courtroom?
- Does Strong Starts implementation lead to an increase in program-eligible children living at home with their original caregiver 12 months after case filing in the Strong Starts courtroom?

Since the program aims to effect change at the courtroom level, we include families who are eligible for but not selected into the program in the treatment group.

Methods

Using administrative court data from the New York State Office of Court Administration for a single New York City county, this evaluation leverages near-random case assignment to implement a natural experiment triple difference interrupted time series design. To explore whether program implementation led to any change in outcomes for eligible cases in the Strong Starts courtroom, the analysis creates and compares cohorts along three dimensions:

- Judge: Strong Starts judge/another judge;
- Eligibility: Strong Starts eligible/Strong Starts ineligible; and
- Time: Pre-implementation/post-implementation.

Outcomes

Removal at Filing We estimate a highly significant 25 percentage point reduction in removals at filing for Strong Starts-eligible children in the Strong Starts courtroom following program implementation.

With Original Caregiver at 12 Months We estimate a 9 percentage point increase in the number of children residing with their original caregiver 12 months after filing for Strong Starts-eligible families in the Strong Starts courtroom following program implementation. While this difference is not statistically significant, we believe it is meaningful, based on the concurrent significant decrease in removals at filing, absence of any increase in later removals, and lack of a significant decrease in reunifications.

Conclusions

Qualifying children were significantly less likely to be removed from their initial caregiver at case filing following program implementation, and there are indications that they were more likely to reside at home with their initial caregiver 12 months after filing. The findings from this study combined with the results from a 2021 evaluation showing a decrease in subsequent child welfare court episodes for Strong Starts participants (Sharlein 2021) paint a picture of an intervention that leads to more stable outcomes for children, better family court outcomes for their respondent caregivers, and the preservation of the attachment relationship. The targeted support on Strong Starts cases and expanded practitioner knowledge applicable to all eligible cases form a plausible mechanism for the estimated decrease in removals, and these results underscore the importance of Strong Starts' inclusion of families with children under court-ordered supervision in their programming.

Chapter 1
Introduction

Infant-Toddler Court Teams

Infants and toddlers are overrepresented in the United States' expansive child welfare system (James Bell Associates 2009). Children aged three and under are more likely than children of other ages to enter the child welfare court system as the subject of a neglect or abuse petition in a court filing, and once an infant or toddler is under court jurisdiction, they are more likely to be removed from their original caregiver and placed in out-of-home care (see Figure 1). The removal of a subject child[1] initiates an attachment disruption. Such disruptions during early childhood have been linked to enduring impacts on emotional regulation, psychopathology, and social cognition (Coan 2016; Shonkoff and Phillips 2000; Sroufe 2005; Thompson 2016). The removal also introduces a scenario in which attachment relationships need to be nurtured between the child and both the original and placement caregivers. To address the exigencies of all parties, child protective services (CPS) must place the child with a responsible adult; monitor and support the placement; monitor and provide services for the respondent parent;[2] and provide logistics, time, and space for interaction between the respondent parent and the subject child.

A removal also triggers a concurrent and often prolonged permanency court process. The reunification of a subject child in out-of-home care with their original caregiver is always the initial permanency goal stipulated by CPS and is considered the preferred permanency outcome (Child Welfare Information Gateway 2017). However, CPS may initiate a court proceeding to terminate the parental rights of the original caregiver after an extended period of monitoring (when a child has been in out-of-home care for 15 of 22 months as stipulated by the Adoptions and Safe Families Act)[3]—an antecedent to the permanency resolution of adoption.[4] An infant or toddler placed in out-of-home care is less likely to be reunified than

---

[1] A subject child is the child who is the subject of a neglect or abuse petition.
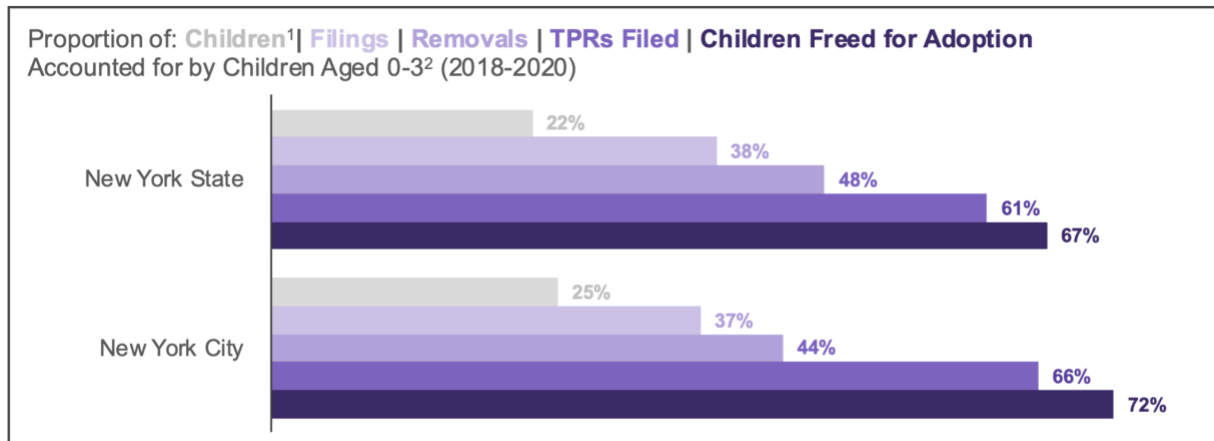[2] Respondent caregivers are the original caregivers against whom a neglect or abuse allegation is made.
[3] PL 105-89 Sec. 103
[4] Children can also achieve permanency through custody or guardianship with a fit and willing caregiver. These options are usually explored as an alternative to a termination of parental rights when it is determined that reunification is no longer a viable outcome.

their older peers and the relationship with their original caregiver is more likely to be severed through a Termination of Parental Rights (TPR) proceeding (see Figure 1). The resulting extended stays in the child welfare system are associated with an increased likelihood of additional attachment disruption for the subject child, and later adverse outcomes such as poor physical and mental health (Afifi et al. 2020; Casey Family Programs 2023).

Figure 1. Overrepresentation of Young Children Throughout the Family Court Process

Proportion of: Children[1] | Filings | Removals | TPRs Filed | Children Freed for Adoption
Accounted for by Children Aged 0-3[2] (2018-2020)

New York State
- 22%
- 38%
- 48%
- 61%
- 67%

New York City
- 25%
- 37%
- 44%
- 66%
- 72%

[1] Estimate based on US 2020 Census ACS 5-Year Estimates Subject Tables
[2] Age at initial CP Filing for all court metrics

Data Sources: US 2020 Census ACS 5-Year Estimates Subject Tables; NYS Unified Court System, Universal Case Management System – Family Court Data Metrics

The increasing overrepresentation of families with young children at every stage of the child welfare court process has not gone unnoticed. Over the last 20 years, targeted Infant-Toddler Court Team (ITCT) programs, often supported by the national ZERO TO THREE network, have popped up throughout the country to better support and address the needs of those families. With few exceptions, these programs have worked with children and families post-removal, introducing principles of Early Childhood Mental Health (ECMH) and attachment, and working towards reunification (Casanueva et al. 2017; Casanueva et al. 2019; James Bell Associates 2009; McCombs-Thornton and Foster 2012).

The Strong Starts Court Initiative

One exception to the post-removal intervention norm is the Center for Justice Innovation's Strong Starts Court Initiative (hereafter, Strong Starts) in New York State, where both families of children under court-ordered supervision and those that have experienced a removal are eligible for programming.

The Strong Starts program model pairs a judge who receives special training in ECMH with a clinical coordinator with specific expertise in ECMH and attachment to create a specialized

court team. Families can be referred to the program by the judge if the case is initially filed in the Strong Starts judge's courtroom (regardless of removal status), the oldest subject child is three or younger, the family has no prior child welfare court involvement, and the program has availability (coordinator caseloads are capped at 20). The respondent parents, in consultation with their attorneys,[5] must consent to become program participants. The clinical coordinator connects participants with targeted clinical and social services, with a particular focus on the child-parent relationship. They also convene regular court team clinical conferences that bring CPS, attorneys, and service providers together with the family to problem-solve around barriers, discuss case progress, and engage in collaborative planning and decision-making. After each clinical conference the coordinator provides a court report to the judge and attorneys ahead of the next court appearance. Strong Starts supports its case-specific components with an ongoing training series on ECMH and childhood well-being for the court practitioner community.

Strong Starts was piloted in 2015 with a single coordinator/judge pairing in one of New York City's (NYC) five family courts (one in each borough) and has since grown to incorporate 11 judges and 13 coordinators operating in all five NYC boroughs and New York's Erie and Westchester counties. Strong Starts' theory of impact[6] posits that the introduction of the clinical coordinator and the increased focus on collaboration and the principles of ECMH will lead to an increase in what we define as family preservation: the prevention of attachment disruptions, strengthened and/or repaired attachments between children and their original caregivers, and increased stability for those caregivers.

Previous ITCT Research

In a 2021 evaluation of the Strong Starts Court Initiative, judges and attorneys reported that interaction with the program increased both their understanding of ECMH and their awareness of available resources for families with young children. In the same evaluation, participating caregivers noted a similar increase in their understanding of ECMH, improved relationships with their subject children, and an understanding and belief in the fairness of the court process. These outcomes and the evaluation's indication that Strong Starts participation is associated with a reduced likelihood of re-entry into the child welfare system support Strong Starts' theory of impact (Sharlein 2021). However, the evaluation did not

---

[5] In New York City, respondent parents are assigned an attorney at the initial court proceeding which occurs within a day of case filing.
[6] Strong Starts theory of impact is similar to a theory of change and was developed in 2021 with the support of the University of Oregon's Early Childhood Precision, Innovation, and Shared Measurement team.

examine program impacts on case resolution type or family preservation. Impact analyses can be notoriously difficult to conduct in a child welfare setting and comes with myriad design challenges.

Methodological Challenges

Cohort and outcome construction are primary challenges in child welfare research design. Child welfare cases, especially those with a removal and subsequent termination of parental rights, can take years to resolve. Analyzing a dataset limited to children who have exited court jurisdiction inevitably excludes those challenging cases that remain open for years, provides a skewed picture of both time-to and type of resolution, and privileges shorter duration resolutions—such as reunification (McCombs-Thornton and Foster 2012). One potential solution is constructing datasets using entry cohorts—that is, by using groupings of children entering court jurisdiction during specified time periods. Researchers should select analyses that consider the open-ended and extended nature of child protective outcomes (e.g., survival analysis, point-in-time outcome proportions that include rather than censor pending outcomes).

Control group construction poses an even larger challenge. A randomized controlled trial (RCT)—in which eligible families are randomly selected into or out of an intervention, tracked, then compared—is considered the gold standard experimental design, but RCTs are rarely practical. Even if otherwise appropriate, they are time and resource-intensive and may require withholding the intervention from eligible families following a lengthy information and consent process (McCombs-Thornton and Foster 2012). To avoid these potential challenges, evaluators frequently employ quasi-experimental designs that approximate randomization to create an appropriate comparison group for analysis.

Researchers studying ITCTs often create comparison groups using data from historic nationally representative datasets, such as the National Survey of Child and Adolescent Well-Being (e.g., Casanueva et al. 2017; Casanueva et al. 2019; McCombs-Thornton and Foster 2012). Unfortunately, this approach will frequently lead to flawed results, since child welfare practice varies widely from state to state (JBS International 2020), county to county, and even courtroom to courtroom. Such local variation means that national samples are not representative of local contexts and, as such, cannot provide an appropriate comparison in studies of localized interventions.

Even researchers with access to a contemporary local comparison group[7] need to consider the impact ITCT implementation may have on non-participant families. ITCTs are typically designed such that information on ECMH and available resources is applied by practitioners, particularly the ITCT judge, in non-ITCT cases. Additionally, selection into an ITCT is rarely truly random. Since programs have limited capacity, there is typically professional discretion in candidate referral, and parents must consent. For example, a parent's lawyer is unlikely to support a potentially time- and resource-intensive intervention for their client if they believe the case can resolve quickly in their client's favor without it, and a judge is unlikely to refer a case to an ITCT if they hold a similar belief. Additionally, the voluntary nature of participation may indicate sampling bias; particularly motivated parents—or those with other unique characteristics—may be more likely to agree to participate and also more likely to have favorable outcomes. Thus, real-world ITCT implementation often means that even local, contemporary comparison cohorts have shortcomings as comparison groups. One potential way to address these shortcomings is by incorporating a pre-implementation baseline group, in addition to a contemporary comparison group.

---

[7] Faria et al. 2020 and Sharlein 2021 are examples of contemporary cohort comparison use in the literature.

Chapter 2
Research Design, Data, and Methods

This study seeks to assess Strong Starts' impact on family preservation.[8] Specifically, we investigate the impact Strong Starts program implementation had on family preservation for eligible families appearing before a single Strong Starts judge in one NYC family court. Since the program aims to effect change at the courtroom level as well as at the level of individual families, the inclusion in the treatment group of families who are eligible for but not selected into the program allows for a more complete assessment of program impacts.

We use two court process measures to assess family preservation:

- **Removal at Filing** What impact did Strong Starts implementation have on the removal of program-eligible children from their original caregiver at the start of cases in the Strong Starts judge's courtroom?[9]

- **With Original Caregiver at 12 Months** What impact did Strong Starts implementation have on program-eligible children appearing in the Strong Starts judge's courtroom living at home with their original caregiver 12 months after case filing?[10]

Rather than use a more traditional permanency measure such as reunification at 12 months, we chose to look at the proportion of children with their original caregiver 12 months after filing regardless of removal status. This construction both allows for a larger sample by combining supervision cases with removal cases and provides a more holistic perspective on family and attachment preservation in courtroom practice.

---

[8] Defined for the purposes of this paper as the prevention of attachment disruptions, strengthened and/or repaired attachments between children and their original caregivers, and increased stability for those caregivers.

[9] Removals occurring within two weeks of case filing

[10] Regardless of their original removal status

Taken together, these measures can help us understand how likely a child is to remain and/or end up with their original caregiver over the course of a child welfare court episode, preserving the original attachment relationship.

Case Assignment and Sampling Frame

The study uses administrative data derived from the New York State Unified Court System's Universal Case Management System (UCMS) for a single NYC borough (county). This NYC family court has 11 judges who preside over approximately 3,000 new neglect and abuse cases each year, around 300 of which meet the eligibility criteria for Strong Starts. Cases filed by NYC's Family Court Legal Services (FCLS)—the legal branch of the city's child protection agency—are assigned to judges based on a daily rotating intake calendar. This is the mechanism by which both Strong Starts-eligible and Strong Starts-ineligible cases come before the Strong Starts judge. The sample data for this study is drawn from all neglect and abuse filings in the family court over a six-year period, three years before and three years following Strong Starts program implementation. The data also include earlier neglect or abuse petitions involving any subject child or respondent parent; age and gender; family relationships; and any activity that occurred up to 18 months after case filing, including allegations, removals, permanency, and related termination of parental rights proceedings.

The Triple Difference Design

The Triple Difference (DDD) is a controlled interrupted time series (ITS) quasi-experimental research design (Reeves et al. 2017). Basic ITS designs compare outcome trends before and after the implementation of an intervention. For the purposes of this study, a basic ITS design would compare the outcomes for Strong Starts eligible families appearing before the Strong Starts judge prior to program implementation to the outcomes for Strong Starts eligible families appearing before the same judge following program implementation.

To ensure that a change in the judge's approach towards all families unrelated to Strong Starts implementation is not responsible for any observed differences, we introduce a control and look at the Strong Starts judge's other cases to see if similar changes in outcomes occurred. This design would be considered a difference-in-differences (DiD). To further ensure that a court-wide shift in the treatment of cases with young children is not responsible for any observed differences, in a DDD we introduce a second control and look to see if similar changes in outcomes occurred in other courtrooms.

This analytic approach relies on the parallel trend assumption (in statistical terms, this assumption is called the "key identifying assumption"). For these analyses to have causal interpretations (see Table 1), an outcome variable for the treatment group needs to trend in the same direction over time prior to treatment (or program) implementation as that outcome variable for the control group. In other words, they'd need to appear parallel in a time series. The assumption is typically assessed in the literature by visually inspecting trend plots, though many authors do not assess parallel trends at all (Olden and Møen 2022). For this study, the relationship of the difference between an outcome variable for the Strong Starts judge's pre-implementation eligible and ineligible cases and the other judges' pre-implementation eligible and ineligible cases would need to be consistent prior to the introduction of Strong Starts. We provide the plots needed to assess the assumption for each outcome variable in the results section of this paper and note where we believe it is met.

Table 1. Conditions for Causal Interpretation of DDD

DDD Coefficient is Significant	Parallel Trend Assumption is Met	Causal Interpretation Justified
Yes	Yes	Yes
Yes	No	No
No	Yes	No
No	No	No

Appendix A includes technical details of the statistical approach and more discussion of study design.

Analysis Cohorts

Since cases in this NYC family court are assigned based on a rotating daily intake calendar, which is considered near-random assignment statistically, we can treat this analysis as a natural experiment. The three dimensions used for the analysis are:

- Judge: Strong Starts judge/another judge;
- Eligibility: Strong Starts eligible/Strong Starts ineligible; and
- Time: Pre-implementation/post-implementation.

Dividing the data along these three dimensions produces the eight cohorts used for the analysis (see Table 2 for cohort descriptions and Figure B in Appendix B for a case assignment decision tree).

Table 2. Dimensions and Cohorts used for DDD Analyses

Dimension A: Judge	Dimension B: Eligibility	Dimension C: Time	
		Pre-Implementation	**Post-Implementation**
Strong Starts Judge	**Strong Starts Eligible**	Strong Starts eligible children appearing in the Strong Starts judge's courtroom pre-implementation	Strong Starts eligible children appearing in the Strong Starts judge's courtroom post-implementation
	Strong Starts Ineligible	Strong Starts ineligible children appearing in the Strong Starts judge's courtroom pre-implementation	Strong Starts ineligible children appearing in the Strong Starts judge's courtroom post-implementation
Other Judge	**Strong Starts Eligible**	Strong Starts eligible children appearing in other judge's courtroom pre-implementation	Strong Starts eligible children appearing in other judge's courtroom post-implementation
	Strong Starts Ineligible	Strong Starts ineligible children appearing in other judge's courtroom pre-implementation	Strong Starts ineligible children appearing in other judge's courtroom post-implementation

Cohort Characteristics[11]

Table 3 displays the demographic characteristics of each DDD cohort. The table also notes where there are statistically significant differences between cases heard by the Strong Starts judge and other judges; this is to assess our treatment of judge assignment as near-random. While there are some significant or near-significant differences, almost all are of minimal magnitude with no apparent explanation that would undermine the assumption of neutral case assignment. The exceptions to this are the *number of family court episodes for both child* and *respondents* among program-ineligible cases following implementation. These differences (still small in magnitude: 0.3 average previous family court episodes per child for the Strong Starts judge as compared with 0.2 for other judges, and 0.4 previous family court episodes per respondent for the Strong Starts judge as compared with 0.3 for other judges) likely result from New York City Family Court's adherence to the One Family, One Judge principle. This principle ensures that a family appears before a single judge for their initial and any subsequent child protective episodes when possible and appropriate (New York City Family Court 2023). Thus, the longer a judge sits on the bench, the more likely they will hear a subsequent episode for a family.

---

[11] Appendix A presents a discussion of adjustments made to cohort composition for the purposes of analysis.

Of the 90 eligible cases appearing in the Strong Starts courtroom following implementation, 44 (49%) were enrolled in the program. Strong Starts participants accounted for 61% (16/26) of the cases with a removal at filing (prior to Strong Starts participation) and 44% (28/64) of the court-ordered supervision cases.[12]

---

[12] Although these figures do not reflect the primary analysis, they support the notion that courtroom practitioners are more hesitant to approve of a resource-intensive intervention they may see as an unnecessary burden on court-ordered supervision cases that are likely to be brief. None of the Strong Starts judge's 16 eligible cases that closed in under six months were enrolled in the program.

Table 3. Demographic Characteristics for Cases in each DDD Cohort[1]

| | Pre-Implementation Cases | | | | Post-Implementation Cases | | | |
| | (A) Strong Starts Eligible | | (B) Other Cases | | (C) Strong Starts Eligible | | (D) Other Cases | |
N	Strong Starts Judge 106	Other Judge 682	Strong Starts Judge 992	Other Judge 5,699	Strong Starts Judge 90	Other Judge 821	Strong Starts Judge 911	Other Judge 8,893
Demographics[2]								
Mean child age at filing (years)[m]	0.9	1.0	7.6	7.9	1.0	1.0	8.0	8.0
Mean respondent age (years)[3, m]	27.2	27.0	34.9*	35.5	28.9	28.0	35.7	35.7
Mean # of respondents	1.4	1.3	1.3	1.3	1.3	1.3	1.3+	1.3
Mean # of siblings	0.3	0.3	2.2***	1.9	0.2	0.2	1.9	2.0
Mean # of previous Family Court episodes[4]								
Subject child	0.0	0.0	0.2*	0.2	0.0	0.0	0.3***	0.2
Respondent[5]	0.0	0.0	0.4	0.3	0.0	0.0	0.4***	0.3
Child gender[6, m]								
Male	44%	50%	48%	50%	63%**	49%	46%+	49%
Female	56%	50%	52%	50%	37%**	51%	54%+	51%

+p<.10   *p<.05   **p<.01   ***p<.001

[1] Statistical tests compare Strong Starts judge and other judge within A, B, C, and D.

[2] ANOVA test used to determine significance.

[3] Mean age if multiple respondents.

[4] Cases are only considered Strong Starts eligible if there is no previous family court episode for the respondent (as a respondent) or subject children.

[5] Maximum number of episodes if multiple respondents.

[6] Gender is coded as a binary (male/female) variable. Chi-square test used to determine significance.

[m] Missing values: mean child age at filing: 1, 0.0%; mean respondent age: 280, 1.5%; child gender: 25, 0.1%.

Results, Discussion, and Conclusion

Results

Table 4 and Figure 2A show the estimated impact of Strong Starts implementation on family preservation for Strong Starts-eligible children appearing in the Strong Starts judge's courtroom; we used the charts in Figure 2B to assess whether each measure meets the parallel trend assumption (see Appendix A for additional information on the model and results).

Table 4. Effects of Strong Starts Implementation on Family Preservation Measures

	DDD Coefficients
Removals at filing	-0.252***
p-value	(<.001)
95% Confidence Interval	(-0.389, -0.114)
ϵ	0.070
N	17,898
With original caregiver at 12 months[1]	0.091
p-value	(.220)
95% Confidence Interval	(-0.055, 0.237)
ϵ	0.074
N	15,263

+p<.10,* p<.05, ** p<.01, ***p<.001
[1] Excludes cases filed in the year preceeding Strong Starts implementation.

Removals at Filing

As shown in Table 4, we estimate a 25 percentage point reduction in removals at filing following Strong Starts implementation. This reduction is highly statistically significant, indicating that the estimated difference is almost certainly a real association and not a result of chance. Additionally, this analysis meets the parallel trend assumption (Figure 2B). The high level of statistical significance and meeting of the parallel trend assumption suggest that the effect is causal; that is, Strong Starts implementation caused the estimated reduction in removals at filing.

Figure 2. Interpretation of Results: DDD Estimates and Parallel Trend Assessment

For the measured effect estimated by a DDD to be considered causal, the DDD coefficient must be significant (A) *and* the measure must pass the parallel trend assumption (B). The DDD for **Removals at Filing** meets both criteria, while the DDD for **With Original Caregiver at 12 Months** meets neither.

A. Effects of Strong Starts Implementation on Family Preservation Measures (DDD Coefficient by measure)

A shaded (purple) dot indicates increased family preservation, the shaded line represents the 95% confidence interval for the estimate.
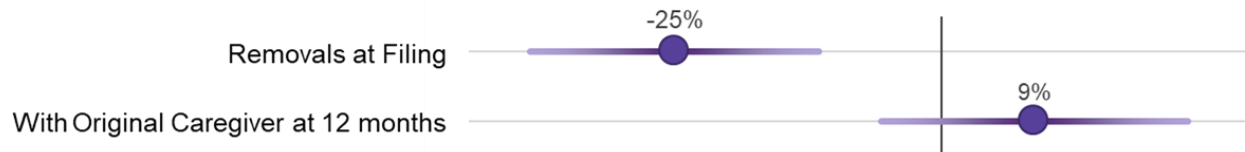
B. DDD Trends for Parallel Trend Assumption Assessment

The Y-axis is the difference in the relationship between Strong Starts eligible outcomes and Strong Starts ineligible outcomes for the Strong Starts judge and other judges in the county. The Pre-implementation trend needs to be parallel to the X-axis (time) for results to be considered causal.
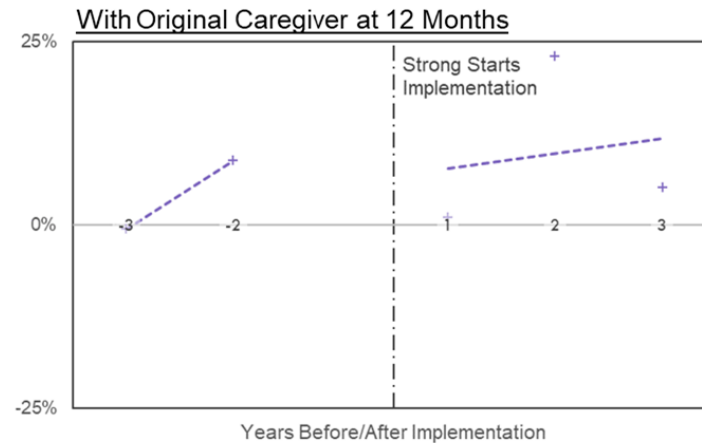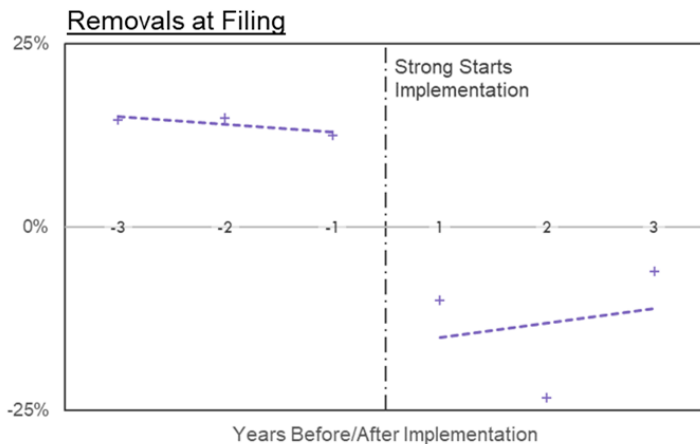
With Original Caregiver at 12 Months

We estimate a 9 percentage point increase in the number of children residing with their original caregiver 12 months after filing. However, this estimate did not reach statistical significance, with an estimated one in four probability that this observation is only due to chance. Also, as shown in Figure 2B, the parallel trend assumption is violated, meaning that any association between Strong Starts implementation and change in this outcome may not be causal.

However, there is still good reason to believe the estimated increase in the proportion of children residing with their original caregiver 12 months after filing is meaningful. A reduction in removals at filing among Strong Starts eligible children appearing in the Strong Starts judge's courtroom could be accompanied by either (a) an increase in removals following a period of court-ordered supervision, or (b) a reduction in the proportion of those children removed to out-of-home care who were reunified with their families within 12 months. That is, the positive program impact on fewer removals at filings could be undermined by negative impacts on longer term outcomes in terms of more removals after that initial filing or fewer reunifications among those children who are removed. We checked to see if either of these was the case and observed no statistically significant change; in fact, the proportion of children removed following a period of supervision also decreased (Table 5). If fewer children are removed and there is no change in the likelihood of reunification among the children who are removed, it follows that children would be more likely to reside with their original caregiver 12 months after filing.

Table 5. Changes in Post-Supervision Removals and Post-Removal Reunification Among Eligible Cases in the Strong Starts Courtroom

	Pre-Implementation	Post-Implementation	Chi-Square Test P-Value
Children Removed Following Supervision	14% (n=9)	8% (n=7)	0.20^{ns}
N	63	90	
Children Reunified w/in 12 Months of Filing	44% (n=20)	36% (n=12)	0.47^{ns}
N	45	33	

+$p<.10$,* $p<.05$, ** $p<.01$, ***$p<.001$, $^{ns}p>.10$ (not significant)

Discussion and Conclusion

Strong Starts program implementation successfully led to an increase in family preservation for Strong Starts eligible families appearing before the Strong Starts judge. Following program implementation, qualifying subject children were significantly less likely to be removed from their initial caregiver at case filing and there are indications that they were more likely to reside at home with their initial caregiver 12 months after filing. The findings from this study combined with the results from the 2021 evaluation showing a decrease in subsequent child welfare court episodes for Strong Starts participants (Sharlein 2021), paint the picture of an intervention that leads to more stable outcomes for children, better family court outcomes for their respondent caregivers, and the preservation of the attachment relationship between the two.

These results underscore the importance of Strong Starts' inclusion of families with children under court-ordered supervision and support ZERO TO THREE's recently updated guidance encouraging ITCTs to accept families of children without a removal (Hudson 2017). A removal decision is an extremely high-stakes moment in a child welfare proceeding. Strong Starts coordinators anecdotally note that judges and practitioners see the program as something of an alternative to removal for cases that are on the margins; when practitioners perceive a family to be at risk, the coordinator provides an extra layer of expertise and support that may make practitioners more comfortable with keeping the family intact. We also know from the 2021 evaluation that practitioners apply the knowledge of available resources and early childhood mental health they learn from Strong Starts to other cases, which might lead them to make similar decisions on eligible cases that are not selected into the program (Sharlein 2021). Taken together, the targeted support and expanded practitioner knowledge form a plausible mechanism for the estimated decrease in removals for Strong Starts eligible children appearing before the Strong Starts judge.

While this evaluation's results are promising, they only cover the implementation of Strong Starts in a single courtroom of a single courthouse. It is possible that this courtroom was uniquely situated for a successful implementation of the Strong Starts program and that Strong Starts implementation in other courtrooms has not had the same impact. This study is a starting point, and future research should explore the effect of program implementation at other sites. It would be ideal to conduct studies that utilize a similar design at those sites; however, it may not always be appropriate. Other courts may not have a bench of judges large and/or consistent enough to create a valid comparison group for the Strong Starts judge, or case assignment may not be random. In these circumstances, a pre/post-implementation

design that compares outcomes for a Strong Starts judge's program-eligible families to outcomes for that judge's Strong Starts-ineligible families would be more appropriate.

As Strong Starts expands across New York State and similar ITCTs expand around the country, it is important for them to consider how best to leverage their resources to maximize their impact on the families they serve. The results of this study suggest that these programs can maximize their impact on family preservation the overrepresentation of young children in care by including families of children under court-ordered supervision in the program, thereby preventing attachment disruptions and circumventing the extended and often fraught permanency process.

References

Afifi, Tracie O., Samantha Salon, Isabel Garcés, Shannon Struck, Janique Fortier, Tamara Taillieu, Ashley Stewart-Tufescu, Gordon J. G. Asmundson, Jitender Sareen, and Harriet L. MacMillan. 2020. "Confirmatory Factor Analysis of Adverse Childhood Experiences (ACEs) Among a Community-Based Sample of Parents and Adolescents." *BMC Pediatrics* 20:178. doi: [10.1186/s12887-020-02063-3](10.1186/s12887-020-02063-3).

Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan. 2004. "How much should we trust differences-in-differences estimates?" *Quarterly Journal of Economics* 119(1): 249–75. doi: [10.1162/003355304772839588](10.1162/003355304772839588).

Casanueva, Cecilia, Sarah Harris, Christine Carr, Chelsea Burfeind, and Keith Smith. 2017. *Final Evaluation Report of the Quality Improvement Center for Research-Based Infant-Toddler Court Teams*. Research Triangle Park, NC: RTI International. Retrieved April 18, 2024 from https://www.zerotothree.org/wp-content/uploads/2019/12/Final-Evaluation-QICCT.pdf.

Casanueva, Cecilia, Sarah Harris, Christine Carr, Chelsea Burfeind, and Keith Smith. 2019. "Evaluation in multiple sites of the Safe Babies Court Team approach." *Child Welfare* 97(1): 85-107.

Casey Family Programs. 2023. *Issue Brief: Strong Families: What Impacts Placement Stability?* Retrieved September 15, 2023 from https://www.casey.org/media/23.07-QFF-SF-Placement-Stability-Impacts.pdf.

Child Welfare Information Gateway. 2017. *Bulletin for Professionals October 2017: Supporting Successful Reunifications.* Retrieved September 13, 2023 from https://www.childwelfare.gov/pubPDFs/supporting_reunification.pdf.

Coan, James A. 2016. "Chapter 12: Toward a Neuroscience of Attachment." Pp. 242-69 in *Handbook of Attachment, Third Edition: Theory, Research, and Clinical Applications*, edited by Jude Cassidy and Phillip R. Shaver. New York: The Guilford Press.

Faria, Ann-Marie, Jill Bowden, Jameela Conway-Turner, Jingtong Pan, Tina Ryznar, Laura Michaelson, Taletha Derrington, and Jill Watson. 2020. *The Safe Babies Court Team Evaluation: Changing the Trajectories of Children in Foster Care*. Washington, DC:

American Institute for Research. Retrieved April 18, 2024 from https://www.air.org/sites/default/files/Safe-Babies-Court-Team-Evaluation-FINAL-092520.pdf.

Hudson, Lucy. 2017. *A Guide to Implementing the Safe Babies Court Team Approach*. Washington, DC: ZERO TO THREE. Retrieved April 18, 2024 from https://ecc.flcourts.gov/content/download/837615/file/0%20A%20Guide%20to%20Implementing%20the%20SBCT%20Approach.pdf.

James Bell Associates. 2009. *Evaluation of the Court Teams for Maltreated Infants and Toddlers: Final Report*. Submitted to the Office of Juvenile Justice and Delinquency Prevention, U.S. Department of Justice. Retrieved April 18, 2024 from https://www.zerotothree.org/wp-content/uploads/2019/12/James-Bell-Associates-Evaluation-of-the-Court-Teams-for-Maltreated-Infants-and-Toddlers-2009..pdf.

JBS International, Inc on behalf of the Children's Bureau. 2020. *Child and Family Services Reviews Aggregate Report, Round 3: Fiscal Years 2015-2-18*. Retrieved March 23, 2023 from https://www.acf.hhs.gov/sites/default/files/documents/cb/cfsr_aggregate_report_2020.pdf.

McCombs-Thornton, Kimberly L., and E. Michael Foster. 2012. "The Effect of the ZERO TO THREE Court Teams Initiative on Types of Exits from the Foster Care System – A Completing Risks Analysis." *Children and Youth Services Review* 34:169-78. doi: 10.1016/j.childyouth.2011.09.013.

New York City Family Court. 2023. *Looking Forward—A Vision for Our Future*. Retrieved May 15, 2023 from https://ww2.nycourts.gov/COURTS/nyc/family/vision.shtml.

Olden, Andreas and Jarle Møen. 2022. "The Triple Difference Estimator." *The Econometrics Journal* 25(3): 531–553. doi: 10.1093/ectj/utac010.

Reeves, Barnaby C., George A. Wells, and Hugh Waddington. 2017. "Quasi-Experimental Study Designs Series-Paper 5: A Checklist for Classifying Studies Evaluating the Effects on Health Interventions—A Taxonomy without Labels." *Journal of Clinical Epidemiology* 89:30-42. doi: 10.1016/j.jclinepi.2017.02.016.

Shonkoff, Jack P., and Deborah A. Phillips, eds. 2000. *From Neurons to Neighborhoods: The Science of Early Childhood Development.* Washington, DC: National Academy Press. doi: [10.17226/9824](#).

Sharlein, Jeffrey. 2021. *Helping the Youngest Start Life Strong: A Mixed Methods Evaluation of the Strong Starts Court Initiative.* New York, NY: Center for Justice Innovation. Retrieved March 23, 2023 from https://www.innovatingjustice.org/sites/default/files/media/document/2021/SSCI_eval_full_report_FINAL4.pdf.

Sroufe, L. Alan. 2005. "Attachment and Development: A Prospective, Longitudinal Study from Birth to Adulthood." *Attachment & Human Development* 7(4):349-367. doi: [10.1080/14616730500365928](#).

Thompson, Ross A. 2016. "Chapter 16: Early Attachment and Later Development: Reframing the Questions." Pp. 330-48 in *Handbook of Attachment, Third Edition: Theory, Research, and Clinical Applications*, edited by Jude Cassidy and Phillip R. Shaver. New York: The Guilford Press.

Appendix A.
Details on Model, Analysis Cohorts, and Results

Statistical Approach

The DDD can be expressed in an equation as the difference between 2 DiDs (Olden and Møen 2022):

$$\delta_{triple} = [(\bar{Y}_{J=1, E=1, Post=1} - \bar{Y}_{J=1, E=0, Post=1}) - (\bar{Y}_{J=1, E=1, Post=0} - \bar{Y}_{J=1, E=0, Post=0})]$$
$$- [(\bar{Y}_{J=0, E=1, Post=1} - \bar{Y}_{J=0, E=0, Post=1}) - (\bar{Y}_{J=0, E=1, Post=0} - \bar{Y}_{J=0, E=0, Post=0})]$$

- ***where*** *J=1: Strong Starts judge | J=0: other judge | E=1: Strong Starts eligible | E=0: Strong Starts ineligible | Post=1: post-implementation | Post=0: pre-implementation | $\bar{Y}$=measured proportion of cases (e.g., removal rate)*

or as a regression (Olden and Møen 2022):[13]

$$Y = \beta_0 + \beta_1 J + \beta_2 E + \beta_3 Post + \beta_4(J*E) + \beta_5(J*Post) + \beta_6(E*Post) + \beta_7(J*E*Post) + X + \epsilon$$

- ***where*** *Y=outcome of interest | J(0,1)=judge dummy (SSCI judge or other judge) | E(0,1)=eligibility dummy | Post(0,1)=time period dummy | X=vector of control variables | ϵ=error*
- ***β_7****, the regression coefficient for the 3-way interaction term, is the triple-difference estimator for the effect of the program on eligible families in the Strong Starts judge's court part post-implementation*
- *control variables **X** include child's age at filing, respondent age at filing (average age if multiple respondents[14]), number of respondents, number of siblings, number of previous child protective family court episodes for subject child, number of previous*

---

[13] Regression formulas for the DiDs included in the expanded results are:
- **Model 1** Comparison between Strong Starts eligible cases and other cases in Strong Starts Courtroom: $Y = \beta_0 + \beta_1 E + \beta_2 Post + \beta_3(E*Post) + X + \epsilon$; &
- **Model 2** Comparison between Strong Starts eligible cases in the Strong Starts and other courtrooms: $Y = \beta_0 + \beta_1 J + \beta_2 Post + \beta_3(J*Post) + X + \epsilon$

where β_3 = DiD estimator.

[14] For the <1% of cases with more than two respondents recorded, we average the age of the oldest and youngest respondents.

child protective family court episodes for respondent (maximum number if multiple respondents), and child gender

Olden and Møen show that in DDD settings where there are only two groups (i.e., in this case, Strong Starts judge as opposed to another judge), estimating the DDD as a regression using the post-implementation dummy as the only time variable and calculating non-clustered standard errors is optimal. Therefore, we have used this approach in the current evaluation. However, although the authors show it is the best approach, we acknowledge their caution that the calculated DiD and DDD estimators are still prone to over-rejection of the null hypothesis, meaning that they are more likely to detect effects when no effect exists than the stated level of statistical significance suggests (Bertrand et al. 2004; Olden and Møen 2022). All analyses were conducted using SPSS software.

Cohort Adjustments

For the purposes of this study, cases are considered Strong Starts eligible if the oldest subject child is two years old or younger and the family has no prior child welfare court involvement. This differs from the criteria outlined in the introduction, which require the oldest subject child to be no older than three years old. We made this adjustment to account for the program's observed demographics and create a more accurate cohort and comparison; no three-year-olds participated in the program throughout this study's time frame. Additionally, exceptions were sometimes made to the stated eligibility criteria: 17% (9/53) of participants in the full data set do not meet stated Strong Starts criteria. These participant exceptions are not drawn into this study's Strong Starts eligible sample for two reasons: (1) the intervention was not designed for them, and (2) their inclusion would necessitate a less representative expansion of the sample criteria.

For the second outcome (with original caregiver at 12 months), we excluded all filings in the year leading up to implementation from the analysis. This is to avoid including pre-implementation cases whose outcomes may have been impacted by Strong Starts implementation.

Expanded DDD Results

Typical presentation of DDD results include the related DiD models as shown in Table 6. Model 1 compares Strong Starts eligible cases to ineligible cases within the Strong Starts Strong Starts judge's courtroom; Model 2 compares Strong Starts eligible cases in the Strong Starts judge's courtroom to eligible cases in other courtrooms; Model 3 is the DDD.

Table A. Effects of Strong Starts Implementation on Family Preservation Measures, Expanded

	Model 1 DiD Strong Starts Eligible versus Other Cases in Strong Starts Courtroom	Model 2 DiD Strong Starts Eligible in Strong Starts Courtroom versus Other Courtrooms	Model 3 DDD
Removals at filing	-0.244***	-0.236**	-0.252***
p-value	(<.001)	(.002)	(<.001)
95% Confidence Interval	(-0.376, -0.112)	(-0.382, -0.090)	(-0.389, -0.114)
ϵ	0.067	0.074	0.070
N	2,076	1,672	17,898
With original caregiver at 12 months[1]	0.150*	0.062	0.091
p-value	(.033)	(.448)	(.220)
95% Confidence Interval	(0.0012, 0.287)	(-0.098, 0.221)	(-0.055, 0.237)
ϵ	0.070	0.081	0.074
N	1,725	1,394	15,263

+p<.10,* p<.05, ** p<.01, ***p<.001
[1] Excludes cases filed in the year preceeding Strong Starts implementation.

The DDD coefficient for **removals at filing** indicates that there was a significant 25.2 (±13.7, p<.001) percentage point reduction following Strong Starts implementation that meets the parallel trend assumption (Figure 2). DiD coefficients and significance levels for Models 1 and 2 are similar to those for the DDD model (Table A).

A 9.1 (±14.6, p=.220) percentage point increase in **children residing with their original caregiver 12 months after filing** was also estimated but did not reach statistical significance or meet the parallel trend assumption (Figure 2). DiD coefficients for Models 1 and 2 follow a similar trend as the DDD coefficient, but Model 1 has a larger estimated impact that shows a higher degree of significance (Table A). This indicates a court-wide relative increase in program-eligible children residing with their original caregiver—independent of Strong Starts implementation—led to the larger apparent impact of program implementation seen in Model 1, which does not account for court-wide trends.

Appendix B.
Decision Tree for Case Assignment and Cohort Development

Figure B. Case Assignment, Cohort Development, and Triple Difference Development