The author(s) shown below used Federal funding provided by the U.S. Department of Justice to prepare the following resource:

Document Title:	**Oklahoma Model Standards Study (OKMSS)**
Author(s):	**Margaret Lloyd Sieger, PhD, Jon D. Phillips, PhD**
Document Number:	311037
Date Received:	12/2025
Award Number:	2018-DC-BX-0007

Oklahoma Model Standards Study (OKMSS)
Final Report

Margaret Lloyd Sieger, PhD
University of Kansas School of Medicine
Department of Population Health
msieger@kumc.edu
3901 Rainbow Boulevard
Kansas City, KS 66160

Jon D. Phillips, PhD
University of Connecticut
School of Social Work
jon.phillips@uconn.edu
38 Prospect Street
Hartford, Connecticut 06103

Contents

Introduction

This report is divided into seven sub-studies:

1. [Assessing Substance Use Treatment as a Mediator in Family Treatment Court Outcomes](#)
2. [Assessing Addiction Severity as a Moderator of Family Treatment Court Outcomes](#)
3. [Family Treatment Court Cost Analysis](#)
4. [Development and Pilot Testing of the Model Standards Implementation Scale (MSIS)](#)
5. [Key FTC Practices Associated with Reunification](#)
6. [Survey Findings of Caregivers with Maltreatment Cases](#)
7. [Assessing Cross-systems Collaboration as a Predictor of Implementation Success](#)

For each sub-study, we describe the purpose and research questions, timeframes and procedures for data collection, analytic methods, results, limitations, and conclusions. The research presented in this evaluation received human subjects approval from the University of Connecticut (UConn) Institutional Review Board and the National Institute of Justice Human Subjects Protection Officer. A data use agreement among UConn, Oklahoma Department of Human Services, and Oklahoma Department of Mental Health and Substance Abuse Services was executed to ensure the privacy and confidentiality of the administrative records used in this research.

Before presenting each sub-study in its entirety, this report begins with a [summary of accomplishments](#), a [summary of key findings](#), and a list of [policy and practice recommendations](#).

Summary of Accomplishments

Development of the Family Treatment Court (FTC) Implementation Tools
- We developed and pilot tested the **Model Standards Implementation Scale (MSIS)**, the first validated instrument for assessing alignment with Family Treatment Court (FTC) Best Practice Standards (the *Standards*).
- The MSIS includes a comprehensive site visit protocol, with structured interview guides, template documents, and a sophisticated scoring instrument.
- We launched a web-based version of the tool, the Family Treatment Court Implementation Tool (FIT), to support broader dissemination and enhance usability.

Methodological Contributions
- **Cross-system data linkage** was used to overcome limitations of previous research by integrating administrative data from child welfare, substance use treatment, and court agencies (Sub-Studies 1, 2, 3, 5; see *Sub-Study 1: Methods*).
- **Advanced statistical techniques** were employed to address selection bias:
  - Inverse probability weighting (*Sub-Study 1: Methods—IPW Description*)
  - Propensity score matching (Sub-Studies 2, 3; see *Sub-Study 2: Methods—PSM Description*)
  - Propensity score weighting (*Sub-Study 5: Methods—PSW Description*)
- **Survey methodology:**
  - Engagement of hard-to-reach caregivers at the start of their maltreatment case (*Sub-Study 6: Methods*)
  - Engagement of child welfare and family treatment court (FTC) professionals (*Sub-Study 7: Methods*), capturing stakeholder insights often absent from administrative data.

Summary of Key Findings

Sub-Study 1: Assessing Substance Use Treatment as a Mediator in FTC Outcomes
- o Families who participated in FTC were 66% ($p<.01$) more likely to reunify than those who did not, regardless of treatment completion.
- o Caregivers who completed treatment were 182% ($p<.001$) more likely to reunify than those who did not, regardless of FTC participation.
- o Treatment completion did not explain the full impact of FTCs, suggesting these courts support reunification through additional pathways.

Sub-Study 2: Assessing Addiction Severity as a Moderator of FTC Outcomes
- o FTC participants had higher reunification rates across all levels of addiction severity.
- o The benefit of FTC participation was strongest for caregivers with higher addiction severity.
  - ▪ Low addiction severity: FTC caregivers 44% more likely to reunify (non-significant)
  - ▪ Moderate addiction severity: FTC caregivers 120% ($p<.01$) more likely to reunify
  - ▪ High addiction severity: FTC caregivers were 124% ($p<.05$) more likely to reunify

Sub-Study 3: FTC Cost Analysis
- o FTC participants reunified 164 days sooner, on average, resulting in an estimated 29,975 total foster care days saved.
- o FTC implementation costs over the 5-year study were estimated at $2.81 million, while savings from reduced foster care utilization were estimated at $2.76 million.
- o Foster care savings did not fully offset FTC implementation costs, resulting in a net difference of approximately -$55,036 (or -$301 per child).

Sub-Study 4: Development and Pilot Testing of the Model Standards Implementation Scale (MSIS)
- o The MSIS was developed to assess implementation of 67 best practice provisions related to 8 standards for FTCs.
- o The MSIS demonstrated good inter-rater reliability, known-groups validity, and acceptable internal consistency.
- o In FTC pilot testing, 84% of provisions were rated "Fully Implemented" by at least one rater.
- o Poor implementation was most common in Monitoring and Evaluation provisions.

Sub-Study 5: Key FTC Best Practices Associated with Reunification
- o 85% of FTC best practice provisions were positively associated with reunification.
- o Provisions most strongly linked to reunification included:
  1. Evidence-based manualized treatment (OR=167.6; $p<.05$)
  2. Trauma-specific services for children and caregivers (OR=25.4; $p<.05$)
  3. Recovery Supports (OR=7.9; $p<.001$)
  4. Use of Valid and Reliable Screening and Assessment for Caregivers and Families (OR=7.7; $p<.001$)
  5. Alcohol and other Drug Testing Protocols (OR=3.1; $p<.01$)

Sub-Study 6: Survey Findings of Caregivers with Maltreatment Cases
- o Preliminary descriptive analysis (*N*=27) identified potential differences between FTC and non-FTC participants, with FTC participants appearing to show greater readiness for change and greater trauma symptom severity.
- o A larger sample is needed to confirm and further explore these patterns.

Sub-Study 7: Assessing Cross-systems Collaboration as a Predictor of Implementation Success
- o The capacity for interagency collaboration was significantly associated with the extent of overall implementation success and the implementation of four best practice standards, but the relationship was contrary to what was expected, with sites that had less implementation success having greater capacity than those with more implementation success.
- o The findings suggest that as sites achieve higher levels of implementation they may experience collaboration challenges that need to be addressed.

Policy & Practice Recommendations

Based on the findings from this evaluation, we make the following recommendations. Each recommendation is accompanied by the key finding(s) that informed it, *presented in italics*.

Supported by [Sub-Study 1](#):

1) **Expand access to FTCs and invest in supportive services.** *Given their independent effect on reunification*, FTCs should be scaled to reach more families affected by substance use disorders by creating new programs and expanding capacity within existing ones. Funding should also prioritize the family-focused services embedded in the FTC model, such as parenting support, trauma-informed care, and wraparound services, which likely contribute to their overall effectiveness. To improve fairness and procedural justice, FTCs should standardize and monitor referral and enrollment pathways; in our study, *White caregivers were disproportionately overrepresented in the FTC group.*

2) **Sustain and strengthen harm reduction principles in FTC and child welfare settings.** *FTCs increase reunification even among caregivers who do not complete treatment*, highlighting the importance of supporting caregivers during relapse, continuing to engage them in treatment and child visitation, and responding with treatment adjustments rather than sanctions.

3) **Promote treatment completion as a key reunification strategy.** *Treatment completion remains the strongest individual predictor of reunification.* Expanding access to high-quality treatment and reducing barriers to engagement and retention should be a central focus for courts and child welfare agencies.

Supported by [Sub-Study 2](#):

4) **Expand access to FTCs.** Consistent with Sub-Study 1, this study supports continued expansion of FTCs. The results indicate that families across all levels of addiction severity are significantly more likely to achieve family reunification when participating in an FTC.

5) **Use addiction severity assessments to guide screening and prioritization.** Clinical tools like the Addiction Severity Index (ASI) can guide enrollment decisions by identifying families most likely to benefit when capacity is limited. *FTC participants with moderate to high severity experienced the greatest gains in reunification.*

Supported by [Sub-Study 3](#):

6) **Invest in FTCs as a fiscally responsible strategy despite modest net costs.** Policymakers should recognize that this analysis captures only direct foster care savings and excludes broader cross-system benefits, such as improved caregiver recovery, reduced recidivism,

healthcare savings, and enhanced child well-being, which likely provide additional value. *While FTC implementation costs slightly exceeded foster care savings ($55,036 net cost or $301 per child), FTCs represent a sound investment that accelerates family reunification by an average of 164 days.*

Supported by [Sub-Study 4](#):

7) **Use the Model Standards Implementation Scale (MSIS) to support implementation monitoring and continuous improvement.** This study demonstrates that the MSIS is a reliable and valid tool for assessing FTC implementation of the Best Practice Standards (the Standards). This study also reveals that "Monitoring and Evaluation," one of the eight Standards, was the least implemented across sites. The MSIS can help address this gap by providing courts and states with a structured, evidence-based tool to strengthen evaluation and support continuous program improvement.

Supported by [Sub-Study 5](#):

8) **Prioritize implementation of high-impact FTC provisions related to evidence-based treatments, trauma-specific services, and recovery supports to improve reunification outcomes.** State administrators and organizations providing technical assistance should emphasize these practices through training, guidance, and continuous quality improvement efforts. *These provisions were among the most strongly associated with reunification.*

9) **Embed the MSIS into program monitoring and state oversight systems to strengthen implementation and promote reunification.** Using the MSIS can help courts and oversight bodies identify and address gaps in practice, promoting more consistent and effective delivery of FTC best practices. *Findings show that greater implementation of core practices is linked to higher reunification rates.*

10) **Strengthen cross-system partnerships to expand access to supportive services for families.** Jurisdictions should ensure FTC participants have access to a comprehensive range of services, including a full continuum of substance use treatment that is evidence-based and culturally responsive, trauma-specific services, recovery supports, and complementary supports, such as housing and transportation. Achieving this requires strong partnerships across behavioral health, child welfare, and community-based agencies. *Several of the provisions most strongly associated with reunification relied on services delivered through external partners.*

Supported by [Sub-Study 6](#):

11) **Findings may support existing calls to enhance early assessment practices.** While the *Standards* call for timely, comprehensive assessment using structured instruments, this exploratory study offers preliminary support for two specific tools: the SOCRATES 8D Personal Drug Use Questionnaire and the Trauma Symptom Checklist-40 (TSC-40). *FTC participants*

reported greater readiness for change and greater trauma symptom severity, suggesting these instruments may help identify caregiver needs and inform service matching, if findings are replicated.

Supported by [Sub-Study 7](#):

12) **Support child welfare systems in developing policies, structures, and practice that support and enhance interagency collaboration.** *The association between capacity for interagency collaboration and implementation success* indicates that agencies should conduct ongoing assessments of their collaborative practices and relationships with other agencies to identify strengths and areas for improvement as they work to implement and sustain best practices for FTCs. Given that *capacity for interagency collaboration was perceived to be less in sites experiencing moderately high implementation success*, systems should develop and formalize interagency policies, structures, and practices to address barriers to collaboration as they arise. Systems may benefit from learning from systems that have achieved both high implementation success and high capacity for interagency collaboration.

Sub-Study 1: Assessing Substance Use Treatment as a Mediator in Family Treatment Court Outcomes

Study Authors

Margaret Lloyd Sieger, PhD; Jessica Becker, MSW; Jon Phillips, PhD; Cindy Nichols, PhD; Elizabeth J. Goldsborough, MSW; John Prindle, PhD

Complete Citation

Lloyd Sieger, M., Becker, J., Phillips, Nichols, C., Goldsborough, E. J., & Prindle, J. (R&R). Substance use treatment completion does not mediate the relationship between family treatment court participation and reunification: Results from five courts in the Southwestern U.S. *Drug and Alcohol Dependence*.

Study Purpose

Family treatment courts (FTCs) have emerged as a specialized approach to addressing child welfare cases involving caregiver substance use. By bringing together judges, treatment providers, and child welfare professionals, FTCs aim to help caregivers access and complete substance use treatment, with the ultimate goal of family reunification. While case studies published over the past thirty years suggest FTC programs achieve better outcomes than traditional deprived courts (TDCs), methodological limitations in earlier research limit causal inference. Additionally, it is often assumed that FTCs improve reunification outcomes by increasing treatment completion; yet this mechanism has not been empirically tested. This study seeks to strengthen the evidence base by using inverse probability weighting (IPW) to reduce selection bias and generate estimates of the effects of FTC participation on treatment completion and reunification that more closely approximate those from a randomized controlled trial. In addition, we apply mediation analysis to examine whether the effect of FTC participation on reunification is explained by its success in increasing treatment completion among caregivers.

Research Questions

This study had two research questions:

    (1) What is the effect of FTC participation on likelihood of substance use treatment completion?
    (2) What is the effect of FTC participation on likelihood of family reunification?
    (3) To what extent does treatment success mediate the effect of FTC participation on reunification likelihood?

Methods

    Design

This study used a longitudinal, observational research design.

Data Sources and Study Population

Data for this study were collected from six counties in a Southwestern state with operational FTCs from 2018 to 2022. Three administrative data sources were linked to generate a sample of dyadic records that include information on child demographic and case characteristics, caregiver demographic and substance use treatment characteristics, and FTC participation status. The data sources included: 1) Oklahoma Department of Human Services (DHS) foster care records, 2) Oklahoma Department of Mental Health and Substance Abuse Services (DMHSAS) treatment admission records, and 3) a list of FTC participants provided by Oklahoma DMHSAS, which included their FTC enrollment and discharge dates.

Timeframe

This study examined child welfare cases with adjudication hearing dates that occurred between October 1, 2018 and September 30, 2022. The DHS extract includes records with removal dates as early as April 30, 2015, and return dates through May 25, 2023. The treatment admission data include adults admitted on or after July 7, 2016, and discharged on or before March 28, 2024. FTC participation records include admission dates beginning December 11, 2013, and discharge dates through April 10, 2023.

Data Linkage and Sample Selection

Probabilistic and Deterministic Linking Procedures

Perpetrator records and substance use treatment admission records were linked using probabilistic matching procedures (see Figure 1). The goal of this linkage was to identify corresponding substance use treatment records for adults identified as perpetrators in the foster care records, thus enabling integration of caregivers' substance use treatment variables in the outcome modeling. Linkage was conducted using Probabilistic Record Linkage for Families (PRLF), an application built on Python RecordLinkage package that incorporates several machine learning algorithms of probability scoring (Prindle et. al., 2024). Scoring accounts for dimensions of similarity and dissimilarity using first name, last name, birthdate, and sex. A match was retained if the probability score exceeded 0.70 using an XGBoost classifier model. In instances in which more than one potential match exceeded this threshold (n=16), the record with the higher probability score was retained. This process yielded 831 linked adults, representing 10.8% of perpetrators, 4.4% of adult treatment records, and 95.9% of FTC admission records.

These 831 records were then deterministically merged using a perpetrator ID variable to 9,007 child foster care records. Each foster care record represents a unique perpetrator-child-referral number combination, where the referral number corresponds to a maltreatment report and subsequent foster care placement. Because a single perpetrator may be associated with multiple children, this merge yielded 2,022 perpetrator-child dyads.

Figure 1: Data Linkage Sample Sizes
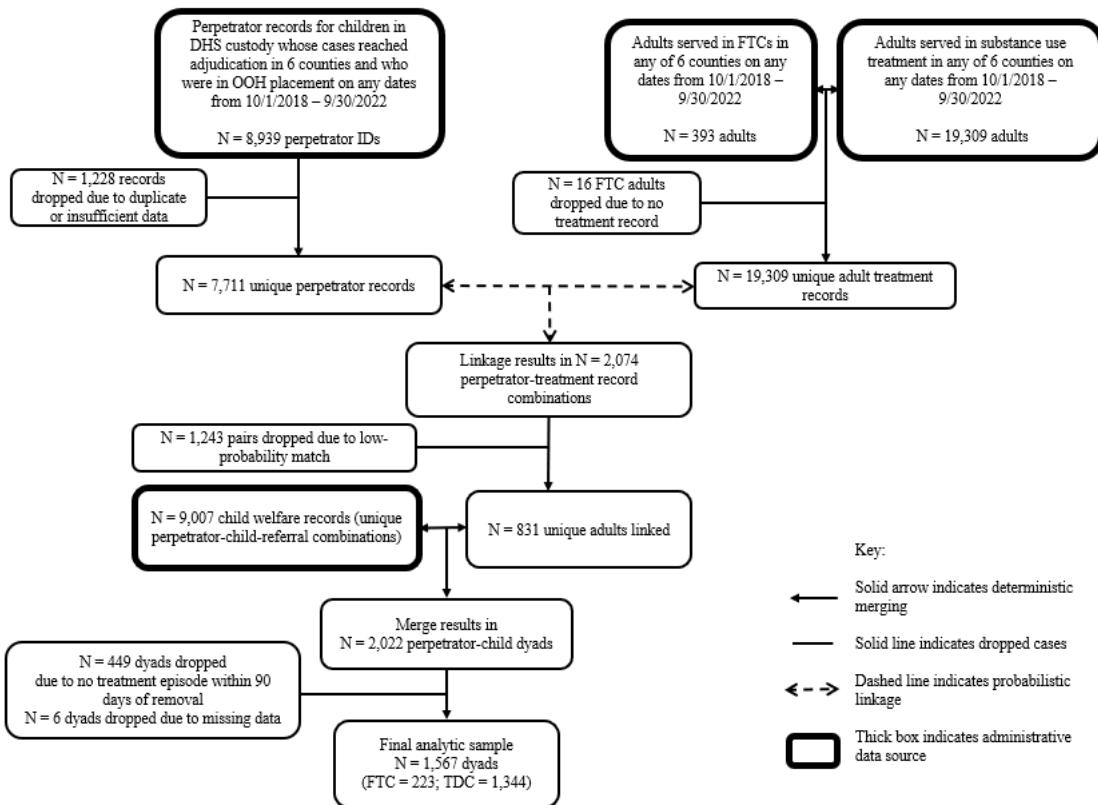
Treatment Episode Selection Criteria

For perpetrators with multiple treatment episodes (up to three per perpetrator), we implemented a systematic approach to select the single episode most likely connected to the child welfare case, based on temporal alignment with the child's removal. Preference was given to treatment episodes that overlapped with the foster care placement period, as these likely reflect treatment initiated in response to the child welfare case. Overlap was defined as treatment episodes with admission dates before the child's return date and/or discharge date after the child's removal date. When multiple episodes overlapped or none did, we selected the episode with the admission date closest to the removal date.

If no overlapping episodes existed in the data, we retained non-overlapping treatment episodes occurring within 90 days of removal, selecting the episode with the admission date closest to the removal date. These episodes are plausibly related to the child welfare case because treatment may begin shortly before removal as families try to address concerns leading to placement, or shortly after removal due to delays in treatment access, assessment, or caregiver readiness. Treatment episodes that did not overlap and had admission dates more than 90 days before or after removal were flagged as unrelated to the child welfare case and their treatment variables were cleared. These dyads were considered to have missing treatment data for analysis, as the available episodes were less likely to reflect the relevant treatment context.

Exclusions and Final Analytic Sample

We excluded 449 dyads for whom no treatment data were retained based on the selection criteria described above. An additional six dyads were excluded due to missing data: two lacked perpetrator race data and four lacked Addiction Severity Index (ASI) scores because they were under the age of 18 and had instead been assessed using the Teen ASI. In total, 455 dyads (22.5% of the merged sample) were excluded. The final analytic sample included 1,567 dyads, comprising 223 in the FTC group (189 children and 113 caregivers) and 1,344 in the TDC group (1,156 children and 703 caregivers). Each observation (i.e., dyad) represents a unique perpetrator-child-referral number combination.

Some children and perpetrators appear in multiple dyads, which were retained because each dyad reflects a distinct treatment episode (with its own treatment completion status) and a separate foster care episode (with its own reunification outcome). Four records reflected two perpetrator-child duplicates. The duplicate dyads were retained because each was associated with a different referral number, indicating separate cases that occurred at different timepoints and may have differed in context, treatment needs, and the ages of caregiver and child.

Measures

Dependent Variable

Reunification was derived from the placement exit reason variable from administrative foster care records. Originally recorded in eight categories, this variable was recoded into a binary indicator, where cases were coded as "1" if the exit reason was "Reunification" or "Reunified with Custodial Parent" and "0" for all other reasons: "Adoption," "Child Aged Out/Emancipation," "Custody to Relative," "Guardianship – Non-Relative," "Guardianship – Relative," and "Tribal Jurisdiction."

Mediator Variable

Substance use treatment completion was derived from the treatment discharge type variable from administrative treatment admission records. Originally recorded in 12 categories, this variable was recoded into a binary indicator, where cases were coded as "1" for participants discharged as "Completed Court Treatment" or "Completed Treatment" and "0" for all other discharge types: "Administrative Discharge," "Broke Rules," "Death," "Failed to Begin Treatment," "Incarcerated," "Left Against Counselor Advice," "Medical," "Moved," "Transferred to another Treatment Facility," and "Treatment Incompatibility."

Independent Variable

FTC participation was coded as "1" for participants enrolled in an FTC during the study period and "0" for participants enrolled in a traditional child welfare court (i.e., traditional deprived court, or TDC).

Control Variables

As recommended by [Brookhart et al. (2006)](#), variables that prior studies have found to be associated with the outcome of interest (reunification) were used to determine the propensity scores and inverse probability weights. Overall, the IPW model included seventeen covariates related to caregiver and child demographics, child welfare case characteristics, and caregiver substance use treatment. Demographic covariates included continuous variables for child age and caregiver age, a three-category variable for caregiver race, and a binary indicator for caregiver sex.

Child welfare covariates included a four-category variable for total number of allegations in the maltreatment report, a three-category variable for number of placements during the removal, and a binary indicator for kinship placement.

Caregiver substance use treatment covariates consisted of a three-category variable for level of care recommendation; binary indicators for alcohol, cannabis, an opioid, and a stimulant as primary substances of choice; continuous Addiction Severity Index (ASI) scores across drug, employment, family, and psychiatric domains; and a four-category variable for number of lifetime treatment admissions. The decision to categorize certain continuous variables, such as number of allegations and placements, was made to limit the influence of outliers.

Procedures

Stata Version 18.0 was used to clean and analyze the data.

IPW (Inverse Probability Weighting)

To overcome the lack of randomization into the treatment condition (and subsequent differences between FTC and TDC dyads), inverse probability weights were applied to two modeling approaches: one estimating unstandardized coefficients and another calculating odds ratios to aid interpretability. These weights were derived in Stata, where propensity scores were calculated using logistic regression to estimate the probability of FTC participation based on demographic, child welfare, and treatment-related covariates.

Inverse probability weights assigned greater weight to individuals who were less likely to be in their respective groups, thereby balancing covariate distributions between the FTC and TDC groups and reducing bias in the estimation of FTC participation effects. Two key metrics were used to assess covariate balance: the standardized difference in weighted means (standardized mean differences; SMDs) and the variance ratios of weighted variance (VRs). According to [Rubin (2001)](#), covariates are considered balanced when the absolute values of SMDs are below 0.25 and VRs fall between 0.5 and 2. [Nguyen et al. (2017)](#) recommended a more conservative threshold for SMDs, suggesting that covariates with absolute SMDs greater than 0.10 should be adjusted for to address residual imbalance.

Descriptive Characteristics

Demographic, child welfare, and treatment characteristics of FTC and TDC caregivers and children were compared to assess differences between groups before and after applying inverse probability weights. Before weighting, chi-squared tests assessed differences in categorical variables, and one-way analysis of variance (ANOVA) was used to compare group means for continuous variables. After weighting, survey-weighted analyses were conducted. Categorical variables were assessed using weighted proportions and chi-squared tests. Continuous variables were analyzed using weighted means and survey-weighted regression, as one-way ANOVA cannot be applied with survey weights.

Weighted Logistic Regression Models

To examine the relationship among FTC participation, treatment completion, and reunification, two models were estimated in Mplus, both applying the IPW. The first model was a weighted logistic regression that assessed whether FTC participation predicted reunification. The second model was a weighted logistic regression model that tested whether treatment completion mediated the relationship between FTC participation and reunification. To address imbalances remaining after applying the weights, both models included covariates with SMDs greater than 0.10, consistent with [Nguyen et al.'s (2017)](#) recommendation. The mediation model was tested with counterfactually-defined causal effects as recommended by [Muthen and Muthen (2017)](#). The maximum likelihood estimator was used and confidence intervals for the indirect effect of FTC participation on reunification were calculated using 10,000 bootstrap replications.

Post-hoc analyses indicated the models met most of the assumptions of logistic regression (i.e., absence of multicollinearity and outliers, linearity of the logit, and independence of errors; [Stolzfus, 2011](#)). Post-hoc analyses did not reveal any evidence of multicollinearity or outliers among the predictor variables. Linearity of the logit was not tested as there were no continuous predictor variables. Because perpetrators could be associated with more than one child (i.e., in the case of siblings), we conducted a sensitivity analysis with one randomly selected observation from each perpetrator (n=816). These results are omitted below, but findings were substantively unchanged, suggesting that any intra-class correlation did not materially affect our conclusions.

Findings

Unweighted Descriptive Characteristics

Table 1 compares demographic, child welfare, treatment characteristics, and outcome variables for FTC and TDC caregivers and children prior to applying the IPW. Results demonstrate several significant differences between the unweighted groups. FTC participants were younger, with caregivers averaging 30.1 years compared to 32.7 years ($p<.001$), and their children averaging 3.2 years compared to 4.2 years ($p<.001$). FTC participants were more likely to be White (70.9% vs. 59.2)

and less likely to be Black or Native[1] ($p<.01$), and were also more likely to be female (75.8% vs. 69.1%; $p<.05$).

FTC participants tended to have fewer maltreatment allegations overall ($p<.05$), and there were significant differences in the placement patterns ($p<.05$). Children in FTC cases most commonly experienced 3-4 placements during the removal (45.7%), while placement patterns were more variable in the TDC group. They were also more likely to be placed with family members (59.2% vs. 50.0%; $p<.05$).

Treatment recommendations also differed significantly ($p<.001$). FTC participants were less likely to be referred to crisis or detox services or to outpatient care, and more likely to be referred to community living, halfway houses, or residential treatment. There were also significant differences in primary substance: FTC participants were less likely to report alcohol ($p<.001$) and more likely to report a stimulant ($p<.01$), with no significant differences in reporting cannabis or an opioid as their primary substance.

On the ASI, FTC participants scored higher on the drug ($p<.001$) and family ($p<.001$) domains, with no group differences on the employment or psychiatric domains. There were also significant differences in lifetime treatment admissions ($p<.05$), with FTC participants more likely to have 4-5 admissions and TDC participants more likely to have 6 or more.

Finally, FTC participants were more likely to achieve family reunification (50.2% vs. 36.5%; $p<.001$) and to complete substance use treatment (46.2% vs. 34.3%; $p<.01$).

Table 1: Unweighted Descriptive Characteristics (N=1,567)

Variable	FTC Participants (N=223) Mean (SD) or n (%)	TDC Participants (N=1,344) Mean (SD) or n (%)	Significance
Child age at removal	3.21 (3.71)	4.24 (4.27)	$p<.001$
Caregiver age	30.07 (5.56)	32.66 (6.93)	$p<.001$
Caregiver race			$p<.01$
White	158 (70.85)	796 (59.23)	
Black	40 (17.94)	314 (23.36)	
Native	25 (11.21)	234 (17.41)	
Caregiver female sex	169 (75.78)	928 (69.05)	$p<.05$
Total # of allegations			$p<.05$
1	149 (66.82)	787 (58.56)	

---

[1] Native includes individuals identified as American Indian, Alaska Native, Native Hawaiian, or other Pacific Islander

Table 1: Unweighted Descriptive Characteristics (N=1,567)

Variable	FTC Participants (N=223) Mean (SD) or n (%)	TDC Participants (N=1,344) Mean (SD) or n (%)	Significance
2	51 (22.87)	360 (26.79)	
3	21 (9.42)	138 (10.27)	
4 or more	2 (0.90)	59 (4.39)	
Total # of placements			*p*<.05
1-2	73 (32.74)	481 (35.79)	
3-4	102 (45.74)	495 (36.83)	
5 or more	48 (21.52)	368 (27.38)	
Any kinship placements	21 (9.42)	138 (10.27)	*p*<.05
Treatment level of care recommendation			*p*<.001
Community living or residential treatment	88 (39.46)	252 (18.75)	
Crisis or detox	12 (5.38)	177 (13.17)	
Outpatient	123 (55.16)	915 (68.08)	
Primary substance of choice			
Alcohol	20 (8.97)	293 (21.80)	*p*<.001
Cannabis	31 (13.90)	217 (16.15)	N.S.
Opioid	34 (15.25)	159 (11.83)	N.S.
Stimulant	134 (60.09)	639 (47.54)	*p*<.01
ASI domain scores			
Drug	6.83 (1.85)	6.28 (2.33)	*p*<.001
Employment	4.93 (2.66)	4.71 (3.04)	N.S.
Family	6.35 (2.05)	5.72 (2.52)	*p*<.001
Psychiatric	5.26 (2.71)	5.08 (2.84)	N.S.
# Lifetime treatment admissions			*p*<.05
1	30 (13.45)	221 (16.44)	
2-3	62 (27.80)	407 (30.28)	
4-5	72 (32.29)	311 (23.14)	
6 or more	59 (26.46)	405 (30.13)	
Reunified by study end	112 (50.22)	490 (36.46)	*p*<.001
Completed treatment	103 (46.19)	461 (34.30)	*p*<.01

Weighted Descriptive Characteristics

Table 2 compares demographic, child welfare, treatment characteristics, and key study variables for FTC and TDC caregivers and children, after applying the IPW using survey-weighted analyses. Results demonstrate greater balance of covariate distribution between the groups. No comparisons remain significantly different except for treatment level of care recommendation (*p*<.001) and lifetime treatment admissions (*p*<.05). FTC participants remained less likely to be referred to crisis or detox,

more likely to have 4-5 treatment admissions, and less likely to have 6 or more admissions. While FTC participants were no longer significantly more likely to complete treatment, they remained significantly more likely to achieve family reunification (49.1% vs. 36.5%; *p*<.01).

Table 2 also presents the results of covariate balance diagnostics between FTC and TDC participants after weighting. As shown, the SMDs for all covariates were well below Rubin's (2001) threshold of 0.25, with absolute values ranging from 0.01 to a maximum of 0.11. Most covariates also met Nguyen et al.'s (2017) more conservative threshold of 0.10; only one covariate slightly exceeded this criterion: total number of allegations (absolute SMD=0.10). Similarly, VRs for all covariates fell within the acceptable range of 0.5 to 2, with values ranging from 0.66 to 1.16.

Table 2: Weighted Descriptive Characteristics (N=1,567)

Variable	FTC Participants (N=200) Mean or Weighted Percentage (Linearized SE)	TDC Participants (N=1,367) Mean or Weighted Percentage (Linearized SE)	SMD	VR	Significance
Child age at removal	3.70 (0.34)	4.09 (0.11)	-0.09	0.86	N.S.
Caregiver age	32.16 (0.89)	32.28 (0.19)	-0.02	1.05	N.S.
Caregiver race			-0.03	0.93	N.S.
White	60.98 (4.78)	60.94 (1.33)			
Black	24.79 (4.76)	22.63 (1.14)			
Native	14.24 (3.31)	16.43 (0.99)			
Caregiver female sex	67.81 (4.85)	70.10 (1.24)	-0.05	1.05	N.S.
Total # of allegations			-0.11	0.79	N.S.
1	63.75 (4.11)	59.89 (1.34)			
2	24.89 (3.64)	26.22 (1.20)			
3	9.85 (2.26)	9.81 (0.80)			
4 or more	1.51 (1.06)	4.08 (0.52)			
Total # of placements			-0.04	0.84	N.S.
1-2	32.89 (4.33)	36.06 (1.32)			
3-4	46.11 (4.50)	36.78 (1.33)			
5 or more	21.00 (3.29)	27.17 (1.22)			
Any kinship placements	46.28 (4.36)	51.17 (1.37)	-0.10	1.00	N.S.
Treatment level of care recommendation			-0.01	1.16	*p*<.001
Community living or residential treatment	26.60 (3.21)	21.43 (1.19)			
Crisis or detox	3.56 (1.17)	12.96 (0.92)			
Outpatient	69.85 (3.43)	65.60 (1.33)			
Primary substance of choice					
Alcohol	23.61 (5.27)	19.96 (1.06)	0.09	1.13	N.S.
Cannabis	14.98 (2.87)	15.75 (0.99)	-0.02	0.96	N.S.
Opioid	14.70 (2.75)	12.49 (0.94)	0.07	1.15	N.S.
Stimulant	45.06 (4.27)	49.25 (1.38)	-0.08	0.99	N.S.

ASI domain scores					
Drug	6.39 (0.17)	6.36 (0.06)	0.01	0.68	N.S.
Employment	4.85 (0.19)	4.74 (0.08)	0.04	0.66	N.S.
Family	6.01 (0.17)	5.81 (0.07)	0.08	0.72	N.S.
Psychiatric	5.21 (0.19)	5.11 (0.08)	0.04	0.78	N.S.
# Lifetime treatment admissions			0.02	0.81	*p<.05*
1	12.01 (2.86)	16.04 (1.00)			
2-3	30.39 (4.32)	30.69 (1.28)			
4-5	33.58 (4.34)	22.75 (1.14)			
6 or more	24.02 (3.42)	30.51 (1.27)			
Reunified by study end	49.05 (4.49)	36.53 (1.33)			*p<.01*
Completed treatment	42.38 (4.42)	34.57 (1.31)			N.S.

Note: SMD=standardized mean differences; VR= variance ratio of weighted variance. According to Rubin (2001), SMD<0.25 and a VR between 0.5 and 2 indicate acceptable balance. According to Nguyen et al. (2017), variables with SMD>0.1 should be included as controls to address residual imbalance.

Weighted Logistic Regression Results

Following the recommendation of Nguyen et al. (2017), we included any covariate with an absolute SMD greater than 0.10 as a control in the weighted regression models. Based on the balance diagnostics in Table 2, this applied to only one variable, total number of allegations.

Table 3 provides the results of the weighted logistic regression model. The findings indicate that FTC participants are 1.66 times more likely to achieve reunification when compared to TDC participants (*OR*=1.66, *p*<.01).

Table 3: Results of the weighted logistic regression examining the likelihood of family reunification (N=1,567)

Variable	OR	SE	95% CI	p
FTC	1.66	0.31	1.14 – 2.40	*p<.01*
Total # of allegations (ref.=1)				
2	1.13	0.14	0.88 – 1.45	N.S.
3	0.92	0.17	0.64 – 1.32	N.S.
4 or more	0.57	0.18	0.31 – 1.04	N.S.

Note: Interpret odds ratios as follows: (OR-1)*100 = % change in outcome, e.g., (1.66-1)*100 = 66% increase in likelihood of reunification among FTC group, controlling for covariates.

Weighted Mediation Model

Figures 2 and 3 illustrate the results of the weighted mediation model evaluating whether treatment completion mediates the relationship between FTC participation and family reunification. Figure 2 provides the unstandardized coefficients and Figure 3 presents odds ratios.

The model revealed significant direct effects on reunification for both FTC participation (β=0.10; *OR*=1.53; *p*<.05) and treatment completion (β=1.04; *OR*=2.82; *p*<.001). This indicates that FTC participants were 1.53 times more likely to achieve reunification compared to TDC participants, independent of treatment completion. Furthermore, treatment completers were 2.82 times more likely to achieve reunification compared to TDC participants, independent of FTC participation. FTC participation was not found to have a significant direct effect on treatment success (β=0.34; *OR*=1.41), nor was it found to have a significant indirect effect on reunification (β=0.02; *OR*=1.09). Overall, the results of the mediation model do not provide evidence that treatment completion mediates the relationship between FTC participation and reunification.

Figure 2: Results of mediation model with unstandardized coefficients and standard errors (N=1,567)
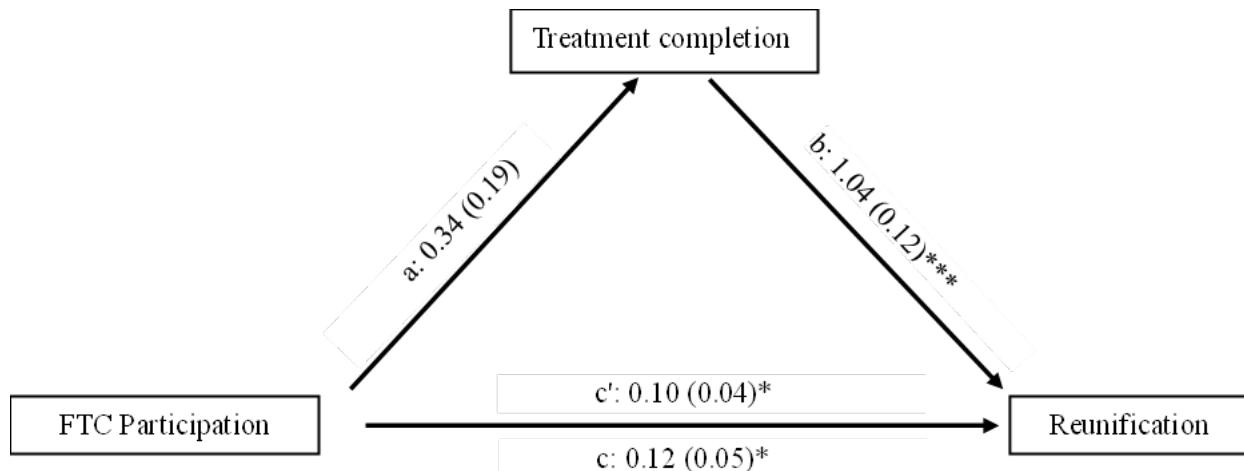
Note: Presented coefficients are unstandardized with standard errors in parentheses. A is the direct effect of FTC participation on treatment completion, b is the direct effect of treatment completion on reunification, c' is the direct effect of FTC participation on reunification, and c is the total effect of FTC participation on reunification. To aid in interpretation, exponentiated coefficients (odds ratios) are presented in figure 3.
* p<.05 *** p<.001

Figure 3: Results of mediation model with exponentiated coefficients (odds ratios) (N=1,567)
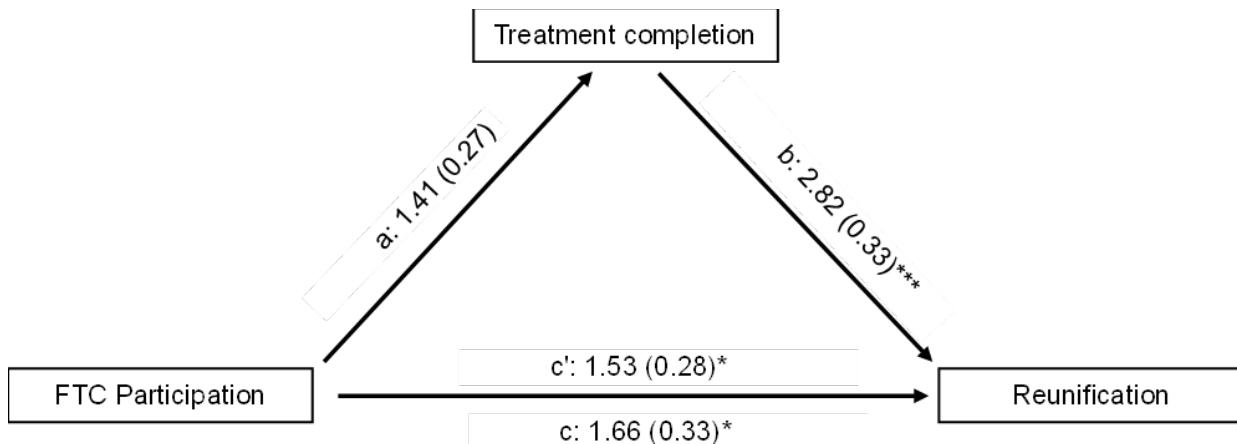
Note: Presented coefficients are exponentiated (odds ratios) with standard errors in parentheses. A is the direct effect of FTC participation on treatment completion, b is the direct effect of treatment completion on reunification, c' is the direct effect of FTC participation on reunification, and c is the total effect of FTC participation on reunification. Interpret as follows: (OR-1)*100 = % change in outcome, e.g., (1.66-1)*100 = 66% increase in likelihood of reunification among FTC group, controlling for treatment completion status.
* p<.05 *** p<.001

Limitations

Our study is limited in several ways. First, the sample was restricted to one state and only five FTCs, limiting generalizability to other regions and jurisdictions. Second, while robust, the research design cannot match the rigor of a randomized controlled trial, and estimates should not be interpreted as causal. Although IPW improved the balance between the FTC and TDC participants groups, unobserved confounders may still bias the results. While our treatment dataset included two trauma-related measures for caregivers, both were excluded due to substantial missing data (49% for trauma assessment scores and 20% for Adverse Childhood Experiences (ACE) scores). Prior research has shown that trauma can influence both substance use treatment engagement and child welfare outcomes ([Clark et al., 2020](#); [Kisiel et al., 2009](#); [Thompson, 2018](#)). Treatment readiness and criminogenic risk, understudied yet important concepts in child welfare and FTC research, may also contribute to FTC participation and reunification but were not measured in the current study. Additionally, our treatment data only included 'treatment episodes,' and did not include all types of substance use therapies, such as medications for opioid use disorder. We were likewise unable to control for access to other concurrent supports and services.

Probabilistic matching procedures were used to link perpetrator records and treatment records, which may have introduced measurement error due to potential mismatched or unlinked records. Additionally, because reunification was measured as a child-level outcome rather than a dyad-level outcome in the available administrative data, it was not always possible to identify which specific caregiver was the subject of reunification. For children with both FTC and non-FTC caregivers, both dyads would be classified as achieving reunification if the child reunified, even if reunification occurred with only one caregiver. Similarly, we could not confirm whether the reunification occurred with the caregiver represented in the linked treatment data.

Finally, the study period coincided with the COVID-19 pandemic, which significantly disrupted maltreatment reporting, court operations, and treatment delivery, potentially affecting the findings.

Conclusion

By applying cross-system data linkage and IPW to address selection bias and unmeasured caregiver-level factors that have limited earlier research, these findings add to the growing body of robust evidence supporting this intervention. Our rigorous quasi-experiment found that FTC-involved families were over two-thirds more likely to reunify compared to matched controls (*OR*=1.66). Additionally, we found that substance use treatment completion did not mediate the path to reunification, suggesting that the FTC program exerts an independent treatment effect unrelated to the effect of substance use treatment.

The fact that families were found to achieve reunification without successfully completing treatment may point to the benefit of integrating harm reduction principles in FTC settings. Our finding that White caregivers were disproportionately over-represented in the pre-weighted treatment group points to ongoing concerns related to equitable and equivalent referrals and entries into these programs. For FTCs to truly meet the needs of this population, programs must continue to strengthen best practice standards implementation, including ongoing fidelity monitoring.

Sub-Study 2: Assessing Addiction Severity as a Moderator of Family Treatment Court Outcomes

Study Authors

Jon Phillips, PhD; Jessica Becker, MSW; Margaret Lloyd Sieger, PhD

Study Purpose

Families involved with the child welfare system due to parental substance use disorder often face challenges across multiple areas of life, including mental health, employment, housing, and legal involvement. These problems are frequently driven or exacerbated by addiction. While previous research, including findings from Sub-Study 1, has established that FTC participation is associated with higher rates of family reunification, less is known about whether the effectiveness of FTCs varies depending on the severity of a caregiver's addiction.

This study examines whether the impact of FTC participation on the likelihood of family reunification is moderated by the severity of a caregiver's substance use addiction. We also examine two secondary outcomes to provide additional context about children's experiences following removal: *permanency* (any legal permanency, including reunification, adoption, or guardianship) and *alternative permanency* (adoption or guardianship only). These terms are used throughout the remainder of the study. Findings help clarify which families benefit from FTC programs, and whether some families benefit more than others when compared to the traditional child welfare court process (i.e., traditional deprived courts, or TDCs). Results may also inform the potential utility of a severity measure to support FTC screening and referral decisions by helping identify families most likely to benefit from the intensive support offered through FTC program.

Research Questions

This study had two research questions:

(1) How does the effect of FTC participation on likelihood of reunification differ by caregiver addiction severity, and which severity group or groups experience the greatest benefit from FTC participation?
(2) How does the effect of FTC participation on permanency vary by caregiver addiction severity?
(3) How does the effect of FTC participation on alternative permanency vary by caregiver addiction severity?

Methods

Design

This study used a quasi-experimental longitudinal design.

Data Sources

Data for this study were collected from six counties in a Southwestern state with operational FTCs from 2018 to 2022. Three administrative data sources were linked to generate a sample of dyadic records that include information on child demographic and case characteristics, caregiver demographic and substance use treatment characteristics, and FTC participation status. The data sources included: 1) Oklahoma Department of Human Services (DHS) foster care records, 2) Oklahoma Department of Mental Health and Substance Abuse Services (DMHSAS) treatment admission records, and 3) a list of FTC participants provided by Oklahoma DMHSAS, which included their FTC enrollment and discharge dates.

Timeframe

This study examined child welfare cases with removal dates that occurred between January 1, 2018 and June 1, 2022. The DHS extract includes records with removal dates as early as April 30, 2015, and return dates through May 25, 2023. The treatment admission data include adults admitted on or after July 7, 2016, and discharged on or before March 28, 2024. FTC participation records include admission dates beginning December 11, 2013, and discharge dates through April 10, 2023.

Participants

Perpetrator records and substance use treatment records were linked using probabilistic matching procedures described in Sub-Study 1. A total of 831 adult records with a non-missing perpetrator ID were successfully linked. These records were then deterministically matched to 9,007 child foster care records using a perpetrator ID variable, resulting in 2,022 perpetrator-child dyads. We excluded 449 dyads for whom no treatment data were retained based on the treatment episode selection criteria described in Sub-Study 1.

Forty-one observations were excluded that had removal dates prior to the study start date (January 1, 2018). Two observations were excluded due to missing perpetrator race information, and four were excluded due to missing Addiction Severity Index (ASI) scores because the caregivers were under the age of 18 and had instead been assessed using the Teen ASI. The ASI is a clinical tool used to assess problem severity in seven key areas of life for individuals with substance use disorders.

From the remaining 1,526 observations, propensity score matching (PSM) was used to generate a matched analytic sample of 521 observations, comprising 217 in the FTC group (183 children and 113 caregivers) and 304 in the TDC group (295 children and 242 caregivers). Each observation (i.e., dyad) represents a unique perpetrator-child-referral number combination, where the referral number corresponds to a maltreatment report and subsequent foster care placement.

Some children and perpetrators appear in multiple dyads, which were retained because each dyad reflects a distinct treatment episode (with its own treatment completion status) and a separate foster care episode (with its own permanency outcome). Two records reflected a single perpetrator-child duplicate. Both records were retained because each was associated with a different referral

number, indicating separate cases that occurred at different timepoints and may have differed in context, treatment needs, and the ages of caregiver and child.

Measures

Dependent Variables

Three binary dependent variables were used in the analysis: reunification, permanency, and alternative permanency. Each was derived from the placement exit reason variable in the foster care administrative data, which was originally recorded in eight categories.

Reunification was the primary outcome of interest. The variable was coded as "1" if the exit reason was "Reunification" or "Reunified with Custodial Parent," and "0" for all other exit reasons: "Adoption," "Child Aged Out/Emancipation," "Custody to Relative," "Guardianship – Non-Relative," "Guardianship – Relative," and "Tribal Jurisdiction."

Permanency, used as a secondary outcome, represents any legal permanency, including reunification, adoption, and guardianship. It was coded as "1" if the exit reason was "Reunification," "Reunified with Custodial Parent," "Adoption," "Guardianship – Non-Relative," or "Guardianship – Relative," and "0" for all other exit reasons.

Alternative permanency, also a secondary outcome, refers to adoption or guardianship. It was coded as "1" if the exit reason indicated "Adoption," "Guardianship – Non-Relative," "Guardianship – Relative," or "Custody to Relative," and "0" for all other exit reasons. These categories were combined into a single outcome because they all represent forms of legal permanency outside of reunification.

Moderator Variables

To examine whether addiction severity moderates the relationship between court type and study outcomes, we created three variables that classified participants based on their court type (FTC vs. TDC) and addiction severity level (low, moderate, high). FTC participation was determined using the list of FTC participants, with participants coded as "1" if enrolled during the study period, and "0" otherwise (TDC participants).

Each variable represents a specific combination of court type and addiction severity, allowing us to directly compare FTC to TDC within each severity level:
- FTC/low (1=FTC participant with low addiction severity; 0= TDC participant with low addiction severity)
- FTC/moderate (1=FTC participant with moderate addiction severity; 0=TDC participant with moderate addiction severity)
- FTC/high (1=FTC participant with high addiction severity; 0= TDC participant with high addiction severity)

Addiction severity levels were derived from treatment admission records, which included continuous ASI scores across seven domains: alcohol, drug, employment, family, legal, medical, and psychiatric. These scores, documented during the clinical assessment process at treatment admission, ranged from 0-9, with higher scores indicating greater severity and need for treatment. According to Carise (2007, p. 8), ASI scores are interpreted as follows:

- 0-1: No real problem, treatment not indicated
- 2-3: Slight problem, treatment probably not necessary
- 4-5: Moderate problem, some treatment indicated
- 6-7: Considerable problem, treatment necessary
- 8-9: Extreme problem, treatment absolutely necessary

A threshold of "6 or above" was used to identify clinically significant domains, as it indicates a considerable problem requiring treatment. We then summed the number of domains where each participant scored at or above this threshold. The resulting distribution was examined and divided into tertiles, creating three relatively equal groups. This approach yielded the following addiction severity categories:

- Low Severity: Scored 6 or above on 0-3 ASI domains
- Moderate Severity: Scored 6 or above on 4-5 ASI domains
- High Severity: Scored 6 or above on 6-7 ASI domains

This classification structure allows for comparison of outcomes between FTC and TDC participants while controlling for addiction severity level, enabling an assessment of whether FTC participation is associated with different outcomes across varying levels of treatment need.

Time Variables

For children who achieved permanency, time to permanency was calculated as the number of days between their removal date and their foster care exit date (both obtained from foster care records). Children who did not exit foster care during the study period were treated as censored cases, with their time calculated as the number of days between their removal date and May 26, 2023 (one day after the latest observed return date in the dataset).

Control Variables

For consistency, the same variables used to determine the propensity scores and inverse probability weights in Sub-Study 1 were used to determine the propensity scores in this study (see *Sub-Study 1: Method—Measures—Control Variables* for full variable descriptions).

For the alternative permanency survival analyses, the lifetime treatment admissions variable was collapsed from four categories to three to address quasi-complete separation in the Cox regression models. Quasi-complete separation occurs when a predictor variable perfectly or near perfectly predicts the outcome for a subset of observations, leading to unstable parameter estimates (Allison, 2008). This adjustment was necessary because no children with caregivers who had only one

admission experienced alternative permanency, resulting in unstable estimates in the Cox models, similar to the complete separation problem.

Procedures

Stata Version 18.0 was used to clean and analyze the data.

Propensity Score Matching (PSM)

PSM was performed twice using Stata's "psmatch2" command. This command estimated propensity scores, representing the probability of enrollment in an FTC (FTC=1), using a logistic regression model where the indicator of FTC participation was regressed on the following covariates: child age, adult age, adult race, adult sex, total number of allegations in the maltreatment report, number of placements during the removal, any kinship placement, level of care recommendation, indicators for primary substance of choice (alcohol, marijuana, opioid, stimulant), ASI scores across drug, employment, family, and psychiatric domains, and number of lifetime treatment admissions.

Following the recommendations for the psmatch2 package, cases were seeded and randomly sorted prior to each PSM attempt to ensure that the order of variables did not influence results and to enable replicability (Leuven & Sianesi, 2003). In the first round of matching, no caliper was applied, allowing FTC participants to be matched with the closest comparison cases regardless of the distance between their propensity scores. In the second round, a caliper size of one-quarter of the standard deviation of the estimated propensity scores was applied (Guo & Fraser, 2010), limiting matches to control participants with similar scores. Nearest neighbor matching (1:2) with replacement was used, matching each FTC participant to two closest comparison cases that fell within the specified caliper.

Of 218 FTC cases, 217 matched to at least one comparison case, yielding a matched sample of 521 observations, comprising 217 in the FTC group and 304 in the TDC group. After finalizing the matching approach, bivariate analyses were repeated to assess covariate balance. Two covariates, treatment level of care recommendation and lifetime treatment admissions, remained significantly different between groups and were included as control variables in subsequent analyses.

Descriptive Characteristics

Demographic, child welfare, and treatment characteristics of FTC and TDC caregivers were compared using the full sample of 1,526 observations (FTC=218, TDC=1,308) and the matched sample (N=521, FTC=217, TDC=303) to assess differences between groups before and after matching. Chi-squared tests were used to assess differences in categorical variables, and one-way analysis of variance (ANOVA) was applied to compare group means for continuous variables.

Survival Analysis Models

Cox proportional hazards models with robust standard errors were used to estimate the effect of FTC participation on the study outcomes. This approach, which is a type of survival analysis, was chosen because it accounts for both time and censoring (i.e., cases in which the outcome had not been

achieved by the end of the observation period but still could occur later). This approach is preferred when not all participants experience the outcome of interest (e.g., reunification) during the study period.

The Cox models estimate hazard ratios (HRs), which represent the relative likelihood of an event occurring on any given day of observation, comparing two groups. In this study, an HR less than 1 indicates a lower likelihood of achieving the outcome of interest for FTC participants relative to TDC participants, while an HR greater than 1 indicates a higher likelihood. For example, when examining reunification, HRs greater than 1 reflect a greater likelihood of reunification on a given day for FTC participants. The percent change in hazard was approximated as 1-HR (for HRs<1) or HR-1 (for HRs>1).

To answer Research Question 1, we estimated separate models for the likelihood of reunification within each addiction severity group (low, moderate, and high). The magnitude of the hazard ratios was compared across groups to identify which group or groups experienced the greatest benefit of FTC participation. To answer Research Questions 2 and 3, we applied the same stratified approach to estimate the effect of FTC participation on likelihood of permanency and likelihood of alternative permanency.

For each research question, both unadjusted and adjusted models were run. The adjusted models controlled for level of care recommendation and number of lifetime treatment admissions.

Findings

Pre-match Descriptive Characteristics

Results revealed several significant differences between the unmatched groups. FTC participants were younger; caregivers averaged 29.98 years compared to 32.78 years ($p<.001$), and their children averaged 3.17 years versus 4.30 years ($p<.001$). FTC participants were more likely to be White (71.8% vs. 59.6%) and less likely to be Black (18.4% vs. 23.0% or American Indian/Alaskan Native or Native Hawaiian/other Pacific Islander (9.6% vs. 17.6%) ($p<.01$). There was no significant difference in sex between FTC and TDC participants.

FTC participants were more likely to have only one maltreatment allegation and less likely to have multiple reports compared to TDC participants ($p<.05$), while the number of placements during the removal did not differ significantly between groups. FTC children were more likely to be placed with family members (59.2% vs. 50.1%; $p<.05$).

Treatment recommendations also differed significantly ($p<.001$). FTC participants were less likely to be referred to crisis or detox (5.5% vs. 13.2%) or to outpatient services (54.4% vs. 67.7%), but more likely to be referred to community living, halfway houses, or residential care (39.9% vs. 19.2%). FTC participants were less likely to report alcohol (9.2% vs. 21.8%; $p<.001$) and more likely to report a stimulant (59.6% vs. 47.9%; $p<.01$) as their primary substance. There was no significant difference in the likelihood of reporting marijuana (13.8% vs. 15.6%) or an opioid (15.6% vs. 11.9%) as the primary substance.

FTC participants scored higher than TDC participants on the drug (*p*<.001) and family (*p*<.001) domains of the ASI, while there were no significant differences on the employment or psychiatric domains. FTC participants were more likely to have 4-5 lifetime treatment admissions (33.0% vs. 23.2%; *p*<.05), while other admission categories showed smaller differences between the groups.

Post-match Descriptive Characteristics

After matching, differences between FTC and TDC participants were largely reduced. However, treatment recommendations remained significantly different (*p*<.05). FTC participants were similarly likely to be referred to outpatient services, but less likely to be referred to crisis or detox (5.5% vs. 12.2%) and more likely to be referred to community living, halfway houses, or residential care (39.6% vs. 32.6%). Additionally, FTC participants were significantly more likely to have 4-5 lifetime treatment admissions and less likely to have 2-3 lifetime treatment admissions (*p*<.01).

Sample Characteristics

Table 4 shows the sample characteristics by court type. The average child age at removal was 3.3 years, and the average caregiver age was 30.0 years. Most caregivers identified as White (69.5%), followed by Black/African American (18.5%) and American Indian/Alaskan Native or Native Hawaiian/other Pacific Islander (11.9%). Three-quarters (75.1%) of caregivers were female. Addiction severity levels were similar across court types, with a majority categorized as low severity (55.7%), followed by moderate (24.0%), and high (20.4%).

Table 4: Sample Characteristics by Court Type (Matched Sample, N=521)

Variable		Total Sample	FTC	TDC	
		Mean (SD) or n (%)	Mean (SD) or n (%)	Mean (SD) or n (%)	Significance
Demographics	Child age at removal	3.30 (3.65)	3.18 (3.70)	3.38 (3.62)	N.S.
	Caregiver age	29.98 (5.54)	29.98 (5.54)	29.97 (5.55)	N.S.
	Caregiver race				N.S.
	White	362 (69.48)	156 (71.89)	206 (67.76)	
	Black/African American	97 (18.62)	40 (18.43)	57 (18.75)	
	American Indian/Alaskan Native or Native Hawaiian/other Pacific Islander	62 (11.90)	21 (9.68)	41 (13.49)	
	Caregiver sex				N.S.
	Male	130 (24.95)	54 (24.88)	76 (25.00)	
	Female	391 (75.05)	163 (75.12)	228 (75.00)	

Addiction Severity Levels	Low	290 (55.66)	118 (54.38)	172 (56.58)	N.S.
	Moderate	125 (23.99)	60 (27.65)	65 (21.38)	
	High	106 (20.35)	39 (17.97)	67 (22.04)	

Outcome by Court Type and Addiction Severity

Table 5 shows outcomes and demographics by court type and addiction severity level. Across all severity groups, FTC participants were more likely to achieve reunification, with the largest differences observed among the moderate and high severity groups. However, FTC participants achieved reunification faster than TDC participants only in the high severity group. Reunification was slower among FTC participants in the low and moderate severity groups.

Across all severity groups, FTC participants were more likely to achieve some type of permanency than TDC participants. The largest differences were observed among those with low and high addiction severity. FTC participants achieved permanency faster than TDC participants in all severity groups, with the biggest difference in the high severity group.

FTC participants were more likely to achieve alternative permanency (i.e., adoption or guardianship) than TDC participants only in the low severity group, and they were less likely in the moderate and high severity groups. FTC participants achieved alternative permanency faster than TDC participants in the low and moderate severity groups, but took longer in the high severity group.

Demographic characteristics were largely similar across groups, though a few patterns emerged. FTC participants had higher proportions of White caregivers than TDC participants in the low and high severity groups, but a lower proportion in the moderate group. Conversely, Black caregivers were more represented in FTCs in the moderate and high severity groups, but less represented in the low group. Caregivers identifying as American Indian, Alaskan Native, Native Hawaiian, or other Pacific Islander were more represented in FTCs at the moderate level, but less represented at the low and high levels. In TDCs, the proportion of female caregivers increased with addiction severity, while FTCs maintained a relatively stable gender distribution across severity levels. Child and caregiver ages showed minimal variation across court type and addiction severity level.

Table 5: Outcomes and Demographics by Court Type and Addiction Severity Level (Matched Sample)

Variable	FTC Low Severity (n=118)	TDC Low Severity (n=172)	FTC Moderate Severity (n=60)	TDC Moderate Severity (n=65)	FTC High Severity (n=39)	TDC High Severity (n=67)
	Mean (SD) or n (%)	Mean (SD) or n (%)	Mean (SD) or n (%)	Mean (SD) or n (%)	Mean (SD) or n (%)	Mean (SD) or n (%)

Outcomes	Reunified	47 (39.83)	59 (34.30)	40 (66.67)	25 (38.46)	21 (53.85)	22 (32.84)
	Time to Reunification (Days) if Achieved	641.38 (204.603)	631.36 (245.00)	670.98 (232.17)	505.28 (245.25)	603.33 (200.64)	808.09 (318.72)
	Achieved Permanency	99 (83.90)	123 (71.51)	48 (80.00)	51 (78.46)	32 (82.05)	48 (71.64)
	Time to Permanency (Days) if Achieved	732.88 (285.34)	769.35 (305.70)	691.19 (251.81)	711.24 (325.58)	719.09 (262.32)	785.73 (267.34)
	Achieved Alternative Permanency	53 (44.92)	64 (37.21)	8 (13.33)	26 (40.00)	11 (28.21)	27 (40.30)
	Time to Alternative Permanency (Days) if Achieved	803.15 (331.94)	896.56 (302.40)	792.25 (333.78)	909.27 (266.57)	940.09 (225.70)	755.22 (223.60)
Demographics	Child age at removal	3.24 (3.65)	3.25 (3.66)	3.22 (3.85)	3.72 (3.43)	2.93 (3.71)	3.38 (3.70)
	Caregiver Age	29.89 (5.44)	30.13 (5.79)	30.87 (6.30)	29.31 (4.68)	28.87 (4.41)	30.22 (5.70)
	Caregiver race						
	White	95 (80.51)	116 (67.44)	32 (53.33)	47 (72.31)	29 (74.36)	43 (64.18)
	Black/African American	12 (10.17)	32 (18.60)	21 (35.00)	15 (23.08)	7 (17.95)	10 (14.93)
	American Indian/Alaskan Native or Native Hawaiian/other Pacific Islander	11 (9.32)	24 (13.95)	7 (11.67)	3 (4.62)	3 (7.69)	14 (20.90)
	Caregiver sex						
	Male	30 (25.42)	59 (34.40)	16 (26.67)	12 (18.46)	8 (20.51)	5 (7.46)
	Female	88 (74.58)	113 (65.70)	44 (73.33)	53 (81.54)	31 (79.49)	62 (92.54)

Survival Analysis Results

Reunification (Research Question 1)

The survival curves for reunification (Figures 4-6) reveal differences between FTC and TDC cases in the likelihood of achieving reunification over time, stratified by addiction severity. These curves account for censored cases, representing children who had not yet reunified by the end of the observation window.

Among participants with low addiction severity (Figure 4), FTC cases reunified faster than TDC participants, with divergence beginning around 500 days after removal and continuing to widen. By the end of the observation period, approximately 30% of FTC cases had not reunified compared to 50% of TDC cases. Among those with moderate addiction severity (Figure 5), a more pronounced difference emerged in favor of FTC participants beginning around 500 days. By approximately 1,400 days, very few FTC cases had not reunified while about 50% of TDC cases had not reunified. Among high addiction severity cases (Figure 6), FTC participants consistently achieved reunification faster than TDC participants through approximately the first 1,400 days, with the advantage emerging around 400 days. However, the curves converged near the end of the observation period, with both groups showing similar proportions of cases not reunified (approximately 30-40%).

Table 6 confirms these visual observations. Among those with low addiction severity, FTC participants were 44% more likely to reunify compared to TDC participants (adjusted HR=1.44), although this difference was not statistically significant. Among those with moderate addiction severity, FTC participants were 120% more likely to reunify compared to TDC participants (adjusted HR=2.20; $p<.01$). Among those with high addiction severity, FTC participants were 124% more likely to reunify compared to TDC participants (adjusted HR=2.24; $p<.05$).

Figure 4: Survival Curve of Reunification for Caregivers with Low Addiction Severity (0-2 ASI Domains=6+) (n=290)
Matched Sample: FTC (n=118), TDC (n=172)
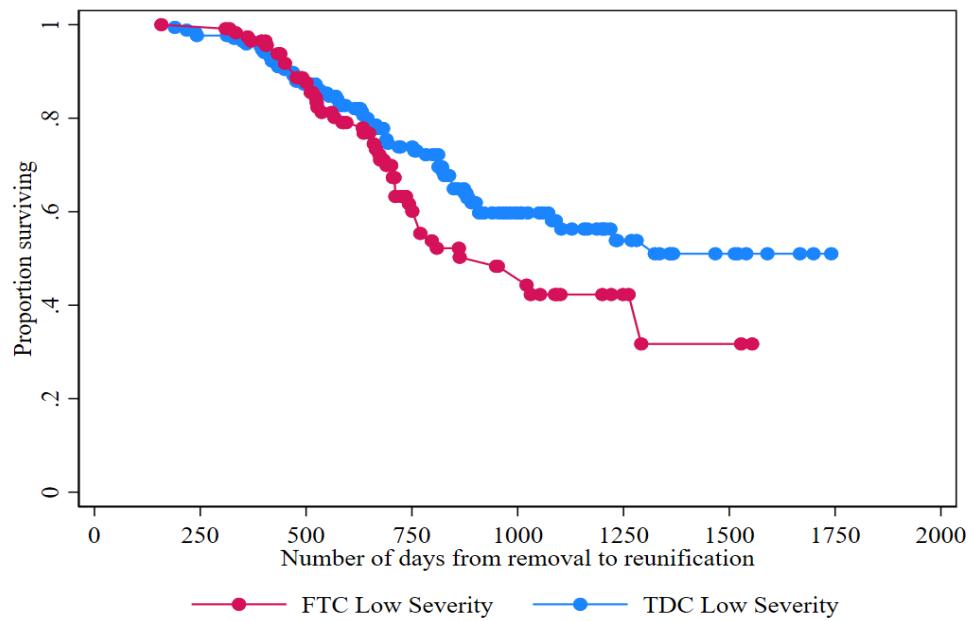
Figure 5: Survival Curve of Reunification for Caregivers with Moderate Addiction Severity (3-4 ASI Domains=6+) (n=125)
Matched Sample: FTC (n=60), TDC (n=65)
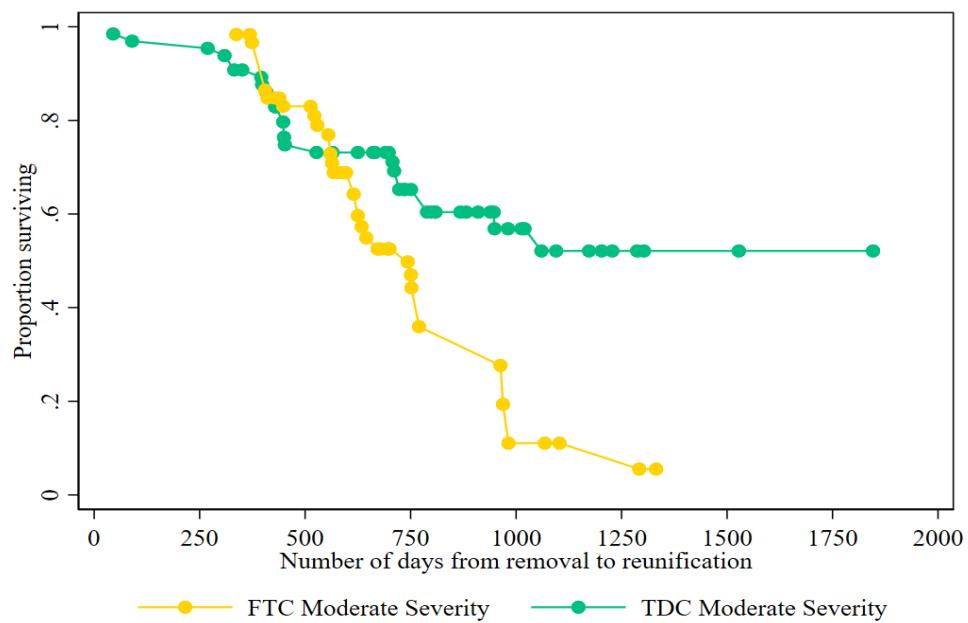
Figure 6: Survival Curve of Reunification for Caregivers with High Addiction Severity (5-7ASI Domains=6+) (n=106)
Matched Sample: FTC (n=39), TDC (n=67)
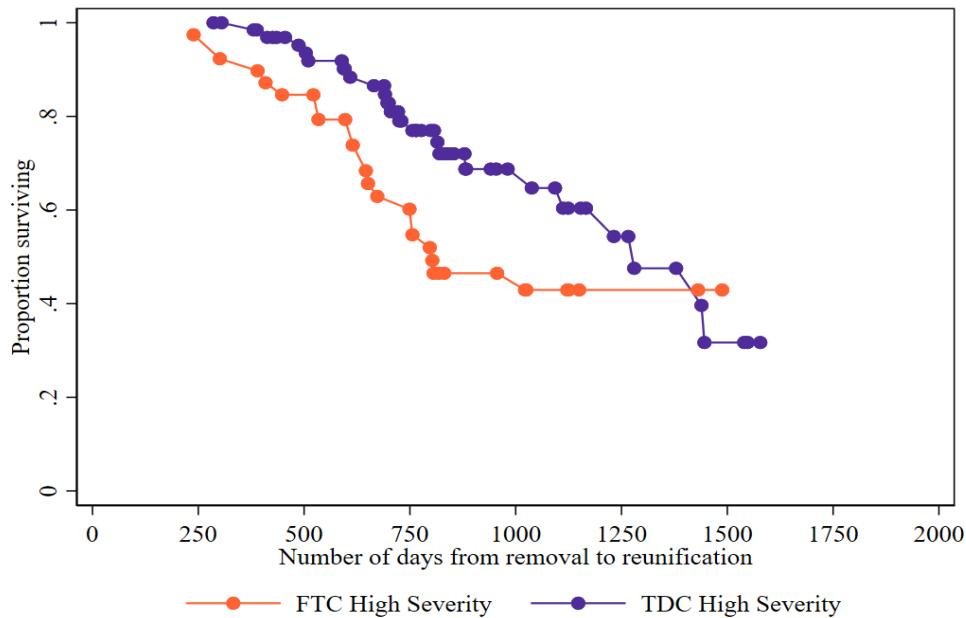
Table 6: Proportional Hazards Model for Reunification: FTC vs. TDC Across Addiction Severity Levels (Matched Sample, N=521)

Variable	Unadjusted Hazard Ratio (Robust SE)	Unadjusted Hazard Ratio (Robust SE)	Unadjusted Hazard Ratio (Robust SE)	Adjusted Hazard Ratio (Robust SE)	Adjusted Hazard Ratio (Robust SE)	Adjusted Hazard Ratio (Robust SE)
Low-Severity FTC (ref.=Low-Severity TDC)	1.45 (0.28)			1.44 (0.29)		
Moderate-Severity FTC (ref.=Moderate-Severity TDC)		2.15** (0.56)			2.20** (0.60)	
High-Severity FTC (ref.=High-Severity TDC)			1.69 (0.52)			2.24* (0.71)
Treatment level of care						

recommendation (ref.=outpatient)			
Community living or residential treatment	0.56** (0.12)	0.74 (0.19)	0.32** (0.12)
Crisis or detox	0.25 (0.28)		0.34* (0.15)
# Lifetime treatment admissions (ref=1)			
2-3	1.29 (0.39)	0.85 (0.44)	0.42 (0.54)
4-5	1.16 (0.35)	0.82 (0.46)	0.30 (0.39)
6 or more	0.84 (0.29)	0.88 (0.47)	0.34 (0.42)

Note: Models are stratified by addiction severity level. Group sizes: FTC Low (n=118), FTC Moderate (n=60), FTC High (n=39), TDC Low (n=172), TDC Moderate (n=65), TDC High (n=67). Adjusted models control for treatment level of care and number of lifetime treatment admissions.
* *p*<.05; ** *p*<.01; *** *p*<.001

Permanency (Research Question 2)

The survival curves for permanency (Figures 7-9) reveal differences between FTC and TDC cases in the likelihood of achieving any legal permanency outcome (including reunification, adoption, or guardianship) over time across addiction severity levels. These curves account for censored cases, representing children who had not yet achieved permanency by the end of the observation window.

Among participants with low addiction severity (Figure 7), FTC cases achieved permanency faster than TDC cases throughout the observation period, with divergence beginning around 500 days after removal and continuing to widen. By approximately 1,500 days, virtually no FTC cases had not achieved permanency, whereas a portion of TDC cases had not achieved permanency. A similar pattern appears among those with moderate addiction severity (Figure 8). Group differences emerge around day 500, and by approximately 1,250 days, no FTC participants had not achieved permanency. In contrast, a portion of TDC cases had not achieved permanency even after 1,750 days. Among high addiction severity cases (Figure 9), the curves show a smaller but still visible separation, with FTC participants achieving permanency slightly faster than TDC participants. By the end of the observation period, both groups showed similar proportions of cases not achieving permanency (approximately 15%).

Table 7 confirms these visual observations. Among those with low addiction severity, FTC participants were 55% more likely to achieve permanency compared to their TDC counterparts (adjusted HR=1.55; *p*<.01). Among those with moderate addiction severity, FTC participants were 52% more likely to achieve permanency compared to TDC participants (adjusted HR=1.52; *p*<.05). Among those with high addiction severity, FTC participants were 32% more likely to achieve permanency compared to TDC

participants (HR=1.32), although this difference was not statistically significant, consistent with the limited separation observed in the survival curves.

Figure 7: Survival Curve of Permanency for Caregivers with Low Addiction Severity (0-2 ASI Domains=6+) (n=290)
Matched Sample: FTC (n=118), TDC (n=172)
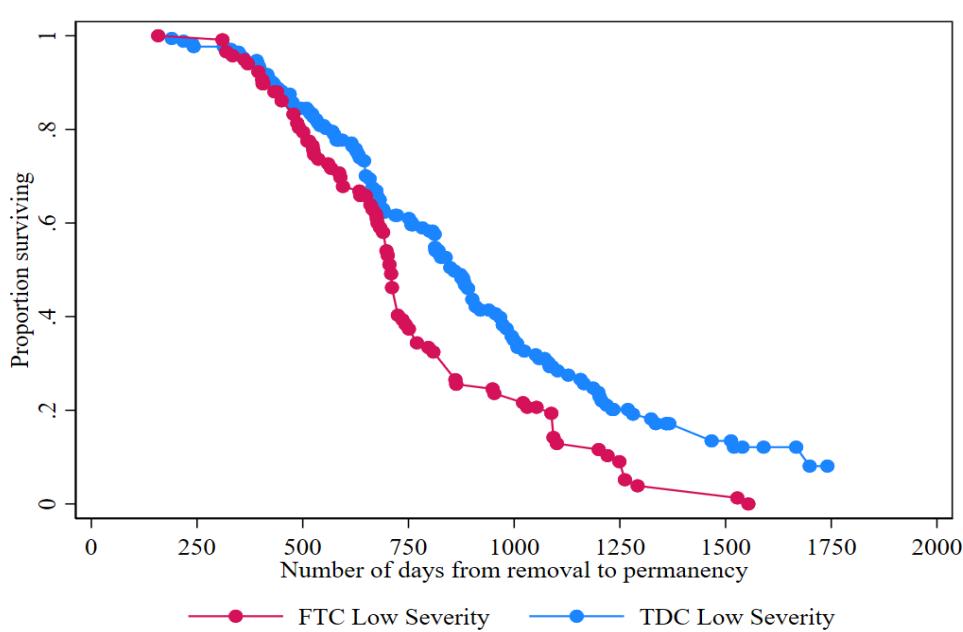
Figure 8: Survival Curve of Permanency for Caregivers with Moderate Addiction Severity (3-4 ASI Domains=6+) (n=125)
Matched Sample: FTC (n=60), TDC (n=65)
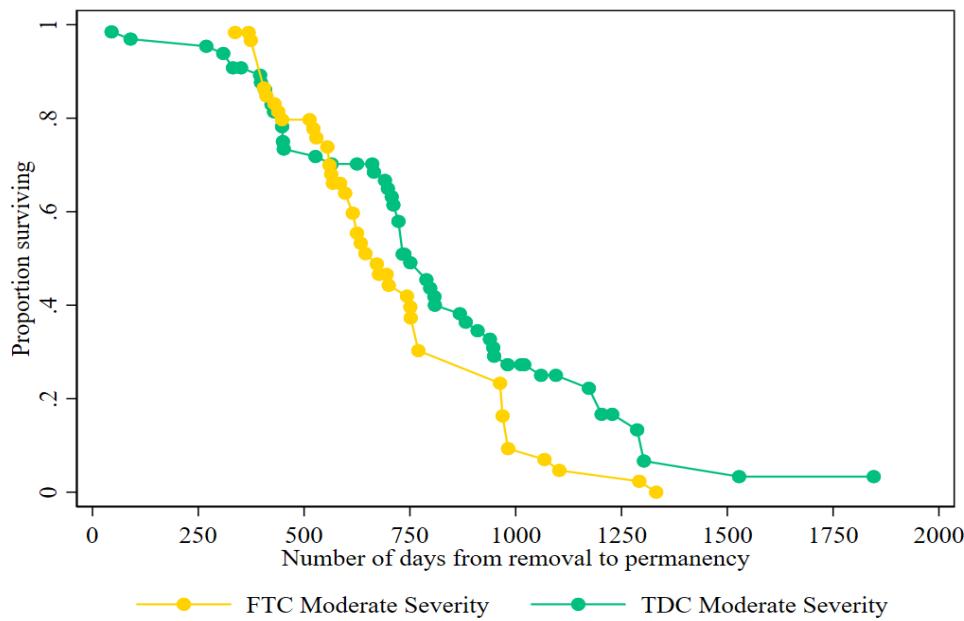
Figure 9: Survival Curve of Permanency for Caregivers with High Addiction Severity (5-7 ASI Domains=6+) (n=106)
*Matched Sample: FTC (n=39), TDC (n=67)*
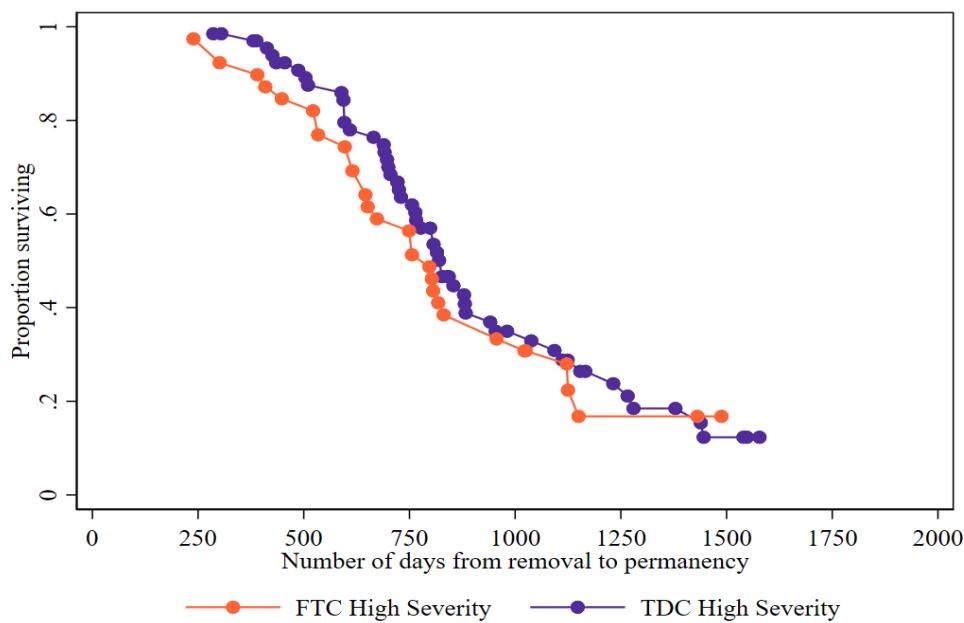
Table 7: Proportional Hazards Model for Permanency: FTC vs. TDC Across Addiction Severity Levels (Matched Sample, N=521)

Variable	Unadjusted Hazard Ratio (Robust SE)	Unadjusted Hazard Ratio (Robust SE)	Unadjusted Hazard Ratio (Robust SE)	Adjusted Hazard Ratio (Robust SE)	Adjusted Hazard Ratio (Robust SE)	Adjusted Hazard Ratio (Robust SE)
Low- Severity FTC (ref.=Low- Severity TDC)	1.58*** (0.21)			1.55** (0.21)		
Moderate- Severity FTC (ref.=Moderate- Severity TDC)		1.45 (0.29)			1.52* (0.31)	
High- Severity FTC (ref.=High- Severity TDC)			1.14 (0.27)			1.32 (0.29)
Treatment level of care recommendation (ref.=outpatient)						
Community living or residential treatment				0.86 (0.12)	0.79 (0.17)	0.63 (0.24)
Crisis or detox				0.51 (0.22)		0.84 (0.33)
# Lifetime treatment admissions (ref=1)						
2-3				1.21 (0.27)	1.20 (0.61)	1.29 (1.83)
4-5				1.13 (0.24)	1.01 (0.54)	0.75 (1.06)
6 or more				1.15 (0.27)	1.08 (0.56)	0.73 (1.02)

Note: Models are stratified by addiction severity level. Group sizes: FTC Low (n=118), FTC Moderate (n=60), FTC High (n=39), TDC Low (n=172), TDC Moderate (n=65), TDC High (n=67). Adjusted models control for treatment level of care and number of lifetime treatment admissions.
* $p<.05$; ** $p<.01$; *** $p<.001$

Alternative Permanency (Research Question 3)

The survival curves for alternative permanency (i.e., adoption of guardianship) (Figures 10-12) reveal differences between FTC and TDC cases in the likelihood of achieving this outcome over time across addiction severity levels. These curves account for censored cases, representing children who had not yet achieved alternative permanency by the end of the observation window.

Among participants with low addiction severity (Figure 10), FTC cases achieved alternative permanency faster than TDC participants**,** with divergence beginning around 750 days after removal and widening after 1,200 days. By approximately 1,600 days, virtually no FTC cases had not achieved this outcome, while almost 20% of TDC cases had not achieved this outcome by the end of the observation period. Among those with moderate addiction severity (Figure 11), TDC cases initially achieved alternative permanency faster between 750-1,100 days. After that point, both groups showed similar rates. By approximately 1,300 days, no FTC cases had not achieved alternative permanency, while about 5% of TDC cases had not achieved this outcome even after 1,750 days. Among high addiction severity cases (Figure 12), TDC participants achieved alternative permanency faster, but FTC cases caught up around 1,125 days. By 1,500 days, both groups had similar proportions of cases not achieving alternative permanency (approximately 40%).

Table 8 confirms these visual observations. Among those with low addiction severity, FTC participants were 72% more likely to achieve alternative permanency compared to TDC participants (adjusted HR=1.72; *p*<.01). Among those with moderate addiction severity, FTC participants were 38% less likely to achieve adoption/guardianship compared to TDC participants (adjusted HR=0.62), although this difference was not statistically significant. Among those with high addiction severity, FTC participants were 29% less likely to achieve alternative permanency (adjusted HR=0.71), again a non-significant difference.

Figure 10: Survival Curve of Adoption/Guardianship for Caregivers with Low Addiction Severity (0-2 ASI Domains=6+) (n=290)
Matched Sample: FTC (n=118), TDC (n=172)
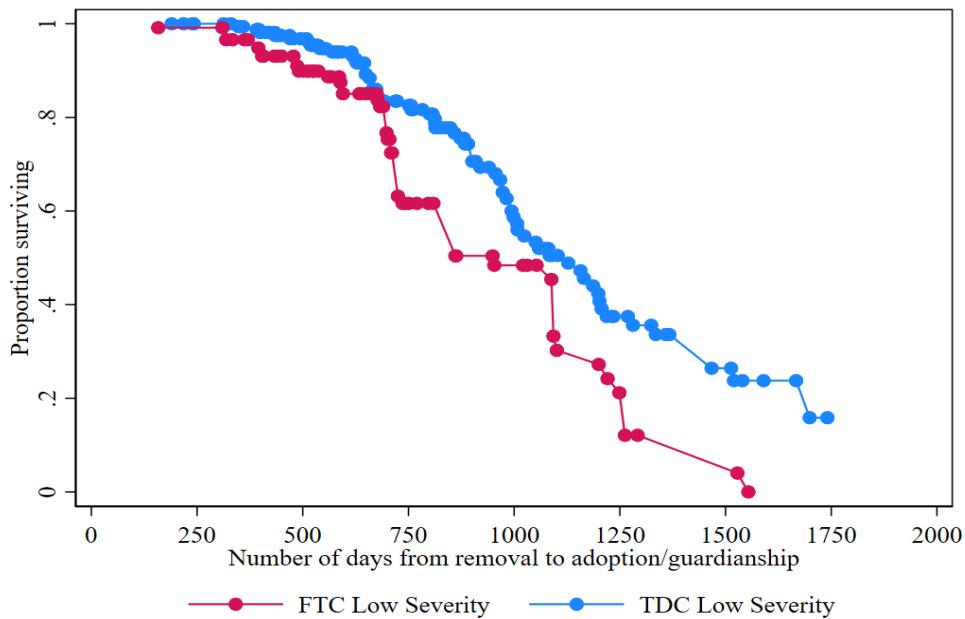
Figure 11: Survival Curve of Adoption/Guardianship for Caregivers with Moderate Addiction Severity (3-4 ASI Domains=6+) (n=125)
Matched Sample: FTC (n=60), TDC (n=65)
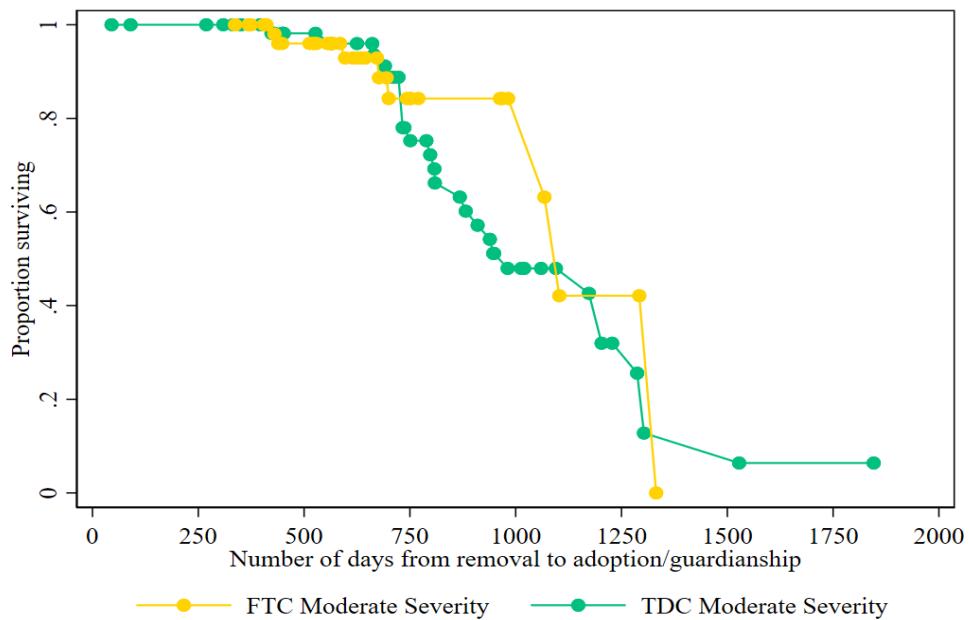
Figure 12: Survival Curve of Adoption/Guardianship for Caregivers with High Addiction Severity (5-7ASI Domains=6+) (n=106)
Matched Sample: FTC (n=39), TDC (n=67)
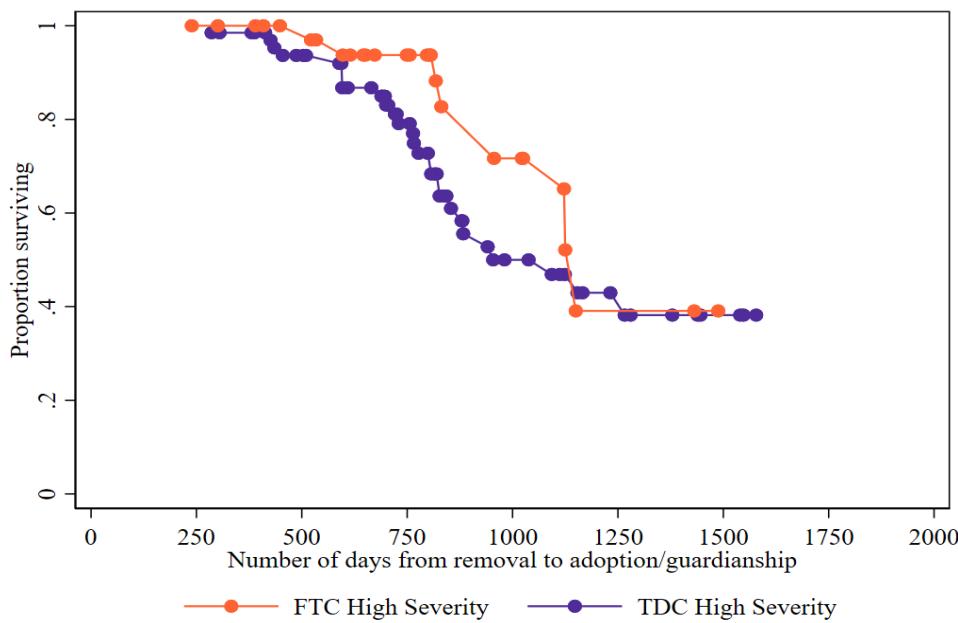
Table 8: Proportional Hazards Model for Adoption/Guardianship: FTC vs. TDC Across Addiction Severity Levels (Matched Sample, N=521)

Variable	Unadjusted Hazard Ratio (Robust SE)	Unadjusted Hazard Ratio (Robust SE)	Unadjusted Hazard Ratio (Robust SE)	Adjusted Hazard Ratio (Robust SE)	Adjusted Hazard Ratio (Robust SE)	Adjusted Hazard Ratio (Robust SE)
Low- Severity FTC (ref.=Low-Severity TDC)	1.75** (0.31)			1.72** (0.31)		
Moderate-Severity FTC (ref.=Moderate-Severity TDC)		0.60 (0.23)			0.62 (0.25)	
High- Severity FTC (ref.=High-Severity TDC)			0.66 (0.22)			0.71 (0.23)
Treatment level of care recommendation (ref.=outpatient)						
Community living or				1.22 (0.23)	0.89 (0.33)	2.54 (2.26)

residential treatment			
Crisis or detox	0.71 (0.34)		3.69 (3.46)
# Lifetime treatment admissions (ref=1-3)			
4-5	1.01 (0.21)	0.70 (0.27)	0.49 (0.21)
6 or more	1.34 (0.32)	0.84 (0.33)	0.43 (0.17)

Note: Models are stratified by addiction severity level. Group sizes: FTC Low (n=118), FTC Moderate (n=60), FTC High (n=39), TDC Low (n=172), TDC Moderate (n=65), TDC High (n=67). Adjusted models control for treatment level of care and number of lifetime treatment admissions.
* $p<.05$; ** $p<.01$; *** $p<.001$

Differential Impact of FTC by Outcome and Addiction Severity Level

Figure 13 displays the adjusted hazard ratios from the Cox regression models for each study outcome by addiction severity level, controlling for treatment level of care and lifetime treatment admissions. By visually comparing the magnitudes of the hazard ratios, the figure highlights the variation in likelihood of achieving various outcomes by addiction severity level (Research Questions 1-3) and which group or groups benefit most from FTC participation in terms of increased likelihood of reunification (Research Question 1).

Reunification became increasingly likely with higher addiction severity. Among caregivers with low addiction severity, FTC participation was associated with a non-significant increase in the likelihood of reunification (HR=1.44). This effect strengthened for those with moderate severity (HR=2.20; $p<.01$) and remained elevated in the high severity group (HR=2.24; $p<.05$), suggesting that caregivers with moderate to high addiction severity experienced the greatest benefit from FTC participation.

In contrast, the effect of FTC participation and permanency weakened as addiction severity increased. Hazard ratios were significant among caregivers with low (HR=1.55; $p<.01$) and moderate severity (HR=1.52; $p<.05$), but declined to a non-significant HR of 1.32 among those with high severity.

For alternative permanency, FTC participation was significantly associated with a higher likelihood of this outcome only among caregivers with low addiction severity (HR=1.72; $p<.01$). Among those with moderate (HR=0.62) and high (HR=0.71) addiction severity, hazard ratios fell below 1.0, suggesting a lower likelihood of alternative permanency at higher levels of addiction severity; however, these estimates were not statistically significant.

Figure 13: Adjusted Hazard Ratios: FTC vs. TDC by Addiction Severity Level (Matched Sample, N=521)
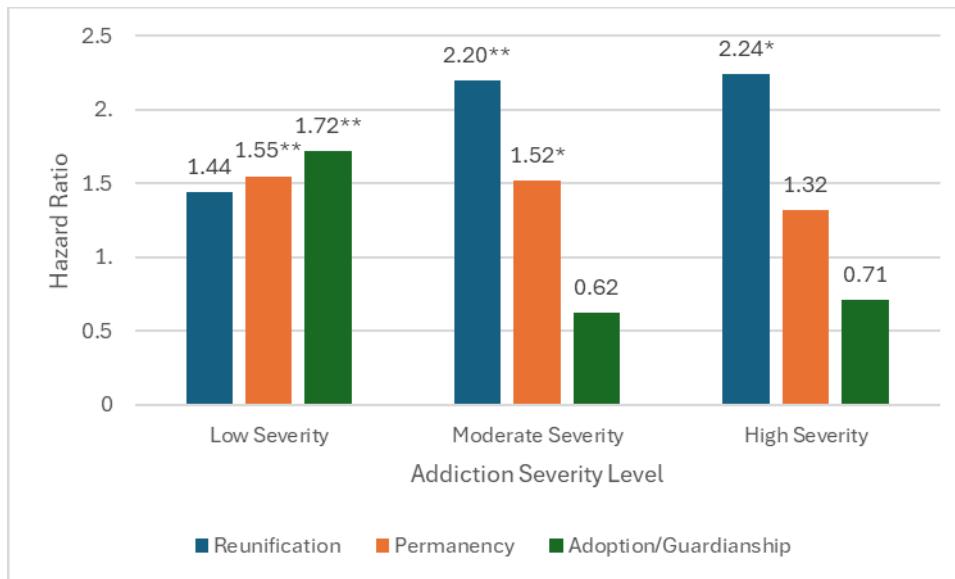
Note: Models control for treatment level of care and number of lifetime treamtent admissions. Group sizes: Group sizes: FTC Low (n=118), FTC Moderate (n=60), FTC High (n=39), TDC Low (n=172), TDC Moderate (n=65), TDC High (n=67). Adjusted models control for treatment level of care and number of lifetime treatment admissions.
* $p<.05$; ** $p<.01$; *** $p<.001$

Limitations

Our study is limited in several ways. First the sample was restricted to one state and only five FTCs, limiting generalizability to other regions and jurisdictions. Second, while robust, the research design cannot match the rigor of a randomized controlled trial, and estimates should not be interpreted as causal. Although PSM improved the balance between the FTC and TDC participant groups, unobserved confounders may still bias the results.

Third, the child welfare administrative data used in this study did not include measures of criminogenic risk, highlighting a broader disciplinary mismatch between criminal justice and child welfare research. Although our theoretical framing draws from adult drug court literature, where criminogenic risk is central to conceptualizing the model's "target population," child welfare systems do not typically assess these constructs. This is primarily because it is unknown whether key criminogenic risk constructs, namely antisocial attitudes and antisocial peers, are implicated in risk of repeat maltreatment. Our analysis relied on the ASI, which measures important service needs included in typical criminogenic and family risk assessments but does not measure these other criminogenic risk concepts. Consequently, our analysis lacks precision in assessing the extent to which "high risk, high need" individuals are better suited to FTC programs.

Additionally, our treatment data only included 'treatment episodes,' and did not include all types of substance use therapies, such as medications for opioid use disorder. We were likewise unable to control for access to other concurrent supports and services.

Probabilistic matching procedures were used to link perpetrator records and treatment records, which may have introduced measurement error due to potential mismatched or unlinked records. Additionally, because reunification was measured as a child-level outcome rather than a dyad-level outcome in the available administrative data, it was not always possible to identify which specific caregiver was the subject of the reunification. For children with both FTC and non-FTC caregivers, both dyads would be classified as achieving reunification if the child reunified, even if reunification occurred with only one caregiver. Similarly, we could not confirm whether the reunification occurred with the caregiver represented in the linked treatment data.

Finally, the study period coincided with the COVID-19 pandemic, which significantly disrupted maltreatment reporting, court operations, and treatment delivery, potentially affecting the findings.

Conclusion

This study examined whether addiction severity influences FTC effectiveness in promoting family reunification. Findings indicate that FTC participation led to greater increases in reunification rates among caregivers with moderate to high addiction severity, compared to those with low severity, relative to their counterparts in TDCs. While FTC participants with low addiction severity showed a modest, non-significant advantage in reunification (44%), those with moderate and high addiction severity experienced substantially greater benefits (120%; $p<.01$ and 124%; $p<.05$, respectively).

In addition to reunification, the study examined two secondary outcomes: permanency (any legal permanency outcome including reunification, adoption or guardianship) and alternative permanency (adoption or guardianship only). FTC participants were generally more likely than TDC participants to achieve permanency, especially among those with low and moderate addiction severity. However, the impact of FTC participation on alternative permanency was attenuated. Only caregivers with low addiction severity saw a statistically significant benefit, suggesting alternative permanency rates were similar across court types. This pattern may reflect the higher reunification rates among FTC participants with moderate and high severity, which reduced the number of children eligible for alternative permanency.

Together, the findings suggest that FTC participation is associated with increased reunification across all levels of caregiver addiction severity, underscoring the importance of broad access to FTCs. However, the benefits were greatest among caregivers with moderate to higher severity, suggesting that, in settings with limited capacity, these families may be particularly strong candidates for FTC participation. This study provides preliminary evidence that substance use disorder severity measures, such as the ASI, could inform the screening and referral process by helping identify families most likely to benefit from FTC when reunification is the primary goal.

Sub-Study 3: Family Treatment Court Cost Analysis

Study Authors

Margaret Lloyd Sieger; Jessica Becker, MSW; PhD; Jon Phillips, PhD

Study Purpose

The purpose of this sub-study is to understand whether family treatment court (FTC) participation was associated with cost offset, or savings, resulting from less foster care utilization on average. To accomplish this purpose, this study has two primary aims. One, this study aims to calculate the average "days saved" in foster care for FTC participants compared to traditional deprived court (TDC) comparison cases. Two, this study aims to determine the costs incurred or offset as a result of differences in foster care utilization.

Research Questions

This study had two research questions:

    (1) How many days in foster care were saved, on average, for FTC participants compared to matched TDC cases, based on time to reunification?

    (2) What are the cost implications of foster care days saved due to earlier reunification among FTC participants?

Methods

    Design

This study used a quasi-experimental longitudinal design.

    Data Sources and Study Population

Data for this study were collected from six counties in a Southwestern state with operational FTCs from 2018 to 2022. Three administrative data sources were linked to generate a sample of dyadic records that include information on child demographic and case characteristics, caregiver demographic and substance use treatment characteristics, and FTC participation status. The data sources included: 1) Oklahoma Department of Human Services (DHS) foster care records, 2) Oklahoma Department of Mental Health and Substance Abused Services (DMHSAS) treatment admission records, and 3) a list of FTC participants provided by Oklahoma DMHSAS, which included their FTC enrollment and discharge dates. FTC site-level budget data was obtained from project partners at Oklahoma DMHSAS.

    Timeframe

This study examined child welfare cases with removal dates that occurred between January 1, 2018 and June 1, 2022. The DHS extract includes records with removal dates as early as April 30, 2015, and return dates through May 25, 2023. The treatment admission data include adults admitted on or after July 7, 2016, and discharged on or before March 28, 2024. FTC participation records include admission dates beginning December 11, 2013, and discharge dates through April 10, 2023.

Participants

Perpetrator records and substance use treatment records were linked using probabilistic matching procedures described in Sub-Study 1. A total of 831 adult records with a non-missing perpetrator ID were successfully linked. These records were then deterministically matched to 9,007 child foster care records using a perpetrator ID variable, resulting in 2,022 perpetrator-child dyads. We excluded 449 dyads for whom no treatment data were retained based on the treatment episode selection criteria described in Sub-Study 1.

Forty-one observations were excluded that had removal dates prior to the study start date (January 1, 2018). Two observations were excluded due to missing perpetrator race information, and four were excluded due to missing Addiction Severity Index (ASI) scores because the caregivers were under the age of 18 and had instead been assessed using the Teen ASI. The ASI is a clinical tool used to assess problem severity in seven potential key areas of life for individuals with substance use disorders.

From the remaining 1,526 observations, propensity score matching procedures described in Sub-Study 2 were used to generate a final matched sample of 521 observations, comprising 217 in the FTC group (183 children and 113 caregivers) and 304 in the TDC group (295 children and 242 caregivers). Each observation represents a unique perpetrator-child-referral number combination, where the referral number corresponds to a maltreatment report and subsequent foster care placement.

Some children and perpetrators appear in multiple dyads, which were retained because each dyad reflects a distinct treatment episode (with its own treatment completion status) and a separate foster care episode (with its own reunification outcome). Two records reflected a single perpetrator-child duplicate. Both records were retained because each was associated with a different referral number, indicating separate cases that occurred at different timepoints and may have differed in context, treatment needs, and the ages of caregiver and child.

Measures

Dependent Variable

Reunification was derived from the placement exit reason variable from administrative foster care records. Originally recorded in eight categories, this variable was recoded into a binary indicator, where cases were coded as "1" if the exit reason was "Reunification" or "Reunified with Custodial Parent" and "0" for all other reasons: "Adoption," "Child Aged Out/Emancipation," "Custody to Relative," "Guardianship – Non-Relative," "Guardianship – Relative," and "Tribal Jurisdiction."

Time Variable

Time was measured as the number of days from the child's removal to their exit from foster care. Children who did not exit by the end of the observation window (May 26, 2023) were treated as censored, with time calculated as the number of days between their removal date and May 26, 2023 (one day after the latest observed return date in the dataset).

Independent Variable

FTC participation was measured using data from the list of FTC participants. Individuals enrolled in an FTC at any point during the study period were coded as "1" while those not enrolled were coded as "0" (TDC participants).

Control Variables

For consistency, the same variables used to determine the propensity scores and inverse probability weights in Sub-Study 1 were used to determine the propensity scores in this study (see *Sub-Study 1: Method—Measures—Control Variables*).

Procedures

Stata Version 18.0 was used to clean and analyze the data.

To answer Research Question 1, life tables were estimated. Life tables are a type of survival analysis that estimate the proportion of cases not yet experiencing ("have survived") a specific event. In this analysis, daily survival probabilities were estimated from the removal date through the end of the study period (May 26, 2023), treating reunification as the event of interest and censoring children who did not achieve reunification during the observation window.

In the context of survival analysis, failure was defined as a child reunifying with their caregiver. Children who exited foster care through other means (e.g., adoption, guardianship) were coded as non-failures (non-events). Children who remained in care at the end of the observation window were treated as censored.

Separate life tables were generated for the FTC and TDC groups. These life tables produced survival curves, representing the cumulative probability of not yet achieving reunification on each day of observation. We then used cubic regression to model each group's survival curve. Specifically, we regressed the cumulative survival probability on time since removal using a cubic specification for each group (Johnson-Motoyama et al., 2013). The cubic terms were statistically significant and retained for both models. These equations describe the predicted survival trajectory (i.e., probability of remaining in foster care) for FTC and TDC cases, respectively (Figure 14).

Figure 14: Cubic Equations for FTC and TDC Groups

$$Y_i = \beta_{1i} \times X_i + \beta_{2i} \times X_i^2 + \beta_{3i} \times X_i^3 + \alpha_1$$

Y_1 = Proportion of children not yet reunified in group i (i = 1, FTC; i = 2, TDC)

X = The time of observation (days) in group i

β_{ij} = coefficient of X^j in group i (j = 1, 2, 3)

α_i = intercept in group i

Model Fit:

FTC group: $F(3, 113) = 2351.69$, $p < 0.001$, Adjusted R-squared = 0.9838

$$Y_1 = -.0001357 \times X + -1.32e{-}06 \times X^2 + 6.90e{-}10 \times X^3 + 1.143578$$

TDC group: $F(3, 218) = 5987.01$, $p < .001$, Adjusted R-squared = 0.9878

$$Y_2 = -.0003076 \times X + -3.53e{-}07 \times X^2 + 1.86e{-}10 \times X^3 + 1.085027$$

To estimate the average number of foster care days saved per child, we calculated the integral of the difference between the two fitted survival curves. This integral represents the area between the curves and reflects the cumulative difference in the probability of not yet achieving reunification between the FTC and TDC groups over the observation period (Figure 15).

Figure 15: Integral Calculation of Days Saved

$$Total\ Difference\ (survival) = \int_0^{1845}[(-.0003076 \times X\_ + -3.53e-07 \times X\_{}^2 + 1.86e-10 \times X\_{}^3 + 1.085027) - (-.0001357 \times X\_ + -1.32e-06 \times X\_{}^2 + 6.90e-10 \times X\_{}^3 + 1.143578)]dx$$

To answer Research Question 2, we used the days-saved estimate from the survival analysis to compute cost avoidance. First, we calculated the total costs associated with operating FTCs using site-level budget data obtained from project partners at Oklahoma DMHSAS. Reported costs were from Fiscal Year 2025 (July 1, 2024—June 30, 2025) and categorized into: (1) Administrative costs, including FTC coordinator salaries, attorney fees, participant incentives, travel, and office supplies; and (2) Treatment Sole Source costs, which reflect non-billable provider expenses such as additional urinalysis testing, specialized assessments, staff training and travel, and other program-specific needs.

Findings

      Foster Care Days Saved

Figure 16 displays the survival curves for both FTC and TDC groups, showing the cumulative probability of not yet achieving reunification over time. These curves account for censored cases, specifically children who had not exited foster care by the end of the observation window. FTC cases achieved reunification faster than TDC participants throughout the observation period, with divergence beginning around 500 days after removal and continuing to widen over time.

Figure 16: Survival Curve of Reunification: FTC vs. TDC (N=521)
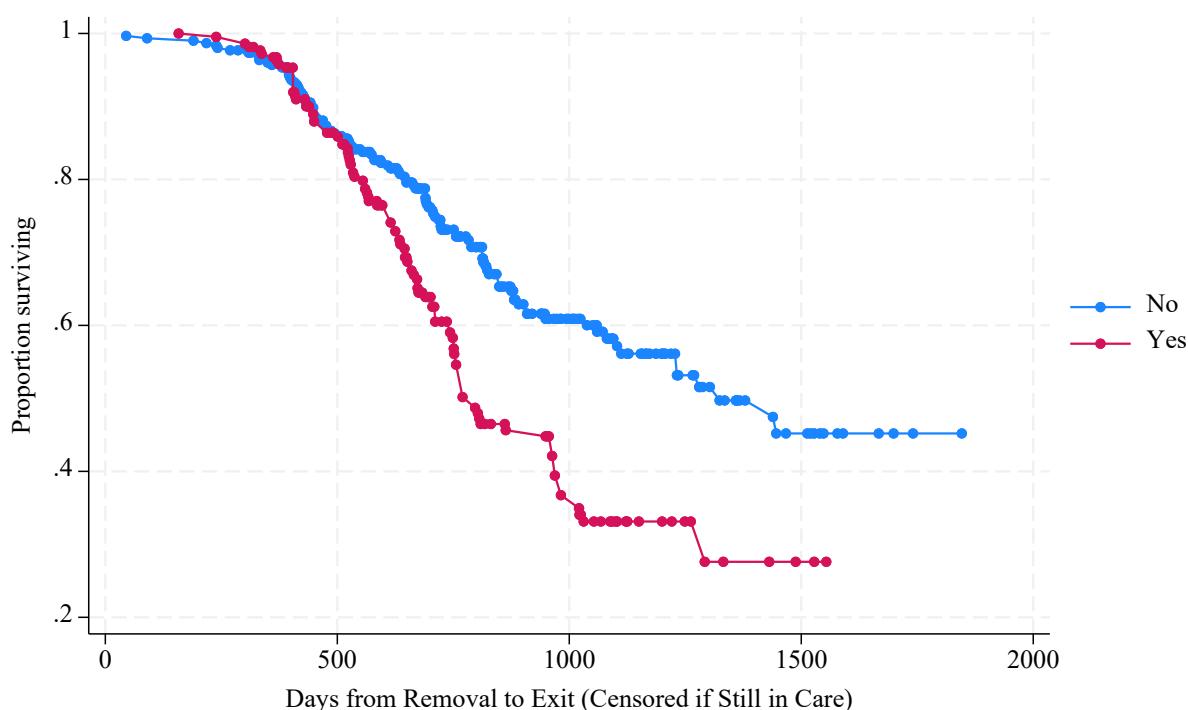
To answer Research Question 1 (foster care days saved) we used the integral calculation of the difference between the fitted cubic survival curves. Findings revealed that FTC children reunified an average of 163.8 days sooner than matched TDC cases. When extrapolated to the 183 FTC children in the study sample, this translates to an estimated 29,975.4 total foster care days saved over the course of the study period. The area between the survival curves in Figure 16 represent this cumulative difference in time to reunification and forms the basis for the cost avoidance estimate presented in Table 9.

Table 9: Integral Difference between FTC and Matched Cases on Days to Reunification

Time Frame	FTC	TDC	Integral Difference at Upper Limit (Days Saved Per Child)

	%	%	
0-180	0.0	0.7	-11.58
180-360	2.8	4.3	-7.72
360-540	19.5	15.9	2.66
540-720	39.4	25.3	16.37
720-900	54.3	36.5	30.26
900-1080	67.3	40.3	41.14
1080-1260	67.3	46.1	45.84
1260-1440	73.3	51.6	41.19
1440-1620	73.3	54.8	24.00
1620-1800	73.3	54.8	-8.89
UL: 1845	73.3	54.8	163.78

Cost Analysis Results

To answer research Question 2 (cost analysis), we first estimated total FTC implementation costs over the 5.05-year study period (1,845 days) using the FY2025 administrative and Treatment Sole Source expenditures. These costs varied across FTC sites, reflecting differences in staffing, service delivery models, and program infrastructure. For example, one FTC site incurred higher costs due to allocating funds for a peer recovery support specialist embedded in court operations, a role not adopted in other FTC. In contrast, FTCs located in urban counties often had greater access to community-based resources, which helped lower their overall expenditures compared to FTCs located in rural counties. Table 10 displays the FY2025 expenditures by FTC site and the combined total across all five FTCs.

The total reported FY2025 implementation cost was $556,918.00. To estimate cumulative costs over the full study period, we multiplied this amount by 5.05, representing the 1,845-day study window divided by a 365.25-day year (1,845 ÷ 365 ≈ 5.05 years). This yielded an estimated $2,812,435.90 in total FTC implementation costs.

Table 10: FY2025 FTC Implementation Costs and Study Period Estimate

Site	Administrative	Treatment Sole Source	Total Cost
Site A	$65,700.00	$5,000	$70,700.00
Site B	$72,000.00	$150,594.00	$222,594.00
Site C	$60,434.00	$10,000.00	$70,434.00
Site D	$84,000.00	$35,000.00	$119,000.00
Site E	$66,250.00	$7,940.00	$74,190.00
FY2025 Total			$556,918.00
Study Period Total (FY2025 Total × 5.05 years)			$2,812,435.90

We next estimated the potential reduction in foster care cost costs associated with earlier reunification among FTC participants. This estimate was calculated by multiplying the daily cost of foster care in Oklahoma ($92) by the average number of foster care days saved per FTC child (163.78)

and the number of FTC participants who utilized foster care (183). The resulting estimate reduction in foster care costs was $2,757,400.08, as shown in Table 11.

Foster Care Cost Reduction = $92 × 163.78 × 183 = $2,757,400.08

We then subtracted FTC implementation costs from the projected reduction in foster care costs to calculate the net cost difference. The resulting net difference was -$55,035.82, indicating that foster care cost reductions associated with earlier reunification did not fully offset the cost of FTC implementation.

Next Cost Difference = $2,757,400.08 - $2,812,435.90 = -$55,035.82

Lastly, we divided the estimated net cost difference by the number of FTC children in the study sample to estimate the net cost difference per child.

Next Cost Difference per Child = -$55,035.82 ÷ 183 = -$300.74

Table 11: Costs and Cost Avoidance

Domain	Cost	N Units	Total
Foster care for 183 children	$92/day	163.78 days × 183 children	$2,757,400.08
FTC Implementation Costs	$556,918.00	5.05 years	$2,812,435.90
Estimated Net Cost Difference			-$55,035.82
Estimated Net Cost Difference Per Child			-$300.74

Limitations

Our study is limited in several ways. First, the sample was restricted to one state and only five FTCs, limiting generalizability to other regions and jurisdictions. Second, while robust, the research design cannot match the rigor of a randomized controlled trial, and the estimated differences in time to reunification and associated foster care costs should not be interpreted as causal effects of FTC participation. Although IPW improved the balance between the FTC and TDC participants groups, unobserved confounders may still bias the results. While our treatment dataset included two trauma-related measures for caregivers, both were excluded due to substantial missing data (49% for trauma assessment scores and 20% for Adverse Childhood Experiences (ACE) scores). Prior research has shown that trauma can influence both substance use treatment engagement and child welfare outcomes ([Clark et al., 2020](#); [Kisiel et al., 2009](#); [Thompson, 2018](#)). Treatment readiness and criminogenic risk, understudied yet important concepts in child welfare and FTC research, may also contribute to FTC participation and permanency but were not measured in the current study. Additionally, our treatment data only included 'treatment episodes,' and did not include all types of substance use therapies, such as medications for opioid use disorder. We were likewise unable to control for access to other concurrent supports and services.

Probabilistic matching procedures were used to link perpetrator and treatment records, which may have introduced measurement error due to potential mismatched or unlinked records. Additionally, because reunification was measured as a child-level outcome rather than a dyad-level outcome in the available administrative data, it was not possible to determine with certainty which specific caregiver was the subject of reunification. Some children have two caregivers, including cases where one caregiver participated in FTC and the other did not. In these cases, both dyads were classified as achieving reunification if the child reunified, even if reunification occurred with only one caregiver. Similarly, we could not confirm whether the reunification occurred with the caregiver represented in the linked treatment data.

In estimating the costs of FTCs, we applied FY2025 implementation cost data to the entire 5.05-year study period. Because FY2025 reflects the most recent and likely highest-cost year, this approach likely overestimates total FTC costs. At the same time, we limited our estimation of savings to foster care utilization only. However, the effects of a parent's substance use recovery and a child's reunification are likely to extend across multiple systems, including healthcare, behavioral health, education, and the justice system, which were not captured in this analysis due to feasibility constraints. Together, these decisions reflect a deliberately conservative approach to estimating the net cost implications of FTC participation and likely understate its broader fiscal and societal value.

Finally, the study period coincided with the COVID-19 pandemic, which significantly disrupted maltreatment reporting, court operations, and treatment delivery, potentially affecting the findings.

Conclusion

This study examined whether participation in FTCs was associated with cost offsets resulting from earlier reunification and reduced foster care utilization. Findings revealed that children whose caregivers participated in FTC reunified approximately 164 days sooner, on average, than matched cases in TDC, resulting in an estimated 29,975 foster care days saved across the FTC sample. These days saved were associated with an estimated $2.76 million reduction in foster care costs.

These savings, however, did not fully offset the estimated $2.81 million in FTC implementation costs accrued over the 5.05-year study period. The resulting net difference was approximately $55,036, or $301 per child, suggesting that FTCs operated at a modest net cost when considering foster care savings alone. However, this calculation does not account for potential longer-term or cross-system benefits, such as improved caregiver recovery, reduced recidivism, or enhanced child well-being, which were beyond the scope of this analysis.

Taken together, these findings suggest that FTCs appear to be a promising and fiscally responsible strategy for supporting families affected by substance use disorders. Although foster care savings did not fully offset FTC implementation costs during the study period, FTC participation was associated with significantly earlier family reunification and shorter stays in foster care, indicating meaningful value for families and child welfare systems.

Sub-Study 4: Development and Pilot Testing of the Model Standards Implementation Scale (MSIS)

Study Authors

Margaret Lloyd Sieger, PhD; Jessica Becker, MSW; Kelly Earles, M.Ed.; Karin Thompson-Wise, MBA; Kaitlin Hagain

Complete Citation

Lloyd Sieger, M. H., Becker, J., Earles, K., Thompson-Wise, K., & Hagain, K. (2023). The development and pilot testing of a family treatment court best practices assessment: The Model Standards Implementation Scale. *Family Court Review*. https://doi.org/10.1111/fcre.12739

Study Purpose

In 2019, the Family Treatment Court (FTC) Best Practice Standards (the *Standards)* were published to clarify attributes of FTC programs associated with superior child, caregiver, and family outcomes. The *Standards* cover the breadth of FTC operations including program structure and leadership, substance use treatment and complementary services, and behavioral responses to participants. This study aimed to develop an instrument (the Model Standards Implementation Scale; "the MSIS") that stakeholders can use to assess implementation of the *standards* by individual FTCs. This study had three primary objectives: (1) document the development process of the MSIS, (2) evaluate its psychometric properties, including inter-rater reliability (IRR), internal consistency, and face, content, and known-groups validity, (3) present findings from pilot testing across multiple sites. By creating a tool that balances usability with scientific validity, this study aims to help FTC stakeholders systematically evaluate and enhance their programs' alignment with best practices.

Research Questions

This study had four research questions:

(1) What is the inter-rater reliability (IRR) of the MSIS?
(2) To what extent does the MSIS demonstrate internal consistency?
(3) To what extent does the MSIS demonstrate face, content, and known-groups validity?
(4) What are the implementation levels of different standards across courts?

Methods

    Design

This study used a longitudinal, psychometric validation design.

    Data Sources

Court-level implementation data was collected through annual evaluations using the newly developed measure, the MSIS. Each evaluation assessed the court's implementation of 67 best practice provisions, with one implementation score assigned per provision.

Timeframe

This study draws on pilot MSIS evaluation data collected from January 1, 2021 to December 31, 2022.

Participants

This project involved six counties in a southwestern state. At the time of data collection, these were the only counties in the state implementing FTC programs. Five FTC programs were in operation across the six counties, with one program serving a two-county region. All counties also operated at least one TDC. The counties ranged from rural to urban, with population densities ranging from fewer than 30 to over 1,000 persons per square mile. The FTC programs ranged in longevity from 2 to 23 years.

Recruitment involved emailing each courtroom's judge with a project overview and description of the procedures related to the evaluation. The sample included five FTC programs, 12 TDCs, 14 judges, and six treatment providers, all assessed across five FTC jurisdictions (three rural and two urban).

Measures

The MSIS assesses courts' implementation of 67 best practice provisions based on data collected from four sources:
(1) Interview with court or child welfare professionals,
(2) Interviews with treatment professionals,
(3) Observations of pre-court staffings and status hearings, and
(4) Documents from the court and treatment provider.

These data sources aim to capture information on both the court program's origins and current practices (that occurred within the prior 12 months), as covered by the *Standards*. Provisions are scored based on their level of implementation using a three-point scale:
- Fully Implemented (3)
- Partially Implemented (2)
- Not Yet Implemented (1)

Each provision is rated independently by two raters using data from at least two sources. For example, Rater 1 evaluates Provision 1A based on interview data and rates it again based on observation data; Rater 2 performs the same process. Raters were instructed to make ratings based on what was stated in the interview or document or observed in the staffing or status hearing. Even if the rater believed a given provision was being implemented, but no data collected during the site visit supported that conclusion, the rater was instructed to base the rating on the data. Moreover, if information yielded from one data source (e.g., interview question) suggested full implementation, and information yielded from another data source (e.g., observation) suggested no implementation, raters

were instructed to record the rating appropriate to the data source. In other words, the provision would have two ratings—full implementation as assessed by interview question and no implementation as assessed by observation. For each provision, individual ratings from different data sources and raters are converted into a single overall provision score, which can take values of 1, 1.5, 2, 2.5, or 3, reflecting a range from no implementation (1) to full implementation (3). This approach ensures implementation levels are validated through multiple sources of data and independent raters.

Procedures

The procedures detailed in this section describe the procedures that we used to evaluate implementation of the *Standards*. Although the MSIS is only intended to evaluate FTCs, both FTCs and TDCs were assessed in order to establish known groups validity for the purposes of this paper. A complete description of procedures for using the MSIS, including the instrument itself, is archived with the National Criminal Justice Reference Service (NCJRS)(available at https://www.ojp.gov/ncjrs/virtual-library/abstracts/family-treatment-court-model-standards-implementation-scale).

Interview and observational data were collected during in-person or virtual site visits to each court. Both raters participated in each interview and observation session. Each rater was instructed to complete their own scoring sheet, independently from the other rater, typically during, but no more than 2 days after, the site visit. Each site visit included the following:

- 90-min interview with FTC coordinator
- Observation of an entire FTC staffing meeting and status hearing
- 90-min interview with TDC case manager (or other key informant)
- Observation of an entire TDC judge's status hearing
- 45-min interview with treatment provider(s)

Analysis

Stata Version 18.0 was used to clean and analyze the data.

IRR was calculated using the two rater Cohen's Kappa statistic for all interviews and observations. Benchmarks for adequate IRR according to Landis and Koch (1977) were as follows: 0-20%=Slight agreement; 21-40%=Fair agreement; 41-60%=Moderate agreement 61-80%=Substantial agreement; 81-100%=Almost perfect agreement.

Internal consistency for items measuring each *Standard* was calculated using Cronbach's alpha. George and Mallery's (2003) commonly used criteria for Cronback's alpha suggest the following: >0.9=Excellent, >0.8=Good, >0.7=Acceptable, >0.6=Questionable, >0.5=Unacceptable.

Three types of validity were assessed: face, content, and known group validity. Face validity considers whether an instrument appears to be measuring what it intends to measure, while content validity is the extent to which an instrument adequately covers the range of subject matter (Fitzpatrick et al.,

1998). In this study, these two concepts ask whether the MSIS is measuring the *Standards*, and whether all elements of the *Standards* are covered.

Unlike face and content validity, known group validity is a quantitative assessment of validity that evaluates the extent to which a measure is able to distinguish between two different populations it should theoretically differentiate. In this study, known groups validity was established if the MSIS scores significantly differ between the FTCs and the TDCs. Mean differences for each *Standard* across two groups were measured using one-way ANOVA.

Other forms of validity, such as convergent and divergent validity—the extent to which scores on this instrument are similar or different from other instruments measuring similar or different concepts— are not testable at this time because there are no other published instruments for measuring FTC *Standards* implementation.

Development of the instrument

Publication of the *Standards* was the field's first opportunity to create a simple, practice-friendly measurement instrument that states, counties, and individual courts can use to assess their programs' level of FTC fidelity and the association between FTC fidelity and client outcomes. The MSIS was developed in three stages and subsequently refined after initial implementation. Stage one involved distilling measurable key concepts from each provision. The *Standards* provide a brief (one to two paragraphs) explanation of each provision followed by a longer description of its rationale and key considerations. The authors described the "essential component" of each provision explanation to serve as the definition for that provision. The essential component paraphrases the brief explanation provided in the *Standards*. For example, the essential component from provision 1A, "Multidisciplinary & Multisystemic Collaborative Approach," is coordination and collaboration between the court, child welfare, substance use treatment, mental health treatment, children's services and related health, education, and social service systems.

Next, the methods of assessment were determined. The three possible methods of assessment were interview, observation, and document review. Interview questions could pertain to court and/or treatment professionals' observations during court staffings or court hearings, and document review could be based on documents obtained from a court or the treatment provider.

Thereafter, interview questions were developed, observation criteria were defined, and the type of information expected to be found in the documents was specified.

The MSIS was implemented using REDCap, a secure, web-based application used for data collection, storage, and transmission.

Findings

      Inter-Rater Reliability (IRR)

Table 12 presents IRR findings for five types of data collection completed by two raters: interviews with court personnel, interviews with treatment personnel, observations during FTC staffings, observations during hearings, and review of documents from both the court and the corresponding treatment provider. Aside from the 13 null values that could not be computed due to too few rating categories, all but one IRR value met the criteria for moderate agreement and over two-thirds met the criteria for substantial agreement. Nearly a third of the IRR ratings met the criteria for an almost perfect agreement. The ratings with the highest IRR were document review, which may be explained by both the large number of missing documents and corresponding missing scores as well as the objective nature of reviewing the documents for criteria. Conversely, the ratings with the lowest IRR were observations of hearings, which may be explained by the more subjective nature of observing live processes for criteria. It is worth noting that IRR generally improved over time in both FTC and TDC settings.

Table 12: Inter-Rater Reliability

Court	Interview Court	Interview Treatment	Observation Staffing	Observation Hearing	Document Review
FTC 1	77.8%	77.8%	93.3%	38.9%	86.4%
FTC 2	73.4%	83.3%	66.7%	72.2%	86.4%
FTC 3	60.9%	88.9%	53.3%	72.2%	100.0%
FTC 4	62.5%	88.9%	66.7%	61.1%	90.9%
FTC 5	70.3%	88.9%	66.7%	66.7%	88.6%
TDC 1	62.9%	77.8%	73.3%	50.0%	+
TDC 2	66.7%	77.8%	+	55.6%	100.0%
TDC 3	66.7%	77.8%	+	44.4%	100.0%
TDC 4	64.1%	66.7%	+	47.1%	88.6%
TDC 5	64.1%	66.7%	+	55.6%	88.6%
TDC 6	64.1%	66.7%	+	61.1%	88.6%
TDC 7	62.5%	88.9%	+	55.6%	100.0%
TDC 8	59.4%	55.6%	+	61.1%	96.5%
TDC 9	57.8%	55.6%	+	+	95.5%
TDC 10	59.4%	55.6%	+	64.7%	95.5%
TDC 11	57.8%	55.6%	+	55.6%	95.5%
TDC 12	59.4%	55.6%	+	50.0%	95.5%

Internal Consistency

Table 13 presents Cronbach's alpha scale reliability coefficients for each *Standard*. Coefficients for each *Standard* range from 0.67 (*Standard* 3) to 0.94 (*Standard* 1). All but two coefficients (*Standards* 3 and 5) meet the criteria for acceptable internal consistency, and two coefficients (*Standards* 1 and 7) meet the criteria for excellent internal consistency ([George & Mallery, 2003](#)).

Table 13: Cronbach's Alpha

Standard	Scale Reliability Coefficient	Number of Items in the Scale	Average Interitem Covariance
1	0.94	10	0.22
2	0.70	6	0.12
3	0.67	5	0.09
4	0.76	4	0.16
5	0.68	9	0.02
6	0.83	10	0.05
7	0.92	14	0.17
8	0.78	4	0.17

Validity

Face validity was established through an iterative process involving the research team and other stakeholders familiar with the *Standards* ("MSIS reviewers"). MSIS reviewers were asked to consider each item and the measure as a whole in their assessment of its clear and straightforward approach to measuring the *Standards.* Feedback from MSIS reviewers was incorporated prior to pilot-testing. Furthermore, use of the instrument in several different counties and a multitude of courts—both FTC and DTC—established that the instrument has face validity from practitioners and court personnel who were interviewed, observed, and asked to provide documents as a part of the pilot study.

To establish content validity, a table was created aligning each provision with each corresponding item (at least two MSIS items using two data collection sources for each provision). Key concepts in the provision were highlighted and cross-referenced to language included in each item. All authors and an additional research assistant reviewed each provision and its corresponding items. Changes to the instrument resulting from this cross-referencing review process were incorporated prior to pilot-testing.

Table 14 presents the results of one-way ANOVA tests conducted to assess known groups validity, based on 2021-2022 data, by examining differences between MSIS scores by court type. All *Standards* show statistically significant differences in scores in FTCs compared to TDCs. *Standard* 3 scores significantly higher for FTC at the $p<.01$ level while the scores for all other *Standards* are significantly higher for FTC at the $p<.001$ level.

Table 14: One-Way ANOVA Test Showing Differences Between MSIS Scores by Court Type

Std	Structure	FTC Mean (SD)	TDC Mean (SD)	Test Statistic
1	Organization and structure	2.73 (0.24)	1.79 (0.21)	$F(1,15) = 75.34$, $p<.001$
2	Role of the judge	2.85 (0.12)	2.07 (0.21)	$F(1,15) = 60.74$, $p<.001$
3	Ensuring equity and inclusion	1.72 (0.19)	11.38 (0.18)	$F(1,15) = 11.94$, $p<.01$
4	Early identification, screening, and assessment	2.80 (0.12)	2.16 (0.16)	$F(1,15) = 61.34$, $p<.001$
5	Timely, high-quality, and appropriate SUD treatment	2.95 (0.05)	2.74 (0.08)	$F(1,15) = 33.64$, $p<.001$

6	Comprehensive case management, services, and supports for families	2.82 (0.06)	2.44 (0.16)	F(1,15) = 24.40, *p*<.001
7	Therapeutic responses to behavior	2.90 (0.11)	2.05 (0.16)	F(1,15) = 114.19, *p*<.001
8	Monitoring and evaluation	2.25 (0.32)	1.46 (0.31)	F(1,15) = 22.43, *p*<.001

Model Standards Implementation (MSIS) Scores

Table 15 presents MSIS scores for all provisions averaged across FTC and TDC courts. FTC programs demonstrated full implementation (two ratings of 3) for 25 (37.3%) of the Provisions, and ratings of at least 2.5 (at least one rating of 3) for 56 (83.6%) of the Provisions. Several FTCs demonstrated ratings of at least 2.5 for all provisions within certain *Standards*, including *Standard* 2 (Role of the Judge), *Standard* 4 (Early Identification), *Standard* 5 (Treatment), and *Standard* 7 (Responses to Behavior). TDC programs demonstrated ratings of at least 2.5 for 22 (32.8%) of the Provisions. FTC scored significantly higher than TDCs for 37 (55.2%) provisions.

Poor implementation (ratings of less than 2) was observed for only four (6.0%) of the provisions in the FTC. The poorly implemented provisions were within *Standard* 3 (Equity & Inclusion) and *Standard* 8 (Monitoring & Evaluation).

Table 15: Provision MSIS Scores (Mean & SD)

Std	Provision	FTC Mean (SD)	TDC Mean (SD)	Test Statistic
Std. 1 Organization & Structure				
1A	Collaborative	2.80 (0.45)	1.96 (0.45)	F(1,15) = 12.38, p < 0.01
1B	Partnerships	2.80 (0.27)	2.13 (0.23)	F(1,15) = 27.97, p < 0.001
1C	Multidisciplinary	2.80 (0.45)	1.91 (0.20)	F(1,14) = 31.59, p < 0.001
1D	Governance	2.20 (0.84)	1.04 (0.14)	F(1,15) = 23.45, p < 0.001
1E	Mission	3.00 (0.00)	1.13 (0.23)	F(1,15) = 330.88, p < 0.001
1F	Communication	2.90 (0.22)	2.63 (0.23)	F(1,15) = 5.25, p < 0.05
1G	Interdisciplinary Ed	2.30 (0.45)	1.92 (0.29)	F(1,15) = 4.53, p = 0.050
1H	Approach	2.60 (0.55)	2.23 (0.52)	F(1,14) = 1.72, p = 0.211
1I	FTC Manual	2.90 (0.22)	1.25 (0.45)	F(1,15) = 58.83, p < 0.001
1J	FTC Staffing	3.00 (0.00)	1.67 (0.25)	F(1,15) = 141.18, p < 0.001
Std. 2 Role of the Judge				
2A	Partners	2.90 (0.22)	1.33 (0.44)	F(1,15) = 54.90, p < 0.001
2B	Decision Making	3.00 (0.00)	2.75 (0.45)	F(1,15) = 1.47, p < 0.244
2C	Participation	3.00 (0.00)	1.17 (0.58)	F(1,15) = 48.53, p < 0.001
2D	Interaction	2.90 (0.22)	2.17 (0.33)	F(1,15) = 20.83, p < 0.001
2E	Training	2.70 (0.45)	2.29 (0.50)	F(1,15) = 2.50, p < 0.135
2F	Assignment Length	2.60 (0.55)	2.71 (0.58)	F(1,15) = 0.13, p = 0.728
Std. 3 Equity & Inclusion				
3A	Admissions	2.80 (0.27)	1.00 (0.00)	F(1,15) = 571.76, p < 0.001

3B	Retention Rates	1.40 (0.55)	1.71 (0.45)	F(1,15) = 1.47, p = 0.244
3C	Treatment	1.10 (0.22)	1.25 (0.45)	F(1,15) = 0.49, p = 0.496
3D	Responses	1.10 (0.22)	1.45 (0.52)	F(1,14) = 2.07, p = 0.173
3E	Team Training	2.20 (0.76)	1.50 (0.21)	F(1,15) = 9.26, p < 0.01
Std. 4 Early Identification, Screening, and Assessment				
4A	Target Population	2.70 (0.45)	1.00 (0.00)	F(1,15) = 191.25, p < 0.001
4B	Standardized	2.80 (0.27)	1.50 (0.53)	F(1,13) = 26.15, p < 0.001
4C	Screening with Parent	3.00 (0.00)	3.00 (0.00)	
4D	Screening with Child	2.50 (0.50)	2.75 (0.26)	F(1,15) = 1.89, p = 0.189
4E	Barrier Resolution	3.00 (0.00)	2.42 (0.29)	F(1,15) = 19.65, p < 0.001
Std. 5 Timely, High-Quality, and Appropriate SUD Treatment				
5A	Timely Access	2.90 (0.22)	2.33 (0.49)	F(1,15) = 5.93, p < 0.05
5B	Needs Matching	3.00 (0.00)	3.00 (0.00)	
5C	Continuum of Care	3.00 (0.00)	2.83 (0.39)	F(1,15) = 0.88, p = 0.362
5D	Integrated Treatment	3.00 (0.00)	2.79 (0.26)	F(1,15) = 3.15, p = 0.096
5E	Family-Centered	2.80 (0.27)	2.67 (0.25)	F(1,15) = 0.97, p = 0.339
5F	Gender-Responsive	3.00 (0.00)	2.88 (0.23)	F(1,15) = 1.47, p = 0.244
5G	Pregnant Women	2.80 (0.45)	2.46 (0.26)	F(1,15) = 4.04, p = 0.063
5H	Culture	2.90 (0.22)	2.42 (0.29)	F(1,15) = 11.08, p < 0.01
5I	Evidence-Based	3.00 (0.00)	3.00 (0.00)	
5J	Medication-Assisted Treatment	3.00 (0.00)	2.96 (0.14)	F(1,15) = 0.40, p = 0.536
5K	Drug Testing	3.00 (0.00)	2.50 (0.00)	
5L	Qualifications	3.00 (0.00)	3.00 (0.00)	
Std. 6 Comprehensive Case Management, Services, and Supports for Families				
6A	Case Management	3.00 (0.00)	2.21 (0.33)	F(1,15) = 26.99, p < 0.001
6B	Family Involvement	2.70 (0.27)	2.46 (0.45)	F(1,15) = 1.22, p = 0.286
6C	Recovery Supports	3.00 (0.00)	2.92 (0.19)	F(1,15) = 0.88, p = 0.362
6D	Visitation	2.30 (0.27)	2.13 (0.23)	F(1,15) = 1.88, p = 0.191
6E	Parenting Programs	3.00 (0.00)	2.79 (0.26)	F(1,15) = 3.15, p = 0.096
6F	Reunification	3.00 (0.00)	2.58 (0.47)	F(1,15) = 3.80, p = 0.070
6G	Trauma Services	3.00 (0.00)	3.00 (0.00)	
6H	Children's Services	3.00 (0.00)	2.58 (0.36)	F(1,15) = 6.49, p < 0.05
6I	Parent Services	2.90 (0.22)	2.33 (0.25)	F(1,15) = 19.62, p < 0.001
6J	Early Intervention	2.70 (0.27)	2.33 (0.33)	F(1,15) = 4.85, p < 0.05
6K	SUD Intervention	2.40 (0.65)	1.50 (0.52)	F(1,15) = 9.12, p < 0.01
Std. 7 Therapeutic Responses to Behavior				
7A	Child and Family Focus	2.90 (0.22)	2.71 (0.33)	F(1,15) = 1.36, p = 0.262
7B	Adjustments	2.90 (0.22)	2.42 (0.19)	F(1,15) = 20.06, p < 0.001
7C	Service Modifications	3.00 (0.00)	2.21 (0.26)	F(1,15) = 45.50, p < 0.001

7D	FTC Phases	2.80 (0.27)	1.73 (0.47)	$F(1,14) = 22.31$, $p < 0.001$
7E	Incentives & Sanctions	2.80 (0.45)	1.00 (0.00)	$F(1,15) = 214.41$, $p < 0.001$
7F	Equity	2.70 (0.27)	2.13 (0.48)	$F(1,15) = 6.11$, $p < 0.05$
7G	Certainty	2.90 (0.22)	1.00 (0.00)	$F(1,14) = 868.66$, $p < 0.001$
7H	Advance Notice	3.00 (0.00)	2.00 (0.77)	$F(1,14) = 8.02$, $p < 0.05$
7I	Timely Response	3.00 (0.00)	1.63 (0.43)	$F(1,15) = 48.53$, $p < 0.001$
7J	Be Heard	2.90 (0.22)	2.21 (0.50)	$F(1,15) = 8.65$, $p < 0.05$
7K	Demeanor	3.00 (0.00)	2.96 (0.14)	$F(1,15) = 0.40$, $p = 0.536$
7L	Child Safety	2.80 (0.45)	2.88 (0.31)	$F(1,15) = 0.16$, $p = 0.695$
7M	Substance Use	3.00 (0.00)	2.17 (0.69)	$F(1,15) = 7.12$, $p < 0.05$
7N	Discharge	2.90 (0.22)	1.50 (0.39)	$F(1,14) = 55.49$, $p < 0.001$
Std. 8 Monitoring & Evaluation				
8A	Electronic Database	2.90 (0.22)	1.96 (0.58)	$F(1,15) = 11.95$, $p < 0.01$
8B	Quality Improvement	1.80 (0.57)	1.75 (0.69)	$F(1,15) = 0.02$, $p = 0.889$
8C	Best Practices	2.10 (0.55)	1.13 (0.23)	$F(1,15) = 28.55$, $p < 0.001$
8D	Evaluation Methods	2.20 (0.45)	1.00 (0.00)	$F(1,15) = 95.29$, $p < 0.001$

Limitations

Given the complexity of the *Standards*, a major challenge in developing the MSIS was translating its extensive content into a practical assessment tool. The *Standards* outline eight broad areas with 67 specific provisions, detailing critical aspects of FTC operations. Capturing every detail would have resulted in an instrument that was too lengthy and subjective for practical use. To balance specificity with feasibility, we systematically distilled measurable concepts from each provision, retaining only essential components and removing redundancy or ambiguous elements.

Once we defined what to measure, we determined how to measure it. Implementation fidelity studies use various methods, including self-report, observations, and recordings, each with strengths and limitations (Breitenstein et al., 2010). In FTCs, knowledge gaps among professionals posed challenges—treatment providers specializing in children may not be familiar with adult services, and court coordinators may lack systems-level knowledge. We found that group interviews with multiple professionals, rather than only one professional, provided a more complete picture. To address inconsistencies, we used the highest rating for each provision across multiple data sources.

Beyond development, implementation of the MSIS posed challenges. The evaluation process is resource-intensive, requiring significant time from raters and FTC team members. Overburdened courts may hesitate to allocate resources, but we found that providing feedback reports incentivized participation. Courts were eager to see their adherence to the *Standards* and areas for improvement, and remained engaged in subsequent evaluations to track improvements.

Additionally, the fact that data collection occurred during the COVID-19 pandemic also introduced obstacles, as courts transitioned between virtual and in-person formats. Virtual settings presented

issues like internet disruptions and procedural confusion, initially lowering inter-rater reliability (IRR). When courts returned to in-person operations, IRR dipped again before improving with practice. These findings suggest that continued MSIS use will further enhance reliability and promote better adherence to the *Standards.*

Conclusion

The MSIS exhibits good IRR, face validity, content validity, and known group validity, as well as acceptable internal consistency. Our findings suggest the FTC programs examined are implementing many provisions with high levels of fidelity, while other provisions remain poorly implemented. Overall, 37% of provisions were rated Fully Implemented by both raters at each of the five FTCs evaluated, and 84% of provisions were rated Fully Implemented by at least one rater. Only 6% of Provisions were rated "Not Yet Implemented" by at least one rater.

The high and low ratings clustered within certain *Standards.* While *Standard* 2 (Role of the Judge), 4 (Early Identification), 5 (Treatment), and 7 (Therapeutic Responses) were rated as Fully Implemented by at least on rater for 100% of provisions, *Standard* 3 (Equity & Inclusion) and 8 (Monitoring & Evaluation) were rated "Not Yet Implemented" by at least on rater for 60% and 25% of provisions, respectively. Low implementation of *Standards* 3 and 8 among the courts evaluated in this study suggests that courts may struggle to implement the data collection and quality improvement practices emphasized in these *Standards.*

As FTCs proliferate, it is important that states and individual courts assess the extent to which their program conforms to the *Standards* and identify areas of improvement in order to achieve the desired outcomes. The MSIS is one newly developed approach to assessing implementation that demonstrates adequate reliability and validity. Future longitudinal research using the tool across settings and states is needed.

Tool Evolution and Future Directions

Following this study, we refined the MSIS based on pilot testing results and feedback. This process included revising, adding, and deleting interview questions, observation criteria, and document review items to improve clarity, relevance, and usability. The tool was subsequently restructured from its original REDCap format into a Word-based data collection instrument and an Excel-based scoring instrument. In this updated version, evaluators administered the items in the data collection instrument and then entered their ratings into the scoring instrument, which included the new feature of automatic overall provision score calculation. The transition from REDCap to Word and Excel formats increased public accessibility by eliminating the need for specialized database access. This version is now publicly available through NIJ.

Additionally, the research team developed a beta web-based version of the tool, called the Family Treatment Court Implementation Tool (FIT), designed to enhance usability, streamline data collection and scoring, and provide automated, tailored feedback to courts. As of September 2025, the FIT is publicly accessible at no cost for pilot testing (https://thefittool.com/); however, public availability

and pricing may change as the tool is further developed. Future research is needed to evaluate how this enhanced version of the tool supports courts' evaluation of the *Standards*, to identify additional features that may further improve its utility, and to determine whether high alignment with the *Standards* predicts positive family and treatment outcomes.

Sub-Study 5: Key Family Treatment Court Practices Associated with Reunification

Study Authors

Jessica Becker, MSW; Margaret Lloyd Sieger, PhD; Jon Phillips, PhD

Study Purpose

This study examines the relationship between the implementation of [Family Treatment Court (FTC) Best Practice Standards](#) (the *Standards)* and reunification in child welfare cases. A substantial body of empirical research, including findings from [Sub-Study 1](#) and [Sub-Study 2](#), demonstrates that FTCs outperform traditional child welfare settings with respect to reunification. However, FTC programs vary widely in their structure and practices, making it challenging to identify the specific elements that drive their effectiveness. The *Standards* provide a detailed framework of 67 evidence-based provisions grouped into eight core standards, yet the impact and relative importance of individual provisions on reunification has not been empirically assessed. This study quantifies the association between each best practice provision and reunification outcomes, identifying ten key provisions that most strongly predict family reunification. A focus on implementation of these practices may help FTCs maximize their effectiveness in supporting participants and their families.

Research Questions

This study had two research questions:

(1) What is the association between the level of implementation of each FTC best practice provision and family reunification?
(2) Which FTC best practice provisions are most strongly associated with family reunification?

Methods

Design

This study used a longitudinal, observational research design.

Data Sources and Study Population

Data for this study were collected from six counties in a Southwestern state with operational FTCs from 2018 to 2022. Three administrative data sources were linked to generate a sample of dyadic records that include information on child demographic and case characteristics, caregiver demographic and substance use treatment characteristics, and FTC participation status. The data sources include: 1) Oklahoma Department of Human Services (DHS) foster care records, 2) Oklahoma Department of Mental Health and Substance Abuse Services (DMHSAS) treatment admission records,

and 3) a list of FTC participants provided by Oklahoma DMHSAS, which included their FTC enrollment and discharge dates. In addition, court-level implementation data were collected using the Model Standards Implementation Scale (MSIS). The research team conducted MSIS evaluations across 18 courts, including five FTCs and 13 traditional deprived courts (TDCs). The MSIS assessed each court's level of implementation across 67 best practice provisions, assigning one score per provision during each annual evaluation.

Timeframe

This study examined child welfare cases with adjudication hearing dates that occurred between October 1, 2018 and September 30, 2022. The DHS extract includes records with removal dates as early as April 30, 2015, and return dates through May 25, 2023. The treatment admission data include adults admitted on or after July 7, 2016, and discharged on or before March 28, 2024. FTC participation records include admission dates beginning December 11, 2013, and discharge dates through April 10, 2023. MSIS evaluation data were collected from 2021 to 2025; however, this sub-study only includes data collected during 2021 and 2022 to align with the timeframe for the adjudication hearings.

Participants

Perpetrators' child welfare records and substance use treatment records were linked using [probabilistic matching procedures](#) described in Sub-Study 1. A total of 831 adult records with a non-missing perpetrator ID were successfully linked. These records were then deterministically matched to 9,007 child foster care records using a perpetrator ID variable, resulting in 2,022 perpetrator-child dyads.

These linked administrative records were subsequently merged with court-level MSIS implementation data, collected annually from 2021-2022, using judge, year, and court program type (FTC or TDC). Administrative records from prior to 2021 were matched to the earliest available MSIS evaluation from 2021. Of the 2,022 observations, 592 were excluded because the court listed in the record was not included in the MSIS pilot study. One additional case was excluded due to missing Addiction Severity Index (ASI) scores because the caregiver was under the age of 18 and had instead been assessed using the Teen ASI. The ASI is a clinical tool used to assess problem severity in seven potential areas of life for individuals with substance use disorders.

Next, we excluded 356 observations for whom no treatment data were retained based on the [treatment episode selection criteria](#) described in Sub-Study 1. The final analytic sample included 1,073 dyads, comprising 168 in the FTC group (146 children and 85 caregivers) and 905 in the TDC group (754 children and 483 caregivers). Each observation represents a unique perpetrator-child-referral number combination, where the referral number corresponds to a maltreatment report and subsequent foster care placement.

Some children and perpetrators appear in multiple dyads, which were retained because each dyad reflects a distinct treatment episode (with its own treatment completion status) and a separate foster care episode (with its own reunification outcome). Two records reflected a single perpetrator-

child duplicate. Both records were retained because each was associated with a different referral number, indicating separate cases that occurred at different timepoints and may have differed in context, treatment needs, and the ages of caregiver and child.

Measures

Dependent Variable

Reunification was derived from the placement exit reason variable from administrative foster care records, which was originally recorded in eight categories. The variable was coded as "1" if the exit reason was "Reunification" or "Reunified with Custodial Parent" and "0" for all other exit reasons: "Adoption," "Child Aged Out/Emancipation," "Custody to Relative," "Guardianship – Non-Relative," "Guardianship – Relative," and "Tribal Jurisdiction."

Independent Variables

Each court was scored on its alignment with the *Standards* at the provision level. A total of 67 provisions were rated on a 1 to 3 scale, with possible increments of 0.5 (1=Not Yet Implemented; 3=Fully Implemented). These provision scores, derived from the MSIS served as the independent variables in the analysis. Further detail on the MSIS and data collection procedures is provided in Sub-Study 4.

Variables Used in Propensity Score Weighting (PSW) Model

For consistency, the same variables used to determine the propensity scores and inverse probability weights in Sub-Study 1 were used to determine the PSW in this study (see Sub-Study 1: Method— Measures—Control Variables).

Procedures

Stata Version 18.0 was used to clean and analyze the data.

Propensity Score Weighting (PSW)

Propensity scores were estimated using logistic regression to predict the probability of FTC participation based on demographic, child welfare, and treatment-related covariates. These scores represented each participant's predicted probability of FTC participation given their observed characteristics. The scores were then applied as probability weights in subsequent analyses to adjust for baseline differences between FTC and TDC participants. Individuals with a higher likelihood of FTC participation ("FTC-likely participants") contributed more to the estimation, focusing the analysis on the population most likely to enroll in FTC. References to the weighted sample reflect estimates for this FTC-likely population.

At the same time, retaining the full sample preserved statistical power, ensured representation of the broader child welfare population, and allowed inclusion of individuals who resembled FTC participants but did not enroll, representing a relevant population that could be targeted for future FTC recruitment or access.

Propensity scores ranged from 0.006 to 0.677 (M=0.16, SD=0.12), with a right-skewed distribution (skewness=1.17), consistent with the low baseline probability of FTC participation given the 5:1 TDC-to-FTC ratio in the sample. Most participants (75%) had scores below 0.23, and only 10% exceeded 0.33.

Descriptive Characteristics

Demographic, child welfare, and treatment characteristics of FTC and TDC caregivers and children were compared to assess differences between groups. Chi-squared tests were used to evaluate differences in categorical variables, and one-way analysis of variance (ANOVA) was conducted to compare group means for continuous variables. Because the propensity score weights were used to focus the analysis on FTC-likely individuals rather than to achieve covariate balance between groups, only unweighted descriptive statistics are reported, and covariate balance diagnostics are not presented.

Weighted Logistic Regression Models

Using the propensity score weights, we conducted 67 separate logistic regression models to examine the association between each individual MSIS provision score and the likelihood of reunification (Research Question 1). We identified the 10 provisions from these models with the largest effect sizes (i.e., odds ratios [OR]) as the most strongly associated with reunification (Research Question 2).

Findings

Cross-Site Outcome Comparison Results

Table 16 shows unweighted reunification rates by court type across the five sites included in the analysis. In all sites, FTC participants had higher reunification rates than their TDC counterparts. Absolute differences ranged from 0.06 in Site E to 0.19 in Site C, corresponding to a 6% and 19% advantage for FTC participants, respectively. On average, 51.2% of FTC participants reunified, compared to 41.7% of TDC participants. To facilitate comparison across sites, we calculated a standardized percent difference as the difference between the TDC group's rate and the FTC group's rate, divided by the TDC rate. This approach has been used in cross-site evaluations ([Carey et al., 2012](#)) to account for variation across sites. We found that the standardized percent difference in reunification was 41.7% higher for FTCs. To better understand this advantage, we next examined the relationship between each of the 67 best practice provisions and reunification to identify which were more strongly associated with reunification.

Table 16: Unweighted Reunification Rates by Court Type and Site (N=1,107)

Site	FTC Reunification Rate	FTC n	TDC Reunification Rate	TDC n
Site A	0.33	27	0.18	99
Site B	0.70	20	0.55	106
Site C	0.53	68	0.34	499
Site D	0.57	23	0.41	49
Site E	0.47	30	0.41	152

Unweighted Descriptive Characteristics

Table 17 compares demographic, child welfare, treatment characteristics, and reunification outcomes for FTC and TDC caregivers and children prior to applying the propensity score weights. Results demonstrate several significant differences between the unweighted groups. FTC participants were significantly younger, with caregivers averaging 30.1 years compared to 32.9 years ($p<.001$), and their children averaging 3.0 years versus 4.2 years ($p<.001$). FTC participants were more likely to be White (73.8% vs. 60.3%) and less likely to be Black or Native ($p<.01$), and were also more likely to be female (77.4% vs. 68.3%; $p<.05$).

There were no significant differences between groups in number of maltreatment allegations or placement patterns. Children of FTC participants were more likely to be placed with family members (64.3% vs. 46.9%; $p<.001$).

Treatment recommendations also differed significantly ($p<.001$). FTC participants were less likely to be referred to crisis or detox services and to outpatient care, and more likely to be referred to community living, halfway houses, or residential treatment ($p<.001$). There were also significant differences in primary substance: FTC participants were less likely to report alcohol ($p<.001$) and more likely to report a stimulant ($p<.01$), with no significant difference in reporting cannabis or an opioid as their primary substance.

On the ASI, FTC participants scored higher on the drug ($p<.01$), family ($p<.001$), and psychiatric ($p<.05$) domains, with no group differences on the employment domain. There were also significant differences in lifetime treatment admissions ($p<.05$), with FTC participants more likely to have 4-5 admissions and TDC participants more likely to have 6 or more.

Finally, FTC participants were more likely to achieve family reunification (51.2% vs. 36.1%; $p<.001$).

Table 17: Unweighted Descriptive Characteristics (N=1,107)

Variable	FTC Participants (N=173) Mean (SD) or n (%)	TDC Participants (N=934) Mean (SD) or n (%)	Significance
Child age at removal	3.04 (3.62)	4.22 (4.31)	$p<.001$
Caregiver age	30.12 (5.85)	32.86 (6.91)	$p<.001$
Caregiver race			$p<.01$
White	124 (73.81)	546 (60.33)	

Table 17: Unweighted Descriptive Characteristics (N=1,107)

Variable	FTC Participants (N=173) Mean (SD) or n (%)	TDC Participants (N=934) Mean (SD) or n (%)	Significance
Black	23 (13.69)	182 (20.11)	
Native	21 (12.50)	177 (19.56)	
Caregiver female sex	130 (77.38)	618 (68.29)	*p*<.05
Total # of allegations			N.S.
1	103 (61.31)	520 (57.46)	
2	46 (27.38)	243 (26.85)	
3	17 (10.12)	100 (11.05)	
4 or more	2 (1.19)	42 (4.64)	
Total # of placements			N.S.
1-2	61 (36.31)	317 (35.03)	
3-4	70 (41.67)	327 (36.13)	
5 or more	37 (22.02)	261 (28.84)	
Any kinship placements	108 (64.29)	424 (46.85)	*p*<.001
Treatment level of care recommendation			*p*<.001
Community living or residential treatment	74 (44.05)	187 (20.66)	
Crisis or detox	10 (5.95)	107 (11.82)	
Outpatient	84 (50.00)	611 (67.51)	
Primary substance of choice			
Alcohol	14 (8.33)	192 (21.22)	*p*<.001
Cannabis	21 (12.50)	133 (14.70)	N.S.
Opioid	22 (13.10)	105 (11.60)	N.S.
Stimulant	107 (63.69)	451 (49.83)	*p*<.01
ASI domain scores			
Drug	6.84 (1.74)	6.26 (2.36)	*p*<.01
Employment	4.97 (2.73)	4.75 (2.96)	N.S.
Family	6.40 (2.00)	5.69 (2.58)	*p*<.001
Psychiatric	5.55 (2.56)	5.02 (2.86)	*p*<.05
# Lifetime treatment admissions			*p*<.05
1	19 (11.31)	153 (16.91)	
2-3	51 (30.36)	272 (30.06)	
4-5	55 (32.74)	203 (22.43)	
6 or more	43 (25.60)	277 (30.61)	
Reunified by study end	86 (51.19)	327 (36.13)	*p*<.001

Weighted Logistic Regression Results

Figure 17 illustrates the distribution of effect sizes across the 67 provisions in relation to reunification outcomes. Effect sizes were categorized based on the following benchmarks defined by Rosenthal (1996): weak=below 1.5 (above reverse 0.67); small=1.5 (reverse of 0.67); medium=2.5 (reverse of 0.4); large=4 (reverse of 0.25). Most provisions were positively associated with reunification, typically showing weak to small positive effects, though some demonstrated medium to large associations.

Of the 67 provisions, 57 (85.1%) were positively associated with reunification (OR>1), including 37 (55.2%) with statistically significant effects. Conversely, nine provisions (13.4%) were negatively associated with reunification (OR<1), though all had weak to small effect sizes and only two reached statistical significance. One provision (5L) showed no association with reunification (*OR*=1).

Figure 17: Distribution of Effect Sizes Across 67 FTC Best Practices (Weighted FTC-Likely Sample, N=1,107)
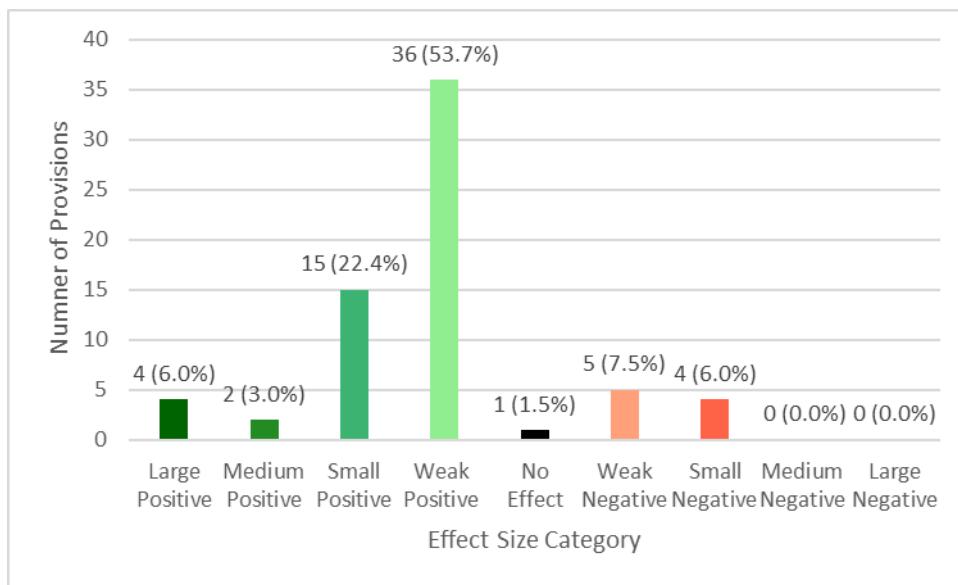
Table 18 lists all 67 best practice provisions and their corresponding odds ratios and significance levels for associations with reunification, based on a weighted analysis focused on FTC-likely participants (Research Question 1). Odds ratios reflect the change in the odds of reunification associated with a half-point increase in implementation score, which ranged from 1.0 (Not Yet Implemented) to 3.0 (Fully Implemented). For example, a half-point increase in the implementation score for Standard 1 Provision A, which reflects the extent to which a court implements a multidisciplinary and multisystemic collaborative approach, was associated with 1.3 times greater odds of reunification.

Table 18: Odds Ratios for Reunification Associated with Implementation of 67 Best Practices (Weighted FTC-Likely Sample, N=1,107)

Best Practice Provision	Standard and Provision Name	Odds Ratio	n
1A	Multidisciplinary & Multisystemic Collaborative Approach	1.34	1,054
1B	Partnerships, Community Resources & Support	1.32	1,054
1C	Multidisciplinary Team	1.52*	1,008
1D	Governance Structure	1.37*	1,054
1E	Shared Mission & Vision	1.33**	1,054
1F	Communication & Information Sharing	0.97	1,054
1G	Cross-Training & Interdisciplinary Education	1.14	1,054
1H	Family-Centered, Culturally-Relevant, and Trauma-Informed Approach	1.28	955
1I	Policy & Procedure Manual	1.33**	1,054
1J	Pre-Court Staffing & Review Hearing	1.60***	1,054
2A	Convening Partners	1.29*	1,054
2B	Judicial Decision Making	1.50	1,054
2C	Participation in Pre-Court Staffings	1.38***	1,054
2D	Interaction with Participants	1.21	1,054
2E	Professional Training	1.03	1,054
2F	Length of Judicial Assignment to FTC	1.08	1,054
3A	Equitable FTC Program Admission Practices	1.50***	1,054
3B	Equitable FTC Retention Rates and Child Welfare Outcomes	1.00	1,054
3C	Equitable Treatment	0.78	1,054
3D	Equitable Responses to Participant Behavior	0.90	955
3E	Team Training	1.88***	1,054
4A	Target Population, Objective Eligibility and Exclusion Criteria	1.48***	1,054
4B	Standardized and Systematic Referral, Screening, and Assessment Process	1.25	853
4C	Use of Valid and Reliable Screening and Assessment for Caregivers and Families	7.71***	1,054
4D	Use of Valid, Reliable, and Developmentally Appropriate Screening and Assessment for Children	0.62*	1,054
4E	Identification and Resolution of Barriers to Recovery and Reunification	1.16	1,054
5A	Timely Access to Appropriate Treatment	1.23	1,054
5B	Treatment Matches Assessed Need	0.50**	1,054
5C	Comprehensive Continuum of Care	2.72**	1,054

Table 18: Odds Ratios for Reunification Associated with Implementation of 67 Best Practices (Weighted FTC-Likely Sample, N=1,107)

Best Practice Provision	Standard and Provision Name	Odds Ratio	n
5D	Integrated Treatment of Co-Occurring Substance Use and Mental Health Disorders	1.02	1,054
5E	Family-Centered Treatment	1.38	1,054
5F	Gender-Responsive Treatment	1.32	1,034
5G	Treatment for Pregnant Women	1.91*	1,054
5H	Culturally Responsive Treatment	2.09**	1,054
5I	Evidence-Based Manualized Treatment	167.63*	1,054
5J	Medication Assisted Treatment	0.41	1,054
5K	Alcohol and Other Drug Testing Protocols	3.08**	1,054
5L	Treatment Provider Qualifications	1.00	1,054
6A	Intensive Case Management and Coordinated Case Planning	1.13	1,054
6B	Family Involvement in Case Planning	1.04	1,054
6C	Recovery Supports	7.89***	1,054
6D	High-Quality Parenting Time (Visitation)	1.16	1,054
6E	Parenting and Family-Strengthening Programs	1.31	1,054
6F	Reunification and Related Supports	1.13	1,054
6G	Trauma-Specific Services for Children and Caregivers	25.44*	1,054
6H	Services to Meet Children's Individual Needs	1.66*	1,054
6I	Complementary Services to Support Caregivers and Family Members	2.10**	1,054
6J	Early Intervention Services for Infants and Children Affected by Prenatal Substance Exposure	1.07	1,054
6K	Substance Use Prevention and Intervention for Children and Adolescents	1.32*	1,054
7A	Child and Family Focus	0.75	1,054
7B	Treatment Adjustments	2.12*	1,054
7C	Complementary Service Modifications	1.68*	1,054
7D	FTC Phases	1.65***	955
7E	Incentives and Sanctions to Promote Engagement	1.32**	1,054
7F	Equitable Responses	1.46*	1,054
7G	Certainty	1.43***	955
7H	Advance Notice	1.63***	955
7I	Timely Response Delivery	1.45**	1,054
7J	Opportunity for Participants to be Heard	1.10	1,054
7K	Professional Demeanor	2.33*	1,054
7L	Child Safety Interventions	0.59	1,054
7M	Use of Addictive or Intoxicating Substances	1.42**	1,054

Table 18: Odds Ratios for Reunification Associated with Implementation of 67 Best Practices (Weighted FTC-Likely Sample, N=1,107)

Best Practice Provision	Standard and Provision Name	Odds Ratio	n
7N	FTC Discharge Decisions	1.60***	955
8A	Maintain Data Electronically	1.36*	1,054
8B	Engage in Process of Continuous Quality Improvement	1.17	1,054
8C	Evaluate Adherence to Best Practices	1.41*	1,054
8D	Use of Rigorous Evaluation Methods	1.43*	1,054

Note: Odds ratios>1 indicate a positive association with reunification; <1 indicate a negative association.

* *p<.05, ** p<.01, *** p<.001*

Top 10 FTC Best Practice Provisions Associated with Reunification

The provisions highlighted below represent the ten most strongly associated with reunification, ranked by effect size (Research Question 2). All odds ratios exceed 2.0 and are statistically significant, distinguishing them from the remaining provisions, which showed smaller effects. Odds ratios indicate how the likelihood of reunification changed with higher implementation. For example, Standard 5 Provision I, which had the largest effect size, was associated with odds of reunification that were 167.6 times higher for each half-point increase in implementation score. Asterisks denote statistical significance: **$p<.01$, ***$p<.001$.

1. 5I: Evidence-Based Manualized Treatment (*OR=167.6**)
   - Partner treatment providers use evidence-based, manualized treatments.
   - Fidelity to the treatment model is assessed on a regular basis.
   - Providers are trained, certified (when applicable), and clinically supervised.

2. 6G: Trauma-Specific Services for Children and Caregivers (*OR=25.4**)
   - Trauma-specific interventions are available to FTC participants and their children.
   - These interventions are evidence-based or evidence informed and provided by trained professionals with fidelity.
   - FTC participants and their children are screened or assessed for trauma.

3. 6C: Recovery Supports (*OR=7.9***)
   - The FTC links participants with professionally trained or certified recovery specialists or with peer support recovery coaches.
   - The FTC team helps participants build a community-based recovery support network.
   - A range of peer support options are offered, rather than requiring a specific group.

4. 4C: Use of Valid and Reliable Screening and Assessment for Caregivers and Families (*OR=7.7***)

- Valid and reliable instruments are used to screen and assess families referred to FTC.
- Assessments guide program eligibility, case planning, treatment level of care, and service referrals.

5. 5K: Alcohol and Other Drug Testing Protocols (*OR=3.1***)
     - A standardized drug testing protocol specifies the frequency (minimum of two times per week), scheduling, randomization, observation, and breadth of testing.
     - The protocol outlines processes for confirmation, notification, and dissemination of test results.
     - Testing is conducted to monitor both licit and illicit substance use.

6. 5C: Comprehensive Continuum of Care (*OR=2.7***)
     - Participants have access to a continuum of substance use disorder treatment that includes outpatient treatment, intensive outpatient treatment, partial hospitalization, residential or inpatient treatment, and medical detox.
     - Medication management is available at each level of care.

7. 7K: Professional Demeanor (*OR=2.3**)
     - FTC team members interact with participants, children, and families in a respectful and professional manner.

8. 7B: Treatment Adjustments (*OR=2.1**)
     - Team considers whether non-compliance reflects a therapeutic problem before issuing a sanction.
     - Treatment type, level of care, and dosage are adjusted based on clinical needs, including substance use and mental, physical, social, or emotional health.
     - Adjustments are made in consultation with clinical professionals.
     - Treatment adjustments are not used as incentives or sanctions.

9. 6I: Complementary Services to Support Caregivers and Family Members (*OR=2.1***)
     - FTC participants have access to a comprehensive range of complementary support services (e.g., child care, employment, educational, domestic violence, legal, transportation, food, clothing, housing, medical, and dental care).
     - Complementary services are selected based on formal assessment and tailored to individual needs.
     - Services promote engagement and retention in treatment, sustained recovery, and permanency.

10. 5H: Culturally Responsive Treatment (*OR=2.1***)
     - Treatment services and practices are respectful of and responsive to the cultural and linguistic needs of FTC participants.

Limitations

Findings of this study should be interpreted in the context of several limitations. First, the sample was restricted to one state and only six counties, limiting generalizability to other regions and jurisdictions. Second, the relatively small number of FTC participants (n=168) limited our ability to examine variation within the FTC population itself. While PSW was used to give greater weight to individuals with characteristics similar to FTC participants, this statistical adjustment cannot substitute for a larger sample of actual FTC participants. The weighting approach enabled a focused analysis of the population most likely to participate in FTC but was not designed to balance the FTC and TDC groups. Moreover, while the covariates selected for weighting were grounded in the literature and predictive of FTC participation in our data, they may not fully capture all relevant factors that determine FTC eligibility or likelihood of participation.

Third, the data include inherent clustering at both the court and family levels. The analysis used court-level predictors (implementation scores) to model an individual-level outcome (reunification), but the small number of courts in the sample limited our ability to statistically adjust for clustering. In addition, some children were part of the same family units, introducing possible correlation in reunification outcomes due to sibling relationships.

Fourth, the observational design precludes causal inference. Because courts were not randomly assigned to implement specific best practices, the observed associations may reflect correlation rather than a direct causal effect. In particular, implementation of one provision may be correlated with other practices—measured or unmeasured—that may be driving improvements in reunification outcomes. Additionally, MSIS scores were collected in 2021 following delays caused by the COVID-19 pandemic, and were retroactively applied to cases adjudicated as early as 2018. This assumes that court practices remained relatively stable over time, which may not fully reflect actual shifts in implementation.

Fifth, the MSIS tool itself was refined during the evaluation period through an iterative process informed by pilot testing and evaluator feedback. Revisions included the addition, removal, and clarification of interview questions, observation criteria, and document review items to improve clarity, relevance, and usability. Although the content of the implementation criteria (i.e., the best practice provisions) remained unchanged, these refinements to the instrument may have introduced some inconsistencies in how provisions were scored across courts or over time.

Sixth, probabilistic matching procedures were used to link perpetrator records and treatment records, which may have introduced measurement error due to potential mismatched or unlinked records. Additionally, because reunification was measured as a child-level outcome rather than a dyad-level outcome in the available administrative data, it was not always possible to identify which specific caregiver was the subject of reunification. For children with both FTC and non-FTC caregivers, both dyads would be classified as achieving reunification if the child reunified, even if reunification occurred with only one caregiver. Similarly, we could not confirm whether the reunification occurred with the caregiver represented in the linked treatment data.

Finally, the study period coincided with the COVID-19 pandemic, which significantly disrupted maltreatment reporting, court operations, and treatment delivery, potentially affecting the findings.

Conclusion

This study provides the first empirical examination of the relationship between implementation of specific FTC best practice provisions and family reunification outcomes. FTC participants had a 41.7% higher reunification rate than their TDC counterparts, and 85% of the 67 provisions were positively associated with reunification. Although nine provisions were negatively associated with reunification, all had weak to small effect sizes.

The weighted analysis identified a core set of best practice provisions most strongly linked to improved reunification. The most impactful practices focused on evidence-based manualized treatment, trauma-specific services for children and caregivers, robust recovery supports, valid and reliable screening and assessment, and standardized alcohol and drug testing protocols.

These findings offer actionable guidance for FTC programs. Courts may wish to prioritize implementation of these high-impact practices to improve reunification outcomes. While the results reinforce the effectiveness of the FTC model, they also highlight the potential for targeted implementation of key practices to maximize impact. Future research should examine whether these relationships persist across different jurisdictions and populations, and explore other key outcomes such as treatment completion.

Sub-Study 6: Survey Findings of Caregivers with Maltreatment Cases

Study Authors

Jessica Becker, MSW; Margaret Lloyd Sieger, PhD

Study Purpose

This study aimed to explore the baseline characteristics of caregivers involved in child maltreatment cases, with a focus on comparing participants in family treatment courts (FTCs) to those in traditional deprived courts (TDCs). Specifically, we sought to understand both the demographic profiles and clinical characteristics of these populations, examining differences in substance use patterns, mental health status, and trauma symptoms between FTC and TDC participants. By collecting data through comprehensive surveys administered at the time of initial appearance hearings, the study aimed to establish baseline differences between caregivers who ultimately chose to participate in FTCs versus those who proceeded through traditional child welfare court process (TDCs).

Research Questions

This study had two research questions:

(1) What are the demographic and descriptive characteristics of FTC participants compared to TDC participants?
(2) How do FTC participants compare to TDC participants in terms of substance use, mental health, and trauma symptoms?

Methods

Design

This study utilized a cross-sectional survey design.

Timeframe

This study collected survey responses from May 6, 2021 to May 10, 2023.

Participants

Participants were recruited from six counties in a Southwestern state with operational FTCs during the study period. Eligible participants were at least 18 years old, had a child in custody of the Department of Human Services (DHS), and were in the early stages of their child maltreatment case. To ensure that participants completed the survey at the start of their case, they were asked to complete the survey within 60 days of receiving the recruitment flyer, which was distributed at the initial appearance hearing. As a result, participants completed the survey before making the decision to participate in an FTC. Participation was voluntary.

Figure 18 depicts the number of respondents who attempted, screened in, and completed the survey. In total, 30 legitimate attempts to unique individuals were made to complete the survey. Two of the legitimate attempts were screened out; one because they started the survey more than 60 days from their initial appearance and one because they indicated they did not have a child in DHS custody. Thus, 28 respondents (93.33% of legitimate attempts) screened in. One person chose not to sign the consent form, required to proceed to the survey. A total of 27 respondents (90%) completed the survey.

Figure 18: Survey Completion Flow Chart

```
┌─────────────────────┐
│ 30 Legitimate Survey│
│      Attempts       │
└──────────┬──────────┘
           │
           ▼
┌─────────────────────┐        ┌─────────────────────┐
│   28 Screened In    │───────▶│   2 Screened Out    │
└──────────┬──────────┘        └─────────────────────┘
           │
           ▼
┌─────────────────────┐
│ 27 Completed Survey │
└─────────────────────┘
```
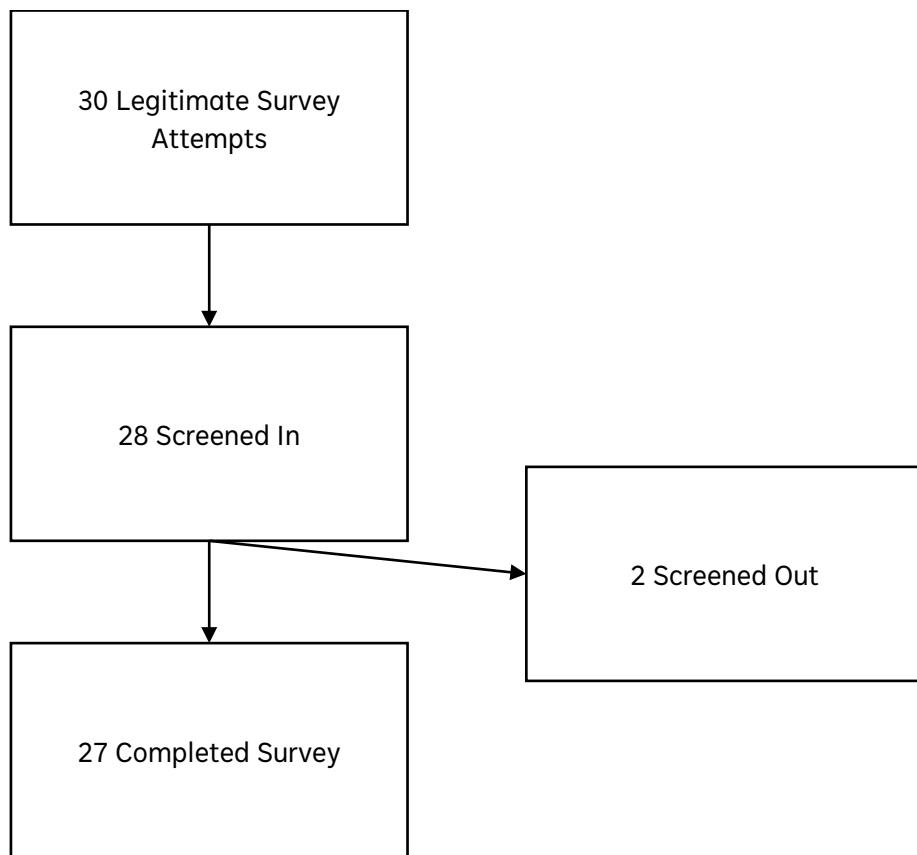
Recruitment Efforts and Challenges

Our goal was to recruit a total sample of 1,125-1,500 adult-child dyads. Recruitment flyers in English and Spanish describing the survey and eligibility criteria were distributed to court case managers (or individuals in an equivalent position) and members of the research team who distributed them to potential participants at the time of their initial appearance hearing for their child maltreatment case. Each flyer contained a unique identification code (Flyer ID) and instructions for participants to text a research team members' cell phone number. The research team member would verify the Flyer ID's validity before sending the survey link. Prior to starting the screener, participants selected their preferred language (English or Spanish) for completing both the screener and survey.

Recruitment for this survey faced significant challenges despite multiple strategies to engage participants. Data collection occurred during the COVID-19 pandemic, when court proceedings were cancelled and/or transitioned to online platforms, reducing in-person interactions and limiting opportunities for recruitment as originally planned. Additionally, the number of cases reported may have decreased during this time, affecting the potential pool of participants.

To increase participation, we increased the incentive from $30 to $60 and posted large flyers in courthouses, treatment facilities, and DHS offices. We revised our recruitment flyers distributed individually to participants to include a QR code, allowing participants to directly access the survey without needing to contact the research team. Further, we collaborated with court and treatment personnel to identify additional recruitment strategies, which included "Lunch and Learn," educational videos, and recruiting participants at any hearing or interaction with court, DHS, or treatment professionals during the first 60 days of the caregiver's child maltreatment case.

Despite these efforts, we were only able to recruit 27 complete responses, including 5 caregivers who ultimately participated in an FTC. As a result, this study will only report high-level descriptive statistics, as the small sample size limits our ability to draw broader conclusions about caregivers in FTC and TDC settings. Furthermore, to maintain participant confidentiality given the small sample size, specific counts across analytic categories are not reported.

Data Sources

Electronic survey data was collected using Research Electronic Data Capture (REDCap), which is a HIPAA-compliant web-based data collection and storage system. A list of FTC participants provided by Oklahoma DMHSAS was used to identify which participants ultimately participated in an FTC, after survey completion.

Measures

1. *Screener:* The screener collected the following information to determine participant eligibility:
   a. Date of birth: Verified participants' age to confirm they met the minimum requirement of 18 years.
   b. Child in DHS Custody: Confirmed participants currently had a child in DHS custody, a requirement for eligibility.
   c. Date of Recruitment Flyer: Identified when the participant received the flyer to estimate their initial appearance date. Only those who received the flyer within 60 days of completing the screener were eligible.
   d. Flyer ID: Tracked unique codes assigned to each recruitment flyer. These codes were used to validate eligibility and identify the county of adjudication, as flyers were only distributed to caregivers with DHS cases in Oklahoma counties served by an active FTC.
   e. Previous Participation: Verified that participants had not previously completed the survey, as this would disqualify them from participation.

2. *Demographic Questionnaire*: This instrument obtained the following demographic information:
   - Respondent:
     a. Race
     b. Ethnicity
     c. Sex
     d. Living situation
     e. Education level
     f. Employment status
     g. Income
     h. Previous DHS involvement
   - Respondent's youngest child currently in DHS custody:
     a. Race
     b. Ethnicity
     c. Sex
     d. Prenatal substance exposure
     e. Special needs status
3. *Addiction Severity Index (ASI) Self-Report Questionnaire:* This instrument assesses problems associated with substance use and generates composite scores ranging from 0 (no reported problems) to 1 (highest level of reported problems) across the following domains:
   a. Employment Status
   b. Medical Status
   c. Psychiatric Status
   d. Family/Social Status
   e. Alcohol Use
   f. Drug Use
   g. Legal Status
4. *SOCRATES 8D Personal Drug Use Questionnaire*: This instrument assesses readiness for change among individuals who use alcohol or drugs, providing subscale scores in the following categories:
   a. Recognition (range: 7-35): Higher scores indicate acknowledgement of problems and a desire for change.
   b. Ambivalence (range: 4-20): Higher scores reflect uncertainty and potential openness to reflection.
   c. Taking Steps (range: 8-40): Higher scores are predictive of successful behavior change.
5. *Center for Epidemiologic Studies Depression Scale (CES-D)*: This instrument assesses depressive symptoms, with total scores ranging from 0 to 60. Higher scores indicate more severe depressive symptoms.
6. *Trauma Symptom Checklist-40 (TSC-40)*: This instrument assesses trauma-related problems, with higher scores indicating greater severity of trauma-related problems. The instrument provides a total score ranging from 0 to 120, as well as subscale scores across the following categories:

     a.   Dissociation (range: 0-18)
     b.   Anxiety (range: 0-27)
     c.   Depression (range: 0-27)
     d.   SATI (Sexual Abuse Trauma Index) (range: 0-21)
     e.   Sleep Disturbance (range: 0-18)
     f.   Sexual Problems (range: 0-24)

Findings

Demographic Characteristics

The mean age of participants was 32.84 (SD=7.19), ranging from 22 to 48. The majority of participants identified as White non-Hispanic (59%), while the remainder identified as American Indian, Black, Latino, or other racial backgrounds. Over two-thirds of the participants were female. Among the sample of 27 participants, five ultimately participated in an FTC.

Due to small and unequal sample sizes, Mann-Whitney U and Fisher's exact tests were used to compare group differences on continuous and categorical variables, respectively. No statistically significant differences were observed between groups in age, sex, or racial/ethnic distribution.

Risk Characteristics

Table 19 reports mean scores and standard deviations for clinical assessment measures, including ASI domains, SOCRATES 8D subscales, CES-D total score, and TSC-40 subscales and total score for the entire sample. The clinical assessments revealed mild to moderate difficulties across multiple domains. On the ASI, Employment Status showed the highest average severity score (0.60), while Medical Status, Alcohol Use, and Drug Use domains indicated the lowest scores, indicating minimal reported problems.

The SOCRATES 8D subscales reflected low to moderate readiness to change. Participants scored an average of 18.00 on Recognition and 25.73 on Taking Steps, suggesting limited acknowledgement of problems, but moderate engagement in change behavior. Ambivalence scores were lower on average (8.27), indicating minimal uncertainty about drug use problems.

The mean total score on the CES-D was 16.85, reflecting mild depressive symptoms. For the TSC-40, subscale scores were consistently low. The Sexual Problems subscale had the lowest mean score (3.32), while Sleep Disturbance had the highest (6.54). The mean total score for the TSC-40 was 25.91, suggesting mild trauma-related symptoms in the sample.

Table 19: Mean Scores for Clinical Assessment Instruments (N=27)

Instrument	Domain/Subscale	Mean Score (SD)
ASI Self-Report Questionnaire		
	Employment Status	0.60 (0.34)
	Medical Status	0.01 (0.05)

	Psychiatric Status	0.24 (0.23)
	Family/Social Status	0.27 (0.25)
	Alcohol Use	0.00 (0.00)
	Drug Use	0.04 (0.06)
	Legal Status	0.19 (0.33)
SOCRATES 8D		
	Recognition	18.00 (10.37)
	Ambivalence	8.27 (4.79)
	Taking Steps	25.73 (12.36)
CES-D		
	Total Score	16.85 (14.00)
TSC-40		
	Dissociation	3.60 (4.08)
	Anxiety	5.85 (6.07)
	Depression	6.91 (6.74)
	SATI (Sexual Abuse Trauma Index)	3.92 (4.93)
	Sleep Disturbance	6.54 (5.49)
	Sexual Problems	3.32 (4.87)
	Total Score	25.91 (26.64)

Due to small and unequal sample sizes, Mann-Whitney U tests were conducted to compare group differences. FTC participants scored higher scores on the ASI Self-Report Questionnaire Drug Use domain, lower on the Family/Social Status domain, and similar on the Employment Status, Medical Status, Psychiatric Status, Alcohol Use, and Legal Status domains. However, none of the ASI Self-Report Questionnaire domain differences were statistically significant. FTC participants scored higher on all SOCRATES 8D subscales (Recognition, Ambivalence, and Taking Steps), with only the Taking Steps subscale showing statistical significance ($p<.05$). On the CES-D, FTC participants scored slightly higher on the total score, though this difference was not statistically significant. For the TSC-40 subscales, FTC participants scored higher across all subscales, with the Sexual Problems approaching statistical significance ($p=0.07$).

Limitations

Our study is limited in several ways, and as such, the results should be interpreted with caution. First, the sample size was small (N=27) and unequal across groups. As a result, findings are not generalizable. Furthermore, the small sample size, particularly within the FTC group (n=5), reduced the statistical power of the analyses and likely contributed to the lack of significant findings in several comparisons despite observed differences. Future studies, with larger samples are needed to understand baseline differences between caregivers who end up participating in an FTC and caregivers who do not.

Conclusion

Due to significant recruitment challenges and a small sample size (N=27), this study's findings are primarily descriptive in nature. While some differences were observed between FTC and TDC participants, particularly in readiness for change and trauma symptoms, the limited sample size prevents drawing broader conclusions about these populations. Future research with larger samples is needed to better understand the characteristics of caregivers who participate in FTCs versus traditional dependency courts.

Sub-Study 7: Assessing Cross-systems Collaboration as a Predictor of Implementation Success

Study Authors

Jon D. Phillips, PhD; Elizabeth J. Goldsborough, MSW; Cindy Nichols, PhD

Study Purpose

Preliminary analyses of previously collected data (see sub-study 4) revealed considerable variability in how family treatment court (FTC) programs are implemented, including the extent to which they adhere to the [Best Practice Standards (the *Standards*)](#). The primary aim of this study is to identify factors that support or hinder implementation of the *Standards*, a critical step in fostering environments where FTCs can successfully operate.

This study is guided by the Consolidated Framework for Implementation Research (CIFR), a leading implementation science theory [(Damschroder et al., 2022)](#). The CFIR outlines five key domains of factors that influence implementation outcomes in healthcare and social service settings: intervention characteristics, outer setting, inner setting, characteristics of individuals, and implementation process. Several of these domains highlight the importance of cross-systems collaboration at both the agency and professional levels, aligning with multiple *Standards* that require agencies to develop joint protocols and procedures for professionals to work together. Given the importance of cross-systems collaboration when implementing interventions like FTCs, our study examined whether the capacity for collaboration at the professional-level (i.e., interprofessional collaboration) and agency-level (i.e., interagency collaboration) were associated with the successful implementation of the *Standards*.

Research Questions

This study had the following two research questions:

(1) Is capacity for interprofessional and interagency collaboration associated with the extent of implementation of best practice standards?
(2) What are the relative contributions of interprofessional and interagency collaboration capacities to the implementation of best practice standards?

Methods

Design

This study utilized a cross-sectional survey design.

Timeframe

This study collected survey responses from April 2, 2025 to June 6, 2025.

Recruitment and Study Population

We used a purposive convenience sampling strategy to recruit child welfare professionals from one Southwestern state with eight operational FTCs. Eligible participants were professionals who held one or more of the following roles:

1. A dedicated member of an FTC team.
2. A professional in the community who works with children, parents, or families involved with the child welfare system (direct FTC involvement not required).
3. A member of a steering or executive committee for FTCs.
4. A director or administrator of an agency that serves children, parents, or families involved with the child welfare system (direct FTC involvement not required).

Contact lists for potential respondents were developed in collaboration with project partners at Oklahoma Department of Mental Health and Substance Abuse Services (OK DMHSAS). FTC teams provide OK DMHSAS updated rosters annually, which served as the foundation for identifying key stakeholders, including steering and executive committee members and agency leadership. This process yielded a primary list of 157 professionals, including agency leaders.

Recruitment materials were distributed by the researchers through two strategies:

1. Direct outreach to professionals: Research staff emailed a standardized recruitment message, which included a study description and a link to an online survey.
2. Agency-led distribution: Administrators in regional offices for Oklahoma Department of Human Services (DHS) and behavioral healthcare agencies distributed the same recruitment message to their staff via email.

Participation was open to professionals serving the child welfare population in any of the eight counties with an operational FTC, regardless of direct involvement with FTC. Professionals were encouraged to forward the survey invitation to eligible colleagues. Follow-up reminder emails were sent by the research team to increase response rates.
Participants who completed the survey received a $60 gift card

Final Study Sample

Figure 19 depicts the number of respondents who attempted, screened in, and completed the survey. In total, there were 188 attempts to complete the survey. Sixteen of the 188 attempts were screened out because the respondent did not work in the counties of interest. One respondent was removed from the sample because they did not consent to participate, and nine respondents were removed because they did not answer any surveys questions other than those used to determine eligibility. Overall, a total of 162 respondents were included in the sample.

Figure 19: Survey Completion Flow Chart

```
┌─────────────────────────┐
│                         │
│  188 Survey Attempts    │
│                         │        ┌─────────────────────────┐
└───────────┬─────────────┘        │   16 Screened Out       │
            │         ─────────────▶   Did not work in        │
            │                      │   counties of interest   │
            ▼                      └─────────────────────────┘
┌─────────────────────────┐
│                         │
│  172 Screened In        │
│                         │        ┌─────────────────────────┐
└───────────┬─────────────┘        │   10 Screened Out       │
            │                      │                         │
            │         ─────────────▶   1 did not consent;     │
            │                      │   9 did not answer survey│
            ▼                      │   questions             │
┌─────────────────────────┐        └─────────────────────────┘
│                         │
│  162 Completed          │
│  Survey                 │
│                         │
└─────────────────────────┘
```
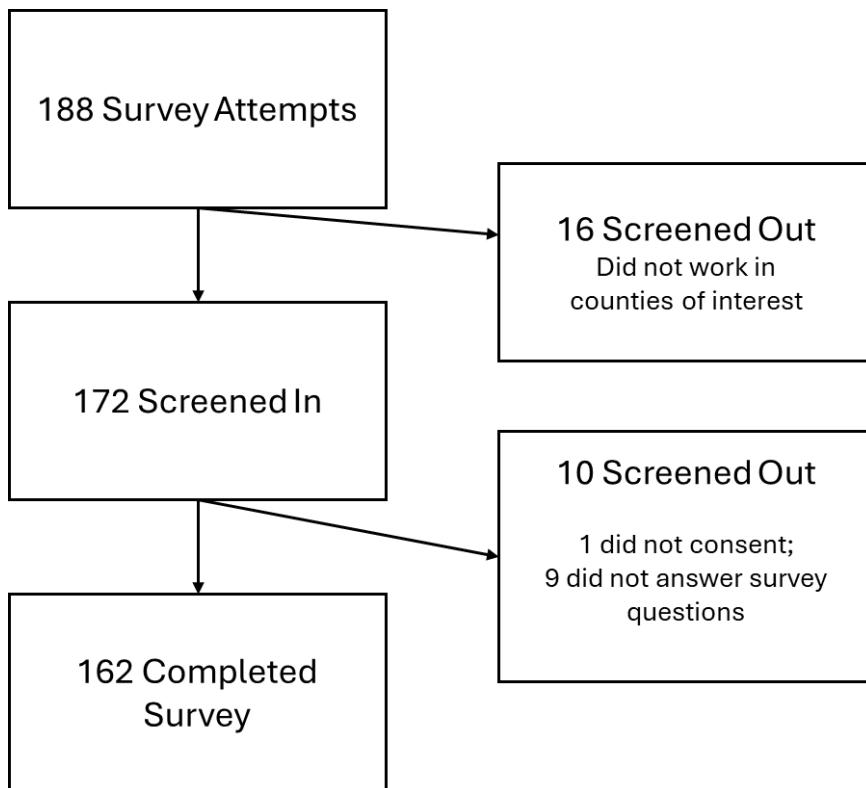
Data Sources

Two data sources were used for this study. *Electronic survey data* was collected from professionals using Qualtrics, which is a secure, web-based platform data collection and storage system. The survey data included measures of capacity for interprofessional and interagency collaboration, as well as questions about respondents' professional and demographic backgrounds. *Model Standards Implementation Scale (MSIS) data* was used to assess courts' implementation of best practice standards. The MSIS data contains ratings for 67 provisions that are grouped into 8 standards. The data was collected from interviews with court professionals, child welfare professionals, and treatment professionals; observations of pre-court staffings and status hearings; and documents from the courts and treatment providers (see sub-study 4 for additional information about the development and application of the MSIS).

The MSIS measures the extent of implementation of 67 provisions using a three-point scale:
- Fully Implemented (3)
- Partially Implemented (2)
- Not Yet Implemented (1)

Each provision is rated independently by two raters using data from at least two sources. For example, Rater 1 evaluates Provision 1A based on interview data and rates it again based on observation data; Rater 2 performs the same process. Raters were instructed to make ratings based

only on the data available. If information yielded from one data source (e.g., interview question) suggested full implementation, and information yielded from another data source (e.g., observation) suggested no implementation, raters were instructed to record the rating appropriate to the data source. In other words, the provision would have two ratings—full implementation as assessed by interview question and no implementation as assessed by observation. For each provision, individual ratings from different data sources and raters are converted into a single overall provision score, which can take values of 1, 1.5, 2, 2.5, or 3, reflecting a range from no implementation (1) to full implementation (3). This approach ensures implementation levels are validated through multiple sources of data and independent raters.

Measures

Dependent Variables

Capacity for interprofessional collaboration was measured with a slightly shortened version of the Interprofessional Socialization and Valuing Scale (ISVS; King et al., 2016). The ISVS asks respondents to indicate the extent to which they possess various competencies for interprofessional collaboration on a 7-point scale (1=not at all; 7=to a very great extent). Items are averaged to create a continuous measure of capacity for interprofessional collaboration. Three items were omitted in this study because they focused on collaboration with clients, not professionals. Inter-item reliability for the 18 items used was excellent (α=0.95)

Capacity for interagency collaboration was captured with a shortened version of the Interorganizational Collaborative Capacity Instrument (ICC; Jansen et al., 2008). The ICC contains 52 items which ask respondents to indicate the extent to which they agree that their agency has various structures, procedures, policies, and resources that support and promote interagency collaboration on a 6-point scale (1=strongly disagree; 3=neither agree nor disagree; 6=strongly agree). Items are averaged to create a continuous measure of capacity for interagency collaboration. Given concerns that the length of the ICC would lead to survey attrition, not all items were used. Ten items were selected for exclusion because psychometric testing previously conducted by the scale developer, which were provided to the researchers of the current study, indicated that their exclusion would not negatively impact scale reliability and, in some instances, would improve it. In the current study, inter-item reliability for the 42 items used was excellent when assessing the capacity for interagency collaboration in each county's court (α=0.97), Department of Human Service office (α=0.97), and behavioral healthcare agency (α=0.98).

Independent Variables

The extent of implementation of best practice standards in each site was measured at an overall level and for each standard separately using the MSIS data. A two-step process was followed to measure the extent of overall implementation. First, the percentage of provisions (n=67) that were fully implemented was calculated for each site. Second, these percentages were divided into tertiles to create a categorical variable with three levels: moderate implementation success (55% to 61% of the provisions fully implemented), moderately high implementation success (73% to 87% of the provisions fully implemented), and high implementation success (92% of the provisions fully implemented). This

two-step process was followed to create a series of categorical variables measuring the extent of implementation of each of the eight standards: (1) Organization and Structure (10 provisions); (2) Role of the Judge (6 provisions); (3) Equity and Inclusion (5 provisions); (4) Early Identification, Screening, and Assessment (5 provisions); (5) Timely, High-Quality, and Appropriate Substance Use Disorder Treatment (12 provisions); (6) Comprehensive Case Management, Services, and Supports for Families (11 provisions); (7) Therapeutic Responses to Behavior (14 provisions); and (8) Monitoring and Evaluation (4 provisions).

Respondent Characteristics

Respondents were asked to provide information about their professional roles and demographic characteristics to better understand whose perspective was represented in the sample. This included the type of agency or program where they worked (e.g., Department of Human Services, Behavioral Healthcare, or courts), their position in their agency or court, whether they served on a dedicated FTC team or FTC steering committee, the number of years they have worked in their position, and the number of years they have worked in their agency or court.

Procedures

Stata Version 18.0 was used to clean and analyze the data. To describe the sample, counts and percentages were calculated for binary and categorical variables, and means and standard deviations were calculated for continuous variables. Multivariate analysis of variance (MANOVA) was used to test for differences between groups (moderate implementation success, moderately high implementation success, and high implementation success) in their capacity for collaboration. MANOVA examines differences between groups in an unobserved composite dependent variable, which is a weighted liner combination of the observed dependent variables (capacity for interprofessional collaboration and capacity for interagency collaboration in the current study). A MANOVA was run for overall implementation and the implementation of each of the eight standards. MANOVA is an omnibus test that can identify whether groups significantly differ from one another, but it provides little information about the nature of the difference. Therefore, if the *p*-value for the MANOVA was statistically significant (< .05) or approached significance (>.05 and < .10), descriptive discriminant analyses and analyses of variance (ANOVA) were run to explore the differences further. Differences that approached significance were included for further examination given that this was an exploratory study and there was concern that the small sample size, particularly the small number of sites, limited statistical power. Descriptive discriminant analysis was used to identify the number of significant discriminant functions (i.e., unobserved composites of the observed dependent variables) that best separated the groups. In addition, standardized function coefficients produced by the descriptive discriminant analyses were examined to assess the overall and relative contributions of the observed dependent variables to the discriminant function (i.e., how capacity for interprofessional collaboration and capacity for interagency collaboration each contributed to group differences). A one-way ANOVA was then conducted for each observed dependent variable to determine which groups differed from one another and how their mean scores for the dependent variables differed (i.e., how they differed in their capacity for interprofessional collaboration and capacity for interagency collaboration).

Findings

Sample Characteristics

Characteristics of respondents in the sample are provided in Table 20. Overall, there was good representation from the three agencies or entities of interest (Department of Human Services, Behavioral Health, and the courts), and a large proportion of the sample had experience working with FTCs, having been a member of a dedicated FTC team and/or a steering committee. Respondents also had sufficient time working in their child welfare system and their agency to provide an informed perspective on their capacity for interprofessional collaboration and their agency's capacity for collaboration.

Table 20: Sample Characteristics (N=162)

Variable		Total Sample
		Mean (SD) or n (%)
Agency type and position	Department of Human Services	78 (48.2)
	Caseworker	49 (62.8)
	Supervisor	16 (20.5)
	Director/Administrator	12 (15.4)
	Other	1 (1.3)
	Behavioral health	50 (30.9)
	Direct treatment provider	34 (68.0)
	Supervisor	5 (10.0)
	FTC coordinator	4 (8.0)
	Director/Administrator	7 (14.0)
	Courts	34 (21.0)
	Attorney/Asst. DA	13 (38.2)
	CASA	9 (26.5)
	Judge	3 (8.8)
	FTC coordinator	2 (5.9)
	Director/Administrator	4 (11.8)
	Other	3 (8.8)
FTC involvement	FTC team member	78 (48.1)
	FTC coordinator	6 (3.7)
	Steering committee member	43 (26.5)
Tenure	Years in position	4.76 (5.46)
	Years in agency/court	6.92 (6.37)
Site	A	19 (11.7)
	B	25 (15.4)
	C	32 (19.8)

D	24 (14.8)
E	16 (9.9)
F	34 (21.0)
G	12 (7.4)

MANOVA Results

Table 21 provides the results of the MANOVAs. The findings indicate that there were significant differences between sites with moderate implementation success, moderately high implementation success, or high implementation success in their capacity for interprofessional and interagency collaboration when examining overall implementation and 4 of the 8 standards (Organization & Structure; Equity & Inclusion; Timely, High-Quality, and Appropriate SUD Treatment; and Comprehensive Case Management, Services, and Supports.

Table 21: Results of MANOVAs and post hoc descriptive discriminant analysis (N=162)

		Descriptive Discriminant Analysis	
Capacity for collaboration	MANOVA test statistic	# of functions	Std. discriminant coefficients
Overall Implementation	$F(4,310)=2.33, p=.056$	1	
Capacity for IPC			-0.45
Capacity for IAC			-0.74
Std 1: Organization & Structure	$F(4,310)=2.07, p=.085$	1	
Capacity for IPC			0.88
Capacity for IAC			0.23
Std 2: Role of the Judge	$F(4,310)=1.15, p=.333$	N/A	
Capacity for IPC			N/A
Capacity for IAC			N/A
Std 3: Equity & Inclusion	$F(4,310)=2.95, p=.021$	1	
Capacity for IPC			0.08
Capacity for IAC			-1.03
Std 4: Early Identification, Screening, & Assessment	$F(4,310)=1.48, p=.209$	N/A	
Capacity for IPC			N/A
Capacity for IAC			N/A
Std 5: Timely, High-Quality, and Appropriate SUD Treatment	$F(4,310)=3.14, p=.015$	1	
Capacity for IPC			0.06
Capacity for IAC			-1.02

Std 6: Comprehensive Case Management, Services, & Supports	$F(4,310)=3.26, p=.012$	1		
Capacity for IPC				-0.76
Capacity for IAC				-0.89
Std 7: Therapeutic Responses to Behavior	$F(4,310)=1.80, p=.128$	N/A		
Capacity for IPC				N/A
Capacity for IAC				N/A
Std 8: Monitoring & Evaluation	$F(4,310)=1.32, p=.262$	N/A		
Capacity for IPC				N/A
Capacity for IAC				N/A

Note: Std=standard; IPC=interprofessional collaboration; IAC=interagency collaboration; Post hoc discriminant analysis conducted if p-value for MANOVA was < 0.10.

Overall and Relative Contributions of Dependent Variables

Table 21 also provides results of the descriptive discriminant analyses which were only conducted if *p* < .10 for the MANOVA. For each area of implementation, there was only one significant discriminant function. Furthermore, for all but one area of implementation, the standardized discriminant function coefficient for capacity for interagency collaboration was larger than the coefficient for capacity for interprofessional collaboration. This indicates that capacity for interagency collaboration contributed to group differences more than capacity for interprofessional collaboration. The reverse was true, however, when examining the implementation of standard 1 (Organization & Structure).

ANOVA Results

The results of the post-hoc ANOVAs, which were only conducted if *p* < .10 for the MANOVA, are contained in Table 22. Means and standard deviations for the dependent variables are provided for all groups and areas of implementation, regardless of the MANOVA results, for descriptive purposes.

There are two noticeable patterns to the ANOVA results. First, with the exception of standard 6 (Comprehensive Case Management, Services, & Supports), the ANOVAs indicated that groups significantly differed in their capacity for interagency collaboration, but no significant difference was found in their capacity for interprofessional collaboration. For standard 6, significant group differences were found for capacity for interagency collaboration and capacity for interprofessional collaboration. This pattern aligns with the results of the descriptive discriminant analysis which suggested that capacity for interagency collaboration had a larger impact than capacity for interprofessional collaboration on group differences, and that the difference between the two types of collaboration was less pronounced for standard 6. It is also noticeable that, in Table 22, the means and standard deviations are higher for capacity for interprofessional collaboration when compared to those for capacity for interagency collaboration. In other words, respondents tended to rate their capacity for interprofessional collaboration higher than their agency's capacity for collaboration, which likely explains why it was not found to be related to implementation success.

The second pattern to the ANOVA results is that, contrary to expectations, for all areas where the ANOVA indicated group differences, the capacity for interagency collaboration was significantly higher in sites with moderate implementation success rather than sites with moderately high or high implementation success. Stated differently, these findings indicate that there was *less* implementation success in the sites with greater capacity for collaboration. It should be noted, however, that for two areas of implementation (standard 3 and standard 6), the capacity for interagency collaboration was also greater in sites with high implementation success compared to sites with moderately high implementation success.

There are several possible explanations for the inverse relationship between capacity for interagency collaboration and implementation success. First, this study relied on respondents' perceptions of collaboration. Perhaps respondents in sites that achieved moderate implementation success have inflated perceptions of their agency's capacity for collaboration. Consequently, respondents in these sites may see less of a need to collaborate when implementing FTCs, which could limit implementation success. Relatedly, respondents in sites that experienced greater implementation success may hold high standards or be more self-critical, leading them to rate their capacity for interagency collaboration lower. Second, given that FTCs require a great deal of cross-systems collaboration, respondents in sites with moderately high or high implementation success may be more likely to experience collaboration challenges and recognize gaps in their capacity for collaboration, which could lead them to have lower assessments of their collaborative capacity. The fact that, for some areas of implementation, (standard 3 and standard 6), sites with high implementation success had greater capacity for interagency collaboration than those with moderately high implementation success, may indicate that these sites have found ways to resolve collaboration challenges and develop policies, structures, or practices that increase their capacity for collaboration.

It should also be noted that the ANOVAs for overall implementation indicated significant group differences in capacity for interagency collaboration even though the *p*-value for the MANOVA for overall implementation approached significance (*p* = .056). However, the ANOVAs for standard 1 (Organization & Structure) did not detect significant group differences despite the fact that the *p*-value for the MANOVA indicated a difference may exist (*p* = .085). Overall, the ANOVA results provide additional support for concluding that capacity for interagency collaboration is related to overall implementation but fail to support the link between capacity for collaboration and the implementation of standard 1.

Table 22: Results of post hoc ANOVAs (N = 162)

Capacity for collaboration	Extent of Implementation success			ANOVA test statistic
	Moderate M (SD)	Moderately high M (SD)	High M (SD)	
Overall Implementation				
Capacity for IPC	6.01 (0.62)	5.74 (0.83)	6.02 (0.84)	$F(2,86)=2.22$, *p*=.115

Capacity for IAC	5.63 (0.80)[MH]	5.20 (0.94)	5.59 (0.95)	$F(2,158)=3.89$, $p=.023$
Std 1: Organization & Structure				
Capacity for IPC	6.01 (0.62)	6.05 (0.76)	5.74 (0.88)	$F(2,83)=2.32$, $p=.105$
Capacity for IAC	5.63 (0.80)	5.42 (0.85)	5.33 (1.03)	$F(2,156)=1.80$, $p=.168$
Std 2: Role of the Judge				
Capacity for IPC	5.84 (0.82)	5.99 (0.61)	6.05 (0.76)	N/A
Capacity for IAC	5.44 (0.98)	5.59 (0.80)	5.42 (0.85)	N/A
Std 3: Equity & Inclusion				
Capacity for IPC	6.00 (0.66)	5.80 (0.83)	5.92 (0.93)	$F(2,54)=1.14$, $p=.328$
Capacity for IAC	5.60 (0.79)[MH]	5.13 (0.99)	5.74 (0.93)[MH]	$F(2,156)=5.70$, $p=.004$
Std 4: Early Identification, Screening, & Assessment				
Capacity for IPC	6.08 (0.59)	5.94 (0.70)	5.83 (0.86)	N/A
Capacity for IAC	5.77 (0.76)	5.48 (0.80)	5.33 (1.01)	N/A
Std 5: Timely, High-Quality, & Appropriate SUD Treatment				
Capacity for IPC	5.98 (0.68)	5.99 (0.82)	5.80 (0.83)	$F(2,159)=1.07$, $p=.347$
Capacity for IAC	5.55 (0.80)[H]	5.79 (0.85)[H]	5.13 (0.99)	$F(2,156)=6.37$, $p=.002$
Std 6: Comprehensive Case Management, Services, & Supports				
Capacity for IPC	6.00 (0.66)[MH]	5.62 (0.83)	6.02 (0.84)	$F(2,159)=3.68$, $p=.028$
Capacity for IAC	5.60 (0.79)[MH]	5.04 (1.00)	5.59 (0.95)[MH]	$F(2,156)=5.31$, $p=.006$
Std 7: Therapeutic Responses to Behavior				
Capacity for IPC	5.99 (0.63)	6.16 (0.62)	5.79 (0.86)	N/A
Capacity for IAC	5.58 (0.82)	5.59 (0.87)	5.37 (0.97)	N/A
Std 8: Monitoring & Evaluation				
Capacity for IPC	5.98 (0.68)	5.76 (0.79)	6.02 (0.84)	N/A
Capacity for IAC	5.55 (0.80)	5.26 (0.99)	5.59 (0.95)	N/A

Notes: Std = standard; IPC = interprofessional collaboration; IAC = interagency collaboration; Post hoc ANOVA conduced if p-value for MANOVA was < 0.10. . [M]Significantly greater than moderate; [MH]Significantly greater than moderately high; [H]Significantly greater than high implementation

Limitations

This study has several limitations which should be considered when interpreting the findings. The small sample size, particularly the small number of sites, reduced the statistical power of the analyses and may have resulted in a lack of statistical significance when examining group differences. The findings may not be generalizable to FTCs outside of the six counties where this study was conducted.

This study used a cross-sectional research design which did not adequately control for many threats to internal validity. Consequently, a causal link between capacity for collaboration and implementation success cannot be established. Finally, items were excluded from some measures to reduce the burden on survey respondents and increase response rates. Although the scales used in the study had excellent inter-item reliability, it is possible that the findings may have been different if all items had been included.

Conclusion

The findings from this study provide insight into factors associated with the implementation of best practice standards for FTCs. They highlight the impact of interagency collaboration on successful implementation, particularly the implementation of policies and practices related to equity and inclusion; timely, high-quality, and appropriate SUD treatment; and comprehensive case management, services, and supports. However, the inverse direction of the relationship challenges the assumption that greater collaboration leads to better implementation. The fact that sites experiencing greater implementation success report less capacity for interagency collaboration may be an indication that these sites are experiencing collaboration challenges and require additional support to sustain their implementation. Future research should aim to develop a better understanding of how collaboration evolves across different stages of implementation, how it is perceived and measured by FTC professionals, and what types of collaborative structures and practices are most effective in supporting long-term success in FTCs.