HHS Public Access
Author manuscript
J Subst Abuse Treat. Author manuscript; available in PMC 2023 September 01.

Published in final edited form as:
J Subst Abuse Treat. 2022 September ; 140: 108829. doi:10.1016/j.jsat.2022.108829.

Using structured implementation interventions to improve referral to substance use treatment among justice-involved youth: Findings from a multisite cluster randomized trial

Steven Belenko[1], **Richard Dembo**[2], **Danica K. Knight**[3], **Katherine S. Elkington**[4], **Gail A. Wasserman**[4], **Angela A. Robertson**[5], **Wayne N. Welsh**[1], **James Schmeidler**[6], **George W. Joe**[3], **Tisha Wiley**[7]

[1]Temple University

[2]University of South Florida

[3]Texas Christian University

[4]Columbia University and New York State Psychiatric Institute

[5]Mississippi State University

[6]Icahn School of Medicine at Mount Sinai

[7]National Institute on Drug Abuse

Abstract

Introduction.—Youth involved in the justice system have high rates of alcohol and other drug use, but limited treatment engagement. JJ-TRIALS tested implementation activities with community supervision (CS) and behavioral health (BH) agencies to improve screening,

CORRESPONDING AUTHOR: Steven Belenko.

Statement of Authors' Contributions
Steven Belenko oversaw development of the entire manuscript, wrote sections of the article, assisted with data analyses, and did general editing. Richard Dembo had main responsibility for the two-level analyses, assisted with the analyses testing the main hypotheses, drafted parts of the Methods and Results sections, and did general editing. Danica Knight wrote sections of the Introduction, Methods, and Discussion sections, advised on the data analyses, and did general editing. Katherine Elkington drafted sections of the Introduction and Discussion sections, assisted in revising the Introduction, and did general editing. Gail Wasserman drafted sections of the Introduction and Discussion sections, advised on some of the data analyses, and did general editing. Angela Robertson drafted sections of the Introduction and Discussion sections and helped edit parts of the Methods section. Wayne Welsh drafted parts of the Methods section, and did general editing. James Schmeidler assisted with the analyses of the main hypotheses, and helped draft and edit sections of the Methods and Results sections. George Joe designed and conducted the analyses of the main hypotheses, and drafted sections of the Methods and Results section. Tisha Wiley drafted sections of the Discussion section.

ClinicalTrials.gov registration # NCT02672150

Credit Author Statement
Conception and design of study: S. Belenko, D. Knight, G. Wasserman, T. Wiley, A. Robertson, K. Elkington
acquisition of data: S. Belenko, D. Knight, G. Wasserman, A. Robertson, R. Dembo
analysis and/or interpretation of data: S. Belenko, R. Dembo, G. Joe, J. Schmeidler
Drafting the manuscript: S. Belenko, R. Dembo, D. Knight, G. Wasserman, K. Elkington, W. Welsh, T. Wiley, A. Robertson, G. Joe, J. Schmeidler
revising the manuscript critically for important intellectual content: S. Belenko, R. Dembo, G. Joe, J. Schmeidler, G. Wasserman, K. Elkington.

identification of substance use service need, referral, and treatment initiation and engagement, guided by the BH Services Cascade and EPIS frameworks. This paper summarizes intervention impacts on referrals to treatment among youth on CS.

Methods.—This multisite cluster-randomized trial involved 18 matched pairs of sites in 36 counties in seven states randomly assigned to core or enhanced conditions after implementing the core intervention at all sites for six months. Enhanced sites received external facilitation for local change team activities to reduce unmet treatment needs; Core sites were encouraged to form interagency workgroups. The dependent variable was percentage referred to treatment among youth in need (N=14,012). Two-level Bayesian regression assessed factors predicting referral across all sites and time periods. Generalized linear mixed models using logit transformation tested two hypotheses: (H1) referrals will increase from baseline to the experimental period, (H2) referral increases will be larger in enhanced sites than in core sites.

Results.—Although the intervention significantly increased referral, condition did not significantly predict referral across all time periods. Youth who tested drug positive, had an alcohol/other drug–related or felony charge, were placed in secure detention or assigned more intensive supervision, or who were White were more likely to be referred. H1 ($p < .05$) and H2 ($p < .0001$) were both significant in the hypothesized direction. Interaction analyses comparing site pair differences showed that findings were not consistent across sites.

Conclusions.—The percentage of youth referred to treatment increased compared with baseline overall, and enhanced sites showed larger increases in referrals over time. However, variations in effects suggest that site-level differences were important. Researchers should carry out mixed methods studies to further understand reasons for the inconsistent findings within randomized site pairs, and how to further improve treatment referrals across CS and BH systems. Findings also highlight that even when CS agencies work collaboratively with BH providers to improve referrals, most justice-involved youth who need SU services are not referred.

Keywords

Juvenile justice; Services cascade; Implementation science; Treatment referral

1. Introduction

Youth involved in the justice system have substantially higher rates of alcohol, marijuana, and cocaine use compared with youth in the general population (Miech et al., 2018; Mieczkowski et al., 1998; Mulvey et al., 2010). Likewise, the prevalence of substance use disorders (SUD) is substantially higher among justice-involved youth: 25%–50% of justice-involved youth (McClelland et al., 2004; Teplin et al., 2002; Wasserman et al., 2010; White et al., 2019) report SUD compared to 4.3% of adolescents in the community aged 12–17 (SAMHSA, 2017).

Among justice-involved youth, untreated behavioral health (BH) disorders increase the likelihood of continued contact with both juvenile and adult justice systems (Cuellar et al., 2006; Hoeve, McReynolds & Wasserman, 2013), as well as the escalating severity of offenses (Hoeve, McReynolds, McMillan & Wasserman, 2013). Yet research shows that these youth often do not access treatment, even after identification of need. Across a range of justice settings (community supervision, detention, and incarceration) only 50% to 84% of those with BH needs access services (Aalsma et al., 2012; Abram et al., 2008; Chapman et al., 2006; Colins et al., 2010; Fazel et al., 2008; Sedlak & Bruce, 2010; Teplin et al., 2005; Wasserman et al., 2021; White et al., 2016); access rates for substance use–related services are even lower than for mental health services (White et al., 2019).

We know less about the BH service needs of the subpopulation of youth in the juvenile justice system (JJS) who are on community supervision (CS; e.g., probation). But data from CS agencies showed that 70% of youths were screened, more than half were in need of treatment, but only about a fifth of those in need were referred to treatment (Wasserman et al., 2021). These findings are cause for alarm because 75% of justice-involved youth are supervised in the community (Hockenberry & Puzzanchera, 2019). The field, thus, needs effective strategies for improving substance use (SU) identification and linkage to treatment services that can inform efforts to more efficiently and effectively address the SUD needs of justice-involved youth (Belenko et al., 2017). The current paper reports key findings from the Juvenile Justice-Translational Research on Interventions for Adolescents in the Legal System (JJ-TRIALS) multisite project, which tested the impacts of a structured implementation intervention on unmet substance use service needs among adolescents under CS (Knight et al., 2016; Wiley et al., 2015); see project description below.

1.1. The Behavioral Health Services Cascade as a framework for system change

The Behavioral Health Services Cascade (hereinafter the Cascade) provides a guiding framework to facilitate research and practice about how best to identify and address SU treatment need within and across agencies, and to document progress toward increasing referral and subsequent service receipt (Belenko et al., 2017; Dennis et al., 2019). The Cascade comprises six distinct interrelated activities that are essential for identifying SU problems and linking youth to clinical services. These activities include screening and assessment, identification of need, referral to treatment, initiation, engagement, and continuing care. As in a typical cascading service system, receipt of services in later steps is typically contingent on receipt of prior steps (e.g., only those in need should be referred). In this way, the Cascade also provides a measurement framework to assess whether interventions are able to reduce unmet service needs (Dennis et al., 2019; Knight et al., 2016).

The idea that as youth transition across service sectors (from identification of need and referral in the JJS to initiation in the treatment system), coordination is required between the justice system and treatment providers is implicit in the Cascade (Welsh et al., 2021). Also implicit is that interventions can be developed and tested that seek to increase the proportion of youth moving through the Cascade. At each step of the Cascade, a clear need exists for rethinking how agencies approach the identification and management of justice-involved youth's SU needs, alone and in collaboration with their local partners (Belenko et al., 2017; Wasserman et al., 2021), particularly across service systems.

1.2 How youth under community supervision move through the Cascade

Individual juvenile probation officers (PO) are typically responsible for ensuring compliance with court recommendations for youth placed on CS, functioning as gatekeepers (Stiffman et al., 2004), and determine how and where treatment recommendations will be carried out (unless an agency has a specific drug and alcohol unit or caseloads). CS agencies utilize written documents or state-mandated policy to guide these decisions, or rely mostly on PO discretion (Bowser et al., 2018), resulting in broad individual and agency variation in Cascade-related practices (Belenko et al., 2018; Knight et al., 2019). Unfortunately, POs are often insufficiently trained on identification and management of SUDs and have limited knowledge about local adolescent treatment programs.

A complex interplay of organizational, staff, and youth factors is associated with Cascade activities (Davis et al., 2016; Fagan et al., 2015; Godley et al., 2005; Karriker-Jaffe, 2011; Wasserman et al., 2021). For example, most POs do not routinely use evidence-based active treatment referral strategies (e.g., making the initial appointment, providing transportation; Knight et al., 2019), but are more likely to do so in organizations that value innovation and flexibility (Knight et al., 2019). Other Cascade activities are differentially influenced by staff training and knowledge/attitudes as well as youth-specific factors (e.g., age, supervision status) and community systems (e.g., availability of adolescent treatment, insurance, resources; Belenko et al., 2017; Taxman & Belenko, 2012; Wasserman et al., 2021).

Training CS staff in best practices on SU and treatment is a necessary but not sufficient strategy. Compared with those who are trained, untrained POs are less likely to recommend services for youth charged with alcohol- or drug-related offenses (Hoeve et al., 2014), who have prior offenses (Wasserman et al., 2008), or with high risk for recidivism (White, 2019). Lack of training reduces effective juvenile probation case management and appropriate provision of services for targeting criminogenic needs (Haqanee et al., 2015). Agencies participating in a systems-level intervention, including evidence-based screening, BH training, inter-agency agreements, and materials to facilitate referrals, increased referrals and referred youth were almost three times as likely to access services than before the intervention (Wasserman et al., 2009). When POs were trained to interpret and report on results of clinical assessments, almost half the youth on their caseloads received treatment referrals; among substance use disordered youth, those referred to services were 66% less likely to recidivate than those not referred (Hoeve et al,. 2014). These studies suggest that training on appropriate identification and linkage strategies, coupled with agency support for such policies, can improve service receipt among youth on CS.

1.3. Multilevel factors are likely to affect Cascade outcomes

Analyses of Cascade outcomes using baseline data from the period prior to initiation of the JJ-TRIALS study found that both youth- and site-level factors were related to these outcomes (Wasserman et al., 2021). Multivariate multilevel regression analyses found that male, older, and white youth were more likely to be in need of treatment, as were those under stricter JJ supervision. Referral was predicted mainly by being under a higher level of JJ supervision. In addition to youth factors, agency- or community-level factors likely exist that influence Cascade outcomes, reflecting the complex impacts of systems, organizational, staff, and youth characteristics (Davis et al., 2016; Fagan et al., 2015; Godley et al., 2005; Karriker-Jaffe, 2011; Wasserman et al., 2021). For example, screening, determination of treatment need, and referral to treatment may be influenced by staff training and knowledge/ attitudes related to SUDs and treatment, as well as community systems issues (e.g., availability of adolescent treatment, insurance, resources; Belenko et al., 2017; Taxman & Belenko, 2012; Wasserman et al., 2021).

1.4 Implementation science and implementation interventions: Guiding organizational and systems change

Because justice-involved youth with SU issues must move from the JJ system to the BH system for SU treatment, improving SU identification and service linkage involves a systems shift in how SU-related services are coordinated across and within agencies (Welsh et al., 2021). Prior work shows the importance of cross-service-sector collaborations to address SU, mental health, and HIV service needs of adult and juvenile justice populations (Belenko, Visher, et al., 2013; Elkington et al., 2020; Friedmann et al., 2013; Pearson et al., 2014). To achieve this, individuals representing both CS and BH agencies must work together to develop processes for determining and executing cross-agency referrals, recording receipt of services, and identifying gaps in service receipt (at an agency or community level). However, cross-system collaboration is often hindered by conflicting goals of justice and treatment systems (e.g., zero-tolerance supervision vs. harm reduction approaches; Fletcher et al, 2009; Lehman et al., 2009), and successful collaboration requires effective communication and shared goals across agencies that likely operate under distinct cultures (Belenko et al., 2013; Henderson & Taxman, 2009; Roman et al., 2011; Welsh et al., 2021).

The justice system has relied upon implementation science to strengthen collaborations with treatment systems and to promote innovative service practices for substance disordered adults (Belenko et al., 2013; Friedmann et al., 2012; Shafer et al., 2014;). Promising implementation approaches that address barriers to cross-system collaboration and service access include process improvement strategies such as local change teams (LCTs) and data-driven decision-making (DDDM). LCTs that include external facilitation and involve multiple agencies and staff (Aarons et al., 2014; Fixsen et al., 2009; Kets de Vries, 2005; Marchant & Young, 2001; Powell et al., 2012) are well-suited to implementation of cross-system initiatives in which organizational cultures and varying approaches to collaboration may be barriers to practice change (Belenko, Visher, et al., 2013; Hoffman et al., 2012).

Similarly, by focusing on systematic data collection and interpretation, as in DDDM, an agency can monitor practice improvement to address problems and implement change (Knight et al., 2016; Marsh et al., 2006). JJ settings have used DDDM to guide system-wide reform to reduce recidivism and system costs (Chayt, 2012; DART, n.d.; Dwyer et al., 2012).

The JJ-TRIALS study was informed by the above considerations in developing and testing whether a set of implementation activities involving staff of CS and BH agencies would improve screening, identification of SU service need, referral, treatment initiation, and treatment engagement among youth under CS (Knight et al., 2016). The study used the Exploration, Preparation, Implementation, and Sustainment (EPIS) framework (Aarons et al., 2011; Becan et al., 2018). EPIS has considerable empirical support and utilization in cross-system initiatives, and can identify the necessary structures and processes within a given system to support implementation of evidence-based practice (Moullin et al., 2019). It comprises four sequential phases of activities: *exploration* (determining population needs and specific programs available); *preparation* (planning how to integrate the program into existing organizational workflow); *implementation* (implementing the program with fidelity); and *sustainment* (establishing the program as part of ongoing activities). The EPIS framework also considers the multilevel nature of service systems and addresses outer context (e.g., county funding) and inner context (e.g., organizational functioning, staff attitudes) across each stage. Examining inner and outer context allows one to explore barriers to implementation and sustainability within and across settings, yielding lessons for broader implementation, dissemination, and scale-up.

This paper focuses on the early part of the Cascade, primarily occurring within the JJS, focusing on screening, determination of treatment need, and referral as primary outcomes. The JJ-TRIALS study design (see Methods) was a cluster-randomized trial involving CS agencies in 36 counties in 7 states. The core condition included five interventions implemented at all sites during a 6-month, pre-randomization period (Knight et al., 2016): (1) staff orientation meetings, (2) needs assessment/system mapping, (3) behavioral health training, (4) site feedback reports, and (5) goal achievement training. Following these core activities, the coordinating center randomly assigned matched pairs of sites within states into core or enhanced conditions. The study matched pairs of sites within states to achieve equivalency based on local population, number of youth referred to the JJS, and number of agency staff. The study used optimal randomization to find the most balanced pattern of assignment across sites within each RC. To do so, the study team ran 10,000 permutations of possible site assignments to study condition within each pair. For each of these, the study computed multivariate Hotelling's T^2 to assess the degree of balance both within and across all RCs. The study selected the final randomization design from a pool of the top 2% of permutations balancing across all characteristics. The JJ-TRIALS research centers were blinded as to assignment until after randomization occurred

Enhanced sites received external facilitation for local change team activities, whereas core sites were encouraged to form interagency workgroups to address their process improvement goals. External facilitators worked with the LCT to lead team meetings, provide encouragement and feedback, and assist the LCTs with addressing their goals and completing their activities (see Knight et al., 2016 for additional details on the study design). Previous research has demonstrated the value of the implementation strategies used in JJ-TRIALS for improving client outcomes in various service settings (see e.g., Aarons et al., 2014; Belenko et al., 2013; Dwyer et al., 2012; Elkington et al., 2020; Hoffman et al., 2012; Powell et al., 2012).

The goals of this paper are to (1) assess the impact of the core intervention on Cascade outcomes through referral, (2) assess the additional impact of external facilitation (enhanced intervention) on Cascade outcomes through referral, and (3) analyze changes in Cascade outcomes over time. Two hypotheses focus on the percentage of CS youth in need of treatment who are referred for SU treatment by CS agencies: *Hypothesis 1*: Compared with the baseline period, the percentage of youth in need who are referred to treatment will increase during the experimental period; *Hypothesis 2*: Compared with the core intervention sites, the percentage of youth in need who are referred to treatment will be greater in enhanced intervention sites (over time relative to baseline) during the experimental period.

2. Methods

The JJ-TRIALS Cooperative included six research centers (RC), one coordinating center (CC), and one funding/oversight entity (NIDA), with participation by juvenile justice agency partners in seven states. Each RC's Institutional Review Board and state/local juvenile justice agencies reviewed and approved all protocols.

2.1. Overview of the JJ-TRIALS Protocol

JJ-TRIALS sought to: (1) increase identification of SU and linkage to treatment among justice-involved youth and (2) test the effectiveness of two bundles of implementation strategies aimed at increasing screening, assessment, referral, and treatment (Knight et al., 2016). The Cascade (Belenko et al., 2017) guided intervention content and operationalization of agency/youth outcomes while EPIS (Aarons et al., 2011) informed selection of intervention components (e.g., needs assessment during an exploration phase and external facilitation during an implementation phase) and general study design (e.g., assessment timing and content; Becan et al., 2018).

Guided by a site feedback report prepared by each RC, sites selected process improvement goals and were trained by the research team on using data to inform this goal selection; the research team developed DDDM templates and tools for staff to use in implementing change. Site workgroups had flexibility in selecting the Cascade activity on which to focus. Most chose to target services deeper into the Cascade, specifically referral (39%) and treatment initiation (39%); more specifically, 47% sought to standardize referral procedures and 44% wanted to improve awareness/training on referral practices (Becan et al., 2020).

Of the original 36 sites, one site withdrew from the study and one site had very limited Cascade data (their matched sites were also dropped), and two sites primarily served detention populations. Thus, 30 sites (15 matched pairs) were available for the analyses (with the exception of the two-level analysis of referral predictors as described below).

2.2. Data sources and data management

We utilized five data sources, corresponding to different components of the EPIS model: U.S. Census data (outer context), staff survey data (inner context), monthly site check-in (MSC) data (inner context), management report data (inner context), and youth service record data obtained from probation agencies (Cascade outcomes).

2.2.1. U.S. Census data—The study assessed community disadvantage using several county-level variables. The research team obtained county-level measures on poverty and urbanicity from U.S. census data, including percentage of families with children under

18 years of age living in poverty (2015 data), percentage of children under age 18 years who were uninsured (2015 data), and percentage of residents in urban areas (2010 census; Wasserman et al., 2021).

2.2.2. Staff survey data—We surveyed staff on their perceptions of their organization's climate, the importance of interagency collaboration, and beliefs about and use of practices regarding SU screening, assessment, referral, and support for treatment (Knight et al., 2019). Survey data collected at baseline from CS staff at each site are included in the analyses in this paper. Survey domains and sample items can be found in Knight et al. (2016). Details on survey procedures can be found in Knight et al. (2016) and Knight et al. (2019). A total of 462 staff in 30 sites completed surveys (79.0% of those who consented). Of these, 69.5% worked in a CS unit (e.g., juvenile probation office), 12.0% in a behavioral unit operated by the JJ agency, and 18.5% worked in an undefined unit within the CS agency.

2.2.3. Monthly site check-in (MSC) data—Following the baseline period, monthly site check-in (MSC) phone calls by RC staff with a representative from each site continued through the end of the study. The MSC documented site activities reflecting progress toward achieving each agency's selected goals, collected information about process improvement efforts, and monitored any changes to services based on data systems, staffing, and policy modifications (Knight et al., 2016). On average, sites completed a mean of 17.1 months of MSCs out of the maximum 20 months.

2.2.4. Site staff management data—A project management report (Knight et al., 2016) tracked key elements of participation in various study events over time for each site, including the number of attendees at the leadership orientation meeting and the number and percentage of CS staff who completed BH trainings provided by the research team to agency staff.

2.2.5. Youth service records—Between March 2014 and November 2017, we obtained deidentified youth service record data (e.g., movement along the Cascade) from CS agencies for 29,049 youth entering probation under CS in 30 counties in 7 states (Dennis et al., 2019)[i]. All youth entering the sample during the data collection period had to have at least one referral for a new offense, and the study excluded them if they had a previous episode of CS ending fewer than 30 days earlier (to avoid overlap in services or sanctions associated with prior probation episodes). The study also excluded youth who were diverted from probation without receiving any CS. Records were further excluded for youth with fewer than 90 days between referral to JJ and the final record abstraction for the site, as they lacked opportunity to move through the entire Cascade. The study had no other exclusion criteria.

Dennis et al. (2019) provides a description of the youth sample. Briefly, 73% of the sample was male; 63% were aged 15–17 and 36% 11–14; 49% were white, 47% Black, and 21%

[i]Although 36 sites initially participated, one site withdrew prior to randomization to study condition, and its matched pair was excluded from the analyses; referral data were not available for another site so it and its matched pair were excluded. Two other sites were dropped because subjects were primarily held in secure detention rather than being under community supervision.

Hispanic; 30% were charged with a property crime, 25% with a violent crime, and 14% with a drug or alcohol crime; and 56% were charged with a misdemeanor, 33% with a felony.

The study extracted a set of 72 measures related to demographics, charges, case processing, recidivism risk, and Cascade outcomes from a mix of electronic databases plus either scanned or physical paper probation records, using common definitions and specifications (see Dennis et al. [2019] for details on data cleaning and data management procedures). Inconsistencies occurred across sites in the completeness and quality of youth record data; the median number of variables available per record was 49 out of 72 (68 percent; Dennis et al., 2019).

2.3. Dependent variables: Stages of the Behavioral Health Services Cascade

This article focuses on three stages of the Cascade: screened, in need of treatment, and referred to treatment. *Screened*: Any indication in the youth record that a formal screening instrument was administered; if the screening results indicated a potential SU problem, the study coded this variable 1, if not we coded as 0. *In need of treatment*: The youth record indicated a need for SU treatment based on the results of a screening instrument, positive drug test, clinical assessment, or other sources of information such as self-disclosure of substance use. *Referred to treatment*: Among youth in need of treatment, an indication in the youth record that he/she was referred by the probation agency to SU treatment at any point during their supervision. This was the main dependent variable in the analyses (identified need and referred = 1, Identified need but not referred = 0). A total of 14,012 youth records across the 30 sites indicated a need for treatment.

2.3.1. Independent variable—The main independent variable was study condition (core vs. enhanced) to which the study randomized each of the 18 matched site pairs within states. The team based matching on the number of youth aged 10–19 in the general population in the county, number of youth on CS, number of CS staff, and whether site agencies used standardized screening/assessment and evidence-based treatment. Randomization occurred at two levels (a) one of three project start dates, and (b) study condition (see Knight et al. [2016] for details in randomization procedures).

2.3.2. Covariates—Youth-level and site- and county-level covariates were selected based on the conceptual and research literature discussed earlier, and informed by previous analyses of JJ-TRIALS data (Robertson et al., 2020; Wasserman et al., 2021).

Youth level: Within-level predictor variables included: being adjudicated delinquent, screened positive for SU, age at time of referral to probation, gender, African American, Hispanic, alcohol/other drug charge, placed in secure detention for the current offense, maximum charge level (felony vs. misdemeanor), and supervision level (“more” = on formal probation supervision, parole, or juvenile drug court; all other statuses were coded as “less” (e.g., diversion, informal probation).

Site and county level: Between-level predictors at the county level included the two community disadvantage factors derived from exploratory factor analyses using U.S. Census data (see below). At the site (or agency) level, measures included (1) the four organizational climate, interorganizational relationship, program needs, and use of referral to treatment factors, reduced from the baseline staff survey items (see Knight et al., 2019), (2) percentage of staff completing BH training, (3) number of months staff working with local change team members were added, (4) number of months with no reported staffing factors/issues, (5) number of months with BH program staff attendance at meetings, and (6) attendance at leadership orientation meeting. The study collected measures 3 through 5 through the MSC interviews (Becan et al., 2020), and measures 2 and 6 from the final management reports (Knight et al., 2016).

Time period: The research team divided the study into several phases based on the implementation stage of the experiment. The baseline period was approximately six months (varying slightly by site depending on data availability) prior to the onset of the JJ-TRIALS project. The pre-randomization period was approximately six months, when the pre-randomization intervention was implemented at each site. The experimental period began when sites were randomized into core and enhanced conditions. This was further divided into early experimental (first 6 months), late experimental (last 6 months), and maintenance (6 months) periods. The rationale for this was that it was likely to take some time before the site staff teams could develop and implement their new processes or procedures. We anticipated that any impacts on Cascade outcomes would not be noticeable until later in the experimental period. Finally, maintenance was a six-month period after the formal experiment ended; the RCs continued to conduct monthly site check-ins and collect youth data, but without active research involvement. Study staff extracted records data for youth entering CS during all these time periods.

2.4. Handling missing data

To include as many records as possible, the study team included blank data fields for the Cascade stages of screening, in need of treatment, and referral as "no" responses. This represents a "lower bound" conservative estimate because blank fields could represent a true negative (e.g., referral not needed, no-referral of someone in need of treatment) or true missing (e.g., referral was made but not entered), and were used for the main analyses. We developed a second, "upper bound" estimate for these Cascade measures for sensitivity analyses using hot deck imputation to replace missing data (Little & Rubin, 2019). Missing data were replaced with the median of the nearest 20 valid (nonmissing) values, which provided unbiased estimates of the mean and standard error at the group level; see Dennis et al., 2019 for more details on these procedures. Hot deck imputation is one of the more widely used and accepted methods for missing data replacement when using large survey and other epidemiological data sets (Andridge & Little, 2010; Little & Rubin, 2019). It is particularly useful when many variables exist, each of which might require different analytic models and assumptions, for which there is not always a basis (or data) to make accurately in multisite research. Comparisons of multiple methods of replacing missing data (Staviseth et al 2019; Little & Rubin, 2019) have found that they performed similarly well when sample sizes were large ($n \geq 1000$) as we have here, and that more advanced imputation methods only achieved a consistent advantage when sample sizes were small ($n \leq 200$).

We also assessed the extent to which youth record data were missing at random to justify imputation, comparing inter-item correlations of the lower bound estimates (not imputed) and the upper bound (imputed) for the variables used in the imputation. Across 81 comparisons, the inter-item correlations for the Cascade steps differed by r=0.10 or less, providing reasonable evidence to meet the assumptions that data were missing at random.

2.5. Analysis plan

The analyses proceeded in several steps: (1) exploratory factor analyses (EFA) of between-level variables reflecting (a) social-community disadvantage and (b) organizational climate, interorganizational relationships, program needs, and use of referral; (2) percentages of substance-using youth referred to treatment compared across sites; (3) overall referral percentages compared across the five study time periods; (4) two-level analysis testing the overall effects of providing facilitation to local change teams in addition to the core intervention; and (5) multilevel analyses testing the two main hypotheses of changes in referral over time and interactions with matched site pairs. The study team used SPSS v25 to conduct the EFAs, and produce the referral percentage by site and experimental time period descriptive tables. We used Mplus 8.3 (Muthén & Muthén, 1998–2018) to conduct the two-level Bayesian regression analysis (Heck & Thomas, 2015; Koehrsen, 2018; Lynch, 2010) and SAS v9.4 Proc GLIMMIX to perform the multilevel condition by time interaction analyses.

2.5.1. Data reduction—The EFAs involved principal axis factoring (PAF), which uses estimated communalities in the major diagonal. This initial estimate assumes the communality of each variable is equal to the square multiple regression coefficient of that variable with the other variables. The PAF factoring method replaces the main diagonal of the correlation matrix (which consists of all 1s) by these initial communality estimates. PAF is then applied to this revised version of the correlation matrix. From the EFA results, we selected factors with eigenvalues greater than 1.0 and that explained the predominant amount of variance in each EFA model. To improve interpretation of the factor structure, we performed an oblique rotation of the factor solution, which permitted correlated factors. The study then calculated regression factor scores and saved them for further use in our analyses.

The research team conducted an EFA, using principal axis factoring (PAF) and oblique rotation, on the community disadvantage variables. Results indicated a two-factor solution, explaining 69.4% of the variance, best fit the data and was meaningfully interpreted. Pattern matrix results indicated percent of families with children under the poverty line in 2015, percent of families receiving SNAP/food stamp benefits, and percent of single parent households with children in 2015 loaded highly on Factor 1, *poverty*. Variables reflecting percent of families with no health insurance in 2015, percent of adults who were high school graduates as of 2015, percent urban in 2010, and aggregate Medicare spending in the county per capita in 2015 reflected a second factor labelled *health-related issues* (Factor 1-Factor 2 correlation, r=-0.093). The Kaiser-Meyer-Olkin measure of sampling adequacy (KMO) was 0.56, indicating minimal factor adequacy.[ii]

We conducted a similar EFA on the measures of organizational climate, interorganizational relationship, program needs, and use of referral measures. A four-factor solution, accounting for 69.9% of the variance, best fit the data and was meaningfully interpreted. Pattern matrix review indicated a Factor 1 related to *organizational climate* (e.g., performance feedback, communication); Factor 2 reflected *interorganizational relationships* (e.g., effectiveness of relationships, quality of communication); Factor 3 related to *innovation* (e.g., adaptability, need for program improvement); Factor 3 was *program needs*; Factor 4 *(Referral)* reflected referral activities (e.g., use of active referrals, use of standard referrals). The average correlation among the four factors was low (r=0.001). The factor solution barely met minimal standards for factor adequacy (0= 0.49); hence, caution in interpreting the results is indicated.[iii]

2.5.2. Two-level regression analysis of predictors of referral—Preliminary two-level analysis (within level = youth factors, between level = site factors) found an intraclass correlation coefficient (ICC) of 0.246 for study condition, and a design effect (DS) of 86.5 (average cluster size=348.6); the latter indicates clustering in the data needs to be considered (Muthén & Muthén, 2018). Accordingly, we conducted a two-level logistic regression analysis examining the factors affecting referral to treatment, a dichotomous outcome variable. These analyses were at the case level, meaning that if a youth received a new JJ referral during the study period that was handled as a new case; 17% of youth had more than one JJ referral.

Initial estimation of the two-level model using linear regression (a widely used frequentist method) could not be achieved due to no within-cluster (site) variation for several predictor variables. Therefore, we conducted a Bayesian two-level regression analysis of the predictors of referral. For each coefficient in the regression model, its Bayesian one-tailed p-value is a test of the direction of an effect, obtained from an estimation procedure that assumes the null hypothesis is false (see Marsman & Wagenmakers, 2017). The one-tailed p-value aids in interpretation and inference and does not gauge model fit.

To assess fit for the Bayesian analysis, which uses a Markov Chain Monte Carlo (MCMC) iterative process to estimate the model, we evaluated the results in two ways. First, to make sure the MCMC algorithm produced a Markov chain that converged "to the appropriate density (the posterior density) and that "'mixes' well throughout the support of the density" (Lynch, 2010:132), we obtained a potential scale reduction factor (PSR) developed by Gelman and Rubin (1992a, b). A PSR less than 1.1 is considered good, and a PSR of 1.000 is ideal (Lynch, 2010; Zitzmann & Hecht, 2019). The PSR for our two-level Bayesian regression model was 1.001. To ensure that we obtained convergence, we re-estimated the model with double the number of iterations, with results still indicating convergence. Second, we obtained an estimate of the fit of the model to the data by obtaining a posterior predictive p-value (PPP), which is a test of overall model fit (Depaoli, 2021). Unlike frequentist estimates derived from maximizing the likelihood function, Bayesian estimates

[ii]Due to space concerns, detailed tables reporting these results are omitted. Copies of these results can be obtained from the senior author upon request.
[iii]Due to space concerns, detailed tables reporting these results are omitted. Copies of these results can be obtained from the senior author upon request.

are obtained from the posterior distribution (Koehrsen, 2018: 6). In Bayesian analysis, PPP is a natural byproduct of the MCMC approximation, calculated using posterior predictive distributions of the same sample size and of the same likelihood as the original data. At each MCMC iteration, a new set of data is generated based on updated parameter estimates. An excellent-fitting model is expected to have a PPP value around 0.5 (Lynch, 2010), and a low PPP indicates that the model is not appropriate for the data and that there is misspecification (Asparouhov & Muthén, 2010b).

Preliminary analysis indicated sizable within cluster (site) variation for several variables planned for inclusion in the two-level model. For example, screened positive for drugs, a critical variable in our analysis, had an overall missing value rate of 15%, which ranged from 0% to 73% cross sites. These findings raised concern that the missing data were not MAR and could not appropriately be estimated by direct maximum likelihood or data imputation procedures. As Jakobsen et al. (2017:9) assert, when “data are MNAR, no methods exist to handle missing data appropriately”.

Accordingly, we used listwise deletion for several reasons: (1) listwise deletion is “less biased than multiple imputation or maximum likelihood when data are missing on predictor variables in regression analysis” (Allison, 2002; Allison, 2014, p. 1); (2) an included variable with a substantial amount of missing data (drug test results) was considered a critical variable in the analysis, as it produced highly valid data. Even though listwise deletion reduced the data set by 51%, resulting in some loss of statistical power and generalizability, the number of remaining cases (n=6972) provided sufficient statistical power to reliably estimate effects (Brown et al., 2009). The number of cases per site ranged from 19 to 1004, with, as noted earlier, a mean site size of 348.6. The study team dropped these data from three sites from this analysis due to substantial missing data on several covariates, leaving 27 sites for these analyses.

2.5.3 Tests of interactions and time period: Main hypotheses—Because the youth (level 1) were nested within the treatment sites (level 2), we used a multilevel analysis to test the hypothesis that referral to treatment for those in need was differentially affected by intervention type over time. The factors defining the independent variables were intervention (core vs. enhanced) and project implementation phase (baseline, pre-randomization, early experimental, late experimental, and maintenance). Because the dependent variable was a dichotomy (referral to treatment or not) and the data were nested within paired treatment sites, the study used a generalized linear mixed model to test the research hypotheses related to intervention and implementation phase.

We employed a multilevel logit model (Anderson & Aitkin, 1985; Wong & Mason, 1985). The logic for using these models is similar to that for nonbinary multilevel models (Bryk & Raudenbush, 1992). These include correction for biases in the estimation of the parameters due to clustering of subjects within the groups; calculation of correct standard errors, confidence intervals, and significance tests; and the correct estimates of the variances and covariances of random effects, which enables the decomposition of the total variance in the outcome variable into portions associated with each level (Guo & Zhao, 2000).[iv]

For multilevel models of binary dependent variables, two common estimation methods exist (Guo & Zhao (2000): marginal quasi-likelihood (MQL) and penalized quasi-likelihood (PQL). Parameter estimation for binary data models can be downwardly biased, especially with commercial packages such as HLM and MLn, but analyses using PQL will be relatively accurate (Guo & Zhao, 2000). This procedure takes into consideration both under-dispersion and over-dispersion of a parameter used in calculating the conditional variance, where under- or over-dispersion can lead to unreliable estimates of the standard errors (Gui & Zhao, 2000).

In contrast to the two-level analyses described earlier, for youth who had multiple entries during the observation period (17% of all youth), we selected the first probation episode. This strategy avoided effects on subsequent screening and referral outcomes, but reduced the unduplicated youth sample to 10,514 in need of treatment. For the primary dependent variable (referral to treatment of a youth in need), an initial analysis without independent variables estimated the ICC, the proportion of variance due to site differences, as 0.204, indicating the 20.4% of the variance in the dependent variable was between sites.

The independent variables were *sitepair* (15 pairs of sites) *intervention* (core vs. enhanced condition), and two comparisons among implementation time periods (baseline, early experimental, late experimental, maintenance; the pre-randomization period was excluded). One analysis compared baseline against the combined later three phases, a dichotomy termed *time*. A second analysis compared referrals among the three postrandomization phases, which was termed *laterphases*, to further examine how the intervention affected referrals over time.

We initially performed the analyses with the intercepts as random (allowing for differences in site means) and all independent variables as fixed effects (not shown), and then conducted a second analysis with intercepts and *sitepair* random, thereby controlling for unobserved heterogeneity. For this analysis with *sitepair* random, the type 3 fixed effects tests of *sitepair*, *intervention*, and *intervention*sitepair* reflected extrapolation to a larger population of matched pairs of sites. The analysis for implementation periods provided tests of two hypotheses: (1) combined experimental period phases will have higher rates of referral to treatment than baseline referral of youth in need (time effect); (2) the difference between referral rates of the combined experimental periods and baseline will be larger in the enhanced than core intervention conditions (*time*intervention* interaction).

A wide variation existed in referral rates among sites (Appendix A), attributable to the impact of *sitepair*, *intervention*, and *intervention*sitepair*. We conducted several analyses to explore whether interactions among these factors might also affect referrals. An interaction with a mean square that is substantially smaller—not similar to or larger—than an effect implies the effect is consistent. The effect's result applies in general across the levels of the other effects in the interaction, but this does not rule out rare exceptions. Consistency reinforces the interpretation of significance, or lack of significance, of a hypothesis. For

[iv]The basic multilevel model for a dependent binary variable with one independent variable is $\log[p_{ij}/(1 - p_{ij})] = \beta_0 + \beta_1 x_{ij} + U_j$, where U_j is the random effect at level 2. Multiple independent variables with their interactions are easily included.

a significant hypothesis, lack of consistency implies at least one level of the other effect has a result in the opposite direction from the hypothesis. For a nonsignificant hypothesis, lack of consistency implies levels of the other effect have substantial results in both the hypothesized and opposite directions. Lack of consistency weakens the interpretation of significance, or lack of significance, of a hypothesis.

As a test of the robustness of the findings, similar to the two-level analyses, the team repeated interaction analyses using the imputed referred to treatment outcome variable (see above).

3. Results

3.1. Percentage screened and in need of treatment

We first present results for the baseline and all four intervention time periods for screening, in need of treatment, and referral (Table 1). Compared to the rate of youth screened during baseline (71.5%), the overall screening rates across the four periods of the intervention were slightly lower but remained relatively stable. The percentage of youth screened was higher in the enhanced sites during the baseline period compared to the core sites and remained higher throughout the study. Screening declined over time in the core sites, decreasing from 65.3% at baseline to 50.0% during the maintenance period. By contrast, the percentage of screened youth in the enhanced sites remained relatively stable.

Slightly fewer than half of youth (48.2%) were determined to be in need of treatment (based on results of the screening, positive drug test, or drug charge). This percentage was similar in the core and enhanced sites overall, did not vary by time in the enhanced sites, but decreased over time in core sites (most likely because of the decrease in screening over time).

3.2. Referral to treatment among youth in need of treatment

For the main outcome variable of referral to treatment among youth in need of treatment, we present overall findings by time period and condition (Table 2). Among youth determined to be in need (N=14,012), only 26.1% were referred to treatment overall. The percentage referred was higher in the core sites at baseline and remained higher throughout the study. The percentage of youth in need who were referred to treatment increased in the enhanced sites after randomization, from 18.0% to 23.8% once facilitation began, peaking at 27.5% during the late experimental period. Similarly, among core sites, referrals among youth in need increased from 27.1% at baseline to a peak of 33.9% during the late experimental period.

Overall, referrals for youth in need of treatment increased 8.1 percentage points from baseline to the late experimental period, with a 3.1 percentage point decline during the maintenance phase. The core condition had a 6.8 percentage point increase in referrals for youth with a treatment need from baseline to the late experimental period, with a 0.2 percentage point increase in referrals during the maintenance phase. By contrast, the enhanced condition had a 9.5 percentage point increase in referrals for youth needing treatment between baseline and the late experimental period, with a 5.6 percentage point decline in the maintenance phase. Comparing the baseline period to the experimental periods (T3–T5), referrals increased across all sites from 22.5% at baseline to 28.7%.

These overall findings mask considerable variation across sites. As Appendix A shows, there are significant differences in site referral rates within and across the seven states. Site referral rates for youth in need range from 5.4% (State 4, Site 5) to 86.9% (State 2, Site 2).

3.2.1. Post-experimental decay—
At the beginning of the six-month maintenance period, the facilitators ceased working with the enhanced site local change teams. During that period, an overall decrease occurred in the percentage of youth in need who were referred to treatment in enhanced sites (from 27.5% to 21.9%), but little change in core sites (Table 2). Ten of 15 enhanced sites showed a decrease in referral percentage, compared with 8 of 15 core sites.

3.3. Predictors of referral for those in need

Table 3 presents results of the two-level Bayesian regression analysis, with site as the cluster variable and referral to treatment (among those in need) as the dependent measure. This analysis incorporates the entire study (five time periods). Overall, model fit results indicated an excellent fit of the two-level model (PPP=0.499). The initial analysis, based on 50,000 iterations, was replicated 100,000 times, with parameter estimates remaining unchanged.

Experimental condition was not a significant predictor of referral to treatment overall. The study identified several significant individual- and between-level predictors of referrals. At the individual level, youth who tested positive, who had an alcohol/other drug–related charge, were placed in secure detention, were arrested on a felony charge, or were assigned to receive more intensive supervision were more likely to be referred to treatment. By contrast, Hispanic and African American youth were less likely to be referred to treatment.

At the between (site) level, staff in agencies that emphasized the importance of referrals were significantly more likely to refer substance using youth to treatment. The percentage of staff completing BH training was negatively related to referrals.

3.3.1. Sensitivity analysis—
We used an imputed estimate of referral to treatment for substance using youth (0=not referred, 1=referred), described in the Methods section, and (a) compared the concordance in referral for the two measures, and (b) estimated a two-level Bayesian regression model using the imputed referral variable as the dependent measure. There was 92.3% concordance between the original and imputed measures. The two-level regression found parameter estimates similar to those in Table 3. The study found no significant intervention effect. The study replicated initial analysis involving 50,000 iterations for 100,000 iterations, with similar parameter estimates.[v]

3.4. Interactions among sitepairs, time, and intervention

The multilevel analysis of variance model for design included tests of two directional hypotheses, (1) *time* and (2) *intervention*time*. Table 4 displays the ANOVA Type III fixed

[v]Full results are available from the senior author on request.

effects model results for time, with *sitepair*, *intervention*, and their interaction as random effects.

For Hypothesis 1 (time), the least square means were .336 (.130) and .361 (.130) for the baseline and experimental phases, respectively. The two-sided F test approaches statistical significance ($p < .06$); the one-sided test for the directional hypothesis is statistically significant ($p < .03$), since the result is in the hypothesized direction, indicating that sites showed an overall higher referral rate during the combined experimental periods than during baseline. The p-value reflects both the sample size and the effect size of the difference between the baseline and the experimental periods. The sample size was very large, and the p-value was not very significant, which implies a very small effect size. The analysis of variance effect size (Cohen's f) for design in Table 4, was calculated as 0.03 directly from the estimated means and error mean square. Converting this into the more familiar units of Cohen's d, yielded $d = 2f = 0.06$.

The interactions that include time are *sitepair*time*, *intervention*time*, and *sitepair*intervention*time*. All their mean squares are at least as large as *time*, implying that the overall conclusion from the hypothesis has exceptions. Some matched site pairs, interventions, and sitepair*intervention combinations had higher referral rates in baseline than in the experimental periods. These statistics do not identify the exceptions, only that a pattern of exceptions exist to the generality; Appendix A displays the site variation in the percentage of youth referred. For example, the *intervention*time* results suggest two equivalent interpretations: (1) the two study conditions differ in their changes over time, and (2) the levels of time differ in their differences in intervention. Since both intervention and time are dichotomous, in the context of the significant result for Hypothesis 1, this interaction has a pattern of the enhanced sites having higher referral rates over the experimental periods than baseline, and—to a smaller extent—the core sites having higher referral rates at baseline than in experimental periods. Thus, despite the significant result for Hypothesis 1, the strong interaction implies that its conclusion does not apply to both interventions. For the other interactions, finding some *sitepair* or *sitepair*intervention* combinations with higher baseline than later phase referral rates is also expected.

For Hypothesis 2 (*Intervention*time*), the test is highly significant ($p < .0001$), and in the hypothesized direction. The least square means corresponding to this test for the Core condition were .344 (.183) and .317 (.183) for baseline and experimental phases, respectively. For the enhanced condition the least square means were .327 (.183) and .405 (.183) for baseline and experimental phases, respectively. Post hoc tests showed that the difference between baseline and experimental phases was not significant for the core condition ($t=1.57$, $p<.116$) but was for the enhanced condition ($t=-3.90$, $p<.0001$). The overall conclusion is that the combined experimental period referral rates were relatively higher than the baseline phase in the enhanced condition but not in the core condition.

The *sitepair*intervention*time* interaction had a mean square not much smaller than intervention*time. The pattern of this interaction suggests that a few matched site pairs may exist for which referrals over the combined experimental periods were relatively higher than at the baseline phase. In fact, of the 12 significant changes, seven occurred among the enhanced condition, with six significant increases from baseline to the combined experimental period and one significant decrease. The core condition had three significant increases from baseline and two significant decreases.

The above analysis combines the three later phases into a single group; an additional comparison of interest is whether the referral results among these three phases differ. Significant differences would indicate instability for Hypothesis 1, since differences among phases' differences from baseline would also exist. A significant *intervention*laterphases* interaction, showing that the intervention had different effects in the three later phases, would indicate instability for Hypothesis 2. Table 5 displays the results of the analysis for *laterphases*, with *sitepair*, *intervention*, and their interaction as random effects.

The significant *laterphases* difference ($p < .002$) indicates a difference in the average referrals among the three experimental periods—another source of instability for Hypothesis 1. The least square means were .386 (.494), .392 (.494), and .315 (.494) for the early experimental, late experimental, and maintenance phases, respectively. The smaller mean squares for interactions that include *laterphases* imply that these differences among the three experimental periods also apply consistently across *sitepairs*, *intervention*, and *sitepair*intervention* combinations. Instability for Hypothesis 1 from differences among experimental periods supplement the instability indicated by the interactions for *time* presented above.

The nonsignificant ($p = .16$) *intervention*laterphases* interaction indicates that for the differences among the experimental periods, average effects of the intervention did not differ substantially from what would be expected by chance. The one interaction that includes *intervention*laterphases* had a somewhat smaller mean square. Neither of these results imply instability of Hypothesis 2 results due to differences among the experimental periods. The result concerning stability for Hypothesis 2 supplements the conclusion of stability from the interaction presented above.

In summary, for Hypothesis 1 a significant ($p < .03$) overall difference in the hypothesized direction occurred. Combined experimental periods had a higher rate than baseline in referral to treatment of youth in need. However, these differences had substantial variability, with exceptions in the direction opposite to that hypothesized. Hypothesis 2 had a significant difference ($p < .0001$) in the hypothesized direction—the differences between referral rates of combined experimental periods and baseline were larger in enhanced than in core sites. A few exceptions to these differences—with differences in referral rates being larger for the core than enhanced conditions—could not be ruled out.

4. Discussion

Guided by the EPIS framework for promoting system change and the Cascade framework for analyzing identification of SU need and linkage to services, this study compared the effectiveness of two implementation strategies for improving referral to SU treatment among youth under CS. The study largely confirmed the hypotheses, despite substantial variation in referral practices across study sites over time. Compared with the baseline period, the percentage of youth in need who were referred to treatment increased during the experimental period, and the increase was greater in sites receiving enhanced support. Also of note is that the percentage of youth screened for substance use declined over time in the core sites while remaining stable in the enhanced sites. The overall positive effect compared with baseline reflects that the JJ-TRIALS intervention (which involved needs assessment, site feedback reports, extensive trainings, and support for selecting process improvement goals but no external facilitation of implementation teams) was relatively robust, but that structured facilitation provided additional benefits in terms of increased referral rates relative to baseline.

In an earlier report on baseline data (Wasserman et al., 2021), only about one in five youths in need were referred to treatment. Relative to baseline, this study found that the JJ-TRIALS intervention increased the rate of referral to treatment of those with identified substance use needs, from 22.5% at baseline to 28.7% during the experiment (time periods 3–5). These results suggest that a set of activities that target staff training and guidance in setting up inter-agency teams to work toward defined practice goals and feedback about agency movement toward those goals can help to improve management of youths' substance use needs. While the more intensive facilitation offered to enhanced sites was associated with somewhat higher gains, most youth in need were not referred to treatment, and the removal of external facilitation activities resulted in a greater drop off in these improved practices than in sites without this facilitation.

Although the apparent decay in referral rates after enhanced facilitation ended is notable, two possible explanations should be considered. First, youth referred during the maintenance phase had less opportunity to receive a referral than youth in earlier cohorts; data collection in this period was truncated at six months because of project time constraints, resulting in less time for youth to receive referrals. Wasserman et al. (2021) found that during the baseline period the average length of time between intake and treatment referral was 25.7 days, although only 44% were referred within 30 days of screening. Thus, some youth entering in months 4–6 of the maintenance period would not have had sufficient time to receive a referral, so that for some, a referral remained possible in the weeks after data collection ended. On the other hand, the decrease may have resulted from the removal of external facilitation, suggesting that interventions that rely on external facilitators to drive change may need to pay particular attention to ensuring that LCTs develop internal capacity to provide sustainability over the long-term. Although external facilitators in JJ-TRIALS were instructed to train and mentor a local champion and transition leadership to this individual at the end of the 12-month experimental period, the degree to which sites did this successfully varied. Future studies should explore these and other potential reasons for post-experimental decay.

The inconsistency of the intervention effects across the randomized matched pairs suggests that site-level differences exist that produced variation in the intervention's impacts. For example, some core sites showed greater improvement in referrals than their matched enhanced sites. Future studies should utilize qualitative and site management data to help understand the reasons behind these inconsistent findings, which may relate to factors associated with agency leadership buy-in, staff motivation, and interorganizational relationships between probation and community treatment partners.

When controlling for factors that could affect referrals across the entire study period, we found no overall significant effect of facilitation (enhanced condition). This finding suggests that youth and agency-level factors may supersede or impede process improvement efforts if not adequately addressed. For example, level of supervision predicted increased referrals across sites, suggesting that youth who are "deeper" into the system are more likely to receive SU treatment referrals (Wasserman et al., 2021). The lack of consistent attention to following up on the identified SU needs of youths receiving a lower level of supervision is of concern, because the greatest proportion of justice-involved youth are processed at lower supervision levels (Sickmund & Puzzanchera, 2014) and access to needed BH services can prevent further offending (Cuellar et al., 2006; Hoeve, McReynolds, & Wasserman, 2013; Hoeve, McReynolds, McMillan, & Wasserman, 2013). Other predictive youth factors were consistent with findings from previous analyses of baseline period data (Wasserman et al., 2021): testing positive for a drug, having an alcohol/other drug–related charge, being placed in secure detention, and arrested on a felony charge were positively related to referral to treatment. Hispanic and African American youth were less likely to be referred to treatment.

Interestingly, in preliminary work we found that sites selecting referral-related Cascade goals as the target of their improvement efforts (39% of sites; Becan et al., 2020) did not differ in referral rates from those that chose other Cascade targets (e.g., screening, treatment initiation). Compared to referral rates for the entire sample (Table 1), youth in need were more likely to receive a referral over time, suggesting that sites in both conditions (and even more so in the enhanced condition) showed greater differentiation regarding who should be referred, perhaps adjusting screening and identification procedures accordingly. Sites where staff expressed greater emphasis on the importance of referrals were significantly more likely to refer youth to treatment. The percentage of staff completing behavioral health training was negatively related to referrals, perhaps reflecting that agencies recognized that staff needed this type of training.

The JJ-TRIALS study is the first to leverage the Cascade to develop locally defined and tailored targets of practice change. Moreover, Cascade data informed the use of DDDM by CS and BH agencies to not only identify the initial intervention targets but to continue the examination and refinement of practice change innovation so as to improve targeted outcomes (e.g., identification of need, referral). Adult and JJ settings have used DDDM to guide system-wide reform to reduce recidivism and system costs (Chayt, 2012; Dwyer et al., 2012).

The successful use of external facilitators and local change teams (LCTs) to accomplish system process change has been documented in the adult correctional system (Belenko, Visher, et al., 2013; Shafer et al., 2014) and more recently in juvenile CS settings to increase access to HIV/STI testing (Elkington et al., 2020; Huang et al., 2020). The current study is the first to use LCTs to achieve practice change between juvenile CS and BH systems. LCTs are an especially efficient implementation strategy for establishing strong communication between agencies (Bowser et al., 2018), especially those with differing cultures or priorities (Belenko, Visher, et al., 2013; Hoffman et al., 2012; Hurlburt et al., 2014).

Recently, EPIS theory has been expanded to consider the role of bridging factors that support interaction between internal and external contexts (Lengnick-Hall et al., 2020). Our results may suggest that certain types of bridging activities (such as those offered by an external facilitator, compared to those generated by the more “bootstrapping” approach available to core sites) may have different impacts on the uptake and sustainability of practice change. Agencies seeking to improve coordination with other community partners may want to consider the degree of external guidance those efforts will require. We should note, however, that the approach to change with an external facilitator required much greater effort on the part of the LCT in the enhanced arm to continue to implement alone, versus practice change in the core arm, which may have been more in keeping with sites’ internal capacity and thus more sustainable. Future research should explore the impact of external facilitation on LCT dynamics, locus of (change) control, and ultimately on sustainability of practice change. Such work can highlight the role of additional facilitation to address transition of the change process to minimize decay.

More generally, sustainability of new interventions and practices has been a challenge in health care and other service settings, and a gaps remains in research about effective sustainability strategies (Hailemariam et al., 2019; Shelton et al., 2018), particularly in low-resource settings such as juvenile justice (Ritchie et al., 2015). The current study contributes to the literature in noting that although external facilitation in combination with an interagency change team improved outcomes, decay occurred once the study removed facilitation (Stetler et al., 2006). Models such as blended facilitation (i.e., external and local facilitation) have proven effective in clinical settings (Kirchner et al., 2014; Ritchie et al., 2015), and factors such as adaptation and alignment, leadership support, strategic action plans, and resource allocation to build organizational capacity (Hailemariam et al., 2019; Shelton et al., 2018) may achieve sustained practice change in juvenile justice settings.

The JJ-TRIALS intervention yielded improvements in the proportion of identified youths provided with service referrals, but referral rates remained relatively low at the end of the study. Enhanced sites moved from an unadjusted rate of 18% during baseline to a high of 27.5% during the intervention, a 53% improvement; core sites went from a higher rate of 27.1% during baseline to a high of 34.0% during the intervention, a 25% improvement. The overall low rate of service referrals for those in need was unrelated to the availability of county BH services (Wasserman et al., 2021), or to other county indicators of economic well-being. On the other hand, this multi-state, multi-site study did not require specific policy initiatives at each site, nor consequences for better or worse progress toward best practice goals. Agencies wishing for a higher and more consistent level of practice change should consider ways to formalize and incentivize process improvement goals, and to more closely monitor those efforts. Reducing structural barriers to treatment referral and engagement, such as stigma, treatment capacity, insurance barriers, and family support and engagement are also likely to improve referral rates.

Across all study periods, this study found lower service referral rates for African American and Hispanic youth, even after need had been identified. This finding is consistent with longstanding reports of lower behavioral health service use for non-white youths in the general community (Alegria, 2011), although whether such a pattern reflects lack of available services in some communities or implicit referral bias remains unclear (Hall et al., 2015).

4.1. Limitations

This study had some limitations. Site recruitment procedures meant that the study included only CS agencies willing to participant in JJ-TRIALS. Although sites were from diverse states and regions, they were not randomly selected and thus not necessarily representative of all juvenile probation agencies, which limits external validity. However, the characteristics of the youth and agencies were diverse and similar to those reported in a survey of juvenile justice agencies from a nationally representative sample of counties (Scott et al., 2019).

Even with study participation commitments from CS and BH agency leadership, sites varied in staff involvement in implementation activities. For example, BH staff engagement in the LCT or interagency workgroup varied widely. Further, each RC was responsible for delivering all implementation interventions within their state, including facilitating enhanced sites' LCTs. Variation in how well LCT facilitation was implemented may have accounted for differences in outcomes and we will explore this in future mixed methods papers.

Finally, we must acknowledge that collecting data from JJ agencies presented methodological challenges resulting in missing or inconsistent administrative youth data (Dennis et al., 2019). For example, some data were only available in free text format or scanned documents; definitions of juvenile justice statuses or actions sometimes varied across sites; dates of key Cascade events (e.g., date of referral to treatment) were often missing or inconsistent; and agency staff were sometimes overburdened and unable to respond to data queries from the research teams or provide documentation. The RC and CS partners were able to work through many but not all these issues with the collaborating agencies. Therefore, we used a conservative approach to handling missing data, assuming that blank records meant that a Cascade event did not occur (Dennis et al., 2019). However, we obtained similar results using hot deck imputation of missing Cascade data. Nonetheless, improvements in the quality of ongoing data collection and data systems are crucial for supporting future research on Cascade outcomes in the JJS.

4.2. Conclusion

Overall, these results demonstrate both the complexity and challenges of executing systems change in a complex environment where those in need of services move between CS and BH systems. On one hand, results are promising, showing an overall increase in referrals among those in need and suggesting that the complex JJ-TRIALS intervention led to improvements in referral practices. Further, results suggest that improvements were greater in enhanced sites, suggesting that active facilitation can promote and sustain positive outcomes. On the other hand, some core sites outperformed enhanced sites, suggesting that for these core sites, improvements were possible without additional facilitation, based only on the basic set of implementation activities. This high degree of site variability suggests that a range of inner and outer context factors (e.g., state/court policy, service availability, funding, staff practices, implementation fidelity, staff engagement in change practices) that the current study did not examine may account for variation in referral rates. Future studies should examine the reasons for site heterogeneity.

Although the JJ-TRIALS intervention demonstrated the capacity to improve how systems collaborate to identify and respond to substance use–related needs, it also highlights that most justice-involved youth who need SU services are not referred, even when JJ systems work with community BH providers to implement improvements. Most juvenile CS agencies already screen most youth for BH issues (Scott et al., 2019). Further increasing referral is likely to require a comprehensive effort involving implementing policies to utilize screening results to systematically determine and act upon identified SU needs. Using screening practices to drive closer collaboration with BH partners is necessary to increase referral and treatment access and effectively address unmet substance use service needs among justice-involved youth.

Acknowledgements

All persons who have made substantial contributions to the work reported in the manuscript (e.g., technical help, writing and editing assistance, general support), but who do not meet the criteria for authorship, are named in the Acknowledgements and have given us their written permission to be named. If we have not included an Acknowledgements, then that indicates that we have not received substantial contributions fromnon-authors.

Funding

This study was funded under the Juvenile Justice Translational Research on Interventions for Adolescents in the Legal System project (JJ-TRIALS) cooperative agreement, funded by the National Institute on Drug Abuse (NIDA), National Institutes of Health (NIH). The authors gratefully acknowledge the collaborative contributions of NIDA's scientific officer, Dr. Tisha Wiley, and support from the following grant awards: Chestnut Health Systems (U01DA036221); Columbia University (U01DA036226); Emory University (U01DA036233); Mississippi State University (U01DA036176); Temple University (U01DA036225); Texas Christian University(U01DA036224); and University of Kentucky (U01DA036158).

APPENDIX A: Overall Percentage Referred to Treatment among Youth in Need of Treatment, by Site and Matched Pair

SITE	N	% REFERRED
RC 1, Site 1 (E)	170	57.1
RC 1, Site 2 (C)	319	22.9
RC 1, Site 3 (E)	233	22.3
RC 1, Site 4 (C)	262	31.3
RC 2, Site 1 (E)	197	70.6
RC 2, Site 2 (C)	188	68.1

SITE	N	% REFERRED
RC 2, Site 3 (E)	205	62.0
RC 2, Site 4 (C)	183	86.9
RC 3, Site 1 (E)	89	56.2
RC 3, Site 2 (C)	315	35.9
RC 3, Site 3 (E)	680	28.5
RC 3, Site 4 (C)	287	9.4
RC 3, Site 5 (E)	120	43.3
RC 3, Site 6 (C)	69	15.9
RC 4, Site 1 (E)	1177	24.7
RC 4, Site 2 (C)	780	34.9
RC 4, Site 3 (E)	826	13.1
RC 4, Site 4 (C)	224	37.5
RC 4, Site 5 (E)	1853	4.5
RC 4, Site 6 (C)	1160	55.5
RC 5, Site 1 (E)	852	19.6
RC 5, Site 2 (C)	995	12.3
RC 5, Site 3 (E)	559	11.4
RC 5, Site 4 (C)	1297	14.8
RC 6, Site 1 (E)	138	37.7
RC 6, Site 2 (C)	303	31.7
RC 6, Site 3 (E)	107	51.4
RC 6, Site 4 (C)	199	20.6
RC 6, Site 5 (E)	77	58.4
RC 6, Site 6 (C)	148	29.1

Chi-square= 2526.13, df=30, $p < .001$

REFERENCES

Aalsma MC, Tong Y, Lane K, Katz B, & Rosenman MB (2012). Use of outpatient care by juvenile detainees upon community reentry: effects of mental health screening and referral. Psychiatric Services, 63(10), 997–1003. 10.1176/appi.ps.201100550. [PubMed: 22911470]

Aarons GA, Fettes DL, Hurlburt MS, Palinkas LA, Gunderson L, Willging CE, & Chaffin MJ (2014). Collaboration, negotiation, and coalescence for interagency-collaborative teams to scale-up evidence-based practice. Journal of Clinical Child & Adolescent Psychology, 43(6), 915–28. 10.1080/15374416.2013.876642. [PubMed: 24611580]

Aarons GA, Hurlburt M, & Horwitz SM (2011). Advancing a conceptual model of evidence-based practice implementation in public service sectors. Administration and Policy in Mental Health and Mental Health Services Research, 38(1), 4–23. 10.1007/s10488-010-0327-7. [PubMed: 21197565]

Abram KM, Paskar LD, Washburn JJ, & Teplin LA (2008). Perceived barriers to mental health services among youths in,, detention. Journal of the American Academy of Child and Adolescent Psychiatry, 47(3), 301–8. 10.1097/CHI.0b013e318160b3bb. [PubMed: 18216738]

Alegria M, Carson NJ, Goncalves M, & Keefe K (2011). Disparities in treatment for substance use disorders and co-occurring disorders for ethnic/racial minority youth. Journal of the American Academy of Child and Adolescent Psychiatry. 50(1) 22–31. [PubMed: 21156267]

Allison P (2002). Missing data. Thousand Oaks, CA: Sage.

Allison P (2014). Listwise deletion: It's NOT evil. Retrieved from https://statisticalhorizons.com/listwise-deletion-its-not-evil. Accessed 7/30/19.

Anderson DA & Aitkin M (1985). Variance component models with binary response: interviewer variability. Journal of the Royal Statistical Society: Series B, 47, 203–10. https://www.jstor.org/stable/2345561.

Andridge RR, & Little RJ (2010). A review of hot deck imputation for survey non-response. International statistical review, 78(1), 40–64. [PubMed: 21743766]

Asparouhov T & Muthen B (2010). Bayesian Analysis of Latent Variable Models using Mplus. Version 4. Mplus Technical Report http:\\www.statmodel.com.

Becan JE, Bartkowski JP, Knight D, Wiley T, DiClemente R, Ducharme L, … Aarons GA (2018). A model for rigorously applying the Exploration, Preparation, Implementation, Sustainment (EPIS) framework in the design and measurement of a large scale collaborative multi-site study. Health & Justice, 6(1), 9. 10.1186/s40352-018-0068-3. [PubMed: 29654518]

Becan JE, Fisher JH, Johnson ID, Bartkowski J, Seaver R, Gardner S, Aarons GA, Renfro TL, Muiruri R, Blackwell L, Piper KN, Wiley TA, & Knight DK (2020). Improving substance use services for juvenile justice-involved youth: Complexity of process improvement plans in a large scale multi-site study. Administration and Policy in Mental Health and Mental Health Services Research, 47(4), 501–514. 10.1007/s10488-019-01007-z. [PubMed: 31927648]

Belenko S, Johnson I, Taxman F & Rieckmann T (2018). Probation staff attitudes toward substance abuse treatment and evidence-based practices. International Journal of Offender Therapy and Comparative Criminology, 62, 313–333. 10.1177/0306624X16650679. [PubMed: 27220361]

Belenko S, Knight D, Wasserman GA, Dennis M, Wiley T, Taxman FS, Oser C, Dembo R, Robertson A, & Sales J (2017). The Juvenile Justice Behavioral Health Services Cascade: A new framework for measuring unmet substance use treatment services needs among adolescent offenders. Journal of Substance Abuse Treatment, 74, 80–91. 10.1016/j.jsat.2016.12.012. [PubMed: 28132705]

Belenko S, Hiller M, & Hamilton L (2013). Treating substance use disorders in the criminal justice system. Current Psychiatry Reports, 15(11), 414. 10.1007/s11920-013-0414-z. [PubMed: 24132733]

Belenko S, Visher C, Copenhaver M, Hiller M, Melnick G, O'Connell D, et al. (2013). A cluster randomized trial of utilizing a local change team approach to improve the delivery of hiv services in correctional settings: study protocol. Health and Justice, 1(8), 1–11. 10.1186/2194-7899-1-8.

Bowser D, Henry BF, Wasserman GA, Knight D, Gardner S, Krupka K, Grossi B, Cawood M, Wiley T, & Robertson A (2018). Comparison of the overlap between juvenile justice case processing and screening & referral to behavioral health services. Journal of Applied Juvenile Justice Services, 5, 97–125.

Browne WJ, Golalizadeh Lahi M, & Parker RMA (2009). A guide to sample size calculations for random effect models via simulation and the MLPowSim software package. University of Bristol.

Bryk AS & Raudenbush SW (1992). Hierarchical linear models: Applications and data analysis methods. Newbury Park, CA: Sage

Cain MK & Zhang Z (2019). Fit for a Bayesian: An Evaluation of PPP and DIC for Structural Equation Modeling. Structural Equation Modeling: A Multidisciplinary Journal, 26, 39–50.

Chapman JF, Desai RA, & Falzer PR (2006). Mental health service provision in juvenile justice facilities: Pre- and post-release psychiatric care. Child and Adolescent Psychiatric Clinics of North America, 15(2), 445–58. 10.1016/j.chc.2005.11.002. [PubMed: 16527665]

Chayt B (2012). Juvenile justice and mental health: a collaborative approach. http://modelsforchange.net/publications/350. Accessed 11 Apr 2016.

Colins O, Vermeiren R, Vreugdenhil C, van den Brink W, Doreleijers T, & Broekaert E (2010). Psychiatric disorders in detained male adolescents: a systematic literature review. The Canadian Journal of Psychiatry, 55(4), 255–63. 10.1177/070674371005500409. [PubMed: 20416149]

Cuellar AE, McReynolds LS, & Wasserman GA (2006). A cure for crime: Can mental health treatment diversion reduce crime among youth? Journal of Policy Analysis and Management, 25, 197–214. 10.1002/pam.20162. [PubMed: 16465707]

DART: Data Analytic Recidivism Tool. http://recidivism.cityofnewyork.us/index.php?m=about. Accessed 11 Apr 2016.

Davis JP, Dumas TM, Wagner EF, & Merrin GJ (2016). Social ecological determinants of substance use treatment entry among serious juvenile offenders from adolescence through emerging adulthood. Journal of Substance Abuse Treatment, 71, 8–15. 10.1016/j.jsat.2016.08.004. [PubMed: 27776683]

Dennis ML, Smith CN, Belenko S, Knight D, McReynolds L, Rowan G, Dembo R, DiClemente R, Robertson A, & Wiley T (2019). Operationalizing a behavioral health services cascade of care model: Lessons learned from a 33-site implementation in juvenile justice community supervision. Federal Probation, 83, 52–64. [PubMed: 34366438]

Depaoli S (2021). Bayesian Structural Equation Modeling. New York: Guilford Press.

Dwyer AM, Neusteter SR, & Lachman P (2012). Data-driven decisionmaking for strategic justice reinvestment. http://www.urban.org/sites/default/files/alfresco/publication-pdfs/412543-Data-Driven-Decisionmaking-forStrategic-Justice-Reinvestment.PDF. Accessed 11 Apr 2016.

Elkington KS, Spaulding A, Gardner S, Knight D, Belenko S, Becan JE, Robertson AA, Oser C, & DiClemente R (2020). A system-level intervention to encourage collaboration between juvenile justice and public health agencies to promote HIV/STI testing. AIDS Education and Prevention, 32(4), 337–355. 10.1521/aeap.2020.32.4.337. [PubMed: 32897134]

Fagan AA, Wright EM, & Pinchevsky GM (2015). A multi-level analysis of the impact of neighborhood structural and social factors on adolescent substance use. Drug and Alcohol Dependence, 153, 180–186. 10.1016/j.drugalcdep.2015.05.022. [PubMed: 26049206]

Fazel S, Doll H, & Långström N (2008). Mental disorders among adolescents in juvenile detention and correctional facilities: A systematic review and meta-regression analysis of 25 surveys. Journal of the American Academy of Child and Adolescent Psychiatry, 47(9), 1010–19. 10.1097/CHI.ObO13e31817eecf3. [PubMed: 18664994]

Fixsen DL, Blasé KA, Naoom SF, & Wallace F (2009). Core implementation components. Research on Social Work Practice, 19(5), 531–40. 10.1177/1049731509335549.

Fletcher B, Lehman W, Wexler H, Melnick G, Taxman F, & Young D (2009). Measuring collaboration and integration activities in criminal justice and substance abuse treatment agencies. Drug and Alcohol Dependence,103(Suppl 1), S54–S64. 10.1016/j.drugalcdep.2009.01.001. [PubMed: 20088023]

Friedmann PD, Ducharme LJ, Welsh W, Frisman L, Knight K, Kinlock T, et al. (2013). A cluster randomized trial of an organizational linkage intervention for offenders with substance use disorders: study protocol. Health & Justice,1(6). 10.1186/2194-7899-1-6.

Friedmann PD, Hoskinson R, Gordon M, Schwartz R, Kinlock T, Knight K, et al. (2012). Medication-assisted treatment in criminal justice agencies affiliated with the criminal justice-drug abuse treatment studies (cj-dats): availability, barriers, and intentions. Substance Abuse, 33(1), 9–18. 10.1080/08897077.2011.611460. [PubMed: 22263709]

Gelman A & Rubin DB (1992a). Inference from iterative simulation using multiple sequences. Statistical Science, 7, 457–472.

Gelman A & Rubin DB (1992b). A single series from the Gibbs sampler provides a false sense of security. In Bernardo JM, Berger JO, Dawid AP, & Smith AF (Eds.), Bayesian Statistics 4 (p. 592–612). New York, NY: Oxford University Press.

Godley MD, Kahn JH, Dennis ML, Godley SH, & Funk RR (2005). The stability and impact of environmental factors on substance use and problems after adolescent outpatient treatment for cannabis abuse or dependence. Psychology of Addictive Behaviors, 19(1), 62. 10.1037/0893-164X.19.1.62. [PubMed: 15783279]

Guo G & Zhao H (2000). Multilevel modeling for binary data. Annual Review of Sociology, 26, 441–462. 10.1146/annurev.soc.26.1.441.

Hailemariam M, Bustos T, Montgomery B, Barajas R, Evans LB, & Drahota A (2019). Evidence-based intervention sustainability strategies: a systematic review. Implementation Science,14, 57. [PubMed: 31171004]

Haqanee Z, Peterson-Badali M, & Skilling T (2015). Making “What Works” work: Examining probation officers’ experiences addressing the criminogenic needs of juvenile offenders. Journal of Offender Rehabilitation, 54(1), 37–59. 10.1080/10509674.2014.980485.

Hall WJ, Chapman MV, Lee KM, et al. (2015). Implicit racial/ethnic bias among health care professionals and its influence on health care outcomes: a systematic review. American Journal of Public Health,105, e60–e76.

Heck RH & Thomas SL (2015). An introduction to multilevel modeling techniques. 3rd edition. New York: Routledge.

Henderson CE, & Taxman FS (2009). Competing values among criminal justice administrators: the importance of substance abuse treatment. Drug and Alcohol Dependence, 103(1), S7–S16. 10.1016/j.drugalcdep.2008.10.001. [PubMed: 19054632]

Hockenberry S, & Puzzanchera C (2019). Characteristics and trends of delinquency cases resulting in probation. Washington DC: U.S. Department of Justice, Office of Juvenile Justice and Delinquency Prevention.

Hoffman KA, Green C, Ford JH II, Wisdom JP, Gustafson DH, & McCarty D (2012). Improving quality of care in substance abuse treatment using five key process improvement principles. Journal of Behavioral Health Services and Research, 39(3), 234–244. [PubMed: 22282129]

Hoeve M, McReynolds LS, Wasserman GA (2013). The influence of adolescent psychiatric disorder on young adult recidivism. Criminal Justice and Behavior, 40(12), 1368–1382. 10.1177/0093854813488106.

Hoeve M, McReynolds LS, McMillan C, Wasserman GA (2013). The influence of mental health disorders on severity of re-offending in juveniles. Criminal Justice and Behavior, 40(3), 289–301. 10.1177/0093854812459639.

Hoeve M, McReynolds LS, & Wasserman GA (2014). Service referral for juvenile justice youths: Associations with psychiatric disorder and recidivism. Administration and Policy in Mental Health and Mental Health Services Research, 41(3), 379–389. 10.1007/s10488-013-0472-x. [PubMed: 23397231]

Hoffman KA, Green CA, Ford JH II, Wisdom JP, Gustafson DH, & McCarty D (2012). Improving quality of care in substance abuse treatment using five key process improvement principles. The Journal of Behavioral Health Services & Research, 39(3), 234–44. 10.1007/s11414-011-9270-y. [PubMed: 22282129]

Huang S, Gardner S, Piper KN, Coleman AS, Becan JE, Robertson AA, Elkington KS (2020). Bridging systems to implement HIV/STI programming within the juvenile justice system: Strengths and challenges of a local change team approach. AIDS Education and Prevention, 32(5), 432–453. 10.1521/aeap.2020.32.5.432. [PubMed: 33112673]

Hurlburt M, Aarons GA, Fettes D, Willging C, Gunderson L, & Chaffin MJ (2014). Interagency collaborative team model for capacity building to scale-up evidence-based practice. Children and Youth Services Review, 39, 160–168. 10.1016/j.childyouth.2013.10.005. [PubMed: 27512239]

Jakobsen JC, Gluud C, Wetterslev J, & Winkel P (2017). When and how should multiple imputation be used for handling missing data in randomised clinical trials- a practical guide with flowcharts. BMC Medical Research Methodology, 17, 1–10. [PubMed: 28056835]

Karriker-Jaffe KJ (2011). Areas of disadvantage: A systematic review of effects of area‐ level socioeconomic status on substance use outcomes. Drug and Alcohol Review, 30(1), 84–95. 10.1111/j.1465-3362.2010.00191.x. [PubMed: 21219502]

Kets de Vries MFR (2005). Leadership group coaching in action: The zen of creating high performance teams. Academy of Management Perspectives, 19(1), 61–76. 10.5465/ame.2005.15841953.

Kirchner JE, Ritchie MJ, Pitcock JA, Parker LE, Curran GM, & Fortney JC (2014). Outcomes of a partnered facilitation strategy to implement primary care–mental health. Journal of General Internal Medicine, 29(4), 904–912. [PubMed: 25355087]

Knight DK, Belenko S, Wiley T, Robertson A, Arrigona N, Dennis M, Bartkowski J, McReynolds L, Becan J, Knudsen H, Wasserman G, Rose E, DiClemente R, Leukefeld C, & JJ-TRIALS Cooperative. (2016). Juvenile Justice—Translational Research on Interventions for Adolescents in the Legal System (JJ-TRIALS): A cluster randomized trial targeting system-wide improvement in substance use services. Implementation Science, 11, 57. 10.1186/s13012-016-0423-5. [PubMed: 27130175]

Knight DK, Joe GW, Morse DT, Smith C, Knudsen H, Johnson I, Wasserman GA, Arrigona N, McReynolds LS, Becan JE, Leukefeld C, & Wiley TRA (2019). Organizational context and individual adaptability in promoting perceived importance and use of best practices for substance use. Journal of Behavioral Health Services & Research, 46(2), 192–216. 10.1007/s11414-018-9618-7. [PubMed: 29777368]

Koehrsen W (2018). Introduction to Bayesian linear regression. https://towarfdsdatascoemnce.com.

Lengnick-Hall R, Wilging C, Hurlburt M, Fenwick K, & Aarons GA (2020). Contracting as a bridging factor linking inner and outer contexts during EBP implementation and sustainment: a prospective study across multiple U.S. public sector service systems. Implementation Science. 15, 43. 10.1186/s13012-020-00999-9. [PubMed: 32527274]

Lehman WE, Fletcher BW, Wexler HK, & Melnick G (2009). Organizational factors and collaboration and integration activities in criminal justice and drug abuse treatment agencies. Drug and Alcohol Dependence, 103(Suppl 1), S65–72. 10.1016/j.drugalcdep.2009.01.004. [PubMed: 19307068]

Little RJ, & Rubin DB (2019). Statistical analysis with missing data. Hoboken, NJ: John Wiley & Sons.

Lukman PA, Abdullah S, & Rachman A (2018). Bayesian logistic regression and its application for hypothyroid prediction in post-radiation nasopharyngeal cancer patients. Journal of Physics: Conference Series, 1725, 1–9.

Lynch SM (2010). Introduction to applied Bayesian statistics and estimation for social scientists. New York: Springer.

Marchant M & Young KR (2001). The effects of a parent coach on parents’ acquisition and implementation of parenting skills. Education and Treatment of Children, 24(3), 351–73. http://www.jstor.org/stable/42899667.

Marsh JA, Pane JF, & Hamilton LS (2006). Making sense of data-driven decision making in education: evidence from recent RAND Research (RAND Corporation Occasional Paper Series). Santa Monica, CA: RAND Corporation. http://www.rand.org/content/dam/rand/pubs/occasional_papers/2006/RAND_OP170.pdf. Accessed 11 Apr 2016

Marsman M & Wagenmakers, Eric-Jan (2017). Three insights from a Bayesian interpretation of the one-sides p value. Educational and Psychological Measurement, 77, 529–539. [PubMed: 29795927]

McClelland GM, Elkington KS, Teplin LA, & Abram KM (2004). Multiple Substance Use Disorders in Juvenile Detainees. Journal of the American Academy of Child and Adolescent Psychiatry, 43(10), 1215–1224. 10.1097/01.chi.0000134489.58054.9c. [PubMed: 15381888]

Miech R, Schulenberg J, Johnston L, Bachman J, O’Malley P, & Patrick M (2018). National adolescent drug trends in 2018. Ann Arbor, MI: Monitoring the Future.

Mieczkowski T, Newel R, & Wraight B (1998). Using hair analysis, urinalysis, and self-reports to estimate drug use in a sample of detained juveniles. Substance Use and Misuse, 33(7), 1547–1567. 10.3109/10826089809069812. [PubMed: 9657415]

Moullin JC, Dickson KS, Stadnick NA, Rabin B, & Aarons GA (2019). Systematic review of the exploration, preparation, implementation, sustainment (epis) framework. Implementation Science, 14(1), 1. 10.1186/s13012-018-0842-6. [PubMed: 30611302]

Mulvey EP, Schubert CA, & Chaissin L (2010). Substance use and delinquent behavior among serious adolescent offenders. US Department of Justice, Office of Justice Programs, Office of Juvenile Justice and Delinquency Prevention.

Muthén LK & Muthén BO (2018). Mplus user's guide, 8th edition. Los Angeles, CA: Muthén & Muthén.

Pearson FS, Shafer MS, Dembo R, del Mar Vega-Debién G, Pankow J, Duvall JL, Belenko S, Frisman LK, Visher CA, & Pich M (2014). Efficacy of a process improvement intervention on delivery of hiv services to offenders: a multisite trial. American Journal of Public Health, 104(12), 2385–91. 10.2105/AJPH.2014.302035. [PubMed: 25322311]

Powell BJ, McMillen JC, Proctor EK, Carpenter CR, Griffey RT, Bunger AC, Glass JE, & York JL (2012). A compilation of strategies for implementing clinical innovations in health and mental health. Medical Care Research and Review, 69(2), 123–57. 10.1177/1077558711430690. [PubMed: 22203646]

Ritchie MJ, Kirchner JE, Parker LE, Curran GM, Fortney JC, Pitcock JA, Bonner AM, & Kilbourne AM (2015). Evaluation of an implementation facilitation strategy for settings that experience significant implementation barriers. Implementation Science, 10, A46.

Robertson A, Fang Z, Weiland D, Joe G, Gardner S, Dembo R McReynolds L Dickson M, Pankow J, Dennis M, & Elkington K (2020). Recidivism among justice-involved youth: Findings From JJ-TRIALS. Criminal Justice and Behavior, 47, 1059–1078. 10.1177/0093854820922891. [PubMed: 35846112]

Roman JK, Butts JA, & Roman CG (2011). Evaluating systems change in a juvenile justice reform initiative. Children and Youth Services Review, 33, S41–S53. 10.1016/j.childyouth.2011.06.012.

Scott CK, Dennis ML, Grella CE, Funk RR, & Lurigio AJ (2019). Juvenile justice systems of care: Results of a national survey of community supervision agencies and behavioral health providers on services provision and cross-system interactions. Health & Justice, 7(1), 11. 10.1186/s40352-019-0093-x. [PubMed: 31201642]

Sedlak A, & Bruce C (2010). Youth's characteristics and backgrounds. US Department of Justice, Office of Justice Programs, Office of Juvenile Justice and Delinquency Prevention.

Shafer MS, Prendergast M, Melnick G, Stein LA, Welsh WN, & the CAW. (2014). A cluster randomized trial of an organizational process improvement intervention for improving the assessment and case planning of offenders: A study protocol. Health & Justice, 2(1), 1. 10.1186/2194-7899-2-1.

Shelton RC, Cooper BR, & Stirman SW (2018). The sustainability of evidence-based interventions and practices in public health and health care. Annual Review of Public Health, 39, 55–76.

Sickmund M & Puzzanchera C (2014). Juvenile offenders and victims: 2014 national report. Pittsburgh, National Center for Juvenile Justice.

Stavseth MR, Clausen T, & Røislien J (2019). How handling missing data may impact conclusions: A comparison of six different imputation methods for categorical questionnaire data. SAGE Open Medicine, 7, 2050312118822912. [PubMed: 30671242]

Stetler CB, Legro MW, Rycroft-Malone J, Bowman C, Curran G, Guihany M, et al. : (2006). Role of external facilitation in implementation of research findings: A qualitative evaluation of facilitation experiences in the Veterans Health Administration. Implementation Science, 1, 23. [PubMed: 17049080]

Stiffman AR, Pescosolido B, & Cabassa LJ (2004). Building a model to understand youth service access: the gateway provider model. Mental Health Services Research, 6(4), 189–98. 10.1023/b:mhsr.0000044745.09952.33. [PubMed: 15588030]

Substance Abuse and Mental Health Services Administration. (2017). Key substance use and mental health indicators in the United States: Results from the 2016 National Survey on Drug Use and Health. https://www.samhsa.gov/data/

Taxman F & Belenko S (2012). Implementing evidence-based practices in community corrections and addiction treatment. New York: Springer.

Teplin LA, Abram KM, McClelland GM, Dulcan MK, & Mericle AA (2002). Psychiatric disorders in youth in juvenile detention. Archives of General Psychiatry, 59(12), 1133–1143. 10.1001/archpsyc.59.12.1133. [PubMed: 12470130]

Teplin LA, Abram KA, McClelland GM, Washburn JJ, & Pikus A (2005). Detecting mental disorder in juvenile detainees: Who receives services. American Journal of Public Health, 95, 1773–1780. 10.2105/AJPH.2005.067819. [PubMed: 16186454]

Wasserman GA, McReynolds LS, Taxman FS, Belenko S, Elkington KS, Robertson AA, Dennis ML, Knight DK, Knudsen HK, Dembo R, Ciarleglio A, & Wiley TRA (2021). The missing link(age): Multilevel contributors to service uptake failure among youths on community justice supervision. Psychiatric Services, 72(5, 546–554. 10.1176/appi.ps.202000163. [PubMed: 33765861]

Wasserman GA, McReynolds LS, Whited AL, Keating JM, Musabegovic H, & Huo Y (2008). Juvenile probation officers' mental health decision making. Administration and Policy in Mental Health and Mental Health Services Research, 35(6), 410–422. 10.1007/s10488-008-0183-x. [PubMed: 18642071]

Wasserman GA, McReynolds LS, Whited AL, Keating JM, Musabegovic H, & Huo Y (2009). Evaluating Project Connect: Improving juvenile probationers' mental health and substance use service access. Administration and Policy in Mental Health and Mental Health Services Research, 36(6), 393–405. 10.1007/s10488-009-0229-8. [PubMed: 19582568]

Wasserman GA, McReynolds LS, Schwalbe CS, Keating JM, & Jones SA (2010) Psychiatric disorder, comorbidity and suicidal behavior in juvenile justice youth. Criminal Justice and Behavior, 37(12); 1361–1376. 10.1177/0093854810382751.

Welsh WN, Dembo R, Lehman WEK, Bartkowski JP, Hamilton L, Leukefeld CG, & Wiley T (2021). Critical factors influencing interorganizational relationships between juvenile probation and behavioral health agencies. Administration and Policy in Mental Health and Mental Health Services Research, 48(2), 233–249. 10.1007/s10488-020-01066-7. [PubMed: 32666324]

White LM, Aalsma MC, Sayler MP, Hershberger A, Anderson VR, Schwartz K, Dir A, & McGrew JH (2019). Behavioral health utilization among detained adolescents: A meta-analysis of prevalence and potential moderators. Journal of Adolescent Health, 64, 700–708. 10.1016/j.jadohealth.2019.02.010.

White LM, Lau KS, & Aalsma MC (2016). Detained adolescents: mental health needs, treatment use, and recidivism. Journal of the American Academy of Psychiatry and the Law, 44, 200–212. [PubMed: 27236176]

Wiley T, Belenko S, Knight D, Bartkowski J, Robertson A, Aarons G, Wasserman G, Leukefeld C, DiClemente R, & Jones D (2015). Juvenile Justice-Translating Research Interventions for Adolescents in the Legal System (JJ-TRIALS): A multi-site, cooperative implementation science cooperative. Implementation Science, 10 (Suppl 1), A43. 10.1186/1748-5908-10-S1-A43.

Wong GY & Mason WM (1985). The hierarchical logistic regression model for multilevel analysis. Journal of the American Statistical Association, 80, 513–23. 10.2307/2288464.

Zhu X, Stone CA (2021). Bayesian comparison of alternative graded response models for performance assessment applications. Educational and Psychological Measurement, 72, 774–799.

Zitzmann S, & Hecht M (2019). Going beyond convergence in Bayesian estimation: Why precision matters too and how to assess it. Structural Equation Modeling: A Multidisciplinary Journal, 26, 646–661. 10.1080/10705511.2018.1545232.

Author Manuscript

Author Manuscript

Author Manuscript

Highlights

- Multisite cluster-randomized trial of an implementation intervention to reduce unmet substance use service needs among youth on community supervision
- Nearly half of youth were identified as being in need of treatment
- The intervention increased the percentage of referrals among youth in need of treatment overall, relative to baseline
- Overall, youth under more intensive juvenile justice supervision, charged with an alcohol or drug offense or testing positive for a drug, arrested for a felony, or in secure detention were more likely to be referred to treatment; Hispanic and Black youth were less likely than whites to be referred to treatment
- The Enhanced intervention resulted in further increases in referrals relative to baseline
- There were substantial differences in referral rates across sites, and variation in experimental impact across sites
- Even after the intervention, most youth in need of treatment were not referred, indicating that additional interventions and policy changes are needed to improve linkage to treatment of youth under community supervision

Table 1.

% of Total Sample Screened and In Need of Treatment

	Baseline (6 months) (N=7840)	Pre-randomization Activities (6 months) (N=5811)	Early Experimental (6 months) (N=5156)	Late Experimental (6 months) (N=4999)	Maintenance (6 months) (N=5243)	TOTAL (N=29049)
SCREENED						
Enhanced	76.6	74.4	77.4	78.7	75.7	76.5
Core	65.3	60.6	60.1	56.6	50.0	59.1
TOTAL	71.5	67.7	69.3	68.0	63.4	68.3
IN NEED						
Enhanced	49.1	47.7	46.2	47.6	46.5	47.6
Core	55.8	49.4	46.4	46.9	43.2	49.0
TOTAL	52.2	48.5	46.3	47.2	44.9	48.2

Screened: Chi-Square=1009.18, df=1, p<.001. In Need: Chi-Square=5.473, df=1, p<.001

Table 2.

Percentage of Youth with Substance Use Treatment Need who were Referred to Treatment, by Experimental Time Period and Condition

Study Condition	**Time Period**					
	T1 Baseline (6 months) (N=4089)	**T2 Pre-randomization Activities (6 months) (N=2820)**	**T3 Early Experimental (6 months) (N=2386)**	**T4 Late Experimental (6 months) (N=2362)**	**T5 Maintenance (6 months) (N=2355)**	**TOTAL (N=14012)**
Enhanced (N=7283)	18.0	19.7	23.8	27.5	21.9	21.6
Core (N=6729)	27.1	30.8	32.4	33.9	34.1	31.0
TOTAL (N=14012)	22.5	25.2	27.9	30.6	27.5	26.1

NOTE: Dark line indicates randomization of sites into Core and Enhanced conditions

Chi-Square=159.23, df=1, $p<.001$

Table 3.

Two-level Regression Analysis (Bayesian Estimation) of Predictors of Referral for Youth with Indicated Substance Use Treatment Need

				95% C.I.	
	Estimate	**Posterior S.D.**	**One-Tailed P-Value**	**Lower 2.5%**	**Upper 2.5%**
Within Level					
Referred if In Need on:					
Adjudicated delinquent	0.086	0.068	0.104	−0.047	0.219
Screened positive for drug	0.520	0.059	0.000	0.406	0.635
Age	−0.006	0.015	0.357	−0.035	0.025
Gender	0.075	0.049	0.064	−0.020	0.171
Hispanic	−0.091	0.051	0.038	−0.191	0.010
African American	−0.202	0.049	0.000	−0.298	−0.107
Alcohol/Drug Charge	0.296	0.042	0.000	0.214	0.378
In Detention	0.193	0.055	0.000	0.087	0.302
Charge Level	0.101	0.045	0.012	0.012	0.188
Higher supervision level	0.709	0.066	0.000	0.580	0.839
Between Level					
Referred if In Need on:					
Experimental Condition	0.014	0.441	0.487	−0.890	0.883
Poverty factor	−0.107	0.284	0.331	−0.677	0.455
Health Insurance factor	0.371	0.251	0.068	−0.155	0.851
Organizational Climate factor	−0.066	0.073	0.160	−0.207	0.084
Interorganizational Relationships	−0.160	0.187	0.176	−0.515	0.230
Innovation	−0.093	0.358	0.386	−0.785	0.654
Use of treatment referrals	0.510	0.279	0.044	−0.106	1.022
% Staff completing BH Training	−2.367	1.000	0.019	−4.186	−0.177
# months adding staff to LCT	−0.084	0.177	0.296	−0.437	0.274
# months without staffing issues	0.025	0.046	0.274	−0.067	0.119
# months BH staff attending	−0.011	0.048	0.397	−0.107	0.083
% attending leadership meetings	0.191	0.654	0.373	−1.143	1.479
Threshold for referred if in need	1.005	0.897	0.118	−0.747	2.847
Residual Variances					
Referred if In Need	0.391	0.454	0.000	0.131	1.672

Table 4.

Results of Multilevel Analysis of Variance of Time

Effect	Num DF	Den DF	Mean Square	F Value	Pr > F
Sitepair	14	.	11.08	0.13	.
intervention	1	8281	1.14	0.02	0.8906
Sitepair*intervention	14	8281	7.71	0.05	1.0000
Time	1	8281	0.57	3.57	0.0588
Sitepair*time	14	8281	0.73	4.56	<.0001
intervention*time	1	8281	2.52	15.74	<.0001
sitepair*intervention*time	14	8281	1.46	9.15	<.0001

Table 5.

Results of Multilevel Analysis of Variance of Later Phases

Effect	Num DF	Den DF	Mean Square	F Value	Pr > F
Sitepair	14	.	4.99	0.01	.
Intervention	1	4481	3.17	0.01	0.9290
Sitepair*intervention	14	4481	4.29	0.01	1.0000
Laterphases	2	4481	1.16	6.69	0.0013
Sitepair*laterphases	28	4481	0.21	1.20	0.2186
intervention*laterphases	2	4481	0.32	1.85	0.1577
Sitepair*intervention*laterphases	28	4481	0.18	1.05	0.3885