

eScholarship@UMassChan

Sustainability of Evidence-Based Practices: Risk-Need-Responsivity in Probation 7 Years Later

Item Type	Preprint
Authors	Vincent, Gina M.;Drawbridge, Dara C.;Lawson, Spencer G.;Todorovic, Kristina;Perrault, Rachael T.
Citation	Vincent, G. M., Drawbridge, D. C., Lawson, S. G., Todorovic, K., & Perrault, R. T. (2024). Sustainability of evidence-based practices: Risk-need-responsivity in probation 7 years later. <i>Criminal Justice and Behavior</i> , 51(5), 643-666. doi: 10.1177/00938548241229193
DOI	10.1177/00938548241229193
Publisher	SAGE Publications
Rights	Copyright © 2024 The Author(s). Authors' submitted manuscript posted as allowed by the publisher's author archiving and re-use guidelines at https://us.sagepub.com/en-us/nam/journal-author-archiving-policies-and-re-use .
Download date	2025-05-16 17:33:59
Item License	https://journals.sagepub.com/page/policies/text-and-data-mining-license
Link to Item	https://hdl.handle.net/20.500.14038/53398

Sustainability of Evidence-Based Practices: Risk-Need-Responsivity in Probation**7-Years Later**

Gina M. Vincent¹, Dara C. Drawbridge¹, Spencer G. Lawson¹, Kristina Todorovic²,
and Rachael T. Perrault¹

¹UMass Chan Medical School

²University of Southern Indiana

Gina M. Vincent  <https://orcid.org/0000-0002-9315-9584>
Dara C. Drawbridge  <https://orcid.org/0000-0002-7111-8634>
Spencer G. Lawson  <https://orcid.org/0000-0002-9572-169X>
Kristina Todorovic  <https://orcid.org/0000-0001-8054-9682>
Rachael T. Perrault  <https://orcid.org/0000-0001-6695-2643>

Author's Note

Gina M. Vincent, Ph.D., Dara C. Drawbridge, Ph.D., Spencer G. Lawson, Ph.D., and Rachael T. Perrault, M.A., Department of Psychiatry, UMass Chan Medical School; Kristina Todorovic, Ph.D., Department of Psychology, University of Southern Indiana.

The authors thank the probation leadership and officers at the study sites. We also thank our statisticians and research associates, Lisa Lundquist, Cassandra Beinemann, Maxwell Christensen, Danielle Rieger, and Lauren McDowell. This research was funded by the Office of Juvenile Justice and Delinquency Prevention (PI: Vincent, 2016-JF-FX-0057) and the John D. and Catherine T. MacArthur Foundation (07-88401-000-HCD) Models for Change Initiative Research Network.

Correspondence concerning this article should be addressed to Gina Vincent, PhD, UMass Chan Medical School, Department of Psychiatry, 222 Maple Ave, Shrewsbury, MA 01545; Gina.Vincent@umassmed.edu.

Gina M. Vincent, Ph.D. is a Professor and Co-Director of the Law and Psychiatry Program and faculty in the Implementation Science & Practice Advances Research Center (iSPARC) at the UMass Chan Medical School. She studies implementation of risk assessment and risk-need-responsivity, as well as mental health and substance use in justice settings.

Dara C. Drawbridge, Ph.D., is an Assistant Professor in the Law and Psychiatry Program and faculty of the Implementation Science & Practice Advances Research Center (iSPARC) at the UMass Chan Medical School. She provides implementation support for, and studies, the implementation of risk-need-responsivity, risk-needs assessment, and case management in justice settings.

Spencer G. Lawson, Ph.D. is a Senior Research Scientist in the Law and Psychiatry Program and in the Implementation Science & Practice Advances Research Center (iSPARC) at the UMass Chan Medical School. His research considers the practical implications for managing and preventing behavioral health conditions among people in criminal-legal settings, with attention also placed on evaluating the implementation and validity of risk instruments.

Kristina Todorovic, Ph.D., is an Assistant Professor at the University of Southern Indiana. She studies extralegal factors that influence jurors' decisions and perceptions of child witnesses' credibility. Her research contributes to the development of evidence-based investigative interview techniques with children in a forensic context.

Rachael Perrault, M.A., is a Project Director in the Law and Psychiatry Program and in the Implementation Science & Practice Advances Research Center (iSPARC) at UMass Chan Medical School. She studies implementation of risk assessment and risk-need-responsivity, as well as decision-making in juvenile justice settings.

This is the authors' manuscript of the article published in final edited form as:

Vincent, G. M., Drawbridge, D. C., Lawson, S. G., Todorovic, K., & Perrault, R. T. (2024). Sustainability of evidence-based practices: Risk-need-responsivity in probation 7 years later. *Criminal Justice and Behavior*, 51(5), 643-666. doi: 10.1177/00938548241229193

Abstract

This quasi-experimental, pre-post study investigated sustainability of risk-needs assessment (RNA) and the risk-need-responsivity (RNR) practices in five juvenile probation offices across two states seven years after rigorous implementation. The study evaluated adherence to practices and facilitators of sustainability via interviews with leadership ($N = 10$) and probation officers ($N = 84$) and a 7th-year cohort of youth propensity-score matched to a cohort from the 1st-year of implementation ($n = 730$ per state cohort). Probation officers in both states reported relatively high adherence to RNR-based practices and one state significantly improved adherence to its RNA administration policies and the risk principle. Adherence to the risk principle in placement and service referral decisions regressed in the other state. Essential ingredients for sustainment included maintaining staff training and monitoring, evolving RNR practices through adoption of complementary evidence-based practices, and having court personnel who enable best practices.

Keywords: Juvenile Justice, Risk-Needs Assessment, Risk-Need-Responsivity, Sustainability, Implementation

Sustainability of Evidence-Based Practices: Risk-Need-Responsivity in Juvenile Probation 7-Years Later

State and local juvenile justice systems have experienced substantial reform over the last 15 years resulting in significant reductions in juvenile arrest, adjudication, and incarceration (Hockenberry & Puzanchera, 2019) and increased adoption of various evidence-based practices (JJGPPS, March, 2021) to reduce recidivism. One of the most well-studied and widely recommended evidence-based practices in justice settings (National Research Council, 2013) is the risk-need-responsivity (RNR) model (Andrews & Bonta, 2010). The RNR framework has three primary principles. The *risk principle* suggests the system should reserve intensive interventions (e.g., confinement, more services) for the highest risk cases. The *need principle* stipulates case management practices should provide programming that targets individuals' dynamic risk factors that influence their offending (criminogenic needs). The *responsivity principle* indicates programming should consider how well the styles and modes of programming are matched to individuals' attributes that may affect treatment response. For the RNR approach to be fully realized, it requires agencies to adopt a valid risk-needs assessment (RNA) to accurately identify individuals' risk for reoffending and their key dynamic risk factors (e.g., impulsivity, substance use) to target with services.

Despite widespread adoption of RNAs and RNR, there are many challenges faced by systems when implementing these practices, including how to translate information from RNAs into case management decisions. For example, probation studies have shown that case plans are often not linked to the results of RNAs as the RNR model prescribes (e.g., Miller & Maloney, 2013; Viglione et al., 2015). Moreover, probation officers may misunderstand how to use the practices or simply do not buy-in to their value (Ingel et al., 2022). However, translating RNAs and the RNR approach into probation practice requires buy-in from not only probation officers

(POs), but also other key stakeholders throughout a jurisdiction (e.g., judges, attorneys, probation leadership). The representatives of all groups must be engaged in the development of strong policies and implementation procedures (Vincent et al., 2012) to cultivate stakeholder buy-in. Studies of RNAs or RNR use in criminal-legal settings that incorporated strong stakeholder-engagement and guidance from researchers report more positive findings (e.g., Viljoen et al., 2019), than those without this engagement.

While strong implementation-related studies of RNA and RNR are slowly increasing, a critical next step is to investigate whether implementation of these practices are sustained over time. To date, there are very few studies of the sustainability of evidence-based practices or reform efforts of any sort in legal settings (e.g., Miller & Palmer, 2020). Moreover, most studies reporting limited adherence to RNR practices do not disentangle whether this outcome was due to an initial implementation problem versus a sustainability problem. To address this gap, the current study gathered an additional wave of data from juvenile probation offices in two states seven years after comprehensive implementation of RNA and RNR-based practices.

Implementation Fidelity and Sustainability

With the substantial uptake of evidence-based practice reforms in justice settings, it is imperative for criminal-legal research to not measure impacts (such as recidivism) without also measuring fidelity. Implementation strategies are the methods used to enhance adoption, implementation, and sustainability of innovations (e.g., RNR, assessment strategies, new interventions) in practice settings (Proctor et al., 2013). To achieve improved outcomes, new practices need to be paired with effective implementation strategies (e.g., revising written procedures, obtaining stakeholder buy-in, designing decision-support tools) and delivered in a system that supports the new practice (Fixsen et al., 2019). Implementation fidelity is “the degree

to which an intervention was implemented as prescribed in the original protocol or as it was intended by the developers” (Proctor et al., 2011, p. 69). A core component of implementation fidelity is practitioners’ (in this case, judges, POs, probation leadership) adherence to new practices (Carroll et al., 2007), which relates to both their behaviors and their attitudes.

Sustainability refers to whether an evidence-based practice has been incorporated into routine practice over time (Moullin et al., 2020). Trajectories of reform efforts can take many forms, with a common form being to simply adopt the reform in name but never actually relate it functionally to practice (Taylor, 2005). Adoption of RNAs in legal settings are no exception. Several studies of POs indicate they may complete their office’s RNA routinely but then report little to no use of the tool in decision-making (Haas & DeTardo-Bora, 2009; Miller & Maloney, 2013). The handful of observational and cross-sectional studies have shown there is little evidence of POs following RNR principles, such that low risk youth may still get multiple services (Fabelo et al., 2015) or POs spend more supervision time on surveillance-related issues than on addressing needs (Viglione, 2019). When fidelity is lacking, agencies are prone to stop using new practices within two years (Loman et al., 2010). Sustained trajectories of reform, on the other hand, can take different forms. One involves the new practice overtaking whatever preceded it so completely that it is institutionalized as the status quo (sustained implementation). Another trajectory is to achieve a dynamic equilibrium by adjusting and adapting the evidence-based practice to a continually changing environment (innovative; Taylor, 2005).

Studies of sustainability should examine multiple constructs. In their synthesis of the literature, Moore et al. (2017) defined sustainability as encompassing five constructs: (1) for a specific period of time, (2) the program or strategy continues to be delivered and/or (3) the individual (e.g., practitioner) change in behavior necessary for the new strategy has maintained;

(4) the strategy and practitioners' behavior change may evolve or adapt while (5) continuing to produce benefits for individuals/systems. In addition, studies of sustainability must consider contextual factors (Moullin et al., 2020; Stirman et al., 2012) that can act as facilitators or barriers to new practices (e.g., sociopolitical climate, practices of agency partners, leadership; Rocque et al., 2014). In the case of RNR, one must remember that POs do not operate in a vacuum. They operate within a context that either enables their ability to follow RNR (e.g., courts permit time for RNAs to be completed pre-disposition to inform diversion or service decisions), or hinders it (e.g., judges set all expectations in court orders before an RNA).

The Current Study

Implementation science underscores the importance of long-term studies that follow the implementation of RNAs and RNR to assess whether and how practices are sustained. The current study is from the *Long-Term RNR Sustainability Study*, a quasi-experimental study of the sustainability of RNA and RNR practices in five county-run probation offices in two states seven years after their RNA and RNR implementation. Two probation offices were from a Southern state, which implemented the *Structured Assessment of Violence Risk in Youth* (SAVRY; Borum et al., 2006), and three were in a Northeastern state, which implemented the *Youth Level of Service/Case Management Inventory* (YLS/CMI; Hoge & Andrews, 2011). All offices originally followed a researcher-guided, stakeholder-engaged, implementation strategy (Vincent et al., 2012) for their RNA and RNR. The current study compares their seventh-year outcomes (7th-year cohort) to outcomes after their first year of implementation in 2009 and 2010 (1st-year cohort) as reported in the *Risk-Needs Assessment in Juvenile Probation Implementation Study* [RNAJP] (Vincent, Paiva-Salisbury et al., 2012; Vincent et al., 2016).

In 2009 and 2010, the five juvenile probation offices implemented policies based on RNR procedures. The policies of interest to the current study include a) administration of a valid RNA prior to disposition for every youth referred or adjudicated, with a few exceptions, and b) alignment of multiple decisions with the risk principle (e.g., disposition and placement recommendations, number of service referrals) such that low risk youth received little intervention and high-risk youth received intensive intervention. In a quasi-experimental pre-post design measuring outcomes up to 18-months after implementation (1st-year cohort), the RNAJP study reported strong adherence to RNA administration policies in three of five probation offices (Vincent et al., 2016). As a result of high adherence, these effective implementers had significant improvement in at least three of the five outcomes studied (reduced placement rates, less severe dispositions, supervision levels and number of services aligned with risk levels, and recidivism), while the remaining two (one per state) with poor adherence had little change in case processing outcomes. However, the initially poor implementers may have achieved better outcomes over time. It can take two-to-four years for agencies' implementation efforts to have an impact on the people served (Goldstein, 2011).

The current study used data from the five probation offices to investigate the sustainability of RNR practices. The researchers defined sustainability using the first four constructs identified by Moore et al. (2017), which pertain to implementation outcomes as opposed to impacts (fifth construct). For the first construct (period of time), we selected a duration of seven years based on findings from other human service agencies indicating at least seven years was the average duration for sustained implementers (Loman et al., 2010). For the second construct (strategies continue to be delivered), we conducted interviews with

probation office leadership to examine whether the RNA and RNR-practices continued to be delivered over time. For the third construct, we examined sustained practitioner behavior change using three methods: (1) 7th-year POs' self-reported adherence to RNR in their decisions, (2) 1st- vs. 7th-year cohorts' adherence to RNA administration policies (i.e., did every youth receive an RNA at the prescribed time?), and (3) 1st- vs. 7th-year cohorts' adherence to the risk principle in placements and service referrals. Comparisons between the 1st- and 7th-year cohorts quantified whether practices regressed (worse than the first year), sustained (same high adherence as the first year), or improved (better than the first year) over time. Information from leadership interviews documented whether RNR practices evolved or adapted (fourth construct) and identified facilitators and barriers.

The primary research questions were: (1) whether both states continued to deliver their RNA and RNR procedures seven years after implementation despite both having had one initially poor implementer site, and (2) whether both states sustained or improved practitioner behavior change as indicated by the three adherence measures. We hypothesized both states would be continuing to deliver their RNA and RNR-based procedures and would not differ significantly in measures of practitioner behavior change.

Method

Sites (Probation Offices)

The researchers conducted this study in the five RNAJP study county-run probation offices; three in a Northeastern (NE) state and two in a Southern (SO) state. The sites originally were selected because they matched between states in terms of their rural versus urban location, volume of youth, and base rates of any out-of-home placements. All sites implemented the RNA and RNR practices office-wide in 2009 except Site 2 in the Northeastern state (NE Site 2)

initially piloted the YLS/CMI with only select POs, and NE Site 3 piloted the YLS/CMI in just one probation unit to start. The current study procedures accommodated for these exceptions by 1) not conducting a cohort comparison of adherence to the RNA administration policy in NE Site 2 (only youth cases receiving a YLS/CMI were included in the 1st-year cohort, giving the impression of 100% adherence), and 2) the 7th-year cohort in NE Site 3 comprised only youth cases seen in the same probation unit as the 1st-year.

Probation Leadership and Officer Samples

The first author conducted one in-person interview per site with the chief probation officer (or equivalent) and manager ($N = 10$) to document system or programmatic changes occurring over the seven-year period. All but one of these leaders had been working in the office when the RNA was first implemented. Research staff conducted phone interviews with 84 (Southern = 38, Northeastern = 46) of the 87 probation officers and supervisors from the five sites who were responsible for conducting, using, or supervising the use of RNAs. The missing three POs were on leave. Most POs interviewed (91.67%) were responsible for conducting initial RNAs with youth at the time of the referral or adjudication, while the others were responsible for only reassessments or were supervisors who reviewed but did not conduct RNAs. Collectively, the PO sample had an average 7.15 years of experience in their current position, 38 had been working in the office since the 1st-year cohort, and 10 POs had only been hired in the past year.

Table 1 reports the characteristics of the full sample of POs and by state. POs in the Northeastern state reported having had significantly more training (includes booster trainings) on the YLS/CMI ($M = 10.83$, $SD = 7.04$) than POs in the Southern state did on the SAVRY ($M = 2.84$, $SD = 2.33$), had been working in their current position for significantly longer, and were

older, on average. There was not a significant state difference in the number of POs who had been hired after the 1st-year cohort was gathered.

Youth Sample

Researchers received administrative data from the 7th-year cohort of youth using all continuous cases that were expected to receive an RNA according to each office's policy (Northeastern state = youth referred to the court, Southern state = youth adjudicated, with a few office-specific policy exceptions) starting January 1st, 2017 until a sufficient number of cases was obtained in each site (Southern $n = 337$; Northeastern $n = 552$). Researchers initially selected youth for inclusion in the 7th-year cohort who matched youth in the pre-implementation cohort from the original RNAJP study using 1:1 nearest neighbor matching (without replacements) procedures to equate the cohorts. Propensity-score matching (PSM) is a technique commonly used in observational studies to reduce potential bias resulting from differences on relevant characteristics between control and treated groups (Rosenbaum & Rubin, 1983). Because the current study compares only the 1st-year and 7th-year cohorts, researchers evaluated the match between these two cohorts, within-probation offices, using a full matching procedure (Stuart et al., 2011) to ensure these cohorts were balanced within sites. By making use of all individuals present in the dataset, a full matching method entails categorizing individuals into a series of matched sets or subclasses. Each set includes both treated and untreated individuals, ensuring the inclusion of at least one individual from each group. Variables were included in the matching procedure if they were demographic characteristics that were not matched naturally (e.g., gender was constant across years so did not require balancing) or were associated with out-of-home placements (e.g., living situation, substance use and mental health history; Brookhart et al., 2006). Researchers conducted a series of bivariable comparisons to examine associations

between each covariate and outcome, applying a $p < .05$ criterion for covariate inclusion in each sites' propensity score analysis (Schafer & Kang, 2008). To evaluate improvement in covariate balance, absolute standardized mean difference (ASMD) was assessed both before and after matching (Rubin, 2001) by examining the percentage balance improvement (i.e., reduction in imbalance) in the standardized mean difference. PSM results are included in the Supplementary Materials Table SA.

The Southern sites showed similar covariate distributions between the treated and control groups (i.e., only 1 of the 20 total covariates across both sites had $ASMD > 0.20$), suggesting an improvement in fit following matching. However, the Northeastern sites still showed poor distributional balance with certain covariates (i.e., 8 of the 28 total covariates across all three sites had $ASMD > 0.20$). Eleven covariates showed a reduction in balance after matching, indicated by negative percent balance improvement estimates; however, most covariates still had an acceptable ASMD of ≤ 0.20 . In part, the observed imbalance with certain covariates for the Northeastern sites was due to a) the change in sampling in NE Site 3, leading to youth having less serious offense histories in the 7th-year cohort because they were assessed earlier in the process (intake) than in the 1st-year cohort (post-disposition), and b) possible overall shifts in the characteristics of youth entering the system in the 7th-year (e.g., they had higher rates of behavioral health treatment than the 1st-year). Consistent with acceptable threshold values, covariates that showed an ASMD above 0.20 following matching and were significantly related to a respective outcome were included in regression analyses to create double robustness and decrease residual bias (Schafer & Kang, 2008). Weights generated from the PSM procedure also were included in subsequent regression models.

The final matched samples included 297 youth per cohort in the Southern state and 433 per cohort in the Northeastern state. The matched cohorts were roughly 24% female, 50% Black, and 15.50 years of age, on average. Only 28% had a violent index offense and the average number of prior offenses was under one (see Supplementary Materials Table SB).

Measures and Procedures

Risk-Needs Assessment Instruments

The Southern state implemented the SAVRY (Borum et al., 2006), a 24 risk-item factors and six protective factor-item instrument that follows the structured professional judgment (SPJ) approach. The SPJ approach involves raters (e.g., POs) providing a categorical risk rating of a youth's overall risk for violence (i.e., low, moderate, high) based on their professional judgment following appraisal of the relevance of all item ratings. The Northeastern state implemented the YLS/CMI (Hoge & Andrews, 2006), a static/dynamic actuarial instrument. The YLS/CMI differs from the SAVRY in that risk level is determined based on total scores (a sum of 42 risk-items) used to assign youths to an actuarial-based categorical risk level. Both instruments have comparable predictive validity for reoffending, and both significantly predict both violent and non-violent reoffending (Olver et al., 2009). Both instruments require the same information gathering and time to conduct, both had excellent inter-rater reliability among POs in the 1st-year cohort (Vincent, Guy, Fusco, et al., 2012), and there were no differences in their implementation fidelity or impact in the 1st-year cohort (Vincent et al., 2016).

Office Leadership Interview: Continued Delivery of Risk-Needs Assessment and RNR

In-person interviews with leadership at each probation office documented whether the RNAs and RNR continued to be delivered (Question 1) and system or programmatic changes that could have facilitated or prevented the sustainability of RNR-based practices. The audio-

recorded interviews queried significant changes or events over the past seven years in the office and in the state system in the following areas: RNA policies, probation department structure, staff resources, office or stakeholder agency leadership changes (e.g., prosecutor, defense attorney, court), staff training, adoption of new reform practices, and availability of services.

Probation Officer Interview: Adherence to Risk-Needs Assessment and RNR Principles

Consenting POs completed an interview with one of five masters-level research assistants. Interviewers asked POs questions to gauge adherence to RNR-based practices within three case management areas: Disposition recommendations, service referrals, and recommendations following probation violations. Each section started with an open-ended question to understand the factors involved in POs' decision-making without prompting socially desirable responding (e.g., "What information do you consider in your disposition recommendations? What factors are most important?"). Interviewers' audio-recorded, transcribed, and entered the responses into a database. The first author and one interviewer independently coded each of the responses to open-ended questions as 'no mention of the RNA', 'mentioned the RNA directly' (includes reference to the RNA's semi-structured interview), or 'mentioned the RNA indirectly' (mentioned specific dynamic risk factors but did not the RNA or its need areas by name). Final ratings were based on the consensus between coders.

Open-ended questions were followed by four direct questions within each case management area asking POs to rate their use of the SAVRY or the YLS/CMI on an 8-point scale (0 = *Never* to 7 = *Always*). These questions and rating scale were adapted from a national survey designed by Miller and Maloney (2013) and asked how often do they (1) make a recommendation that corresponds with the RNA risk level (or that targets the need areas identified by the RNA for services), (2) make a more restrictive recommendation than the RNA

risk level indicates (or disregard need areas identified on the RNA), (3) make a less restrictive recommendation than the RNA risk level indicates (or target needs not identified on the RNA), and (4) make a recommendation without consulting the RNA. Interviewers asked for an example when POs gave responses of two or higher to questions 1 thru 4. Finally, POs were asked to rate how useful they found the RNA to be for the respective decision on an 8-point scale (0 = *Not at all useful* to 7 = *Extremely useful*).

Administrative Data: Adherence to Risk-Needs Assessment Administration Policies

The researchers used administrative data to record whether and when an RNA had been completed for each youth in both cohorts and whether the timing matched the policy. The NE site's policies were to complete the YLS/CMI with all court referred youth pre-adjudication (with a few site-specific exceptions, such as excluding diverted youth) unless the youth or family refused, in which case the YLS/CMI was to be completed within 10 days after adjudication hearing. The SO site's policies generally stipulated POs were to administer the SAVRY post-adjudication and pre-disposition. SO Site 1 did not require the SAVRY for deferred or unsupervised cases and permitted POs to administer the SAVRY within 10 days of disposition if pre-disposition assessment was not possible. RNA policy adherence rates were calculated as the percentage of youth in each cohort who received an RNA at the time stipulated, divided by the number of youth cases for whom the RNA should have been completed as per the respective sites' policy. The calculations allotted a 10-day window to account for date errors.

Administrative Data: Adherence to the Risk Principle

We evaluated adherence to the risk principle using administrative data from youth in the two cohorts who received the YLS/CMI or SAVRY to examine whether risk level was positively correlated with two case management outcomes: Restrictive placements and number of service

referrals. Restrictive placements included secure (detention, correctional institutions) and non-secure residential facilities, excluding psychiatric or child welfare placements. Services were defined as any community-based or placement-related services aimed at treatment or rehabilitation (e.g., mentoring programs, Functional Family Therapy, counseling), and did not count accountability-related interventions such as community service and electronic monitoring. The study tracked service referrals rather than whether services were received because the referrals were less likely to confound PO and court decisions without service availability or refusals. Placements and service referrals were recorded until each youths' case was closed, or to the end of the study follow-up period, whichever came first. The minimum follow-up periods were held constant for both the 1st- and 7th-year cohorts within each site, ranging from seven (SO Site 1) to 11 months (NE Site 1). However, average lengths of follow-up still varied by cohort, making it essential to control for follow-up days when comparing cohorts.

Data Analyses

To examine whether both states had sustained or improved practitioner behavior change (Question 2) along the three measures of adherence researchers conducted all analyses at the state-level, except adherence to RNA administration policies due to site-specific policy differences. To examine PO self-reported adherence in the 7th-year, we calculated mean PO responses on the 8-point scale questions and the proportion that mentioned using their RNA for each area of decision-making according to researcher consensus ratings. To test the hypothesis of the absence of state-level differences in PO adherence to RNR, analyses first identified covariates to include in state comparisons by examining significant state differences in PO characteristics (using chi-squares and *t*-tests) and testing these in regression models to identify variables significantly associated with the self-report measures of RNR adherence. Next,

researchers tested state differences using chi-squares for dichotomous consensus ratings (logistic regressions if covariates were present) and *t*-tests for POs' 8-point scale ratings (Analyses of Covariance if covariates were present).

For the second measure of practitioner behavior change, chi-squares compared RNA policy adherence rates (number of youth receiving the RNA at the correct time/total number of new youth eligible to receive the RNA) between matched cohorts within sites. Rates were quantified as significantly increasing (improved), staying constant (sustained), or significantly decreasing (regressed). For explanatory purposes, we also calculated overall rates of completion of the RNA (number of cases that received an RNA/total number of youth eligible to receive an RNA), regardless of whether the RNA was completed at the time dictated by policy.

For the last measure of practitioner behavior change, researchers examined adherence to the risk principle within each cohort by determining whether risk level was positively correlated with youth ever receiving a restrictive placement (chi-squares) and their number of service referrals (analysis of variance; ANOVA). Placement and service data analyses excluded youth who were lost at follow-up due to transfer to adult court or moving out of state ($n = 38$) or who never received an RNA, and therefore, did not have a risk level ($n = 198$). Moderated hierarchical regressions (logistic for placement data and linear regression for service data) tested whether cohort membership (1st-year or 7th-year) moderated the associations between risk level and the respective outcome. The first blocks included the PSM weights, days of follow-up (to account for variable follow-up lengths), and any unmatched variables from the PSM identified as covariates. The second blocks included risk level (0 = low, 1 = moderate, and 2 = high/very high) and cohort (0 = 1st-year or 1 = 7th-year) and the third blocks added an interaction term for risk level by cohort. Significant interaction terms indicated adherence either improved or

regressed in the 7th-year, depending on the direction of the term. Non-significant interactions indicated sustained adherence. These patterns were used to gauge state-level differences in the last two measures of behavior change.

Results

Continuing to Deliver the RNA and RNR Practices (Question 1)

According to the leadership interviews, consistent with the hypothesis, both states were using their RNA across all counties seven years after their initial implementation. All five study sites were using the RNAs office-wide, including the two sites that started by piloting the YLS/CMI in only part of their office seven years prior. No probation office amended its policies regarding how RNAs were to be considered in decisions (e.g., case plans should include youths' top need areas identified by the RNA). Only one office significantly changed its RNA administration policy. Four years after initial implementation, NE Site 3 upgraded its policy to complete the YLS/CMI pre-adjudication rather than post-disposition. This was in alignment with the other state offices and enabled risk and needs to influence more case processing decisions.

Programmatic and Contextual Changes Influencing Sustainability (Question 1)

According to statements from Northeastern site leadership, many significant state-level initiatives occurred to promote juvenile justice system reform following adoption of the YLS/CMI. The juvenile act was modified to require juvenile justice agencies to use evidence-based practices and indicated probation offices would not receive grants unless they adopted the YLS/CMI. A state entity designed the Juvenile Justice System Enhancement Strategy, which outlined a process of system reform that included the YLS/CMI, RNR and motivational interviewing, followed by changing youth behavior through quality case management, evidence-based practices, graduated response, and skill-building. This entity provided financial resources

for probation offices to get advanced training on case management and skill-building, including statewide adoption of the Carey Guides Brief Intervention Tool. The state entity and probation chief's committee instituted a process for biannual recertification of state YLS/CMI master trainers and PO booster trainings twice a year on the YLS/CMI and case plan training.

At the site-level, each probation office used state resources to maintain internal master trainers who in turn provided YLS/CMI booster trainings annually. Each probation office instituted routine supervisory reviews of the quality of YLS/CMIs. NE Sites 1 and 2 also implemented supervisory reviews to ensure the services recommended by POs were addressing needs identified on the YLS/CMI. Additionally, NE Site 2 supervisors obtained rigorous supervisor training and added the quality of POs' YLS/CMIs and need-to-service matching in case plans as considerations in job performance evaluations. Since implementing the YLS/CMI in 2009, every site had changes in its probation leadership, eliminated their intensive probation units, adopted graduated response protocols, received motivational interview training, and had changes in their judiciary and assistant district attorneys, some of whom were "tougher on crime". Law enforcement in NE Site 3's district implemented a school diversion program, resulting in substantial reductions in delinquency referrals. This reduction made it feasible for the office to conduct the YLS/CMI with all youth pre- instead of post-adjudication. NE Sites 1 and 2 had officers trained in the Carey Guides to promote needs-based supervision contacts. The chief in NE Site 1 stated that the "job of probation looks different, with the YLS/CMI and motivational interviewing being fully integrated into their work now."

In the Southern state, the two study sites were among a handful of locally operated probation offices but were still affected by some state-level changes. There were two relevant state-level initiatives since their first year of implementation. First, the payer of community

services shifted from the state juvenile justice agency to multiple managed care organizations. These organizations covered evidence-based programs, such as Multi-Systemic Therapy, which led to an increase in their availability and use. Second, in 2013 the state adopted detention standards that reduced the number of youths who could be sent to these facilities. Both sites had changes in their judiciary, probation leadership, and assistant district attorneys, one of which increased use of post-petition diversion (SO Site 1). Both sites had adopted motivational interviewing and graduated response approaches. Unlike the Northeastern state, there were no statewide initiatives to maintain master trainers or promote case planning skills. SO Site 1 gave POs booster trainings on the SAVRY annually, included SAVRY and case plan training as part of the orientation for new POs, and maintained quality assurance by supervisors routinely checking SAVRYs and case plans. Conversely, SO Site 2 only had one SAVRY booster since 2010, and supervisors routinely checked the quality of case plans but not SAVRYs. The Southern state's POs had completed significantly less RNA and RNR training than their Northeastern counterparts (see Table 1).

Probation Officer Behavior Change by State: Self-Reported Adherence (Question 2)

As indicated in Table 1, POs in the Northeastern state had significantly more experience working in juvenile justice and in their current position than POs in the Southern state. Regression models indicated months of experience was significantly associated with POs' use of their RNA in two areas: consensus ratings for POs' use of their RNA in service recommendations; $B = .01(.01)$; $\text{Exp}[B] = 1.01$; 95% CI [1.00, 1.01], $p = .05$; and POs' ratings for making service recommendations without consulting their RNA; $B = .27(.01)$; 95% CI [.001, .01], $p = .02$. Thus, we included months of experience as a covariate for adherence to RNR

in service recommendations to remove the effects of state differences in PO characteristics before testing our hypothesis that states would not differ in POs' self-reported adherence.

Table 2 provides the results of PO interviews overall and by state. The number of PO respondents varied by case management area because not every PO had a role in each area of decision-making. With respect to open-ended questions (first row within each case management area in Table 2), over half of POs (63.77%) referenced the RNA for disposition recommendations, approximately 55% for service recommendations, and 26% when recommending responses to a violation. Overall, POs' mean responses on rating scales indicated the RNA affected their decisions most of the time ($M = 5.14$ for probation violations to $M = 6.19$ for services; $7 = Always$). Moreover, ratings indicated their recommendations infrequently departed from the RNA risk level or need areas ($M = 1.12$ for disregarding needs to $M = 2.10$ for being more restrictive with probation violations). Examples of circumstances in which POs stated they might be more restrictive than the RNAs' risk level included sex offender cases, cases where they felt the family needed intensive services, or when the youth acquired new charges. A few POs provided an example indicating a general orientation towards punishment; "Sometimes you have to do that to make a point to the juvenile or sometimes they just don't get it." The most frequent departure from the RNA was to target needs not identified on the RNA with services ($M = 2.91$, $SD = 1.90$). Examples of such instances according to these POs often involved responsivity factors, such as mental health or learning disabilities, or needs coming to light after the initial assessment; "There may be things that they [youth] don't answer truthfully on the SAVRY, or that their parents don't know about ... Or things that we find out about..." With respect to usefulness ratings, the lowest mean rating was for violations ($M = 4.72$, $SD = 2.06$).

Contrary to the hypothesis that states would not differ, *t*-test or ANCOVA results identified significant state-level differences in POs' self-reported adherence to RNR in every case management area with medium effect sizes. Southern POs gave significantly higher ratings than Northeastern POs when asked if disposition and violation recommendations corresponded with risk level (see Table 2). Northeastern POs were most apt to make more restrictive recommendations following violations than the risk level would indicate. With respect to service recommendations, Northeastern POs mentioned using their RNA almost three times as often as Southern POs upon open-ended questioning ($B = -1.222(.512)$; $\text{Exp}[B] = 0.295$; $p = .017$; 95% CI [.108, .804]), but also reported making recommendations without consulting their RNA more often, on average ($F(2,70) = 4.30$, $p = .017$, $\eta_p^2 = .11$). With respect to usefulness ratings, the only state-level difference was that Southern POs gave significantly higher ratings for the usefulness of the RNA in disposition recommendations than Northeastern POs.

Practitioner Behavior Change: Adherence to RNA Policies (Question 2)

Table 3 provides the YLS/CMI or SAVRY completion and policy adherence rates for each cohort by site with chi-square results comparing policy adherence rates. Contrary to the hypothesis, states significantly differed such that Northeastern sites significantly improved and Southern sites regressed. NE Site 3 had marked improvement with youth in the 7th year cohort being over four times as likely as those in the 1st year cohort to receive the YLS/CMI pre-disposition. NE Site 1 significantly improved by about 10%, which was a smaller effect (OR = 1.64), but this site had less improvement to make than NE Site 3. Change in adherence could not be evaluated for NE Site 2 due to its partial implementation procedure in the 1st year of implementation; however, it had excellent adherence in the 7th year (95.19%). In the Southern state, SO Site 1 significantly improved its rates of policy adherence (OR = 3.01), however, SO

Site 2 regressed from approximately 57% of its youth having a completed SAVRY pre-disposition in the 1st year to only 25% in the 7th year (OR = .25).

Practitioner Behavior Change: Adherence to the Risk Principle (Question 2)

Among youth with a completed YLS/CMI in the Northeastern state's 1st-year cohort ($n = 334$), the risk levels were 48.50% low, 45.21% moderate, 6.29% high. The 7th-year cohort ($n = 412$) had a significantly higher percentage of low-risk (57.52%), and lower percentage of moderate risk (36.17%) youth than the 1st-year cohort ($\chi^2(2) = 6.60$, $V = .04$, 95% CI [.03, .17], $p = .04$). Among youth with a completed SAVRY in the Southern state's 1st-year cohort ($n = 260$), the risk levels were 40.38% low, 43.46% moderate, 16.15% high. The percentage of youth in the 7th-year cohort with a SAVRY ($n = 248$) rated as low risk (30.24%) was significantly lower than the 1st-year cohort ($\chi^2(2) = 5.94$, $\phi = .11$, 95% CI [.04, .19], $p = .05$). SO Site 2 significantly decreased from 13 (1st-year) to only two high-risk youth (7th-year) while the portion of high-risk youth in SO Site 1 increased (14.15% to 27.22%).

Table 4 reports results of adherence to the risk principle for placement and service decisions by cohort and state. Overall rates of youth going to restrictive placements stayed relatively constant in both states between the first and seventh years of RNA implementation (SO = 31.90% versus 32.69%; NE = 18.70% versus 21.09%, respectively). In the Northeastern state, chi-square analyses indicated risk was significantly related to whether youth received a placement in both cohorts (see Table 4). A moderated logistic regression analysis including the PSM weights, days of follow-up, and unbalanced covariates (any violent priors, number of prior offenses), demonstrated risk level was significantly associated with placements at the second block (Exp[B] = 2.41, 95% CI [1.42, 4.07], $p = .001$), and the interaction term between cohort and risk level was not (Exp[B] = 1.14, 95% CI [0.57, 2.29], $p = .71$), indicating sustained

adherence over time. In the Southern state, placements were significantly related to risk level in the 1st-year but not in the 7th-year (see Table 4) where low-risk youth were as likely to be placed (32.00%) as high-risk youth (34.00%). A moderated hierarchical logistic regression including the PSM weights and days of follow-up (there were no covariates) demonstrated risk level was significantly associated with placements at the second block ($\text{Exp}[B] = 1.93$, 95% CI [1.28, 2.91], $p = .002$), but the interaction term between cohort and risk level also was significant ($\text{Exp}[B] = 0.51$, 95% CI [.29, .91], $p = .02$), indicating a significant regression in adherence.

With respect to service referrals, ANOVA results indicated the number of referrals significantly increased across youths' risk level in both states in the 1st year cohorts, but with only a small effect in the Southern state ($\eta^2 = .03$; see Table 4). In the Northeastern state, the association remained significant in the 7th year cohort with a medium effect ($\eta^2 = .08$). A moderated hierarchical linear regression including PSM weights, days of follow-up, and unbalanced covariates (any violent priors, number of prior offenses, any prior mental health treatment, and any prior substance use treatment) demonstrated risk level was significantly associated with service referrals in the Northeastern state at the second block ($\beta = .09$, 95% CI [.03, .54], $t = 2.21$, $p = .03$), and the interaction term between cohort and risk level was not significant ($\beta = -.18$, 95% CI [-.62, .33], $t = -.60$, $p = .55$), indicating the association between risk and services sustained over time. In the Southern state's 7th-year cohort, although the average number of referrals remained in the expected direction with high-risk youth receiving an average of 3.02 ($SD = 2.01$) referrals and low-risk youth receiving 2.23 ($SD = 2.29$), the association was not significant. Consequently, the moderated regression including days of follow-up and PSM weights (there were no other covariates) demonstrated risk level was not significantly associated with services at the second block ($\beta = -.70$, 95% CI [-3.78, 0.41], $t = -1.58$, $p = .11$), but the

interaction term between cohort and risk level was significant ($\beta = -.48$, 95% CI [-8.48, -.18], $t = -2.05$, $p = .04$) indicating adherence regressed. Results of covariate analyses and regressions are included in Supplementary Materials Tables SC and SD.

Discussion

This was the first multi-site study to examine the sustainability of RNR and RNA in probation offices, using four sustainability constructs (Moore, et al., 2017). Prior studies reported POs do not use their RNAs in decisions regularly and/or do not follow the RNR model (e.g., Dyck et al. 2018; Viglione et al., 2015), but whether these findings are a result of initial poor implementation or lack of sustainability is unknown. The current study addressed this knowledge gap by revisiting five probation offices seven years after a rigorous, stakeholder-engaged implementation of RNA and RNR to examine whether states sustained, improved, or regressed over time, as well as facilitators and barriers to sustainability.

Continued Delivery of Risk-Needs Assessment and RNR Principles

Consistent with the hypothesis that both states would have sustained their delivery of RNA and RNR, all five probation offices were still delivering their RNA and some RNR-based practices. In fact, the two offices that originally started with only partial office pilots moved to full implementation of their RNA by the seventh year. Moreover, both states had rolled out implementation of their RNAs in juvenile probation statewide. One initially ineffective probation office (NE Site 3) with an original YLS/CMI administration policy that was not conducive to RNR-based practices vastly improved its policy and adjusted its probation structure to conduct assessments earlier in the process. This agency had considerably poor adherence to practices in its first year (less than 40% of youth had received a YLS/CMI) yet achieved striking gains in implementation fidelity, a 95% RNA

administration adherence rate by year seven, due to policy changes from new leadership.

Evolving and Adapting the Evidence-Based Practices

Since initial implementation, programmatic changes in RNA administration and RNR-based practices occurred indicating agencies adapted practices over time. Leadership interviews revealed clear state differences—the Northeastern sites benefited from statewide initiatives that promoted evidence-based practices. The sociopolitical climate in the state enabled strong implementation of RNR practices that evolved through the adoption of complementary innovations (e.g., motivational interviewing, training POs in skill-building) and statewide support to maintain fidelity to the YLS/CMI and case planning through booster trainings, internal coaching, and quality assurance.

Conversely, in the Southern state, state-level initiatives were fewer, but increased accessibility to evidence-based community services and limited use of detention. Data indicated these changes were not followed by reductions in rates of restrictive placements and the number of services provided to youth increased regardless of risk level. There was no state-level support for quality assurance strategies or staff training. The office that did not sustain internal training or strong quality assurance procedures seemingly experienced a degradation in the validity of SAVRYs given the substantial decline in youth rated as high-risk (13 in the 1st-year, 2 in the 7th-year) and increase in youth rated low-risk (56.52%). This risk composition is inconsistent with national reductions in arrest (presumably for lower risk youth) and was inconsistent with the other Southern office.

Probation Officer Behavior Change: Self-Reported Adherence to RNR

The hypothesis that states would not differ in PO reported adherence was partially supported. Overall, POs reported relatively high adherence to the risk principle for

disposition recommendations and to the need principle (> average 5 on an 8-point scale, with higher numbers indicating stronger adherence). However, the Southern state was significantly less likely than the Northeastern state to report use of their RNA in service decisions in their open-ended responses. The limited use of the RNA in case planning may partially explain the Southern states' regressed adherence to the risk principle in numbers of services referred; by the 7th-year, low-risk youth were receiving about as many services as high-risk youth in the South. Open-ended questions across POs indicated departures from criminogenic needs in service decisions were generally to address responsivity factors (e.g., mental health needs), which is consistent with RNR and may represent an evolution in practice since both states adopted mental health screening. Adherence to the risk principle when recommending responses to probation violations was low across the board. POs reported lower perceptions of the usefulness of RNAs in probation violation recommendations, possibly explaining the lack of change in restrictive placement rates in both states (most occur following a violation of probation rather than at disposition). The favorable perceptions of the RNAs usefulness in other decisions may be a sign that RNR practices had been institutionalized to some degree in both states, and POs were not merely complying as has been found nationally (Miller & Maloney, 2013) but were in fact bought in. An experimental study of POs across this Northeastern state provides further evidence of strong PO adherence to risks and needs, with a few departures (Miller & Palmer, 2020).

Broad Practitioner Behavior Change: Adherence to RNA Policies and Risk Principle

We examined broader practitioner behavior change via adherence to administration of the RNA and adherence to the risk principle, which is dependent on both PO behavior and an enabling court environment. For example, in states like the Southern state where

probation attempts to complete an RNA with youth's post-adjudication and pre-disposition, whether the RNA is completed within this timeframe is dependent on the court's willingness to allow the time in-between hearings. Only judges have the authority to impose restrictive placement decisions, which may or may not follow PO recommendations. Likewise, the number of services to which youth are referred is somewhat controlled by POs but most juvenile courts include the services youth are to attend in the court orders, meaning it is partly or wholly determined by the court.

Inconsistent with the hypothesis that states would not differ, there were meaningful state differences in the maintenance of practitioner behavior change. Despite the apparent institutionalization of the RNA and RNR among POs in both states, administrative data painted a different picture of the sustained fidelity of practices. The Northeastern state significantly improved its adherence to administering the RNA to every youth on time (pre-adjudication) and sustained its strong adherence to the risk principle in decisions. In part, this was due to a change in probation leadership in NE site 3, which led to a significant change in administration policy (from post to pre-disposition). Incidentally, this new chief was the only leader who was not employed in the office for the 1st year cohort.

Results were opposite in the Southern state. Leadership interviews indicated there was less buy-in to the RNR approach among their new judges and district attorneys than in the previous court administration, which was reflected in their administrative data. The Southern office (SO Site 2) with the strictest policy, which was reliant on judges bifurcating adjudication and disposition hearings, had a significant regression in policy adherence (57% to 25%) and consequently, the lowest RNA completion rate of any site (75% of eligible youth). SO Site 1 had improved administration adherence, but this was mainly the result of

an influx in youth handled informally, in which case the SAVRY was not required. SO Site 1's administration policy was more feasible than SO Site 2 in that POs could complete the RNA within 10 days post-disposition if necessary.

The approach of conducting RNAs post-disposition may have been more feasible than pre-disposition RNAs but it was clearly not more effective. Collectively across both sites in the Southern state, adherence to the risk principle in case management significantly regressed over time. In the 7th-year cohort, low-risk youth received too many services and were just as likely to be placed as high-risk youth. In part, this may have been a result of the seemingly low validity of PO SAVRY ratings in SO Site 2 (only two youth in the 7th-year cohort were rated as high risk), indicating youth may have been falsely designated as low risk. However, the pattern was similar in SO Site 1 where there was no evidence of low-validity SAVRY ratings. It seems courts, and possibly POs, were frequently making decisions without the benefit of a valid RNA in both jurisdictions, either because they did not have it or because they did not see a reason to use it. In sum, the Southern state's sustainability was hindered by system changes and turnover in court personnel resulting in an absence of court buy-in, as well as insufficient quality assurance and staff monitoring.

Study Limitations

This study has several limitations. First, this study speaks more to sustainability of the risk principle than the need or responsivity principles, which are considerably more complex (e.g., Drawbridge et al., 2019; Vieira et al., 2009), involve examining additional stakeholders and barriers due to service availability (Haqanee et al., 2014), and were beyond the scope of this paper. Second, we did not control for site-level differences in the analyses of PO data due to small numbers in some sites (n ranged 8 to 22). Third, we were

unable to achieve acceptable balance in the PSMs between cohorts in two sites but accounted for this by including covariates in the adherence to risk principle cohort-comparisons. Fourth, data from only two time-points is a limitation that inhibits measurement of fluctuations over time, particularly with respect to use of the RNA in placement and service-related decisions. However, it was clear the delivery of RNAs and RNR continued for the entire seven-year period and that buy-in by POs was maintained despite any potential fluctuations. Nonetheless, the lack of multiple time points prevents pinpointing which systemic or practice changes had the largest influence on sustainability.

Adherence rates to administration of the RNA for Northeastern offices, where youth receive the YLS/CMI prior to adjudication, were not able to account for youth or parent refusals to be interviewed prior to an adjudication hearing or for youth not showing up to appointments. Situations like these and occasional unconventional court practices (e.g., youth may be automatically disposed to a placement or sent to adult court) lead to the strong argument that offices should not be expected to have a 100% adherence to RNA administration policies. Eighty-five percent is an effective and feasible benchmark that still results in a positive impact on case processing (Drawbridge et al., 2019). Finally, on the surface, it may appear state differences in sustainability were due to use of different instruments, particularly because the SAVRY involves some judgment. This judgment may have led to inaccurate risk levels or less use of need areas (which are less clear than in the YLS/CMI) leading to greater reliance on judgment by POs in decisions. It is difficult to disentangle effects of instrumentation differences from the effects due to dramatic differences in a) ongoing and training and quality assurance procedures, and b) court and prosecutor buy-in, both of which were significantly poor in one SO site yet strong in all NE

sites. The difference in instrumentation did not result in state-level differences in fidelity to the RNA or RNR procedures in the 1st-year cohorts. This points to the more plausible explanation of sustainability differences being due to dramatic state-level differences in ongoing training and quality assurance practices. For example, the validity of SAVRY risk levels appeared to have degraded in SO Site 2 (only two high-risk youth in the 7th-year) but not in SO Site 1 where more training was provided.

Conclusions and Recommendations

Overall, this study of sustainability of RNA and RNR principles adds further support to single-state findings regarding successful juvenile justice reforms when strong implementation procedures are followed (Rocque et al., 2014). Research consistently demonstrates that effective implementation strategies for evidence-based practices are critical for achieving positive outcomes in real world settings (Fixsen et al., 2009). The initial strategies followed by probation offices in this study included orientations and involvement with multiple stakeholder groups to promote buy-in, development of RNR-based policies and decision-support tools (e.g., case plans aligned with need and responsivity areas, service matrix), and facilitation of a master trainer model and supervisory reviews. The initial implementation and intervention-level outcomes (reference removed for blind review) far surpassed that of post-hoc observational studies in sites with initially strong adherence to their RNA administration procedures (e.g., Dyck et al., 2018; Viglione, 2019). Also, this study demonstrated agencies that struggle with implementation initially can evolve and achieve implementation fidelity and strong outcomes (e.g., NE Site 3).

The state differences in sustainability uncovered by the current study points to facilitators and barriers to sustaining RNR-based practices. These fell into the three categories of drivers identified by implementation science (Fixsen et al., 2019). Drivers included the benefits of

supportive leadership at the state and local levels innovating practices through the adoption of complementary initiatives (leadership drivers), regular local training and supervisory monitoring (competency drivers), and continued buy-in from system partners, namely court officials (organizational drivers). Conversely, the Southern state demonstrated some barriers to sustainability, including turnover in judges and district attorneys in the absence of education from probation leadership about the importance of the RNR-approach, and lack of supervisory quality assurance measures and ongoing training. Key implications of the findings include:

1. Poor implementation outcomes in the first year or two of implementing RNAs and RNR does not preclude later implementation success if the sociopolitical climate and leadership enables and promotes evidence-based practices.
2. Sustaining several aspects of RNR requires behavior change from not only probation officers and strong leadership, but also from court stakeholders (Harvell et al., 2018). Restrictive placements are dependent on the courts, and in large part, so are service referrals due to the inclusion in court orders. These aspects of case management are unlikely to adhere to the risk principle if the court is not bought into and/or is not receiving RNA information pre-disposition unless the courts permit POs to have control over service planning. Studies should attend to whether courts are enabling or disabling.
3. Probation leaders should adopt an ongoing strategy for educating court personnel on their RNA and RNR to manage turnover. Promoting buy-in for evidence-based practices must be ongoing, it is not a “one and done”. State-level support for evidence-based practices can have a strong influence in this regard.
4. Policies to administer the RNA prior to adjudication are more effective and sustainable than post-adjudication/pre-disposition policies because they are not dependent on the

courts delaying the disposition hearing and ensure courts have access to risk and need information prior to disposition decisions. Any pre-adjudicatory procedures must be paired with protections against self-incrimination (see Vincent et al., 2012).

5. Coaching, booster training, and supervisory coaching or quality assurance within probation offices are critical; state-supported training will promote sustainability.
6. Evolution and adaptation are important parts of sustainability (Moore et al., 2017; Taylor, 2005) and are essential as the real-world multi-level, context in which these practices are delivered changes over time. Likewise, our benchmarks for successful implementation and sustainment should account for these changes. We should not expect 100% adherence to RNA administration policies or to the risk and need principle, given the importance of integrating the responsivity principle.

Future research should examine whether agencies maintaining RNR practices sustain the benefits for justice-involved youth (Moore et al., 2017). These benefits may be conceptualized via intervention-level outcomes such as reductions in supervision levels and rates of formal processing, reductions in use of restrictive placements, improved need to service matching, and reductions in recidivism. Studies of the sustainability of justice reform efforts are scarce. Future sustainability studies should focus attention towards developing sophisticated measures of practice adaptations, the impact of changing contextual factors, and which implementation supports are most effective for sustaining practice (e.g., written policies, coaching, stakeholder buy-in, data-tracking and continuous quality improvement methods).

References

- Andrews, D. A., & Bonta, J. (2010). Rehabilitating criminal justice policy and practice. *Psychology, Public Policy, and Law, 16*(1), 39–55. <https://doi.org/10.1037/a0018362>
- Borum, R., Bartel, P., & Forth, A. (2006). *Manual for the Structured Assessment for Violence Risk in Youth (SAVRY)*. Psychological Assessment Resources.
- Brookhart, M. A., Schneeweiss, S., Rothman, K. J., Glynn, R. J., Avorn, J., & Stürmer, T. (2006). Variable selection for propensity score models. *American Journal of Epidemiology, 163*(12), 1149–1156. <https://doi.org/10.1093/aje/kwj149>
- Carroll, C., Patterson, M., Wood, S., Booth, A., Rick, J., & Balain, S. (2007). A conceptual framework for implementation fidelity. *Implementation Science, 2*(1), 40. <https://doi.org/10.1186/1748-5908-2-40>
- Drawbridge, D., Todorovic, K., Winters, G., & Vincent, G.M. (2019). Implementation of risk-need-responsivity principles into probation case planning. *Law and Human Behavior, 43*, 455-467. <https://doi.org/10.1037/lhb0000351>.
- Dyck, H. L., Campbell, M. A., & Wershler, J. L. (2018). Real-world use of the risk-need-responsivity model and the Level of Service/Case Management Inventory with Community-Supervised Offenders. *Law and Human Behavior, 42*, 258-268.
- Fabelo, T., Arrigona, N., Thompson, M. D., Clemens, A., & Marchbanks, M. P. (2015). Closer to home: An analysis of the state and local impact of the Texas juvenile justice reforms. *New York, NY: Council of State Governments Justice Center*.
- Fixsen, D. L., Blase, K. A., Naoom, S. F., & Wallace, F. (2009). Core implementation components. *Research on Social Work Practice, 19*(5), 531-540. <https://doi.org/10.1177/1049731509335549>

Fixsen, D. L., Blase, K. A., & Van Dyke, M. K. (2019). *Implementation practice & science*. Active Implementation Research Network.

Haas, S. M., & DeTardo-Bora, K. A. (2009). Inmate reentry and the utility of the LSI-R in case planning. *Corrections Compendium*, 34(1), 11-16.

Haqanee, Z., Peterson-Badali, M., & Skilling, T. (2014). Making “what works” work: Examining POs’ experiences addressing the criminogenic needs of juvenile offenders. *Journal of Offender Rehabilitation*, 54(1), 37-59.
<https://doi.org/10.1080/10509674.2014.980485>

Harvell, S., Love, H., Pelletier, E., Warnberg, C., Willison, J. B., & Winkler, M. K. (2018). *Bridging research and practice in juvenile probation*.
<https://ojjdp.ojp.gov/sites/g/files/xyckuh176/files/pubs/252234.pdf>

Hockenberry, S., and Puzanchera, C. (2019). *Juvenile Court Statistics 2017*. Pittsburgh, PA: National Center for Juvenile Justice.

Hoge, R. D. & Andrews, D. A. (2011). *Youth Level of Service/Case Management Inventory 2.0*. Multi-Health Systems.

Ingel, S. N., Davis, L. R., Rudes, D. S., Hartwell, T. N., Drazdowski, T. K., McCart, M. R., Chapman, J. E., Taxman, F. S., & Sheidow, A. J. (2022). Misunderstanding and sensemaking among juvenile probation officers working with evidence-based practices. *Victims & Offenders*, 17(7), 975–993. <https://doi.org/10.1080/15564886.2021.1985670>

Juvenile Justice Geography, Policy, Practice & Statistics (JJGPPS). Online. Available: <http://www.jjgps.org/juvenile-justice-services#risk-assessment?year=2020>. Developed by the National Center for Juvenile Justice (NCJJ), with funding from the John D. and Catherine T. MacArthur Foundation.

- Loman, S.L., Rodriguez, B.J., & Horner, R.H. (2010). Sustainability of a targeted intervention package: First step to success in Oregon. *Journal of Emotional and Behavioral Disorders, 18*(3), 178 – 191.
- Miller, J. & Maloney, C. (2013). Practitioner compliance with risk/needs assessment tools: A theoretical and empirical assessment. *Criminal Justice and Behavior, 40*(7), 716-736.
<https://doi.org/10.1177/0093854812468883>
- Miller, J., & Palmer, K. (2020). Juvenile probation officer decision-making in a reforming state: Assessing the application of evidence-based principles. *Criminal Justice and Behavior, 47*(9), 1136-1155. <https://doi.org/10.1177/0093854820925112>
- Moore, J. E., Mascarenhas, A., Bain, J., & Straus, S. E. (2017). Developing a comprehensive definition of sustainability. *Implementation Science, 12*(1), 1-8.
<https://doi.org/10.1186/s13012-017-0637-1>
- Moullin, J. C., Sklar, M., Green, A., Dickson, K. S., Stadnick, N. A., Reeder, K., & Aarons, G. A. (2020). Advancing the pragmatic measurement of sustainment: a narrative review of measures. *Implementation Science Communications, 1*(1), 1-18.
<https://doi.org/10.1186/s43058-020-00068-8>
- National Research Council. (2013). *Reforming Juvenile Justice: A developmental approach*. Washington, DC: The National Academics Press.
<https://doi.org/10.17226/14685>
- Olver, M. E., Stockdale, K. C., & Wormith, J. S. (2009). Risk assessment with young offenders: A meta-analysis of three assessment measures. *Criminal Justice and Behavior, 36*(4), 329-353. doi:10.1177/0093854809331457
- Proctor, E. K., Powell, B. J., & McMillen, J. C. (2013). Implementation strategies:

- recommendations for specifying and reporting. *Implementation Science*, 8(1), 1-11.
<https://doi.org/10.1186/1748-5908-8-139>
- Proctor, E., Silmere, H., Raghavan, R., Hovmand, P., Aarons, G., Bunger, A., Griffey, R., & Hensley, M. (2011). Outcomes for implementation research: conceptual distinctions, measurement challenges, and research agenda. *Administration and Policy in Mental Health and Mental Health Services Research*, 38(2), 65-76.
<https://doi.org/10.1007/s10488-010-0319-7>
- Rocque, M., Welsh, B. C., Greenwood, P. W., & King, E. (2014). Implementing and sustaining evidence-based practice in juvenile justice: A case study of a rural state. *International Journal of Offender Therapy and Comparative Criminology*, 58(9), 1033-1057.
<https://doi.org/10.1177/0306624X13490661>
- Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1), 41-55.
<https://doi.org/10.1093/biomet/70.1.41>
- Rubin, D. B. (2001). Using propensity scores to help design observational studies: Application to the tobacco litigation. *Health Services and Outcomes Research Methodology*, 2(3-4), 169-188. <https://doi.org/10.1023/A:1020363010465>
- Schafer, J. L., & Kang, J. (2008). Average causal effects from nonrandomized studies: A practical guide and simulated example. *Psychological Methods*, 13(4), 279-313.
<https://doi.org/10.1037/a0014268>
- Stirman, S. W., Kimberly, J., Cook, N., Calloway, A., Castro, F., & Charns, M. (2012). The sustainability of new programs and innovations: a review of the empirical literature and recommendations for future research. *Implementation Science*, 7(1), 1-19.

<https://doi.org/10.1186/1748-5908-7-17>

- Stuart, E. A., King, G., Imai, K., & Ho, D. (2011). MatchIt: Nonparametric preprocessing for parametric causal inference. *Journal of Statistical Software*. <https://doi.org/10.18637/jss.v042.i08>
- Taylor, J. E. (2005). *Sustainability: Examining the survival of schools' comprehensive school reform efforts*. Washington, DC: American Institutes for Research.
- Vieira, T. A., Skilling, T. A., & Peterson-Badali, M. (2009). Matching court-ordered services with treatment needs: Predicting treatment success with young offenders. *Criminal Justice and Behavior*, 36(4), 385–401. <https://doi.org/10.1177/0093854808331249>
- Viglione, J. (2019). The risk-need-responsivity model: How do probation officers implement the principles of effective intervention?. *Criminal Justice and Behavior*, 46(5), 655-673. <https://doi.org/10.1177/0093854818807505>
- Viglione, J., Rudes, D. S., & Taxman, F. S. (2015). Misalignment in supervision: Implementing risk/needs assessment instruments in probation. *Criminal Justice and Behavior*, 42(3), 263-285. <https://doi.org/10.1177/0093854814548447>
- Viljoen, J. L., Shaffer, C. S., Muir, N. M., Cochrane, D. M., & Brodersen, E. M. (2019). Improving case plans and interventions for adolescents on probation: The implementation of the SAVRY and a structured case planning form. *Criminal Justice and Behavior*, 46(1), 42-62. <https://doi.org/10.1177/0093854818799379>
- Vincent, G. M., Guy, L. S., & Grisso, T. (2012). *Risk Assessment in Juvenile Justice: A Guidebook for Implementation*. Chicago, IL: John D. & Catherine T. MacArthur Foundation. www.NYSAP.us

- Vincent, G. M., Guy, L. S., Fusco, S.L., & Gershenson, B.G. (2012). Field reliability of the SAVRY with probation officers: Implications for training. *Law and Human Behavior, 36*, 225-236. <https://doi.org/10.1007/s10979-011-9284-2>
- Vincent, G. M., Guy, L. S., Perrault, R. T., & Gershenson, B. G. (2016). Risk assessment matters, but only when implemented well: A multi-site study in juvenile probation. *Law and Human Behavior, 40*, 683-696
- Vincent, G. M., Paiva-Salisbury, M., Cook, N. E., Guy, L. S., & Perrault, R. (2012). Impact of Risk/Needs Assessment on Juvenile Probation Officers' Decision-Making: Importance of Implementation. *Psychology, Public Policy, & the Law, 18*, 546-576.

Table 1
Probation Officer Characteristics by State (N = 84)

	Overall (N = 84)		Northeastern (n = 46)		Southern (n = 38)		Statistic(df)	p	d/O R	95% CI
	n	%	n	%	n	%				
Sample										
# POs interviewed (% all POs)	84	96.55	46	95.83	38	97.44				
# POs interviewed for cohort #1	38	45.24	23	50.00	15	39.50	$\chi^2(1, 84) = .93$.34		
Current Position										
Supervisor	13	15.47	7	15.22	6	15.79				
Field or Intake PO	61	72.62	34	73.91	27	71.05				
Other	10	11.90	5	10.87	5	13.16				
# POs who conduct initial RNAs	77	91.67	45	97.83	32	84.21				
# New POs (< 12 months)	9	10.70	5	10.90	4	10.50	$\chi^2(1, 84) = .03$.96		
Experience										
Months in current position (M, SD)		85.84 (87.71)		103.82 (95.93)		63.97 (71.88)	$t(70) = 2.15$.03	.41	
Months working in JJ (M, SD)		180.34 (108.39)		204.22 (105.30)		151.45 (106.31)	$t(82) = 2.28$.02	.50	
RNA trainings received (M, SD)		7.21 (6.73)		10.83 (7.04)		2.84 (2.33)	$t(56) = 7.22$	< .001	1.52	
Characteristics										
Gender: Female	33	39.29	15	32.61	18	47.37	$\chi^2(1, 84) = 1.90$.17	1.86	[.77, 4.51]
Race/Ethnicity										
Non-Latinx White	57	67.86	39	84.78	18	47.37	$\chi^2(1, 84) = 17.07$	< .001	.45 ^a	[.27, .63]
Non-Latinx Black	20	23.81	3	6.52	17	44.74				
Other (e.g., Latinx, Asian)	7	8.33	4	8.69	3	7.89				
Education										
Some College or Bachelors	52	61.90	29	63.04	23	60.52	$\chi^2(1, 84) = .06$.81	1.11	[.46, 2.69]
Master's degree	32	38.09	17	36.96	15	39.47				
Age (M, SD)		40.65 (9.56)		41.59 (9.39)		39.53 (9.77)	$t(82) = 0.98$.33	.21	

Note. N = Sample size; % = Percent of state sample. Significant differences are in bold. Reference group was Northeastern. ^a = V reported for effect size

Table 2: Probation Officer Interview Ratings for Their Use of Risk-Needs Assessments in Decisions by State

	Overall (<i>N</i> = 84)	NE State (<i>n</i> = 46)	SO State (<i>n</i> = 38)	Statistic(df)	<i>p</i>	<i>d</i> / η_p^2	OR [95% CI]
	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)				
Disposition Recommendations	<i>n</i> = 69	<i>n</i> = 40	<i>n</i> = 29				
Mentioned using RNA ^a	63.77%	65.00%	62.07%	$\chi^2(1) = .06$.80		.88[.33, 2.38]
Rec corresponds with risk	5.52 (1.46)	5.15 (1.66)	6.03 (.94)	<i>t</i> (67) = -2.58	.01	.65	
More restrictive rec than the risk	1.72 (1.35)	1.88 (1.40)	1.52 (1.27)	<i>t</i> (67) = 1.09	.28	.27	
Less restrictive rec than the risk	1.88 (1.53)	2.08 (1.48)	1.62 (1.59)	<i>t</i> (66) = 1.22	.23	.29	
Rec without consulting the RNA	1.00 (1.68)	1.13 (1.86)	.83 (1.42)	<i>t</i> (67) = .72	.47	.18	
Useful for disposition recs	5.03 (1.71)	4.65 (1.85)	5.55 (1.35)	<i>t</i> (67) = -2.23	.03	.56	
Service Recommendations	<i>n</i> = 76	<i>n</i> = 43	<i>n</i> = 33				
Mentioned using RNA	55.27%	69.76%	36.36%	<i>B</i> = -1.22	.02		.29[.11, .80] ^b
Target needs from RNA	6.19 (1.26)	6.24 (1.05)	6.12 (1.49)	<i>F</i> (2,70) = .89	.42	.03 ^c	
Disregard some needs from RNA	1.12 (1.57)	1.40 (1.58)	.76 (1.50)	<i>F</i> (2,70) = 1.70	.19	.05 ^c	
Target needs not identified in RNA	2.91 (1.90)	3.00 (2.08)	2.79 (1.67)	<i>F</i> (2,70) = .70	.50	.02 ^c	
Rec without consulting RNA	1.64 (1.20)	2.10 (2.16)	1.06 (1.62)	<i>F</i> (2,70) = 4.30	.02	.11 ^c	
Useful for service recs	5.63 (1.56)	5.45 (1.48)	5.85 (1.64)	<i>t</i> (73) = -1.10	.28	.25	
Recommending Responses to Probation Violations	<i>n</i> = 69	<i>n</i> = 39	<i>n</i> = 30				
Mentioned using RNA	26.09%	30.77%	20.00%	$\chi^2(1) = 1.02$.31	.56	.56[.18, 1.73]
Rec corresponds with risk	5.14 (1.82)	4.67 (2.02)	5.77 (1.33)	<i>t</i> (67) = -2.58	.01	.64	
More restrictive rec than the risk	2.10 (1.71)	2.46 (1.68)	1.63 (1.65)	<i>t</i> (67) = 2.04	.04	.50	
Less restrictive rec than the risk	1.78 (1.69)	2.03 (1.68)	1.47 (1.68)	<i>t</i> (67) = 1.37	.17	.33	
Rec without consulting the RNA	1.67 (2.25)	2.05 (2.50)	1.17 (1.78)	<i>t</i> (67) = 1.64	.10	.40	
Useful for probation violations	4.72 (2.06)	4.49 (2.23)	5.03 (1.79)	<i>t</i> (67) = -1.10	.28	.27	

Note. Rec = Recommendation. Recommendation ratings ranged from 0 = *Never* to 7 = *Always*, and usefulness ratings ranged from 0 = *Not at all useful* to 7 = *Extremely useful*. Significant findings are in bold. ^a 'Mentioned using RNA' refers to the open-ended questions and consensus ratings. ^b Results are Exp[B] and its 95% CI from logistic regression conducted to control for months of experience working in juvenile justice. Reference group was Northeastern. ^c ANCOVAs conducted to include months in current position as a covariate. Effect size reported is η_p^2

Table 3*Change in Adherence to Risk/Needs Assessment Administration Policies by State and Site*

	1 st Year Cohort		7 th Year Cohort		Comparing Rates of Adherence				Sustainability
	RNA Completion Rate <i>n</i> (%)	Policy Adherence Rate <i>n</i> (%)	RNA Completion Rate <i>n</i> (%)	Policy Adherence Rate <i>n</i> (%)	χ^2 (df)	<i>p</i>	OR	95% CI ^a	
Northeast State (<i>N</i> =433 per cohort)	334 (77.10)	285 (66.90)	412 (95.20)	339 (78.30)					
Site 1 (<i>n</i> = 221)	190 (85.97)	152 (68.78)	209 (94.57)	173 (78.28)	5.13 (1,442)	.02	1.64	[1.07, 2.51]	Improved
Site 2 (<i>n</i> = 104)	104 (NA)	104 (NA)	99 (95.19)	99 (95.19)					NA
Site 3 (<i>n</i> = 108)	40 (37.04)	29 (28.71)	104 (96.30)	67 (62.04)	23.34 (1,209)	< .001	4.06	[2.27, 7.25]	Improved
Southern State (<i>N</i> =297 per cohort)	261 (88.47)	212 (75.18)	248 (83.8)	164 (68.62)					
Site 1 (<i>n</i> = 205)	185 (90.24)	165 (82.50)	180 (87.80)	142 (93.42)	9.24 (1,352)	< .01	3.01	[1.44, 6.30]	Improved
Site 2 (<i>n</i> = 92)	76 (82.61)	47 (57.32)	69 (75.00)	22 (25.29)	17.93 (1,169)	< .001	.25	[.13, .48]	Regressed

Note. NA = not applicable. Completion Rates = number of youth with RNAs completed at any point during supervision/the total number of youth court referrals (in the Northeast) or adjudications (in the South) for which a SAVRY or YLS/CMI was expected as per the sites' policy. Policy adherence Rate = number of youth with RNAs completed at the time dictated by policy (e.g., pre-adjudication, pre-disposition)/the total number of youth court referrals (in the Northeast) or adjudications (in the South) for which it was expected as per the sites' policy. Significant findings are in bold.

^aChi-squares compared rates between the Adherence to Policy Rate cells.

Table 4*Change in Adherence to the Risk Principle in Case Management Areas by State*

% Placed ^a	1 st Year Cohort (Southern <i>n</i> = 257; Northeastern <i>n</i> = 332)						7 th Year Cohort (Southern <i>n</i> = 245; Northeastern <i>n</i> = 403)						Sustained
	Low <i>n</i> (%)	Mod <i>n</i> (%)	High <i>n</i> (%)	χ^2 (<i>df</i> , <i>N</i>)	<i>p</i>	<i>V</i> [95% CI]	Low <i>n</i> (%)	Mod <i>n</i> (%)	High <i>n</i> (%)	χ^2 (<i>df</i> , <i>N</i>)	<i>p</i>	<i>V</i> [95% CI]	
NE State	15 (9.35)	34 (22.67)	13 (61.90)	36.70 (2,332)	< .001	.33 [.22, .46]	21 (9.01)	49 (34.03)	15 (57.69)	55.83 (2,403)	< .001	.37 [.28, .47]	Sustained
SO State	21 (20.00)	40 (35.40)	21 (53.85)	16.13 (2,257)	< .001	.25 [.14, .37]	24 (32.00)	39 (32.50)	17 (34.00)	.06 (2,245)	.97	.02 [.02, .18]	Regressed
Service Referrals ^b	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)	<i>F</i> (<i>df</i> , <i>N</i>)	<i>p</i>	η^2	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)	<i>F</i> (<i>df</i> , <i>N</i>)	<i>p</i>	η^2	
NE State	.89 (1.64)	1.62 (2.20)	2.57 (2.87)	9.71 (2,331)	< .001	.06	1.19 (1.51)	2.03 (1.99)	2.85 (2.17)	17.44 (2,402)	< .001	.08	Sustained
SO State	1.75 (1.74)	2.04 (1.39)	2.49 (1.25)	3.43 (2,256)	.03	.03	2.23 (2.29)	2.37 (2.03)	3.02 (2.01)	2.34 (2,244)	.10	.02	Regressed

Note. *M* = mean, *SD* = standard deviation. Significant findings are in bold.

^aCells represent the percentage of youth within each risk level who received a restrictive placement at any point during supervision.

^bMeans represent the average number of service referrals received by youth within each risk level.

Supplementary Table SA: Propensity-Score Matching Results Following Full Matching Procedure

Southern Site 1: Absolute Standardized Mean Differences Before and After Propensity Score Matching							
Variable	Pre-Matching			Post-Matching			% Improvement
	1 st Year Cohort	7 th Year Cohort	SMD	1 st Year Cohort	7 th Year Cohort	SMD	
	<i>M</i>	<i>M</i>		<i>M</i>	<i>M</i>		
Gender	0.33	0.24	0.19	0.33	0.33	0.01	96.00
Race	0.89	0.85	0.11	0.89	0.88	0.01	87.40
Age at Adjudication	15.10	14.97	0.09	15.10	15.14	0.03	69.50
Age at First Petition	13.85	13.44	0.25	13.85	13.90	0.03	87.80
Violent Index Offense	0.26	0.28	0.04	0.26	0.25	0.01	69.70
Nonviolent Felony	0.19	0.18	0.03	0.19	0.21	0.06	-143.80
Any Violent Prior Petitions	0.20	0.31	0.28	0.20	0.18	0.04	86.90
# Prior Petitions	1.29	1.38	0.04	1.29	1.22	0.03	28.30
Outpatient MH Tx Ever	5.10	127.09	1.75	5.10	10.18	0.07	95.80
Living Arrangement ^a	1.74	1.81	0.16	1.74	1.76	0.06	64.70
Enrolled in School	6.27	11.15	0.07	6.27	6.09	0.00	96.40

Notes. $N_{1st\ Year\ Cohort} = 205$ and $N_{7th\ Year\ Cohort} = 205$. |SMD| = absolute standardized mean difference. Outpatient MH Tx Ever = Any Outpatient Mental Health Treatment Ever (history or current). ^a Living Arrangement was coded as 1 = living with parents or relatives, 2 = living with a single parent/relative, or 3 = other/institution.

Southern Site 2: Absolute Standardized Mean Differences Before and After Propensity Score Matching							
Variable	Pre-Matching			Post-Matching			% Improvement
	1 st Year Cohort	7 th Year Cohort	SMD	1 st Year Cohort	7 th Year Cohort	SMD	
	<i>M</i>	<i>M</i>		<i>M</i>	<i>M</i>		
Race	0.63	0.60	0.06	0.63	0.59	0.07	-14.40
Age at First Petition	12.73	13.50	0.28	12.73	12.48	0.09	68.60
Violent Index Offense	0.21	0.25	0.11	0.21	0.20	0.01	92.60
Any Violent Prior Petitions	0.23	0.28	0.13	0.23	0.22	0.01	92.60
# Prior Petitions	2.24	2.05	0.06	2.24	2.56	0.11	-71.20
Axis I Diagnosis Ever	0.13	0.46	0.97	0.13	0.14	0.02	98.30
Axis II Diagnosis Ever	0.15	0.21	0.15	0.15	0.11	0.13	15.00
Living Arrangement ^a	12.64	1.76	0.10	12.64	1.83	0.10	0.70
Enrolled in School	55.12	3.11	0.23	55.12	4.17	0.22	2.00

Notes. $N_{1st\ Year\ Cohort} = 92$ and $N_{7th\ Year\ Cohort} = 92$. |SMD| = absolute standardized mean difference.

^a Living Arrangement was coded as 1 = living with parents or relatives, 2 = living with a single parent/relative, or 3 = other/institution.

Northern Site 1: Absolute Standardized Mean Differences Before and After Propensity Score Matching							
Variable	Pre-Matching			Post-Matching			% Improvement
	1 st Year Cohort	7 th Year Cohort	SMD	1 st Year Cohort	7 th Year Cohort	SMD	
	<i>M</i>	<i>M</i>		<i>M</i>	<i>M</i>		
Race	0.66	0.26	0.48	0.66	0.49	0.20	58.20
Ethnicity	0.21	0.17	0.10	0.21	0.09	0.30	-198.50
Age at First Petition	15.36	15.27	0.05	15.36	15.19	0.09	-90.10
Violent Index Offense	0.15	0.24	0.24	0.15	0.15	0.01	97.00
Any Violent Prior Petitions	0.08	0.08	0.02	0.08	0.10	0.08	-384.40
# Prior Petitions	0.52	0.40	0.12	0.52	1.14	0.61	-427.00
Axis II Diagnosis Ever	0.10	0.16	0.20	0.10	0.11	0.05	75.40
Living Arrangement ^a	1.71	1.82	0.20	1.71	1.56	0.25	-23.60

Notes. $N_{1st\ Year\ Cohort} = 221$ and $N_{7th\ Year\ Cohort} = 221$. |SMD| = absolute standardized mean difference. ^a Living Arrangement was coded as 1 = living with parents or relatives, 2 = living with a single parent/relative, or 3 = other/institution.

Northern Site 2: Absolute Standardized Mean Differences Before and After Propensity Score Matching							
Variable	Pre-Matching			Post-Matching			% Improvement
	1 st Year Cohort	7 th Year Cohort	SMD	1 st Year Cohort	7 th Year Cohort	SMD	
	<i>M</i>	<i>M</i>		<i>M</i>	<i>M</i>		
Gender	0.20	0.25	0.12	0.20	0.24	0.10	14.20
Race	0.14	0.27	0.28	0.14	0.14	0.02	92.80
Ethnicity	0.05	0.07	0.09	0.05	0.04	0.02	77.90
Age at Adjudication	16.22	16.18	0.03	16.22	16.24	0.01	50.30
Any Violent Prior Petitions	0.01	0.02	0.10	0.01	0.00	0.08	22.50
# Prior Petitions	0.19	0.29	0.12	0.19	0.17	0.03	77.50
Any Axis II Diagnosis Ever	0.06	0.19	0.58	0.06	0.06	0.01	99.00
Outpatient MH Tx Ever	0.90	29.35	13.75	0.90	3.39	1.20	91.30
Outpatient SU Tx Ever	0.64	38.71	21.88	0.64	4.12	2.00	90.90
Living Arrangement ^a	1.91	1.72	0.14	1.91	1.71	0.15	-4.60
Enrolled in School	39.37	13.39	0.13	39.37	8.50	0.16	-18.90

Notes. $N_{1st\ Year\ Cohort} = 104$ and $N_{7th\ Year\ Cohort} = 104$. |SMD| = absolute standardized mean difference, Outpatient MH Tx Ever = Any Outpatient Mental Health Treatment Ever (history or current), Outpatient SU Tx Ever = Any Outpatient Substance Use Treatment Ever (history or current). ^a Living Arrangement was coded as 1 = living with parents or relatives, 2 = living with a single parent/relative, or 3 = other/institution.

Northern Site 3: Absolute Standardized Mean Differences Before and After Propensity Score Matching							
Variable	Pre-Matching			Post-Matching			% Improvement
	1 st Year Cohort	7 th Year Cohort	SMD	1 st Year Cohort	7 th Year Cohort	SMD	
	<i>M</i>	<i>M</i>		<i>M</i>	<i>M</i>		
Gender	0.18	0.21	0.10	0.18	0.14	0.10	-2.60
Race	0.87	0.80	0.19	0.87	0.95	0.20	-4.10
Ethnicity	0.05	0.19	0.66	0.05	0.02	0.11	83.00
Age at Adjudication	16.03	15.80	0.14	16.03	15.97	0.03	76.10
Any Violent Prior Petitions	0.39	0.17	0.46	0.39	0.16	0.47	-2.60
# Prior Petitions	1.38	0.42	0.62	1.38	1.23	0.10	84.40
Outpatient MH Tx Ever	1.79	129.86	41.89	1.79	5.07	1.07	97.40
Living Arrangement ^a	29.50	1.93	0.17	29.50	1.47	0.17	-1.70
Enrolled in School	195.18	0.93	0.49	195.18	1.00	0.49	0.00

Notes. $N_{1st\ Year\ Cohort} = 108$ and $N_{7th\ Year\ Cohort} = 108$. |SMD| = absolute standardized mean difference. Outpatient MH Tx Ever = Any Outpatient Mental Health Treatment Ever (history or current). ^a Living Arrangement was coded as 1 = living with parents or relatives, 2 = living with a single parent/relative, or 3 = other/institution.

Supplementary Table SB: Final Matched Sample Characteristics for Both Cohorts: Overall and by State

	1 st -Year Cohort			7 th -Year Cohort		
	Overall (<i>N</i> = 730)	Southern (<i>n</i> = 297)	Northeastern (<i>n</i> = 433)	Overall (<i>N</i> = 730)	Southern (<i>n</i> = 297)	Northeastern (<i>n</i> = 433)
Gender						
% Female	25.07%	29.29%	22.17%	23.97%	24.24%	23.79%
Race						
% Black	50.96%	78.79%	31.87%	50.68%	75.08%	33.95%
% White	40.68%	20.20%	54.73%	47.26%	23.91%	63.28%
% Other	8.36%	1.01%	13.39%	2.05%	1.01%	2.77%
% Latinx	7.67%	0%	12.93%	9.18%	1.01%	14.78%
Age at Study Start (at time of referral or adjudication)	<i>M</i> = 15.53 <i>SD</i> = 1.84	<i>M</i> = 14.85 <i>SD</i> = 1.95	<i>M</i> = 15.98 <i>SD</i> = 1.61	<i>M</i> = 15.46 <i>SD</i> = 1.79	<i>M</i> = 14.98 <i>SD</i> = 1.72	<i>M</i> = 15.79 <i>SD</i> = 1.77
Index Offense Category						
% homicide	0.14%	0.03%	0.00%	0.00%	0.00%	0.00%
% major sex offense	2.33%	2.69%	2.08%	4.25%	5.05%	3.70%
% robbery or kidnap	3.29%	1.01%	4.85%	3.29%	2.02%	4.16%
% assault/arson intent	22.88%	19.87%	24.94%	20.68%	19.53%	21.48%
% threats or harassment	4.25%	1.01%	6.47%	3.29%	0.34%	5.31%
% minor sex offense	0.14%	0.00%	0.23%	1.10%	0.34%	1.62%
% theft/break & enter/fraud	21.51%	24.24%	19.63%	21.64%	29.63%	16.17%
% arson	0.41%	0.34%	0.46%	0.68%	0.34%	0.92%
% weapons offense	3.42%	4.04%	3.00%	4.11%	3.70%	4.39%
% drug offense	12.47%	6.40%	16.63%	14.38%	5.39%	20.55%
% miscellaneous offenses	19.45%	18.18%	20.32%	20.27%	18.52%	21.48%
% violation	0.27%	0.67%	0.00%	0.00%	0.00%	0.00%
% status offense	8.63%	21.21%	0.00%	6.30%	15.15%	0.23%
% Violent Index Offense	28.63%	23.91%	31.87%	34.90%	31.00%	38.80%
Age at First Petition	<i>M</i> = 14.54 <i>SD</i> = 2.12	<i>M</i> = 13.50 <i>SD</i> = 2.21	<i>M</i> = 15.25 <i>SD</i> = 1.80	<i>M</i> = 14.47 <i>SD</i> = 2.01	<i>M</i> = 13.45 <i>SD</i> = 1.77	<i>M</i> = 15.17 <i>SD</i> = 1.86

% Any Violent Prior Petitions	16.71%	20.54%	14.09%	17.26%	29.97%	8.55%
Mean # Prior Petitions	<i>M</i> = 0.98 <i>SD</i> = 1.73	<i>M</i> = 1.47 <i>SD</i> = 2.21	<i>M</i> = 0.65 <i>SD</i> = 1.21	<i>M</i> = 0.89 <i>SD</i> = 1.51	<i>M</i> = 1.60 <i>SD</i> = 1.87	<i>M</i> = 0.39 <i>SD</i> = 0.93
% Axis I Diagnosis Ever (Y/N)	10.41%	12.79%	8.78% ^a	34.25%	41.75%	29.10%
% Axis II Diagnosis Ever (Y/N)	8.08%	10.44%	6.47% ^a	14.52%	13.13%	15.47%
% Outpatient Mental Health Treatment Ever (Y/N)	26.16%	21.55%	29.33%	31.51%	35.69%	28.64%
% Outpatient Substance Use Treatment Ever (Y/N)	10.14%	3.03%	15.01%	11.92%	3.03%	18.01%
Enrolled in School	90.20%	91.10%	89.60%	93.60%	93.70%	93.50%
Living Arrangement ^a						
% Supervised Setting	25.48%	16.50%	31.64%	23.56%	18.52%	27.02%
% Less Supervised Setting	73.56%	83.16%	66.97%	75.75%	80.47%	72.52%

^a Supervised setting included living with two parents or in an institution or group home, less supervised settings included single-parent households, or living with a relative

Supplementary Table SCa: Logistic Regressions to Identify Covariates for Restrictive Placements by State

	β (SE)	Wald	<i>p</i> -value	Exp(B)	95% CI
Southern State (N = 577)					
Enrolled in school					
Not enrolled		2.66	.26		
Enrolled	-.58 (.35)	2.66	.10	.56	[.28, 1.12]
Graduated/GED	-.54 (1.25)	.18	.67	.59	[.05, 6.78]
Days of follow-up	.01 (.001)	61.33	< .001	1.01	[1.00, 1.01]
Northeastern State (N = 852)					
Ethnicity (Non-Latinx/Latinx)	.38 (.25)	2.19	.14	1.45	[.89, 2.39]
Any prior violent offense	.97 (.25)	14.90	< .001	2.63	[1.61, 4.29]
Living Arrangement					
Parent/relative		.36	.84		
Single parent	.12 (.21)	.35	.56	1.13	[.75, 1.71]
Institution/other	.12 (.42)	.09	.77	1.13	[.50, 2.55]
# Prior Petitions	.48 (.08)	36.70	< .001	1.62	[1.39, 1.89]
Outpatient Mental Health Treatment Ever	.07 (.20)	.11	.74	1.07	[.73, 1.57]
Outpatient Substance Use Treatment Ever	.12 (.23)	.28	.60	1.13	[.72, 1.76]
Enrolled in school					
Not enrolled		3.03	.22		
Enrolled	-.89 (.51)		.08	.41	[.15, 1.12]
Graduated/GED	-.85 (.70)		.23	.43	[.11, 1.69]
Days of Follow-up	.01 (.001)	101.50	< .001	1.01	[1.01, 1.01]

Note. Unmatched variables between the 1st and 7th cohorts within state were tested as covariates with restricted placements. To take a liberal approach, each variable was tested in its own regression model with Days of Follow-up to account for variable follow-up periods across cases. Significant findings are presented in bold.

Supplementary Table SCb: Regressions to Identify Covariates for Number of Services by State

	B (SE)	β	<i>t</i>	<i>p</i> -value	95% CI
Southern State (N = 577)					
School Enrolled	-.14 (.25)	-.02	-.55	.58	[-.62, .35]
Days of Follow-up	.01(< .001)	.40	12.63	< .001	[.01, .01]
Northeastern State (N = 1281)					
Ethnicity	.10 (.17)	.02	.60	.55	[-.23, .44]
Any prior violent offense	.43 (.19)	.07	2.32	.02	[.07, .80]
# Prior Petitions	.11 (.06)	.07	2.05	.04	[.005, .22]
Living Arrangement	.12 (.11)	.04	1.10	.27	[-.09, .33]
Mental health outpatient ever	.42 (.14)	.10	2.93	.004	[.14, .70]
Substance abuse treatment ever	.92 (.18)	.18	5.26	<.001	[.58, 1.26]
Days of Follow-up	.01 (.001)	.34	9.77	<.001	[.004, .01]

Note. Unmatched variables between the 1st and 7th cohorts within state were tested as covariates with number of services. To take a liberal approach, each variable was tested in its own regression model with Days of Follow-up to account for variable follow-up periods across cases. Significant findings are presented in bold.

Supplementary Table SD: Moderated Hierarchical Regressions: Restricted Placements and Number of Services

Logistic Regression - Restrictive Placements					
	β (SE)	Exp(B)	<i>p</i> -value	95% CI	χ^2 (df)
Southern State					
Block 1					56.18(2)***
PSM weights	-.004 (.09)	1.00	.97	[.83, 1.20]	
Follow-up Days	.01 (.001)	1.01	<.001	[1.00, 1.01]	
Block 2 – $\Delta \chi^2$					5.08(2)
PSM Weights	-.005 (.10)	.99	.96	[.82, 1.20]	
Follow-up Days	.01 (.001)	1.01	<.001	[1.00, 1.01]	
SAVRY Risk Level	.32 (.14)	1.38	.03	[1.04, 1.83]	
Cohort	-.11 (.20)	.90	.61	[.60, 1.34]	
Block 3 – $\Delta \chi^2$					5.36(1)
PSM Weights	-.003 (.09)	1.00	.97	[.83, 1.20]	
Follow-up Days	.01 (.001)	1.01	<.001	[1.00, 1.01]	
SAVRY Risk Level	.66 (.21)	1.93	.002	[1.28, 2.91]	
Cohort	.49 (.33)	1.63	.14	[.86, 3.12]	
SAVRY Risk Level*Cohort	-.67 (.29)	.51	.02	[.29, .90]	
Northeastern State					
Block 1					152.68(4)***
PSM Weights	-.01 (.05)	.98	.76	[.89, 1.09]	
Any Violent Priors	.41 (.38)	1.51	.28	[.72, 3.15]	
# Prior Offenses	.45 (.11)	1.57	<.001	[1.26, 1.96]	
Follow-up Days	.01 (.001)	1.01	<.001	[1.01, 1.01]	
Block 2 – $\Delta \chi^2$					30.50(2)***
PSM Weights	-.04 (.06)	.96	.47	[.86, 1.07]	
Any Violent Priors	.25 (.40)	1.29	.52	[.59, 2.83]	
# Prior Offenses	.42 (.12)	1.52	<.001	[1.20, 1.92]	
Follow-up Days	.01 (.001)	1.01	<.001	[1.01, 1.01]	
YLS/CMI Risk Level	.95 (.18)	2.59	<.001	[1.83, 3.67]	
Cohort	.22 (.22)	1.24	.32	[.81, 1.91]	
Block 3 – $\Delta \chi^2$.13(1)
PSM Weights	-.04 (.06)	.96	.45	[.85, 1.07]	
Any Violent Priors	.25 (.40)	1.28	.53	[.59, 2.81]	
# Prior Offenses	.42 (.12)	1.52	<.001	[1.20, 1.20]	

Follow-up Days	.01 (.001)	1.01	<.001	[1.01, 1.01]
YLS/CMI Risk Level	.88 (.27)	2.41	<.001	[1.42, 4.06]
Cohort	.12 (.35)	1.12	.74	[-.56, 2.24]
YLS/CMI Risk Level*Cohort	.13 (.36)	1.14	.71	[-.57, 2.28]

Note. PSM = Propensity Score Match

Linear Regressions – Number of Service Referrals

	B (SE)	B 95% CI	β	<i>t</i>	<i>p</i> -value	R ²	Δ F(df)
Southern State							
Block 1						.02	
PSM Weights	.81 (.72)	[-.61, 2.24]	.05	1.12	.26		
Follow-up Days	-.02 (.01)	[-.03, -.01]	-.12	-2.79	.01		
Block 2 – Δ R ²						.06	10.48(2)***
PSM Weights	.69 (.71)	[-.71, 2.09]	-.04	.97	.33		
Follow-up Days	-.02 (.01)	[-.03, -.01]	-.13	-2.90	.01		
Cohort	6.65 (1.50)	[3.70, 9.60]	.19	4.43	<.001		
SAVRY Risk Level	-1.68 (1.06)	[-3.78, .41]	-.07	-1.58	.11		
Block 3 – Δ R ²						.07	4.20(1)*
PSM Weights	.68 (.71)	[-.71, 2.08]	-.04	-.96	.34		
Follow-up Days	-.02 (.01)	[-.03, -.01]	-.13	-3.06	.01		
Cohort	10.22 (2.30)	[5.71, 14.74]	.30	4.45	<.001		
SAVRY Risk Level	9.12 (5.37)	[-1.44, 19.68]	.38	1.70	.09		
SAVRY Risk Level*Cohort	-4.33 (2.11)	[-8.47, -.18]	-.48	-2.05	.04		
Northeastern State							
Block 1						.16	18.80(6)***
PSM Weights	-.04 (.03)	[-.10, .02]	-.05	-1.24	.22		
Any Violent Priors	.34 (.30)	[-.25, .93]	.05	1.13	.26		
# Prior Petitions	.03 (.09)	[-.15, .22]	.02	.36	.72		
Outpatient MH Tx Ever	.34 (.16)	[.03, .65]	.08	2.13	.03		
Outpatient SU Tx Ever	.81 (.19)	[.43, 1.18]	.16	4.227	<.001		
Follow-up Days	.01 (.001)	[.004, .01]	.32	8.30	<.001		
Block 2 – Δ R ²						.17	3.53(2)*
PSM Weights	-.04 (.03)	[-.11, .02]	-.05	-1.33	.18		
Any Violent Priors	.31 (.30)	[.28, .90]	.05	1.02	.31		

# Prior Petitions	.02 (.09)	[-.16,.20]	.01	.19	.85		
Outpatient MH Tx Ever	.25 (.16)	[-.07, .57]	.06	1.55	.12		
Outpatient SU Tx Ever	.70 (.19)	[.32, 1.09]	.14	3.61	<.001		
Follow-up Days	.01 (.001)	[.003, .01]	.29	7.15	<.001		
Cohort	.25 (.16)	[-.05, .56]	.06	1.62	.11		
YLS/CMI Risk Level	.29 (.13)	[.03, .54]	.09	2.21	.03		
Block 3 – ΔR^2						.17	.36(1)
PSM Weights	-.04 (.03)	[-.10, .02]	-.05	-1.31	.19		
Any Violent Priors	.31 (.30)	[-.28, .90]	.05	1.04	.30		
# Prior Petitions	.02 (.09)	[-.16, .20]	.01	.19	.85		
Outpatient MH Tx Ever	.25 (.16)	[-.07, .57]	.06	1.54	.12		
Outpatient SU Tx Ever	.71 (.19)	[.32, 1.09]	.14	3.63	<.001		
Follow-up Days	.01 (.001)	[.003, .01]	.29	7.10	<.001		
Cohort	.34 (.21)	[-.08, .76]	.08	1.59	.11		
YLS/CMI Risk Level	.65 (.62)	[-.56, 1.86]	.20	1.05	.29		
YLS/CMI Risk Level *Cohort	-.14 (.24)	[-.62, .33]	-.12	-.60	.55		

Note. PSM = Propensity Score Match, Outpatient MH Tx Ever = Any Outpatient Mental Health Treatment Ever (history or current), Outpatient SU Tx Ever = Any Outpatient Substance Use Treatment Ever (history or current).