Washington State Institute for Public Policy
110 Fifth Avenue SE, Suite 214 • PO Box 40999 • Olympia, WA 98504 • 360.664.9800 • www.wsipp.wa.gov

June 2019

Washington State's Aggression Replacement Training for Juvenile Court Youth: *Outcome Evaluation*

In 2018 the Washington State Juvenile Rehabilitation Administration (JRA),[1] in conjunction with the Community Juvenile Accountability Act (CJAA) Oversight Committee, contracted with the Washington State Institute for Public Policy (WSIPP). The contract required WSIPP to conduct an evaluation of the effect of the Washington State Aggression Replacement Training (WSART) program on recidivism outcomes for juvenile court youth.

WSART is a group-based intervention for moderate- and high-risk youth with criminal charges filed in juvenile courts. The program uses cognitive behavioral techniques to teach youth three core components: anger control, moral reasoning, and social skills. WSART was one of the initial programs funded by the Washington Legislature's initiatives to expand evidence-based programming for court-involved youth.[2]

Section I provides additional information about WSART and prior evaluations of its effects on recidivism. Sections II and III describe our evaluation methods, including a discussion of the data used in this report. Section IV presents the general findings of WSART on recidivism. Section V summarizes how the effects vary by the characteristics of program participants and Section VI summarizes how the effects vary by the characteristics of the WSART program. Sections VII and VIII discuss the limitations to this evaluation and the conclusions based on our findings.

Summary

In 2018, WSIPP was contracted to evaluate the effectiveness of the Washington State Aggression Replacement Training (WSART) program at reducing recidivism for court-involved youth. In addition to evaluating the overall effects, WSIPP was asked to evaluate for whom the program, and under what conditions, the program was most effective.

We evaluated the effects of WSART in Washington State courts from 2005 to 2016. We found that, on average, WSART participants were more likely to recidivate than similar youth who did not participate in WSART.

The differences in recidivism for WSART and non-WSART youth were evident in nearly all subpopulations including males, White youth, Black youth, Hispanic youth, younger youth, high-risk youth, moderate-risk youth, youth assessed using the Back On Track risk (BOT) assessment, and youth assessed using the Positive Achievement Change Tool (PACT) assessment. WSART participation reduced recidivism only for females.

WSART effectiveness did not vary based on the average competence determination for trainers in different juvenile courts. However, we found that youth who completed the entire WSART curriculum were significantly less likely to recidivate than youth who participated in but did not complete the WSART program.

Suggested citation: Knoth, L., Wanner, P., & He, L. (2019). *Washington State's Aggression Replacement Training for juvenile court youth: Outcome evaluation.* (Document Number 19-06-1201). Olympia: Washington State Institute for Public Policy.

[1] JRA is a part of the Washington State Department of Social and Health Services. On July 1, 2019, JRA will be relocated to the Department of Children, Youth, and Families.
[2] RCW 13.40.500 through 13.40.540.

I. Background

In 1997, the Washington State Legislature began to invest significantly in juvenile justice evidence-based programs (EBPs) by passing the Community Juvenile Accountability Act (CJAA).[3] The legislature directed WSIPP to identify a range of potential approaches that could cost-effectively reduce recidivism.[4]

Washington's juvenile courts initially implemented four programs in 1999: Functional Family Therapy (FFT), Aggression Replacement Training, Coordination of Services (COS), and Multisystemic Therapy (MST). WSIPP evaluated each of the four programs following implementation in Washington State and concluded that these programs produced cost-effective reductions in recidivism among court-involved youth. For WSART and FFT, statistically and substantively significant reductions in recidivism were found only when the programs were delivered competently.[5] WSIPP recommended continuous evaluations of programs to ensure that they continue to produce the intended effects over time.[6]

In 2018, the CJAA Oversight Committee contracted with WSIPP to "evaluate one CJAA program from the current menu to assess its impact on criminal recidivism." After consultation with WSIPP, CJAA selected WSART as the focus of the current evaluation.[7]

In 2004, WSIPP found that WSART participants were less likely to recidivate with a felony (21%) than WSART-eligible youth who did not participate in the program (25%), but these differences were not statistically significant. Further analyses found that there were significant reductions in recidivism for youth who participated in WSART programs rated highly competent.[8]

In 2017, the Washington State Center for Court Research (WSCCR) published an updated report suggesting that WSART may no longer have a significant effect on the recidivism behaviors of juvenile court youth.[9] For felony recidivism, WSCCR found that WSART participants had significantly *higher* recidivism rates (23%) compared to similar youth who did not participate in WSART (19%).

[3] RCW 13.40.500 – 540.

[4] Aos, S., Phipps, P., Barnoski, R., & Lieb, R. (2001) *The comparative costs and benefits of programs to reduce crime, version 4.0.* (Doc. No. 01-05-1201). Olympia: Washington State Institute for Public Policy.

[5] Barnoski, R. (2004). *Outcome evaluation of Washington State's research-based programs for juvenile offenders.* (Doc. No. 04-01-1201). Olympia: Washington State Institute for Public Policy and Mayfield, J. (2011*). Multisystemic Therapy outcomes in an evidence-based practice pilot.* (Doc. No. 11-04-3901). Olympia: Washington State Institute for Public Policy,

[6] Barnoski, R., Aos, S., & Lieb, R. (2003). *Recommended quality control standards: Washington State research-based juvenile offender programs* (Document No. 03-12-1203). Olympia: Washington State Institute for Public Policy.

[7] The other programs in the CJAA menu are Functional Family Therapy (FFT), Coordination of Services (COS), Education Employment Training (EET), Family Integrated Transitions (FIT), and Multi-Systemic Therapy (MST).

[8] In Barnoski (2004), the felony recidivism rates were 25% for the comparison group youth, 27% for youth who participated in non-competent WSART programs, 20% for youth who participated in competent WSART programs, and 13% for youth who participated in highly competent WSART programs.

[9] Peterson, A. (2017). *Aggression Replacement Training in a probation setting: Outcomes for youths starting treatment January 2010 – September 2012.* Olympia, WA: Center for Court Research, Administrative Office of the Courts.

WSIPP's original evaluation differed considerably from WSCCR's 2017 update. Specifically, WSIPP's initial evaluation of WSART was limited to youth participating in the program in 2000, while WSCCR's evaluation included youth participating in the program from 2010 to 2012. The WSIPP evaluation included youth from 26 juvenile courts while the WSCCR evaluation included youth from only 23 courts. Finally, the WSIPP evaluation had access to court-level competency measures[10] while the WSCCR report did not include competency measures at the court- or trainer-level.

The current study reexamines the effectiveness of WSART using a holistic approach that includes ten years of WSART participants and that also incorporates trainer adherence ratings.[11]

WSART

Washington State's Aggression Replacement Training is a cognitive behavioral program for moderate- and high-risk youth with a history of aggressive or anti-social behaviors. The program teaches three core components: anger control, moral reasoning, and social skills. The components are taught in a classroom setting by a certified WSART trainer.

WSART trainers work with the trainees to establish group norms; introduce new skills and to continuously review skills discussed in previous classes; facilitate skills practice activities; and provide positive feedback during practice sessions. Like other cognitive-behavioral models, ART trainers focus on teaching youth to acknowledge the limitations in their current thoughts, emotions, and behaviors. Trainers then offer new skills and coping mechanisms to promote future prosocial behavior for the youth.

Washington State
Aggression Replacement Training

Program Overview and Eligibility

Program Overview

WSART is a 10-week, 30-hour program administered to groups of 8–12 juvenile court youth. Each group meets three times per week for one-hour classes. In total, the program comprises 30 classes which are indicative of a full session. Each class is administered by a certified WSART trainer.

WSART uses repetitive learning techniques to teach three core components:

- Anger control
- Moral reasoning
- Social skills

Eligibility

Current eligibility for WSART is based on scores from the Positive Achievement Change Tool (PACT). Eligible youth must be classified as moderate- or high-risk and must have one of more of the following:

- A current or prior adjudication for a weapon, violent misdemeanor, or felony conviction;
- A score of at least 2 on Domain 11 (Aggression);*
- A score of at least 5 on Domain 10 (Attitudes/Behavior);* or
- A score of at least 4 on Domain 12 (Social Skills).*

*Relevant domain details are available in Appendix VII.

[10] Barnoski (2004) classified each court's WSART program as "competent" or "not competent" based on consultation with the WSART statewide coordinator.

[11] Formal trainer adherence ratings differ from the program level classifications used in prior evaluations (e.g., Barnoski (2004)). See Appendix I for more details.

Through role-playing, the anger control component teaches youth skills to address feelings of anger and provides them with alternatives to aggression in difficult situations. In the values-oriented moral reasoning component, youth collectively discuss responses to simulated moral dilemmas to identify the need for fairness, justice, and empathy for others. The social skills component uses demonstrations by the trainers and practice sessions to help youth develop prosocial interpersonal skills.

Juvenile court youth who are sentenced with local sanctions[12] are screened for WSART eligibility using the Washington State juvenile court risk assessment—the Positive Achievement Change Tool (PACT).[13] The placement of youth into eligible EBPs is at the discretion of the juvenile court judge and the juvenile probation counselor (JPC).[14]

Research Questions

In addition to evaluating the overall effectiveness of WSART for reducing recidivism among court-involved youth, CJAA asked WSIPP to identify whether WSART is more or less effective for particular populations and under particular conditions. Our evaluation includes three research questions.

1. Does WSART significantly reduce the likelihood of recidivism?

We test the effect of WSART on general recidivism in addition to the unique effects of WSART on certain types of recidivism (i.e., misdemeanor, felony, and violent felony recidivism).

2. For whom is WSART most effective?

We test whether the effects of WSART vary between high-risk youth and moderate-/low-risk youth, between males and females, and between youth from different racial groups.

3. Under what conditions is WSART most effective?

We test whether the effects of WSART depend on the competency of the trainer administering the session and between the youth who complete the full program and the youth who do not complete the full program.

[12] "Local sanctions" means one or more of the following: a) 0-30 days of confinement (detention); b) 0-12 months of community supervision; c) 0-150 hours of community restitution; or d) $0-$500 fine; RCW 13.40.020. Local sanctions are used with diversion, deferred, and conviction dispositions.

[13] Youth who are transferred to the custody of the DSHS to serve a period of confinement in a Juvenile Rehabilitation Facility are not evaluated using the PACT and would not be eligible for community programs such as WSART.

[14] See Appendix II for more information about the juvenile court processes for assigning youth to EBPs.

II. Data

This evaluation uses data from four different sources: 1) the Juvenile Assessment Research Database (ARD), 2) the WSIPP Criminal History Database (CHD), 3) the WSART Session Database, and 4) the WSART trainer adherence and competency evaluations. Exhibit 1 provides an overview of each data source.

We identified our sample of WSART-eligible youth using the ARD data. We linked assessments from the ARD to court cases in the CHD using youth identifiers, the date of the assessment, and the case adjudication date. When possible, we linked assessments for youth who participated in WSART to the records in the WSART session database. We then linked trainer adherence and competency records to youth based on court and year.

The different datasets used in this evaluation do not have a common identifier to link the court case to a specific risk assessment. Appendix I provides full details on the methods we used to compile our final dataset. We identified 43,625 assessments in the ARD for youth who were eligible for WSART from 2005–2015. We removed 13,865 WSART-eligible assessments (31.78%) from the sample because we could not identify a reliable match to a court case.[15] We removed an additional 4,741 assessments (10.87%) while coding the data for analysis.

[15] Youth should receive an assessment within 30 days of adjudication in a juvenile court. However, some courts complete youth assessments prior to adjudication and some youth may be assessed more than 30 days after their adjudication. We considered matches to be unreliable if the assessment start date was not reasonably close to an adjudication date (i.e., more than five months apart) since we could not be certain that the assessment was associated with a particular court case. Matching to the incorrect case could affect measures of recidivism (e.g., saying a youth did not recidivate when really they did) or selection into WSART (e.g., if we have the incorrect characteristics of the current offense).

Exhibit 1

WSART Evaluation Source Data Details

Data name	Data source	Information included	Sample coverage	Years available
Assessment Research Database (ARD)	WAJCA; AOC	Positive Achievement Change Tool (PACT) assessment information; referrals to and participation in evidence-based programs.	Juvenile court youth with a guilty, diversion, or deferred disposition who are sentenced to local sanctions.	2005 – 2018
Criminal History Database (CHD)	WSIPP	Combines court data from AOC, Incarceration and Community Supervision data from DOC, and Residential Confinement data from DSHS.	All court-involved adults and juveniles in Washington State.	1992 – 2018
WSART Session Database	WAJCA; AOC	WSART session and class information including attendance, class content, and venue/administrator data.	All WSART sessions administered for juvenile court youth.	2012* – 2018
WSART Trainer Adherence Data	Snohomish County	Annual reviews of WSART trainer competency and adherence to program curriculum.	WSART lead trainers who administer sessions for at least one year.	2010 – 2018

Notes:

*The WSART database was created in 2012 but implementation in individual courts was gradual. The last counties (Douglas, Cowlitz, and Lewis) joined in 2016.

WAJCA = Washington Association of Juvenile Court Administrators and AOC = Administrative Office of the Courts.

III. Evaluation Methodology

This evaluation compares the recidivism outcomes of youth who participate in WSART (treatment group) to recidivism outcomes of youth who do not participate in WSART (comparison group). Ideally, we would randomly assign eligible youth to participate in WSART or to receive no treatment. With random assignment, we would be confident that any differences in recidivism outcomes could be attributed to participation in the program rather than systematic differences in the characteristics of the youth. Initial research on WSART took advantage of waitlists for program participation to approximate randomization.[16]

Participation in WSART is no longer assigned based on a wait-list approach. The absence of random assignment introduces the possibility of "selection bias," or the possibility that youth who do participate in WSART are systematically different from youth who do not participate in WSART. For example, if youth with more complex needs are eligible for both FFT and WSART, juvenile probation counselors may be more likely to place youth in FFT instead of WSART because FFT provides more comprehensive intervention strategies than WSART.

If the systematic differences between the treatment and comparison groups are responsible for differences in recidivism outcomes, standard techniques of analysis may inappropriately conclude that participation in WSART increases or decreases the likelihood of recidivism.

Continuing the previous example of selection bias, if youth who have more complex needs and who are likely to be placed in FFT are also more likely to recidivate, participation in FFT may have a large effect on recidivism. If analyses find that youth who participated in FFT instead of WSART are less likely to recidivate than youth with less complex needs who were placed in WSART, it does not necessarily mean that WSART does not also reduce the likelihood of recidivism.[17] Failure to account for these selection biases may lead to an underestimate of the effect of WSART.

In the absence of random assignment, we use an advanced statistical technique, propensity score matching (PSM), to identify youth in the comparison group who are most similar to youth in the treatment group. PSM matches youth in each group only on observed characteristics (i.e., characteristics that are measured and included in our data). It is possible that differences in unobserved characteristics may still exist. However, statistical research confirms that PSM is capable of producing results that are comparable to the results found under conditions of random assignment.[18]

[16] Barnoski (2004). Under a wait-list approach, eligible youth are assigned to the program until the program is full. Remaining eligible youth are assigned to the control group and do not participate in the program. This approach approximates randomization since participation in treatment (in this case, WSART) is essentially random and not based on characteristics of the youth (e.g., age, ethnicity, sex).

[17] Standard regression techniques would attempt to capture this bias by controlling for participation in different programs.

[18] Campbell, C. & Labrecque, R.M. (2018). *Panacea or poison: Can propensity score modeling (PSM) methods replicate the results from randomized control trials (RCTs)?* Final Summary Report, Portland, Oregon: Criminal Justice Policy Research Institute.

Treatment and Comparison Groups

The treatment group includes youth who were assigned to and started WSART between January 1, 2006, and June 30, 2016. We include youth who started WSART regardless of whether or not they completed the program.[19]

The comparison group includes youth who were eligible to participate in WSART (see Appendix VII) but who did not start WSART between January 1, 2006, and June 30, 2016.[20] Youth in the comparison group may have started a non-WSART EBP or may not have started any EBP. Consequently, our study includes a "treatment as usual" comparison group.[21]

Youth could have multiple juvenile court cases and, consequently, multiple PACT assessments during the study time frame. For youth who had multiple WSART-eligible assessments, we randomly selected one of the assessments to be included in the analyses. If the randomly chosen assessment included participation in WSART, we included the youth in the treatment group.[22] If the randomly chosen assessment did not include participation in WSART, we included the youth in the comparison group.

Methods

Our final population of WSART-eligible youth includes 7,561 treatment youth and 16,354 comparison youth.[23] We use propensity score matching (PSM) to identify comparison youth who are most similar to treatment youth on observable characteristics. The goal is to establish a final matched sample that is similar on all characteristics except for participation in WSART.

PSM includes three steps. First, we estimate the probability that a youth participates in WSART, referred to as the "propensity score." Second, we match youth from the treatment group to youth in the comparison group who have a similar propensity score, creating a final matched sample for analyses. Third, we assess the effect of WSART participation on recidivism by conducting regression analyses using the matched sample.

[19] We estimate the treatment effect on the treated in order to avoid selection bias that may be related to completion. For example, youth who complete WSART may be more motivated to address their problematic behaviors and less likely to recidivate. We also excluded seven youth for whom the ARD data indicated they started WSART but there was no WSART start date recorded.

[20] We exclude youth who were eligible for WSART but who started WSART after June 30, 2016. We also excluded youth from the comparison group if they were deceased, if they were incarcerated or committed to JRA, if they did not start WSART because their whereabouts were unknown, if they were on warrant status, or if they moved out of state. See Appendix II for more details.

[21] This approach differs from previous evaluations of juvenile court programs conducted by WSIPP. In today's current juvenile court environment, treatment as usual is likely to involve some EBP or local treatment services. In Appendix V, we test the robustness of our findings when using a "no treatment" comparison group; the results were consistent with the results presented in this report.

[22] See Appendix II for additional details. We checked our findings using each youth's first assessment rather than a random assessment and the results were nearly identical.

[23] See Appendix I and Appendix II for details about how we arrived at these final sample sizes.

Calculating the Propensity Score

Propensity score methods are used to minimize selection bias related to observed characteristics. We calculated the propensity score using logistic regression to predict the likelihood of participating in WSART. We included variables for all available characteristics that may be related to participation in WSART. Our final model includes characteristics of the youth, characteristics of the index offense, overall risk and protective scores from the domains from the PACT risk assessment, and individual factor scores from PACT risk assessment items related to EBP eligibility.[24]

Constructing the Matched Sample

We match treatment youth to the comparison youth with the most similar propensity score. Youth in the comparison group were matched to only one youth in the treatment group. We removed youth from the treatment group if we could not identify a good match with youth in the comparison group. Our final matched sample included 6,535 treatment youth and 6,535 comparison youth.[25]

Exhibits 2 shows the means and percentages for characteristics included in the propensity score model for the treatment and comparison groups after matching for demographic, index offense characteristics, and overall risk information. Additional details on the covariate balance before and after matching are available in Appendix II.

[24] See Appendix II for more details.

[25] After matching, we removed seven youth from the treatment group who had a propensity score that was greater than the largest propensity score for all youth in the comparison group (also known as being outside the range of common support). We removed an additional 1,101 youth from the treatment group because there was not a youth in the comparison group with a propensity score within the optimal caliper when matching without replacement. See Appendix II for more information.

Exhibit 2

Covariate Balance: Demographic and Index Case Characteristics

Variable	WSART Mean	Non-WSART Mean	Bias
Demographic and current offense characteristics			
Age at assessment			
12 and under	1.9%	1.9%	-0.360
13	5.9%	6.1%	-0.758
14	12.1%	12.4%	-1.248
15	21.7%	22.2%	-1.285
16	28.5%	28.1%	0.769
7	26.8%	26.3%	1.095
18 and older	3.1%	2.9%	1.202
Race/ethnicity			
White	63.2%	63.0%	0.352
Black/African American	12.8%	12.8%	-0.091
Asian/Pacific Islander/Native Hawaiian	2.8%	2.8%	0.186
American Indian/Alaskan Native	3.7%	3.7%	-0.323
Hispanic	15.9%	16.1%	-0.387
Other race/unknown	1.6%	1.6%	0.216
Sex			
Male	74.0%	74.3%	-1.117
Female	26.0%	25.7%	1.020
Most serious offense grade			
Misdemeanor	66.9%	67.5%	-1.259
Felony	20.5%	20.3%	0.660
Violent felony	12.6%	12.2%	0.986
Most serious offense type			
Person	32.0%	32.9%	-1.924
Property	40.6%	40.3%	0.635
Sex	1.9%	1.7%	1.165
Drug	6.8%	6.6%	0.867
Other offense	18.7%	18.5%	0.426
Risk level			
High-risk	53.2%	53.8%	-1.185
Moderate-/low-risk	46.8%	46.2%	0.656
Total number of eligible EBPs	1.865	1.879	-1.805
Assessment version			
PACT	47.5%	48.0%	-0.980
BOT	52.5%	52.0%	1.265

Note:
Bias measures represent the standardized mean difference * 100. Values greater than 25 indicate severe imbalance. Values greater than 10 indicate moderate imbalance.

IV. Findings

We analyze the effect of WSART participation on four outcomes:

- Any recidivism,
- Misdemeanor recidivism,
- Felony recidivism, and
- Violent felony recidivism.

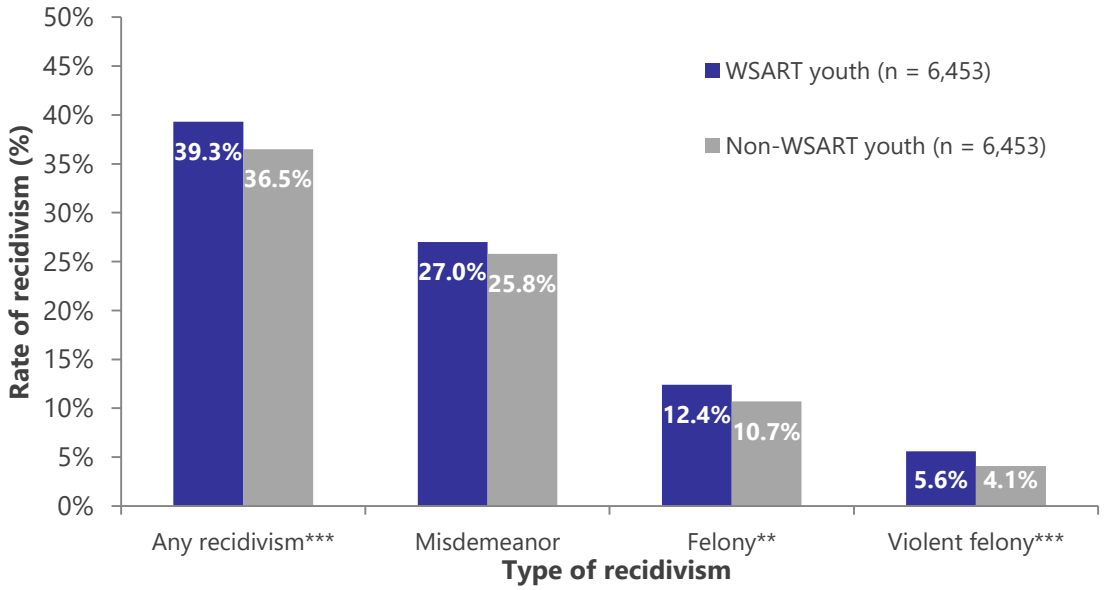
For misdemeanor, felony, and violent felony recidivism, we classified youth based on the most serious offense in the youth's first recidivism event.[26]

Felony recidivism includes all felony offenses, and violent felony recidivism is a subset of general felony recidivism.

The results are presented in Exhibit 3.[27] Overall, youth who participated in WSART were significantly more likely to recidivate than similar youth who did not participate in WSART.[28] WSART participants were significantly more likely than non-participants to recidivate with a felony offense or a violent felony offense. There were no significant differences in misdemeanor recidivism.

Exhibit 3

18-Month Adjusted Recidivism Rates, by Treatment Status



Note:
* p < 0.05, ** p < 0.01, *** p < 0.001

[26] See Appendix III for more details.

[27] We conducted logistic regression analyses on the matched samples and included all of the covariates used in the original propensity score models to adjust for any residual imbalance between the matched groups. Exhibit 3 shows the regression-adjusted probabilities of recidivism.

[28] Statisticians often rely on a metric, the p-value, to determine whether an effect is significant. The p-value is a measure of the likelihood that the difference could occur by chance—values range from 0 (highly significant) to 1 (no significant difference). By convention, p-values less than 0.05 (a 5% likelihood that the difference could occur by chance) are considered statistically significant.

V. Subgroup Analyses

We replicated our analyses with various subpopulations to identify whether the effects of WSART differed for particular groups of court-involved youth. Our subgroup analyses included comparisons of the effect of WSART on recidivism for males and females; for White, Black, and Hispanic youth; for older and younger youth; and for high-risk and moderate- or low-risk youth.

The characteristics that determine selection into WSART may differ for demographic subgroups. To account for these differences, we calculated new propensity scores and conducted individual evaluations of the effects of WSART on recidivism for each subpopulation. For each subgroup, we developed models to identify matched samples of treatment and comparison group youth with a similar likelihood of participating in WSART.

Exhibits 5–8 present the regression-adjusted probabilities of recidivism by subgroup. Separate findings are presented for overall recidivism, misdemeanor recidivism, felony recidivism, and violent felony recidivism.

Exhibit 4
Subgroup Findings

Sex

Females
WSART participants were significantly less likely to recidivate than similar non-WSART youth.

Males
WSART participants were significantly more likely to recidivate than similar non-WSART youth. WSART youth were significantly more likely to recidivate with a felony or a violent felony than non-WSART youth.

Race/ethnicity

White
WSART participation had no overall effect on recidivism for White youth. However, White youth who participated in WSART were more likely to recidivate with a violent felony than similar non-WSART youth.

Black
There were no significant differences in recidivism for WSART participants and similar non-WSART youth.

Hispanic
WSART participants were more likely to recidivate than similar non-WSART youth, but the differences were not significant for any specific type of recidivism.

Age

Youth ages 15 and under
WSART participants were significantly more likely to recidivate than similar non-WSART youth. Differences were significant for felony and violent felony recidivism.

Youth ages 16 and older
There was no significant difference in overall recidivism. WSART participants were more likely to recidivate with a violent felony offense.

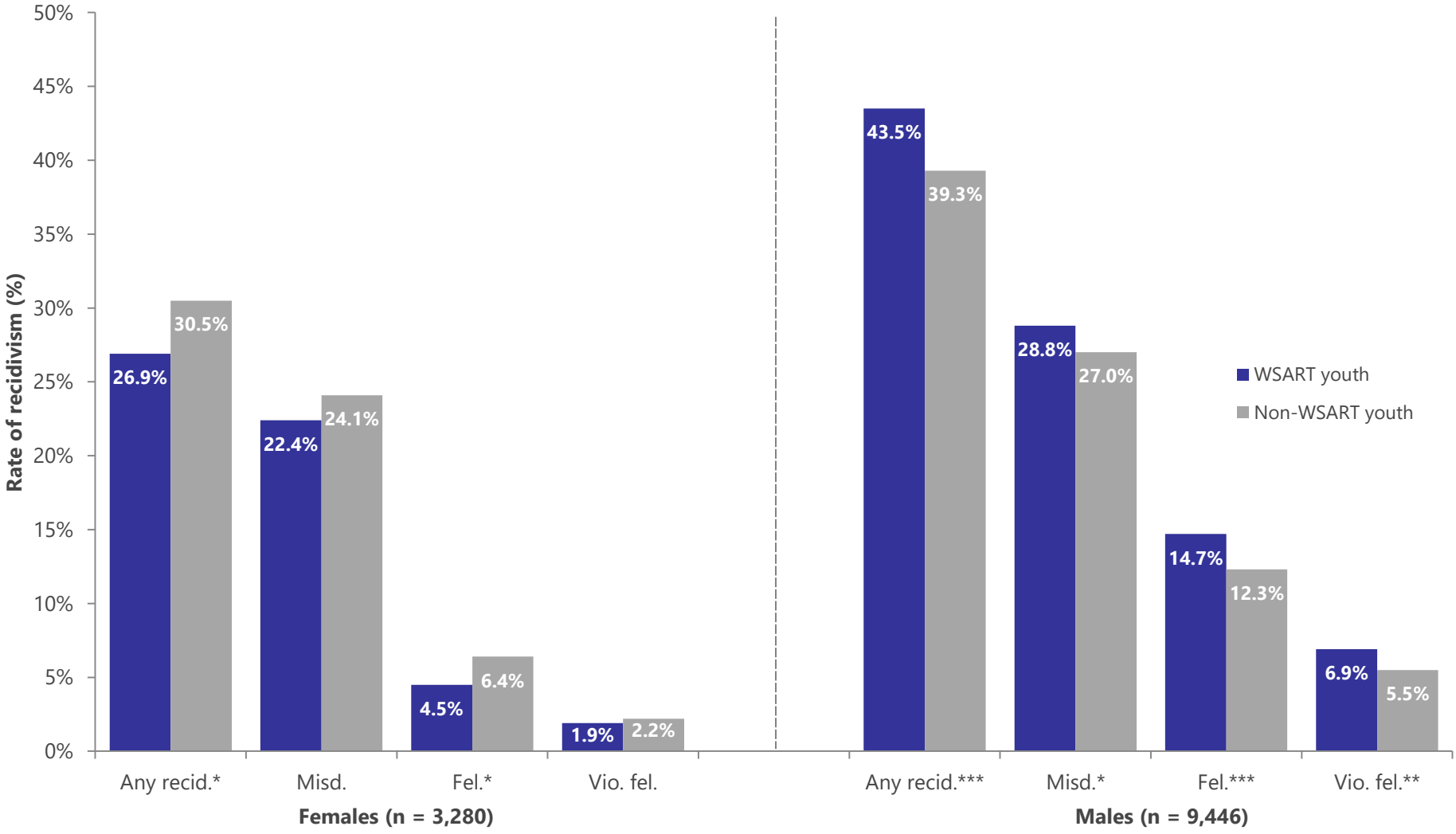
Risk level

High-risk
WSART participants were significantly more likely to recidivate than similar non-WSART youth. Differences were significant for felony and violent felony recidivism.

Moderate-/low-risk
There was no significant difference in overall recidivism. WSART participants were more likely to recidivate with a felony or violent felony offense.

Exhibit 5

18-Month Regression-Adjusted Recidivism Rates, by Sex and Treatment Status



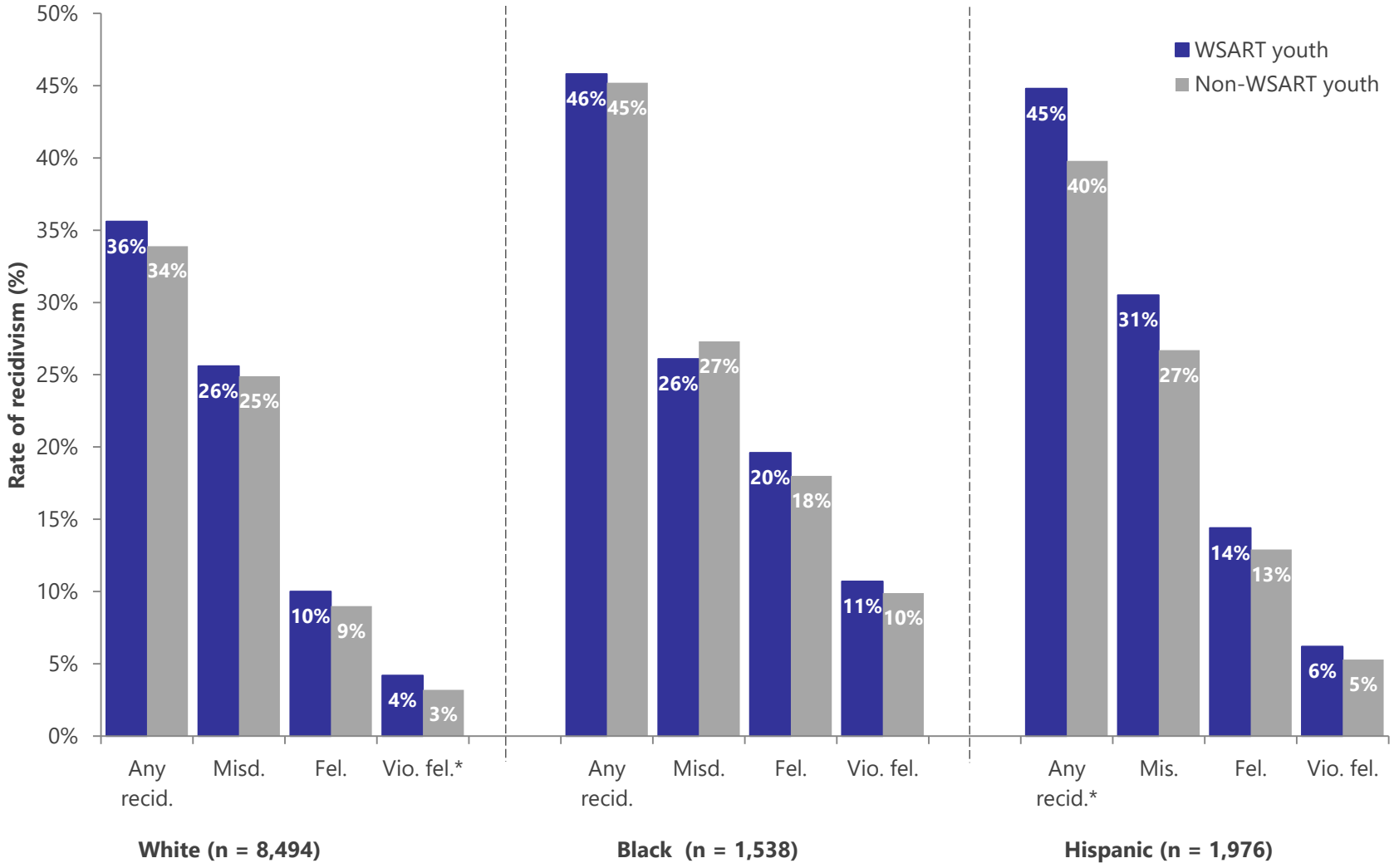
Note:

* p < 0.05, ** p < 0.01, *** p < 0.001

Each sex category represents both an independent sample and an independent analysis.

Exhibit 6

18-Month Regression-Adjusted Recidivism Rates, by Race/Ethnicity and Treatment Status



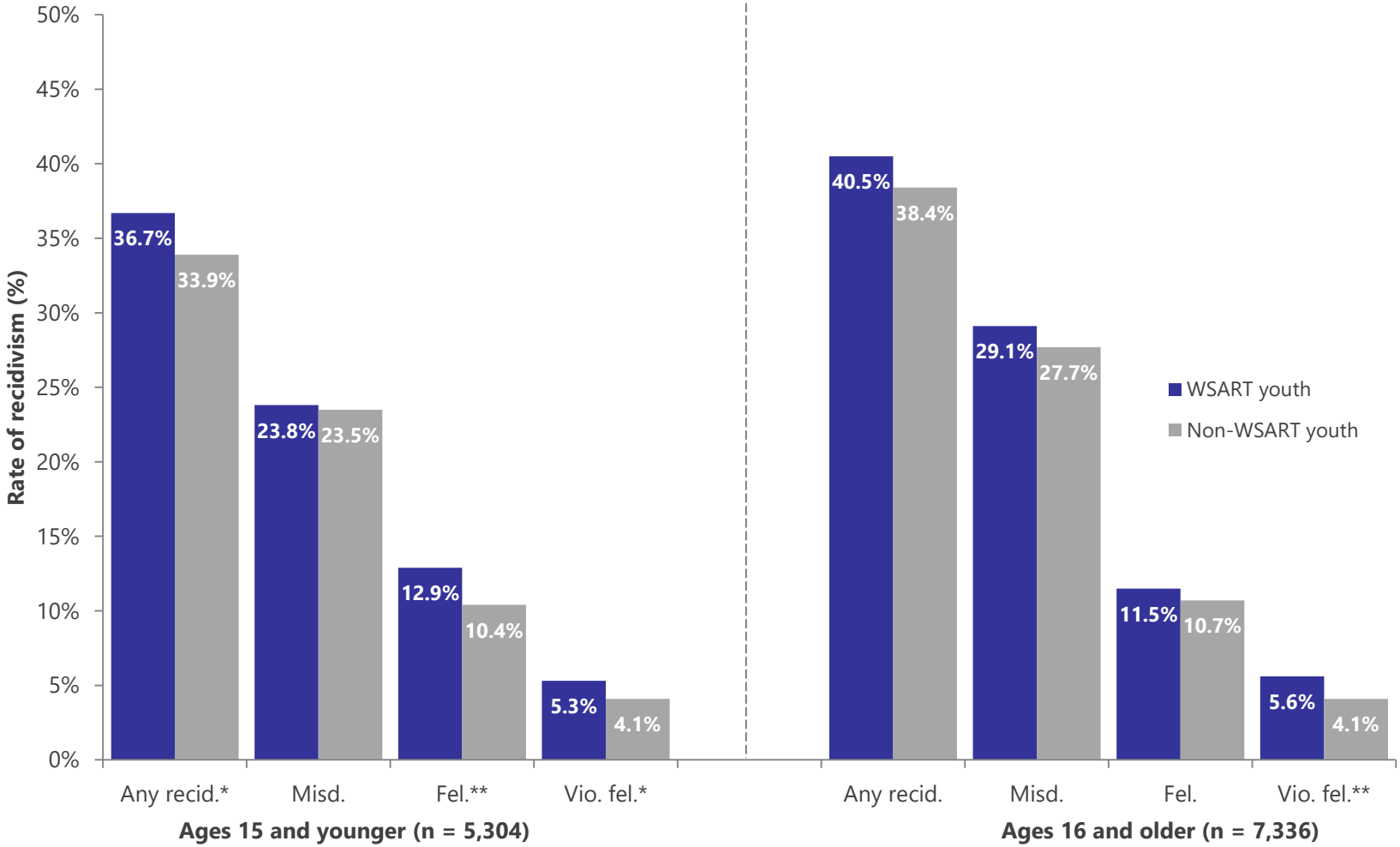
Notes:

* p < 0.05, ** p < 0.01, *** p < 0.001

Each racial category represents both an independent sample and an independent analysis.

Exhibit 7

18-Month Regression-Adjusted Recidivism Rates, by Age and Treatment Status



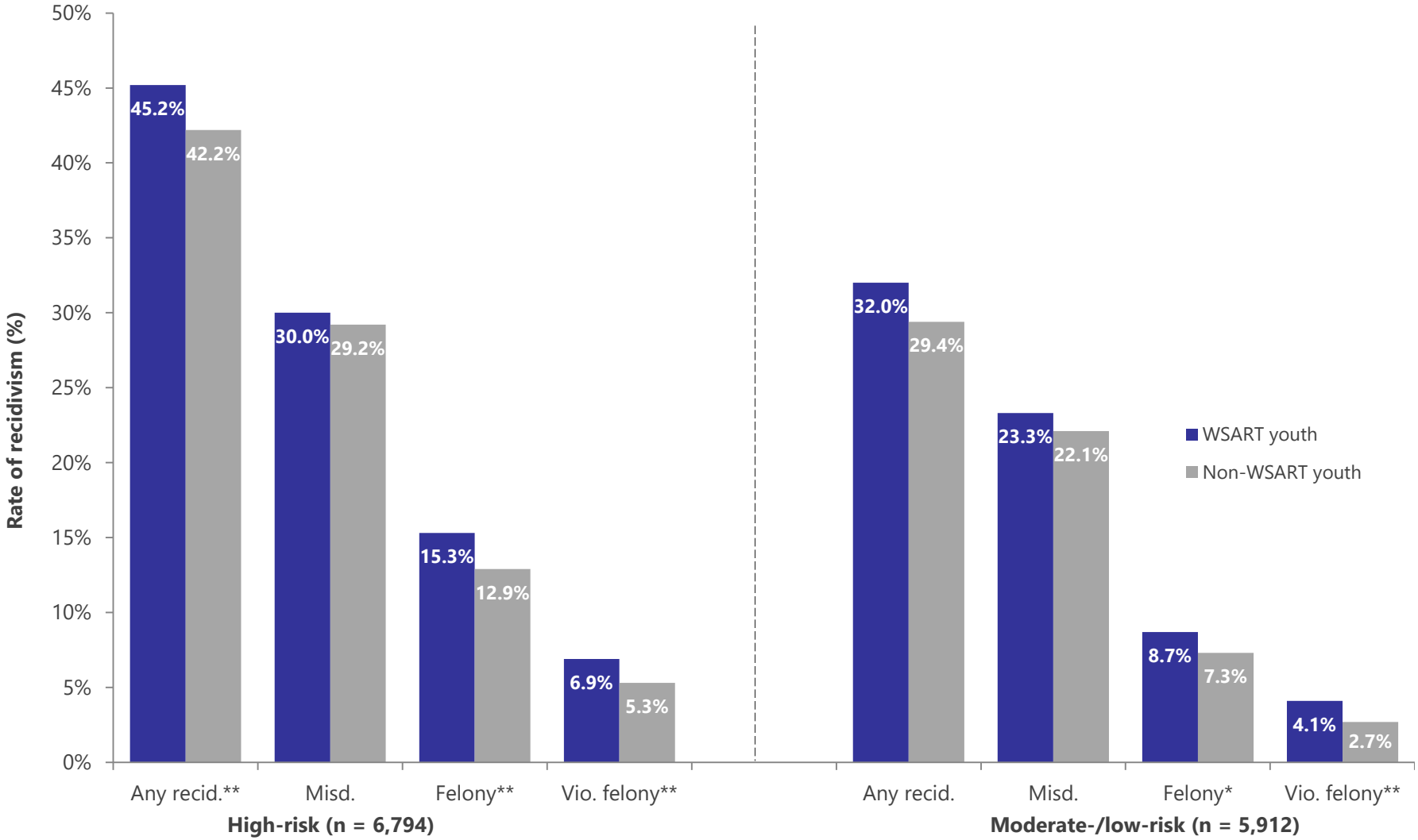
Notes:

* p < 0.05, ** p < 0.01, *** p < 0.001

Each age category represents both an independent sample and an independent analysis.

Exhibit 8

18-Month Regression-Adjusted Recidivism Rates, by Risk Level and Treatment Status



Notes:

* p < 0.05, ** p < 0.01, *** p < 0.001

Each risk category represents both an independent sample and an independent analysis.

VI. WSART Characteristics

The effects of WSART on recidivism may vary by characteristics of the session, selection into the program, or exposure to the program's components. We examined whether the effects of WSART varied by 1) the competency of the trainer, 2) the risk assessment used to screen for eligibility, and 3) the youth's exposure to the entirety of the program curriculum.

Trainer Competence

Previous analyses of EBPs show that trainer competence may moderate the effectiveness of juvenile court programs. We tested whether or not the effects of WSART varied based on competency of the trainers administering WSART in a particular court.

In WSIPP's original evaluation of WSART,[29] measures of individual trainer competence were not available. Rather, WSIPP relied on broad, court-level classifications of program competence based on determinations made by the statewide WSART coordinator. Since 2011, the statewide WSART coordinator has collected evaluations of individual trainer competence, allowing us to examine a data-driven approach to measuring the importance of competent trainers.

WSART trainers are overseen by WSART statewide consultants and are evaluated on an annual basis. Trainers are assessed based on their adherence to the curriculum and to the manualized methods for each of the three core components of WSART (i.e., anger control, social skills, and moral reasoning). Adherence for each core component includes up to 26 measurements. Some examples of adherence measures include, "Were achievements rewarded?" and "Did each youth express how the social skill could be personally useful?"

In addition to adherence to the curriculum and manualized methods, trainers are assessed on their overall competence. For each of the three core components, trainers are rated from not competent to highly competent for six characteristics:

1) Adherence;
2) Clarity (understandable by youth);
3) Simplicity (without unnecessary complication or explanation);
4) Objectivity (presented as factual content without moralizing or debate);
5) Pacing (not too fast, not too slow); and
6) Engagement (interesting and involving).

The competence scores sum to an overall score between 0 and 54.

[29] Barnoski (2004).

Trainers may adhere to the curriculum and manualized requirements but still present information to the youth in an incompetent manner. As such, the adherence scores are not directly related to the overall competence classification.

For this evaluation, we relied on overall trainer competence rather than adherence scores. We received trainer competence information for 25 courts from 2011 to 2016. We were unable to directly match trainers to youth in the ARD database. Consequently, we calculated court-level competence by taking the average competence score for all trainers in a court in a given year.[30] Using a subsample of WSART-eligible youth, we repeated our process of calculating propensity scores and identifying the probability of recidivism on a balanced, matched sample. The results from these analyses are presented in Exhibit 9.

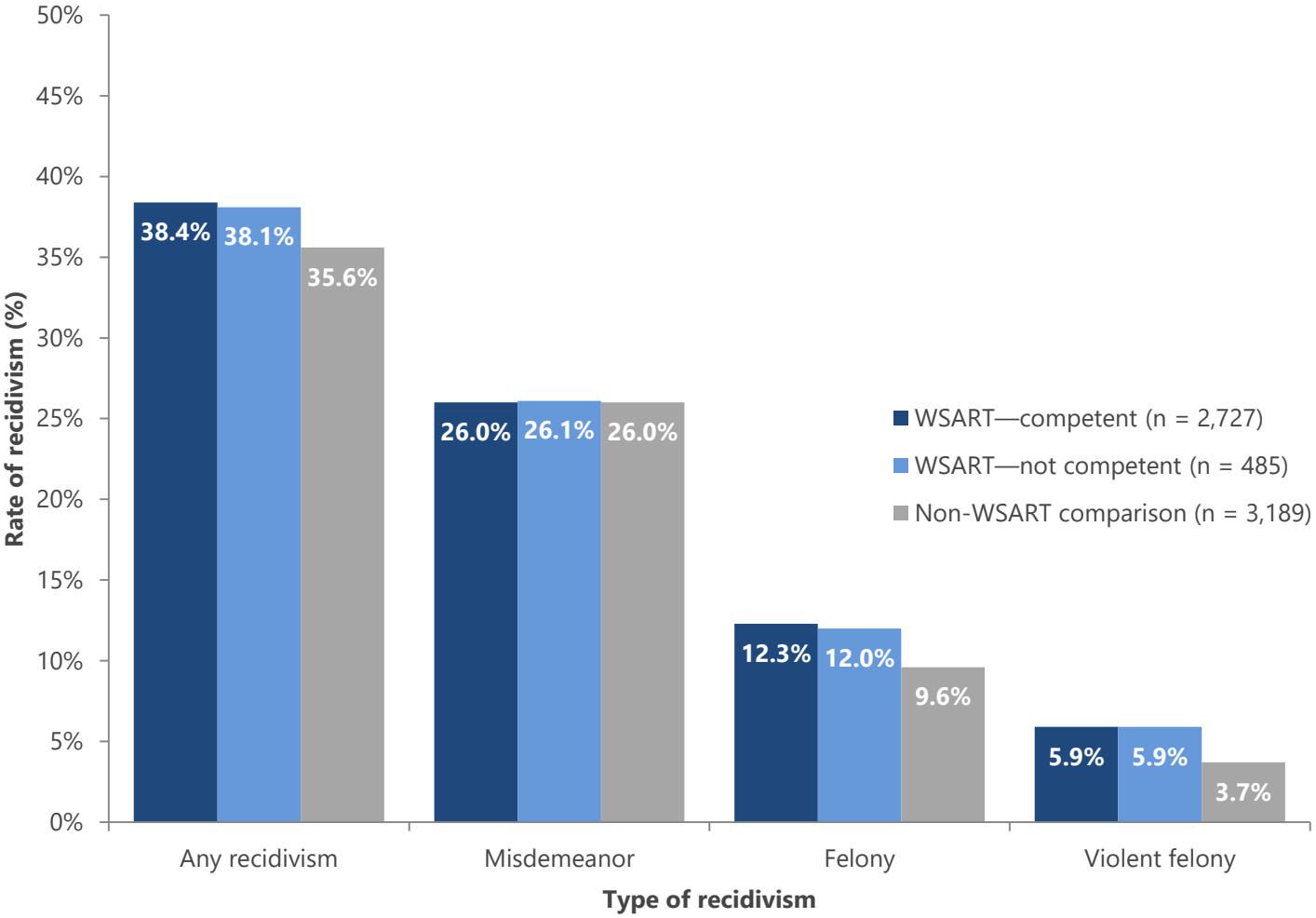
The probability of recidivism was almost identical for youth who participated in WSART in courts with competent trainers and youth who participated in WSART in courts with trainers who were not competent.

We were unable to assess the direct relationship between specific trainer competence classifications and youth in different WSART classes. Appendix I provides an additional discussion regarding our inability to draw reliable conclusions about trainer competence and recidivism given the currently available data.

[30] See Appendix I for additional details. Our measure of court-level competence differs from Barnoski (2004) because we directly calculated the measure using trainer evaluations rather than relying on the judgment of the statewide WSART coordinator.

Exhibit 9

18-Month Regression-Adjusted Recidivism Rates, by Court-Level Trainer Competence



Note:
* p < 0.05, ** p < 0.01, *** p < 0.001

Assessing Eligibility

Early evaluations of WSART found a significant reduction in recidivism for WSART participants. However, more recent evaluations, including this report, found that WSART does not reduce recidivism and may actually increase recidivism.

It is possible that the effects of WSART have changed over time. It is also possible that other changes in Washington State's juvenile courts could account for the change in effectiveness over time. Of particular importance, Washington State switched from the Back on Track (BOT) risk assessment system to the PACT risk assessment system in 2011. While the BOT and the PACT are largely the same, the PACT did include several new questions, new weighting mechanisms for risk factors, and new eligibility criteria for juvenile court EBPs. Since the risk assessment instrument is directly tied to eligibility for and assignment to EBPs, it is possible that the change in assessments affected the populations receiving WSART and the overall outcomes for WSART participation.

To test whether the effects of WSART changed over time, we split our sample based on the version of the risk assessment that the youth received. Functionally this split our sample into youth who were court-involved before 2011 and youth who were court-involved during or after 2011 (see Exhibit 10).

Exhibit 10

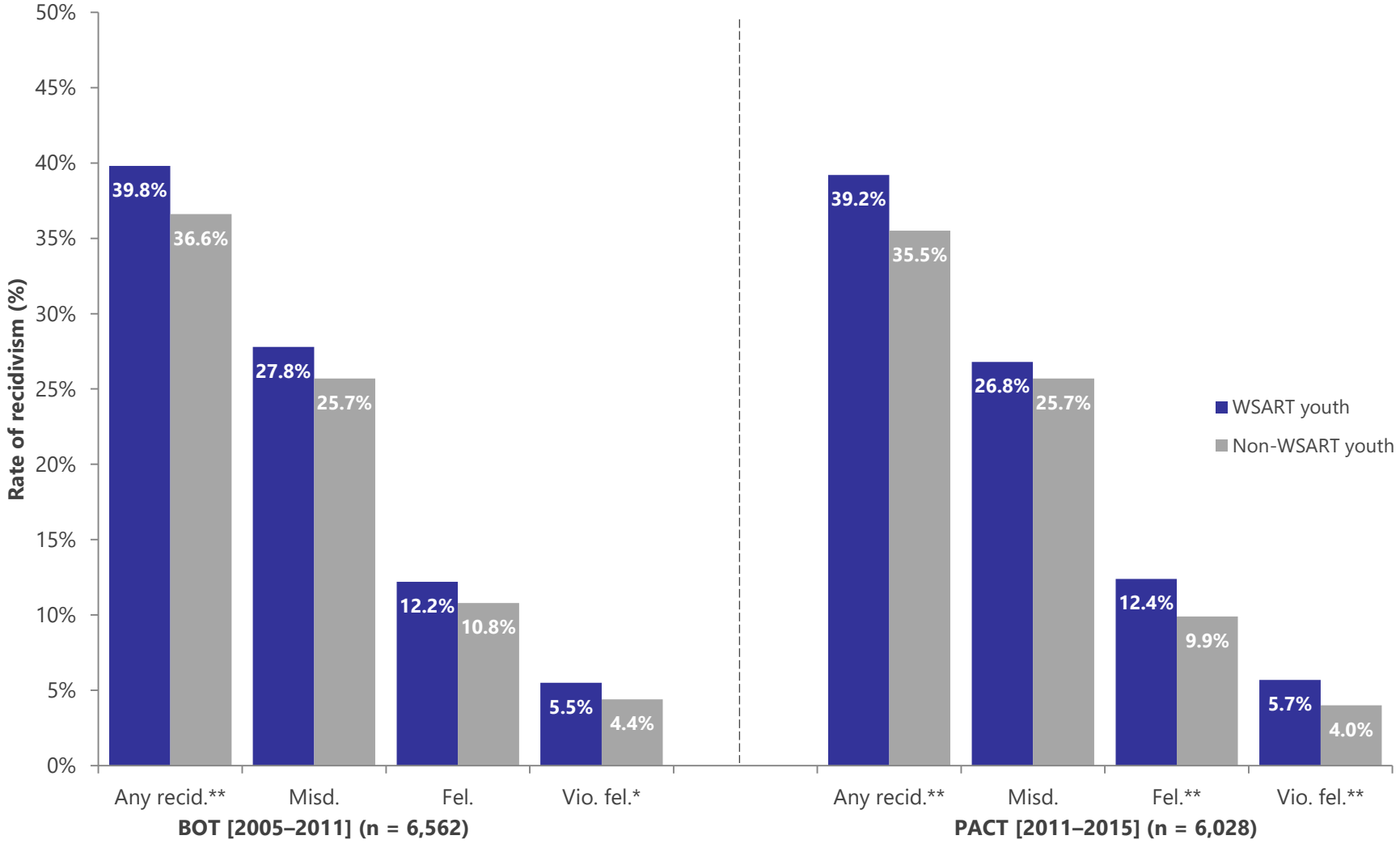
Year of Assessment, by Assessment Type for Full Sample (n = 23,915)

Year	BOT	PACT
2005	1,439	0
2006	2,572	0
2007	2,671	0
2008	2,605	0
2009	2,445	0
2010	2,229	0
2011	132	2,179
2012	5	2,035
2013	1	1,925
2014	0	1,790
2015	0	1,887

Exhibit 11 presents the effects of WSART on recidivism for youth assessed under the BOT assessment system and the effects for youth assessed under the PACT assessment system. Overall, youth who participated in WSART under the BOT system were marginally more likely to recidivate than similar non-WSART youth, but there were no significant differences in any specific type of recidivism. In contrast, youth who participated in WSART under the PACT system were significantly more likely to recidivate than similar non-WSART youth. Additionally, WSART youth under the PACT system had significantly higher rates of both felony and violent felony recidivism compared to similar non-WSART youth.

Exhibit 11

18-Month Adjusted Recidivism Rates, by Assessment Version and Treatment Status



Notes:

* p < 0.05, ** p < 0.01, *** p < 0.001

Each assessment category represents both an independent sample and an independent analysis.

Program Completion

Program effectiveness may depend on whether or not the youth receives the entirety of the treatment. Youth who fail to attend classes or who are removed from the program due to behavioral issues may not be exposed to enough of the WSART curriculum to internalize the skills necessary to desist from future offending.

Exhibit 12 shows the unadjusted recidivism rates for the full sample of WSART-eligible youth. Overall, youth who started WSART but did not complete WSART had a higher rate of recidivism than youth who completed WSART and youth who did not participate in WSART. Youth who completed the entire WSART program had a lower rate of recidivism than youth who did not participate in WSART.

Exhibit 12

Unadjusted Recidivism, by WSART Participation and Completion (n = 23,915)

	% Recidivism
WSART completer	33.89%
WSART non-completer	49.80%
Non-WSART youth	40.46%

Analyses of program completion must account for additional selection bias. Specifically, it is possible that youth who complete the program are more motivated to change and are less likely to recidivate than youth who are less motivated to change. To account for this bias, we limit our sample to youth who participated in WSART and matched youth who failed to complete the program to similar youth who successfully completed the program.[31]

Of the 7,543 youth who started WSART, 5,565 (73.78%) completed the full program.

In addition to selection bias, these analyses must account for the possibility of reverse causality. For this analysis, we are interested in the effects of program completion on recidivism. It is possible that some youth failed to complete the program because they committed a new offense and were forced to leave the program because they were confined to a detention facility or committed to a Juvenile Rehabilitation facility. We limited our analyses to youth who dropped out of the program, who did not meet minimum attendance requirements, or who were removed from the program by the trainer (n = 1,862).

Due to the disproportionate number of youth who completed WSART (n = 5,565), we used nearest neighbor matching with replacement to identify a similar non-completer for each completer.[32] Without replacement, our analyses would be limited to a maximum of 1,862 of the youth who completed the WSART program.[33]

The findings for program completion are presented in Exhibit 13. Youth who successfully completed the WSART program were significantly less likely to recidivate than similar youth who did not complete the program. Youth who successfully completed WSART had recidivism rates exactly eight percentage points lower than those who did not complete (34.2% for successful completers and 42.2% for non-completers). The differences were significant for misdemeanor, felony, and violent felony recidivism.

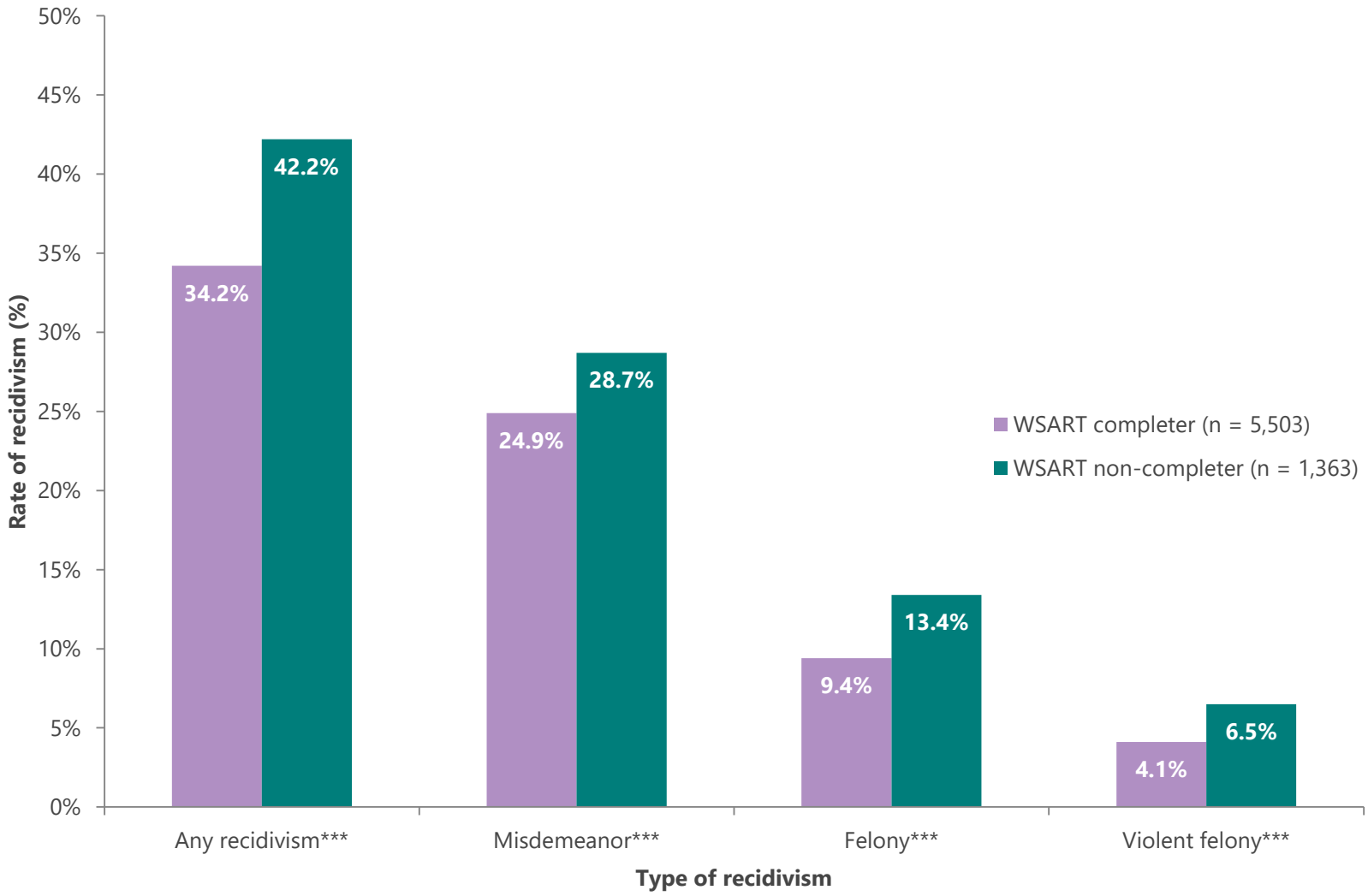
[31] See Appendix VI for additional tests.

[32] See Appendix II for additional details.

[33] See Appendix VI for additional details.

Exhibit 13

18-Month Regression-Adjusted Recidivism Rates, by Completion Status



Note:

* p < 0.05, ** p < 0.01, *** p < 0.001

VII. Limitations

Our ability to evaluate the effectiveness of WSART was limited by the availability of reliable data and the characteristics of the youth included in our sample. In addition, we made several decisions in this evaluation that affect the ability to compare these findings to previous evaluations. This section provides a brief discussion of these limitations and their impact on our overall conclusions.

Availability of Reliable Data

Our analyses were limited to the 67.8% of WSART-eligible youth who we were able to reliably match to a court case in the Criminal History Database. To date, there is no unique identifier in the assessment and EBP data that can be used to identify the corresponding court case in AOC's court records. Our review of the data found that characteristics of the current offense (e.g., seriousness of the offense and/or type of offense) were significantly related to selection to WSART. Consequently, being able to include information about the current offense is necessary to eliminate important selection bias from the analyses of the effects on recidivism.

Ideally, we would be able to evaluate the session characteristics and trainer competence for the specific session that each youth participated in. However, given the gradual implementation of the session database and the prevalence of missing trainer evaluations, we were unable to reliably evaluate the effects of session or trainer characteristics. We constructed data-driven court-level competence measures, which is an improvement from the strictly qualitative rankings used in early evaluations. However, we were unable to validate whether we had a complete trainer evaluation for all trainers who administered a WSART session in each year. Thus, our rankings may not accurately reflect the competence even at the court level.

In addition to CJAA-funded EBPs, juvenile court youth may participate in treatment programs funded and administered by the local juvenile court or other forms of treatment such as in-patient drug treatment. We did not have access to data capturing participation in non-CJAA funded EBPs. It is possible that WSART-eligible youth (both in the treatment or comparison groups) participated in alternative forms of treatment that affected the overall recidivism outcomes.

Characteristics of Youth in Sample

Propensity score matching seeks to match treatment youth (i.e., WSART participants) with highly similar comparison youth (i.e., non-WSART participants). However, if youth included in the treatment group are highly dissimilar from youth in the comparison group, it may not be possible to create matched pairs for all treatment youth.

For our main analyses and all sensitivity analyses, there were treatment youth with propensity scores that were significantly higher than any comparison group youth. As such, we had to remove some of the WSART youth who were most likely to be assigned to participate in WSART.

In our analysis of the overall effects on recidivism, we dropped 1,108 youth who participated in WSART (14.7% of youth who started WSART). Exhibit 14 shows the unadjusted rate of recidivism for treatment youth who were included and treatment youth who were excluded in our main analyses.

WSART youth for whom we were unable to find a comparable non-WSART youth were less likely to recidivate than youth who were matched to non-WSART youth and who were included in our main analyses. However, we used alternative methods to evaluate the effect of WSART on the full sample and the findings were consistent with those presented in the main report (see Appendix IV).

Exhibit 14

Unadjusted Recidivism Rate for WSART Participants Included and Excluded in Main Analyses
(n = 7,561)

	% Recidivism
Included in Sample (n = 6,453)	39.33%
Excluded from Sample (n =1,108)	30.96%

Comparability to Prior Evaluations

Our evaluation is not directly comparable to the prior evaluations of WSART[34] for several reasons. First, we use a treatment as usual approach for selection of our comparison group, while prior evaluations used a no-treatment comparison group. If other forms of treatment also work to reduce recidivism, it is possible that our comparison group would have lower rates of recidivism than comparison groups in previous evaluations. We tested alternative approaches to specifying our comparison group to include only those youth who did not participate in any CJAA-funded EBP and the results were consistent with the findings in the main report (see Appendix V).

For this evaluation, we measured type of recidivism based on the most serious offense in the first recidivism event during the follow-up period. Previous evaluations measured recidivism as the most serious offense type from all recidivism events during the follow-up period. It is possible that the lower rates of felony and violent felony recidivism in our evaluation are due to the differences in measurement. We discuss these differences in depth in Appendix III.

[34] Barnoski (2004) and Peterson (2017).

VIII. Conclusion

Combined, the analyses in this report provide in-depth information regarding the effectiveness of WSART in Washington State juvenile courts. Overall, WSART does not appear to be an effective program for reducing the likelihood of recidivism among court-involved youth. This section summarizes our findings in the context of our research questions.

1) Does WSART significantly reduce the likelihood of recidivism?

For overall recidivism, WSART participants were more likely to recidivate than similar youth who were eligible for, but who did not participate in, WSART (39.3% vs. 36.5%, respectively). WSART participants were significantly more likely to recidivate with a felony or violent felony offense, but there were no significant differences for misdemeanor recidivism.

2) For whom is WSART most effective?

The effectiveness of WSART varied by sex, race/ethnicity, and risk level. Females who participated in WSART were significantly less likely to recidivate than similar females who did not participate in WSART. However, males who participated in WSART were significantly more likely to recidivate than males who did not participate in WSART.

There were few significant differences between WSART participants and non-WSART participants by race/ethnicity. White youth who participated in WSART were significantly more likely to recidivate with a violent felony than non-WSART youth. Overall, Hispanic youth who participated in WSART were significantly more likely to recidivate than Hispanic youth who did not participate in WSART. Despite the absence of statistical significance, WSART participants had higher rates of recidivism than non-WSART participants for all racial groups.

The effects of WSART varied between older and younger youth. Youth aged 15 or younger who were assigned to WSART were significantly more likely to recidivate generally and with a felony or violent felony than youth aged 15 or younger who did not participate in WSART. For overall recidivism, there were no significant differences between WSART youth and non-WSART youth who were 16 years old or older. These findings suggest that younger youth may not be as responsive to the WSART program.

Overall, WSART participation led to significantly higher rates of recidivism for high-risk youth but had no significant effect for moderate- and low-risk youth. For high-risk youth, WSART participants were more likely than non-WSART youth to recidivate with a felony and violent felony offense. For moderate- and low-risk youth, WSART participants were significantly more likely to recidivate with a felony or violent felony offense.

We were unable to assess the effects of WSART on recidivism for the youth who were most likely to be assigned to WSART. In our sample, the 5% of WSART-eligible youth who were most likely to be assigned to WSART were assigned to WSART. It is possible that WSART does have a positive

effect on this population, and that the overall lack of an effect of WSART is associated with the courts' increasing reliance on WSART for youth with less complex needs.[35]

3) Under what conditions is WSART most effective?

The negative effects of WSART were significant under the PACT assessment system (2011–2015) and the BOT assessment system (2005–2011). However the differences for felony and violent felony recidivism were more pronounced under the PACT system than the BOT system. We were unable to identify why the effects of WSART changed over time. Session-level and trainer-level data were not available for WSART sessions prior to 2011. Without these data, we are unable to test whether characteristics of the sessions or competence of WSART trainers changed over time.

Although prior evaluations found that trainer competence moderates the effect of WSART, our analyses found no differences in the rate of recidivism for youth in competent WSART court programs and non-competent WSART court programs. However, given the data limitations, it would be inappropriate to conclude with certainty that trainer competence does not moderate WSART program effectiveness.

Although we found no significant differences in recidivism based on court-level trainer competence, we did find significant differences based on completion of the WSART program. Specifically, youth who participated in the full WSART program were significantly less likely to recidivate than youth who participated in but did not complete the WSART program. We conducted additional sensitivity analyses (see Appendix VI) and found that youth who completed WSART were no more likely to recidivate than youth who did not participate in WSART at all.

The findings for completion suggest that dosage matters for WSART effectiveness. It is possible that there are characteristics of WSART trainers that are likely to keep youth engaged and motivated to attend class that are not measured by the current trainer adherence and competence evaluations. Our findings suggest that WSART programs should evaluate methods for keeping youth engaged and that by increasing retention rates, courts may be able to increase the effectiveness of the WSART programs.

Finally, the changing effectiveness of WSART may be related to changes in the underlying juvenile court population, changes in juvenile court processing, or changes in the risk assessment system and placement into juvenile court EBPs. A comprehensive evaluation of larger juvenile court changes over time may assist in understanding why WSART is not an effective program for reducing recidivism rates with current juvenile court populations.

[35] It is possible that our findings represent a net-widening effect. If courts are increasingly expanding their WSART programs, the types of youth who are assigned to WSART may change over time. For example, courts may be increasingly assigning younger youth to participate in WSART. Our findings may actually reflect issues related to eligibility and placement into WSART rather than an overall ineffectiveness of WSART. Future studies should examine how the risk profiles of youth have changed over time and how the courts' use of EBPs has changed over time.

Appendices

Washington State's Aggression Replacement Training for Juvenile Court Youth: *Outcome Evaluation*

Appendices

I. Data

For our evaluation of Washington State Aggression Replacement Training (WSART), our main analyses required data that identify youth who participated in WSART and comparable youth who did not, and the corresponding criminal history and recidivism for each youth. Our subanalyses explored under what conditions WSART is most effective and required data that identify WSART session-level information (e.g., class participation and attendance) and trainer adherence and competency information. This section of the appendix describes our different data sources and the process of combining them into an analytic data set.

Information regarding WSART referrals and completion, as well as information for candidate comparison youth, is housed in the Juvenile Assessment Research Database (ARD). Criminal history and recidivism data came from WSIPP's Criminal History Database (CHD). The WSART Session Database contains session- and class-level data and the WSART Quality Assurance Coordinator in Snohomish County maintains trainer data regarding adherence and competency.

Juvenile Assessment Research Database (ARD)

The Juvenile Assessment Research Database (ARD), housed and maintained by the Administrative Office of the Courts (AOC), contains information from the actuarial risk assessment tool used in the 33 juvenile courts in Washington State. The risk assessment used by the courts was developed to determine the risk of recidivism, to identify those who would most benefit from rehabilitation efforts, and to aid in the development of case management plans to rehabilitate justice-involved youth. The ARD contains useable assessment information starting in January 2005, and our evaluation used data available as of October 2018.

Risk Assessment Tool and Process

The risk assessment tool produces an overall risk-level classification as well as scores on 12 domains concerning specific youth risks and needs.[36] Together, the risk-level classification and the domain scores determine eligibility for each of the available evidence-based programs (EBPs). The original tool was developed by WSIPP in consultation with the Washington State Association of Juvenile Court Administrators and was named the Washington State Juvenile Court Assessment (WAJCA).[37] The tool was updated in 2011 and renamed the Positive Achievement Change Tool (PACT). While questions and scoring have changed slightly across versions, each version is administered using the same process.[38]

At adjudication, a youth is administered a *prescreen* assessment to identify their risk-level classification. Youth found to be moderate- or high-risk after the prescreen assessment receive an *initial* assessment approximately 30 days after the court disposes the case. The initial assessment is an expanded version of the prescreen and involves a structured interview by a juvenile probation counselor (JPC) with both the youth and family. Scores on the initial assessment are the basis for determining EBP eligibility. As youth complete interventions and their supervision requirements, *reassessments*[39] may be administered to evaluate any changes in the youth's risk level or needs. A *final* assessment is administered near the end of supervision to evaluate any changes to their risk level or needs.

Eligibility and Referrals to Evidence-Based Programs

WSART is one of six evidence-based programs (EBPs) funded by CJAA in the juvenile courts. The other five interventions are Coordination of Services, Education and Employment Training, Functional Family Therapy, Multisystemic Therapy-Family Integrated Transitions, and Multisystemic Therapy. Each program has specific eligibility criteria based on different domain scores and individual item scores.

Youth may be eligible for multiple programs, and eligibility for a program does not guarantee that a youth is referred to and starts a program. Other factors that determine intervention receipt are the geographic availability of interventions, the youth's prior participation in EBPs, the parental capacity to meet the participation requirements of the intervention, and the general JPC discretion. In addition to the risk assessment domain scores, the ARD also includes intervention start and completion dates, completion status, and reasons for non-completion.

[36] The 12 domains are: criminal history, demographics, school, use of free time, employment, relationships, family, alcohol and drugs, mental health, attitudes, aggression, and social skills.

[37] Barnoski, R. (2004). *Washington State Juvenile Court Assessment Manual, Version 2.1.* Olympia: Washington State Institute for Public Policy

[38] WSIPP received an updated scoring guide for the PACT from the Washington State Case Management Assessment Process (CMAP) Coordinator to adjust for differences in the scores for PACT assessments in our sample.

[39] In the WAJCA, reassessments were administered at different points of the youth's supervision. These reassessments often occurred after successful completion of a program and did not represent an offense resulting in a new juvenile court case. In the PACT, reassessments describe a new offense and correspond to a different court case than the case that prompted their initial assessment. With this information, we include reassessments in our sample if the assessment version was the PACT.

ARD Data Structure

Information captured by the risk assessment is in separate data tables in the ARD that is linked by unique identifiers for youth and assessments (individuals can have more than one assessment). We used four tables within the ARD: *Domain Score, Assessments, Interventions,* and *Person*.

The *Domain Score* table contains assessment-level records of the aggregate scores on the 12 domains in the PACT, and the *Assessments* table contains individual items comprising each domain. We used youth and assessment identifiers to link assessment records to domain records for complete information on each assessment.

The *Interventions* table contains EBP eligibility, program start and completion dates, and reasons for program non-starts and non-completions. Data are at the intervention-level, with a record for each intervention the youth is eligible for in a given assessment. The file is restructured to the assessment-level and merged with assessment scores using youth and assessment identifiers.

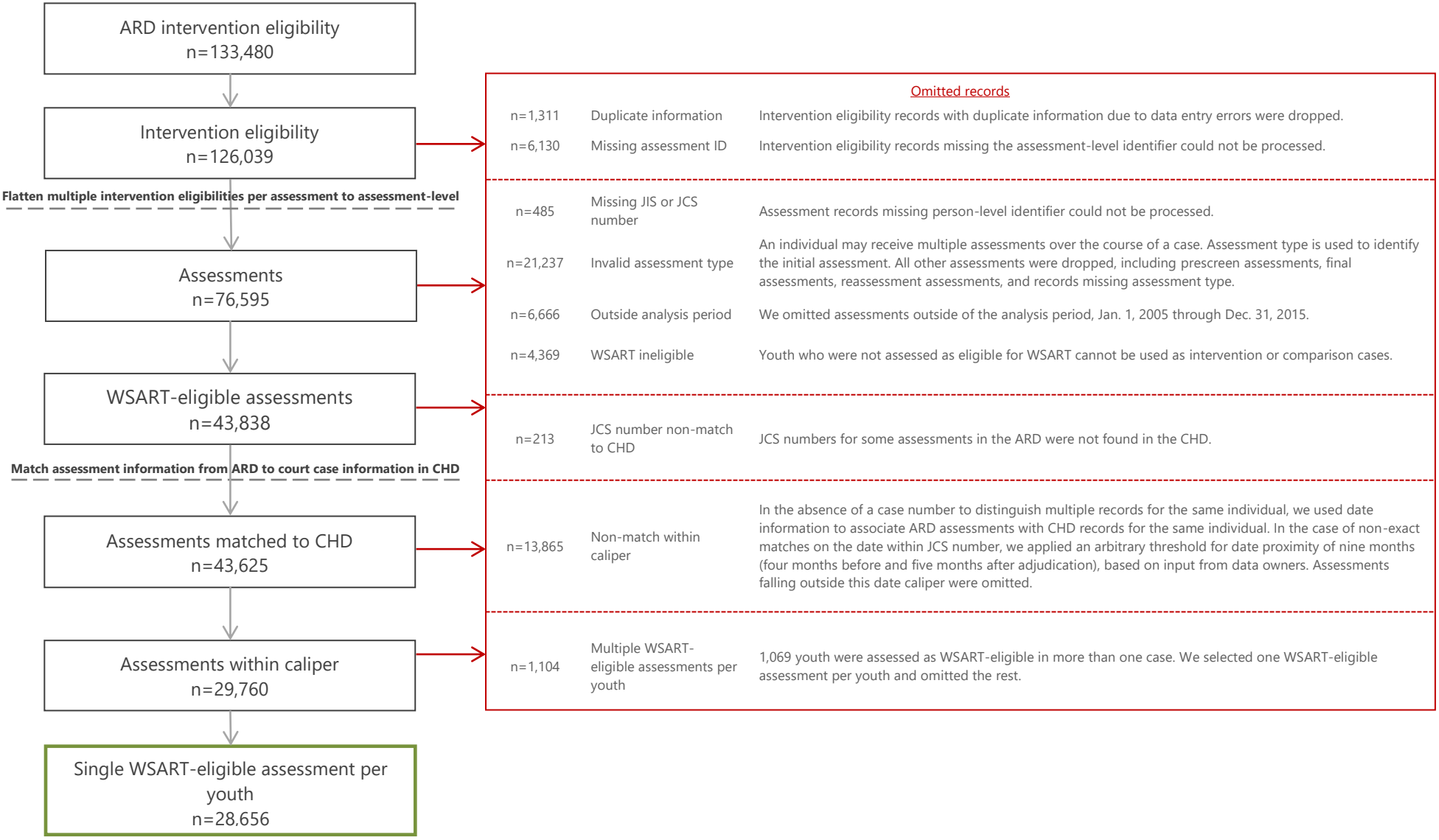
Finally, we linked data from the *Person* table, including person characteristics (first name, last name, date of birth, sex, and race), court information, and youth identifier variables for linking with criminal history data.[40]

Exhibit A1 demonstrates the data processing of the ARD, the assessments omitted from our analysis due to missing information, the process of restructuring information to achieve an assessment-level dataset, and the selection of assessments indicating WSART eligibility that move forward to be associated with criminal history and recidivism data.

[40] Unique youth identifiers used by AOC are the Judicial Information System (JIS) and Juvenile and Correction System (JCS) numbers.

Exhibit A1

Data Processing to Obtain WSART-Eligible Sample with Risk Assessment and Criminal History Data



Notes:
Assessment Research Database (ARD)
Criminal History Database (CHD)
Judicial Information System (JIS) or Juvenile and Corrections System (JCS) number
Washington State Aggression Replacement Training (WSART)

Criminal History Database (CHD)

WSIPP's Criminal History Database (CHD) combines data from several Washington State agencies: court data from the Administrative Office of the Courts, residential confinement data from Juvenile Rehabilitation at the Department of Social and Health Services, and incarceration in state prisons and community supervision data from the Department of Corrections. The CHD allows researchers to create criminal history and recidivism measures for all justice-involved youth in the state. WSIPP updates the CHD on a quarterly basis, and our analyses used information in the CHD as of February 2019.

ARD/CHD Compiled Dataset

ARD and CHD records can be reliably associated at the youth level using the Judicial Information System (JIS) and Juvenile and Correction System (JCS) numbers. However both data sources can include multiple records for a given youth—e.g., the ARD can include multiple assessments for a given case or separate cases, and the CHD can include multiple cases over time. Unfortunately, there is no court case identifier to link records between these two data sources (there is a court case number in the CHD but not the ARD). In the absence of a case identifier, we use date variables—associating the date the assessment was started from the ARD[41] and the date of adjudication from the CHD[42]—using an inexact matching method. There is substantial variability from one case to the next in the sequence and timing of assessment and adjudication. Typically, but not always, the initial assessment follows adjudication. The amount of time between these events varies widely. Based on input from WSART and PACT stakeholders, we applied a caliper to define the maximum temporal distance for an acceptable match between assessment and adjudication dates. Assessments falling within four months before adjudication[43] or five months after adjudication were assumed to be events within the same court case. When multiple records fell within the caliper, we selected the closest temporal match. Exhibit A1 accounts for the data lost when establishing the link between the ARD and the CHD.

Our final step in data processing is to select a single case per youth to avoid including youth who received or were eligible for WSART more than once (outlined in green in Exhibit A1). Appendix II further discusses our approach to selecting a single WSART eligibility in such cases, as well as propensity score methods for comparison group selection.

WSART Trainer Adherence and Competency

For WSART, the EBP Trainer Adherence Tool was developed to measure adherence to the program model and overall trainer competence.[44] WSART trainers are required to submit video footage of a class taught for each curriculum component for a WSART consultant to evaluate annually. Trainers identified as competent retain the designation for a year, until their next review. For trainers deemed not competent, improvement plans and follow-up video submissions are required to bring trainers into good standing. The process aims to maintain adherence to the WSART model and ensure overall trainer competence to support positive outcomes for youth.

---

[41] The variable *assessstarteddate* comes from the *Domain Scores* data table. It was selected over the *assessstarteddate* in the Interventions data table because it was missing less often in the Domain Scores table compared to the Interventions table.
[42] The variable *bestadjudicationdate* is calculated first using the adjudication date associated with a case. If the adjudication date is missing, we use the disposition date. If the disposition date is missing, we take the sentence date. If the sentence date is missing, we take the file date.
[43] If a match occurred before the adjudication date, the start date of the assessment had to occur after the offense date.
[44] WSIPP's 2004 evaluation found that youth with competent trainers and therapists had the lowest rates of recidivism. Youth with trainers and therapists deemed not competent had higher rates of recidivism, higher even than youth who did not receive the program. Barnoski (2004).

The tool measures adherence on the three components of the WSART curriculum (anger control, moral reasoning, and social skills) and overall trainer competency. Each of the three curriculum components of WSART receives a score out of 100, with higher scores representing greater fidelity to the curriculum component. The scores from these three sections are not necessarily representative of competency. A trainer may show adherence to the curriculum components of WSART model but may not deliver the components competently. For competency, a trainer receives an overall score out of a possible 54 points, with higher scores representing greater competency. The overall competency score comes from scores on six criteria[45] for each of the three curriculum components on a scale of zero to three, with higher scores indicating greater competence.[46] Importantly, the adherence tool does not provide cut points for the overall competency scores and competence classifications.

The WSART Quality Assurance Coordinator, Christopher Hayes, maintains the EBP Trainer Adherence Tool scores. The trainer scores we received were available from 2010 to 2018 from 25 juvenile courts. In addition to the scores, each tool reports the name of the trainer, the county or court of the trainer, the date of the session submission, and the specific class submitted for review.

We had several concerns about the reliability of the data included in the trainer adherence tools. First, the assessment tools do not include a manual and there are no quality checks to ensure that the information is complete or accurate. Second, there is no guide indicating what score should be associated with a particular competence classification. Third, there were many adherence tools that included the measures of adherence for the three curriculum components but did not include any scores for overall competence. In some instances, reviewers provided a competence classification on each of the three curriculum adherence components despite the fact that competence is supposed to be measured independent from curriculum adherence. Fourth, it was unclear whether the dates on an adherence assessment represented the dates of the videotaped sessions or the date on which the therapist was reviewed by a WSART consultant. Most adherence tools included only a year, but they did not specify if the year was a calendar year or a fiscal year.

To validate the data included in the trainer adherence tools, we sought additional information on trainer competence. Juvenile Rehabilitation provided WSIPP with WSART Consultant Annual Reports[47] for fiscal years 2014 to 2018. Consultant Annual Reports include trainer names, court, and overall competency ratings for trainers under the supervision of each consultant. The classifications reported in the annual reports determine whether a trainer is assigned to a formal improvement plan.

We classified trainers as highly competent, competent, borderline competent, or not competent using the classifications in the annual report. In instances where the trainer's classification was missing on the annual report, we used the overall competency score from adherence assessment. Exhibit A2 reports the ranges of scores that qualify trainers for each of the four categories of competency.

[45] The six criteria for competency are 1) adherence to the model, 2) clarity, 3) simplicity, 4) objectivity, 5) pacing, and 6) engagement.
[46] A score of zero is *not competent*, a score of one is *borderline competent*, a score of two is *competent*, and a score of three is *highly competent*.
[47] The WSART Consultant Annual Reports are submitted to the Juvenile Rehabilitation (JR) Juvenile Court Administrator and Statewide WSART Quality Assurance specialist. The report provides a summary of information for all WSART trainers for a single consultant.

Exhibit A2

Competency Classifications for WSART Trainers

Competency rating	Scores range
Not competent	0-13
Borderline competent	14-27
Competent	28-41
Highly competent	42-54

There are many instances in which WSART trainers have missing or incomplete competency scores.[48] Given our concerns of reliability, we used the average competency score at the court-level. With the court and year information, we calculated the average competence rating for a given court in a given year. This strategy resulted in 127 competency designations.

WSART Session Data

WSART trainers compile data on each WSART session using the WSART Session Database (hereafter "Session Database"). The Session Database includes youth-level, class-level, and session-level information for each WSART course. The information in the database is reported by WSART trainers and the database is housed and maintained by the AOC.

The Washington State Juvenile Court Administrators established the Session Database in 2012 in order to collect more information about the administration of WSART in juvenile courts. Use of the database was gradually introduced to juvenile courts beginning in 2012. Exhibit A3 indicates the year that each juvenile court began using the Session Database.

[48] For example, a trainer who did not remain in their position long enough to complete an annual review would be missing a competency score. Those trainers, who may have led WSART sessions with youth, were never evaluated and never received a competency designation.

Exhibit A3

Session Database Onboarding Year by Juvenile Court

Juvenile court	Database start (month/year)
Snohomish	10/2012
Yakima	12/2012
Spokane	01/2013
Chelan	06/2013
Island	06/2013
Adams	09/2013
Clark	09/2013
Grant	09/2013
Mason	09/2013
Okanogan	09/2013
Thurston	01/2014
Whatcom	01/2014
Jefferson	06/2014
Pierce	09/2014
Benton/Franklin	01/2015
Asotin/Garfield	01/2015
King	02/2015
Walla Walla	08/2015
Kitsap	11/2015
Douglas	01/2016
Cowlitz	06/2016
Lewis	06/2016

Session Data Structure

WSIPP received 19 separate files of data. Some files included information for each youth in each class of a WSART session (e.g., the attendance file); some files included information about each class in a session (e.g., the class file); some files included information about the overall WSART session (e.g., session general comments and session files); and some files included information about the individuals in a WSART session (e.g., person file, session staff file, and student file). We compiled the data to establish a final dataset that included information about each youth for each class within each WSART session.

Matching Session and ARD Data

We matched youth in the Session Database to youth in the ARD database using the JCS number. Youth may be included in the ARD database multiple times. In order to match the WSART Session Database records to the ARD record that corresponds to the youth's participation in WSART, we identified matches using the session start date from the Session Database and the WSART start date from the ARD records. We counted records as a match if the dates were within 30 days of one another.

As expected, some counties were disproportionately represented in our matched sample due to the gradual implementation of the WSART database. Clark, Pierce, Snohomish, Spokane, and Yakima accounted for 65% of the matches between the Session Database records and the ARD records. Within our larger ARD sample, the aforementioned five counties accounted for only 49% of all WSART participants.

We were unable to match all youth in the ART sessions. While we could use the session data to construct class- or session-level demographic characteristics, such as the percent of females or percent of White youth in each session, we would have incomplete youth-level records to test the effects of session composition on WSART program effectiveness.

For the youth who we were able to match, we looked for variance in individual-level characteristics such as the number of classes in a session. We found that all sessions included 30 classes, and most youth who successfully completed the program had perfect or near-perfect attendance records. The ARD data includes a variable to account for why a youth may not successfully complete the WSART program. ARD records indicating unsuccessful completion due to lack of participation were generally consistent with individual-level attendance records from the session database.

Due to concerns about generalizability, missing data, and the absence of unique information in the Session Database, we ultimately decided not to use session data in any of our analyses. The session data do appear to be more complete in recent years (e.g., 2016, 2017, and 2018). However, we were unable to use recent cohorts because of the need for an 18-month follow-up period to calculate recidivism outcomes. As more data become available, more comprehensive analyses may be possible using the WSART Session Database.

II. Study Group Selection and Matching Procedures

This study uses administrative data to examine whether or not WSART significantly reduced the likelihood of recidivism among moderate-/low- and high-risk court-involved youth from 2005 to 2016. During these years, youth were not randomly assigned to participate in WSART. In the absence of randomization, this study uses propensity score matching, a quasi-experimental method, to approximate randomization and minimize the presence of selection bias.

Propensity score matching identifies youth who did not participate in WSART but who are similar to youth who did participate in WSART on observable characteristics. This section of the appendix describes the process of identifying youth in the treatment group and comparison group and the propensity score methods used to match youth in each group.

Study Groups

We selected treatment group youth by identifying PACT assessments that included a referral to and participation in WSART from January 1, 2006, to June 30, 2016. We selected comparison group youth by identifying risk assessments that included a referral to WSART but that did not include participation in WSART.

Youth with Multiple Assessments

Given the length of our study time frame (ten years), it is possible that the same youth could have multiple juvenile court cases and, consequently, multiple risk assessments. Including multiple assessments for the same youth is problematic for two reasons. First, youth may be included in the treatment and comparison groups if they are eligible for WSART in multiple assessments but did not always participate in WSART. Second, youth with multiple assessments may be higher-risk for recidivism than youth with only one assessment. Including multiple assessments for the same youth could bias the sample such that the findings would be driven disproportionately by a small population of high-risk youth. To address these concerns, we randomly selected only one assessment for youth with multiple assessments for our sample and excluded all other assessments.[49]

No Treatment vs. Treatment as Usual

Our study is not able to use a true control group of youth who receive no treatment. Rather, youth in the comparison group may have participated in a different EBP, no EBPs, or other forms of treatment programs offered by local juvenile courts. Prior analyses of WSART and other juvenile court EBPs restricted their analyses to youth who did not participate in any EBPs.[50] By comparing EBP participants to individuals who do not participate in EBPs, these studies attempt to estimate the effect of a particular EBP to no treatment (a true control group).

---

[49] An alternative approach is to select only the first assessment for each youth. We examined both approaches and found no significant differences in the characteristics of the samples when using the first assessment or a random assessment. We decided to choose a random assessment to minimize the potential for bias that may be associated with unobserved characteristics that we could not examine. Additionally, selecting the first assessment for youth with multiple assessments may artificially increase the likelihood of recidivism in our sample. Since youth with multiple assessments committed multiple offenses at different times, selecting the first assessment would systematically include assessments for which treatment failed and the youth recidivated. For further validation, we replicated our analyses on the sample of youth for which we selected the first assessment rather than a random assessment. The overall findings for recidivism were consistent with the findings presented in this report.

[50] Peterson (2017) and Fumia, D., Drake, E., & He, L. (2015). *Washington's Coordination of Services program for juvenile offenders: Outcome evaluation and benefit-cost analysis* (Doc. No. 15-09-1901). Olympia: Washington State Institute for Public Policy.

Limiting the comparison group to youth who do not participate in EBPs could bias the findings in two ways. First, youth who do not participate in any EBPs may be more likely to have committed less serious offenses, resulting in relatively short periods of supervision. Often times, youth do not participate in any EBPs if their prescribed length of supervision is not long enough to complete an EBP. Second, youth who do not participate in a juvenile court EBP may be more likely to participate in substance use or mental health treatment. Unfortunately, non-CJAA funded treatment programs are not recorded in administrative court data, and the absence of CJAA-funded treatments may not actually mean the youth received no treatment.

Similarly, prior studies of some juvenile court EBPs constructed comparison groups using youth from juvenile courts that do not offer the EBP of interest.[51] However, WSART is one of the most widely available juvenile court EBPs, offered in 23 of the 33 juvenile courts. There are differences between the courts—for example, non-WSART courts are more rural. These differences indicate potential unobservable differences between youth who live in WSART court jurisdictions and those who do not. Drawing comparison group youth from these non-WSART courts would likely bias our findings.[52]

For this study, we assess the effect of WSART compared to "treatment as usual." We construct our comparison group from youth in all courts that started at least one youth in a WSART program. Washington State juvenile courts have invested heavily in EBPs for moderate- and high-risk youth. In the absence of WSART, treatment as usual would include a portfolio of other CJAA-funded EBPs (e.g., FFT, MST, EET) and non-CJAA funded treatments (e.g., substance abuse treatment or mental health treatment). By estimating the effects using a treatment as usual approach, our study isolates the likelihood of recidivism for WSART-eligible youth in Washington State if WSART did not exist.[53]

Reasons for Not Starting WSART

The intervention data included some information about why youth did not participate in a program for which they were eligible. We excluded youth from the comparison group if they did not start WSART for the following reasons: deceased, awaiting or involved in in-patient drug treatment, committed to JRA, currently in WSART, incarcerated, no starter information, whereabouts unknown, on warrant status, or moved or is moving out of state. The number of youth dropped for each reason is detailed in Exhibit A4.

If youth were deceased, committed to JRA, incarcerated, or moved out of state, we would not have reliable data regarding recidivism. As such, these youth may appear to not recidivate when they are either not at risk (i.e., confined or deceased) or would not be included in Washington State court data if they did recidivate. For youth who were missing or who were on warrant status, their failure to participate in WSART was due to an event that occurred after their disposition, but prior to the start of a WSART course. Finally, we excluded youth if the data noted that WSART starter and completer information was missing since we cannot be certain that these youth did not participate in WSART.

[51] Fumia et al. (2015) and Miller et al. (2015).

[52] For example, courts in smaller, more rural counties may not offer WSART. However, youth from smaller, more rural counties may be less likely to recidivate generally. As such, analyses showing a higher likelihood of recidivism among WSART participants may actually be capturing the higher likelihood of recidivism among court-involved youth in larger, more urban jurisdictions. A review of OFM statistics found that nine of the ten juvenile courts that do not offer WSART were ranked among the 15 (out of 32) smallest court jurisdictions by population size. Furthermore, three of these nine courts had jurisdiction over multiple, rural counties (Ferry, Pend Orielle, and Stevens; Klickitat and Skamania; and Pacific and Wahkiakum).

[53] We conducted sensitivity analyses that limit the comparison group youth to only those youth who do not participate in any EBPs. See Appendix IV for additional details.

Other Data Exclusions

While coding the data for analyses, we made additional selection decisions that resulted in the removal of some youth from both the treatment and comparison groups. These decisions are summarized in Exhibit A4. First, we removed youth for whom we did not have an adequate follow-up period to measure recidivism. For these youth, the at-risk date did not allow for a complete 30-month follow-up (18-month follow-up and 12-month adjudication period). Second, we excluded youth who were listed as starting WSART but for whom we did not have a WSART start date available in the data. Third, we removed youth who were less than 10 years old or greater than 21 years old on the date they were assessed. Washington State recommends that WSART not be used for youth younger than 12 years old, but it is still possible that youth ages 10 and 11 were assigned to WSART in some courts.[54] During the timeframe of the sample included in our evaluation, Washington State juvenile courts were able to maintain jurisdiction of youth up to age 21. Thus, it is possible that some youth were assessed and assigned to an EBP up to age 21.

Exhibit A4

Cases Removed During Data Processing

Reason for removal	WSART youth	Non-WSART youth	Total
Data from CHD and ARD match	7,658	20,998	28,656
Insufficient follow-up period	36	--	36
Missing ART start date	7	--	7
Youth aged < 10	27	53	80
Youth aged > 21	14	45	59
Missing index offense characteristics	13	16	29
Missing risk assessment information	--	1	1
Courts with no WSART starters[1]		77	77
Reason for not starting WSART:			4,452
In-patient drug treatment		1,136	
Committed to JRA		706	
Currently in WSART		31	
Deceased		28	
Incarcerated		77	
No starter or completer information		801	
Whereabouts unknown		114	
Youth on warrant status		935	
Moved or is moving out of state		624	
Total final sample	7,561	16,354	23,915

Fourth, we removed youth for whom necessary independent variables were missing. Specifically, we removed youth who were missing information about the index offense characteristics and youth who were missing risk assessment information. Finally, we removed youth from Washington State juvenile courts that did not use WSART during the sample timeframe.

[54] Communication with Washington State's Statewide WSART coordinator.

Matching Procedures

Our evaluation relies on administrative data to assess the effects of WSART on recidivism outcomes for court-involved youth from 2006 to 2016. During this time, youth were not randomly assigned to participate in WSART. The absence of random assignment introduces the possibility of "selection bias," or the possibility that youth who do participate in WSART are systematically different from youth who do not participate in WSART. In order to minimize selection bias and to approximate the conditions of random assignment, we use propensity score matching (PSM) to match youth in the treatment group with similar youth in the comparison group.

Ideally, we would directly match youth from the treatment group to youth in the comparison group on all covariates related to treatment assignment and the outcome. However, when the list of observed characteristics related to treatment assignment is large, it is unlikely that a substantial number of exact pairs will exist in both samples. This common matching dilemma is referred to as the "curse of dimensionality."[55] Propensity score techniques overcome this limitation by matching individuals on a single score that represents a vector of observed characteristics. Research demonstrates that, when used appropriately, PSM can produce unbiased estimates that are similar to estimates derived under conditions of random assignment.[56]

PSM includes three steps: 1) estimating the probability that a youth participates in WSART, or the "propensity score," 2) matching youth from the treatment group to youth in the comparison group using the propensity scores to create a balanced, matched sample, and 3) assessing the effects of WSART participation on recidivism by conducting regression analyses using the matched sample. This portion of the appendix describes the first two steps of the process. Appendix III describes the process for estimating the effects of WSART using regression analyses.

First we calculate the propensity score using the following equation:

$$p(x_i) = \Pr(T_i = 1|x_i) = \left(\frac{e^{(\alpha+\beta_1 x_{i1}+\cdots+\beta_k x_{ik})}}{1+e^{(\alpha+\beta_1 x_{i1}+\cdots+\beta_k x_{ik})}}\right)$$

In this equation, $p(x_i)$ is the probability of individual i being in the treatment group (T) given a vector of k covariates identified in Exhibit A8. The propensity score is derived from a logistic regression model in which a represents the intercept, β_j represents the coefficient for covariate x and e is the base of the natural logarithm. After estimating the logistic regression, we use the predicted probabilities from the model as our estimates of the propensity score.

The model predicting assignment to treatment should include covariates related to selection into the treatment group that may also affect the likelihood of recidivism. In addition, the model should include only the covariates that are measured prior to selection into treatment. For example, we would not include a covariate indicating whether or not a youth completed an EBP for their current case since completion is measured after a youth is initially assigned to and begins participating in an EBP.

We included three types of covariates in our selection model: youth demographics, characteristics of the index offense/current case, and characteristics from the risk assessment domains. We tested multiple

[55] Apel & Sweeton (2009).
[56] Rosenbaum & Rubin (1983); Apel & Sweeton (2009); and Campbell & Labrecque (2018).

specifications of the risk assessment data including a model that used the overall domain score (calculated as the risk factor score minus the protective factor score), a model that used the total risk factor and total protective factor scores from each domain, and a model that used the individual factor score for each factor in each domain. Although the model that included each factor in each domain was the most predictive of participation in WSART, many domain items were excluded from the model due to collinearity with other domain items. Ultimately we selected a model that included the overall risk and protective scores from each domain and also included individual domain items that were directly related to eligibility for CJAA EBPs. Exhibits A5–A7 provide a description of the characteristics included in the propensity score model.

Exhibit A5

Propensity Score Matching Covariate Description: Youth Demographics and Case Characteristics

Characteristic	Data source	Description	Values
Youth demographics			
Sex	ARD; CHD	Sex of youth reported on assessment. If no sex was recorded, we used the sex reported in the court data.	Male Female
Race	ARD; CHD	Race of youth reported on assessment. If race was missing, we used the race reported in the court data.	Black/African American American Indian/Alaskan Native Asian/Pacific Islander/Native Hawaiian White
Ethnicity	CHD	Indicator of whether the youth was recorded as being of a Hispanic.	Hispanic Non-Hispanic
Age	ARD; CHD	We use age at the date on which the assessment was started.	11 - 21
Case characteristics			
Court	ARD	Court in which the case was adjudicated.	Juvenile Courts
Offense type	CHD	Offense category for the most serious offense in the index case.	Person Property Sex Drug Other
Offense grade	CHD	Grade of the most serious offense in the index case.	Violent felony Felony Misdemeanor
Type of disposition	CHD	Type of disposition for the index case.	Conviction Diversion Deferred disposition

Note:
CHD = Criminal History Database; ARD = Assessment Research Database.

Exhibit A6

Propensity Score Matching Covariate Description: Assessment Characteristics

Characteristic	Static or dynamic	Risk or protective	Values
Assessment characteristics			
Risk level classification			Moderate-/Low-risk; High-risk
Social history score (pre-screen)			0-18
Assessment version			PACT; BOT
Assessment start year			FY 2005 - FY 2016
Domain scores			
Domain 1: Criminal History	Static	Risk	0-31
Domain 3a: School History	Static	Risk	0-5
Domain 3a: School History	Dynamic	Risk	**0, 2**
Domain 3a: School History	Static	Protective	**0, 2**
Domain 3a: School History	Dynamic	Protective	**0, 2**
Domain 3b: Current School	Dynamic	Risk	0-22
Domain 3b: Current School	Dynamic	Protective	0-17
Domain 4a: Historic Use of Free Time	Static	Protective	**0-4**
Domain 4b: Current Use of Free Time	Dynamic	Risk	**0-1**
Domain 4b: Current Use of Free Time	Dynamic	Protective	0-6
Domain 5a: Employment History	Static	Risk	**0-2**
Domain 5a: Employment History	Static	Protective	**0-5**
Domain 5b: Current Employment	Dynamic	Risk	**0-1**
Domain 5b: Current Employment	Dynamic	Protective	0-9
Domain 6a: History of Relationships	Static	Risk	**0-3**
Domain 6a: History of Relationships	Static	Protective	**0-4**
Domain 6b: Current Relationships	Dynamic	Risk	0-8
Domain 6b: Current Relationships	Dynamic	Protective	0-10
Domain 7a: Family History	Static	Risk	0-13
Domain 7a: Family History	Static	Protective	**0-4**
Domain 7a: Family History	Dynamic	Risk	**0-1**
Domain 7a: Family History	Dynamic	Protective	**0-1**
Domain 7b: Current Family	Dynamic	Risk	0-34
Domain 7b: Current Family	Dynamic	Protective	0-23
Domain 8a: Alcohol and Drug History	Static	Risk	0-22
Domain 8a: Alcohol and Drug History	Static	Protective	0-8
Domain 8a: Alcohol and Drug History	Dynamic	Risk	**0-1**
Domain 8a: Alcohol and Drug History	Dynamic	Protective	**0, 3**
Domain 8b: Current Alcohol and Drug Use	Dynamic	Risk	0-24
Domain 8b: Current Alcohol and Drug Use	Dynamic	Protective	**0-2**
Domain 9a: Mental Health History	Static	Risk	0-9
Domain 9a: Mental Health History	Static	Protective	0-5
Domain 9a: Mental Health History	Dynamic	Risk	**0-2**
Domain 9a: Mental Health History	Dynamic	Protective	0, 1, 4, 5
Domain 9b: Current Mental Health	Dynamic	Risk	**0-4**
Domain 9b: Current Mental Health	Dynamic	Protective	**0-3**
Domain 10: Attitudes and Behaviors	Dynamic	Risk	0-23
Domain 10: Attitudes and Behaviors	Dynamic	Protective	0-18
Domain 11: Aggression	Dynamic	Risk	0-13
Domain 11: Aggression	Dynamic	Protective	0-8
Domain 12: Skills	Dynamic	Risk	0-18
Domain 12: Skills	Dynamic	Protective	0-28

Note:
Domains that are bolded were analyzed using categorical variables. Domains that are not bolded were analyzed using continuous variables.

Exhibit A7

Propensity Score Matching Covariate Description: EBP Eligibility

Eligibility item	Description	Values
ART eligibility		
Domain 1, item 4	Weapons referrals	None; one or more
Domain 1, item 5	Violent misdemeanor referrals	None; one or more
Domain 1, item 6	Violent felony referrals	None; one or more
FIT eligibility		
Domain 9A, item 1	History of suicidal ideation	Yes; No
Domain 9A, item 7	History of mental health problems	Yes; No
Domain 9A, item 14	Current mental health problems	Yes; No
Domain 9B, item 1	Current suicidal ideation	Yes; No
Domain 9B, item 3	Current mental health treatment prescribed	Yes; No
Domain 9B, item 4	Current mental health medication prescribed	Yes; No
Domain 9B, item 5	Current mental health problems interfere with work with youth	Yes; No
Domain 8A, item 1	History of alcohol use	Yes; No
Domain 8A, item 2	History of drug use	Yes; No
Domain 8A, item 6	Current use of alcohol and/or drugs	Yes; No

Note:
Eligibility associated with overall domain scores are captured in the assessment variables. The eligibility for FFT, MST, and EET, are captured in the overall domain scores and do not have any additional item-specific eligibility criteria.

Exhibit A8 summarizes the results of our logistic regression model predicting selection into WSART. Our model also includes dummy variables for courts and for year of assessment to control for systematic differences in selection into WSART associated with fixed court characteristics or systematic differences based on year of court involvement. The results for the court and year fixed effects are suppressed from this exhibit but are available upon request.

Exhibit A8

Summary of Logistic Regression Predicting Assignment to WSART (N = 23,915; R^2= 0.150)

Variable	Sig[#]	Variable	Sig[#]
Demographic and current offense characteristics		Current offense and general risk characteristics	
Age at assessment	***	**Most serious offense grade**	***
Race/ethnicity		**Most serious offense type**	***
Sex		**Risk level**	
Risk assessment characteristics		Total number of eligible EBPs	*
Social history score		**Assessment version**	***
Conviction record, static risk	***	**Juvenile court**	*
School history static risk		**Assessment year**	*
School history, dynamic risk		EBP eligibility characteristics	
School history, static protective	**	**Criminal record—weapon referral**	
Current school, dynamic risk		**Criminal record—one against-person misd. referral**	
Current school, dynamic protective		**Criminal record—two or more against-person misd. referrals**	**
Historic use of free time, static protective	**	**Criminal record—one or two against-person felony referrals**	
Current use of free time, dynamic risk		**Criminal record—three or more against-person felony referrals**	
Current use of free time, dynamic protective		**Mental health history—suicidal ideation**	
Employment history, static risk		**Mental health history—treatment/medication, score 1**	**
Employment history, static protective		**Current mental health status**	*
Current employment, dynamic risk		**Current mental health—suicidal ideation**	
Current employment, dynamic protective		**No current mental health treatment prescribed**	
History of relationships, static risk		**Currently attending mental health treatment**	
History of relationships, static protective	***	**Mental health treatment currently prescribed, not attending**	
Current relationships, dynamic risk		**No current mental health medication prescribed**	
Current relationships, dynamic protective		**Currently taking mental health medication**	
Family history, static risk		**Mental health medication currently prescribed, not taking**	
Family history, static protective		**Current mental health problem interferes with youth**	**
Family history, dynamic risk	**	**History of alcohol use, static risk**	*
Current living arrangements, dynamic protective	*	**History of drug use, static risk**	**
Current living arrangements, dynamic risk			
Alcohol and drug history, static risk	***		
Alcohol and drug history, static protective			
Alcohol and drug history, dynamic risk	*		
Current alcohol and drug use, dynamic risk	***		
Current alcohol and drug use, dynamic protective	*		
Mental health history, static risk			
Mental health history, static protective			
Mental health history, dynamic risk			
Current mental health, dynamic risk			
Current mental health, dynamic protective			
Attitudes/behaviors, dynamic risk	***		
Attitudes/behaviors, dynamic protective	**		
Aggression, dynamic risk	***		
Aggression, dynamic protective			
Skills, dynamic risk	***		
Skills, dynamic protective	***		

Notes:

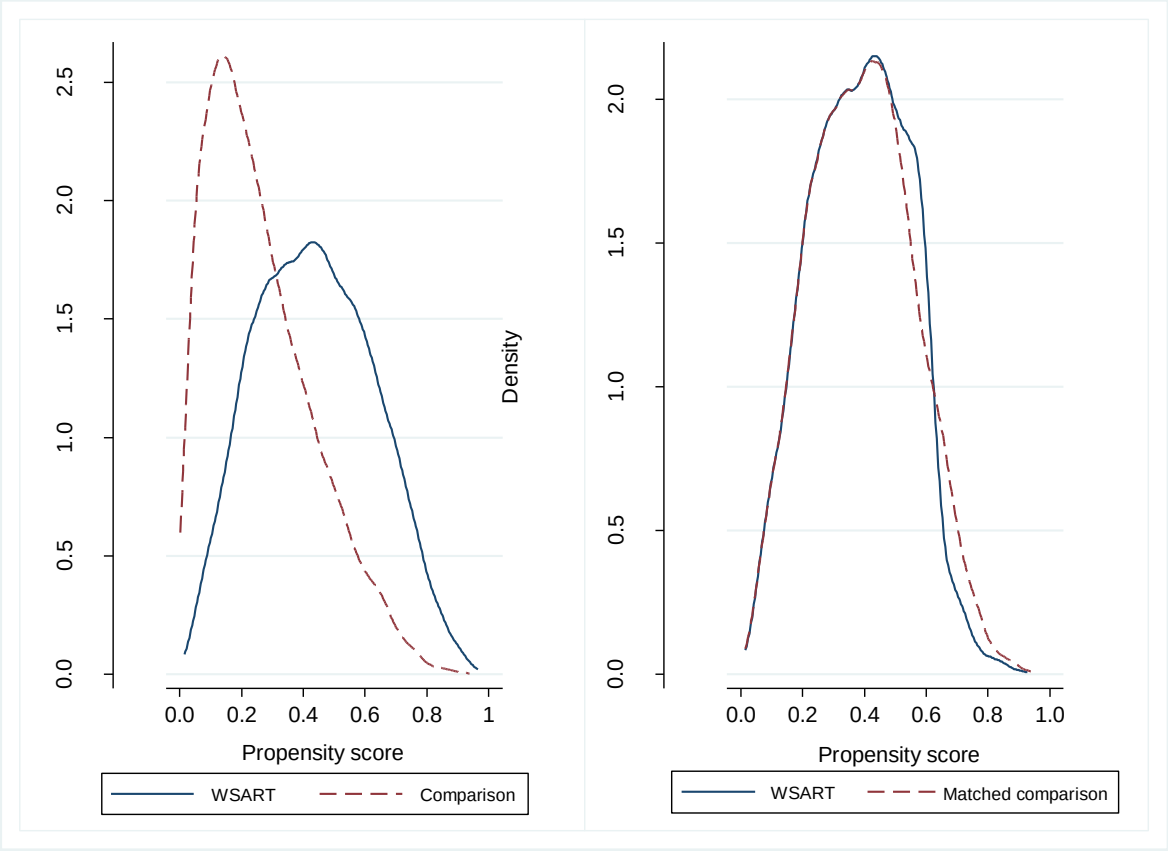
[#] Covariates may be continuous or categorical. Bold variables were included as categorical variables. For categorical variables, significance indicators represent the highest level of significance for any of the individual classifications in the logistic regression model.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Prior to matching, we evaluated the presence of common support between the propensity scores for youth in the treatment group and youth in the comparison group. Exhibit A9 displays the distributions of the propensity scores for youth in each group. The propensity scores for youth who participated in WSART were disproportionately higher than the propensity scores for youth who did not participate in WSART.

Exhibit A9

Distribution of Propensity Scores by Treatment Status, Before and After Matching



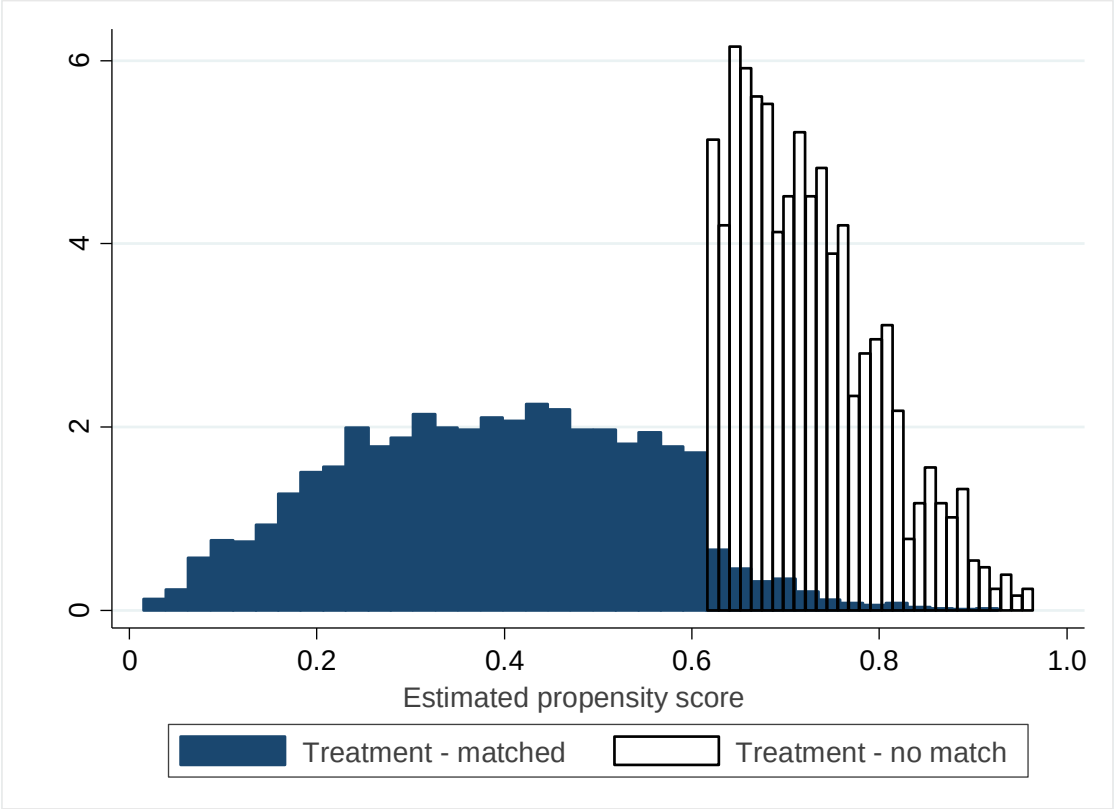
Our preferred method of matching for this evaluation was 1:1 nearest neighbor matching with a caliper and without replacement. There are multiple alternative methods of matching with propensity scores, but prior studies suggest that 1:1 matching with a caliper is the most likely to produce causal estimates that are similar to the results from a randomized control trial.[57] This method ensures that each treatment youth is matched with a similar comparison group and that the youth in each group are included only once. We examine multiple methods of matching (e.g., without a caliper) and propensity score weighting in Appendix IV.

Exhibit A9 shows how PSM identifies samples of youth who are highly similar in the likelihood to participate in treatment. However, we were unable to find good matches for youth in the treatment group who were the most likely to participate in WSART. Our final matched sample excluded 1,108 treatment youth for which there were no youth in the comparison group with a similar propensity score within the specified caliper. Exhibit A10 shows the distribution of propensity scores for youth in the treatment sample who were included and who were excluded in the matched sample. To test whether these exclusions affected the overall results using alternative propensity score methods in Appendix IV.

[57] Campbell & Labrecque (2018).

Exhibit A10

Distribution of Propensity Scores for Matched and Unmatched Treatment Youth



We examined the quality of our propensity score model and matching methods by examining the balance of the covariates in our samples. To assess balance, we examined the standardized difference (SD) using the following equation:

$$\text{SD} = 100 \times \left(\frac{\overline{X}_i - \overline{X}_{i,j}}{\sqrt{(s_i^2 + s_j^2)/2}} \right)$$

The SD calculates the difference in the mean of the covariate for the treated and comparison groups divided by the pooled standard deviation for the covariate. We calculate the SD for each covariate before and after matching. SD values greater than 20 indicate a significant imbalance between the groups. Exhibit A11 and Exhibit A12 display the SD for each covariate before and after matching. After matching, all covariates were balanced at an acceptable level. The effectiveness of our matching method was particularly evident in the significant reduction in bias for the average propensity score value for WSART and non-WSART youth before matching (97.32) and after matching (-3.11).

Exhibit A11

Covariate Balance: Demographic and Index Case Characteristics

Variable	Before matching			After matching		
	WSART Mean	Non-WSART Mean	Bias	WSART Mean	Non-WSART Mean	Bias[#]
Demographic and current offense characteristics						
Propensity score	0.437	0.260	97.32	0.387	0.393	-3.11
Age at Assessment						
12 and under	2.0%	1.4%	4.72	1.9%	1.9%	-0.36
13	7.0%	3.7%	14.44	5.9%	6.1%	-0.76
14	13.6%	8.2%	17.45	12.1%	12.4%	-1.25
15	23.0%	16.8%	15.65	21.7%	22.2%	-1.29
16	27.5%	26.3%	2.61	28.5%	28.1%	0.77
17	24.2%	34.7%	-23.09	26.8%	26.3%	1.10
18 and older	2.7%	8.9%	-26.70	3.1%	2.9%	1.20
Race/ethnicity						
White	64.1%	60.9%	6.49	63.2%	63.0%	0.35
Black/African American	12.8%	14.0%	-3.57	12.8%	12.8%	-0.09
Asian/Pacific Islander/Native Hawaiian	3.0%	2.8%	1.12	2.8%	2.8%	0.19
American Indian/Alaskan Native	3.5%	4.1%	-3.28	3.7%	3.7%	-0.32
Hispanic	15.2%	15.5%	-0.67	15.9%	16.1%	-0.39
Other race/unknown	1.5%	2.7%	-8.62	1.6%	1.6%	0.22
Sex						
Male	73.9%	74.2%	-0.87	74.0%	74.3%	-1.12
Female	26.1%	25.8%	0.87	26.0%	25.7%	0.74
Most serious offense grade						
Misdemeanor	66.6%	68.7%	-4.37	66.9%	67.5%	-1.26
Felony	20.0%	19.8%	0.67	20.5%	20.3%	0.66
Violent felony	13.3%	11.6%	5.38	12.6%	12.2%	0.99
Most serious offense type						
Person	35.2%	26.3%	19.36	32.0%	32.9%	-1.92
Property	39.5%	38.8%	1.42	40.6%	40.3%	0.64
Sex	1.7%	2.7%	-7.17	1.9%	1.7%	1.17
Drug	6.3%	9.4%	-11.45	6.8%	6.6%	0.87
Other offense	17.4%	22.8%	-13.67	18.7%	18.5%	0.43
Risk level						
High-risk	52.6%	57.1%	-8.97	53.2%	53.8%	-1.19
Moderate-/low-risk	47.4%	42.9%	8.97	46.8%	46.2%	1.19
Total number of eligible EBPs	1.857	1.89	-4.50	1.865	1.879	-1.81
Assessment version						
PACT	51.8%	36.1%	32.07	47.5%	48.0%	-0.98
BOT	48.2%	63.9%	-32.07	52.5%	52.0%	0.98

Note:

[#] Bias measures represent the standardized mean difference. Values greater than 25 indicate severe imbalance. Values greater than 10 indicate moderate imbalance.

Exhibit A12

Covariate Balance: Risk Assessment Domains and EBP Eligibility Risk Factors

Variable	Before matching			After matching		
	WSART Mean	Non-WSART Mean	Bias[#]	WSART Mean	Non-WSART Mean	Bias[#]
Risk assessment characteristics						
Social history score	8.47	8.38	3.02	8.45	8.48	-0.87
Conviction record, static risk	7.62	8.84	-31.65	7.77	7.82	-1.36
School history, static risk	3.71	3.58	9.20	3.69	3.70	-0.39
School history, dynamic risk	0.09	0.14	-15.35	0.10	0.10	0.92
School history, static protective	0.09	0.12	-9.45	0.09	0.09	-0.72
School history, dynamic protective	0.91	0.86	15.35	0.90	0.90	-0.92
Current school, dynamic risk	8.41	7.43	18.60	8.28	8.35	-1.23
Current school, dynamic protective	4.13	4.04	2.65	4.13	4.14	-0.11
Historic use of free time, static protective score 1	0.15	0.14	3.63	0.15	0.15	-1.85
Historic use of free time, static protective score 2	0.32	0.33	-1.34	0.32	0.33	-0.13
Historic use of free time, static protective score 3	0.18	0.18	-0.37	0.18	0.18	1.34
Historic use of free time, static protective score 4	0.17	0.17	-0.43	0.17	0.17	1.31
Current use of free time, dynamic risk	0.26	0.27	-2.59	0.26	0.26	-0.32
Current use of free time, dynamic protective	2.29	2.24	3.14	2.30	2.30	0.41
Employment history, static risk score 1	0.03	0.05	-10.61	0.03	0.03	2.07
Employment history, static risk score 2	0.02	0.02	-4.29	0.02	0.02	-0.12
Employment history, static protective	0.63	0.87	-17.36	0.67	0.67	0.07
Current employment, dynamic risk	0.00	0.01	-2.17	0.00	0.00	-0.73
Current employment, dynamic protective	1.44	1.89	-23.08	1.52	1.50	1.35
History of relationships, static risk score 1	0.06	0.06	2.44	0.06	0.06	-0.20
History of relationships, static risk score 2	0.69	0.68	2.42	0.68	0.68	-0.23
History of relationships, static risk score 3	0.18	0.20	-3.54	0.19	0.19	0.28
History of relationships, static protective score 1	0.40	0.37	7.53	0.39	0.41	-2.58
History of relationships, static protective score 2	0.29	0.29	-1.28	0.29	0.28	1.64
History of relationships, static protective score 3	0.11	0.13	-5.58	0.11	0.11	0.77
History of relationships, static protective score 4	0.05	0.07	-6.98	0.06	0.05	1.99
Current relationships, dynamic risk	3.91	3.79	6.28	3.88	3.89	-0.79
Current relationships, dynamic protective	3.89	3.89	0.21	3.89	3.86	1.21
Family history, static risk	2.89	3.07	-7.15	2.92	2.92	-0.29
Family history, static protective score 1	0.13	0.13	-1.66	0.13	0.13	-0.42
Family history, static protective score 2	0.29	0.28	1.50	0.29	0.28	0.96
Family history, static protective score 3	0.37	0.35	3.04	0.36	0.37	-0.91
Family history, static protective score 4	0.15	0.16	-1.40	0.15	0.15	0.77
Family history, dynamic risk	0.01	0.03	-18.40	0.01	0.01	0.12
Current living arrangements, dynamic protective	9.98	9.94	1.01	10.06	10.02	1.00
Current living arrangements, dynamic risk	10.14	9.51	11.50	10.01	10.09	-1.49
Alcohol and drug history, static risk	5.76	7.05	-20.37	6.06	6.00	1.00
Alcohol and drug history, static protective	1.25	1.38	-9.24	1.28	1.30	-1.08
Alcohol and drug history, dynamic risk	0.71	0.74	-6.81	0.72	0.72	0.48

Variable	Before matching			After matching		
	WSART Mean	Non-WSART Mean	Bias[#]	WSART Mean	Non-WSART Mean	Bias[#]
Current alcohol and drug use, dynamic risk	5.81	6.57	-10.99	6.03	5.99	0.62
Current alcohol and drug use, dynamic protective score 1	0.10	0.13	-9.55	0.11	0.10	0.79
Current alcohol and drug use, dynamic protective score 2	0.01	0.02	-4.21	0.01	0.01	0.79
Mental health history, static risk	2.05	2.01	1.70	2.04	2.05	-0.76
Mental health history, static protective	3.67	3.71	-2.91	3.68	3.68	0.36
Mental health history, dynamic risk score 1	0.34	0.33	1.27	0.34	0.34	-0.79
Mental health history, dynamic risk score 2	0.01	0.01	-1.30	0.01	0.01	0.83
Current mental health, dynamic risk score 1	0.09	0.10	-1.82	0.09	0.10	-0.86
Current mental health, dynamic risk score 2	0.03	0.03	1.43	0.03	0.04	-2.37
Current mental health, dynamic risk score 3	0.01	0.02	-2.50	0.01	0.01	0.00
Current mental health, dynamic risk score 4	0.00	0.01	-3.11	0.00	0.00	0.98
Current mental health, dynamic protective score 1	0.08	0.07	3.64	0.08	0.08	0.00
Current mental health, dynamic protective score 2	0.06	0.06	0.49	0.07	0.07	-0.32
Current mental health, dynamic protective score 3	0.03	0.03	2.26	0.04	0.03	0.79
Attitudes/behaviors, dynamic risk	8.26	7.76	10.90	8.07	8.12	-1.12
Attitudes/behaviors, dynamic protective	5.31	5.91	-13.65	5.51	5.47	0.81
Aggression, dynamic risk	5.38	4.68	24.70	5.17	5.21	-1.23
Aggression, dynamic protective	1.79	2.25	-20.59	1.93	1.93	-0.08
Skills, dynamic risk	9.03	6.77	41.11	8.42	8.58	-3.00
Skills, dynamic protective	6.50	8.59	-42.11	6.99	6.83	3.08
Criminal record—weapon referral	0.10	0.11	-6.04	0.10	0.10	0.30
Criminal record—1 against-person misd. referral	0.28	0.27	1.07	0.27	0.28	-0.52
Criminal record—2+ against-person misd. referrals	0.16	0.17	-0.51	0.16	0.16	-0.71
Criminal record—1 or 2 against-person felony referrals	0.19	0.20	-3.37	0.18	0.18	-0.63
Criminal record—3+ against-person felony referrals	0.00	0.01	-6.22	0.00	0.00	0.00
Mental health history —suicidal ideation	0.25	0.24	2.75	0.25	0.25	-0.11
Mental health history —treatment/medication, score 1	0.18	0.17	2.86	0.17	0.18	-0.82
Mental health history —treatment/medication, score 2	0.13	0.14	-2.48	0.13	0.14	-0.46
Current mental health status	0.30	0.29	3.22	0.30	0.30	-0.92
Current mental health—suicidal ideation	-0.63	-0.64	2.25	-0.63	-0.62	-1.31
No current mental health treatment prescribed	0.08	0.07	1.16	0.07	0.08	-1.06
Currently attending mental health treatment	0.13	0.12	3.19	0.13	0.13	0.75
Mental health treatment currently prescribed, not attending	0.05	0.05	-0.82	0.05	0.05	-0.56
No current mental health medication prescribed	0.11	0.10	3.46	0.10	0.10	-0.36
Currently taking mental health medication	0.11	0.11	0.93	0.11	0.11	-0.05
Mental health medication currently prescribed, not taking	0.04	0.04	-1.05	0.04	0.04	0.32
Current mental health problem interferes with youth	-0.62	-0.62	0.06	-0.62	-0.61	-1.43
History of alcohol use, static risk	1.85	2.17	-15.32	1.93	1.91	0.93
History of drug use, static risk	3.95	4.64	-15.91	4.11	4.07	0.90

Note:
[#] Bias measures represent the standardized mean difference. Values greater than 25 indicate severe imbalance. Values greater than 10 indicate moderate imbalance.

III. Estimating the Effects of WSART on Recidivism

In this study, we used four measures of recidivism: any recidivism, misdemeanor recidivism, felony recidivism, and violent felony recidivism. We coded type of recidivism based on the most serious offense in each youth's first recidivism event. Consistent with standards previously published by WSIPP, we measured recidivism as any criminal offense committed within 18 months following the youth's at-risk date that resulted in a court legal action (i.e., conviction, diversion, or deferred disposition) in a Washington State court. We measured recidivism using an 18-month follow-up period and a 12-month adjudication period.

For youth who participated in WSART, the at-risk date was the WSART start date. For youth who did not participate in WSART, we had to construct an equivalent at-risk date. If we were to measure recidivism for the comparison group beginning at the adjudication date, we would potentially include recidivism events for the comparison group that would not be considered for similar youth in the treatment group. That is, if a youth in the treatment group committed an offense prior to starting WSART, it would not be identified as a recidivism event. To ensure that we were not biasing the recidivism measures in the comparison group, we calculated the average time between the adjudication date and the start of WSART for youth in the treatment group. We started the follow-up period for youth in the comparison group based on the average length of time to WSART participation for youth in the treatment group. In our analyses, this meant the at-risk date for the comparison group youth was 97.8 days or 3.2 months after the adjudication date.

We coded two measures of recidivism: any recidivism and type of recidivism. While previous evaluations[58] analyzed type of recidivism by selecting the most serious offense committed within the entire follow-up period, we limited type of recidivism to the most serious offense in the first recidivism event. This change in measurement does not affect the results for any recidivism but may lead to lower rates of felony and/or violent felony recidivism. We use first recidivism event because we currently do not have access to the data necessary to construct adjusted follow-up periods based on subsequent criminal justice contacts. If some individuals are sentenced to detention or a short-term commitment in a Juvenile Rehabilitation facility, we would no longer have a complete 18-month follow-up period in which the youth was at-risk for recidivism. Unequal follow-up periods introduce the possibility of statistical bias that would affect the reliability of our conclusions. For example, if a youth commits a felony and is committed to a JR facility for ten months, they would not have the opportunity to commit a violent felony during the follow-up period.[59]

Additionally, we are interested in estimating the causal effects of WSART. If a youth recidivates with a serious misdemeanor after participating in WSART and is referred to another treatment program (such as FFT) for the new offense, any subsequent offense may be causally related to their experience with the later treatment program and not their experience with WSART. Without using structural equation models or alternative statistical techniques to account for additional interventions following a recidivism event, we do not feel it is appropriate to attribute all future behaviors to an individual's participation in WSART.

[58] Barnoski (2004) and Peterson (2017).
[59] Knoth, L., Wanner, P., & He, L. (2019). *Washington State recidivism trends: FY 1995– FY 2014.* (Doc. No. 19-03-1901). Olympia: Washington State Institute for Public Policy.

We analyzed the effects of participation in WSART on recidivism using logistic regression. The dependent variable was a binary indicator of whether or not the youth recidivated within the follow-up period. We used the Stata margins command to conduct pairwise comparisons between the likelihood of recidivism for youth in the treatment group (WSART participants) and youth in the comparison group (WSART-eligible youth who did not participate in WSART).

We first calculated the raw differences in recidivism for the matched samples. In order to account for any residual imbalance remaining after the propensity score match, we estimated "doubly robust" regression models whereby we included all of the covariates included in the original propensity score models.[60] The results presented in the main report are based on the regression-adjusted models using 1:1 nearest neighbor matching without replacement.

[60] Publications on PSM find that doubly robust methods (i.e., controlling for matching covariates in the outcome model) are sufficient for eliminating residual bias and producing reliable standard errors in the outcome model. Subsequently, we report results using the analytical standard errors from doubly robust methods rather than reporting our results using bootstrapped standard errors. See Ho, D.E., Imai, K., King, G., & Stuart, E.A. (2007). Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Political analysis, 15*(3), 199-236.

IV. Sensitivity Analyses—Propensity Score Matching

We completed multiple sensitivity tests to check the consistency of our findings with different specifications of the propensity score match. Using the previously established propensity scores, we conducted the propensity score match using 1:1 nearest neighbor without replacement and without a caliper; 1:1 nearest neighbor without replacement, with a caliper, and matched in descending order;[61] 1:1 nearest neighbor with replacement with a caliper; and 3:1 nearest neighbor with replacement. In addition, we tested propensity score weighting methods to examine models that maintained the full treatment and control samples.

For methods that matched with replacement, we included frequency weights in our regression models to account for the fact that observations in the comparison group may serve as comparisons for multiple observations in the treatment groups. For methods using a caliper, we estimated the appropriate caliper as 0.2 times the standard deviation of the logit of the propensity score.[62]

Our findings were generally consistent across all of the propensity score matching and propensity score weighting methods.

[61] Matching in a descending order means that the algorithm will first select a match for the youth with the highest propensity score in the treatment group. We tested this method to ensure that we obtained the maximum number of matches for youth with the highest propensity to be assigned to WSART.

[62] Rosenbaum, P.R., & Donald B.R. (1985). Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *The American Statistician, 39*(1), 33-38 and Austin, P.C. (2011). Optimal caliper widths for propensity-score matching when estimating differences in means and differences in proportions in observational studies. *Pharmaceutical Statistics, 10*(2),150–161.

Exhibit A13

Differences in Recidivism Using Alternative Propensity Score Matching and Weighting Specifications

	Any recidivism					Misdemeanor recidivism					Felony recidivism					Violent felony recidivism				
	Non-WSART	WSART	% point diff.	SE	Sig.	Non-WSART	WSART	% point diff.	SE	Sig.	Non-WSART	WSART	% point diff.	SE	Sig.	Non-WSART	WSART	% point diff.	SE	Sig.
Unmatched	39.0%	41.3%	2.4%	0.007	0.001	27.8%	28.5%	0.7%	0.007	0.297	11.2%	13.0%	1.8%	0.005	0.000	4.6%	5.9%	1.3%	0.004	0.000
1:1 No replacement, no caliper	36.0%	38.5%	2.5%	0.008	0.001	25.6%	26.5%	0.9%	0.007	0.201	10.4%	12.0%	1.6%	0.005	0.001	4.1%	5.6%	1.4%	0.003	0.000
1:1 No replacement, with caliper	**36.5%**	**39.3%**	**2.8%**	**0.008**	**0.000**	**25.8%**	**27.0%**	**1.2%**	**0.008**	**0.107**	**10.7%**	**12.4%**	**1.7%**	**0.005**	**0.002**	**4.1%**	**5.6%**	**1.5%**	**0.004**	**0.000**
1:1 No replacement, with caliper, descending	36.2%	38.7%	2.5%	0.008	0.001	25.7%	26.6%	0.9%	0.007	0.223	10.4%	12.1%	1.7%	0.005	0.001	4.2%	5.6%	1.4%	0.004	0.000
1:1 With replacement and caliper	35.3%	38.1%	2.8%	0.007	0.000	25.1%	26.4%	1.2%	0.007	0.078	10.1%	11.8%	1.7%	0.005	0.001	4.0%	5.5%	1.5%	0.003	0.000
3:1 With replacement and caliper	36.1%	38.1%	2.0%	0.008	0.020	25.9%	26.3%	0.4%	0.008	0.626	10.2%	11.8%	1.6%	0.005	0.003	4.0%	5.5%	1.5%	0.004	0.000
ATT weighting	35.8%	38.0%	2.2%	0.008	0.005	25.7%	26.3%	0.6%	0.007	0.383	10.2%	11.7%	1.6%	0.005	0.002	4.1%	5.5%	1.4%	0.003	0.000
IPTW weighting	38.8%	41.3%	2.5%	0.008	0.003	27.7%	28.2%	0.5%	0.008	0.489	11.1%	13.1%	2.0%	0.006	0.000	4.5%	5.8%	1.3%	0.004	0.001

Notes:
Calipers were calculated as 0.2 times the standard deviation of the log of the propensity scores. 1:1 and 3:1 matching were conducted using nearest neighbor greedy matching.
The results presented are from regression-adjusted logistic regressions that include all covariates specified in the propensity score model in addition to control measures for each of the non-WSART evidence-based programs a youth started for a given assessment.
The bold model represents our preferred results.

V. Sensitivity Analyses—Estimating the Effects of WSART on Recidivism

In order to assess whether or not WSART is effective at reducing recidivism, we would ideally compare WSART participants to WSART-eligible youth who receive no treatment. However, most moderate- and high-risk youth sentenced in juvenile courts receive some type of treatment. Our analyses assessed the effectiveness of WSART compared to treatment as usual (TAU). TAU for moderate- and high-risk youth includes a portfolio of juvenile court EBPs (e.g., FFT, MST, and EET) and non-EBPs (e.g., substance use disorder programs and mental health treatment). Previous research suggests that findings from effectiveness evaluations using TAU be affected by the population included in the comparison group.[63]

We conducted additional analyses to examine whether our results were driven by the characteristics of TAU for the comparison group. First, we restricted the analyses to treatment youth who were eligible only for WSART (5,900 out of 7,561). Second, we restricted the analyses to treatment youth who were eligible only for WSART (5,900) and restricted the comparison group to youth who did not participate in any other CJAA-funded juvenile court EBP (14,280 out of 16,354).

WSART vs. No Juvenile Court EBP

We replicated our methods from the main analyses, using propensity score models to match each treatment youth to the most similar comparison youth. We matched treatment to comparison youth using one-to-one nearest neighbor without replacement and calculated a new caliper using the variance of the new propensity scores. We used the same covariates as the model included in the main analyses. Analyses of post-matching balance showed that all covariates were balanced after matching.[64]

[63] Löfholm, C.A., Brännström, L., Olsson, M., & Hansson, K. (2013). Treatment-as-usual in effectiveness studies: What is it and does it matter? *International Journal of Social Welfare, 22*(1), 25-34.
[64] Results available upon request.

Exhibit A14

Differences in Recidivism Using Alternative Sample Specifications

	Any recidivism					Misdemeanor recidivism					Felony recidivism					Violent felony recidivism				
	Non-WSART	WSART	% point diff.	SE	Sig.	Non-WSART	WSART	% point diff.	SE	Sig.	Non-WSART	WSART	% point diff.	SE	Sig.	Non-WSART	WSART	% point diff.	SE	Sig.
WSART-only treatment	37.1%	38.5%	1.4%	0.01	0.126	26.7%	26.9%	0.3%	0.01	0.735	10.5%	11.5%	1.1%	0.01	0.078	1.1%	5.3%	1.1%	0.00	0.006
WSART-only treatment, no EBP comparison	36.3%	39.3%	3.1%	0.01	0.001	26.1%	27.2%	1.1%	0.01	0.182	10.2%	12.2%	2.0%	0.01	0.001	1.3%	5.5%	1.3%	0.00	0.002

Notes:
We use 1:1 nearest neighbor matching without replacement and include a caliper calculated as 0.2 times the standard deviation of the log of the propensity scores. 1:1 and 3:1 matching were conducted using nearest neighbor greedy matching.
The results presented are from regression-adjusted logistic regressions that include the covariates specified in the propensity score model.

The first set of findings restricts the treatment group to test whether the findings from the main analyses were driven by a youth's participation in multiple EBPs (e.g., WSART and FFT). The second set of findings provides the most restrictive review of WSART effectiveness by comparing the youth who received only WSART to youth who did not start any CJAA-funded EBP. It is important to note that these analyses are based only on the EBP participation associated with a youth's initial risk assessment. It is possible that youth in the treatment group or comparison group had a reassessment for which they were referred to a CJAA-funded EBP. However, the current ARD database does not have a direct link between initial assessments, reassessments, and final assessments.

These findings speak to the effectiveness of WSART compared to no EBP treatment. However, these findings do not reflect the overall effects of WSART in Washington State. Almost 22% of the treatment youth participated in multiple EBPs including WSART. In addition, we did not have information about youths' participation in non-EBP treatment programs. These findings do not generalize to the overall effects of WSART in real-world conditions. Regardless, the results are not substantively different from our main findings.

VI. Sensitivity Analyses—Subgroup Analyses

In order to test whether or not WSART has different effects for subpopulations, we replicated our propensity score matching and recidivism analyses for different subgroups. An alternative approach would be to just estimate the marginal effects of WSART for different subgroups using the main analyses conducted on the full sample. However, conducting independent analyses by subgroup has two main advantages. First, if there are differences in the selection bias by subgroup, conducting independent propensity scores and matching ensures that the models minimize selection bias equally for each specific subgroup. Second, by matching within subgroup, we maximize balance in the covariates for youth in the treatment group and youth in the comparison group, establishing the most reliable comparison for the effects on each subpopulation.

The main report includes figures for the marginal effects of WSART by sex, age, race/ethnicity, risk level, and type of risk assessment. Exhibit A15 shows the percentage point differences and significance for each type of recidivism by subgroup. Each of the subgroups were analyzed separately using 1:1 nearest neighbor matching without replacement using a caliper of 0.2 standard deviations of the propensity scores for the overall subgroup. We checked for covariate balance with each subgroup match and found acceptable levels of balance for each matched group.

For race/ethnicity, we were unable to analyze American Indian/Alaskan Native, Asian/Pacific Islander/Native Hawaiian, or other race/unknown separately due to inadequate sample sizes Exhibit A16 shows the unadjusted rates of recidivism and sample sizes by racial/ethnic group and treatment group.

Our main model and sensitivity testing by subgroups include fixed effects for court and year of assessment. Similar to our concerns with some racial groups, we did not conduct individual models by court or by year due to limitations associated with maintaining an adequate sample size. However, Exhibit A17 shows the unadjusted rates of recidivism and sample sizes by juvenile court and treatment group. We were unable to analyze the juvenile courts separately due to inadequate sample sizes.

Exhibit A15

Differences in Recidivism Using Alternative Sample Specifications (1:1 NN Without Replacement, With a Caliper)

	Any Failure					Misdemeanor Failure					Felony Failure					Violent Felony Failure				
	Non-WSART	WSART	% point diff.	SE	P-value	Non-WSART	WSART	% point diff.	SE	P-value	Non-WSART	WSART	% point diff.	SE	P-value	Non-WSART	WSART	% point diff.	SE	P-value
Sex																				
Male	39.3%	43.5%	4.2%	0.01	0.000	27.0%	28.8%	1.8%	0.01	0.044	12.3%	14.7%	2.4%	0.01	0.000	5.5%	6.9%	1.4%	0.01	0.003
Female	30.5%	26.9%	-3.6%	0.02	0.018	24.1%	22.4%	-1.7%	0.01	0.222	6.4%	4.5%	-1.9%	0.01	0.017	2.2%	1.9%	-0.4%	0.01	0.491
Race																				
White	33.9%	35.6%	1.7%	0.01	0.096	24.9%	25.6%	0.7%	0.01	0.418	9.0%	10.0%	0.9%	0.01	0.133	3.2%	4.2%	1.0%	0.00	0.016
Black	45.2%	45.8%	0.6%	0.02	0.794	27.3%	26.1%	-1.2%	0.02	0.579	18.0%	19.6%	1.6%	0.02	0.382	9.9%	10.7%	0.8%	0.02	0.575
Hispanic	39.8%	44.8%	5.0%	0.02	0.016	26.7%	30.5%	3.8%	0.02	0.053	12.9%	14.4%	1.4%	0.02	0.336	5.3%	6.2%	0.9%	0.01	0.391
Age																				
15 and younger	33.9%	36.7%	2.7%	0.01	0.026	23.5%	23.8%	0.2%	0.01	0.841	10.4%	12.9%	2.5%	0.01	0.002	4.1%	5.3%	1.1%	0.01	0.047
16 and older	38.4%	40.5%	2.1%	0.01	0.055	27.7%	29.1%	1.4%	0.01	0.169	10.7%	11.5%	0.8%	0.01	0.253	4.1%	5.6%	1.5%	0.01	0.002
Assessment Type																				
PACT	35.5%	39.2%	3.7%	0.01	0.002	25.7%	26.8%	1.1%	0.01	0.323	9.9%	12.4%	0.8%	0.01	0.002	4.0%	5.7%	1.7%	0.01	0.002
BOT	36.6%	39.8%	3.2%	0.01	0.005	25.7%	27.8%	2.0%	0.01	0.058	10.8%	12.2%	0.8%	0.01	0.078	4.4%	5.5%	1.1%	0.01	0.036
Risk Level																				
High Risk	42.2%	45.2%	3.1%	0.01	0.008	29.2%	30.0%	0.8%	0.01	0.447	12.9%	15.3%	2.4%	0.01	0.003	5.3%	6.9%	1.6%	0.01	0.004
Moderate/Low Risk	29.4%	32.0%	2.6%	0.01	0.025	22.1%	23.3%	1.2%	0.01	0.243	7.3%	8.7%	1.5%	0.01	0.032	2.7%	4.1%	1.4%	0.01	0.003

Exhibit A16

18-Month Unadjusted Recidivism Rates, by Race/Ethnicity and Treatment

Race/ethnicity	WSART		Non-WSART	
	Total n	% Recidivism	Total n	% Recidivism
White	9,964	37.9%	4,844	34.7%
Black	2,286	47.1%	965	45.6%
American Indian/Alaskan Native	676	38.8%	265	39.2%
Asian/Pac. Islander/Native Hawaiian	454	35.9%	224	37.9%
Hispanic/Latino	2,527	44.4%	1,150	44.8%
Other/unknown	447	48.5%	113	47.8%

Exhibit A17

18-Month Unadjusted Recidivism Rates by Juvenile Court and Treatment

Juvenile court	WSART		Non-WSART	
	Total n	% Recidivism	Total n	% Recidivism
Adams	34	52.9%	57	49.1%
Asotin/Garfield	64	48.4%	31	25.8%
Benton/Franklin	397	43.1%	1,185	43.5%
Chelan	225	33.3%	313	33.5%
Clark	526	46.8%	1,222	40.0%
Cowlitz	275	36.4%	510	37.1%
Douglas	62	46.8%	44	52.3%
Grant	142	49.3%	385	48.8%
Grays Harbor	41	56.1%	189	45.5%
Island	144	29.2%	95	42.1%
Jefferson	127	19.7%	46	37.0%
King	579	43.2%	3,103	43.4%
Kitsap	365	31.2%	627	38.0%
Kittitas	45	35.6%	91	41.8%
Lewis	203	31.0%	249	32.1%
Mason	125	34.4%	183	33.3%
Okanogan	159	45.3%	233	45.1%
Pacific/Wahkiakum	38	36.8%	80	35.0%
Pierce	1,068	34.5%	2,098	38.5%
Skagit	19	73.7%	140	60.0%
Snohomish	808	35.9%	1,487	39.1%
Spokane	821	33.6%	828	36.1%
Stevens	20	30.0%	102	35.3%
Thurston	371	36.9%	1,188	35.6%
Walla Walla	86	48.8%	260	39.6%
Whatcom	268	49.3%	494	41.1%
Whitman	24	29.2%	30	40.0%
Yakima	505	39.2%	831	47.8%

We were not able to match specific trainers to particular WSART youth. Rather, we used available information on trainer competence scores to calculate the average trainer competence rating for each court in each fiscal year. We then examined whether youth who participated in WSART had different outcomes based on the court and year in which they participated in WSART. Exhibit A18 shows the regression-adjusted rates of recidivism for youth trained in a court with trainers who were, on average, competent or highly competent, youth trained in a court with trainers who were, on average, borderline competent or not competent, and youth who did not participate in WSART.

Exhibit A18

Regression-Adjusted Differences in Recidivism, by Average Trainer Competence

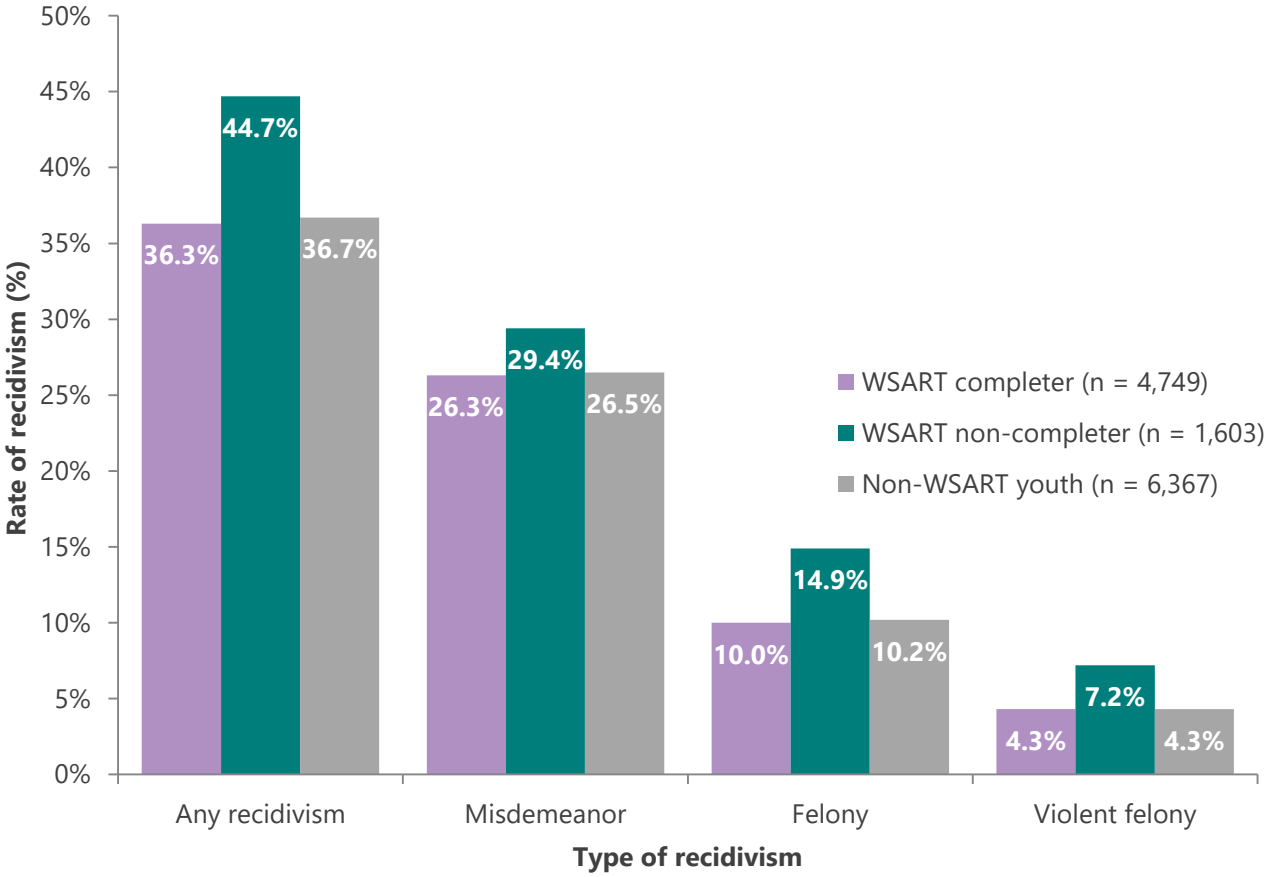
	Non-WSART	WSART not competent	% pt. diff.	SE	p-value	Non-WSART	WSART competent	% pt. diff.	SE	p-value	WSART competent	WSART not competent	% pt. diff.	SE	p-value
Any failure	35.6%	38.1%	2.5%	0.024	0.303	35.6%	38.4%	2.8%	0.012	0.022	38.4%	38.1%	0.3%	0.024	0.890
Misdemeanor failure	26.0%	26.1%	0.1%	0.023	0.961	26.0%	26.0%	0.0%	0.011	0.992	26.0%	26.1%	-0.1%	0.023	0.958
Felony failure	9.6%	12.0%	2.3%	0.016	0.132	9.6%	12.3%	2.7%	0.008	0.001	12.3%	12.0%	0.4%	0.016	0.821
Violent felony failure	3.7%	5.9%	2.2%	0.011	0.046	3.7%	5.9%	2.2%	0.006	0.000	5.9%	5.9%	0.0%	0.012	0.994

Notes:
We use 1:1 nearest neighbor matching without replacement and include a caliper calculated as 0.2 times the standard deviation of the log of the propensity scores.
The results presented are from regression-adjusted logistic regressions that include all covariates specified in the propensity score model.

Finally, we assessed the effect of completion on recidivism. First, we evaluated whether there were differences in the effects of WSART compared to non-WSART participants based on whether or not the youth completed the full program. Exhibit A19 shows the regression-adjusted recidivism rates for WSART completers, WSART non-completers, and non-WSART youth using the original matched sample from the main analyses.

Exhibit A19

18-Month Regression-Adjusted Recidivism Rates, by Completion Status



Notes:
Differences between completers and non-WSART youth were not significant for any outcome.
Differences between completers and non-completers were significant at $p < 0.001$ for any recidivism, felony recidivism, and violent felony recidivism and significant at $p < 0.05$ for misdemeanor recidivism.
Difference between non-completers and non-WSART youth were significant at $p < 0.001$ for any recidivism, felony recidivism, and violent felony recidivism and significant at $p < 0.05$ for misdemeanor recidivism.

It is possible that youth who complete the WSART program are more motivated and less likely to recidivate than youth who do not complete the WSART program. By matching WSART youth to non-WSART youth, we do not account for the potential within-group differences for WSART youth. In order to test whether the differences in recidivism were caused by systematic differences between completers and non-completers, we completed separate analyses in which we matched WSART completers to WSART non-completers based on their propensity to complete the program. We used the same covariates from the main analyses to model new propensity scores for only the WSART participants. We then matched WSART completers to WSART non-completers based on their propensity to complete the program and used logistic regression to assess the effect of completion on recidivism.

We also had to account for the possibility of reverse causality. Specifically, we were interested in knowing whether dropping out has a different effect on recidivism than completing the entire WSART program. As such, we removed youth who did not complete the program because they recidivated. Overall, the majority of youth who started WSART did successfully complete WSART. In order to maximize our sample of WSART completers, we used 1:1 nearest neighbor matching *with* replacement and a caliper. Exhibit A20 shows the regression-adjusted percentages of recidivism and significance tests for differences between completers and non-completers.

Exhibit A20

Regression-Adjusted Percentage in Recidivism, by Completion

	WSART completer	WSART no completion	% pt. diff.	SE	p-value
Any failure	34.2%	42.2%	-8.0%	0.009	0.000
Misdemeanor failure	24.9%	28.7%	-3.7%	0.008	0.000
Felony failure	9.4%	13.4%	-4.0%	0.006	0.000
Violent felony failure	4.1%	6.5%	-2.4%	0.004	0.000

Notes:
We use 1:1 nearest neighbor matching with replacement and include a caliper calculated as 0.2 times the standard deviation of the log of the propensity scores.
The results presented are from regression-adjusted logistic regressions that include all covariates specified in the propensity score model.

Taken together, the completion analyses suggest that completing the full WSART program may have a uniquely beneficial effect on recidivism. The differences in recidivism may not be explained solely by underlying motivations or systematic differences in the youth who complete and youth who do not complete WSART. These findings suggest that programs may be able to increase the effectiveness of the WSART program if they can identify strategies to increase motivation and engagement of youth in order to prevent youth from dropping out of the program.

The risk assessment data included in the propensity score models include a wide range of characteristics that likely influence whether or not a youth will complete the program (e.g., number of prosocial and anti-social friends, relationship with family members, participation in prosocial activities, employment, and belief in using verbal aggression to resolve conflict). However, propensity score matching is not a panacea for eliminating selection bias. The ability to eliminate bias is particularly difficult when modeling completion that is plausibly based on individual, trainer, or session characteristics that we cannot observe. For the main analysis, we modeled assignment to treatment where assignment is based on the judgment of a juvenile probation counselor. The decision of whether or not to assign a youth to WSART should be based largely on the information included in the assessment. For these analyses, we are modeling the likelihood that a youth will complete the program, which is ultimately driven by characteristics that are not necessarily included in the risk assessment such as an individual youth's motivation, number or type of youth in the class, or trainer competence and engagement. Insofar as the decision to drop out of a program is driven by unobservable characteristics that are highly correlated with the observed characteristics, the propensity score models should eliminate associated bias. However, there are likely some unobservable characteristics that are not captured by these models and which may explain some of the difference between youth who complete the program and youth who do not complete the program.

Exhibit A21 provides a summary of the covariates that were significantly different between the full sample of youth who completed WSART and youth who started but did not complete WSART. After matching, the covariates were sufficiently balanced, eliminating the bias associated with these differences. Future analyses should examine the relationship between session and trainer characteristics and attrition from the WSART program.

Exhibit A21

Summary of Significant Differences Between Full Sample of Completers (N = 5,565) and Non-Completers (N=1,978)

Variable	Sig.[#]	Variable	Sig.[#]
Demographic and current offense characteristics		Current offense and general risk characteristics	
Age at assessment		**Most serious offense grade**	***
Race/ethnicity	***	**Most serious offense type**	***
Sex		**Risk level**	
Risk assessment characteristics		Total number of eligible EBPs	*
Social history score	***	**Assessment version**	***
Conviction record, static risk	***	**Juvenile court**	*
School history, static risk	***	**Assessment year**	*
School history, dynamic risk	***	EBP eligibility characteristics	
School history, static protective	***	**Criminal record—weapon referral**	
Current school, dynamic risk	***	**Criminal record—one against-person misd. referral**	
Current school, dynamic protective	***	**Criminal record—two or more against-person misd. referrals**	**
Historic use of free time, static protective	***	**Criminal record—one or two against-person felony referrals**	***
Current use of free time, dynamic risk	***	**Criminal record—three or more against-person felony referrals**	***
Current use of free time, dynamic protective	***	**Mental health history—suicidal ideation**	
Employment history, static risk		**Mental health history—treatment/medication, score 1**	*
Employment history, static protective	*	**Current mental health status**	*
Current employment, dynamic risk		**Current mental health—suicidal ideation**	
Current employment, dynamic protective	***	**No current mental health treatment prescribed**	***
History of relationships, static risk	***	**Currently attending mental health treatment**	***
History of relationships, static protective	***	**Mental health treatment currently prescribed, not attending**	***
Current relationships, dynamic risk	***	**No current mental health medication prescribed**	*
Current relationships, dynamic protective	***	**Currently taking mental health medication**	*
Family history, static risk	***	**Mental health medication currently prescribed, not taking**	*
Family history, static protective	***	**Current mental health problem interferes with youth**	
Family history, dynamic risk	**	**History of alcohol use, static risk**	***
Current living arrangements, dynamic protective	***	**History of drug use, static risk**	***
Current living arrangements, dynamic risk	***		
Alcohol and drug history, static risk	***		
Alcohol and drug history, static protective	***		
Alcohol and drug history, dynamic risk	***		
Current alcohol and drug use, dynamic risk	***		
Current alcohol and drug use, dynamic protective			
Mental health history, static risk			
Mental health history, static protective			
Mental health history, dynamic risk	*		
Current mental health, dynamic risk	*		
Current mental health, dynamic protective	**		
Attitudes/behaviors, dynamic risk	***		
Attitudes/behaviors, dynamic protective	***		
Aggression, dynamic risk	***		
Aggression, dynamic protective	***		
Skills, dynamic risk	*		
Skills, dynamic protective	**		

Notes:

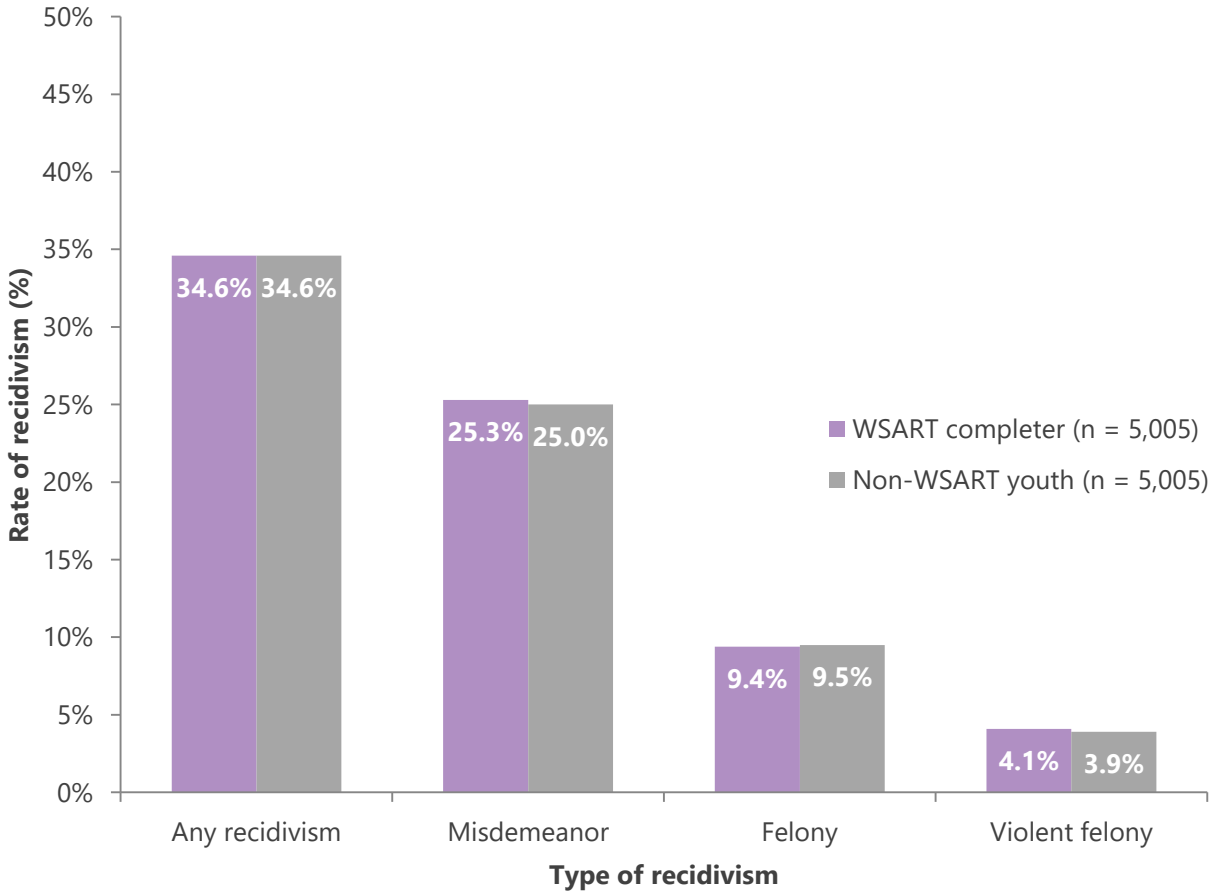
[#] Covariates may be continuous or categorical. Bold variables were included as categorical variables. For categorical variables, significance indicators represent the highest level of significance for any of the individual classifications in the logistic regression model.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

As a final sensitivity test, we analyzed a separately matched sample of WSART participants who completed the program to non-WSART youth. This test allowed us to examine whether WSART completers had outcomes that were significantly different from non-WSART youth. We replicated the analyses from the main model whereby we matched youth using 1:1 nearest neighbor without replacement with a caliper equivalent to 0.2 times the standard deviation of the log of the propensity scores. Exhibit A22 depicts the regression-adjusted recidivism rates for the matched sample of youth who participate in WSART and successfully complete the program compared to youth who did not participate in WSART.

Exhibit A22

18-Month Regression-Adjusted Recidivism Rates, by Completion



Note:
Differences between completers and non-WSART youth were not significant for any outcome

There were no significant differences in recidivism between WSART completers and non-WSART youth. Because we use a treatment as usual comparison group, the absence of significant differences suggests that WSART is equally effective as the portfolio of alternative interventions (e.g., case management, FFT, EET, etc.).

VII. Evidence-Based Program (EBP) Eligibility

Washington State Juvenile Courts use a dynamic risk-needs assessment tool to determine eligibility for evidence-based programming. The assessment was originally known as the Back-On-Track (BOT) assessment but was later renamed the Positive Change Achievement Tool (PACT). Youth are initially assessed using a pre-screen instrument that includes 21 different factors capturing criminal history and social history. The pre-screen classifies youth into three levels: low-risk, moderate-risk, and high-risk. Youth classified as moderate- or high-risk are required to complete the full risk-needs assessment.

The juvenile court assessment gathers additional information for 12 domains:

1) Criminal history
2) Demographics
3) School
4) Use of free time
5) Employment
6) Relationships
7) Family
8) Alcohol and drugs
9) Mental health
10) Attitudes
11) Aggression
12) Social skills

Youth are assigned to participate in EBPs based on their overall risk-level classification, scores on domains from the full assessment, and scores on specific items within the domains. The eligibility for EBPs changed slightly between the BOT and the PACT versions of the assessment. Exhibit A23 summarizes the eligibility criteria for CJAA-funded EBPs.

Exhibit A23

Washington State Juvenile Court Eligibility for Moderate- and High-Risk EBPs

Domain	BOT	PACT
Aggression Replacement Training (ART)		
Risk level	Moderate- or high-risk	Moderate- or high-risk
At least one of the following:		
Domain 1, Criminal history	Static risk factor score of at least one for a weapon (item 4), violent misdemeanor (item 5), or felony conviction (item 6)	Static risk factor score of at least one for a weapon (item 4), violent misdemeanor (item 5), or felony conviction (item 6)
Domain 11, Aggression	Dynamic risk score of at least 2 out of 13	Items 2, 3, and 4 - dynamic risk factor of at least 2
Domain 10, Attitudes/behavior	Dynamic risk score of at least 5 out of 23	items 6-10 - dynamic risk score of at least 5
Domain 12, Skills	Dynamic risk score of at least 4 out of 18	All items except 2 - dynamic risk score of at least 4
Functional Family Therapy (FFT)		
Risk level	Moderate- or high-risk	Moderate- or high-risk
Domain 7b, Current living	Dynamic risk score of at least 8 out of 34	Dynamic risk score equal to or greater than 6
Multi-Systemic Therapy (MST)		
Risk level	High-risk	High-risk
Domain 7b, Current living	Dynamic risk score of at least 8 out of 34	Dynamic risk score equal to or greater than 8
Education and Employment Training (EET)		
Risk level	--	Moderate- or high-risk
Age	--	15 to 18
At least one of the following:		
Domain 3A, School history	--	Static risk score of 4 or 5
Domain 3B, Current school status	--	Dynamic risk score between 7 and 22
Domain 5A, Employment history	--	Static protective score is 0 or 1
Domain 5B, Current employment	--	Dynamic protective score is 0-2
Family Integrated Transitions (FIT)		
Risk level	--	Moderate- or high-risk
At least one of the following:		
Domain 9A, Mental health history	--	History of suicidal ideation (item 1)
Domain 9A, Mental health history	--	History of mental health problems (item 7)
Domain 9A, Mental health history	--	Current mental health problem status (item 14)
Domain 9B, Current mental health	--	Current suicidal ideation (item 1)
Domain 9B, Current mental health	--	Mental health treatment prescribed (item 3)
Domain 9B, Current mental health	--	Mental health medication prescribed (item 4)
Domain 9B, Current mental health	--	Mental health problems interfere with treating the youth (item 5)
At least one of the following:		
Domain 8A, Alcohol and drugs	--	Any past alcohol use (item 1)
Domain 8A, Alcohol and drugs	--	Any past drug use (item 2)
Domain 8A, Alcohol and drugs	--	Current alcohol and/or drug use (item 6)

Note:
Only ART, FFT, and MST were offered when the BOT assessment was used.

For aggression replacement training, domains 10, 11, and 12 primarily drive eligibility. Exhibit A24 provides additional details about the items measured in each relevant domain. Full details for the scoring of individual factors included in each domain are available in the Washington State Juvenile Court Assessment Manual, Version 2.1.[65]

Exhibit A24

Factors Included in PACT Domains 10–12

Domain 10: Attitudes/behaviors	
1	Primary emotion when committing crime(s)
2	Primary purpose for committing crime(s)
3	Optimism
4	Impulsiveness; acting before thinking
5	Belief in control over anti-social behavior
6	Empathy, remorse, sympathy, or feelings for the victim(s) of criminal behavior
7	Respect for property of others
8	Respect for authority figures
9	Attitude toward pro-social rules/conventions in society
10	Accepts responsibility for anti-social behavior
11	Youth's belief in successfully meeting conditions of court supervision
Domain 11: Aggression	
1	Tolerance for frustration
2	Hostile interpretation of actions and intentions of others in a common non-confrontational setting
3	Belief in yelling and verbal aggression to resolve a disagreement or conflict
4	Belief in fighting and physical aggression to resolve a disagreement or conflict
5	Reports/evidence of violence not included in criminal history
6	Reports of problem with sexual aggression not included in criminal history
Domain 12: Skills	
1	Consequential thinking
2	Goal setting
3	Problem-solving
4	Situational perception
5	Dealing with others
6	Dealing with difficult situations
7	Dealing with feelings/emotions
8	Monitoring of internal triggers, distorted thoughts, that can lead to trouble
9	Monitoring of external triggers, events or situations, that can lead to trouble
10	Control of impulsive behaviors that get youth into trouble
11	Control of aggression

[65] Washington State Institute for Public Policy. (2004). *Washington State Juvenile Court Assessment Manual Version 2.1.* Olympia, WA. An updated manual for the PACT is not available at this time. Additional details are available upon request.

Acknowledgements

The authors would like to thank staff at AOC, JR, and Snohomish County Superior Court for helping us obtain the data needed to conduct this evaluation. We are particularly grateful to Christopher Hayes, Cory Redman, Chad Connors, and David Sattler for answering our various questions regarding juvenile court operations, WSART program characteristics, and the overall coverage of the evaluation. Finally, we are thankful for the CJAA advisory committee for their feedback on the final report.

For further information, contact:
Lauren Knoth at 360.664.9805, lauren.knoth@wsipp.wa.gov

Document No. 19-06-1201