Published in final edited form as:

Int J Forensic Ment Health. 2012 ; 11(3): 165–180. doi:10.1080/14999013.2012.737407.

Assessment of Multiple Risk Outcomes, Strengths, and Change with the START:AV: A Short-Term Prospective Study with Adolescent Offenders

Jodi L. Viljoen,
Simon Fraser University

Jennifer L. Beneteau,
Simon Fraser University

Erik Gulbransen,
Simon Fraser University

Etta Brodersen,
Simon Fraser University

Sarah L. Desmarais,
North Carolina State University

Tonia L. Nicholls, and
University of British Columbia and BC Mental Health and Addiction Services

Keith R. Cruise
Fordham University

Abstract

The Short-Term Assessment of Risk and Treatability: Adolescent Version (START:AV; Nicholls, Viljoen, Cruise, Desmarais, & Webster, 2010; Viljoen, Cruise, Nicholls, Desmarais, & Webster, in preparation) is a clinical guide designed to assist in the assessment and management of adolescents' risk for adverse events (e.g., violence, general offending, suicide, victimization). In this initial validation study, START:AV assessments were conducted on 90 adolescent offenders (62 male, 28 female), who were prospectively followed for a 3-month period. START:AV assessments had good to excellent inter-rater reliability and strong concurrent validity with Structured Assessment of Violence Risk in Youth assessments (SAVRY; Borum, Bartel, & Forth, 2006). START:AV risk estimates and Vulnerability total scores predicted multiple adverse outcomes, including violence towards others, offending, victimization, suicidal ideation, and substance abuse. In addition, Strength total scores inversely predicted violence, offending, and street drug use. During the 3-month follow-up, risk estimates changed in at least one domain for 92% of youth, and 27% of youth showed reliable changes in Strength and/or Vulnerability total scores (reliable change index, 90% confidence interval; Jacobsen & Truax, 1991). While these findings are promising, a strong need exists for further research on the START:AV, the measurement of change, and on the role of strengths in risk assessment and treatment-planning.

The Short-Term Assessment of Risk and Treatability (START; Webster, Martin, Brink, Nicholls, & Desmarais, 2009; Webster, Martin, Brink, Nicholls, & Middleton, 2004) is a clinical guide designed to aid in the assessment and management of short-term risks that adults involved in mental health, forensic, or correctional settings may experience, including

---

risk for violence towards others, self-harm, suicide, victimization, substance abuse, self-neglect, and unauthorized leave. Compared to other adult risk assessment instruments, the START was developed with an emphasis on strength factors in addition to vulnerabilities or risk factors, and dynamic or modifiable factors that are relevant to treatment-planning. Furthermore, rather than focusing solely on an individual's risk of violence, the START aims to provide a more integrated assessment of multiple adverse events relevant to care (e.g., self-harm, suicide, victimization, substance abuse).

The START uses a structured professional judgment (SPJ) model of risk assessment, meaning that rather than relying on total scores, evaluators are guided to make a final risk estimate of low, moderate, or high risk on each of the outcomes after systematically considering an individual's current strengths, vulnerabilities, and prior behaviors. In making these final risk estimates, evaluators may place a heavier emphasis on some factors than others and consider case-specific factors that are not included in the START items. In this respect, the START functions as an *aide mémoire* (see Webster, Douglas, Eaves, & Hart, 1997) orienting professionals to evidence-based factors relevant to risk across the eight domains, while allowing evaluators to add their own knowledge regarding how these factors may be best understood in the context of an individual case.

Although the START was developed for use with adults, the authors received numerous queries pertaining to the use of the START with adolescents. Consequently, the Short-Term Assessment of Risk and Treatability: Adolescent Version (START:AV; Nicholls, Viljoen, Cruise, Desmarais, & Webster, 2010; Viljoen, Cruise, Nicholls, Desmarais, & Webster, in preparation) was developed to respond to this clinical interest (see Viljoen, Cruise, Nicholls, Desmarais, & Webster, this issue). At the time the START:AV was developed, several well-validated measures, such as the Structured Assessment of Violence Risk in Youth (SAVRY; Borum, Bartel, & Forth, 2006), existed to assess adolescents' risk of violence. For that reason the START team was reluctant to add another measure to the field. However, attention to broader adverse outcomes, such as self-harm, suicide, victimization, and substance abuse, was lacking. Despite widespread agreement that risk may be particularly dynamic during adolescence (Borum, 2003; Grisso, 1998; Prentky & Righthand, 2003), there were few adolescent measures that were designed to assess *short-term* risk. In addition, although several adolescent risk assessment tools included some protective factors, there remained a need for a more balanced emphasis on risk factors and strength factors. Thus, the adolescent version of the START aimed to complement existing adolescent risk assessment approaches through its focus on multiple adverse outcomes, short-term dynamic assessment, and strength as well as vulnerability factors.

To adapt the START for adolescents, a stepwise, developmentally-informed approach was adopted. First, a measure development team was formed by bringing together members of the adult START (Nicholls, Desmarais, Webster) alongside individuals with clinical and research expertise with adolescents (Cruise, Viljoen). Second, to avoid creating an approach that replicated existing measures, a literature review was conducted to evaluate the potential value of such a measure. Third, a set of developmentally-informed principles were established to guide the adaptation of the START for adolescents. These principles emphasized the need for the adolescent version of the START to a) include developmentally-appropriate risk and protective factors, b) capture the multiple systems that adolescents are embedded within through the inclusion of family, peer, school, community, and individual factors, and c) ground risk assessments in an understanding of normative adolescent development.

Drawing from these principles, the START:AV added several items pertaining to family and peer systems (e.g., Parenting and Home Environment, Social Support from Caretakers and

Other Adults, Relationships with Peers), and adjusted the item anchors and outcomes to take into account adolescents' contexts and lesser maturity compared to adults. For instance, whereas the adult START includes an item on Occupational functioning, this item was revised for adolescents to focus on school as well as employment. Furthermore, the START:AV manual (Viljoen et al., in preparation) compiles adolescent-specific research and highlights developmental issues, thus aiming to facilitate a developmentally-informed assessment. For a more thorough discussion of the rationale and development of the START:AV see Viljoen et al, (this issue).

Research on the Adult START

Although the current study, along with Desmarais et al. (this issue), are the first studies to examine the psychometric properties of the START:AV, a number of studies have examined the adult START. Results on the adult START cannot be generalized to the START:AV, but speaks to the overall framework and may help identify critical research questions to be examined in studies on the START:AV. As such, we briefly summarize findings on the adult START below.

With respect to inter-rater reliability, results generally have found intraclass correlation coefficients (ICCs) to be in the "excellent" range for both Strength and Vulnerability total scores and final risk estimates on the adult version of the START (ICCs > .80s; Desmarais, Nicholls, Wilson, & Brink, 2012; Nicholls, Brink, Desmarais, Webster, & Martin, 2006; Wilson, Desmarais, Nicholls, & Brink, 2010; but see Viljoen, Nicholls, Greaves, de Ruiter, & Brink, 2011 who found somewhat lower inter-rater reliability). Growing research support also has been obtained for the predictive validity of the adult version of the START. Studies have indicated that START final risk estimates and total scores can predict aggression towards others, including physical aggression and violence, verbal aggression, aggression towards objects, and sexually inappropriate behavior (Braithwaite, Charette, Crocker, & Reyes, 2010; Chu, Thomas, Ogloff, & Daffern, 2011; Desmarais et al., 2012; Gray et al., 2011; Nicholls et al., 2006; Nonstad et al., 2010; Wilson et al., 2010).

Research on the ability of the adult version of the START to measure other outcomes of interest, such as self-harm and victimization, is more limited but still quite promising. Gray and colleagues (2011) reported that START final risk estimates predicted self-harm, self-neglect, and victimization in psychiatric inpatients, although Strength and Vulnerability total scores did not. Braithwaite and colleagues (2010) found that START Strength and Vulnerability total scores predicted aggression towards others, unauthorized leave, and substance abuse, but did not predict self-harm, suicidality, self-neglect, or victimization in a small sample of forensic inpatients ($n = 34$). Thus, as research on the START:AV unfolds, it will be particularly important to examine the extent to which the START:AV predicts not only violence but also broader adverse outcomes, as well as its potential utility in non-forensic and community settings.

Given that the assessment of strengths is a major focus of the START, several studies have explored this aspect in greater detail. In user satisfaction surveys, clinicians reported that they found the adult START's attention to both strengths and vulnerabilities to be clinically useful (Doyle, Lewis, & Brisbane, 2008; Kroppan et al., 2011). In addition, research has found Strength total scores to predict lower levels of aggression (Desmarais et al., 2012; Wilson et al., 2010) and more successful community reintegration (Viljoen et al., 2011). Desmarais et al. (2012) found that the Strength total scores were able to add incremental validity beyond that of the Vulnerabilities total scores in predicting aggression. In contrast, other studies have not obtained evidence for incremental validity of strength ratings, possibly as a result of the high inverse correlations between Strength and Vulnerability total

scores ($r > -.80$; Braithwaite et al., 2010; Viljoen et al., 2010; Wilson et al., 2010). Therefore, as research on the START:AV develops, the role of strength factors requires careful attention.

Finally, several studies have examined the dynamic nature and the “shelf-life” of START assessments. Wilson et al. (2010) found that START Strength and Vulnerability total scores were more effective in predicting risk over a short-term follow-up (i.e., 9 months or shorter) than a long-term follow-up (i.e., 12 months), supporting the START’s emphasis on regular reassessment. In addition, Nonstad et al. (2010) found increases in mean strength scores, and decreases in mean vulnerability scores among forensic psychiatric patients during the course of treatment. Given that many view adolescents’ risks as particularly dynamic (Borum, 2003; Grisso, 1998; Prentky & Righthand, 2003), it will be important for research to examine the ability of START:AV assessments to capture change.

The Present Study

Although studies with the adult version of the START support the reliability and validity of START assessments completed in both research and practice, these findings cannot be generalized to the newly-developed adolescent version of the measure, the START:AV. The current study was the first to examine the validity of the START:AV, and had five primary aims. First, we tested the reliability of the START:AV, including both inter-rater reliability as well as internal consistency. Second, to evaluate concurrent validity, we examined the association of START:AV assessments to other risk assessment tools and measures of protective factors (i.e., SAVRY, Borum et al., 2006; Developmental Assets Profile, Search Institute, 2005). Third, we investigated the predictive validity of START:AV assessments, specifically their ability to predict multiple adverse outcomes (i.e., violence towards others, general reoffending, suicidal ideation, non-suicidal self-harm, victimization, and substance abuse) over a short-term (3-month) follow-up period. Fourth, we dedicated specific attention to strengths, including not only the ability of strengths to predict outcomes, but also correlations between strength and vulnerability ratings, the incremental validity of Strength total scores, and the extent to which risk level moderated the relationship between Strength total scores and outcomes (see Fergus & Zimmerman, 2003). Finally, as the START:AV was designed to be dynamic in nature, we investigated 3-month changes in the START:AV ratings through group-based statistics (e.g., *t*-tests) as well as individually-oriented approaches (e.g., reliable change index; Jacobson & Truax, 1991).

Method

Sample

The present sample ($n = 90$) consists of 28 female and 62 male adolescent offenders who were between the ages 13 and 18 at the time of the first administration of the START:AV ($M = 16.38$, $SD = 1.15$). While the majority of the sample (55.6%, $n = 55$) identified themselves as Caucasian, 27.8% ($n = 25$) described themselves as at least partly First Nations/Aboriginal, 8.9% ($n = 8$) as Asian, 6.6% ($n = 6$) as Hispanic, 5.6% ($n = 5$) as East Indian, and 14.4% ($n = 13$) as other races/ethnicities. Index offences varied, and involved violence, such as assault or robbery (61.1%; $n = 55$); property offences, such as theft or mischief (35.6%; $n = 32$); and other offences, such as violations or weapons-related offences (27.8%; $n = 25$). The majority of youth did not have prior charges or convictions (74.4%; $n = 67$, 82.2%; $n = 74$, respectively).

Follow-up data were available for 81 of the 90 youth (90.0%). Of those without follow-up data, four had not yet reached the 3-month follow-up point at the time of data analysis, and five had missed their interview and did not have any file information to review. In 15 cases,

the interviews were not able to be completed and only official justice records were available. Participants who missed the 3-month follow-up interview did not significantly differ from the other participants in terms of START:AV scores, age, gender, or ethnicity.

Procedures

Ethics approval for the study was granted by the university and the probation offices. To be eligible to participate, youth had to be between ages 12 and 18 at the time of the initial baseline interview and adjudicated by the courts. Adolescents were sampled from ten probation offices throughout the region as part of a larger longitudinal study on adolescents' mental health, risks, and strengths.¹ Efforts were made to inform all youth at the offices about the study through study liaisons, probation officers, posters, and flyers. Of the eligible youth approached ($N = ?$), 31.6% declined participation because they were not interested ($n = 116$). In addition, some youth were unable to be contacted because their whereabouts were unknown (7.1%, $n = 26$), or guardian consent could not be obtained, often as a result of not being able to contact guardians to obtain consent (8.4%, $n = 31$). However, the ethnic and gender distribution of our sample is very similar to national and provincial rates for youth involved in the justice system, suggesting that our sample is fairly representative in this respect (Calverley, Cotter, & Halla, 2010).

Parents or other legal guardians provided active consent (rather than implied consent) for youths' participation after being informed about the study by phone and mail, and youth provided assent/consent. Interviews with youth were conducted at the youth's probation office, or a quiet public place, such as a coffee shop. Youth were reimbursed for their participation with gift cards after each interview (\$20 CAD for the baseline interview and \$15 CAD for follow-up interviews). Following each interview, interviewers reviewed the youth's official youth justice records, which ranged in length and quality, but typically contained a running log of the probation officer's contact with the youth, records of attendance at various programs and services, psychiatric reports, police reports, and interviews with the family and/or the youth.

Interviewers were Master's- or PhD-level graduate students, or Honours Psychology students or graduates ($n = 12$). All interviewers had previously completed course work pertaining to clinical and offender populations, and had held volunteer or paid work positions with adolescent offenders. Interviewer training consisted of approximately three days of instructional sessions on the interview process and administration of the START:AV and the SAVRY, followed by the completion of five or more practice cases on these risk measures; five practice cases involved vignettes or mock data, and one (or more) involved an in-person practice case in which a trainee was paired with an experienced rater to jointly interview and assess an adolescent study participant. Prior to working independently, trainees were required to demonstrate adequate inter-rater reliability (which was defined as scores within 5 points of the total scores on consensus or gold standard rating for each risk assessment measure).

Measures

START:AV—The START:AV (Nicholls et al., 2010; Viljoen et al., in preparation) is a structured professional judgment approach designed to assess short-term risks for multiple adverse outcomes, including risk of violence toward others, suicide, self-harm, victimization, substance abuse, unauthorized leave, self-neglect, and general offending. Although general offending was not included as an outcome in the original adult version of

---

¹The START:AV was implemented mid-way through this larger study. Thus, the sample size of the existing analyses is smaller than that of the larger study.

the START, it was added to the START:AV because general offending is a frequent concern in youth justice and mental health contexts². The START:AV includes 23 dynamic and treatment-relevant items that pertain to the youth and his/her social context (e.g., Social Skills, Emotional State, Substance Use, Support from Caregivers and Other Adults, Support from Peers, Parenting and Home Environment). Each item is rated based on the past 3 months using a 3-point scale for Strength and a separate 3-point scale for Vulnerability (0 = *minimally present*, 1 = *moderately present*, 2 = *maximally present*). Evaluators can identify Strength and Vulnerability items (current or historical) that are particularly relevant to risk management for a particular youth; these are referred as key items (for Strength items) or critical items (for Vulnerability items). After completing item ratings, evaluators consider the youth's history of adverse events. Then, using all of the available information and their structured professional judgment, evaluators make final risk estimates for the youths' short-term risk of adverse events; this risk estimate takes into account both dynamic factors as well as the youths' history. Separate risk estimates (low, moderate, high) are made for each of the outcomes examined by the START:AV. In the current study, Strength and Vulnerability total scores were prorated if there were four or fewer missing items. At the time of the study the full START:AV manual (Viljoen et al., in preparation) was not completed; thus, assessments were conducted using the START:AV abbreviated manual (Nicholls et al., 2010).

Concurrent Measures of Risks and Strengths—The Developmental Assets Profile (DAP; Search Institute, 2004) is a 58-item, self-report questionnaire that assesses strengths or assets in personal, social, school, community, and family domains. Items are rated on a 4-point Likert-type scale ranging from “*not at all*” to “*almost always*.” The DAP has good concurrent validity (Search Institute, 2004), and in our sample, coefficient alpha for its domain scales were all greater than .79, except for the School domain which had an alpha coefficient of .69.

The SAVRY (Borum et al., 2006) is a structured professional judgment risk assessment tool designed to assess risk for violence in adolescents. It is comprised of 24 risk factors in historical, social-contextual, and individual/clinical domains, as well as six protective factors (e.g., resilient personality traits, strong commitment to school). A body of research supports its concurrent and predictive validity (e.g., Olver, Stockdale, & Wormith, 2009). In the present study, the inter-rater reliability for the SAVRY risk total was strong; the intra-class correlation coefficient for a single rater was .90 (absolute agreement; McGraw & Wong, 1996).

Outcomes—Outcomes were measured both continuously as well as dichotomously through both validated self-report measures and official reoffense records. While many risk assessment studies rely on dichotomous outcomes only, predictive power is typically diminished when outcomes are dichotomized, and there is a loss of information regarding individual differences (MacCallum, Zhang, Preacher, & Rucker, 2002).

Violence and offending behaviors: Violence and offending behaviors were assessed with the Self Report of Offending (SRO; Knight et al., 2004; Huizinga, Esbensen, & Weiher, 1991). We examined SRO Total scores, which consist of 23 items, as well as youth's scores on the SRO Aggressive Offenses scale, which consists of 10 items (e.g. “how many times in the past few months have you taken something from another person by force, using a weapon?”). Although SRO items are often coded dichotomously, the present study coded responses on a three-point scale to obtain greater information on the frequency of

²The addition of the General Offending risk estimate to the adult START is anticipated in version 2.0.

reoffending (0 = *Never*, 1 = *1 time*, 2 = *2 or more times*). The SRO has been found to have good psychometric properties across gender and ethnic groups (Knight et al., 2004). In the present sample, coefficient alpha was .90 for SRO Total, and .78 for SRO Aggressive Offenses, indicating acceptable internal consistency (i.e., defined as alpha > .70; George & Mallery, 2003). In addition, we examined official youth justice records to determine the presence or absence of arrests for any offense or a violent offense during the follow-up period.

Victimization: Victimization was assessed with the Overt Victimization scale of the Problem Behavior Frequency Scale (PBF Overt Victimization Scale; Multisite Violence Prevention Project, 2004), a 6-item scale that measures the frequency of physical and verbal victimization in the past thirty days. It asks youth how many times they have been hit, pushed or shoved, yelled at, threatened to be hit or physically harmed, threatened or injured with a weapon, and asked to fight. Items were coded on a 3-point scale (e.g. 0 = *Never*, 1 = *1 to 2 times*, 2 = *3 or more times*) and summed to create a total score. In the current study, coefficient alpha was .83. We also dichotomized scores to examine the presence or absence of victimization.

Suicidal ideation, suicidal behavior, and non-suicidal self-harm: Suicidal ideation was measured with the Suicide Ideation scale of the Massachusetts Youth Screening Instrument-Second Version (MAYSI-2; Grisso & Barnum, 2003). This scale is comprised of five self-report items, with questions covering thoughts and intentions about suicidal behavior during the past several months, as well as depressive symptoms that may increase the risk for suicide. It has been found to have good concurrent validity, construct validity, test-retest reliability, and internal consistency (Archer, Simonds-Bisbee, Spiegel, Handel, & Elkins, 2010; Archer, Stredny, Mason, & Arnau, 2004; Grisso, Barnum, Fletcher, Cauffman, & Peuschold, 2001). In the present sample, coefficient alpha was .85.

In addition, youth were asked about suicide attempts and non-suicidal self injury, drawing from questions used in adolescent health surveys (Carolina Population Center, 1999). Youth were first asked broadly about suicidal *and non-suicidal* forms of self-injurious behavior (i.e., “In the past few months, how often have you hurt yourself on purpose? This can include different things like cutting or burning yourself, taking an overdose, or banging your head on purpose *even if you weren’t trying to kill yourself*.”) In order to differentiate suicidal and non-suicidal behaviors, we then asked youth specifically about suicide attempts. From these questions, we created a non-suicidal self-injury (NSSI) variable, which captured youth who had engaged only in non-suicidal self-harm and did not attempt suicide. We examined suicide attempts as a separate variable.

Substance abuse: Substance abuse was measured with the Alcohol/Drug Use Scale of the MAYSI-2. The Alcohol/Drug Use scale is comprised of eight items with questions related to negative consequences of use and the presence of characteristics thought to represent risk factors for abuse. It has good psychometric support in samples of adolescent offenders (e.g., Archer et al., 2004). In the present sample, coefficient alpha was .88. In addition, as a dichotomous measure of substance use, participants were asked if they had been drunk, used marijuana, and/or used street drugs during the past 30 days using questions from the Drug and Alcohol Use – Teen Conflict Survey (DAU; Bosworth & Espelage, 1995). Given that alcohol and marijuana use is fairly common during adolescence, we also created a dichotomous variable for street drug use.

Analyses

Reliability and Validity—To examine inter-rater reliability, intra-class correlation coefficients (ICCs) were calculated on a random sample of 12 cases (13.3%) using a random effects model (McGraw & Wong, 1996). Given that the START:AV can be used by both individual professionals or as part of a multidisciplinary team, we presented ICCs for single raters (ICC₁) as well as the average ICCs across raters (ICC₂). For final risk estimates, we also calculated rates of inter-rater agreement. Alpha coefficients were examined as a measure of internal consistency. Because items on risk assessment tools are sampled to measure diverse risk factors rather than a single unitary construct, internal consistency may not be a necessary feature of many risk assessment tools (Douglas, Skeem, & Nicholson, 2011). However, it might inform decisions regarding whether the creation of a total score is justifiable for research purposes (see Nunnally, 1967, p. 261).

Concurrent validity was examined through bivariate correlations with SAVRY total and section scores, and DAP asset domain scores. To test predictive validity, we first conducted bivariate correlations between START:AV Total scores, risk estimates, and outcome scales. Preliminary examination of the data revealed substantial positive skew on the SRO Total and Aggressive Offenses scales.³ Given that Pearson correlations can be affected by non-normality in data (Blair & Lawson, 1982; Kowalski, 1972), a square root transformation was conducted to normalize this data after setting the minimum value at 1 (Osborne, 2002). For dichotomous outcomes, we also completed receiver operating characteristic curve (ROC) analyses to generate area under the curve scores (AUCs; Mossman, 1994).

Changes in START:AV Scores—Several methods were used to examine changes in START:AV ratings. First, we calculated stability coefficients, and compared mean total scores for the group at baseline and at the follow-up using paired sample t-tests. In addition, to provide an individually-oriented analysis of change, we examined the proportion of adolescents who showed changes in risk estimates, and calculated Jacobson and Truax's reliable change indices (RCI; Jacobson & Truax, 1991) for Strengths and Vulnerability total scores.⁴ The RCI differentiates several individual patterns of change (i.e., reliable increases, reliable decreases, and no reliable change), and addresses measurement error by calculating whether an individual showed more change than one would expect based on chance or error alone. Compared to other approaches, the RCI is at least as conservative in measuring declines in scores, and is more conservative in measuring reliable improvements (Marsden et al., 2011). Although it is not without limitations (Lunnen & Ogles, 1998), several critiques recommend the use of the RCI (Bauer, Lambert, & Nielsen, 2004; Maassen, 2001; Marsden et al., 2011) and it is one of the most commonly-used methods to measure individual change in clinical contexts (Ogles, Lunnen, & Bonesteel, 2001).

Results

New START:AV Items: Correlational Analyses

The START:AV includes several items that were not in the adult version of the START (Webster et al., 2004; 2009). In order to place greater attention on family and peer systems,

³The original SRO Total score had a skewness of 2.30 and a kurtosis of 5.79. In contrast, normal distributions have skewness and kurtosis values between 0 and +/- 1 (Osborne, 2002). The square root transformation produced distributions that more closely approximated a normal distribution (i.e., skewness and kurtosis for transformed scores ranged between 0 and +/-1.3).

⁴The RCI was calculated by: $RCI = \frac{X_2 - X_1}{S_{diff}}$, where X_1 is the START:AV strength or vulnerability total score at baseline and X_2 is the total score at the 3-month follow-up (Jacobson & Truax, 1991). S_{diff} is the standard error of measurement of the two scores and is calculated as: $S_{diff} = \sqrt{2(SE)^2}$, and SE (SEM) was calculated as: $SEM = s_x(\sqrt{1 - r_{xx}})$, where s_x is the standard deviation of scores and r_{xx} is the reliability of a test.

the START:AV authors added Parenting and Home Environment (Item 21), and subdivided Relationships and Social Support into new items focused on caretakers and other adults (i.e., Items 2a and 11a) and peers (Items 2b and 11b). Thus, in our initial analyses, we examined the correlations between these new items to determine if they contributed unique information to the assessment.

Correlations between Relationships from Caretakers and Other Adults (Item 2a) and Relationships from Peers (Item 2b) were significant but fell in the small range according to Cohen's (1988) criteria ($r = .23$ and $.28$ respectively for Strength and Vulnerability ratings, $p < .01$). Correlations between Social Support from Caretakers and Other Adults (Item 11a) and Social Support from Peers (Item 11b) were moderate for Strength and Vulnerability ratings ($r = .32$ and $.50$ respectively, $p < .01$). Parenting and Home Environment (Item 21) was moderately correlated with Relationships with Caretakers and Other Adults (Item 2a; $r = .50$ and $.31$ respectively for Strength and Vulnerability, $p < .01$) and Social Support from Caretakers and Other Adults (Item 11a; $r = .41$ and $.56$, $p < .01$). Given that these correlations were generally moderate rather than large, the new START:AV items appear to provide some unique information. As such, we retained them in the subsequent analyses and in the calculation of Strength and Vulnerability total scores.

Descriptive Statistics: Strength and Vulnerability Scores

Means and standard deviations for START:AV strength and vulnerability item ratings and total scores are shown in Table 1.⁵ Strength and Vulnerability total scores did not differ significantly for male and female youth. Notably, all of the youth in our sample were identified as showing moderate or maximal strengths in *five* or more domains on the START:AV, with strength scores ranging from 7 to 40 out of a total possible score of 44. In fact, 85.6% ($n = 77$) of youth were rated as having moderate or maximal strengths on at least 10 items, and 53.3% ($n = 48$) had moderate or maximal strengths on 15 or more items.

Strength and Vulnerability total scores had a strong inverse correlation ($r = -.74$, $p < .01$). At the item level, correlations between Strength and Vulnerability ratings ranged from $-.32$ (Item 10: External Triggers) to $-.76$ (Item 15: Rule Adherence), as shown in Table 1. Based on Cohen's (1988) criteria for interpreting correlation sizes, 11 items had strength-vulnerability correlations that fell into the moderate range (i.e., $r = .30$ to $.50$) whereas 12 items had strength-vulnerability correlations that were large (i.e., $r > .50$).

Interrater Reliability and Internal Consistency Reliability

Coefficient alpha was .89 for the Strength and .89 for the Vulnerability total scores, which is considered good (George & Mallery, 2003). ICCs are often interpreted as follows: $< .40$ – poor; $.40$ to $.59$ – fair; $.60$ to $.74$ – good; and $.75$ to 1.00 – excellent (Cicchetti & Sparrow, 1981). Using these criteria, ICC₁s (single raters) for the Strength and Vulnerability total scores fell in the “excellent” range (see Table 2). For risk estimates, ICC₁s were slightly lower, with half falling in the excellent range and the other half falling in the good range; however, ICC₂s (averaged across raters) all fell in the excellent range. All disagreements on risk estimates involved ratings of low/moderate or moderate/high; none were major disagreements between ratings of low and high risk.

⁵Given that few youth were taking any medications ($n = 12$, 13.3%), we did not include this item in the calculation of the START:AV Strength and Vulnerability total scores for this particular sample. This is unlikely to have changed our results, given that only 2 youth received scores of 2 for strength or vulnerability on this item.

Concurrent Validity with SAVRY and DAP

As shown in Table 3, the START:AV Strength total scores had large positive correlations with scores from the Protective Factors section of the SAVRY and the DAP Asset domain scores. In addition, the START:AV Vulnerability total scores had large positive correlations with the Risk Factors section and total scores on the SAVRY, all of which provides evidence of concurrent validity.

Prevalence Rates of Outcomes

Self-reported adverse events were common during the follow-up. Overall, 79.0% of participants ($n = 49$) reported substance use (being drunk or use of illegal drugs), and 66.7% ($n = 42$) were victimized. Based on self-report, 68.2% of youth ($n = 45$) reported that they had committed a reoffense of any kind during the past 3-months, and 54.5% ($n = 36$) had committed a violent offense. However, official justice records detected much lower rates of reoffending; 28.8% ($n = 21$) of youth were detected for any reoffending, and 19.2% ($n = 14$) for violence (see also Brame et al., 2004). Although lifetime rates of suicide attempts were high (29.0%, $n = 20$), only one youth made a suicide attempt (1.6%) during the follow-up. Seven youth (11.4%) engaged in NSSI only during the follow-up. Given the low rate of suicide attempts during the follow-up, subsequent analyses focused on suicide ideation and NSSI rather than on suicide attempts. We only found one gender difference in dichotomous or continuous outcome variables: a greater proportion of female youth than male youth reported substance use (94.7% vs. 72.1%), $\chi^2(1, n = 62) = 4.08$, Fisher's Exact Test $p = .05$.

Predictive Validity

Prior to testing the predictive validity of the START:AV, we first examined whether gender moderated the predictive validity of total scores or risk estimates through stepwise multiple regression (Baron & Kenny, 1986; Holmbeck, 1997). We did not find significant moderator effects. Thus, in our remaining analyses, we collapsed subsamples of male and female youth.

Vulnerability Total Scores—As shown in Table 4, the Vulnerability total scores significantly predicted each of the outcomes when measured with continuous scales. Vulnerability total scores were also significantly correlated with all dichotomous outcomes except for NSSI (see Table 5).

Strength Total Scores—Strength total scores inversely predicted scores on SRO Total Offenses and Aggressive Offenses scales, with trends towards significance on the PBF Overt Victimization scale (see Table 4). Strength total scores were also inversely correlated with several dichotomous outcomes, including arrests for a violent offense or any offense and use of street drugs (see Table 5), but they were not significantly correlated with NSSI, or the Suicide Ideation and Alcohol/Drug Use scales on the MAYSI-2.

For outcomes in which both Strength and Vulnerability scores were significant (i.e., SRO Total and Aggressive Offenses scale, official records of any and violent offending, street drug use on DAU), we tested incremental validity of Strength total scores by entering Vulnerability total scores in the first step followed by Strength total scores in the second step. Strength total scores did not add incremental validity in these analyses.

Next, we tested whether the relationship between Strength total scores and continuous outcomes were moderated by youths' risk levels. Based on a median split on Vulnerability total scores, youths' risk categories were classified as high risk (i.e., scores above the median score of 23.5) or lower risk (i.e., scores below 23.5). Moderator analyses were conducted using stepwise linear regression for continuous outcome variables and stepwise

logistic regression for dichotomous outcome variables (Baron & Kenny, 1986; Holmbeck, 1997). We entered Strength total scores and Risk category into the equation, followed by the interaction term between the Strength total scores and the Risk category. We did not find any significant moderator effects. Although Strength total scores appeared to have a somewhat stronger buffering effect on SRO Total scores and SRO Aggressive Offenses scale for youth who were high risk ($r = -.19$ and $-.20$ respectively) compared to those who were lower risk ($r = .02$ and $.07$ respectively), this difference did not reach significance.

Risk Estimates—Each risk estimate was correlated with the *relevant* outcome on the continuous scales (see Table 6). For instance, the risk estimate for harm to others significantly predicted the Aggressive Offense Scale on the SRO, and the risk estimate for victimization predicted the Overt Victimization Scale of the PBF. Moreover, as shown in Table 6, the pattern of correlations was generally stronger between the risk estimates for relevant outcomes than broader outcomes, suggesting the specificity of risk estimates. Despite significant correlations, AUCs for some dichotomous outcomes (e.g., violence and offending, as measured by official records) did not reach significance at $p < .05$ (see Table 5).

Given that both risk estimates and Vulnerability total scores were significant predictors for a number of outcomes, we tested whether the risk estimates showed incremental validity over the Vulnerability total scores through linear regression for continuous outcome variables and logistic regression for dichotomous outcome variables. We entered Vulnerability total scores in the first step followed by the risk estimate that was relevant to a particular outcome in the second step. Risk estimates added incremental validity over Vulnerability total scores for predictions of substance abuse on the MAYSI-2 Alcohol/Drug Use scale, $\beta = .39$, $t = 3.27$, $p < .01$, $R^2\Delta = .13$, $F\Delta(1, 64) = 10.67$, $p < .01$; victimization on the PBF Overt Victimization scale, $\beta = .25$, $t = 2.07$, $p = .04$, $R^2\Delta = .05$, $F\Delta(1, 60) = 4.30$, $p = .04$; and suicide ideation on the MAYSI-2 Suicide Ideation scale, $\beta = .48$, $t = 4.60$, $p < .001$, $R^2\Delta = .23$, $F\Delta(1, 64) = 21.18$, $p < .001$. However, risk estimates did not add incremental validity for predictions of offending on the SRO Total scores and Aggressive Offenses scale. On the dichotomous outcomes, they added incremental validity only for predictions of substance use (self-report on DAU), $B = 1.71$, $Wald = 9.77$, $p = .002$, Odds Ratio = 5.54.

Changes in START:AV Scores

The 3-month stability coefficient was high for the Strength and Vulnerability total scores ($r = .87$ and $.77$ respectively, $p < .05$). Based on paired samples t -tests, the mean Vulnerability total score decreased significantly from baseline ($M = 21.47$, $SD = 8.37$) to follow-up ($M = 19.51$, $SD = 8.34$), $t(63) = 2.75$, $p < .01$, $d = 0.46$), whereas the mean Strength total score did not show significant group-level change (baseline: $M = 20.35$, $SD = 7.60$; follow-up: $M = 19.89$, $SD = 7.47$).

Individual-oriented analyses identified higher rates of change. As shown in Table 6, the proportion of youth who showed changes in risk estimates (i.e., either increases or decreases) ranged from 19.0% (for the suicide risk estimate) to 46.2% (for the general reoffending risk estimate) (see Table 7). Overall, 92.1% ($n = 58$) of participants were rated to have changed in at least one risk domain, and approximately half changed in two or more domains (50.8%, $n = 32$).

Based on RCIs, a youth needed to show a change (increase or decrease) of 5.94 points in their Strength total scores or a change of 8.58 points in their Vulnerability total scores for it to be considered a reliable change at 95% CI, and changes of 4.97 and 7.18 points respectively for Strength and Vulnerability total scores when a 90% CI was used. Using these criteria, 15.9% of the sample ($n = 10$) showed reliable change in either Strength or

Vulnerability total scores when a 95% CI was applied; 26.9% ($n = 17$) showed reliable change when a 90% CI was applied (see Table 8).

Discussion

The START:AV (Nicholls et al., 2010; Viljoen et al., full manual in preparation) is an adolescent adaptation of the START (Webster et al., 2004, 2009), a clinical guide designed to assist in assessment and management of short-term risks that individuals in mental health and justice settings often experience. The current study is the first to examine the START:AV's inter-rater reliability, concurrent and predictive validity, and ability to assess short-term change in strengths and vulnerabilities. We employed a short-term prospective design and sampled youth from adolescent offenders on probation.

Assessment of Risks for Multiple Adverse Outcomes

In justice settings, risk assessments and interventions often focus on adolescents' risk of violence towards others. However, the results of this study illustrate that other adverse outcomes are common among justice-involved youth, consistent with other studies (Morris, Harrison, Knox, & Tromanhauser, 1995; Penn, Esposito, Schaeffer, Fritz, & Spirito, 2003). Even over a short 3-month follow-up period, a high proportion of adolescents in our sample experienced victimization (67%), and engaged in alcohol or drug use (80%) and offending (68%). This reinforces the need for a broader perspective on adolescents' risks and needs, and greater cohesiveness in risk management and treatment strategies.

A key question, however, is whether risks for these multiple adverse outcomes can be assessed with a single risk assessment measure, or if instead, separate measures are needed for each outcome? Our results suggest that it may be possible to integrate the assessment of several risk domains in a single tool. As evidence, START:AV risk estimates (made using structured professional judgment) and Vulnerability total scores predicted multiple adverse outcomes examined during the 3-month follow-up, including scores on self-report measures of aggressive offenses, total offenses, suicide ideation, alcohol and drug use, and victimization. In addition, Strength total scores inversely predicted violence, any offending, and street drug use.

While there is considerable overlap in Vulnerability and Strength factors across outcomes (Kooyman, Dean, Harvey, & Walsh, 2007; Webster, Nicholls, Martin, Desmarais, & Brink, 2006; Webster et al., 2009), some differences do exist (Becker & Grilo, 2007; Gray et al., 2003). The START:AV strives to accommodate these differences by utilizing a SPJ model whereby professionals can give emphasis to factors that are especially important for particular outcomes in their final risk estimates for individual cases. In support of this model, we found that risk estimates predicted relevant or matched outcomes more accurately than broader outcomes. For instance, the risk estimate for victimization had stronger associations with the victimization outcome measure than with the measures of the other types of outcomes. As such, START:AV risk estimates may help raters to develop more precise risk estimates than those yielded by total scores alone.

Furthermore, we found that START:AV risk estimates added incremental validity to Vulnerability total scores in the prediction of several outcomes, including substance abuse, victimization, and suicide ideation (see also Desmarais et al., 2012; Gray et al., 2012). On the other hand, risk estimates did not add incremental validity for the prediction of offending or most dichotomous outcomes (with the exception of substance use). Notably, however, the SPJ model asserts that while total scores and SPJ may *generally* have high correlations, SPJ judgments may differ from total scores in instances in which there are unique considerations that lead an assessor to conclude that risk is high despite a low score, or vice versa (Webster

& Hucker, 2007). Thus, future research on this issue should examine individual instances of disagreements rather than concentrating solely on group-level analyses.

While the START:AV assessments predicted numerous outcomes (including suicide ideation, violent and general offending, victimization, and substance abuse), total scores and risk estimates did not predict NSSI. This may, in part, stem from the limited power and the low base rate of NSSI; although the AUC between Vulnerability total scores and NSSI was .70 ($p = .08$), this did not reach significance at $p < .05$. Also, we relied on several questions pertaining to NSSI and suicidal behavior; however, several validated measures for NSSI have since been developed (Muehlenkamp, Cowles, & Gutierrez, 2010; Nock, Holmberg, Photos, & Michel, 2007). Furthermore, given that NSSI is distinct from suicide with respect to its functions and some of its correlates (Brausch & Gutierrez, 2010; Nock, Joiner, Gordon, Lloyd-Richardson, & Prinstein, 2006), it may be that raters were less attuned to NSSI as a possible outcome.

With respect to inter-rater reliability, our results are promising. The inter-rater reliability of the START:AV assessments fell in the excellent range for Strength and Vulnerability total scores. Although the inter-rater reliability of risk estimates appeared somewhat lower than that of total scores when measured with ICCs, this may be partly attributable to the limited range of risk estimate ratings (i.e., low, moderate, high) and the fact that the risk estimate is a single item, whereas the total score is comprised of multiple items (Vincent, Guy, Fusco, & Gershenson, 2012). Notably, all discrepancies in risk estimates involved differences between low/moderate and moderate/high rather than major differences (i.e., between ratings of low and high risk). Given that our training process was fairly extensive and involved at least four START:AV practice cases, future research is needed to determine if these findings will generalize to briefer trainings and field assessments. We found that inter-rater reliability improved throughout the training, and that the inclusion of a real-world practice case (in which a trainee was paired with an experienced rater) was an important addition to vignette-based training cases.

Assessment of Strengths with the START:AV

The START:AV places equal emphasis on strength factors in addition to vulnerability factors through its inclusion of 23 items which are rated separately for both strengths and vulnerabilities. The purpose of the START:AV Strength items is multifold; these items are intended to help refine predictions of adverse outcomes, aid in treatment-planning, and provide a more balanced description of youth. This study found several sources of support for strength items in the START:AV. First, we found moderate correlations between the START:AV Strength total scores and the DAP assets scores, and large correlations between the START:AV Strength total scores and SAVRY Protective Factors section score, providing evidence for concurrent validity.

Second, consistent with its aim of yielding a balanced and strengths-informed description of individual adolescents, the START:AV identified each adolescent as having some strengths. In fact, all adolescents in our sample had five or more areas of moderate or maximal strengths and over 50% had 15 or more strengths. Desmarais and colleagues (this issue) found similarly high rates of strengths in their field study of START:AV assessments. In comparison, several studies with adolescent offenders have found that approximately 40–50% of adolescents were not identified as having any of the six protective factors on the SAVRY (Rennie & Dolan, 2010; Penney, Lee, & Moretti, 2010; Viljoen et al., 2012; see also Lodewijks, de Ruiter, & Doreleijers, 2010). Thus, even though the START:AV is highly correlated with the SAVRY, the START:AV may identify a greater number of strengths as a result of its greater number of strength items and its 3-point scale.

Third, Strength total scores achieved significant predictive validity (inversely) for predictions of violence, any offending, and street drug use, and approached significance for victimization. However, whereas Vulnerability total scores performed well across outcomes, Strength total scores did not predict some outcomes, such as suicidal ideation. Although further research is needed, it could be that unique strength factors exist for specific outcomes.

Consistent with research on the adult START (Webster et al., 2004; 2009), the correlations between Strength and Vulnerability total scores were high. Because START:AV assessments involve evaluating the same items separately for both strengths and vulnerabilities, this finding is not surprising. However, high correlations have also been reported for other approaches that use separate items for strengths and vulnerabilities. For instance, studies have found the correlation between the HCR-20 (Webster et al., 1997) and the SAPROF (de Vogel, de Ruiter, Bouman, & de Vries Robbé, 2009) to range from $-.69$ to $-.78$ (Abidin, 2012; de Vries Robbé, de Vogel, & de Spa, 2011), and the correlation between the SAVRY Protective Factor section and the SAVRY Risk Total to be $-.69$ (Spice, Viljoen, Gretton, & Roesch, 2010). In contrast to these findings, Desmarais and colleagues (this issue) found much lower correlations between Strength and Vulnerability total scores on the START:AV ($r = -.22$), suggesting a need for further knowledge regarding the distinction between strength and vulnerability factors on the START:AV and other measures.

Overall, Strength total scores were less robust predictors of outcomes than final risk estimates and Vulnerability total scores. Research on other instruments, such as the adult START and SAVRY, has similarly found mixed findings regarding the incremental validity of strengths and protective factors over risk factors (Desmarais et al., 2012; Lodewijkset al., 2010; Viljoen et al., 2011; Wilson et al., 2010). This nomothetic finding, however, should not in any way diminish the focus on continued work on strengths. Prediction of outcomes is only one potential reason to assess strength factors. Consistent with practice guidelines, strength factors may help provide a more balanced description of patients and clients (e.g., American Psychological Association, 2006; 2007; American Psychological Association Task Force on Evidence-Based Practice for Children and Adolescents, 2008; Royal College of Psychiatrists, 2008). Also, they may potentially aid in treatment-planning, help improve therapeutic alliance, and combat therapeutic pessimism in service providers (de Ruiter & Nicholls, 2011).

Furthermore, the relationship between strengths and outcomes may be complex. For instance, previous research has indicated that the relationship between Strength total scores and outcomes may differ depending on youths' risk levels (Lodewijks et al., 2010), consistent with a buffering model (Garmezy, Masten, & Tellegen, 1984; Fergus & Zimmerman, 2005). In the present study, the interaction between risk level and strengths did not reach significance, possibly due to limited power. Thus, future research should continue to test various models of strengths in larger samples (Fergus & Zimmerman, 2005), and examine whether relationships between strengths and outcomes differs based on gender, age, follow-up length, outcome, or other variables. Future research should also examine the role of strengths (on the START:AV and other tools) in predicting *positive* outcomes, such as school achievement and quality of life, and investigate the relevance of strengths to treatment, such as whether the implementation of a measure such as the START:AV leads to increased attention to strengths in treatment plans.

Examination of Dynamic Items and Change Scores on START:AV Assessments

Given that the START:AV focuses on dynamic factors, we examined changes in START:AV scores over a 3-month follow-up period. It is important to note that this is not a treatment study per se. While interventions are a fairly common component of regular youth

justice services, the type, intensity, and recency of treatments varied considerably. In addition, our follow-up period was short. Thus, we expected moderate rather than high levels of change in such a short period. Also, we predicted that while some youth might show improvements, others may experience deteriorations (i.e., increased Vulnerability total scores).

We utilized several data analytic methods to examine change, including several group-based methods to examine overall patterns of change (i.e., stability coefficients) as well as individual-based methods (e.g., Jacobsen & Truax RCI). These methods yielded somewhat different results. In particular, stability coefficients for strengths and vulnerability totals across the two time periods were fairly high ($> .75$). In contrast, individually-based analyses revealed a higher degree of change, as it was possible to differentiate patterns of increase and decrease. According to risk estimates, almost all participants (92%) showed increased or decreased risk during the 3-month follow-up in at least one of the five risk domains examined in this study. Thus, assessors perceived almost all youths' risks to have changed to some extent, potentially providing evidence of the importance of dynamic assessments.

A significant limitation of examining changes in risk estimates (or total scores) is that many methods fail to take into account measurement error; it is possible that a difference between ratings of low and moderate, for instance, could simply reflect imperfect inter-rater reliability rather than true changes in risk levels. In addition, raters may have been oriented to look for change. As such, we also calculated RCIs, which take into account measurement error (in this case, inter-rater reliability). Using this method, a moderate but fairly modest number of youth were judged to show reliable increases or decreases in Strength or Vulnerability total scores (16–27% using 95% and 90% CIs, respectively).

Several conclusions can be drawn from these results. First, these findings emphasize that it will be important to carefully consider data analytic methods and the impact they may have on conclusions about the dynamic nature of risk and the capacity of tools to assess change. Second, and most central to the current study, our findings suggest that the START:AV is able to tap into some changes in strength and vulnerability levels, and risk estimates. Even though our follow-up period was short and participants were not selected from a treatment program, a sizable proportion of adolescents showed reliable change over a 3-month follow-up. In the context of effective treatments, we would expect to see higher levels of change.

Although this study provides preliminary support for the START:AV's emphasis on regular reassessments (Viljoen et al., in preparation; see also Webster et al., 2009), future research is needed to refine our understanding and interpretation of change on the START:AV (and other risk assessment measures). Given that the optimal time frame for reassessment remains unclear and may vary by item and context, change should be studied across diverse samples, settings, and follow-up lengths. It would be particularly fruitful to embed the START:AV into studies examining various treatment approaches; this could, for instance, allow for comparisons of rates and patterns of change for youth receiving empirically-supported interventions compared to usual juvenile justice services, and help identify predictors of change. To measure change, we recommend that researchers examine both individual patterns of change as well as group-level changes in scores, and carefully attend to measurement error. The Jacobsen and Truax RCI appears to be one potentially useful approach to explore, but like any approach it has limitations. It enables an interpretation of simply whether a reliable decrease or increase has occurred but does not comment on the size of change, and it does not take into account regression to the mean (Hsu, 1989). Thus, other approaches should also be examined (Bauer et al., 2004; Duncan, Duncan, & Strycker, 2006; Wise, 2004).

Limitations and Future Directions

This study has several methodological strengths; it is prospective, measured outcomes through multiple sources (i.e., official records and self-reports, including both continuous and dichotomous approaches), and utilized both interview data and file information to conduct START:AV assessments. At the same time, it has a number of limitations. First, a proportion of youth in the study missed their scheduled 3-month interviews. There were no significant differences in baseline START:AV scores or demographic characteristics of youth who missed interviews and those who participated, and we obtained official justice records for most of the youth and therefore were still able to examine outcomes for youth who did not complete in-person interviews. However, this is nevertheless a limitation of prospective designs. Second, our sample size of female youth were small ($n = 28$). Although analyses indicated that gender did not moderate the relationship between START:AV assessments and outcomes, this issue merits ongoing attention (see also Desmarais et al., this issue). Third, the present study used the START:AV abbreviated manual (Nicholls et al., 2010) rather than the full START:AV manual (Viljoen et al., in preparation). It is possible that access to the full manual, which includes a research review and more detailed coding instructions, may impact findings (e.g., improved inter-rater reliability for risk estimates). Finally, given that START:AV ratings were completed by trained RAs rather than by professionals in a clinical setting, real-world implementation studies are essential (see Desmarais et al., this issue).

Limitations notwithstanding, this study provides some initial support for the inter-rater reliability, concurrent validity, and prospective validity of START:AV assessments. START:AV final risk estimates and Vulnerability total scores were able to predict not only risks for violence towards others but also broader adverse outcomes that are important in justice settings, such as victimization and suicide ideation, and Strength total scores significantly predicted lower rates of violence, general offending, and serious drug use. START:AV assessments identified a range of strength factors in all of the adolescents in our sample, suggesting that it provides a strength-oriented approach that may facilitate balanced descriptions of adolescents' strengths in addition to their vulnerabilities. In addition, START:AV assessments appear to have utility in capturing change in dynamic factors and risk. Although this study is a promising first step in the validation of the START:AV, there is now a strong need for further research. We hope that future studies will lead to a more refined understanding of the dynamic nature of START:AV assessments, and evaluate not only the role of START:AV strength factors in predicting outcomes, but also their relevance to treatment and risk management efforts.

Acknowledgments

The first author's work on this project was supported by a Michael Smith Foundation for Health Research Career Investigator Award, and a grant from the British Columbia Mental Health and Addictions Research Network (held jointly with the sixth author). The fifth author's work on this project was supported by the National Institute on Drug Abuse (P30DA028807, PI: Roger H. Peters). The sixth author's work is supported by a Michael Smith Foundation for Health Research Career Investigator Award and a Canadian Institutes of Health Research New Investigator Award. The content is solely the responsibility of the authors and does not necessarily represent the official views of the funding agencies.

References

- Abidin, Z. Prospective Validation Cohort Study of structured professional judgement instruments START and SAPROF in forensic mental health services. College of Psychiatry of Ireland Winter Conference; Killarney, Ireland. 2011 Nov.
- American Psychological Association . Evidence-based practice in psychology. *American Psychologist*. 2006; 61(4):271–285.10.1037/0003-066X.61.4.271 [PubMed: 16719673]

- American Psychological Association . Guidelines for Psychological Practice With Girls and Women. American Psychologist. 2007; 62(9):949–979.10.1037/0003-066X.62.9.949 [PubMed: 18085843]
- American Psychological Association Task Force on Evidence-Based Practice for Children and Adolescents. Disseminating evidence-based practice for children and adolescents: A systems approach to enhancing care. Washington, DC: American Psychological Association; 2008.
- Archer RP, Simonds-Bisbee EC, Spiegel DR, Handel RW, Elkins DE. Validity of the Massachusetts Youth Screening Instrument-2 (MAYSI-2) scales in juvenile justice settings. Journal of Personality Assessment. 2010; 92(4):337–348.10.1080/00223891.2010.482009 [PubMed: 20552508]
- Archer RP, Stredny RV, Mason JA, Arnau RC. An examination and replication of the psychometric properties of the Massachusetts Youth Screening Instrument-Second Edition (MAYSI-2) among adolescents in detention settings. Assessment. 2004; 11:290–302.10.1177/1073191104269863 [PubMed: 15486166]
- Baron RM, Kenny DA. The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. Journal Of Personality And Social Psychology. 1986; 51(6):1173–1182.10.1037/0022-3514.51.6.1173 [PubMed: 3806354]
- Bauer S, Lambert MJ, Nielsen S. Clinical significance methods: A comparison of statistical techniques. Journal of Personality Assessment. 2004; 82(1):60–70.10.1207/s15327752jpa8201_11 [PubMed: 14979835]
- Becker DF, Grilo CM. Prediction of suicidality and violence in hospitalized adolescents: Comparisons by sex. The Canadian Journal of Psychiatry/La Revue Canadienne De Psychiatrie. 2007; 52(9):572–580. Retrieved from <http://cat.inist.fr/?aModele=afficheN&cpsidt=19138858>.
- Borum R. Managing At-Risk Juvenile Offenders in the Community: Putting Evidence-Based Principles Into Practice. Journal of Contemporary Criminal Justice. 2003; 19(1):114–137.10.1177/1043986202239745
- Borum, R.; Bartel, P.; Forth, A. Manual for the Structured Assessment of Violence Risk in Youth(SAVRY). Odessa, FL: Psychological Assessment Resources; 2006.
- Blair R, Lawson SB. Another look at the robustness of the product-moment correlation coefficient to population non-normality. Florida Journal of Educational Research. 1982; 24:11–15. Retrieved from <http://psycnet.apa.org/psycinfo/1983-20165-001>.
- Bosworth, K.; Espelage, D. Teen Conflict Survey. Bloomington, IN: Center for Adolescent Studies, Indiana University; 1995. (Unpublished)
- Braithwaite E, Charette Y, Crocker AG, Reyes A. The predictive validity of clinical ratings of the Short-Term Assessment of Risk and Treatability (START). The International Journal of Forensic Mental Health. 2010; 9(4):271–281.10.1080/14999013.2010.534378
- Brame R, Fagan J, Piquero AR, Schubert CA, Steinberg L. Criminal Careers of Serious Delinquents in Two Cities. Youth Violence And Juvenile Justice. 2004; 2(3):256–272.10.1177/1541204004265877 [PubMed: 20119514]
- Brausch AM, Gutierrez PM. Differences in Non-Suicidal Self-Injury and Suicide Attempts in Adolescents. Journal of Youth And Adolescence. 2010; 39(3):233–242.10.1007/s10964-009-9482-0 [PubMed: 19941045]
- Carolina Population Center. National Longitudinal Study of Adolescent Health Wave II Adolescent In-Home Questionnaire Code Book. Carolina Population Center, University of North Carolina; Chapel Hill: 1999. Available at: <http://www.cpc.unc.edu/projects/addhealth/codebooks/wave2>
- Calverley, D.; Cotter, A.; Halla, E. Juristat. Vol. 30. Ottawa: Statistics Canada, Canadian Centre for Justice Statistics; 2010. Youth custody and community services in Canada, 2008/2009. Catalogue no. 85-002-XRetrieved from <http://www.statcan.gc.ca/pub/85-002-x/2010001/article/11147-eng.pdf>
- Chu C, Thomas SM, Ogloff JP, Daffern M. The predictive validity of the Short-Term Assessment of Risk and Treatability (START) in a secure forensic hospital: Risk factors and strengths. The International Journal of Forensic Mental Health. 2011; 10(4):337–345.10.1080/14999013.2011.629715
- Cicchetti DV, Sparrow SA. Developing criteria for establishing inter-rater reliability of specific items: Applications to assessment of adaptive behavior. American Journal of Mental Deficiency. 1981;

86(2):127–137. Retrieved from <http://psycnet.apa.org/psycinfo/1982-00095-001>. [PubMed: 7315877]

- Cohen, J. Statistical power analysis for the behavioral sciences. 2. New Jersey: Lawrence Erlbaum; 1988.
- Costa PT, McCrae RR. Personality in adulthood: A six-year longitudinal study of self-reports and spouse ratings on the NEO Personality Inventory. *Journal of Personality And Social Psychology*. 1988; 54(5):853–863.10.1037/0022-3514.54.5.853 [PubMed: 3379583]
- Desmarais SL, Nicholls TL, Wilson CM, Brink J. Using dynamic risk and protective factors to predict inpatient aggression: Reliability and validity of START assessments. *Psychological Assessment*. 201210.1037/a0026668
- Desmarais SL, Sellers BG, Viljoen JL, Cruise KR, Nicholls TL, Dvoskin J. Pilot implementation and preliminary evaluation of START:AV assessments in secure juvenile correctional facilities. *International Journal of Forensic Mental Health*. this issue
- de Ruiter C, Nicholls TL. Protective factors in forensic mental health: A new frontier. *The International Journal of Forensic Mental Health*. 2011; 10(3):160–170.10.1080/14999013.2011.600602
- de Vogel, V.; de Ruiter, C.; Bouman, Y.; de Vries Robbé, M. SAPROF. Guidelines for the assessment of protective factors for violence risk. English Version. Utrecht, The Netherlands: Forum Educatief; 2009.
- de Vries Robbé M, de Vogel V, de Spa E. Protective factors for violence risk in forensic psychiatric patients: A retrospective validation study of the SAPROF. *The International Journal of Forensic Mental Health*. 2011; 10(3):178–186.10.1080/14999013.2011.600232
- Donovan JE, Jessor R. Structure of problem behavior in adolescence and young adulthood. *Journal Of Consulting And Clinical Psychology*. 1985; 53(6):890–904.10.1037/0022-006X.53.6.890 [PubMed: 4086689]
- Douglas KS, Skeem JL. Violence Risk Assessment: Getting Specific About Being Dynamic. *Psychology, Public Policy, And Law*. 2005; 11(3):347–383.10.1037/1076-8971.11.3.347
- Douglas, KS.; Skeem, JL.; Nicholson, E. Research methods in violence risk assessment. In: Rosenfeld, B.; Penrod, SD.; Rosenfeld, B.; Penrod, SD., editors. *Research methods in forensic psychology*. Hoboken, NJ US: John Wiley & Sons; 2011. p. 325-346.
- Doyle M, Lewis G, Brisbane M. Implementing the Short-term Assessment of Risk and Treatability (START) in a forensic mental health service. *Psychiatric Bulletin*. 2008; 32(11):406–408.10.1192/pb.bp.108.019794
- Duncan, TE.; Duncan, SC.; Strycker, LA. An introduction to latent variable growth curve modeling: Concepts, issues, and applications. 2. Mahwah, NJ US: Lawrence Erlbaum Associates Publishers; 2006.
- George, D.; Mallery, P. SPSS for Windows step by step: A simple guide and reference. 11.0 update. 4. Boston: Allyn & Bacon; 2003.
- Gray NS, Benson R, Craig R, Davies H, Fitzgerald S, Huckle P, Snowden RJ. The Short-Term Assessment of Risk and Treatability (START): A prospective study of inpatient behavior. *The International Journal Of Forensic Mental Health*. 2011; 10(4):305–313.10.1080/14999013.2011.631692
- Gray NS, Hill C, McGleish A, Timmons D, MacCulloch MJ, Snowden RJ. Prediction of violence and self-harm in mentally disordered offenders: A prospective study of the efficacy of HCR-20, PCL-R, and psychiatric symptomatology. *Journal of Consulting And Clinical Psychology*. 2003; 71(3): 443–451.10.1037/0022-006X.71.3.443 [PubMed: 12795569]
- Grisso, T. Forensic evaluation of juveniles. Sarasota, FL US: Professional Resource Press/Professional Resource Exchange; 1998.
- Grisso, T.; Barnum, R. Massachusetts Youth Screening Instrument-Version User's manual and technical report. Sarasota, FL: Professional Resource Press; 2003.
- Grisso T, Barnum R, Fletcher KE, Cauffman E, Peuschold D. Massachusetts Youth Screening Instrument for mental health needs of juvenile justice youths. *Journal of the American Academy of Child and Adolescent Psychiatry*. 2001; 40:541–548.10.1097/00004583-200105000-00013 [PubMed: 11349698]

- Huizinga D, Esbensen F, Weiher A. Are there multiple pathways to delinquency? Denver youth survey. *Journal of Criminal Law and Criminology*. 1991; 82:83–118. Retrieved from http://heinonline.org/HOL/Page?handle=hein.journals/jclcr82&div=13&g_sent=1&collection=journals.
- Hoge, R.; Andrews, DA.; Leschied, A. Youth Level of Service/Case Management Inventory 2.0: YLS/CMI 2.0 Manual. Toronto, Canada: Multi-Health Systems; 2011.
- Holmbeck GN. Toward terminological, conceptual, and statistical clarity in the study of mediators and moderators: Examples from the child-clinical and pediatric psychology literatures. *Journal of Consulting and Clinical Psychology*. 1997; 65(4):599–610.10.1037/0022-006X.65.4.599 [PubMed: 9256561]
- Hsu LM. Reliable changes in psychotherapy: Taking into account regression toward the mean. *Behavioral Assessment*. 1989; 11(4):459–467. Retrieved from <http://web.ebscohost.com>.
- Jacobson NS, Truax P. Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology*. 1991; 59(1):12–19.10.1037/0022-006X.59.1.12 [PubMed: 2002127]
- Kirk DS. Examining the Divergence Across Self-report and Official Data Sources on Inferences About the Adolescent Life-course of Crime. *Journal of Quantitative Criminology*. 2006; 22(2):107–129.10.1007/s10940-006-9004-0
- Knight GP, Little M, Losya SH, Mulvey EP. The self-report of offending among serious juvenile offenders: Cross-gender, cross-ethnic/race measurement equivalence. *Youth Violence and Juvenile Justice*. 2004; 2:273–295.10.1177/1541204004265878 [PubMed: 20119516]
- Kooyman I, Dean K, Harvey S, Walsh E. Outcomes of public concern in schizophrenia. *The British Journal of Psychiatry*. 2007; (Supplement 50):29–36.10.1192/bjp.191.50.s29
- Kowalski CJ. On the Effects of Non-normality on the Distribution of the Sample Product-moment Correlation Coefficient. *Journal of The Royal Statistical Society: Series C (Applied Statistics)*. 1972; 21(1):1. Retrieved from <http://www.jstor.org/stable/pdfplus/2346598.pdf?acceptTC=true>.
- Kroppan E, Nasset MB, Nonstad K, Pedersen TW, Almvik R, Palmstierna T. Implementation of the Short Term Assessment of Risk and Treatability (START) in a forensic high secure unit. *The International Journal of Forensic Mental Health*. 2011; 10(1):7–12.10.1080/14999013.2011.552368
- Lodewijks HB, de Ruiter C, Doreleijers TH. The impact of protective factors in desistance from violent reoffending: A study in three samples of adolescent offenders. *Journal of Interpersonal Violence*. 2010; 25(3):568–587.10.1177/0886260509334403 [PubMed: 19584407]
- Lunnen KM, Ogles BM. A multiperspective, multivariable evaluation of reliable change. *Journal of Consulting And Clinical Psychology*. 1998; 66(2):400–410.10.1037/0022-006X.66.2.400 [PubMed: 9583343]
- Maasen GH. The unreliable change of reliable change indices. *Behaviour Research And Therapy*. 2001; 39(4):495–498.10.1016/S0005-7967(00)00029-2 [PubMed: 11280346]
- MacCallum RC, Zhang S, Preacher KJ, Rucker DD. On the practice of dichotomization of quantitative variables. *Psychological Methods*. 2002; 7(1):19–40.10.1037/1082-989X.7.1.19 [PubMed: 11928888]
- Marsden J, Eastwood B, Wright C, Bradbury C, Knight J, Hammond P. How best to measure change in evaluations of treatment for substance use disorder. *Addiction*. 2011; 106(2):294–302.10.1111/j.1360-0443.2010.03143.x [PubMed: 20925686]
- McGraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients. *Psychological Methods*. 1996; 1(1):30–46.10.1037/1082-989X.1.1.30
- Morris RE, Harrison EA, Knox GW, Tromanhauser E. Health risk behavioral survey from 39 juvenile correctional facilities in the United States. *Journal of Adolescent Health*. 1995; 17(6):334–344.10.1016/1054-139X(95)00098-D [PubMed: 8924439]
- Mossman D. Further comments on portraying the accuracy of violence predictions. *Law And Human Behavior*. 1994; 18(5):587–593.10.1007/BF01499177
- Muehlenkamp JJ, Cowles ML, Gutierrez PM. Validity of the Self-Harm Behavior Questionnaire with diverse adolescents. *Journal of Psychopathology And Behavioral Assessment*. 2010; 32(2):236–245.10.1007/s10862-009-9131-7

- Multisite Violence Prevention Project. Description of measures: cohort-wide student survey, 2004a. Centers for Disease Control and Prevention, National Center for Injury Prevention and Control; Atlanta, GA: 2004. (Unpublished)
- Nicholls TL, Brink J, Desmarais SL, Webster CD, Martin M. The Short-Term Assessment of Risk and Treatability (START): A Prospective Validation Study in a Forensic Psychiatric Sample. *Assessment*. 2006; 13(3):313–327.10.1177/1073191106290559 [PubMed: 16880282]
- Nicholls, TL.; Viljoen, JL.; Cruise, KR.; Desmarais, SL.; Webster, CD. Short-Term Assessment of Risk and Treatability: Adolescent Version (START:AV) (Abbreviated Manual). Coquitlam, Canada: BC Mental Health and Addiction Services; 2010.
- Nock MK, Holmberg EB, Photos VI, Michel BD. Self-Injurious Thoughts and Behaviors Interview: Development, reliability, and validity in an adolescent sample. *Psychological Assessment*. 2007; 19(3):309–317.10.1037/1040-3590.19.3.309 [PubMed: 17845122]
- Nock M, Joiner T, Gordon K, Lloyd-Richardson E, Prinstein M. Non-suicidal self-injury among adolescents: diagnostic correlates and relation to suicide attempts. *Psychiatry Research*. 2006; 144(1):65–72.10.1016/j.psychres.2006.05.010 [PubMed: 16887199]
- Nonstad K, Nasset MB, Kroppan E, Pedersen TW, Nøttestad Ja, Almvik R, Palmstierna T. Predictive validity and other psychometric properties of the Short-Term Assessment of Risk and Treatability (START) in a Norwegian high secure hospital. *The International Journal of Forensic Mental Health*. 2010; 9(4):294–299.10.1080/14999013.2010.534958
- Nunnally, JC. *Psychometric theory*. New York, NY US: McGraw-Hill; 1967.
- Ogles BM, Lunnen KM, Bonesteel K. Clinical significance: History, application, and current practice. *Clinical Psychology Review*. 2001; 21(3):421–446.10.1016/S0272-7358(99)00058-6 [PubMed: 11288608]
- Olver ME, Stockdale KC, Wormith J. Risk assessment with young offenders: A meta-analysis of three assessment measures. *Criminal Justice and Behavior*. 2009; 36(4):329–353.10.1177/0093854809331457
- Osbourne JW. Notes on the Use of Data Transformation. *Practical Assessment, Research & Evaluation*. 2002; 8(6) Retrieved from <http://pareonline.net/getvn.asp?v=8&n=6>.
- Penn JV, Esposito CL, Schaeffer LE, Fritz GK, Spirito A. Suicide attempts and self-mutilative behavior in a juvenile correctional facility. *Journal of the American Academy Of Child & Adolescent Psychiatry*. 2003; 42(7):762–769.10.1097/01.CHI.0000046869.56865.46 [PubMed: 12819435]
- Penney SR, Lee Z, Moretti MM. Gender differences in risk factors for violence: An examination of the predictive validity of the Structured Assessment of Violence Risk in Youth. *Aggressive Behavior*. 2010; 36(6):390–404.10.1002/ab.20352 [PubMed: 20623508]
- Prentky, R.; Righthand, S. *Juvenile Sex Offender Assessment Protocol II (J-SOAP-II) manual*. Washington, DC: U.S. Department of Justice, Office of Justice Programs, Office of Juvenile Justice and Delinquency Prevention; 2003. Retrieved from <https://www.ncjrs.gov/pdffiles1/ojjdp/202316.pdf>
- Rennie CE, Dolan MC. The significance of protective factors in the assessment of risk. *Criminal Behaviour and Mental Health*. 2010; 20(1):8–22.10.1002/cbm.750 [PubMed: 20104473]
- Royal College of Psychiatrists. Rethinking risk to others in mental health services: Final report of a scoping group. London, UK: Royal College of Psychiatrists; 2008. College report CR150
- Search Institute. Developmental assets profile. Preliminary user manual. Minneapolis, MN: Search Institute; 2004.
- Spice A, Viljoen JL, Gretton HM, Roesch R. Psychological assessment for adult sentencing of juvenile offenders: An evaluation of the RSTI and the SAVRY. *The International Journal of Forensic Mental Health*. 2010; 9(2):124–137.10.1080/14999013.2010.501846
- Viljoen S, Nicholls T, Greaves C, de Ruiter C, Brink J. Resilience and successful community reintegration among female forensic psychiatric patients: A preliminary investigation. *Behavioral Sciences & the Law*. 2011; 29(5):752–770.10.1002/bsl.1001 [PubMed: 21796675]
- Viljoen, J.; Beneteau, J.; Brodersen, E.; Gulbrandsen, E.; Cruise, K.; Nicholls, T.; Desmarais, S.; Douglas, K. Assessing Protective Factors in Adolescent Offenders: A Comparison of the SAVRY,

YLS/CMI, and START:AV. Paper presented at the International Association of Forensic Mental Health Services Conference; Miami, FL. 2012 Apr.

Viljoen JL, Cruise KR, Nicholls TL, Desmarais S, Webster CD. Taking stock and taking steps: The case for an adolescent version of the Short-Assessment of Risk and Treatability. 2012 Manuscript under review.

Vincent GM, Guy LS, Fusco SL, Gershenson BG. Field reliability of the SAVRY with juvenile probation officers: Implications for training. *Law And Human Behavior*. 2012; 36(3):225–236.10.1037/h0093974 [PubMed: 22667812]

Webster, CD.; Douglas, KS.; Eaves, D.; Hart, SD. HCR-20. Assessing the risk of violence. Version 2. Burnaby, Canada: Simon Fraser University and Forensic Psychiatric Services Commission of British Columbia; 1997.

Webster, CD.; Hucker, SJ. Violence risk: Assessment and management. New York, NY US: John Wiley & Sons Ltd; 2007.

Webster, CD.; Martin, M-L.; Brink, J.; Nicholls, TL.; Desmarais, SL. Manual for the Short-Term Assessment of Risk and Treatability (START) (Version 1.1). Port Coquitlam, BC, Canada: British Columbia Mental Health and Addiction Services; 2009.

Webster CD, Nicholls TL, Martin M, Desmarais SL, Brink J. Short-Term Assessment of Risk and Treatability (START): The Case for a New Structured Professional Judgment Scheme. *Behavioral Sciences & the Law*. 2006; 24(6):747–766.10.1002/bsl.737 [PubMed: 17171764]

Webster, CD.; Martin, ML.; Brink, J.; Nicholls, TL.; Middleton, C. Manual for the Short-Term Assessment of Risk and Treatability (START), Version 1.0 (consultation ed.). St. Joseph's Healthcare Hamilton; Ontario: Canada–Forensic Psychiatric Services Commission; Port Coquitlam, British Columbia, Canada: 2004.

Wilson CM, Desmarais SL, Nicholls TL, Brink J. The role of client strengths in assessments of violence risk using the Short-Term Assessment of Risk and Treatability (START). *The International Journal of Forensic Mental Health*. 2010; 9(4):282–293.10.1080/14999013.2010.534694

Wise E. Methods for analyzing psychotherapy outcomes: a review of clinical significance, reliable change, and recommendations for future directions. *Journal of Personality Assessment*. 2004; 82(1):50–59.10.1207/s15327752jpa8201_10 [PubMed: 14979834]

Table 1

Means and Intercorrelations Between START:AV Strength and Vulnerability Items

START:AV Items	Strengths <i>M</i> (<i>SD</i>)	Vulnerabilities <i>M</i> (<i>SD</i>)	Correlation Strength & Vulnerability totals scores <i>r</i>	Key Strengths %	Critical Vulnerabilities %
1. Social skills	1.28 (.65)	0.57 (.64)	-.63**	20.0	4.4
2a. Relationships with caretakers and other adults	.87 (.69)	1.00 (.72)	-.54**	4.4	15.6
2b. Relationships with peers	1.06 (.62)	0.98 (.72)	-.40**	4.4	14.4
3. School/work	0.81 (.72)	1.08 (.72)	-.60**	18.9	13.3
4. Recreational	0.76 (.71)	1.13 (.69)	-.62**	16.7	12.2
5. Self-care	1.49 (.59)	0.59 (.58)	-.50**	8.9	3.3
6. Mental state	1.33 (.60)	0.39 (.49)	-.48**	0.0	5.6
7. Emotional state	1.00 (.70)	1.08 (.69)	-.37**	2.2	16.7
8. Substance use	0.69 (.79)	1.19 (.72)	-.65**	10.0	41.1
9. Impulse control	0.40 (.51)	1.26 (.57)	-.47**	0.0	12.2
10. External triggers	0.58 (.65)	1.22 (.67)	-.32**	3.3	17.8
11a. Social support from caretakers and other adults	1.10 (.60)	1.00 (.75)	-.47**	8.9	13.3
11b. Social support from peers	0.67 (.65)	1.46 (.60)	-.44**	3.4	24.4
12. Material resources	1.22 (.68)	0.57 (.69)	-.63**	5.6	5.6
13. Attitudes	0.73 (.67)	1.18 (.66)	-.43**	4.4	15.6
14. Medication	0.83 (.72)	0.83 (.72)	-.59*	0.0	0.0
15. Rule adherence	0.83 (.72)	1.07 (.78)	-.76**	6.7	8.9
16. Conduct	0.71 (.74)	1.18 (.79)	-.67**	3.3	13.3
17. Insight	0.59 (.60)	1.18 (.66)	-.35**	7.8	10.0
18. Plans	0.98 (.65)	0.92 (.67)	-.54**	16.7	2.2
19. Coping	0.50 (.59)	1.16 (.62)	-.44**	11.1	15.6
20. Treatability	0.87 (.67)	0.86 (.70)	-.61**	11.1	6.7
21. Parenting & home environment	0.86 (.70)	1.08 (.75)	-.60**	3.3	21.1
Total Score	19.27 (8.10)	22.12 (8.24)	-.74**	-	-

Note: Medication was eliminated from the total score, as most youth were not taking medication.

Table 2

Inter-rater Reliability of START:AV Assessments

	<u>ICCs</u>		Inter-rater Agreement
	ICC ₁	ICC ₂	
Total Scores			
Strengths Total	.92	.96	-
Vulnerabilities Total	.86	.93	-
Risk Estimates			
Violence Towards Others	.60	.75	58.3%
General Reoffending	.75	.86	88.9%
Self-Harm	.88	.94	91.7%
Suicide	.52	.69	83.3%
Substance Abuse	.60	.75	66.7%
Victimization	.86	.93	91.7%

Table 3

Concurrent Validity of START:AV Assessments with SAVRY and DAP Assessments

	Strength Total Scores	Vulnerability Total Scores
SAVRY		
Risk Total Score	-.72 **	.83 **
Structured Professional Judgment	-.59 **	.67 **
Protective Factors Score	.80 **	-.65 **
DAP		
Personal Assets Score	.31 **	-.36 **
Social Assets Score	.45 **	-.47 **
Family Assets Score	.39 **	-.43 **
School Assets Score	.31 **	-.28 **
Community Assets Score	.38 **	-.34 **

Note:

**
 $p < .01$.

Table 4

Predictive Validity: START:AV Totals and Outcomes on Continuous Self-Report Scales

Outcomes	Correlations	
	Strength Total Scores	Vulnerability Total Scores
Aggressive Offenses Scale on SRO	-.31 *	.54 **
Total Offenses Scale on SRO	-.30 *	.51 **
Suicide Ideation Scale on MAYSI-2	-.01	.25 *
Alcohol/Drug Use Scale on MAYSI-2	-.15	.35 **
Overt Victimization Scale of PBF	-.22 †	.52 **

Note:

† $p < .10$,* $p < .05$,** $p < .01$,*** $p < .001$.

Table 5

Predictive Validity: START:AV Assessments and Dichotomous Outcomes

	Strength Total Score			Vulnerability Total Score			Final Risk Estimate		
	r	AUC	SE of AUC	r	AUC	SE of AUC	r	AUC	SE of AUC
Violence (official records)	-.27*	.73**	.08	.27*	.70*	.07	.23*	.65†	.07
Any Offense (official records)	-.28*	.69*	.07	.32*	.70**	.06	.31*	.69†	.08
Substance Use (self-report on DAU)	-.15	.59	.09	.26*	.66	.09	.51**	.82***	.08
Street Drug Use (self-report on DAU)	-.28*	.71*	.08	.35**	.78**	.07	.30*	.71*	.07
Any Victimization (self-report on PBF Overt Scale)	-.01	.53	.08	.33**	.70*	.07	.23†	.63	.07
Non-suicidal Self-injury	.02	.46	.12	.21	.70†	.08	.20	.67	.11

Note:

† $p < .10$,

* $p < .05$,

** $p < .01$,

*** $p < .001$.

In order to facilitate interpretation, AUC scores for Strength Total scores were reversed to predict the *absence* of outcomes. The risk estimate corresponds to the domain that is relevant to that outcome. For instance, for the victimization outcome, we present correlations on the risk estimate for victimization.

Table 6

Predictive Validity: START:AV Final Risk Estimates and Outcomes on Continuous Self-Report Scales

	Final Risk Estimates				
	Harm to Others	General Offending	Suicide	Substance Abuse	Victimization
Aggressive Offenses Scale on SRO	.33**	.37*	-.06	.23	.02
Total Offenses Scale on SRO	.30*	.42*	.04	.29*	.04
Suicide Ideation Scale on MAYSI-2	-.06	-.14	.50**	.17	-.22
Alcohol/Drug Use Scale on MAYSI-2	.30*	.40**	.06	.47**	.22
Overt Victimization Scale of PBF	.37**	.39*	.19	.32*	.44**

Note:

[†] $p < .10$,

* $p < .05$,

** $p < .01$,

*** $p < .001$. We have bolded and underlined the risk estimate corresponds to the domain that is relevant to that outcome. For instance, for the victimization outcome we bolded and underlined the risk estimate for victimization.

Table 7

Changes in Risk Estimates

Outcomes	Risk Estimate Remained the Same % (<i>n</i>)	Risk Estimate Decreased % (<i>n</i>)	Risk Estimate Increased % (<i>n</i>)
Violence Towards Others	60.3% (38)	22.2% (14)	17.5% (11)
General Reoffending	53.8% (21)	30.8% (12)	15.4% (6)
Suicide	81.0% (51)	12.7% (8)	6.3% (4)
Substance Abuse	66.7% (42)	19.0% (12)	14.3% (9)
Victimization	66.1% (41)	22.6% (14)	11.3% (7)

Table 8
RCI for Changes in Strength and Vulnerability Total Scores for 3-Month Follow-Up

	95% Confidence Interval			90% Confidence Interval		
	% With No Reliable Change	% With Reliable Decrease	% With Reliable Increase	% With No Reliable Change	% With Reliable Decrease	% With Reliable Increase
Strengths Total	88.9%	4.8%	6.3%	82.5%	7.9%	9.5%
Vulnerabilities Total	93.8%	4.7%	1.6%	84.4%	9.4%	6.3%