Young Adult Follow-up of Adolescent Girls in Juvenile Justice Using the Columbia Suicide Severity Rating Scale

Young Adult Follow-up of Adolescent Girls in Juvenile Justice Using the Columbia Suicide

Severity Rating Scale

David C. R. Kerr, PhD

Oregon Social Learning Center

Oregon State University

Brandon Gibson, MS

Leslie D. Leve, PhD

David S. DeGarmo, PhD

Oregon Social Learning Center

Abstract

We studied the reliability and validity of the Columbia Suicide Severity Scale (C-SSRS).

Severely delinquent adolescent girls ($n = 166$) participated in a treatment trial and repeated

assessments over time. Lifetime suicide attempt history was measured using the C-SSRS in early

adulthood ($n = 144$; 7–12 years post-baseline). Nonclinician raters showed strong interrater

reliability using the C-SSRS. Self-, caseworker-, and caregiver-reports of girls' suicide attempt

histories collected at baseline correlated with adult participants' recollections of their baseline

attempt histories. Suicidal ideation measured prospectively across a 7–12 year period was

associated with retrospectively reported suicide attempt across the same period.

Young Adult Follow-up of Adolescent Girls in Juvenile Justice Using the Columbia Suicide

Severity Rating Scale

The reliable and valid assessment of suicide attempt history in adolescents and young

adults is of critical importance to the study of suicide risk and prevention. To promote progress

in this area, experts have emphasized the importance of using a common nomenclature for

suicidal behaviors (O'Carroll, Berman, Maris, Moscicki, Tanney, & Silverman, 1996; Silverman,

Berman, Sanddal, O'Carroll, & Joiner, 2007). Recently there has been clear movement toward

the adoption of clear definitions for suicide attempt and the use of standardized assessment

instruments (Crosby, Ortega, & Melanson, 2011; Posner et al., 2008).

The Columbia Suicide Severity Rating Scale (C-SSRS; Posner et al., 2008) is one such

tool for assessing suicidal thoughts and behaviors. The semi-structured format of the scale directs

interviewers toward information needed to determine whether interviewees' behaviors meet

broadly-accepted (e.g., by the Centers for Disease Control and Prevention; Crosby et al, 2011)

criteria for specific forms of suicidal behavior. The C-SSRS also can be adapted to various time

frames (e.g., past 6 months; since prior interview) needed for research and risk assessment.

The focus of the present study was to investigate the psychometric properties of the C-

SSRS for assessing longterm histories of suicidal behavior; specifically the extent to which: (1)

interviewers classified suicidal behaviors in the same way (*interrater reliability*), (2) C-SSRS

measures were associated with related constructs (*convergent validity*), and (3) C-SSRS

measures were less associated with measures of less related constructs (*divergent validity*). We

did so using data from a longterm follow-up study of two randomized controlled trials (RCT) of

Multidimensional Treatment Foster Care (MTFC), a delinquency-focused intervention for

adolescent girls involved in the juvenile justice system. These girls were not recruited based on

their depression or suicide risk, but nonetheless showed high rates of such problems initially and over time (Chamberlain & Smith, 2003; Kerr et al., under review). MTFC has well documented effects on delinquency outcomes relative to group care (GC) community treatment as usual (Chamberlain, Leve, & DeGarmo, 2007), but was not designed to target depression or suicide risk. Still, girls assigned to MTFC showed greater decreases in depressive symptoms and rates of suicidal ideation than girls assigned to GC in the 7–12 years after baseline; there were no treatment effects on suicide attempt (Kerr et al., under review).

The characteristics and high retention rates of our sample, along with the repeated measures design and unusually long follow-up period positions the present study well to answer questions related to the assessment of suicidal behaviors. In particular, we compare prospective data collected across adolescence with information collected in early adulthood using the C-SSRS on participants' lifetime histories of suicidal behaviors.[1]

The present study can inform assessment in the field of suicidology in several ways. First, the C-SSRS is promising but relatively new. Several studies support the validity of C-SSRS classifications of suicidal behaviors, including Posner and colleagues' (2011) research on three distinct clinical samples of adolescents and adults, some of which were followed for up to 24 weeks. Still, more studies are needed, including research conducted by independent teams, in other settings and populations, with longer timeframes, and in relation to other measurement approaches.

Second, the C-SSRS and other instruments (e.g., standardized diagnostic interviews), including those used to estimate national prevalence figures for suicidal behaviors (Kessler, Borges, & Walters, 1999), aim to assess the occurrence and timing of lifetime suicide attempt

---

[1] The C-SSRS also can be used to assess suicidal behavior across shorter timeframes (e.g., past month), and separately assesses suicidal ideation; its validity for these purposes, however, is not addressed here.

history. Although these instruments rely on the accuracy of long-term retrospection, there is longstanding concern about this approach (e.g., Rogler, Malgady, & Tryon, 1992), and few studies exist that actually evaluate the validity of such retrospections. Thus, our strong long-term retention rates present a relatively unusual opportunity to test the validity of retrospective reports. Relatedly, statistical power to examine such issues is often limited by the low rates of suicidal behaviors in community samples, or during the relatively short follow-up periods of most studies of clinical samples (e.g., Posner et al., 2011). The present study avoids these problems by following high risk individuals for 7-12 years, which is much longer than is typical.

Third, research on suicide risk assessment often is conducted with populations recruited on the basis of acute suicide risk or mood disorder (e.g., Huth-Bocks, Kerr, Ivey, Kramer, & King, 2007; Posner et al., 2011). Additionally, such research often has occurred in the context of hospitalization, when decisions about admission, discharge, and safety planning are primary concerns both for clinical assessment staff and participants themselves. Thus, validity may be compromised when participants are acutely aware of the consequences that screening high or low suicide risk will have on their confinement and treatment. In this way, the present study may support generalizability of the C-SSRS and related instruments by establishing their reliability and validity when used outside these narrow populations and contexts.

Fourth, using self-report measures as the only criteria for evaluating the validity of other self-report measures of suicidal behavior is a limited approach. The present study design permitted comparisons between self-reports and other informants' reports of girls' suicide attempt histories, which is unusual in the field. Finally, in order to inform suicide prevention efforts research must not only focus on individuals who present for help through expected channels (e.g., primary care; outpatient therapy). Such populations critically deserve our

attention, but represent just a portion of those at high risk for suicide. For example, Luoma,
Martin, and Pearson (2002) found that only 38% of suicide decedents under the age of 35 years
had ever had contact with a mental health care professional. Thus, the present study helps widen
the scope and impact of suicide research by focusing on an underserved and understudied clinical
population that shows high rates of suicidal behavior: females in the justice system.

Research Questions

We pursued four primary descriptive and psychometric analyses of the C-SSRS as a
measure of lifetime suicidal behavior histories. First, to increase understanding of the initial
status and outcomes for adolescent girls with serious histories of delinquency, we considered
descriptive lifetime information about the prevalence, number, severity, and timing of their
suicidal behaviors through early adulthood.

Second, we tested interrater reliability for the system developed by Posner and colleagues
(2008) for differentiating among suicidal behaviors and rating the lethality and medical damages
associated with suicide attempts. Agreement among nonclinician interviewers as well as between
interviewers and a clinically trained rater (first author) was evaluated to build on evidence for
reliability that has been established in clinically trained teams (e.g., Brent et al., 2009).

Subsequent analyses focused on testing the convergent validity of the C-SSRS by
comparing participants' retrospective reports during young adulthood with information collected
at prior measurement occasions. Specifically, our third set of analyses focused on the reported
histories of suicide attempt prior to the baseline RCT assessment, and whether young women's
reports of such baseline histories on the C-SSRS corresponded with those reported at baseline by
participants, their caseworkers, and their caregivers. Two forms of divergent validity were
evaluated by testing the extent to which these associations were stronger for baseline attempt

histories than for histories of other (non-attempt) baseline suicidal behaviors or for post-baseline attempt histories.

Fourth and finally, the convergent validity of young women's retrospective reports regarding suicide attempts that occurred during the post-baseline follow-up period was tested by comparing them to depressive symptoms and suicidal ideation collected prospectively across the same period. Using a related approach, Mazza, Catalano, Abbott, and Haggerty (2011) recently found that suicide attempts recalled by community participants at ages 18–19 years were associated with depressive symptoms measured prospectively across the previous 12 years. More specifically, among participants who had attempted suicide, depressive symptoms were higher during the year of their first attempt than were the symptoms of non-attempting youth, and furthermore, were higher than their own depressive symptoms in the years prior to their attempt. We extend these findings using a different approach and population, and by also comparing participants who did and did not attempt suicide on prospectively measured suicidal ideation.

Method

Participants

Adolescent girls ($n = 166$) enrolled in one of two consecutively run RCTs (ns = 81 and 85 for Cohorts 1 and 2, respectively) conducted in the Northwestern United States between 1997 and 2006 to compare MTFC ($n = 81$) and GC ($n = 85$).[2] Participants had been mandated to out-of-home care due to chronic delinquency. Investigators attempted to enroll all referred girls who were 13–17 years old, had at least one criminal referral in the past year, and were placed in out-of-home care within 12 months following referral. Girls who were pregnant at the time of recruitment were excluded. Girls provided assent and their legal guardian provided consent to participate. Assessment personnel were independent from the clinical staff and were blind to

---

[2] The identical group ns for cohorts and group assignments were coincidental.

group assignment.

The girls were 13–17 years old at baseline ($M = 15.30$, $SD = 1.17$); 68.1% were

Caucasian, 1.8% were African-American, 11.4% were Hispanic, 0.6% were Native American,

0.6% were Asian, 16.9% endorsed "multiracial," and 0.6% answered "other/unknown."[3] At

baseline, 63% lived with single-parent families, and 54% lived in families earning less than

$10,000.

Placement in the intervention setting occurred following the baseline assessment. Girls in

Cohorts 1 and 2 were assessed regularly for 36 and 24 months post-baseline, respectively, as part

of the original RCTs. In 2009 assessments began of all participants from Cohorts 1 and 2 who

could be contacted and who consented to enroll in a follow-up study. These assessments

occurred up to five times at 6-month intervals; of the 164 original participants still living, 93% (n

= 152) participated in one or more of these follow-up assessments. The time between the final

assessment associated with the original RCT and the longterm follow-up necessarily differed by

cohort; specifically, the first follow-up assessment occurred a mean (SD) of 9.81 (1.73) and 4.69

(1.16) years post-baseline for girls in Cohorts 1 and 2, respectively. The assessment schedule for

each cohort is described below. The two cohorts were combined for all analyses. Differences in

assessment timing and missing data are accommodated by person-specific timelines and time

weighting used in the growth model. This model also controlled for intervention assignment and

cohort.

Intervention conditions. Both MTFC and GC are described further elsewhere (e.g.,

Chamberlain, 2003). In brief, girls in MTFC were placed with state-certified foster parents

trained to implement a behavioral reinforcement program. In addition, girls received individual

---

[3] Percentages are based on participants' self-reports in early adulthood and thus differ somewhat from those reported in prior 2-year follow-up studies of this sample (Harold et al., 2013; Kerr et al., 2009) that in many cases had to rely on caregiver or caseworker reports.

therapy and their aftercare placement family received parent management training. Interventions were coordinated and closely supervised by a program supervisor. GC represented usual community services for girls being referred to out-of-home care by the juvenile justice system. Average treatment duration (approximately 6 months) did not differ significantly by condition.

Measures

Hierarchical generalized linear model (HGLM) control variables. Control variables used in the HGLM were: *intervention group assignment,* coded -1 (GC) or 1 (MTFC); *cohort,* coded -1 (Cohort 1) or 1 (Cohort 2); *age at baseline,* calculated based on birth date and baseline assessment date; and *time since baseline*, calculated for each assessment based on baseline and assessment date.

Depressive symptoms. Depressive symptoms were measured with the Center for Epidemiologic Studies, Depression Scale (CES-D; Radloff, 1977). Scores range from 0–60 and are based on a sum of 20 items on the frequency of past week symptoms using a 4-point scale (0–3), ranging from *Rarely or none* (0–1 day) to *Most or all of the time* (5–7 days). The CES-D does not measure suicidal thoughts or behaviors. Cohort 1 girls were assessed at baseline, and 3, 6, 12, 15, 18, 30, and 36 months post-baseline, and then again at the 5 young adult follow-up assessments (at initial follow-up, and again 6, 12, 18, and 24 months later). Cohort 2 girls were assessed at baseline, and 3, 12, and 24 months later, and again at the 5 follow-up assesments. Internal consistencies over time were adequate (α = .84 to .94).

Suicidal ideation. Suicidal ideation was measured from adolescence to early adulthood using a face valid item ("During the past week, how much were you bothered by thoughts of ending your life?") from the Brief Symptom Inventory (BSI; Derogatis & Melisaratos, 1983). Girls in Cohort 1 completed the BSI at 3, 6, 12, 18, and 24 months post-baseline, and then again

at the 5 young adult follow-up assessments. Girls in Cohort 2 completed the BSI at baseline, 6, 12, 18, and 24 months post-baseline, as well as at the 5 young adult follow-up assessments. Given the distributions, responses were recoded to the absence ("not at all" coded 0) or presence ("a little bit" to "very much" coded 1) of suicidal ideation at each time point.

Baseline Suicide Attempt History

Each indicator of suicide attempt history (below) was coded as absent (0) or present (1); unless otherwise indicated, items were identical across cohorts and treatment conditions

Self-reported suicide attempt history. Self-report was based on participants' responses on the National Institute of Mental Health Diagnostic Interview Schedule for Children, Version IV (NIMH-DISC-IV; Shaffer, Fisher, Lucas, Dulcan, & Schwab-Stone, 2000). At baseline, girls in Cohort 1 answered the lifetime suicide attempt history question, "*Have you ever, in your whole life, tried to kill yourself?*" Participants in Cohort 2 answered only the item regarding a suicide attempt in the past year at baseline (*"Have you tried to kill yourself in the last year?"*), but answered the past-year and lifetime items at the 12 months follow-up. No Cohort 2 participant who denied a past-year attempt at baseline reported a past-year attempt at the 12-month follow-up. Thus, we considered Cohort 2 girls who had reported either a past year suicide attempt at baseline or a lifetime attempt history at 12 months to have had a positive lifetime suicide attempt history at baseline.

Caseworker report of participant's suicide attempt history. Caseworker report was based on a single interview item asked of each participant's caseworker at baseline: "has she [target child] ever attempted suicide?"

Caregiver report of participant's suicide attempt history. Caregiver report was based on the same interview item: "has she [target child] ever attempted suicide?"

Early Adult Retrospective Histories of Suicide Attempt and Other Suicidal Behavior

Participants were interviewed using the Columbia Suicide Severity Rating Scale (C-SSRS; Posner et al., 2008), Lifetime Version, beginning 12 months after the first young adult follow-up (or at age 18 years, whichever came later), when funding for this additional assessment began. Participants were interviewed again 6 and 12 months later regarding attempts in the intervening months. Thus, participants in Cohorts 1 and 2 were mean (*SD*) ages 26.93 (.25) and 21.95 (.17), respectively, at the time they completed their final C-SSRS, which occurred at means (*SD*s) of 11.67 (.20) and 6.60 (.13) years, respectively, after baseline. Interviews were audiotaped and transcribed for use in reliability evaluations.

The C-SSRS is a semi-structured interview. We adapted the interview for use by a team of non-clinician interviewers by standardizing the probes already provided on the C-SSRS. Probes were used to elicit reporting of all acts potentially meeting criteria for a suicide attempt, according to Posner and colleagues' (2008) definition of "actual attempt." For each act, interviewers then used further probes provided on the C-SSRS to determine which of the following mutually exclusive codes (1–4 are from the C-SSRS) applied to the act: (1) *actual suicide attempt* was "a potentially self-injurious act committed with at least some wish to die, as a result of act"; (2) *aborted attempt* involved being ready to imminently attempt suicide but stopping oneself from starting a potentially self-injurious suicidal act; (3) *interrupted attempt* involved being interrupted by an outside circumstance, such as another person, from starting a potentially self-injurious suicidal act that otherwise would have occurred; (4) *preparatory act* was a behavior (i.e., not just mental planning) enacted to prepare for suicide, such as gathering pills or writing a suicide note; (5) *non-suicidal self-injury*; (6) *suicidal ideation only*; and (7) *other dangerous acts or lifestyles* were events with potentially lethal consequences (e.g., high

speed driving; prostitution) that were not enacted with suicidal intent. Finally, category 8 *other/not applicable* included acts that could not otherwise be coded.

Actual suicide attempts were then rated using C-SSRS actual lethality/medical damage codes, ranging from 0 (*no physical damage or very minor physical damage*) to 4 (*severe physical damage*). We did not utilize code 5 – *death* due to the nature of this study.

For each act, other than those classified as *suicidal ideation only,* or *other dangerous acts*, interviewers queried participants to establish the timing of the acts using anchors regarding age and grade in school, month, or season of the year when possible, and whether acts preceded or followed RCT entry date. Based on retrospective timing estimates provided by participants, and known dates of birth and study entry, we partitioned suicide attempts as having occurred either pre- or post-baseline. Attempts that occurred the year of baseline and lacked specific timepoint information (e.g., month of event) to classify the event in relation to baseline were conservatively attributed to the pre-baseline period. As a further conservative step, validity analyses of post-baseline attempt rate controlled for the number of pre-baseline attempts (log-transformed to reduce skewness) reported on the C-SSRS. Otherwise, pre- and post-baseline attempt history variables each were coded as negative (0) or positive (1). Lifetime suicide attempt history from the C-SSRS was available for 144 participants (87.8% of those living).[4]

Analysis Plan

We used Fleiss (1971) kappa to test interrater reliabilities among multiple raters using categorical codes (type of act). For ordinal codes (injury/severity) we used intraclass correlation coefficient (ICC) with raters as a random effect and the measure as a fixed effect. Other primary

---

[4] One participant (assigned to GC in Cohort 1) is known to have died by suicide. In a separate report of intervention outcomes (author submitted), this participant was coded as having a positive post-baseline suicide attempt history. To maintain comparability across project reports, we did the same here for analyses of whether prospective measures of depressive symptoms and suicidal ideation were associated with post-baseline suicide attempt.

research questions focused on the convergent validity of histories of suicide attempt collected retrospectively in young adulthood using the C-SSRS. For these analyses we used, as validity criteria, previously collected measures of suicide attempt history, suicidal ideation, and depressive symptoms that occurred during relevant time periods. First, we hypothesized there would be a convergence among multiple reporters and data sources regarding the presence of a baseline lifetime suicide attempt history. We evaluated this hypothesis by making pairwise comparisons and then in Mplus version 6.12 (Muthén & Muthén, 1998-2012) by specifying a latent suicide attempt history variable with multiple binary indicators, and noting the strength with which young adult reports of a baseline attempt history on the C-SSRS loaded on the latent variable [using the maximum likelihood with robust standard errors (MLR) estimator]. We tested the divergent content validity of the retrospective reports of suicide attempt on the C-SSRS by considering whether a history of suicidal behavior *other than attempt* was associated with indicators of baseline suicide attempt history; we tested divergent temporal validity by testing whether recollected history of baseline versus post-baseline suicide attempt history was more strongly associated with indicators of baseline attempt history. We also explored which patterns of multi-method/multi-informant agreement regarding a baseline attempt history were most strongly associated with adult recollection of a baseline attempt history.

In our second set of convergent validity tests, we hypothesized that prospective measures of post-baseline suicidal ideation and depressive symptoms would significantly predict the probability that a suicide attempt would be reported (on C-SSRS) to have occurred across the same time interval. Hypotheses were evaluated with hierarchical linear growth models using the HLM7 program (Raudenbush, Bryk, Cheong, Congdon, & du Toit, 2011). In general, the growth model framework in the HLM program is a multilevel regression framework also known as

mixed modeling. Growth modeling is a form of multilevel modeling where the time-varying suicidal ideation and depressive symptoms are repeated measures at Level 1 nested within individuals at Level 2. The time-varying growth rates of suicidal ideation and depressive symptoms are estimated at Level 1 for each individual. These are also the individual differences or variance in trajectories of these outcomes. These trajectories are then regressed on time-invariant predictors at Level 2, which in the present study were baseline age, baseline suicide attempt history, intervention assignment, cohort, and the intervention $\times$ cohort interaction that significantly predicted depressive symptoms in prior reports (e.g., Harold et al., 2013; Kerr et al., under review).

For the Level 1 repeated measures, the HLM program can provide an advantage of estimating growth rates based on each girl's person-specific assessment timeline. Person-specific models provide more precision in the estimates of change than traditional growth models based on fixed growth weights. Further, because suicidal ideation scores and suicide attempt were binary outcomes, we employed hierarchical generalized linear modeling estimation (HGLM) that adjusts the estimation procedure for continuous variable model assumptions to provide appropriate estimates for logistic probability models also known as Bernoulli estimation.

To predict suicide attempt, we entered "time" to estimate growth rates in depressive symptoms and the probability of suicidal ideation as time-varying Level 1 covariates. The covariates included terms that could be interpreted as average levels of depressive symptoms and suicidal ideation, and time-weighted scores representing the increases or decreases in these problems over time (see Singer & Willett, 2003 for discussion of parameterization of level 1 change scores). More formally, for the binary suicide attempt outcome, the probability or log odds of event occurrence is estimated at Level 1 with the following model:

$$\text{Prob (Suicide Attempt C-SSRS)}_{ti} = \beta_{0i} + \beta_{1i} \, (\text{Time})_{ti} + \beta_{2i} \, (\text{DEP})_{ti} + \beta_{3i} \, (\text{Time} \times \text{DEP})_{ti}$$

$$+ \, \beta_{4i} \, (\text{Ideation})_{ti} + \beta_{5i} \, (\text{Time} \times \text{Ideation})_{ti}$$

where DEP is CESD depressive symptoms, Ideation is the BSI measure, and the dependent binary attempt variable is retrospectively reported suicide attempt on the C-SSRS for participant i repeated over time t and is entered in the Level 1 equation. Suicidal attempt is then a function of the baseline intercept β_{0i} for participant i, a nonlinear probability growth rate β_{1i} (Time) for participant i for the duration of her participation. Time was weighted as years since baseline computed from the participant's age at each assessment. After summarizing the individual intercepts (β_{0i}) and growth rate slopes (β_{1i} and β_{2i}) at Level 1, the Level 2 model then regresses intercepts and slopes on predictors as:

$$\text{Suicide Attempt Intercept } \beta_{0i} = \gamma_{00} + \gamma_{01} \, (\text{age})_i + u_{0i}$$

$$\text{Linear Growth } \beta_{1i} = \gamma_{10} + \gamma_{11} \, (\text{age})_i + \gamma_{12} \, (\text{group condition}) \ldots + u_{1i}$$

$$\text{DEP } \beta_{2i} \quad = \gamma_{20} + u_{2i}$$

$$\text{Time} \times \text{DEP } \beta_{3i} \quad = \gamma_{30} + u_{3i}$$

$$\text{Ideation } \beta_{4i} = \gamma_{40} + u_{4i}$$

$$\text{Time} \times \text{Ideation } \beta_{5i} = \gamma_{50} + u_{5i}$$

where γ_{00} is the initial status intercept adjusting for participant age, γ_{01} is the effect of her age on initial status intercept, and u_{0i} is the random error variance. γ_{10} represents the growth rate adjusting for covariates, γ_{11} is the effect of participant age on growth in suicidal ideation, and γ_{12} is the MTFC intervention effect (intent to treat) on growth in ideation. γ_{20} is the time-varying average level effect of a unit increase in the depressive symptom scale score and each corresponding wave of ideation occurrence. The γ_{30}Time $\times$DEP is the time weighted growth in depressive symptoms, or simply the change in these symptoms as a predictor of the probability

of suicide attempt, and similarly, for Ideation γ_{40} and change in Ideation γ_{50}.

<div align="center">

Results

</div>

Descriptive Statistics

In pursuit of the first aim of this study, we describe the rates of suicidal behaviors reported by adolescent participants and other informants at RCT baseline, and by young adult participants at the longterm follow-up assessment.

Baseline histories of suicide attempt. Baseline self-reports indicated that 48% (79 of 165 non-missing observations) of girls had a lifetime history of suicide attempt at baseline. According to self-reports collected in young adulthood (C-SSRS) and using Posner and colleagues' (2008) definition, 30% (50 of 144) had a history of *actual suicide attempt* at baseline. Baseline caseworker and caregiver reports indicate that 41% (65 of 157) and 43% (63 of 147) of participants, respectively, had a history of suicide attempt.

Lifetime histories of suicide attempt and other suicidal behavior (C-SSRS). The 245 acts participants reported on the C-SSRS were coded as follows: 102 actual, 11 interrupted, and 4 aborted attempts, 16 preparatory acts, 53 acts of non-suicidal self-injury, 10 reports that involved suicidal ideation only, 40 other non-suicidal dangerous acts, and 9 acts judged to be not applicable. Of the 102 actual attempts, lethality/medical damages were coded as follows: 40 *none or very minor*, 26 *minor*, 28 *moderate*, 6 *moderately severe*, and 2 *severe*.

The numbers of participants reporting a lifetime history of each behavior of interest[5] on the C-SSRS are presented in Table 1. As can be seen, nearly half reported a suicide attempt; 17% reported a multiple attempt history. Other suicidal acts were less commonly reported than actual attempts. We wish to be clear that the relatively low rate of preparatory acts (12%) does not

---

[5] We do not report the numbers of participants who, in response to probes regarding suicidal behaviors, reported an incident that we coded as "suicidal ideation only"; this would be a misleading estimate as lifetime history of suicidal ideation was not separately queried.

imply that most acts were unplanned, given that actual, interrupted, or aborted attempts that imminently followed preparatory acts were coded as attempts. Non-suicidal self-injury was reported at least once by a quarter of the sample.

For participants who made one or more actual suicide attempts, Table 1 also indicates the maximum lethality/medical damage rating assigned to any of their attempts. Approximately one third of all participants made an attempt resulting in some injury (rating of minor or higher); approximately one fifth made an attempt resulting in at least a moderate injury (i.e., one that should have received medical attention); and 6% reported a moderately severe or severe attempt (i.e., one that required hospitalization for non-psychiatric medical reasons).

Figure 1 shows the timing of first actual suicide attempt as reported on the C-SSRS. Nearly three quarters ($n = 48$) of the 66 who reported a lifetime history of suicide attempt and a third of the total available sample reported first attempting suicide before age 15 years. Nine girls (14% of 66) attempted prior to age 12; few ($n = 5$; 8% of 66) reported a first attempt at age 18 years or older, though the sample was right censored.

Interrater Reliability of Interviewer Ratings on C-SSRS

Next, we report results of interrater reliability analyses on interviewers' C-SSRS ratings collected in young adulthood. Most participants ($n = 94$) reported at least one act in response to probes meant to elicit reporting of all potential suicide attempts; 245 acts were reported and coded. We originally planned to conduct reliability analyses on a random 20% of acts; however, we expanded the pool of reliability cases because some coding categories were used relatively infrequently (e.g., interrupted attempts). We also continued to assess reliability over time to guard against rater drift, which further increased this pool to 108 (44%) randomly-selected acts.

According to a priori training rules, individual raters were required to reach 85%

agreement with the criterion clinical rater's (first author) ratings on practice cases before

conducting interviews independently, and to remain at 85% or greater agreement with the

clinical rater's ratings on ongoing team-wide reliability checks. Across the 108 acts, individual

raters' absolute agreement with the criterion clinical rater's ratings ranged from 90–97%.

We used Fleiss (1971) kappa to derive measures of agreement among the 5 raters (4

interviewers and one criterion clinical rater) on the nominal (categorical) act rating. Kappa for

raters discerning actual attempt from all other reported acts was .91. We also conducted Fleiss

kappas for other key distinctions. Kappa was .90 for discriminating between the suicidal

behaviors (actual, interrupted, and aborted attempts and preparatory acts) and all other

categories. Kappa was .88 for the 5-category distinction among actual, aborted, interrupted

attempts, preparatory acts, or any other type of act. Finally, kappa was .86 for the use of all 8

categories of ratings. Thus, interrater agreement could be considered very good (Altman, 1991).

Next we tested interrater reliability for the ordinal lethality/medical damage code using two-way

mixed effects average intraclass correlation coefficient (ICC) to test absolute agreement. Medical

injuries are only coded for actual attempts, and the ICC could only be run if all raters contributed

a rating. Thus, the ICC was based on the 42 reliability cases that all raters agreed were actual

attempts. Average-measures absolute agreement among raters was very good, ICC = .953.

Convergent and Divergent Validity of Reports of Baseline Suicide Attempt History

Subsequent analyses concerned the validity of reports of suicidal behavior, and in

particular the validity of longterm retrospective self-reports of suicide attempt using the C-SSRS.

To this end, we first conducted tests of whether baseline reports of lifetime suicide attempt

history would be associated with young adult participants' recalled histories of suicide attempt across the same time period.

Convergent validity of baseline reports of suicide attempt history. Table 2 depicts the correspondence among the four measures of baseline suicide attempt history. Among the measures collected at baseline, caseworker reports showed fair to moderate agreement with those of caregivers and girls, κ (138) = .572 and κ (156) = .352, respectively, $p < .001$. Reports by caretakers and girls showed moderate agreement, κ (146) = .422, $p < .001$.

Convergent and divergent validity of C-SSRS reports of lifetime suicide attempt history at baseline. Retrospective reports of lifetime history of baseline suicide attempt collected in early adulthood on the C-SSRS showed fair levels of agreement with the reports that caseworkers, caregivers, and girls made at baseline, κ (135) = .310, κ (130) = .216, and κ (143) = .330, respectively, $p < .05$. In contrast, C-SSRS retrospective reports of lifetime history of baseline suicidal behaviors *other than actual attempt* (i.e., interrupted and aborted attempt, and preparatory act) were weakly related to baseline caseworker, caretaker, and girls' self-reports of suicide attempt, κ (135) = .030, $p = .42$, κ (130) = .087, $p = .159$, and κ (143) = .177, $p < .01$, respectively. Thus, divergent validity was supported in that recollected attempt history was more strongly related to prior self- and other-reported histories of attempt than was recollected history of non-attempt suicidal behaviors. This interpretation is further strengthened by the associations observed between retrospective reports of baseline actual attempts and retrospective reports of these other suicidal behaviors, $r (144) = .230$, $p < .01$, that would tend to thwart detection of discrimination.

A further, multi-method test of the convergent validity of the C-SSRS was conducted by modeling a latent baseline suicide attempt variable comprised of the four categorical indicators

[Model fit Likelihood Ratio χ^2 ($n = 166$, $df = 7$) = 9.455, $p = .22$]. Standardized loadings of the baseline caseworker, caregiver, and self-reports and the young adult retrospective reports (C-SSRS) on this factor were β =.850, .862, .649, and .496, respectively, all $p < .001$; the latter factor loading indicates a moderate level of correspondence between the C-SSRS and the latent baseline suicide attempt factor (Altman, 1991).

Given that the underlying factor is a function of the indicators, we ran exploratory analyses to further probe how the young adult reports (C-SSRS) would load on a factor defined by combinations of two of the three (caseworker, caregiver, and baseline self-reports) other indicators. When caregiver reports were omitted from the model, C-SSRS loaded $\beta = .656$ (other loadings = .693 and .731); when caseworker reports were omitted, C-SSRS loaded $\beta = .503$ (others = .606 and .962); and when baseline self-reports were omitted, C-SSRS loaded $\beta = .433$ (others = .979 and .775). Thus, validity estimates for C-SSRS ranged from .433–.656, $p < .001$.

We probed the divergent temporal validity of young adult retrospections (i.e., the pre- vs. post-baseline distinction) by rerunning the 4-indicator model to include post-baseline suicide attempt recollected in young adulthood as a fifth indicator. Standardized loadings for the first four indicators remained significant ($p < .001$) and nearly identical in magnitude to those above, whereas post-baseline suicide attempt did not load significantly on the factor, $\beta = .204$, $p = .15$. Thus, findings suggested that indicators of suicide attempt history collected at baseline were more strongly associated with long-term recall of baseline than post-baseline suicide attempt history.

Patterns of multi-informant/multi-method agreement. To better understand the patterns of agreement and disagreement among informants and methods, we further considered the subsample of 120 participants who had data from all baseline informants (caseworker,

caregiver, and self-report) and young adult retrospective C-SSRS reports of pre-baseline attempt

history. Participants were classified on the basis of having a positive attempt history according to

no informants ($n = 44$; i.e., a negative history), other-report only ($n = 20$), self-report only ($n =$

12), self- and one other-report ($n = 14$), or self- and both other-reports ($n = 30$). Rates of

retrospectively reported baseline suicide attempt differed significantly by group, and were as

follows: 18% ($n = 8$ of 44), 25% ($n = 5$ of 20), 33% ($n = 4$ of 12), 50% ($n = 7$ of 14), and 60% (n

$= 18$ of 30), respectively, χ^2 ($df = 4$) $= 15.99$, $p < .001$; in broad terms, the pattern suggests a

suicide attempt history initially reported by multiple informants was more likely to be identified

in young adulthood on the C-SSRS than if reported by a single informant.

Convergent Validity of C-SSRS Reports of Post-Baseline Suicide Attempt.

In our fourth and final set of analyses, we tested associations that post-baseline suicide

attempt had with prospectively measured (post-baseline) depressive symptoms and suicidal

ideation. Results are reported in Table 3. Step 1 is reported in Kerr et al. (under review) and is

not relevant to convergent validity; intervention, cohort, and intervention by cohort effects were

not significant predictors of suicide attempt. Participants who were younger at study entry were

at greater risk; older girls were half as likely to attempt suicide after baseline compared to

younger girls. Surprisingly, pre-baseline suicide attempts did not predict subsequent attempt.

Steps 2 and 3 were germane to convergent validity. Step 2 of the model revealed that

participants who showed higher levels of depressive symptoms across the longterm follow-up

period more often reported a suicide attempt during this period. Deviations (e.g., less steeply

decreasing) from the overall trend of decreasing trajectories of depressive symptoms were not

associated with suicide attempt. Step 3 indicated that participants with higher average levels of

suicidal ideation from baseline to young adulthood were more likely to report a suicide attempt

across the same period; furthermore, the association was not explained by concurrent trajectories of depression. Of note, the extremely high odds ratio for suicidal ideation appears to be explained by the large proportion and co-occurrence of zeros among the binary predictors and the binary outcome in the logistic regression. The point-biserial correlation between post-baseline suicide attempt and average suicidal ideation rate across the period was significant r (146) = .320, $p < .001$, and of a more interpretable magnitude. Finally, relative decreases in suicidal ideation over time did not independently predict suicide attempt.

Discussion

In this sample of seriously delinquent adolescent girls we found evidence for the validity of longterm retrospective reports of lifetime suicide attempt history on the C-SSRS. Specifically, young women's retrospective reports of whether they had attempted suicide by middle adolescence (i.e., by baseline of the original study) showed fair to moderate agreement with self- and other-reported lifetime histories of suicide attempt collected in middle adolescence (i.e., at baseline). In support of divergent construct validity, recollected histories of suicidal behaviors other than actual attempt (interrupted and aborted attempt and preparatory act) were weakly associated with prior self- and other-reports of participants' suicide attempt history across the same time period. In support of divergent temporal validity, self- and other-reports of suicide attempt history collected at baseline were more strongly associated with young adults' retrospective reports of their baseline than of post-baseline suicide attempt history. Thus, the lifetime version of the C-SSRS performed well at elucidating and classifying attempt history.

Several points must be kept in perspective regarding whether this evidence directly pertains to the overall validity of the C-SSRS interviewing and coding system for suicide attempt. First, Posner and colleagues' (2008) instrument articulates a narrower operational

definition of suicide attempt than may have been used by participants and informants at earlier timepoints. Thus, imperfect agreement between ratings on the C-SSRS and prior self- and other-reports of suicide attempt is to be expected. Second, our approach confounded agreement of *whether* and *when* a suicide attempt occurred, and therefore contributed to decreased agreement.

Correspondence regarding girls' lifetime histories of attempt was statistically significant for every pair of informants or methods, but was far from perfect for any pair. Correspondence among participating girls, their caregivers, and their caseworkers at baseline regarding girls' lifetime suicide attempt histories was in the moderate range (κ = .35–.57). Particularly vexing were situations in which caregivers or caseworkers reported attempts that girls themselves did not. We can only speculate that some girls may have been motivated to minimize or conceal their histories from evaluators, or that some caregivers or caseworkers may have mistaken non-suicidal self-injurious acts for suicide attempts.

Agreement between participants' reports of baseline suicide attempt history when asked at baseline and again in adulthood was only fair (κ = .330), and lower than that reported in other work; for example, Klimes-Dougan and colleagues (2007) found moderate (κ = .57) agreement between self-reports in adolescence and young adulthood regarding suicidal thoughts and behaviors (an aggregate measure) across comparable time periods. Such findings are consistent with results of a seminal study on the retrospective method in which Henry, Moffitt, Caspi, Langley, and Silva (1994) found that prospectively measured psychosocial variables and mental health symptoms were uncorrelated or only modestly correlated with retrospective measures covering the same time period. They found that the correspondence was stronger for events and objective states and experiences (e.g., injury, height, residence changes), but still not strong for the number or timing of such phenomena. These findings are pertinent to the assessment of

suicide attempt. Attempts are events that sometimes can be externally verified (e.g., if an attempt or its consequences are witnessed or documented). On the other hand, many suicide attempts are unknown to all but one person and thus recollection relies wholly on her/his memory and willingness to disclose it. Additionally, determining that an act was indeed a suicide attempt requires that the psychological state of intent to die be established. While intent can be inferred from some attempt methods (e.g., firearms), intent often can only be determined from the attempters' recollection of her/his state of mind, which again is prone to error in retrospective report (Henry et al., 1994). The factors associated with recall accuracy were beyond the scope of the present study, but may include more severe attempts and poorer psychological functioning at the time of recall (Christl, Wittchen, Pfister, Lieb, & Bronisch, 2006; Klimes-Dougan, Safer, Ronsaville, Tinsley, & Harris, 2007; Prinstein, Nock, Spirito, & Grapentine, 2001).

Failing to get the same answer twice is a familiar challenge for suicide researchers who measure suicidal thoughts and behaviors using multiple methods, multiple informants, or at multiple time points (e.g., Christl et al., 2006; Klimes-Dougan et al., 2007; Miranda, Scott, Hicks, Wilcox, Munfakh, & Shaffer, 2008; Prinstein et al., 2001). Minimizing methodological sources of error (e.g., through consensus in the field regarding operational definitions) is essential, but will not overcome limits to human recall. This perhaps unavoidable multi-method/multi-informant disagreement is troubling to researchers who wish to use suicide attempt history to, for example, identify genetic risk or determine whether inclusion/exclusion are met for a prevention trial. Others will recognize it as evidence that measurement of suicide attempt, like measures of other events, states, and traits in social science, can be enhanced by repeated and multi-method assessment and the use of latent variable modeling strategies. These approaches assume there is error in any single methodology and therefore model uncertainty

statistically. In all, we echo the conclusions of Henry et al., (1994) that the utility and limitations of retrospective methods must be well understood, and we too urge caution in research and clinical contexts that require specificity in timing, number, and accompanying subjective states.

A second set of findings using very different methodology further supported the convergent validity of long-term retrospective reports of suicide attempt on the C-SSRS. Participants who showed higher rates of depressive symptoms and suicidal ideation measured prospectively across the post-baseline period were significantly more likely than others to report on the C-SSRS having made a suicide attempt across that same period. These findings are consistent with those found in a recent study based on a similar method (Mazza et al., 2011) and extend to an observed association between past suicidal ideation and recollected suicide attempt, as well as to a clinical sample. In this way, our study is one of only a few studies that have considered the validity of retrospective reports of whether individuals have attempted suicide, and of an even smaller number that have considered the validity of reports of *when* attempts occurred. Unfortunately, our analyses could not make more temporally specific conclusions.

A third set of key findings from the present study were those supporting that interrater reliability estimates were strong for our team of nonclinician interviewers trained and supervised by a licensed clinician and suicide researcher. There was strong agreement regarding ratings of types of suicidal behaviors (e.g., aborted attempt) and injury/lethality ratings of actual attempts. These findings complement prior support for the interrater reliability of the C-SSRS (Brent et al., 2009) and bode well for the adoption of the measure in research contexts that cannot support an entire team of clinically-trained evaluators. We emphasize, however, that studies using the C-SSRS will require a clinically-trained team member or clinical consultant for the purposes of

monitoring the safety of participants. Further, we note that our reliability data pertain to the *coding* of the C-SSRS from transcripts, and not to the overall administration of the instrument.

Finally, the descriptive information that participants provided on their longterm histories of suicidal behavior was of interest and concern. Some of the findings regarding histories of suicidal behaviors at study entry were consistent with what is known: that juvenile justice involved youth—and girls in particular—are at high risk for suicide attempt (Abram et al., 2008; Rohde, Seeley, & Mace, 1997). However, the information on participants' subsequent (post-baseline) longterm histories of suicidal thoughts and behaviors are unique, to our knowledge. Fortunately, rates of depressive symptoms, suicidal thoughts, and behaviors tended to decline dramatically to early adulthood, but overall were higher across adolescence than what is reported in community studies. Thus, there is clear potential for interventions to further reduce risk among all or an identified subgroup of juvenile justice girls. The appropriate interpretation of these patterns, however, is somewhat obscured by the fact that participants were involved in an intervention trial. Indeed, as reported previously, MTFC was associated with greater decreases in depressive symptoms and suicidal thoughts than GC. MTFC was not found to impact suicide attempt, but prior and present analyses suggest a pattern of effects that are consistent with the possibility; that is, MTFC reduced depressive symptoms and suicidal ideation, and these decreased levels in turn were associated with decreased risk for suicide attempt (Harold et al., 2013; Kerr et al., under review). Future studies should formally test these patterns.

Based on this and other research we suggest that follow-up studies of interventions—even those that do not target suicide, but that aim to change conditions that share etiological roots with suicide (depression, delinquency, substance abuse, dating violence)—should measure suicidal behavior. Doing so in a manner that permits parsing of pre- and post-intervention

behavioral events will help grow the evidence base on the prevention of suicide. Even statistically underpowered studies can contribute to this mission once their findings are aggregated by future meta-analysts.

The present study had many design strengths, but also some important limitations. Foremost, our findings offer no information regarding the incremental utility of the C-SSRS over other brief (i.e., a single, yes/no self-report item) measures of suicide attempt. Second, as suicide attempts across the follow-up periods were not measured in a consistent manner between the two cohorts, and were not measured comprehensively (i.e., with no temporal gaps) within either cohort, we can offer only relatively coarse evidence for validity of the timing of recollected suicide attempts. Third, we cannot assume that findings would generalize to populations with different clinical features or those that are more ethnically diverse or from urban settings.

In conclusion, our findings offer evidence for the interrater reliability and convergent and divergent validity of the C-SSRS as a tool for retrospectively assessing histories of suicidal behaviors. These findings support the utility of uniform definitions of suicidal behavior in research and clinical contexts and the use of the C-SSRS with diverse clinical populations.

References

ABRAM, K. M., CHOE, J. Y., WASHBURN, J., TEPLIN, L. A., KING, D., & DULCAN, M. K. (2008).

Suicidal ideation and behaviors among youths in juvenile detention. *Journal of the*

American Academy of Child and Adolescent Psychiatry, 47, 291–300. doi:

10.1097/CHI.0b013e318160b3ce

ALTMAN, D. G. (1991). *Practical Statistics for Medical Research*. London: Chapman & Hall.

BRENT, D., GREENHILL, L. L., COMPTON, S., EMSLIE, G., WELLS, K., WALKUP, J., …TURNER, J.

B. (2009). The Treatment of Adolescent Suicide Attempters Study (TASA): Predictors of

suicidal events in an open treatment trial. *Journal of the American Academy of Child and*

Adolescent Psychiatry, *48*, 987–996. doi:10.1097/CHI.0b013e3181b5dbe4

CHAMBERLAIN, P. (2003). The Oregon Multidimensional Treatment Foster Care model: Features,

outcomes, and progress in dissemination. In S. Schoenwald & S. Henggeler (Series Eds.)

Moving evidence-based treatments from the laboratory into clinical practice. *Cognitive*

and Behavioral Practice, 10, 303–312.

CHAMBERLAIN, P., LEVE, L. D., & DEGARMO, D. S. (2007). Multidimensional Treatment Foster

Care for girls in the juvenile justice system: 2-year follow-up of a randomized clinical

trial. *Journal of Consulting and Clinical Psychology, 75*, 187–193. doi:10.1037/0022-

006X.75.1.187

CHAMBERLAIN, P., & SMITH, D. K. (2003). Antisocial behavior in children and adolescents: The

Oregon Multidimensional Treatment Foster Care model. In A. E. Kazdin and J. R. Weisz

(Eds.), *Evidence-based psychotherapies for children and adolescents* (pp. 282–300). New

York: Guilford Press.

CHRISTL, B., WITTCHEN, H. U., PFISTER, H., LIEB, R., & BRONISCH, T. (2006). The accuracy of prevalence estimations for suicide attempts: How reliably do adolescents and young adults report their suicide attempts? *Archives of Suicide Research, 10*, 253–263.

CROSBY, A. E., ORTEGA, L., & MELANSON, C. (2011). *Self-directed violence surveillance: Uniform definitions and recommended data elements, Version 1.0.* Atlanta, GA: Centers for Disease Control and Prevention, National Center for Injury Prevention and Control.

DEROGATIS, L., & MELISARATOS, N. (1983). The Brief Symptom Inventory: An introductory report. *Psychological Medicine, 13*, 595–605. doi:10.1017/S0033291700048017

FLEISS, J. L. (1971). Measuring nominal scale agreement among many raters. *Psychological Bulletin*, *76*, 378–382. doi:10.1037/h0031619

HAROLD, G. T., KERR, D. C. R., VAN RYZIN, M., DEGARMO, D. S., RHOADES, K., & LEVE, L. D. (2013). Depressive symptom trajectories among girls in the juvenile justice system: 24-month outcomes of an RCT of Multidimensional Treatment Foster Care. *Prevention Science.* Advance online publication. doi: 10.1007/s11121-012-0317-y

HENRY, B., MOFFITT, T. E., CASPI, A., LANGLEY, J., & SILVA, P. A. (1994). On the "Remembrance of things past": A longitudinal evaluation of the retrospective method. *Psychological Assessment, 6*, 92–101. doi:10.1037//1040-3590.6.2.92

HUTH-BOCKS, A., KERR, D. C., IVEY, A. Z., KRAMER, A. C., & KING, C. A. (2007). Assessment of psychiatrically hospitalized suicidal adolescents: Self-report instruments as predictors of suicidal thoughts and behavior. *Journal of the American Academy of Child & Adolescent Psychiatry, 46*, 387–395.

KERR, D. C., DEGARMO, D. S., LEVE, L. D., & CHAMBERLAIN, P. C. (2012). *Juvenile justice girls' depressive symptoms and suicidal ideation nine years after Multidimensional Treatment Foster Care.* Manuscript submitted for publication.

KERR, D. C. R., LEVE, L. D., & CHAMBERLAIN, P. (2009). Pregnancy rates among juvenile justice girls in two RCTs of Multidimensional Treatment Foster Care. *Journal of Consulting and Clinical Psychology, 77,* 588–593.

KESSLER, R. C., BORGES, G., & WALTERS, E. E. (1999). Prevalence of and risk factors for lifetime suicide attempts in the National Comorbidity Survey. *Archives of General Psychiatry, 56¸* 617–626. doi:10.1001/archpsyc.56.7.617

KLIMES-DOUGAN, B., SAFER, M. A., RONSAVILLE, D., TINSLEY, R., & HARRIS, S. J. (2007). The value of forgetting suicidal thoughts and behavior. *Suicide and Life-Threatening Behavior, 37*, 431–438. doi:10.1521/suli.2007.37.4.431

LUOMA, J. B., MARTIN, C. E., & PEARSON, J. L. (2002). Contact with mental health and primary care providers before suicide: A review of the evidence. *The American Journal of Psychiatry, 159*, 909–916. doi:10.1176/appi.ajp.159.6.909

MAZZA, J. J., CATALANO, R. F., ABBOTT, R. D., & HAGGERTY, K. P. (2011). An examination of the validity of retrospective measures of suicide attempts in youth. *Journal of Adolescent Health, 49*, 532–537. doi:10.1016/j.jadohealth.2011.04.009

MIRANDA, R., SCOTT, M., HICKS, R., WILCOX, H. C., MUNFAKH, J. L. H., & SHAFFER, D. (2008). Suicide attempt characteristics, diagnoses, and future attempts: Comparing multiple attempters to single attempters and ideators. *Journal of the American Academy of Child & Adolescent Psychiatry, 47*, 32–40.

MUTHÉN, L. K., & MUTHÉN, B. O. (1998-2012). *Mplus User's Guide. Seventh Edition*. Los

    Angeles, CA: Muthén & Muthén.

O'CARROLL, P. W., BERMAN, A. L., MARIS, R. W., MOSCICKI, E. K., TANNEY, B. L., &

    SILVERMAN, M. M. (1996). Beyond the Tower of Babel: A Nomenclature for

    Suicidology. *Suicide and Life-Threatening Behavior, 26*, 237–252.

POSNER, K., BRENT, D., LUCAS, C., GOULD, M., STANLEY, B., BROWN, G., … MANN, J. (2008).

    Columbia-Suicide Severity Rating Scale (C-SSRS), Lifetime/Recent Version. Retrieved

    from The Research Foundation for Mental Hygiene, Inc., www.cssrs.columbia.edu.

POSNER, K., BROWN, G. K., STANLEY, B., BRENT, D. A., KSENIYA, V. Y., OQUENDO, M. A.,

    …MANN, J. J. (2011). The Columbia–Suicide Severity Rating Scale: Initial validity and

    internal consistency findings from three multisite studies with adolescents and adults. *The*

    American Journal of Psychiatry, 168, 1266–1277. doi:10.1176/appi.ajp.2011.10111704

PRINSTEIN, M. J., NOCK, M. K., SPIRITO, A., & GRAPENTINE, W. L. (2001). Multimethod

    Assessment of suicidality in adolescent psychiatric inpatients: Preliminary results.

    Journal of the American Academy of Child & Adolescent Psychiatry, 40, 1053–1061.

RADLOFF, L. S. (1977). The CES-D scale: A self-report depression scale for research in the

    general population. *Applied Psychological Measurement, 1*, 385–401.

    doi:10.1177/014662167700100306

RAUDENBUSH, S. W., BRYK, A., CHEONG, Y. F., CONGDON, R., & DU TOIT, M. (2011). *HLM7:*

    Hierarchial linear and nonlinear modeling. Chicago, Il: Scientific Software International.

ROHDE, P., SEELEY, J. R., & MACE, D. E. (1997). Correlates of Suicidal Behavior in a Juvenile

    Detention Population. *Suicide and Life-Threatening Behavior, 27*, 164–175. doi:

    10.1111/j.1943-278X.1997.tb00288.x

ROGLER, L. H., MALGADY, R. G., & TRYON, W. W. (1992). Evaluation of mental health: Issues of memory in the Diagnostic Interview Schedule. *Journal of Nervous and Mental Disease, 180,* 215–222.

SHAFFER, D., FISHER, P., LUCAS C. P., DULCAN, M. K, & SCHWAB-STONE, M. (2000). NIMH Diagnostic Interview Schedule for Children, Version IV (NIMH DISC-IV): Description, differences from previous versions and reliability of some common diagnoses. *Journal of the American Academy of Child & Adolescent Psychiatry, 39,* 28–38. doi:10.1097/00004583-200001000-00014

SILVERMAN, M. M., BERMAN, A. L., SANDDAL, N. D., O'CARROLL, P. W., & JOINER, T. W. (2007). Rebuilding the Tower of Babel: A revised nomenclature for the study of suicide and suicidal behaviors part 2: Suicide-related ideations, communications, and behaviors. *Suicide and Life-Threatening Behavior, 37*, 264–277. doi:10.1521/suli.2007.37.3.264

SINGER, J. D., & WILLETT, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. New York: Oxford University Press.

Table 1

Numbers (%) of participants reporting a lifetime history of behaviors of interest on the C-SSRS.

	n	% (of *n* = 144)
Any actual suicide attempts	66	45.8
1 suicide attempt	41	28.5
2 suicide attempts	20	13.9
3 suicide attempts	3	2.1
5+ suicide attempts	2	1.4
Maximum lethality/medical damage codes for attempts		
very minor or none	17	11.8
minor	18	12.5
moderate	23	16.0
moderately severe	6	4.2
severe	2	1.4
Other coded acts of interest		
Any interrupted attempts	9	6.3
Any aborted attempts	4	2.8
Any preparatory acts	12	8.3
Any non-suicidal self-harm acts	36	25.0

Table 2

Observed agreement among methods and informants regarding baseline suicide attempt history

	Self-report (C-SSRS, adulthood)			Caseworker (baseline)			Caregiver (baseline)		
	Yes	No	Missing	Yes	No	Missing	Yes	No	Missing
Self-report (baseline)									
Yes	35	32	12	44	29	6	46	25	8
No	15	61	10	21	62	3	17	58	11
Missing	0	1	0	0	1	0	0	1	0
Self-report (C-SSRS, adulthood)									
Yes				27	18	5	27	20	3
No				25	65	4	29	54	11
Missing				13	9	0	7	10	5
Caseworker (baseline)									
Yes							45	12	8
No							17	64	11
Missing							1	8	0

Table 3

HGLM unstandardized coefficients for logistic prediction of suicide attempt

	Step 1		Step 2		Step 3	
	beta	OR	beta	OR	beta	OR
Pre-Baseline Attempts	0.257	1.294	0.203	1.226	0.332	1.394
Baseline Age	-0.547**	0.579	-0.541**	0.582	-0.679**	0.507
MTFC	0.132	1.141	0.160	1.174	0.066	1.068
Cohort	-0.332	0.717	-0.299	0.741	-0.385	0.680
MTFC × Cohort	0.010	1.010	0.057	1.059	0.334	1.397
Depressive Symptoms			0.068*	1.071	-0.008	0.992
Δ Depressive Symptoms			0.013	1.014	0.021	1.021
Suicidal Ideation					10.873**	52749.956
Δ Suicidal Ideation					.960	2.612
-2 Log Likelihood	130.482		125.053		110.431	
Cox & Snell R^2	.078		.112		.197	

Note. Coefficients are the unstandardized values. HGLM = hierarchical generalized linear model estimation.
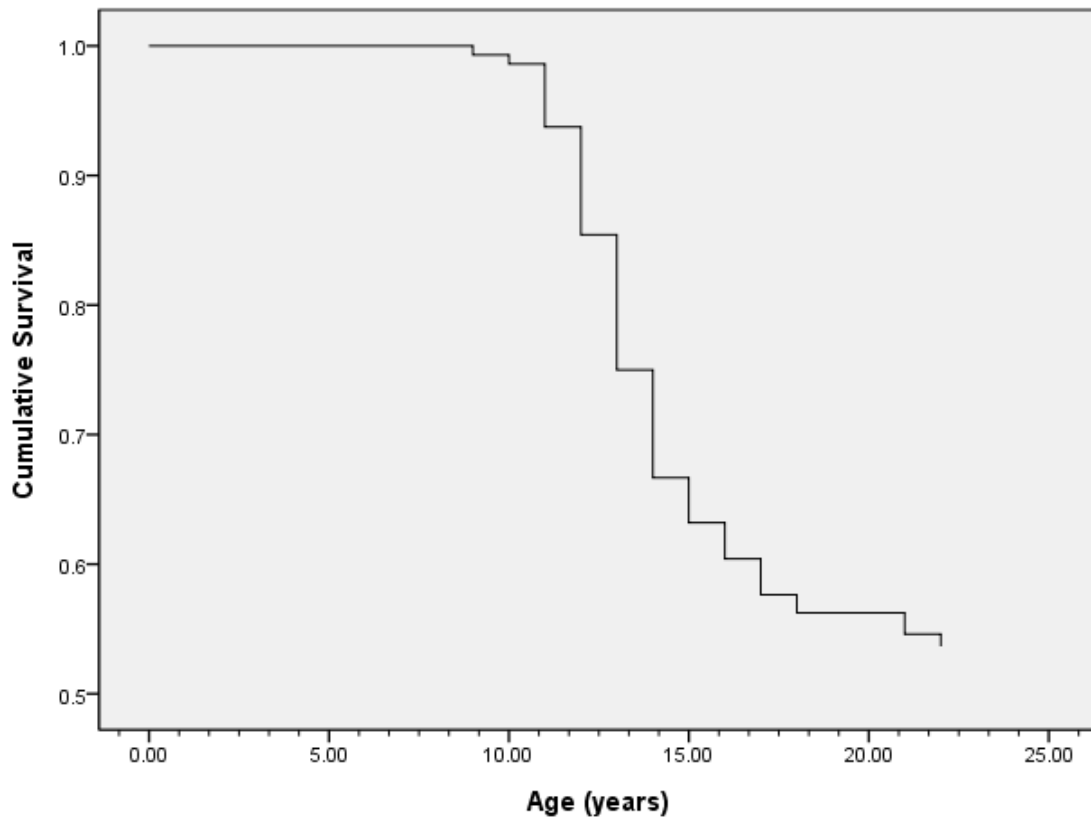
***$p < .001$. **$p < .01$. *$p < .05$.

Figure 1. Retrospectively reported age of first suicide attempt (*n* = 66) in sample of 144 juvenile justice girls followed to early adulthood.