Randomized Control Trial of Big Brothers Big Sisters Mentoring Program for Crime and Delinquency
NCT03495635

(PI: David DuBois; Co-PI: Carla Herrera)

Research and Analysis Plan (Revised 7/11/2022; Note: Original version of the plan was deposited on Open Science Framework website prior to start of study enrollment – see https://osf.io/8ukfv/)

Note: This 7.11.2022 update to our Research and Analysis plan adds an adjustment to impact analyses to take into account non-independence of youth nested within the same family (i.e., siblings) as siblings were randomized (see page 5). Because this adjustment is a more conservative approach, it was deemed appropriate to add despite the 18-month outcome analyses already being in progress. The modification is noted on page 11.

STUDY DESIGN

Overview: This study will rigorously test the Community-Based Mentoring (CBM) program of Big Brothers Big Sisters of America (BBBSA) through a randomized control trial conducted in 16 BBBSA affiliates with approximately 1,356 youth, age 10 or older at enrollment. The measurement strategy will be multi-source/multi-informant. Youth will be the primary informant for most outcomes while parents will be the sole reporter on more "objective" outcomes they would be aware of (e.g., suspensions) and a second reporter on outcomes benefiting from triangulation (e.g., delinquent behavior, aggressive behavior, academic success). Youth and parent surveys will be administered cost-efficiently at baseline and both 18 months and 4 years later; official records of delinquency/crime will be collected twice for all participants: once at the end of the 18-month data collection period, to assess arrests for each youth through 18 months post-baseline, and again 2.5 years later to assess arrests for each youth through 4 years post-baseline; see **Appendix A** for timeline, which is current for all amendments to the original study plan that are referenced in this document). Based on the current age distribution of youth at enrollment into the CBM program, by this latter time point roughly two-thirds of the sample will be at least 16 years old, which is the peak age of onset for serious violence in the U.S. (U.S. Office of the Surgeon General, 2001).

Records data will serve as fairly "objective" measures that have clear policy relevance. However, these data are not without limitations—many delinquent behaviors and crimes go undetected. And even when detected, some do not get officially recorded. The inclusion of self-report data will help address these limitations. Self-reports of negative behavior can be positively biased (i.e., youth report fewer negative behaviors than they have actually engaged in). Yet, they also can capture many minor offenses that are not reflected in police or court records or are simply not caught. Most researchers agree that self-reports are a reliable and valid way to assess delinquent behavior, with youth self-reports of these behaviors correlating fairly highly with official police reports (Farrington et al., 1996; Thornberry & Krohn, 2000), albeit not to a degree suggesting they are without distinctive informational value. In this study, youth reports will be supplemented with parent reports, as research (e.g., Sibley et al., 2010) indicates that at least some groups of youth may underreport delinquent behavior that is known to parents. Youth and parents will also report on risk and protective factors for crime/delinquency, allowing for a test of whether program effects are enhanced or diminished as a function of such factors. These measures, furthermore, can reflect relatively immediate or proximal sources of program impact (e.g., reduced associations with deviant peers) that set the stage for eventual impacts on crime/delinquency. To help minimize bias in reporting of delinquent behavior and associated risk and protective factors, a Certificate of Confidentiality will be obtained from the National Institutes of Health. This Certificate will allow the researchers to refuse to disclose information reported by study participants (e.g., reports of delinquent behavior) in response to legal demands, an assurance that will be highlighted orally and in writing as part of the study consent/assent process. Another potential risk for bias in reporting is that youth and/or parents in the treatment group could be prone to overstating positive outcomes due to a sense of gratitude for services received or other related reasons. To minimize this possibility, care will be taken to frame the research and its aims as broader than that of evaluating the effects of the program (i.e., as a study of social influences on youth development using the study title of "Youth Relationships Study").

At the study's inception, this Research and Analysis plan, once approved by Laura and John Arnold Foundation staff, will be posted on the Open Science Framework website.

Site Selection: One of the most important recent developments in the literature on RCTs (e.g., Green & Glasgow, 2006) has been a call for increased attention to external validity, particularly with respect to likely generalizability of results across settings, populations, and program sites. Indeed, if the aim is to inform policy at a national level, attention to this set of concerns is essential. In view of these concerns, BBBSA affiliates were selected for invitation to participate in this research with the aim of having the resulting group of affiliates be as representative as possible of the BBBSA network of affiliates in terms of size (i.e., large-, mid-, and small-sized affiliates as designated by BBBSA based on numbers of youth served) and other potentially relevant operational and performance characteristics (i.e., proportion of youth served by the affiliate who participate in the CBM program, percentage of expected mentoring relationship support contacts that are completed for youth served in the CBM program, and percentage of mentoring relationships in the CBM program that are sustained for at least one year), characteristics of youth served in the CBM program (i.e., age, gender, family structure and income, % with an incarcerated parent), age of volunteers serving as Big Brothers and Big Sisters in the CBM program, and geographic location (urban vs. non-urban; Northeast, Southwest, Southeast, Midwest, or West). All affiliate characteristics were assessed for the most recent calendar year of operation (i.e., 2015). At the time of selection, a minority of BBBSA affiliates did not make use of the organization's web-based MIS for capturing service delivery data; these affiliates were excluded from consideration as the MIS data will be an important source of information regarding program implementation (see below). An additional group of 50 affiliates operated as "sponsored" affiliates. Rather than operating as free-standing entities, sponsored affiliates operate within the umbrella of a larger organization (e.g., a YMCA). Because sponsored affiliates do not necessarily follow the same program guidelines as standard affiliates, they also were excluded from the pool of affiliates eligible for selection. Two further affiliates were excluded because at the time of site selection they served either only girls or only Jewish youth. Finally, affiliates also were excluded if they paired less than 50 youth with mentors in the CBM program during the most recent calendar year (2015). The aim was to recruit a total of 20 affiliates.

As a first step in the site selection process, a cluster analysis was conducted to divide eligible affiliates into groups based on the above-referenced factors. Specifically, three distinct groupings of affiliates were identified as follows:

- "Larger, Higher-Need-Youth-Serving Affiliates" (n= 36 affiliates; 20% of eligible pool): At the time of site selection, these affiliates tended to serve relatively greater proportions of "high-need" youth (e.g., those with an incarcerated parent) as well as male youth. These affiliates are almost all from BBBSA's mid- and large-size agency alliances and are relatively more likely to be in Southeast and West regions.

- "Northeastern, Larger Affiliates" (n=63; 35% of pool): These affiliates are predominantly those designated by BBBSA as mid- or large-size and are relatively more likely to be in the Northeastern region (although Southwest and Midwest regions are also well represented); relative to those in the above "Larger, Higher-Need" cluster, these affiliates tended to serve youth who are more typical of youth served across all affiliates.

- "Midwest, Small, Non-Urban" (n = 79; 44% of pool): These affiliates were nearly exclusively small size, relatively less likely to be in urban areas, and more likely to be in the Midwest region (although others were represented as well).

Affiliates then were randomly selected from each of these clusters proportional to its size (i.e., total number of affiliates in the cluster), with the constraint that should the level or distribution of any factor (e.g., % male) for the resulting set of affiliates differ significantly ($p < .05$) from those nationally, the most deviant affiliate on that factor was replaced with a different random selection from the same cluster. As a safeguard against lack of geographic representativeness, the maximum number of participating affiliates located within any given state was set at 40% of the affiliates in that state or 4, whichever number was larger. As site recruitment progressed and feasible study enrollment targets began to be formulated for

accepting affiliates, it became apparent that a greater number of larger affiliates would be needed to ensure that the study enrollment goal of 2,500 youth could be reached. Thus, in the later stage of site recruitment, the selection procedures were modified to increase the minimum size threshold (i.e., number of youth paired with a mentor in the CBM program in the past year) to 285, which corresponded to the 15 largest affiliates. Ultimately, 3 of these affiliates were selected and elected to participate in the study, one of these having been selected prior to this change in procedure.

Each selected affiliate was notified via email by the BBBSA Evaluation and Research Partner. Selected affiliates were provided with written materials that summarized information about the study and what would be entailed in being a study site. They also were invited to participate in an informational webinar about the study, facilitated by the research team and national BBBSA staff, as well as to have dedicated communication with the researchers and/or BBBSA staff regarding potential study involvement.

It was emphasized in all communications that involvement in the study was voluntary for all invited affiliates, with care taken to emphasize that a decision to not participate would not adversely affect their relationship with either the organization's national office or the researchers. Affiliates were also informed that they would be provided with funds to support the time of the staff member serving as a Research Liaison as well as with funds to offset costs associated with both initial screening of youth who were ultimately assigned to the study control group and staff time required to consent and administer the baseline surveys to study participants. They were informed in addition that at the end of the study each participating affiliate would receive a summary of impact estimates for their site.

Following selection and invitation of an initial group of 20 affiliates, additional affiliates were invited as needed. Ultimately, 54 affiliates were invited to participate in the study, with 17 agreeing to participate. Because projections indicated that these affiliates would be able to enroll the desired number of youth in the study (i.e., 2,500), recruitment of additional sites was not pursued, the original aim of recruiting 20 sites notwithstanding. Three of the participating affiliates were in the "Larger, Higher Need" cluster, 9 were in the "Northeastern, Larger Affiliates" cluster, and 5 were in the "Midwest, Small, Non-Urban" cluster. One affiliate ("Larger, Higher Need" cluster) subsequently withdrew from the study, prior to the start of study enrollment, due to a serious illness of the affiliate's Executive Director, leaving 16 as the final total of participating affiliates at the start of enrollment. As projections indicated that the study enrollment goal could still be reached with these remaining affiliates, a replacement for the affiliate that dropped from the study was not sought.

To accommodate varying affiliate size, each participating affiliate was given a study enrollment target based on planning discussions that took into account the affiliate's size (i.e., number of youth enrolled in the program in the previous year) and other pertinent considerations (e.g., anticipated increases or decreases in the number of youth served; the number of youth anticipated to be ineligible for the study due, for example, to age, having had a BBBS mentor previously, or a contractual obligation to serve all youth with a given set of characteristics). The study enrollment target for each affiliate was capped at a maximum of 300. This limit was set both to avoid having the burden of study participation being too high for any individual affiliate and to ensure that no affiliate accounted for such a sizable proportion of the total sample that it could be viewed as a threat to the external validity of study findings.

One of the 16 original participating affiliates experienced financial challenges and a change of executive director and subsequently withdrew from study involvement prior to enrolling any participants. Another of the original 16 affiliates similarly closed operations prior to enrolling any participants. This affiliate was small in size and served a predominately rural population; another affiliate with similar characteristics, nearby to the withdrawing affiliate, was invited to join the study and accepted. As study enrollment proceeded, another participating affiliate closed after having already begun to enroll participants in the study. This affiliate was replaced with a nearby affiliate that was assuming responsibility for families that it was in the process of serving; this affiliate also enrolled youth in the study from its own service area. Finally, in an effort to address a slower than anticipated rate of study enrollment, an affiliate serving a large urban area was added to the study with approximately one year remaining in the enrollment period. Ultimately, due to the above factors, a total of 17 sites enrolled youth into the study.

Study Enrollment, Randomization, and Baseline Data Collection (see Appendix B for flowchart): The parent or guardian ("parent") of each youth who presents to a participating affiliate within the enrollment period and meets *program* eligibility criteria (these are generally very broad, e.g., lives in the affiliate's catchment area) will be assessed for *study* eligibility. Study enrollment was initially planned to last 18 months but was increased to 24 months due to a slower than anticipated rate of enrollment. Study eligibility criteria consist of: a) youth being 10 years of age or older; b) youth not having a sibling who is already a study participant, to reduce potential for contamination that could be introduced if the youth were to be enrolled in the study and assigned to the control group, but the youth's sibling previously had been assigned to the treatment group; c) youth not having a severe learning, cognitive or other intellectual disability as reported by the parent; d) parent both speaks and reads English or Spanish; e) youth never having been matched with a Big Brother/Sister through any of the affiliate's programs in the past; f) youth not having a sibling currently receiving services from the affiliate for whom services were initiated (i.e., inquiry was made) prior to start of the study (this includes siblings who are still going through the process of program enrollment or are enrolled, but not yet matched with a Big); and g) youth not falling into a group that the affiliate is excluding from study participation based on previous agreement with the research team. Each affiliate will have the opportunity to exclude up to 4% of study-eligible youth from the research prior to consent and random assignment for any reason deemed appropriate (e.g., perceived high-need of the youth). If study eligible, the parent will be informed that they have the option to either: a) proceed with program enrollment, with the understanding that the youth being matched with a Big Brother/Sister either then or at any point over the duration of the youth's 4 years of study participation will be contingent on the youth being one of the 3 out of 4 study youth who are selected by lottery to receive services; or b) wait for up to 18 months beyond the agency's normal wait time to have program enrollment completed and thus be able to be matched with a Big Brother/Sister, in which case, the parent would be given a list of referrals to non-mentoring youth programs in the community and a tip sheet for connecting youth with supportive adults (these materials were provided to parents of all youth who enrolled in the study, with this being done for the treatment group at the direction of the overseeing Institutional Review Board). Neither study consent nor randomization to condition will be carried out at this stage for two reasons. First, parents typically seek services by phone, which is an inappropriate modality for consent, especially given the limited education level that many of the parents will have. Second, a non-trivial percentage of parents do not complete the enrollment process. Thus, randomizing at this stage would substantially weaken the test of the program's effects. Those who agree to potential study participation will meet, along with their children, with BBBS staff as they would normally as part of program enrollment. However, this process will be modified to progress only as far as necessary to determine that the youth is likely to be eligible and appropriate for the program (ineligibility applies to very few youth, for example those whose parent is unable to fulfill program responsibilities or those having a medical or psychiatric condition that would present safety concerns). At this point, assuming likely program eligibility, program staff acting on behalf of the researchers will obtain formal parent consent and youth assent for study participation.

Each affiliate will be given the option of also recruiting youth from its existing waitlist (i.e., those for whom program enrollment has been completed but who have not yet been matched with a Big Brother/Sister as well as those for whom inquiry was initiated prior to study launch, but program enrollment has not yet been completed). In these instances, a separate phone call will be made to the parent of the youth specifically for study recruitment, and consent/assent will be obtained in an in-person meeting as with new inquiries to the program.

Following consent/assent, program staff will administer baseline surveys to the parent (on paper) and youth (reading questions aloud while the youth listens and marks his or her responses on paper behind a privacy screen). Parents will also complete a brief "administration survey" to gather information that will guide and streamline the 18-month survey administration (e.g., In what format would you/your child like to complete your follow-up survey [online/paper]? In what language? Will you/your child have access to a computer? Will he/she need assistance?). Program staff will also note whether the child needed help with baseline survey administration. The parent and youth will each receive a $10 incentive for completing their baseline surveys.

Random assignment will be conducted, following the survey assessments, at the individual level. The sample allocation ratio will be 3:1 in favor of the treatment group (i.e., three times as many assigned to the treatment as the control group). This ratio was selected for both pragmatic and scientific reasons. Pragmatically, affiliate buy-in to participate in the research will be greater if only a minority of youth are assigned to the control group and thus ineligible to receive services over the course of the study; also, with fewer youth assigned to the control group, there will be less need to reimburse affiliates for lost costs associated with intake procedures for youth who are assigned to the control group. Scientifically, a larger treatment group will provide enhanced opportunities for examining the implications that different combinations of treatment-related factors (e.g., mentor characteristics, relationship quality, match length) may have for program effectiveness (Dumville et al., 2006). Siblings within the same family that enroll in the study will be assigned randomly to treatment or control group in a yoked manner (i.e., using one random assignment envelope in the procedures described below).

For purposes of operationalizing random assignment, each affiliate will be provided with a set of sealed envelopes, each of which includes notification of assignment to either the treatment or control group. Each affiliate will receive a number of envelopes equal to its targeted study enrollment number, plus 20%, with assignments in a ratio of 3 treatment to 1 control (the additional 20% are necessary to ensure that an affiliate is not able to anticipate group assignments at the tail end of its enrollment period). For quality control purposes described below, envelopes will be consecutively numbered (small number written unobtrusively on the back of the envelope). Prior to enrolling a youth into the study, the staff person involved will sign out an envelope through the affiliate's Research Liaison, recording the envelope's number and which family that envelope is assigned to. Staff will open the envelope and share the group assignment with the parent. Affiliates will record each group assignment on a secure online platform (REDCap) and researchers will monitor this information to ensure that each group assignment corresponds to the number on the envelope utilized. In addition, at the conclusion of study enrollment each affiliate will be required to return all opened envelopes (and any unused envelopes) to ensure that all are accounted for and that the number open equals the number of families enrolled.

Staff of each affiliate (including the research liaison and all staff communicating with families about the study) will be trained by the research team in all of the above described study procedures. Each program will also be provided with a detailed manual that outlines all study procedures and includes all study documents.

For each youth assigned to the treatment group, affiliates will use standard procedures to continue the enrollment process following random assignment and match the youth with a volunteer mentor. Youth assigned to the control group will not be served by the affiliate over the course of the youth's four-year participation in the study. However, each youth and his or her parent assigned to the control group will be provided with a list of non-mentoring youth-serving organizations in the community as well as a tip sheet for connecting youth with supportive adults. They also will receive an additional $50 incentive to compensate for the time invested in program enrollment and, subsequently, at both the 18-month and 4-year survey assessments, additional payments of $50 to compensate for time and other resources potentially invested in seeking alternative supports for the youth in the community (e.g., after-school program).

18-Month Survey Assessment: Each youth (and parent) will be approached for re-assessment (regardless of services received) 18 months after the youth's study enrollment, thus allowing for the possibility of capturing effects of 6 additional months of mentoring beyond the 12-month service commitment of each youth's mentor. The 18-month survey assessment will be conducted using a cost-efficient online platform in combination with paper-and-pencil completion of surveys by mail when requested or necessary, guided by the information provided in the baseline "administration" survey (see above). Specifically, each youth's 18-month survey will be completed either online or on paper. In either case, research staff will read instructions and questions to youth over the phone while the youth records her/his responses. This strategy ensures that youth are given reading support when needed. However, they are not asked to respond aloud, which can lead to positively biased responses—particularly in reports of misbehavior. Parents also will complete their surveys using the online platform or, when requested or necessary, on paper,

but without phone assistance unless indicated to be necessary or requested. The youth and parent each will receive $25 for completing their 18-month surveys.

4-Year Survey Assessment: Procedures for the 4-year survey assessment will be the same as those used for the 18-month assessment with minor modifications. Specifically, older youth will be provided the opportunity, where appropriate, to complete the survey without having questions read aloud by research staff; in addition, youth and parent will each receive a slightly higher amount, $30, for completing the 4-year survey.

Study Measures: Official delinquency and crime records will be collected from relevant juvenile and criminal (i.e., adult) court authorities, relying on statewide data repositories when available for efficiency and maximum jurisdictional coverage. Records of all offenses committed by the youth that have been referred to juvenile or criminal court (both crimes and status offenses, such as truancy) up to the point of data collection will be collected, with data collected on two occasions: after 18-month survey data collection has ended and 2.5 years later (so as to ensure availability of arrest data for a period of four years from the time of each youth's date of study enrollment). The format of this information will vary given differences in court information systems. The researchers will work with NCJJ staff to translate data into a common format. Data elements that are expected to be available consistently across jurisdiction include: a) date of arrest; b) date and type of offense; and c) date and nature of adjudication and type of sentence (if any). This approach is not without limitations, as it will exclude offenses that are handled solely by law enforcement (e.g., decline to arrest or arrest but with diversion); however, consultation with NCJJ suggests it would be very challenging to collect adequate data on such incidents due to the multiple law enforcement agencies involved in any given affiliate's catchment area as well as a lack of consistent recording of data across agencies. Questions will be included on the youth and parent surveys, however, that ask about all arrests and police contacts, adapted from the Add Health Study (Bearman et al.,1997). In some instances, it may be possible to obtain official delinquency and crime records only in de-identified form (i.e., without the ability of researchers to identify the youth to whom the data apply); in these cases, at a minimum, the data obtained will include a dichotomous indicator of arrest history as of the date of study enrollment, a dichotomous indicator of whether or not the participant had an arrest during the relevant time window (e.g., 18 months from date of baseline), and treatment/control status. To the extent that is feasible (i.e., does not allow deductive identification of participants) and is allowed by data sharing entities, additional data elements (e.g., age, gender, self-reported delinquency at study baseline) that represent planned and/or potential control variables in study analyses also will be included in the shared dataset to provide access to these variables with the returned deidentified dataset.

Delinquent behavior will be assessed using 13 items from the Add Health Study (Bearman et al., 1997). Each of these items refers to a different type of delinquent behavior (e.g., "steal something worth less than $50", "get into a serious physical fight"). Following Sibley et al. (2010), a parallel version of these items will be completed by the youth's parent. For each item, at baseline, the youth (or parent) will be asked whether he/she (or the child in case of parent) has ever engaged in the behavior described (to capture the youth's full history of delinquent behavior) and, if so, how many times (if any) in the past year (i.e., 0, 1-2, or 3 or more times). At the 18-month follow-up, the latter one-year time frame for reporting will be modified to 18 months to capture all delinquent behavior since baseline and at the 4-year assessment the time frame will be the prior 2.5 years to capture the preponderance of delinquent behavior occurring since the 18-month assessment. For study analyses, involvement in delinquent behavior between study baseline and the 18-month follow-up and during the 2.5 years prior to the 4-year assessment will be coded as yes/no dichotomous measures. The criteria for determining involvement in delinquent behavior will be based on the number and/or severity of the behaviors endorsed by the youth and/or parent. Trends in baseline responses to the delinquency items as well as exploratory factor analyses of item endorsements will be used to inform the development of these criteria. However, researchers will remain blind to treatment vs. control group status of youth in examining the baseline data for this purpose. Originally, it was planned for the criteria for delinquent behavior involvement to be decided on prior to initiation of the 18-month follow-up data collection; however, this did not occur, in part because study enrollment continued during the first several months of follow-up data collection and thus baseline data were not fully collected at the start of the follow-up. Of note, during this interim period, the researchers also conducted explorato-

ry impact analyses based on initial follow-up data (collected for approximately 20% of the sample) for a simple measure that determined delinquent behavior involvement based on whether any of the 13 items were endorsed for the 18-month period since the participant's baseline assessment. The results of these analyses, which varied somewhat depending on point of sample accrual, were generally consistent with a statistically significant effect of assignment to treatment of small to moderate magnitude. Once the full set of baseline data was available, factor analyses of these baseline data supported creating two separate indices of delinquent behavior involvement, one for property-related delinquent behavior and one for violence-related delinquent behavior. Involvement in property-related delinquent behavior will be considered present if either the parent or youth indicate, during the relevant time period (i.e., the preceding 18 months at the 18-month assessment or the preceding 2.5 years at the 4-year assessment), that the youth engaged in one or more of the following 7 behaviors (each asked about in a separate item): take something from a store without paying for it, steal something worth more than $50, go into a house or building to steal something, use or threaten to use a weapon to get something from someone, steal something worth less than $50, sell marijuana (pot) or other drugs, or drive a car without its owner's permission. Involvement in violence-related delinquent behavior will be similarly determined based on parent or youth report of the youth having engaged in any of the following 3 behaviors (again, each asked about in separate items): get into a serious physical fight, hurt someone badly enough to need bandages or care from a doctor or nurse, take part in a fight where a group of your friends was against another group. The remaining 3 items, which asked about running away, deliberately damaging someone else's property, and painting graffiti or signs on someone else's property or in a public place, were not utilized, in part because they failed to load with either of the preceding sets of items in exploratory factor analyses; additional concerns included running away being a status offense that did not fit conceptually with either set of delinquent behaviors, the possibility that painting graffiti could also have captured "sanctioned" art (e.g., public murals), and the potential for reports of deliberately damaging someone else's property to often refer to lower-level mischief, as the endorsement rate for this item was quite high relative to others. In view of the exclusion of these items, the final measures of delinquent behavior were more restrictive than the one explored in preliminary analyses (i.e., some youth regarded as having engaged in delinquent behavior on the exploratory measure may not be regarded as having done so on either of the final measures because their reported delinquent behaviors were limited to those referenced on the excluded items).

Youth will also complete a brief measure of substance use (Herrera, DuBois, & Grossman, 2013), as findings from prior research (Tierney, Grossman, & Resch, 1995) suggest this is a likely area of impact for the CBM program. The item referring simply to alcohol use without parent permission will be modified to reference drinking alcohol to the point of getting drunk (the original item is also being administered but it is not included in any hypothesized outcome measures). For purposes of study analyses, at the 18-month assessment, youth reporting any use of alcohol to the point of intoxication (drunkenness; 1 item), smoking (including use of electronic vapor products; 2 items), or illicit drug use (3 items) during the preceding 18 months will be considered to have involvement in substance use (coded as a 0/1 variable). At the 4-year assessment, in view of normative trends toward substance experimentation among older adolescents, the focus will be on youth reporting recurring substance use; specifically, youth reporting alcohol use to the point of drunkenness at least "once every week or two," illicit drug use at least "once a month," or using tobacco/vaping at least "once or twice a week" during the 6 months preceding the 4-year assessment will be considered to have involvement in substance use (again coded as a 0/1 variable). This information will be gathered at the 4-year assessment via youth report using a six-item assessment of frequency of use over the past 6 months that uses the same 6 items referenced above for the 18-month assessment, excepting the revised time frame for responses; these items will be asked only of youth who respond that they have used a given substance (i.e., as referenced in each item) in the last 2.5 years.

Contextual, behavioral, and attitudinal variables that are established risk or protective factors for delinquency/criminal behavior as well as (at the 4-year timepoint) variables that are positively associated with delinquency/criminal behavior in older adolescence will be assessed through the youth and parent surveys using brief, but well-validated measures. There is a theoretical and empirical basis to hypothesize that mentoring program participation can affect a subset of these factors. These factors will thus also be assessed at 18 months and 4 years as outcomes (see those listed for Secondary Hypotheses H1, H2, and H3 under Analysis below, with H3 focused on variables with the potential to be associated with delinquen-

cy/criminal behavior); the remainder will be included only at baseline to ensure initial equivalence between the treatment and control groups on potential influences on later crime/delinquency and to explore potential moderators of program effects.

With respect to the aforementioned constructs (i.e., risk/protective factors for delinquency/criminal behavior assessed at all time points and those associated with delinquency/criminal behavior that will be assessed at 4 years and included in Secondary Hypothesis H3), the youth survey ("YS") and parent survey ("PS") will assess: **association with deviant peers** (YS; *Association with Delinquent Peers*, Elliott et al., 1996); **school misbehavior** (PS; 3 items from Herrera et al., 2013; dichotomous measure of having been sent to the principal's office for misbehavior, received an in-school detention, and/or been suspended in the last 3 months of school vs. not; at the 4-year assessment, collected only for youth not attending college) and **truancy** (YS/PS; 2 items from Herrera et al., 2013; dichotomous measure indicating a report of having skipped class and/or an entire school day in the last 3 months of school vs. having not done so; at the 4-year assessment, collected only for youth not attending college); **depressive symptoms** (YS; Depressive Symptoms Pediatric Self-Report – Short Form from the *Patient-Reported Outcomes Measurement Information System (PROMIS)*; Irwin et al., 2010); **self-control** (YS/PS; Grasmick et al.,1993); **happiness** (YS; Positive Affect Pediatric Self-Report – Short Form from PROMIS; Forrest et al., 2017); **conventional values** (YS; Belief in the Moral Order scale of the *Communities That Care Youth Survey*; Arthur et al., 2002); **aggressive behavior** (YS/PS; YS: *Aggression Scale*; Orpinas & Frankowski, 2001; PS: *Parent's Checklist from the Fast Track Project: https://fasttrackproject.org/techrept/p/pcl/*); **social skills** (YS; Social Competencies Scale of the *Youth Outcome Measures Online Toolbox*; Muris, 2001); **coping efficacy** (YS; adapted from *Coping Efficacy Scale*; Sandler et al., 2000); **grit** (YS; *Grit Scale for Children*; Duckworth & Quinn, 2009); **goal setting and pursuit** (PS; *Goal Orientation* scale; Child Trends: https://www.childtrends.org/research/research-by-topic/positive-indicators-project/goal-orientation/), **hopeful future expectations** (YS; abbreviated version of the Hopeful Future Expectations Scale; Bowers et al., 2012); **self-esteem** (YS; Global Self-Esteem subscale of brief version of the *Self-Esteem Questionnaire*, DuBois et al., 1996; Silverthorn et al., 2017); **spark development** (YS; DuBois & Keller, 2017), **career exploration** (YS; 2 items adapted from Herrera et al., 2011; dichotomous measure of having engaged in either or both of two types of career exploration in the past 3 months vs. not); **college exploration** (YS; 1 item adapted from Herrera et al., 2011; dichotomous yes/no measure for past 3 months); **volunteering** (YS; 1 item from Herrera et al., 2013; dichotomous yes/no measure for past 3 months); **life satisfaction** (YS; WHO's 2005-06 Health Behaviors in School Age Children Survey: http://filer.uib.no/psyfa/HEMIL-senteret/HBSC/2006_Mandatory_Questionnaire.pdf); **self-advocacy** (YS; Jarjoura et al., 2017); **school engagement** (YS; Behavioral Engagement subscale of the *Engagement versus Disaffection with Learning Scale*; Skinner et al., 2009; at the 4-year assessment, collected only for youth still attending middle/high school or college); **academic success** (YS/PS; 1-item measure of the youth's grades on their most recent report card; adapted from Herrera et al., 2013); **parenting behaviors (**PS; subscales from the *Alabama Parenting Questionnaire*, a well-validated measure of aspects of parenting that are risk or protective factors for delinquency, including involvement, positive parenting, poor monitoring/supervision, and inconsistent discipline [Essau et al., 2006], at the 4-year follow-up, only asked of those parents who report that their child has lived with them at least part of the last year); **family relationships** (PS; *General Functioning scale of the Family Assessment Device*; Epstein et al., 1983; at the 4-year follow-up, only asked of those parents who report that their child has lived with them at least part of the last year); **perceived social support from family members**, **peers, and significant others** (YS; *Multidimensional Scale of Perceived Social Support*; Zimet et al., 1988; at the 4-year follow-up, the 4 items comprising the significant others subscale that refer to a "special person" were modified to specify "a special adult outside of my family" to more clearly differentiate this person from peers and family members); **involvement in out-of-school activities** (PS; Herrera et al., 2007); **receipt of mentoring**, both formal mentoring in a one-to-one or group context to assess control group exposure to non-BBBS mentoring and the presence of a non-parent "mentor-like" adult in the youth's life (YS/PS; Herrera et al., 2013); **youth's risk exposure** (PS; 31-item checklist of risk in 6 areas: mental health; academic challenges; problem behaviors; economic adversity; family risk/stress; and peer difficulties; Herrera et al., 2013). Note: Shortly after shelter-in orders began to be in effect for the COVID-19 pandemic, a question was added to the youth questionnaire asking whether the youth had been in communication with other youth during the previous week, either in-person or otherwise (e.g., online). For those that indicated they had

not been, the questions comprising the youth-reported measure of aggressive behavior (Aggression Scale) were not asked, as these use a 7-day time frame and all refer to interactions with other youth. Instances of "other students" in this scale were also changed to "other kids" to ensure that the scale continued to be relevant to youth who had recently interacted with other youth, but not in the school context.

At the 4-year assessment, the following measures will be added that are relevant to Secondary Hypothesis H3: **Suicidal ideation** (YS; dichotomous measure indicating whether youth reports suicidal ideation during the past 4 years based on response to an item adapted from the *Youth Risk Behavior Survey YRBS*); **Suicide attempt** (YS; dichotomous measure indicating whether youth reports having attempted suicide during the past 4 years based on response to item adapted from the YRBS); **Substance abuse (YS)**: 10 items adapted from the Difficulties subscale of the *Substances and Choices Scale,* Christie et al., 2007) which measures the extent to which substance use has negatively affected the youth's life (e.g., "My alcohol or drug use has led to arguments with the people I live with (family, roommates or caregivers, etc.)") and asked only of youth who reported using any substance other than tobacco or vaping of tobacco in the previous 6 months; **Sexual intercourse without a condom** (YS: dichotomous measure of condom use during last sexual intercourse, adapted from the YRBS 2021, asked only for those who report ever having had sexual intercourse); **Teen pregnancy** (YS; dichotomous measure from the YRBS 2017 adapted to ask about past 4 years, asked only for those who report having ever had sexual intercourse); **Having a sexually transmitted infection (STI)** (YS: dichotomous measure); **Perpetrating dating violence** (YS; 7-item measure of involvement in perpetrating dating violence in the past four years; four items adapted from Miller et al., 2020 and three items adapted from The National Longitudinal Study of Adolescent to Adult Health (Add Health)), asked only of those who report ever having been in a dating relationship); **Discontinuing education prior to high school diploma** (YS/PS; based on single item asking if youth has graduated from high school with discontinuation indicated by youth or parent endorsement of any of the following options (parent survey wording: No, but my child is working on his/her GED, No, but my child received his/her GED, or My child is no longer attending high school and is not currently working on their GED, and lack of discontinuation indicated by endorsement of either of the following options: No, my child is still attending middle or high school or Yes, my child graduated from high school; conflicting youth and parent responses will be treated as missing); **Meaningful engagement in post-secondary education, training, or employment** (YS/PS; a dichotomous measure based on responses to a single item, asked everyone except those currently attending middle or high school, with meaningful engagement indicated by youth or parent endorsement of any of the following options: Attended a 4-year college or university, Attended a 2-year community college, Participated in a job training or career program (including attending a program or school to get certified or licensed to do a particular type of job), Had a full-time job, Had an internship, Enlisted in the military or "Other" response indicating one of these activities based on coding blind to study condition); **College attendance** (YS/PS; a dichotomous measure based on endorsement, for the same item used for meaningful engagement, of either of the following: Attended a 4-year or 2-year college or university); **Stability of living situation** (YS/PS; a dichotomous measure based on response to a single item asking about the consistency of the youth's current housing, with unstable housing indicated by response of "Different places from week-to-week or month-to-month (e.g., hotel, motel, "couch surfing" at friends' or family's homes)" or "Other" response coded as unstable housing based on coding blind to study condition); **Occupational identity** (YS; five items adapted from the *Social Capital Assessment + Learning for Equity (SCALE) Measures,* Search Institute, 2021, assessing the degree to which the youth has a clear sense of their occupational identity); **Specific job/career goal** (YS: dichotomous measure based on youth's response to a single item adapted from Resnjanskij et al., 2021, asking whether there is a specific job or career the youth wants when they get older, and for a description of the job/career if applicable, with the latter responses needing to legitimately refer to a job or career based on review blind to study condition and non-qualifying responses treated as missing); **Extra-familial adult with whom youth can discuss their future** (YS; single item adapted from Resnjanskij et al., 2021); **Progress toward education or career goals** (YS; average of responses to 4 questions adapted from the *SCALE measures*, Search Institute, 2021, asked only of youth who report having an education or career goal, about the extent to which they are making progress toward their education or career goals); **Network support for education/career goals** (YS; average of 5 questions adapted from the *SCALE measures*, Search Institute, 2021, asked only of youth who report having an education or career goal); and **Special interest related to future job/career** (YS; dichotomous measure based on

youth's response to a single item developed for this study, asked only of youth who respond that they have a "spark" or special interest, with those who do not report a spark/special interest coded as a "no" for this measure about special interested related to future job/career).

Implementation data will be obtained from BBBSA's web-based MIS referred to above. These data will include: (1) mentor background information (e.g., race/ethnicity, occupation, prior experience as a BBBS mentor); (2) staff support contacts with mentors, youth, and parents; (3) mentoring relationship length; (4) reasons for not matching youth who remained unmatched by the 18-month and 4-year follow-ups; and (5) standardized assessments of the quality of the mentoring relationship from youth and mentor perspectives at 3 and 12 months and annually thereafter for active relationships. Information on cost per youth served will be obtained using a survey administered to program directors of participating affiliates (Herrera et al., 2007).

Strategies for Minimizing Overall and Differential Study Attrition: Attrition may involve an inability to collect 18-month and 4-year survey data as well as lack of access to official arrest records due to participants moving out of the relevant jurisdiction. Attrition will be minimized by: (1) collecting information at enrollment that will assist with tracking (e.g., detailed current and secondary contact information); (2) mailing a $10 gift card to each youth during the interim between baseline and the 18-month survey assessment and again between the 18-month and 4-year survey assessments, both as a gesture to maintain participant good will and to identify participants who have moved; (3) mailing or emailing birthday cards and annual holiday cards to participants throughout their participation in the research; (4) using monetary incentives for survey completion; (5) keeping survey assessments relatively brief (i.e., 30 minutes or less); (6) building good will with control families by providing them with cash compensation for lost time in program screening and seeking out alternative supports for youth ($150 total—see above) and, as noted above, providing them with a list of alternative programs and a tip sheet for connecting youth with supportive adults at baseline; (7) making strong efforts to collect 18-month and 4-year survey data for all participants; and (8) searching online repositories of publicly-available criminal (i.e., adult) court records using a vendor such as LexisNexis to capture offenses potentially missed through available court records (especially at the final records collection when many participants will have reached adult status).

Attrition will be monitored across both groups (both how many and which youth attrit) as the 18-month and 4-year survey assessments ensue, thus enabling efforts to minimize differential attrition. If, for example, proportionally more controls than treatments appear to be attriting, efforts will be adjusted to ensure that the remaining controls receive additional attention in follow-up efforts. Differential attrition can also occur when the *types* of youth who attrit from the two groups differ, making the sample less comparable than it may have been at baseline. The comparability of the types of controls/treatments that are attriting will be monitored by examining the baseline attributes (e.g., delinquent behavior) of the youth for whom survey and/or records data have been collected to ensure they remain similar on these attributes. If not, efforts will be adjusted to focus more on the subsample that is needed to make the groups more comparable. These strategies require providing a different kind of attention to some group members, which is not ideal because it could yield different kinds of respondents from the two groups at follow-up (the same problem the strategy is trying to counter). However, based on experience with prior similar studies, these strategies are expected to be required to, at most, a minimal extent. To the degree they are needed, there is expected to be a net benefit in terms of a reduced threat to the internal validity for estimates of program impact. In keeping with What Works Clearinghouse [WWC, 2011] standards for acceptable attrition, every effort will be made to achieve an overall attrition rate of no higher than 20% and a differential attrition rate of no more than 5%.

For arrest records, it also will be important to distinguish between youth who continue to reside in the relevant jurisdiction but have no arrests and those who have moved and thus have missing arrest data. To facilitate this determination, participant tracking information will be used with LexisNexis to determine the current address of the youth's consenting parent and thus whether the youth (most likely) continues to reside within the original jurisdictional area.

Analysis: The main study analyses will test primary and secondary hypotheses regarding effects of the opportunity to receive mentoring through the BBBSA CBM program ("program participation") as follows:

Primary H1: Program participation will decrease the likelihood that youth will have a court-related arrest for any of the following types of offenses during the 18 months after study enrollment and during the entire 4 year period of follow-up after study enrollment: person offense, property offense, drug law violation, public order offense, or status offense (i.e., binary variables indicating whether the youth has had an arrest for any of the indicated types of offenses);
Primary H2: Program participation will decrease the likelihood of youths' involvement in both property-related and violence-related delinquent behavior during the preceding 18 months at the 18-month survey assessment and during the preceding 2.5 years at the 4-year survey assessment (i.e., binary variables indicating involvement or not in each type of delinquent behavior as described previously in this plan);
Primary H3: Program participation will decrease the likelihood of youths' involvement in any substance use (i.e., any report of alcohol use to point of drunkenness, tobacco/vaping, or illicit drug use) during the preceding 18 months at the 18-month survey assessment and recurring substance use (i.e., alcohol use to the point of drunkenness "at least once every week or two", illicit drug use "at least once a month", or tobacco/vaping "at least once or twice a week") during the preceding six months at the 4-year survey assessment (i.e., binary variables indicating whether or not the youth reports substance use as defined above);

Secondary H1: Program participation will *decrease* risk factors for delinquent/criminal behavior (i.e., school misbehavior, truancy, aggressive behavior, association with deviant peers, depressive symptoms, averaging youth and parent reports after standardization when both are available for continuous measures or creating a dichotomous measure of parent or youth endorsement vs. endorsement by neither in the case of truancy) as assessed 18 months after enrollment and again 4 years after enrollment;
Secondary H2: Program participation will *increase* protective factors for delinquent/criminal behavior (i.e., conventional values, self-esteem, happiness, life satisfaction, coping efficacy, goal setting and pursuit, grit, social skills, self-advocacy, self-control, peer, parent, and significant other perceived social support, school engagement, academic success, involvement in out-of-school activities, spark development, hopeful future expectations, college exploration, career exploration, volunteering, parenting behaviors, and family functioning, again combining across reporter for the measures with both youth and parent data – i.e., self-control and academic success, the measures of each being treated as continuous measures) as assessed 18 months after enrollment and again 4 years after enrollment; and
Secondary H3: Program participation will decrease longer-term potential correlates of delinquent/criminal behavior during adolescence (i.e., suicidal ideation and attempts, sexual intercourse without a condom, teen pregnancy, having an STI, substance abuse, perpetrating dating violence, discontinuing education prior to receiving a high school diploma [for those 16 years of age or older], not being engaged in post-secondary education, training, or employment [this outcome will be examined only for those who have discontinued their education, but are not working on a GED, or have received their high school diploma or GED], college attendance, living in unstable housing, lack of occupational identity, no job/career goal, no extra-familial adult with whom the youth can discuss their future, little progress toward education or career goals, lower network support for education/career goals, no special interest related to future job/career. All measures will be dichotomous as defined earlier in this document, with the exception of substance abuse, perpetrating dating violence, occupational identity, progress toward education or career goals, education/career network strength, and lack of an extra-familial adult with whom the youth can discuss their future, which will be treated as continuous) 4 years after enrollment.

Generalized linear and nonlinear mixed-effects models (Bryk & Raudenbush, 1992; Fitzmaurice et al., 2004) will be used to test study hypotheses as these models can account for dependence among outcomes due to nesting of youth within both sites and families (i.e., siblings) as well as varying distributions of outcomes (e.g., binary, ordinal, continuous). Intercepts for sites and families within sites will be modeled as random and impact coefficients will be modeled as fixed. Covariates for tests of primary hypotheses will include youth demographics (i.e., age, gender, race/ethnicity, family structure, family income), baseline values of each primary outcome (i.e., history of court-related arrest pre-dating study enrollment, de-

linquent behavior, and substance use [presence or absence of substance use at baseline since recurring substance use was not assessed at baseline in the form it is being assessed at 4-year follow-up]), youth and parent reports of the youth's history, if any, of contacts with law enforcement not leading to arrest, and any other study measure of a potential contextual, behavioral, or attitudinal influence on criminal/delinquent behavior for which there is evidence of a non-negligible association with treatment/control group status (effect size of .05 or greater, per WWC standards, 2011). The same covariates will be used for tests of secondary hypotheses, with the addition of scores on the same measure or, where not available (as will be the case for some of the outcomes assessed at the 4-year follow-up for Secondary H3), the most closely related measure at baseline (i.e., strongest association), excluding those to be included as covariates in all analyses. Note that the analytic approach used will thus be ANCOVA (i.e., control for baseline scores on the outcome measures treated as a covariate) rather than a difference-in-differences approach (e.g., repeated-measures ANOVA).

The multiple tests associated with primary hypotheses will be conducted using the Benjamini-Hochberg (1995) family-wise adjustment (advocated by the WWC) to control for Type I error with the false discovery rate set to .05. At 18 months, the court-related arrest outcome for the period 18 months following study enrollment will serve as the primary outcome for H1, along with the survey-based assessments of delinquent behavior and substance use for H2 and H3, respectively (i.e., n = 4 tests). At 4 years, the court-related arrest outcome for the period of 4 years following study enrollment, property- and violence-related delinquent behavior for the prior 2.5 years on the 4-year follow-up survey, and substance use for the prior 6-month period on the 4-year follow-up survey will supersede these outcomes as assessed at the 18-month follow-up for purposes of family-wise Type I error adjustment (i.e., n = 4 tests). Missing data patterns will be explored and addressed using multiple imputation (Rubin, 1987) as the assumption of ignorable missingness (i.e., missing at random) is reasonable in the context of the rich set of baseline covariate measures that will be available to model the probability of missingness. For the special case of participants with de-identified arrest data (H1), the missing indicator method will be used to accommodate missingness on baseline covariates, an approach which can be expected to yield unbiased results in the context of a randomized control trial (Groenwold et al., 2012). Analyses will adhere to prevailing sets of evidence standards, particularly those of WWC and the Society for Prevention Research (SPR, Flay et al., 2005; e.g., required tests for differential attrition), thus ensuring a high level of rigor. It should be noted that planned primary analyses will, in effect, weight each site proportional to the number of participants from that site.

Estimation of power for the planned tests of program impact are based on the ultimate achieved sample size of approximately 1,350 youth, but are also dependent on a variety of considerations, some of which will be possible to determine only at the stage of data analysis. With this caveat, our estimates make allowance for nesting of data within sites, unequal size of the treatment and control groups, and across-site variation in participant numbers and conservatively assume a relatively high attrition rate of 20%, not taking into account some likely offset of power reductions due to attrition afforded by missing data imputation. For continuous outcomes (i.e., all measures examined in Secondary hypotheses H1, H2, and H3, except those assessed as dichotomous), our estimates indicate that power for tests of impact will be sufficient (.80 or greater) to detect a small treatment effect (Cohen's d = .18 if we assume covariates account for 15% of the variance in the outcome and d = .17 if we assume covariates account for 25% of the variance in the outcome). For binary outcomes (i.e., all outcomes for Primary hypotheses and outcomes assessed as dichotomous for Secondary hypotheses as specified above), assuming the same post-attrition sample size and taking into account unequal size of the treatment and control groups, there will be 80% power to detect the following illustrative population outcome differences between control and treatment groups, respectively: 5% vs. 1.9% (3.1% percentage point difference), 10% vs. 5.2% (4.8 percentage point difference), 20% vs. 12.8% (7.2 percentage point difference), 30% vs. 21.3% (8.7 percentage point difference), 40% vs. 30.5% (9.5 percentage point difference), and 50% vs. 40.2% (9.8 percentage point difference); power will necessarily be lower for those dichotomous outcomes examined in Secondary Hypothesis H3 (e.g., not being engaged in post-secondary education, training, or employment) that will not be applicable to the entire sample as indicated above. Because of a smaller than anticipated sample size, the study has less power than originally envisioned. A related consequence of the smaller sample size is that the number of participants with events (e.g., an arrest) in the tests of primary hypotheses (which all

involve dichotomous measures) will be smaller; with relatively low base rates already anticipated for some outcomes (e.g., arrest), the small number of events has the potential to introduce instability/imprecision into tests of program impact (i.e., regression coefficients could be biased in both positive and negative directions; see, e.g., Peduzzi et al., 1996, and Vittinghoff & McCulloch, 2007). (Note from an earlier version of this plan, prior to the decision to pursue a longer-term follow-up via the 4-year assessment: In part, to help address this concern, we may pursue longer-term follow-up so that the cumulative rates of key events (e.g., arrest) are higher, yielding more precise and stable estimates of program impact. In this case, the later follow-ups will supersede the earlier ones as primary.)

Primary study analyses will be supplemented with sensitivity analyses to determine the robustness of the findings obtained to varying assumptions and analytic approaches. These include: 1) modeling the program impact coefficient in each analysis as a random effect; 2) use of an alternative estimator (PLEASE) in analyses involving binary outcomes because of the potential for some loss of precision when adjusting impact estimates for covariates using a standard estimation approach (Colantuoni & Rosenblum, 2015); 3) using case deletion to account for missing data on outcomes and regression models that control for all potential variables related to the probability of missingness; 4) weighting of observations based on the extent to which study youth enrolled at each site are representative of youth served by BBBSA nationally in its CBM program; 5) weighting of observations by the inverse probability of availability of follow-up survey and/or arrest records data as a check on potential bias introduced by study attrition (Kosse et al., 2016); 6) treating primary outcomes (e.g., delinquent behavior) as continuous/multi-categorical rather than binary; 7) analyses using the undifferentiated measure of involvement in delinquent behavior that was examined for potential impacts with a portion of the sample prior to determining the final measures of delinquent behavior; and 8) where the same outcome measure is available at each of the 3 times of assessment, testing for program impact on change across all 3 time points using longitudinal growth models with growth modelled to reflect change over actual time (i.e., 0, 1.5, and 4 for baseline, 18 months, and 4 years).

Primary analyses also will be supplemented with exploratory analyses. These will include tests of: a) moderation of program effects by baseline measures of youth demographics, arrest history, severity of delinquent behavior, and profile of risk and protective factors for crime/delinquency as determined by a latent profile analysis; b) mediation of program effects on primary outcomes by changes in the risk/protective factors for delinquent/criminal behavior examined in secondary hypotheses (VanderWeele, 2015); c) treatment-on-the-treated effects in which complier average causal effects (CACE) analysis will be used to estimate program effects on primary and secondary outcomes when youth assigned to the treatment condition actually receive mentoring through the program (i.e., experience two or more meetings with an assigned mentor); and d) potential moderation of intervention impacts as a function of the extent to which the follow-up period for a given youth included shelter-in or similar restrictions that would prohibit in-person contact between youth and their mentors in the treatment arm. These analyses add considerable value to the research. In fact, they are all either required (moderation tests for subgroup differences in effects) or framed as highly desirable (mediation and program exposure measurements) in SPR's Standards of Evidence. As further exploratory analyses, program effects on each of the six types of offenses noted in H1 (i.e., person offense, property offense, drug law violation, public order offense, or status offense) will be examined separately; a parallel set of analyses will examine program effects on each of the three types of substance use involved in H2 (i.e., drunkenness, tobacco, and illicit drug use) separately, including initiation of each type of substance use from baseline.

All reports of the research will adhere to CONSORT guidelines for reporting of RCT analyses (http://www.consort-statement.org/). At the study's conclusion, the data will be de-identified and made publicly available on the Open Science Framework website along with survey instruments and any code used to clean and analyze the data.

LITERATURE CITED

Arthur, M., Hawkins, J. D., Pollard, J. A., Catalano, R. F., & Baglioni, A. J. (2002). Measuring risk and protective factors for substance use, delinquency, and other adolescent problem behaviors: The Communities That Care Youth Survey. *Evaluation Review, 26,* 575-601.

Bearman, P. S., Jones, J., & Udry, J. R. (1997). The National Longitudinal Study of Adolescent Health: Research design. Retrieved from http://www.cpc.unc.edu/addhealth.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological), 57,* 289-300.

Bowers, E. P., Geldhof, G. J., Schmid, K. L., Napolitano, C. M., Minor, K., & Lerner, J. V. (2012). Relationships with important nonparental adults and positive youth development: An examination of youth self-regulatory strengths as mediators. *Research in Human Development, 9,* 298–316. http://dx.doi.org/10.1080/15427609.2012.729911

Bryk A, S., & Raudenbush S. W. (1992). *Hierarchical linear models: Applications and data analysis methods*. Newbury Park, CA: Sage.

Christie, G., Marsh, R., Sheridan, J., Wheeler, A., Suaalii-Sauni, T., Black, S., & Butler, R. (2007). The Substances and Choices Scale (SACS)--the development and testing of a new alcohol and other drug screening and outcome measurement instrument for young people. *Addiction, 102*(9), 1390–1398. https://doi.org/10.1111/j.1360-0443.2007.01916.x

Colantuoni, E., & Rosenblum, M. (2015). Leveraging prognostic baseline variables to gain precision in randomized trials. *Statistics in Medicine, 15,* 34, 2602–2617. doi:10.1002/sim.6507

DuBois, D. L., Felner, R. D., Brand, S., Phillips, R. S. C., & Lease, A. M. (1996). Early adolescent self-esteem: A developmental-ecological framework and assessment strategy. *Journal of Research on Adolescence, 6,* 543-579.

DuBois, D. L., & Keller, T. E. (2017). Investigation of the integration of supports for youth thriving into a community-based mentoring program. *Child Development, 88*(5), 1480-1491. doi:10.1111/cdev.12887

Duckworth, A. L. & Quinn, P. D. (2009). Development and validation of the Short Grit Scale (Grit-S). *Journal of Personality Assessment, 91,* 166-174.

Dumville, J. C., Hahn, S., Miles, J. N., & Torgerson, D. J. (2006). The use of unequal randomisation ratios in clinical trials: A review. *Contemporary Clinical Trials. 27,* 1-12.

Essau, C. A., Sasagawa, S., & Frick, P. J. (2006). Psychometric properties of the Alabama Parenting Questionnaire. *Journal of Child and Family Studies, 15,* 597-616.

Epstein, N. B., Baldwin, L. M., & Bishop, D. S. (1983). The McMaster Family Assessment Device. *Journal of Marital and Family Therapy, 9*, 171-180.

Farrington, D. P., Barnes, G. C. & Lambert, S. (1996). The concentration of offending in families. *Legal and Criminological Psychology, 1,* 47-63.

Fitzmaurice, G. M., Laird, N. M., & Ware, J. H. (2004). *Applied longitudinal analysis*. Hoboken, NJ: Wiley-Interscience.

Forrest, C. B., Ravens-Sieberer, U., Devine, J., Becker, B., D., Teneralli, R. E., Moon, J., . . . Bevans, K. B. (2017). Development and evaluation of the PROMIS Pediatric Positive Affect Item Bank, Child-Report and Parent-Proxy Editions. *Journal of Happiness Studies*. Advance online publication. doi:10.1007/s10902-016-9843-9

Flay, B. R., Biglan, A., Boruch, R. F., Castro, F. G., Gottfredson, D., Kellam, E. K. Moscicki, E. K., Schinke, S., Valentine, J. C., & Ji, P. (2005). Standards of Evidence: Criteria for efficacy, effectiveness and dissemination. *Prevention Science, 6,* 151-175.

Grasmick, H. G., Tittle, C. R., Bursik, R. J., & Arneklev, B. J. (1993). Testing the core empirical implications of Gottfredson and Hirschi's general theory of crime. *Journal of Research in Crime and Delinquency, 30,* 5–29. https://doi.org/10.1177/0022427893030001002

Green, L. W., & Glasgow, R. E. (2006). Evaluating the relevance, generalization, and applicability of research issues in external validation and translation methodology. *Evaluation & the Health Professions, 29, 126-153.*

Groenwold, R. H., White, I. R., Donders, A. R., Carpenter, J. R., Altman, D. G., & Moons, K. G. (2012). Missing covariate data in clinical research: when and when not to use the missing-indicator method for analysis. *CMAJ : Canadian Medical Association journal = journal de l'Association medicale canadienne*, *184*, 1265–1269. doi:10.1503/cmaj.110977

Herrera, C., DuBois, D. L., & Grossman, J. B. (2013). *The role of risk: Mentoring experiences and outcomes for youth with varying risk profiles*. New York: A Public/Private Ventures project published by MDRC. Retrieved from http://www.mdrc.org/publication/role-risk

Herrera, C., Grossman, J. B., Kauh, T. J., Feldman, A. F., & McMaken, J. (2007). *Making a difference in schools: The Big Brothers Big Sisters School-Based Mentoring Impact Study*. Philadelphia: Public/Private Ventures.

Herrera, C., Linden, L. L., Arbreton, J. A. & Grossman, J. B. (2011). *Testing the impact of Higher Achievement's year-round out-of-school-time program on academic outcomes*. Philadelphia: Public/Private Ventures.

Irwin, D. E., Stucky, B. D., Langer, M. M., Thissen, D., DeWitt, E. M., Lai, J. S., . . . DeWalt, D. A. (2010). An item response analysis of the pediatric PROMIS anxiety and depressive symptoms scales. *Quality of Life Research, 19,* 595–607. http://dx.doi.org/10.1007/s11136-010-9619-3

Jarjoura, G. R. et al. (2017). *The evaluation of the Mentoring Enhancement Demonstration Program.* Washington, DC: American Institutes for Research. Manuscript in preparation.

Kosse, F., Deckers, T., Schildberg-Hörisch, H., & Falk, A. (2016). *The formation of prosociality: Causal evidence on the role of social environment*. Discussion Paper No. 9861. Bonn, Germany: Institute for the Study of Labor (IZA).

Miller E., Jones K.A., Culyba A.J., Paglisotti, T., Dwarakanath, N., Massof, M., Feinstein, Z., Ports, K. A., Espelage, D., Pulerwitz, J., Garg, A., Kato-Wallace, J., Abebe, K. Z. (2020). Effect of a community-based gender norms program on sexual violence perpetration by adolescent boys and young men: A cluster randomized clinical trial. *JAMA Network Open*, *3*(12), e2028499. doi:10.1001/jamanetworkopen.2020.28499

Muris, P. (2001). A brief questionnaire for measuring self-efficacy in youths. *Journal of Psychopathology and Behavioral Assessment, 23,* 145–149. http://dx.doi.org/10.1023/A:1010961119608

Orpinas P, & Frankowski R. (2001). The aggression scale: a self-report measure of aggressive behavior

for young adolescents. *Journal of Early Adolescence, 21,* 51-68.

Peduzzi P., Concato J., Kemper E., Holford, T. R., & Feinstein, A. R. (1996). A simulation study of the number of events per variable in logistic regression analysis, *Journal of Clinical Epidemiology, 49*, 1373-1379.

Resnjanskij, S., Ruhose, J., Wiederhold, S., & Woessmann, L. (2021). *Can mentoring alleviate family disadvantage in adolescence? A field experiment to improve labor-market prospects*. CESifo Working Paper No. 8870. Available at https://www.cesifo.org/en/publikationen/2021/working-paper/can-mentoring-alleviate-family-disadvantage-adolscence-field

Rubin, D. B. (1987) *Multiple imputation for nonresponse in surveys*. New York: J. Wiley & Sons.

Sandler I. N., Tein J. Y., Mehta P., Wolchik S., & Ayers T. (2000). Coping efficacy and psychological problems of children of divorce. *Child Development, 71,* 1099-118.

Search Institute (2021). *Social Capital Assessment + Learning for Equity (SCALE) Measures: User guide and technical manual.* Available at https://searchinstitute.org/resources-hub/social-capital-assessment-learning-for-equity-scale-measures

Sibley, M. H., Pelham, W. E., Molina, B. S., Waschbusch, D. A., Gnagy, E. M., Babinski, D. E., & Biswas, A. (2010). Inconsistent self-report of delinquency by adolescents and young adults with ADHD. *Journal of Abnormal Child Psychology*, *38*, 645-656.

Silverthorn, N., DuBois, D. L., Lewis, K. M., Reed, A., Bavarian, N., Day, J., . . . Flay, B. R. (2017). Effects of a school-based social-emotional and character development program on self-esteem levels and processes: A cluster-randomized controlled trial. *SAGE Open, 7*(3), 1-12. doi:10.1177/2158244017713238

Skinner, E. A., Kindermann, T. A., & Furrer, C. J. (2009). A motivational perspective on engagement and disaffection: Conceptualization and assessment of children's behavioral and emotional participation in academic activities in the classroom. *Educational and Psychological Measurement, 69,* 493-525. http://dx.doi.org/10.1177/0013164408323233

Thornberry, T. P., & Krohn, M. D. (2000). The self-report method for measuring delinquency and crime. *Criminal Justice*, *4*, 33-83.

Tierney, J. P., Grossman, J. B., & Resch, N. L. (1995). *Making a difference: An impact study of Big Brothers/Big Sisters*. Philadelphia, PA: Public/Private Ventures.

U.S. Office of the Surgeon General. (2001). *Youth violence: A report of the Surgeon General*. Rockville, MD: Author.

VanderWeele, T. (2015). *Explanation in causal inference: Methods for mediation and interaction*. New York: Oxford University Press.

Vittinghoff, E., & McCulloch, C. E. (2007). Relaxing the rule of ten events per variable in logistic and Cox regression. *American Journal of Epidemiology, 165,* 710–718. https://doi.org/10.1093/aje/kwk052

What Works Clearinghouse (2011). *Procedures and standards handbook (Version 2.1)*. Washington, DC: US Department of Education, Institute of Education Sciences. Retrieved from https://ies.ed.gov/ncee/wwc/pdf/reference_resources/wwc_procedures_v2_1_standards_handbook.pdf

Zimet, G. D., Dahlem, N. W., Zimet, S. G., & Farley, G. K. (1988). The Multidimensional Scale of Perceived Social Support. *Journal of Personality Assessment, 52,* 30-41.

http://dx.doi.org/10.1207/s15327752jpa5201_2

Appendix A: Study Timeline

Appendix B: Flowchart for Study Enrollment, Randomization, and Data Collection