Appendix A: Figures

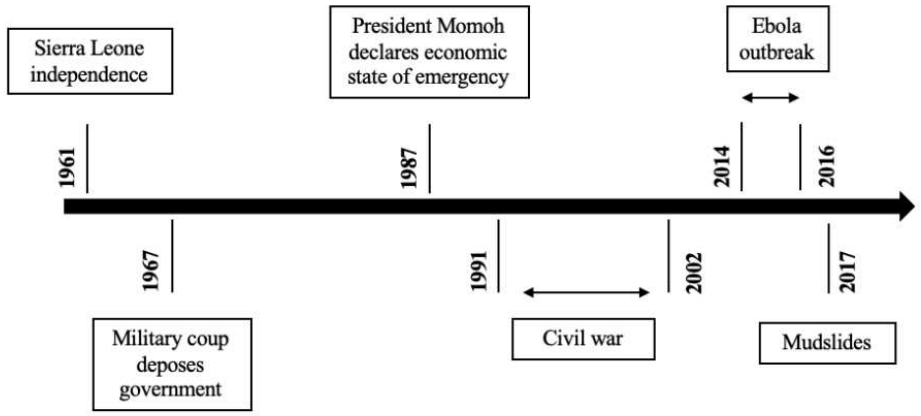

Figure: Social and political upheavals since gaining independence

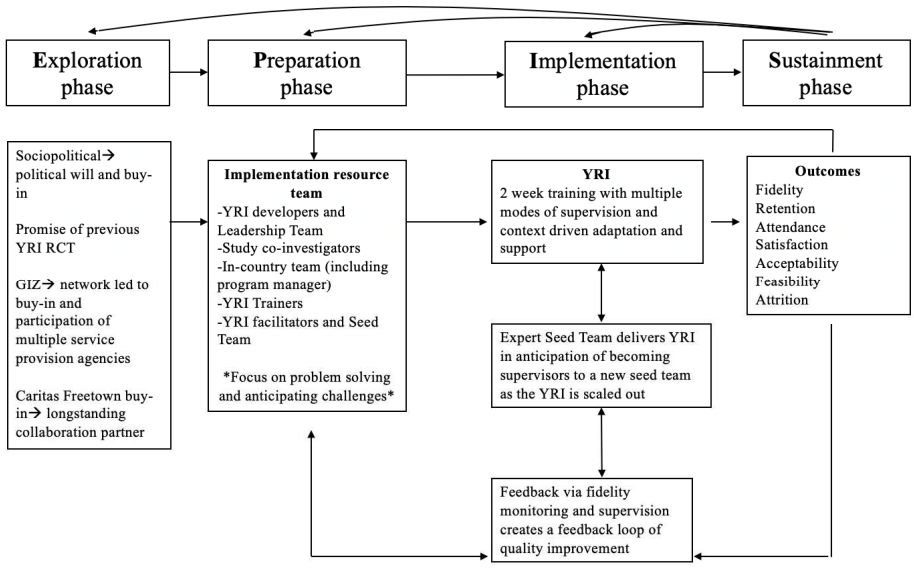

Figure: Application of the EPIS Framework in Youth FORWARD (37)

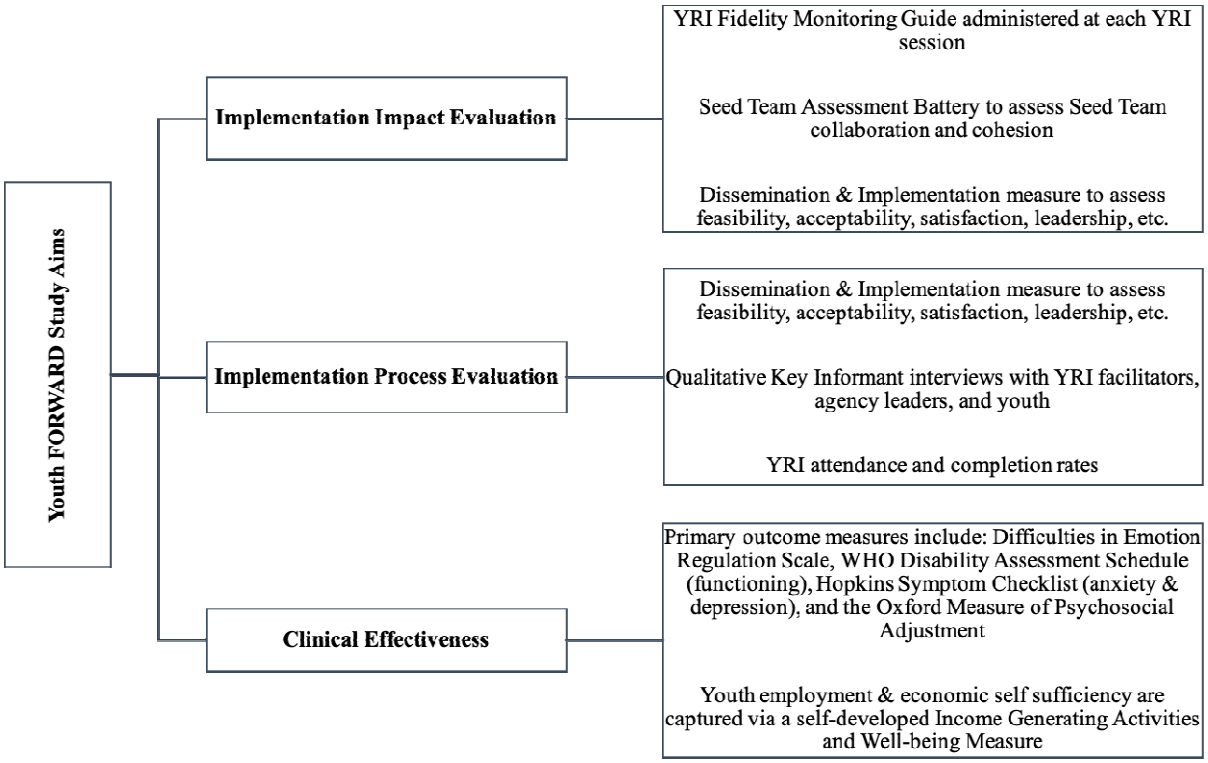

Figure: Scale-out study aims and associated primary outcome measures

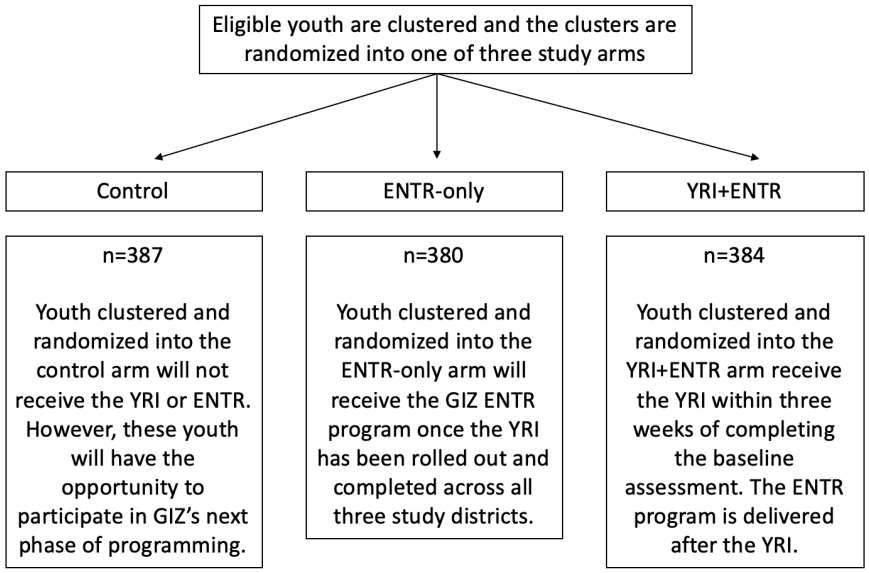

Figure: Description of study arms

Appendix B: Data Collection Plan

Construct	Time Point	Instrument/psychometrics	Respondents
Emotion Dysregulation; Daily Functioning; Coping Skills & Prosocial Attitudes; Social Support; Intimate Partner Relationships; Anxiety, Depression, Stigma & Risk Behaviors	<u>Quantitative</u> • Survey (baseline, post-YRI, post-ENTR, 12-months follow-up) <u>Qualitative</u> • Key Informant Interviews (baseline, post-ENTR, 12-months follow-up) • Focus Groups (12-months follow-up)	• Difficulties in Emotion Regulation • WHO Disability Adjustment Scale • EQ5 Health Questionnaire • Oxford Measure of Psychosocial Adjustment • Brief COPE scale • WHO Quality of Life-BREF • Responses to Stress Questionnaire • Revised Conflict Tactics Scale • Inventory of Socially Supportive Behaviors • Hopkins Symptom Checklist • Everyday Discrimination Scale • Adapted Youth Risk Behavior Survey • Post-Traumatic Stress Disorder Civilian Checklist • Daily Hardships • Goal Commitment Scale	<u>Quantitative</u> Youth (n=1151) <u>Qualitative</u> Key Informants: Youth (n=90), YRI Facilitators (n=12), Agency Leaders (n=2) Focus Groups: Youth (n=40)
Youth employment and economic self-sufficiency	<u>Quantitative</u> • Survey (baseline, post-YRI, post-ENTR, 12-months follow-up)	• Income Generating Activities and Well-Being Measure	Youth (n=1151)
Report on youth functioning and performance	<u>Quantitative</u> • Survey (baseline, post-ENTR, 12-months follow-up)	• Adapted Barkley Deficits in Executive Functioning Scale • Performance Survey adapted from classroom report used in prior YRI RCT • Teacher-Youth Rating Scale adapted from classroom report used in prior YRI RCT • Working and Training Performance Survey (self-created)	Third-party reporters (n=618)
Adoption, Acceptability, Appropriateness, Feasibility, Reach/Access	<u>Quantitative</u> • Survey (baseline, post-ENTR)	• Applied Mental Health Research Implementation Science Measure	Youth (n=764), YRI Facilitators (n=12), Agency Leaders (n=2)
YRI Fidelity	<u>Quantitative</u> Administered for every YRI session	• YRI Fidelity Rating Guide	Filled out by a YRI expert

Appendix C: Detailed Analysis

Quantitative data analysis. The in-country program manager, Caritas Freetown-based data manager, and staff at Innovations for Poverty Action (IPA) will oversee incoming data. This will allow our team to intervene immediately if there are any data entry issues or issues using Survey CTO or REDCap in the field. A team based at Boston College, led by the Research Program on Children and Adversity (RPCA), will lead quantitative data analysis.

Quantitative (clinical effectiveness) data analysis will investigate how youth and service provider outcomes differ between the YRI+ENTR group and the ENTR-only and control groups over time. A multilevel model will be used to compare YRI+ENTR subjects to those in the ENTR-only (and additionally the control group) to assess whether there is greater change in mental health and behavioral outcomes in the groups receiving the YRI. Specifically, we are looking at a change in emotion regulation. To achieve the primary aim for outcomes monitored over time, we will employ a longitudinal modeling approach by means of multilevel modeling (hierarchical linear modeling/HLM, including non-linear models). Here the null hypothesis is that the trajectories (i.e. slopes) of YRI-treated subjects over the study period—pre-assessment, post-YRI, post-ENTR, and at 12 months of follow-up—will not show greater improvement as compared to ENTR-only and control subjects, while the alternative hypothesis is that these slopes will reflect greater improvement at a $p \leq 0.05$ level of statistical significance. With four time points per individual, we will fit quadratic trajectories (if needed, cubic polynomials) using a multilevel model for continuous (or discrete, such as employment status, as necessary) outcomes.

The multilevel approach is suited for our application because we not only will view observations over time as clustered within individuals, but also will view subjects as clustered within intervention groups and those within sites. Specifically, a four-level multilevel model will be employed (level one: time-point; level two: individual; level three: intervention group, level four: site). This approach is extremely flexible and allows the use of unbalanced data, as in this case where some subjects will structurally be observed at three of four time points. It also naturally accommodates subjects who, for whatever reason, are missing data at an observation interval. In order to accommodate the fact that data collection will overlap with ongoing programming occurring the last week of YRI programming and the first week of ENTR training, we will create two new variables, one representing the number of days before the end of the YRI the interview was conducted (0 if the YRI was completed and for all subjects not receiving the YRI) and another for the number of days since the beginning of the ENTR training (0 if before the training and for the control group), which can be used to test for intervention effects.

Multilevel models are very flexible, and variables observed at any level can be used to model individual-level outcomes, including cross-level interactions and mechanisms of change. The role of specific covariates, including demographics, traumatic experiences associated with the 11-year Sierra Leonean civil war, the Ebola epidemic, or other exposures will be explored. Furthermore, for example, at the individual-level, demographic characteristics such as number of sessions attended, age and gender, as well as prior war trauma exposures will be studied as potential modifiers of intervention effectiveness. We will explore characteristics such as site location, and whether the site is rural or urban as moderators of YRI effects. We will also examine fidelity, using a fidelity checklist to score audiotaped sessions, at the treatment group level as potential mechanisms of treatment effectiveness. By scoring audiotaped sessions, we

will be able to rate how well each YRI facilitator communicates core intervention components related to intervention competence and adherence. We will also capture participant attendance rates, and participant satisfaction via our Dissemination and Implementation measures. Our measures of fidelity will be incorporated into mixed models to determine their relationship to our outcomes of interest.

We will be able to use time-varying covariates to model outcome effects over time. We will model the relationship of measured mental health and behavioral problems to employment- and employment-training-related outcomes. This will help inform the validity of the premise of linking YRI to ENTR, that mental health difficulties inhibit successful participation in both employment training and in the economy. Our detailed investigation of implementation will allow us to introduce any of a number of implementation variables, such as fidelity, into our models to test whether they are related to youth outcomes.

Qualitative Data Analysis will be informed by a four-step analytical strategy derived from Thematic Content Analysis. First, we will use open coding to examine key themes in our qualitative data (e.g. suitability of YRI components for ENTR integration, issues affecting youth attendance, support to individuals within the CTA, etc.). Separate code books will be developed for qualitative analyses at the ENTR site level, the facilitator level and the youth level of analysis, although cross-cutting themes triangulated across data source will be of additional interest. Second, we will iteratively develop a coding scheme organized by key themes. Third, two team members trained in using each codebook/coding scheme will independently code 10% of transcripts at the appropriate level of analysis to examine reliability. Poor agreement (i.e. low kappa ratings as scored in MAXQDA) will be grounds for refining the codebook or retraining. We will repeat reliability testing until coding is at 80% agreement for all data sources. These

data coding exercises will provide training opportunities for government and agency partners involved in the Capacity Building Core of the Youth FORWARD initiative, especially as multiple coders working on similar segments of data will only strengthen our code development. Once all coding schemes are operating at high reliability, in a fourth step we will code all datasets in MAXQDA using these robust code books. Results will identify key internal and external factors influencing the integration of the YRI into the ENTR. Themes specific to social support and cross-site problem solving inherent in the CTA model will be of interest as well as questions about multi-agency buy-in and flexibility in using an CTA Seed Team for training and supervision. Data will also illuminate barriers and facilitators to effective integration, as well as organizational and intervention features influencing best practices.

Qualitative and quantitative data will be synthesized and triangulated to understand barriers and facilitators to ENTR participation, integration of YRI into ENTR (AIM 2). We will examine findings from qualitative and quantitative data analyses using "joint displays" to identify areas of synergy; for example, qualitative reports of intervention content deemed useful arraying against quantitative data on factors influencing participation and feedback on the experience of integrating YRI into ENTR taken from routine supervision notes during the CRT. These methods will allow us to examine areas of convergence or divergence in the data. For example, if quantitative administrative data indicate poor attendance and we learn that ENTR training and work schedules are a major barrier to YRI participation under CTA, we might conduct additional qualitative interviews with agency leaders and youth about this topic and explore inner and outer context approaches to address this barrier. If contradictions arise, we may examine qualitative/quantitative data on hypothesized associations to establish relationships that may be

tested further. If our results are inconclusive, we may examine our quantitative data for validity or collect additional data.

Power analysis

To determine the effect of the YRI as delivered by facilitators within the ENTR framework, we assume a "small" to "medium" standardized mean difference (Cohen's d) between treatment conditions of approximately 0.3 for youth outcomes (emotion regulation, social support, daily functioning, externalizing problems, internalizing problems). An effect size of 0.3 is similar to what we observed in our prior RCT of the YRI program in the context of education programs. Table 3 provides a tabled layout of our power calculation for our study design.

Our study will have 1,151 subjects clustered and randomized into three arms i.e. 387-youth in control, 380-youth in ENTR-only and 384-youth in YRI+ENTR. Each arm will have approximately 20-clusters, for a total of 60-clusters. Our power calculations assume two sex-segregated groups of 10-participants each per site with no age stratification (i.e. each cluster will have 20-participants): a group of 10-women and a group of 10-men with no age stratification. Although our multilevel approach to modeling can retain all subjects, including those with missing data points, to accommodate the loss of precision due to loss of subjects from attrition (estimated to be a maximum of 20% at last follow-up), power is estimated accommodating 10% attrition to reflect a likely number of subjects at the midpoint of the study. Under intent to treat, all subjects initially observed will be included in all analyses regardless of their participation in ENTR or YRI+ENTR or other reasons for their loss to follow up.

At a power of .8 and a precision– level (alpha) of $p \leq 0.05$, the minimum detectable effect (MDE) (assuming one-tailed for all hypotheses which are stated in terms that YRI+ENTR will be

superior to ENTR -only or that YRI+ENTR will be *superior to* control) is .27 for comparing YRI+ ENTR to ENTR -only or YRI+ENTR to control. More common two-tailed power calculations for "different from" hypotheses (i.e. YRI+ENTR will be *different from* ENTR -only or that YRI+ENTR will be *different from* control) have slightly larger MDE of .30.

Table: Power Calculations

Total N of Sites	60
Total N of subjects (target)	1,200
N of subjects per site	20
N of subjects per group	10 (9 to account for attrition)
ICC at group level	.05
ICC at site level	.03
R^2 for covariates at site level	.10
Power	.8
MDE at alpha=.05 (two-tailed): YRI+ENTR vs ENTR or Control	.30
MDE at alpha=.05 (one-tailed) YRI+ENTR vs ENTR or Control	.27

The addition of a fourth time point would generally represent power to detect a smaller standardized effect than shown above for a three-point model; however, having only half the data at one time point will have the effect of increasing the size of the standard errors associated with the slope (i.e. making the estimates less precise) and standard power software has no way to account for this uncertainty, so it would be quite conservative to say that power to detect differences between the slopes of the ENTR-only and YRI+ENTR groups would be the same as in a three-point model. Screening failures and withdrawals will be documented throughout

recruitment and the study in each country. Drop-outs will not be replaced. We believe that attrition from the study will be minimal due to our strong partnerships in-country and experience carrying our large-scale research in Sierra Leone. Furthermore, both Caritas and IPA have substantial experience recruiting and tracking “hard to reach” subjects in challenging environments.

Using the same protocols utilized in our prior studies, we will ask the participants to give their postal box number (if they have one) and telephone number (if they have one), and names, addresses, and contact information for up to three people who will always know how to reach them. Participants will be reminded that if we are to contact the people listed, we would never discuss any details about their involvement in the study. We will use these records to track the location of study participants over time. We have effectively used these methods in other research studies, resulting in very low attrition rates. Based on our previous research, we expect attrition from baseline to the end of follow-up to be less than 20%. We will keep careful records for those who drop-out of the study and test for attrition bias based on data we will have prior to study drop-out. To the extent that such bias is present, we will limit generalizations accordingly, or, where possible, introduce statistical adjustments to address bias.