IMPROVING MEASUREMENT OF YOUTH AND YOUNG ADULT DELINQUENCY RISK
FINAL REPORT

August 2021

IMPROVING MEASUREMENT OF YOUTH AND YOUNG ADULT DELINQUENCY RISK

Final Report

August 2021

AID-OAA-M-13-00011

TABLE OF CONTENTS

TABLES AND FIGURES

ACRONYMS

ASU	Arizona State University
AUC	Area Under the Curve
C-YSET	Caribbean Youth Services Eligibility Tool
CFYR	Community, Family and Youth Resilience
ESC	Eastern and Southern Caribbean
GRYD	Gang Reduction and Youth Development
IRB	Institutional Review Board
JYS	Jamaica Youth Survey
MSE	Mean Squared Error
PPI	Poverty Probability Index
RNA	Risk and Needs Assessment
ROC	Receiver Operating Characteristic
RSS	Residual Sum of Squares
SAVRY	Structured Assessment of Violence Risk in Youth
SI	Social Impact, Inc.
USAID	United States Agency for International Development
USC	University of Southern California
Y-RISC	Youth Risk Screen
YSET	Youth Services Eligibility Tool

INTRODUCTION

Programs designed to prevent youth delinquency can only be effective if accurately targeted to at-risk youth. Recognizing the importance of targeting, many delinquency prevention programs employ risk screens designed to identify those who are most at risk and screen-out those who are not. To be effective, a risk screen must meet at least three criteria: it must be short and easy to administer; it must be easy for program administrators to grade and aggregate into a risk score; and it must accurately predict who is most at risk of falling into delinquency in the future.

This assessment seeks to address this challenge in the context of a family counseling program for at-risk youth in the Caribbean. In 2020, Social Impact (SI) concluded an impact evaluation of the United States Agency for International Development (USAID)'s Community, Family and Youth Resilience (CFYR) *Family Matters* program in Guyana, St. Lucia and St. Kitts and Nevis (Diaz-Cayeros et al., 2020). The CFYR *Family Matters* program was a violence prevention program that targeted youth between 10 and 17 years of age who had been assessed as at risk of engaging in crime and violence. The program engaged at-risk youth and their families in 14 months of structured family counseling, specifically adapted for the Caribbean context.

To target the program towards those most at risk of delinquency, program administrators used the Youth Services Eligibility Tool (YSET). The YSET was originally designed by researchers at the University of Southern California (USC) and was later adapted for use in the Caribbean. The Caribbean YSET (C-YSET) uses nine survey modules to determine risk: 1) Weak parental monitoring, 2) Critical life events, 3) Impulsive risk taking, 4) Risky group behaviors, 5) Guilt neutralization, 6) Negative peer influence, 7) Peer delinquency, 8) Self-reported delinquency, and 9) Family antisocial influence. In addition to being used to determine program eligibility, the C-YSET was used as a program evaluation tool to measure changes in youth risk levels from baseline to endline.

The Diaz-Cayeros et al. (2020) evaluation raised several concerns regarding the C-YSET. First, the evaluation found that many program beneficiaries were already well-adjusted, law-abiding youth despite being coded as "high-risk" by the C-YSET, raising questions about its reliability and accuracy. Second, at over 70 questions, the C-YSET was costly to administer and prone to resulting in respondent fatigue and associated declines in data quality. Lastly, the practice of using the C-YSET as a program evaluation tool to measure changes before versus after the program was found to be problematic, because many of the modules in the C-YSET did not map to the theory of change underlying the family counseling intervention.

To address these concerns, this report first rigorously evaluates the predictive performance of the C-YSET using longitudinal data on a sample of 2,393 potentially at-risk youth in Honduras and the Caribbean, drawing on methods from the field of forecasting and machine learning. This analysis confirms the C-YSET's poor predictive performance and reveals that only a small subset of its more than 70 variables are predictive of future delinquency. The second phase of this report develops a revised and streamlined risk assessment tool, the Youth Risk Screen (Y-RISC), which incorporates the strongest predictors from the C-YSET with a select subset of additional risk factors identified by recent research as strongly predictive of future delinquency. We then pre-test the Y-RISC on a sample of 90 young adults across five Caribbean countries to confirm that the tool is well-understood and likely to elicit truthful responses. After summarizing results from the pre-test survey and presenting the final Y-RISC tool, we present a set of guidelines for how implementers should use the Y-RISC to target their programs towards those most in

need of secondary violence prevention programming. Finally, we provide guidelines for how to expand the tool for use as both a risk screen and a pre-/post-program evaluation tool.

BACKGROUND

THE YOUTH SERVICES ELIGIBILITY TOOL (YSET)

The YSET was used for screening youth referred to secondary services as part of the Gang Reduction and Youth Development (GRYD) program, and to gauge the risk of those youth in joining gangs (Hennigan et al., 2014). Later, Creative Associates and USAID adapted the tool to be used in Honduras, El Salvador, Guyana, St. Lucia, St. Kitts and Nevis, and Tunisia.

Like the original YSET, the C-YSET includes 15 modules, nine of which measure risk factors and are used for the risk assessment, as listed below (Kraus et al., 2017). The remaining six modules are used for programmatic purposes, rather than for the risk assessment. The C-YSET and YSET instruments are proprietary; those interested in accessing the instrument should reference Hennigan et al. (2014).

Risk factor modules used to determine eligibility include:

- Antisocial tendencies
- Weak parental supervision
- Critical life events
- Impulsive risk taking
- Guilt neutralization
- Negative peer influence
- Peer delinquency
- Self-reported delinquency
- Family gang influence

In the original YSET designed for use among youth at risk of gang involvement in Los Angeles, youth were scored as "high-risk" if they scored above a certain threshold (or cut point) on four or more of the nine risk factor modules (Hennigan et al., 2014, p. 113). The risk threshold for each module was set to the median score *among those currently involved in gang behavior* (Ibid, p. 113). According to Hennigan et al. (2014), the rationale for setting thresholds for each risk factor to the median among those currently engaged in gang behavior is that this provides a straightforward way to identify who is "at-risk" with respect to a particular risk factor; the rationale for reserving the "high-risk" categorization to those who score above these thresholds on four or more risk factors is that risk is cumulative across different risk factors. As Hennigan et al. (2014) write: "high-risk is not related to the presence of one risk or another, rather it is the accumulation of multiple risks, across multiple domains, that is most clearly associated with gang joining" (p. 133).

In the Caribbean, the C-YSET was used to predict involvement in a broad range of delinquent behaviors, including violent crime, non-violent crime, substance abuse, and gang involvement. As with the original YSET, risk thresholds for each risk factor were calibrated to scores among those currently involved in

delinquent behavior. Those scoring above the threshold on four or more risk factors were categorized as 'at-risk'.[1]

In 2018, Creative Associates contracted Arizona State University (ASU) to assess the reliability and validity of the C-YSET's risk factor scales (Katz, Cheon and Zheng, 2020). The assessment found that certain risk factor modules included in the C-YSET lacked internal validity, and that as a result, the accuracy of the tool was suboptimal.

This study builds on the ASU study in several key ways. First, whereas the ASU study focused on adjusting the thresholds within each module used to classify an individual as "at-risk" for that particular module, this study develops an entirely new risk assessment tool using machine learning algorithms. This risk assessment tool is shorter, simpler to use, and potentially more accurate than the existing C-YSET. We proceed to develop a revised risk assessment tool built around the subset of C-YSET risk factors that are the strongest predictors of delinquency, plus a select number of additional scales that have been shown to be robust predictors of delinquency and are normed for use on Caribbean youth. We then validate the tool via a pretest with a sample of young adults in the Caribbean and provide a set of detailed guidelines to help practitioners use the revised risk assessment tool in the field.

STUDY OBJECTIVES

IMPROVING THE YSET AS A RISK ASSESSMENT TOOL

The first objective of this study is to address the concerns raised in Diaz-Cayeros et al. (2020) about the YSET's ability to accurately predict which adolescent youth are most at risk of delinquency when used as a tool for programmatic targeting.[2] In particular, certain properties of the YSET, including the length of the assessment, redundant or overlapping items, module reliability, and the inclusion of indirect or weak predictors of delinquency, raised doubts about the tool's ability to accurately and reliably predict delinquent behavior. In addition, the evaluation team found qualitative evidence suggesting that youth were not engaged throughout the questionnaire, and that respondents may not have comprehended some of the questions and/or answered accurately. Moreover, a risk assessment tool with a highly complex scoring system may not be effectively adopted by country governments with relatively limited experience in survey data collection and analysis methods. Lastly, because the YSET was developed more than a decade ago, it excludes risk factors identified in the recent literature as robust predictors of delinquency.

In light of these concerns, this study seeks to answer the following questions with regard to using the YSET as a risk assessment tool:

(1) What is the predictive validity of the YSET tool? How accurately can it provide a forecast of future delinquency?

(2) Of the large set of questions included in the YSET, which are the strongest predictors of delinquency? Can a small subset of questions be used as the basis for a shorter, more streamlined

---

[1] For full details on the scoring of the C-YSET, see University of Southern California (2020). Manual: Secondary Prevention, Program Eligibility & Cut Point Calibration. Los Angeles: University of Southern California.

[2] Throughout this study, we use the term "delinquency" to refer to the various types of delinquent behavior measured in the YSET tool: simple and aggravated assault, property crime, theft, and gang involvement.

risk assessment tool that will be easier to use for programmatic targeting, while still retaining enough predictive power of future delinquency?

(3) Can a relatively simple and transparent scoring methodology be proposed for determining program eligibility based on latent risk?

(4) Based on similar risk prediction tools for intervention eligibility, what additional data sources could complement eligibility decisions (e.g., clinical referral, motivational interviewing, etc.)?

VALIDATING THE Y-RISC RISK ASSESSMENT TOOL FOR USE IN ESC COUNTRIES

Building on the analysis above, this assessment develops a revised risk assessment tool, the Y-RISC, that meets three criteria: short and easy to administer; simple to score responses to generate a risk score; and accurate at predicting future delinquency.

To this end, this assessment adopts a data-driven approach to condensing the original YSET into a more streamlined Y-RISC based on the subset of original YSET variables that are the strongest predictors of future delinquency. Furthermore, recognizing that the original YSET was developed more than a decade ago for use among gang-affected communities in Los Angeles, California, this assessment adds additional modules to the Y-RISC, to measure risk factors not included in the original YSET that recent research suggests are highly predictive of delinquency across multiple cultural contexts.

The YSET has been validated for use among youth aged 12 to 18 years. However, because USAID's programming also encompasses young adults aged 18 to 29, this assessment validates the Y-RISC for use among young adults by pretesting the revised risk screen on a sample of 90 young adults from Eastern and Southern Caribbean (ESC) countries. Further details on the methodology for validating the Y-RISC are provided in the "Methodology and Data" section. The goal is to develop a single instrument that can be used across both age groups in order to simplify administration (for instance, for a program that serves youth ages 16-24, having a single risk screen to be used for all respondents avoids confusion and mistakes in instrument selection).

DEVELOPING GUIDELINES FOR USING RISK ASSESSMENT TOOLS

The third and final objective of this study is to develop guidelines for how the risk assessment tool should be used. These guidelines include how to balance the risk score with other considerations relevant to beneficiary selection, such as qualitative assessments made by social workers or program staff. Historically, assessment of violence risk has been characterized by an either-or dichotomy, based on unstructured clinical judgments versus completely structured actuarial (statistical) models, with considerable debate as to the best approach (Skeem & Monahan, 2011). In part, this determination depends on the goal of the risk assessment. We note that the Y-RISC is intended to serve as a secondary prevention tool to identify youth most likely to engage in delinquent behavior, who have had little to no contact with the justice system and would be considered a priority for inclusion in programming. Given that the goal of the revised YSET is to determine eligibility for program inclusion, there are many potential limitations of relying solely on youth self-reported data. The most obvious is when a youth does not tell the truth. For this reason, we recommend that the Y-RISC be used in tandem with an additional referral process, for instance, from parents, school counselors, or social workers. Both sources of information have their own benefits and

drawbacks which should be carefully considered by implementers when making decisions on program inclusion, which are further discussed in the "Guidelines for using the Y-RISC" section.

The study also seeks to develop guidelines for how to adapt the Y-RISC for use as a program evaluation tool rather than a risk assessment tool. The Diaz-Cayeros et al. (2020) evaluation found that the YSET was commonly used without adaptation as a program evaluation tool, to measure changes before versus after the program. However, while the YSET measures a large number of risk factors potentially linked to delinquency, not all of these are suitable to serve as outcomes in a program evaluation. For example, some risk factors are static and do not change over time (e.g., critical life events such as a family member migrating or passing away). Others are not closely linked to the theories of change underlying common interventions for at-risk youth. For example, in the context of family strengthening programs, the YSET may not be capturing intermediate outcomes beyond parental monitoring that are plausibly linked to the program's logic model and important to understand program impact. In some cases, the time frame for self-report conflicts with the duration of the intervention. For example, the delinquency self-report on the YSET asks youth to report over a 6-month period. However, a program may only last for one month, in which case the post-test data would include reporting of behavior before the program began.

With these concerns in mind, this study also proposes a slightly modified version of the Y-RISC adapted for program evaluation. We also discuss how to use a theory of change framework to decide if this modified version should be supplemented with other modules when evaluating a given program. It is expected that these guidelines will help practitioners more effectively capture program impacts when measuring changes using this tool to supplement their overall program evaluation strategy.

The culmination of this analysis is a simpler, streamlined risk assessment tool, along with a set of guidelines for social service workers, enumerators, and program implementers on how to use the tool to assess latent risk. These guidelines provide a more transparent understanding for policymakers on how to aggregate questions into a single score, and whether and when to use cut-offs to categorize individuals into risk categories versus using a simple risk ranking or ordering.

METHODOLOGY AND DATA

Validating the C-YSET's predictive performance requires *panel* data from at least two points in time. Panel data permits us to use the C-YSET scores at time $t = 0$ (baseline) to predict delinquency at time $t = 1$ (endline), and to then calculate the accuracy of these predictions by comparing them to what actually occurred at time $t = 1$.

After reviewing all of the cases where the C-YSET has been used, we identified four datasets that meet this criteria from four distinct countries: Guyana, St. Kitts and Nevis, St. Lucia, and Honduras. Data from Guyana, St. Kitts and Nevis, and St. Lucia were collected as part of an evaluation of the Community, Family and Youth Resilience (CFYR) *Family Matters* program (Diaz-Cayeros et al., 2020). Data from Honduras were collected for an evaluation of the *Proponte Más* family-based violence prevention program (Katz et al., 2019). The total sample size of the combined dataset is 2393 observations.[3]

---

[3] 50 percent (1413) of these observations are from Honduras, 18 percent (431) are from St. Lucia, 16 percent (376) are from Guyana, and the remaining 7 percent (173) are from St. Kitts and Nevis. Although the Honduras data included juveniles as young as 8 years old, significantly younger than our target population of 12- to 18-year-olds, dropping those aged 8 to 11 in Honduras does not substantively change the results reported below on the YSET's predictive power and the selection of the strongest risk factors.

We focus on predicting four measures of delinquency that are of primary interest to policymakers and practitioners:

- Violent crime: aggravated assault, group fights, simple assault, and armed robbery
- Non-violent crime: burglary, robbery, and property destruction
- Drug use: used or sold drugs
- Gang involvement: involved in or associates with a criminal gang[4]

Constituent variables for each of these outcomes come from the C-YSET's self-reported delinquency risk factor module. We merge all four datasets into a single, unified dataset, allowing us to evaluate the YSET's performance across all four countries, in addition to its performance within specific countries.

In compiling this dataset, we elect to draw on data from both control *and* treatment groups for the prediction and variable selection analyses below. We do this for several reasons. The first is sample size: by analyzing both treatment and control, we more than double our sample size, because in Honduras, which accounts for 60 percent of our sample, the data come from a performance evaluation in which all units were given the intervention. If we were to restrict our analysis to data from control groups only, we would be limited to only 483 observations, hardly enough for reliable estimates in the analysis that follows. Second, drawing on treatment data in addition to control data for the Caribbean countries would only bias our predictions if treatment were indeed a strong determinant of delinquency or if it moderated the impact of other risk factors on delinquency. Yet the Diaz-Cayeros et al. (2020) evaluation showed that the CFYR treatment had a precisely estimated null effect on self-reported delinquency.[5] Moreover, analysis of variation suggests that CFYR treatment accounts for only 0.001 of the overall variation in delinquency. Lastly, as a robustness check, we ran all of our models separately on both treatment and control groups, and verified that our core conclusions do not change as a result of this decision.[6]

EVALUATING THE YSET'S PERFORMANCE

CAN THE YSET ACCURATELY PREDICT FUTURE DELINQUENCY?

We begin with a data-driven approach to assessing the predictive power of the C-YSET. To do so, we exploit the panel structure of the data in which data were collected at baseline and endline in each of the four countries covered in this study, as described above. We use the YSET score at time $t = 0$ (baseline) to predict delinquency at time $t = 1$ (endline).

---

[4] Gang involvement was measured using a series of questions about the respondent's "group of friends", including whether they: do illegal things together, steal things that don't belong to them, get into serious fights with other groups, use weapons when fighting, sell or traffic drugs, claim an area or neighborhood as their own, challenge groups from other neighborhoods, and have a reputation for being violent. Respondents who answered affirmatively to three or more of these questions were considered to be involved or associated with a criminal gang. Under this definition, 10 percent of the sample was considered to be gang-involved at time t=0. Breaking results down by country, 7 percent were involved in gangs in Honduras, 13 percent in St. Kitts and Nevis, 10.5 percent in Guyana, and 11 percent in St. Lucia.

[5] The estimated treatment effect was -.016, with a standard error of .04.

[6] In particular, we ran the variable selection procedure described below among i) the full sample, ii) treatment only (including Honduras), and iii) control only. We found that the variables selected and identified in Figure 3 were the same across all three models (though coefficient estimates varied slightly, as we would expect for slightly different samples).

Our approach to categorizing an individual as 'at-risk' differs slightly from the approach used for the C-YSET. Whereas the C-YSET relies on thresholds for each risk factor that are unique to each country and established subjectively by researchers (as described in the Background section), we aggregate the nine YSET risk factor modules additively into a continuous risk score, then use the optimal threshold to categorize an individual as "at-risk" or not, where the optimal threshold is defined as the score at which we maximize the sum of sensitivity + specificity.[7] Defining the optimal cut score in this manner is popular in the field of data science, because it maximizes accuracy while weighting "false positives" and "false negatives" equally.[8]

Table 1: Prediction Accuracy[*] of the Caribbean YSET

	VIOLENT BEHAVIOR	DRUG USE	PROPERTY CRIME	GANG INVOLVEMENT	AVG. ACCURACY
Accuracy	58%	58%	44%	65%	56%
True + accuracy	69%	73%	41%	48%	58%
True – accuracy	48%	49%	46%	69%	53%
False + rate	52%	51%	54%	31%	47%
False – rate	31%	27%	59%	52%	42%
AUC	**0.62**	**0.65**	**0.55**	**0.61**	**0.61**

Note: Accuracy of the C-YSET score at baseline of predicting delinquency at endline.

Table 1 indicates that the C-YSET's overall accuracy rate is 56 percent on average, across all four types of delinquency. The C-YSET delivers a similarly modest "true positive" accuracy rate of 58 percent on average, indicating that it correctly predicts delinquency for 58 percent of individuals who ultimately engaged in delinquency. This modest true positive rate is matched by a relatively high average false positive rate of 47 percent, indicating that the C-YSET falsely predicts delinquency in many non-delinquents. Conversely, the false negative rate is 42 percent on average, indicating that the current scoring system fails to identify many youth who go on to engage in delinquent behaviors.

---

[7] Sensitivity is defined as the "true positive" rate, or the percentage of true delinquents who are correctly identified. Specificity is the "true negative" rate, or the percentage of non-delinquents that are correctly identified.

[8] For background on the selection of optimal thresholds for predicting binary outcomes such as "at-risk", see Fawcett (2006).

In the field of forecasting, prediction performance is most often summarized in the form of a Receiving Operating Characteristic (ROC) curve. A ROC curve plots the true positive rate against the false positive rate and traces out the two types of error as we vary the threshold at which we classify an individual as 'at-risk'. As such, ROC curves are most appropriate for continuous risk scores in which there are many possible thresholds for designating someone as 'at-risk', such as the C-YSET. The ROC curve for a high-performing model would hug the top left corner of the graph, indicating a high true positive rate and a low false positive rate. Conversely, a ROC curve for a low-performing model would run along the 45-degree line, where the true positive rate is equal to the false positive rate, indicating the model is just guessing positive cases at random.

Figure 1 plots the performance of the C-YSET at predicting violent crime, non-violent crime, drug use, and gang involvement. Across all four categories, the ROC curve is close to the 45-degree line, indicating poor performance. An important summary measure of a ROC curve is what is called the Area Under the Curve (AUC), which has a useful interpretation: the probability that, if you choose a positive case at random and a negative case at random, the model scores the positive case higher than the negative case. Thus, an AUC value of .5 is what you would get from random guessing; an AUC above .5 is better than guessing, and AUC lower than .5 is worse than guessing. The average AUC value across all four categories is .61, indicating the C-YSET performs only marginally better than random guessing.
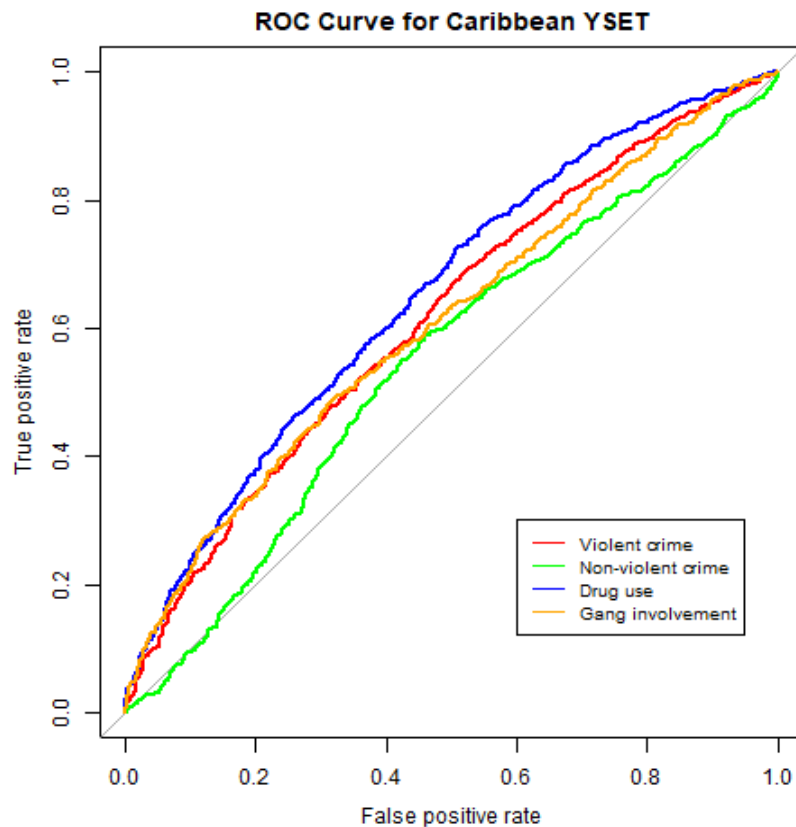
Figure 1 ROC curve of the Caribbean YSET's performance at predicting four types of delinquency: violent crime, non-violent crime, drug use / selling, and gang involvement.

Does the YSET's predictive accuracy vary across countries?

Table 2 displays the average prediction performance of the YSET across the four categories of crime by country. Average performance is poor across all four countries, with average AUC values ranging from a low of .60 to a high of .66.

Table 2: Average Prediction Performance* of the YSET across countries

	GUYANA	HONDURAS	ST. KITTS AND NEVIS	ST. LUCIA
Accuracy	66%	67%	55%	59%
True + accuracy	42%	50%	80%	61%
True – accuracy	76%	75%	44%	58%
False + rate	24%	25%	56%	42%
False – rate	58%	50%	20%	39%
AUC	**0.6**	**0.66**	**0.63**	**0.61**

* Note: Prediction performance of the C-YSET score at baseline at predicting delinquency at endline, across all four categories of crime.

Does the YSET's predictive accuracy vary by age?

We also analyzed the YSET's accuracy by age, splitting the data into two subgroups: 13 years old or younger at baseline (56 percent of our sample), and 14 to 17 years old at baseline (44 percent of our sample). We found that the YSET's accuracy was the same on average in each of these subgroups, with an average AUC of .59 across the four types of delinquency.

LIMITATIONS TO THE CURRENT YSET

We posit three potential explanations for the relatively poor performance of the C-YSET. The first is its length. At over 100 questions, the C-YSET can take up to 60 minutes to administer. In debriefs with survey enumerators during the *Family Matters* evaluation (Diaz-Cayeros et al., 2020), many enumerators reported survey fatigue by respondents and expressed their suspicions that respondents may have been "satisficing" (not thinking questions through and/or putting an effort to give accurate answers) after the first few survey modules, leading to a deterioration of data quality.

The second potential contributor to the C-YSET's poor performance is the inclusion of risk factors that have only a weak or tenuous relationship with delinquency. Including these risk factors in the overall score,

and weighting them equally with more predictive risk factors, will serve to decrease the overall accuracy of the C-YSET. By this logic, streamlining the C-YSET to include only the most predictive risk factors should improve its overall performance. We test this proposition in the next section.

Another potential limitation to the C-YSET's performance is the *exclusion* of risk factors that may be strongly predictive of delinquency. Because the original YSET was developed more than a decade ago and designed to predict gang involvement in Los Angeles, it is unlikely to reflect the latest research on the strongest risk factors for delinquency in the Caribbean.

In the ensuing sections, we attempt to address each of these concerns as we develop the Y-RISC.

IDENTIFYING THE STRONGEST PREDICTORS OF DELINQUENCY

As a first step towards developing a shorter, more streamlined risk assessment tool, we adopt a data-driven approach to identify which risk factors within the C-YSET are the strongest predictors of delinquency – and by extension, which risk factors contribute little to its predictive power and could be omitted from a revised risk assessment tool.

To do this, we draw on methods of variable selection from machine learning, specifically Lasso regression (James et al., 2013). Like linear regression, the Lasso regression model assumes each risk factor has a discrete, additive effect on delinquency. However, unlike regression, the Lasso model penalizes model complexity by "shrinking" the coefficients for weak predictors towards zero. In this way, risk factors that perform poorly are discarded, while those that are strongest receive greater weight.

Formally, the Lasso coefficients and shrinkage parameter minimize the quantity:

$$\min_{\beta,\lambda} \sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij} \right)^2 + \lambda \sum_{j=1}^{p} |\beta_j| = RSS + \sum_{j=1}^{p} |\beta_j|$$

By comparison, linear regression coefficients minimize the first term of this expression, the residual sum of squares (RSS). Higher values of the tuning parameter, λ, place a greater penalty on complexity and serve to shrink coefficients toward zero, thereby eliminating the associated variables from the model.

We identify the optimal tuning parameter λ through a process called cross-validation. Cross-validation involves dividing the data into N randomly selected subsets, training the model on $N-1$ of those subsets, testing it on the remaining subset, and lastly iterating the process N times to generate a prediction for each subset. After optimizing the tuning parameter through cross-validation, we test the performance of the optimal model on a test dataset comprising 20 percent of our total sample size.

A step-by-step outline of this process is as follows:

1. Randomly split the dataset into a training dataset (80% of the sample) and a test dataset (20%)
2. On the training dataset:
    a. For each potential value of the tuning parameter λ:
        i. Split the training sample into five groups
        ii. Fit the model on four of the five groups
        iii. Test accuracy on the fifth group
        iv. Record the Mean Squared Error (MSE)
        v. Repeat steps I to IV 500 times and calculate the average MSE
    b. Identify the tuning parameter with the highest accuracy (lowest MSE) as the optimal tuning parameter
3. On the test dataset:
    a. Fit the model using the optimal tuning parameter
    b. Observe which coefficients survive and therefore constitute the strongest risk factors
    c. Calculate the overall accuracy, true positive rate, false positive rate, true negative rate, and false negative rate

The results of this variable selection procedure are presented in Figure 2, which plots the weights assigned to each risk factor scale, or the absolute value of their Lasso coefficient. Consistent with our prior expectations, past delinquency is the strongest predictor of future delinquency. The second strongest predictor is age, indicating that older juveniles are more likely than younger juveniles to engage in delinquency. The third most important risk factor is whether the individual is out of school, which associates positively with delinquent behavior. The antisocial tendencies scale, impulsive risk-taking scale, and male gender also associate with greater delinquency, but their influence is small. Contrary to our prior expectations, the peer delinquency module receives no weight -- we discuss this surprising finding in further detail in the next section. The remaining variables also receive no weight in the optimal Lasso model.

It is important to note the important distinction between causation and correlation in this analysis: just because a variable is a strong *predictor* of delinquency does not imply that it is a *cause* of delinquency. A variable may predict delinquency for either of two reasons – because it causes delinquency, or because it is associated with other variables omitted from our analysis that cause delinquency. Because we cannot rule out the latter, the results reported in this study should not be interpreted as causal factors that could inform program design. Rather, our predictive analysis is concerned solely with the task of predicting delinquency while being agnostic as to its underlying causes.
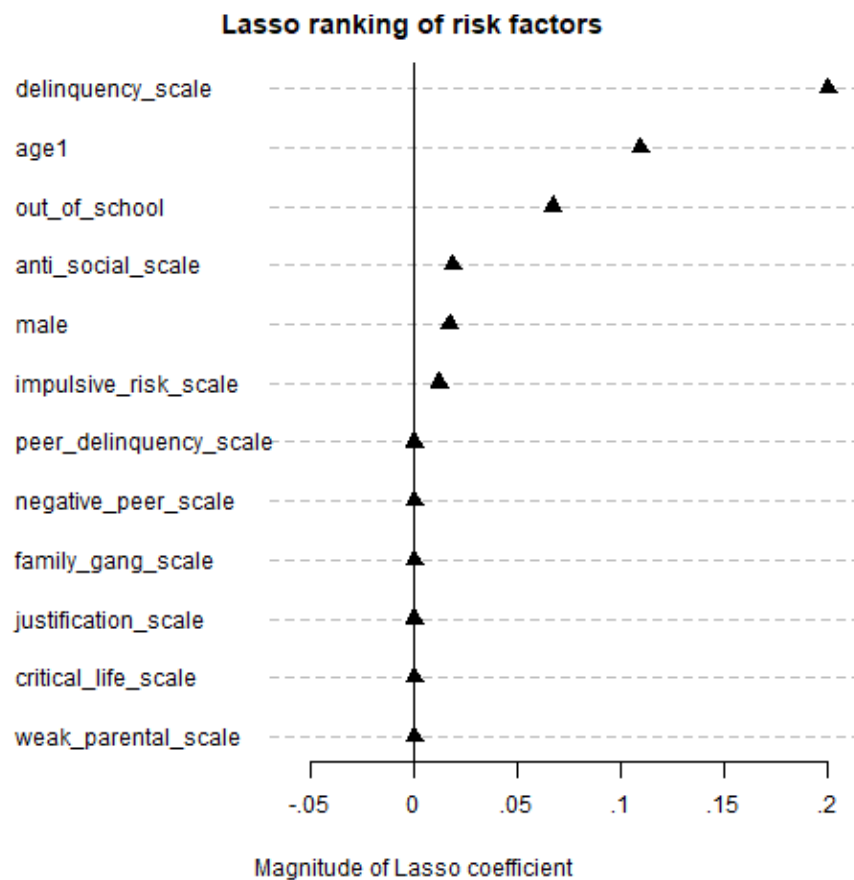
Figure 2 Lasso ranking of risk factors from the C-YSET risk screen

How does this model perform in terms of predicting future delinquency?

Table 3 presents accuracy metrics across all four categories of delinquency. On average, the Lasso model slightly outperforms the C-YSET. Average accuracy improves by four percentage points, from 56 percent in the C-YSET to 60 percent in the Lasso model. Average AUC also improves marginally by .02, from .61 to .63.

Taken as a whole, these results suggest that a simple prediction model consisting of just three risk factors – past delinquency, age, and whether enrolled in school – can slightly outperform the much more complex and time-consuming to administer C-YSET.

Table 3: Out of Sample Prediction Accuracy[*] of The Optimal Lasso Model

	VIOLENT BEHAVIOR	DRUG USE	PROPERTY CRIME	GANG INVOLVEMENT	AVG. ACCURACY
Accuracy	61%	68%	50%	60%	60%
True + accuracy	65%	69%	22%	64%	55%
True – accuracy	57%	67%	69%	59%	63%
False + rate	43%	33%	31%	41%	37%
False – rate	35%	31%	78%	36%	45%
% delinquent at endline	49%	36%	40%	21%	37%
AUC	**0.63**	**0.73**	**0.54**	**0.63**	**0.63**

* Note: Out of sample accuracy of the cross validated lasso model baseline risk factors to predict delinquency at endline.

Notwithstanding these encouraging findings, overall accuracy remains modest for both the C-YSET and the Lasso prediction models. On this point, we highlight a few important considerations. The first is the size of the challenge at hand - predicting which youth will become delinquent one to two years from now is inherently *hard*. Policymakers should remain clear-eyed about both the promise and limitations of risk screens and avoid weighting them above clinical judgment, qualitative assessments by social workers, or other criteria that may influence placement in at-risk programming. We return to this point in the "Guidelines for using the Y-RISC" section.

The second point to emphasize is that the Lasso model was only able to draw on risk factors from the set of nine risk factors measured in the C-YSET, and it remains possible that other relevant risk factors not included in the C-YSET are important and would improve accuracy. With this point in mind, the ensuing section outlines the process we took to develop a revised risk screen, the Y-RISC, which combines the

most predictive risk factors from the C-YSET – past delinquency, age, and school enrollment – with additional risk factors identified by recent research as strongly predictive of delinquency.

DEVELOPING THE Y-RISC

Recognizing that the original YSET was developed to predict gang involvement more than a decade ago for use in Los Angeles, California, this assessment does not limit itself to the risk factors measured in the original YSET. To ensure the Y-RISC reflects the latest research on youth risk factors for delinquency, we review recent research as well as other risk screens to ensure the Y-RISC includes the strongest predictors of delinquency. In addition, because USAID supports programs for both adolescents and young adults, we design the Y-RISC for use among persons aged 10-29.

A literature review of recent research on risk factors for delinquency and related risk screens can be found in Appendix I. We combine insights from this literature review with insights on which of the YSET risk factors performed best at predicting future delinquency from the analysis above to develop the Y-RISC. Table 4 provides the list of Y-RISC modules, and the source and rationale for including each module.

The YSET risk factor modules that performed best in the Lasso analysis for predicting future delinquency were past delinquency, age, and school status (in or out of school). These YSET questions are included in the Y-RISC, in addition to other basic demographic questions such as country and ethnicity. With the exception of the peer delinquency module, the other YSET modules were dropped from the revised tool on account of low predictive power.[9] Although the peer delinquency module did not perform well in the Lasso analysis, we elected to include it in the Y-RISC because risky peer behavior is a leading risk factor according to the literature. In our view, the results of the Lasso analysis do not outweigh the large and robust literature highlighting the influence of peer delinquency on an individual's delinquent behavior.[10]

Table 4: Y-RISC modules, sources, and rationale for inclusion

MODULE	SOURCE FOR QUESTIONS	RATIONALE FOR INCLUSION
Demographics	YSET	Standard demographic questions
Productive engagement	Jamaica Youth Survey	Lacking productive engagement is an important risk factor identified in the literature, and it was missing from the YSET beyond one question about working and one about school. This module adds questions on getting training, volunteering, and helping in a family business.

---

[9] Although the antisocial scale and the impulsive risk scale had non-zero coefficients, they were exceedingly small (<.03), and it was determined that their contribution to predictive power did not justify their inclusion in the revised risk screen.

[10] In our view, there are several plausible explanations for this seemingly counterintuitive result. One possibility is that the peer delinquency module's position relatively late in the YSET survey led to respondent fatigue and "satisficing", reducing the quality of the resulting data. Another possibility is that peer delinquency's contribution to future delinquency is captured by past delinquency, which is already included in the model. And lastly, the non-association of peer delinquency and future delinquency could simply be a chance artifact of the data that would not replicate if we collected data from a new sample.

Poverty	Adapted from the Poverty Probability Index (PPI)[11]	Important risk factor identified in the literature that was missing from the YSET.
Historical factors	SAVRY	Important risk factor identified in the literature that was missing from the YSET.
Decision-making	Jamaica Youth Survey[12]	Important risk factor identified in the literature that was missing from the YSET.
Delinquency	YSET	This module had the best predictive power for future delinquency in the YSET assessment. We removed some redundant questions to make the module shorter.[13]
Social support	Jamaica Youth Survey[14]	Important risk factor identified in the literature that was missing from the YSET.
Peer delinquency	YSET	Although it does not perform well in the Lasso analysis, it is an important risk factor identified in the literature.

Many of the new questions and modules we added to this survey come from the Jamaica Youth Survey (JYS). This is the only survey we found that was used for identifying youth at risk of engaging in risky and violent behavior and was validated in a Caribbean country.

We included JYS's "productive engagement" module because both the literature and the findings from the Lasso analysis above suggest that various forms of productive engagement, including but not limited to school attendance, are important protective factors against delinquency. Importantly, the JYS productive engagement module goes beyond just whether the respondent is in school or working, to also measure other kinds of productive activities they may engage in that would reduce their risk, such as volunteering, getting training, and helping out with the family business.

The JYS "social support" module measures the extent to which respondents have adult mentors, support networks, and are aware of social and work opportunities around them. Social support, or lack thereof, features prominently in the literature on adolescent and young adult delinquency risk. The "Decision-making" module checks whether the respondent thinks of alternative options and the consequences of their actions, for themselves and others, before acting (Meeks-Gardner et al., 2011).

---

[11] For more information on the PPI, see: https://www.povertyindex.org/

[12] These items were adapted from the Cognitive Autonomy and Self-Evaluation (CASE) Inventory (Beckert, 2007).

[13] Based on our subjective assessment of redundant questions, we combined "In the past 6 months, have you been in fights" and "In the past 6 months, have you been in serious group fights" into just "have you been in serious fights where someone got injured?"; we also dropped "In the last 6 months, have you broken into a building to steal something?" because it was redundant with "In the last 6 months, have you stolen or tried to steal something valuable?".

[14] These items were adapted from two measures: The Multidimensional Scale of Perceived Social Support (Canty-Mitchell & Zimet, 2000) and the Presence of Caring Scale (Springer & Phillips, 1992).

Our literature review indicated that historical factors related to an individual's home environment during their childhood are critical determinants of delinquency. To capture this, we draw on questions from the Structured Assessment of Violence Risk in Youth (SAVRY) tool.[15]

Poverty is a consistent correlate of delinquency in the academic literature. Not owning a refrigerator was highly predictive of delinquency in the Jamaica study that used the JYS (Meeks-Gardner et al., 2011). To identify questions that could serve as a reliable indicator of whether the family falls below the poverty line, we looked at the Poverty Probability Index (PPI) for countries that are in the Latin America and Caribbean region, namely the Dominican Republic, El Salvador, Guatemala, Haiti, Honduras, and Paraguay.[16] The questions we selected were strongly predictive of poverty in these countries, and include items such as owning assets like an automobile or refrigerator, and how many rooms the house has.

The final Y-RISC tool, including all modules, questions and answer options, can be found in Table 5. The tool includes 8 modules and 39 questions. Thirty-five items are used for the risk scoring, and the remaining four provide demographic information that does not get scored but should be useful for program administrators. In comparison with the JYS, the Y-RISC is much shorter, has simpler answer options, is easier to score, and takes advantage of new knowledge from research published in the last decade and from our pre-testing.

To ensure the risk assessment tool can be deployed and administered without specialized expertise, we designed the Y-RISC scoring scheme to be as simple as possible. Each of the 36 scoring questions receives an equal weight, with 1 point assigned for the response options associated with delinquency and 0 points assigned for all other responses. The answer options that get 1 point added are marked with 1 in the "Scoring" column. All other answer options do not add any points to the risk score.

Note that under this scoring scheme, the length of a given module dictates how much weight is assigned to that risk factor in the overall score. For example, the past delinquency module has eight questions worth a total of eight points, whereas the social support module has five questions worth a total of five points. The relative length of each section is therefore deliberate and accords with our subjective assessment of the relative predictive power of each risk factor.[17]

---

[15] More information on the SAVRY can be found in the literature review in Appendix I.

[16] For more information on the PPI, see: https://www.povertyindex.org/

[17] That the relative weighting of risk factors is not validated empirically is an important limitation of our study. This limitation highlights the need to more fully validate the Y-RISC using panel data among a sample of potentially at-risk youth. Such data would allow us to assess empirically the relative importance of each set of risk factors (and constituent variables), and to then use this information to inform the weighting assigned to each question. See Kshirsagar et al. (2017) for further details on how this could be done.

Table 5 Y-RISC tool with all modules, questions, and scoring

Question or module name	Question text	Answer options	Scoring
Demographics			
demo_country	In which country do you currently reside?	[Enter country options]	
demo_age	What is your age?[18]		
demo_gender	What is your gender?	Male	1
		Female	
		Nonbinary	
		Other	
demo_ethnicity	What is your race or ethnicity?	[Country specific. Allow for multiple choice]	
Productive engagement	*The following is a list of activities of how young people spend most of their time. Please tell me which you are doing right now:*		
school	In school? Either full time or part time.	Yes	
		No	1
training	Getting job training, vocational training, or completing an internship/apprenticeship?	Yes	
		No	1
work	Working full or part time and/or self-employed?	Yes	
		No	1
family_business	Helping in a family business without pay?	Yes	
		No	1
volunteering	Volunteering or working for the community (without pay)?	Yes	
		No	1
Poverty	*Next, I will ask a few questions about your home. If you live in more than one place, please answer about your primary home.*		
poverty1	Does the place where you live have a refrigerator?	Yes	
		No	1
		Don't know	
		Not applicable	
poverty2		Yes	1

---

[18] Although Age was an important predictor of delinquency in the YSET data, we don't use it for scoring in the Y-RISC because the Y-RISC is designed for youth aged 10 to 29, rather than 10 to 17 in the YSET. Accordingly, we have no empirical basis for assigning scores to those who are over the age of 18.

	Does the place where you live have fewer than three rooms? Counting any kind of room, such as kitchen, bedroom, bathroom, etc.	No	
		Don't know	
		Not applicable	
poverty3	Does anyone in your household own a car or truck?	Yes	
		No	I
		Don't know	
		Not applicable	
Historical factors	*This next section is about how your life was when you were a child. Please tell me yes or no for the following statements:*		
historical_factors1	I saw a lot of violence around me in my home.	Yes	I
		No	
		Refuse to answer	I
historical_factors2	I saw a lot of violence around me in the neighborhood I lived in.	Yes	I
		No	
		Refuse to answer	I
historical_factors3	The people who raised me got in trouble with the law.	Yes	I
		No	
		Refuse to answer	I
historical_factors4	As a kid, I always did well in school and got good grades.	Yes	
		No	I
		Refuse to answer	
historical_factors5	When I was a young kid, my home life was comforting and secure.	Yes	
		No	I
		Refuse to answer	
Decision-making	*These next questions ask about making decisions, that is, the things that you make up your mind about. Please tell me how often you do these things:*		
decision_making1	When I am making up my mind about something important, I think about all the things that could happen (the consequences).	Always	
		Often	
		Sometimes	I
		Never	
decision_making2	I consider different choices before making up my mind about something.	Always	
		Often	
		Sometimes	I

		Never	
decision_making3	I think about how the things I do will affect others.	Always	
		Often	
		Sometimes	I
		Never	
Delinquency	*People sometimes break rules or laws. Some of the questions in this section ask about the rules or laws you may have broken in the last year. Remember, your answers will stay private and will not be shared with anyone outside the research team.*		
delinquency1	In the last year, have you used marijuana or other illegal drugs? Not including alcohol.	Yes	I
		No	
		Refuse to answer	I
delinquency2	In the last year, have you sold illegal drugs?	Yes	I
		No	
		Refuse to answer	I
delinquency3	In the last year, have you purposely damaged or destroyed property that did not belong to you?	Yes	I
		No	
		Refuse to answer	I
delinquency4	In the last year, have you carried a knife, gun, or other weapon for protection?	Yes	I
		No	
		Refuse to answer	I
delinquency5	In the last year, have you stolen or tried to steal something valuable?	Yes	I
		No	
		Refuse to answer	I
delinquency6	In the last year, have you hit someone with the purpose of hurting them?	Yes	I
		No	
		Refuse to answer	I
delinquency7	In the last year, have you attacked someone with a weapon?	Yes	I
		No	
		Refuse to answer	I
delinquency8	In the last year, have you been involved in fights where someone got seriously injured?	Yes	I
		No	
		Refuse to answer	I

Social support	*This next set of questions asks about your relationship with others. Please tell me yes or no for the following statements:*		
social_support1	There are people I can depend on to help me if I really need it.	Yes	
		No	I
social_support2	I have an adult I can turn to for help when I am worried about something or have a problem.	Yes	
		No	I
social_support3	There is a special person in my life who cares about my feelings.	Yes	
		No	I
social_support4	I know where to go if I need advice about something.	Yes	
		No	I
social_support5	I know about opportunities available to me (such as jobs, trainings, religious activities, sports, clubs, etc.)	Yes	
		No	I
Peer delinquency	*This next set of questions asks about whether your friends are getting into trouble. But we don't want to know who specifically is getting into trouble, and we don't want to know anyone's name. Remember, this information is only for research and will not be shared with anyone outside the research team.*		
peer_delinquency1	In the last year, how many of your close friends used marijuana or other illegal drugs? Not including alcohol.	None	
		A few	
		A lot	I
		Refuse to answer	
peer_delinquency2	In the last year, how many of your close friends sold drugs?	None	
		A few	
		A lot	I
		Refuse to answer	
peer_delinquency3	In the last year, how many of your close friends purposely damaged or destroyed property that did not belong to them?	None	
		A few	
		A lot	I
		Refuse to answer	
peer_delinquency4	In the last year, how many of your close friends stole or tried to steal something valuable?	None	
		A few	
		A lot	I

		Refuse to answer	
peer_delinquency5	In the last year, how many of your close friends attacked someone with a weapon?	None	
		A few	
		A lot	1
		Refuse to answer	
peer_delinquency6	In the last year, how many of your close friends have been involved in fights where someone got seriously injured?	None	
		A few	
		A lot	1
		Refuse to answer	
TOTAL SCORE (Valid range is from **0-36**)			

VALIDATING THE Y-RISC

PRE-TESTING

To validate the Y-RISC, we pre-tested the instrument on a sample of 90 young adults aged 18-28 from Guyana, St. Lucia, St. Kitts and Nevis, St. Vincent and the Grenadines, and Antigua and Barbuda. In the first three countries' lists, the sample came from the same families who were contacted for the administration of the C-YSET for the Diaz-Cayeros et al. (2020) evaluation. For the latter two countries, the sample was obtained from USAID and it was young adults who had participated in relevant programming. When enumerators called and a YSET family caregiver answered, the enumerators asked if there was a young adult aged 18-29 in the family, and how to best reach them. We pretested the survey on young adults, rather than adolescents below the age of 18, because the surveys we draw from have already been validated for use among adolescents (see Diaz-Cayeros et al., (2020) and Katz et al., (2019)).

Oftentimes the respondent was one of the participants from the Diaz-Cayeros et al. (2020) evaluation, who completed the C-YSET as an adolescent in 2019 but has since turned 18; in other instances, we surveyed one of their older siblings. We focused on this sample of families with prior exposure to the USAID program to ensure the sample was similar to the type of population where USAID is likely to target programming in the future. The decision was also expedient, insofar as we already had the contact information for this sample and could survey them over the phone in accordance with COVID-19 safety protocols.

Within the survey, we included three sources of feedback. First, after each survey module, enumerators were prompted to note if there were problems along the following categories before they continued to the next module: 1-Confusion, 2-Discomfort, 3-Potential deceit, 4-Other, 5-No problem. These questions served as a quick checklist to make sure that any problems that arose were marked. Second, after the

survey was completed, enumerators asked a set of closed and open-ended feedback questions. And lastly, in 32 of the 90 surveys, we passively recorded key survey modules and the post-survey feedback session so that we could have a sense of respondents' tone and/or levels of hesitancy when answering questions.[19]

IRB approval for the survey was obtained from Social Impact's internal IRB Board. Participants were offered a phone credit worth 5 USD for their participation in the survey. On average, the Y-RISC component of the survey lasted about 15 minutes. Table 6 includes the number of surveys conducted per country.

Table 6: Number of Y-RISC surveys completed

COUNTRY	NUMBER OF SURVEYS
Guyana	30
St. Lucia	22
St. Kitts and Nevis	30
St. Vincent and the Grenadines	6
Antigua and Barbuda	2
TOTAL	**90**

---

[19] Unfortunately, due to technical limitations with our survey software (SurveyCTO) and compatibility across Android and iOS platforms, only three of our six enumerators were able to activate recordings. Accordingly, we only have audio records for 32 of the 90 surveys.

Table 7 summarizes the demographic information on Y-RISC respondents. The age of respondents ranged from 18 to 28 years old, and the mean was 20.2. The sample had slightly more females (53 percent) than males (47 percent). Most of the respondents identified as Black/African descendent/Afro-Caribbean/Afro-Guyanese. The ethnicity categories varied across countries and were aggregated for the purpose of this analysis.

Table 7: Y-RISC respondents' demographic information

	SUMMARY STATISTICS
AGE	
Mean	20.2
Minimum	18
Maximum	28
GENDER	
Male	47%
Female	53%
ETHNICITY	
Black/African descent/Afro-Caribbean/Afro-Guyanese	82%
East Indian/Indo-Caribbean/Indo-Guyanese	3%
Amerindian/Caribbean	2%
Other	12%

PRE-TEST ENUMERATOR FEEDBACK

Table 8 summarizes the data from the enumerator feedback checklist administered after each survey module, showing the percentage of surveys by module where a problem was identified by the enumerators. Not unsurprisingly, a small number of respondents were uncomfortable with some of the questions, particularly around their childhood and violence (historical factors), drug use, delinquency, whether they have people they can count on (social support), and their friends breaking laws (peer delinquency). Enumerators felt that some respondents may have lied in the delinquency and peer delinquency modules because they noticed hesitation before answering. The "Other" problem in the demographics section was due to some respondents in St. Lucia not immediately understanding the word "ethnicity". In response, we have added the word "race" to clarify the meaning of that question. The other issues raised in the "Other" category were mostly about the respondents' hesitating or thinking through before answering.

Table 8: Percent of surveys where enumerators identified problems with modules

MODULE	CONFUSION	DISCOMFORT	POTENTIAL DECEIT	OTHER	NO PROBLEM
Demographics	0%	0%	0%	6%	94%
Productive engagement	0%	1%	0%	0%	99%
Poverty	0%	1%	0%	0%	99%
Historical factors	1%	7%	1%	3%	88%
Decision-making	0%	0%	0%	1%	99%
Delinquency	1%	4%	9%	2%	89%
Social support	0%	2%	0%	1%	97%
Peer delinquency	0%	8%	7%	0%	89%

In addition, after the survey was completed, enumerators were asked more thorough questions regarding their perceived level of honesty and discomfort on the part of the respondents, as well as whether the respondent grew tired towards the end of the interview. The full list of questions asked of enumerators can be found in Appendix II. The enumerators felt that respondents were completely or mostly honest in 95 percent of the surveys. In 8 percent of the surveys, they reported that the respondent seemed to grow tired towards the end.

PRE-TEST RESPONDENT FEEDBACK

Following the main part of the survey, we asked respondents a set of closed and open-ended feedback questions. The full list of feedback questions can be found in Appendix II. In Figure 3 we see that a small share of respondents, 6 percent, said that there were questions that were confusing or difficult to answer. A couple of respondents mentioned the question on ethnicity, one said they did now know what types of crimes their peers might commit, and one mentioned difficulty answering questions about their social support network. The vast majority of respondents (98 percent) reported being able to stay focused during the entire interview.

Were there questions that were difficult to answer?

6%

94%

■ No ■ Yes

Were you able to stay focused during the entire interview?

2%

98%

■ No ■ Yes
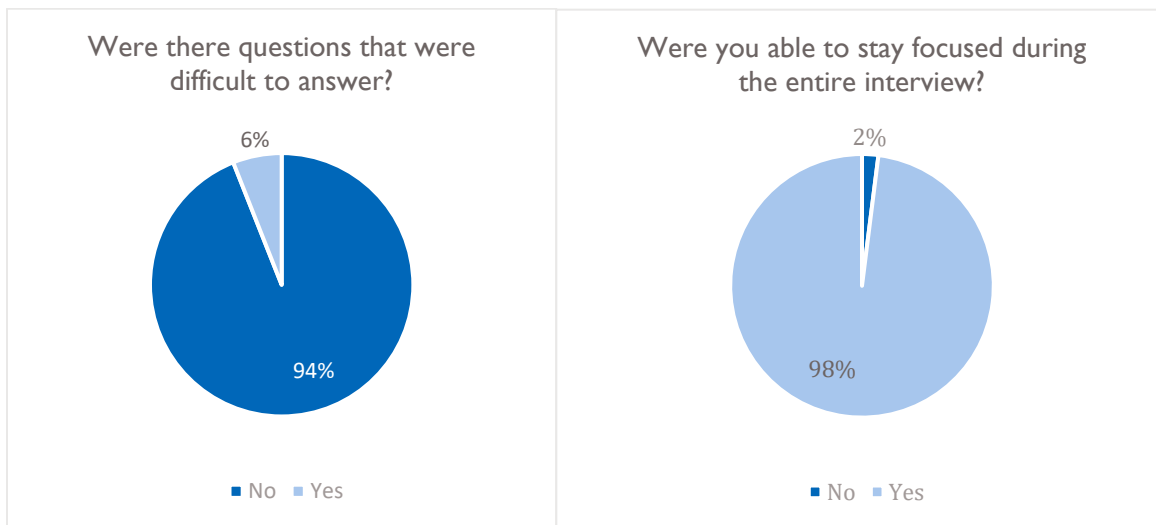
Figure 3: Respondent feedback on question difficulty and ability to stay focused

To get a sense for how truthful the respondents were during the survey, we first asked an indirect question: "Were there questions that you think other respondents would probably lie about?" As shown in Figure 4, most participants, 79 percent, said yes. The most commonly cited questions that they think others would lie about include drugs, involvement in crimes (often mentioning violence and weapons in their answers), and their friends' delinquency. Our key informant interviews revealed that the questions about self-reported delinquency and peer delinquency made the respondents most uncomfortable and likely to lie, because they were worried about reporting their own crimes or snitching on their friends.
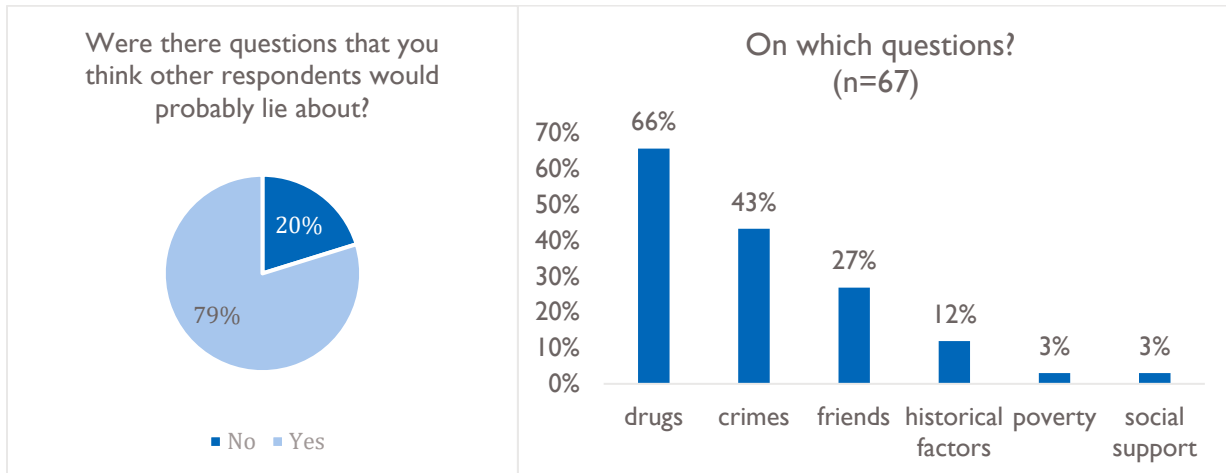
Figure 4: Questions the respondents believe others would lie about

And yet, despite the skepticism about whether others would be willing to admit to past or peer delinquency, a large number of respondents did admit to these actions in the self and peer delinquency modules. For example, as shown in Figure 5, 40 percent of respondents admitted to using illegal drugs in the past year, 21 percent admitted to getting in fights, 14 percent admitted to carrying a gun or knife for protection, 10 percent admitted to destroying property, and 9 percent to selling drugs. In terms of peer delinquency (see Figure 6), 61 percent reported that their friends used illegal drugs, 27 percent said their friends sold drugs, and 17 percent said that some of their friends committed aggravated assault in the past year.
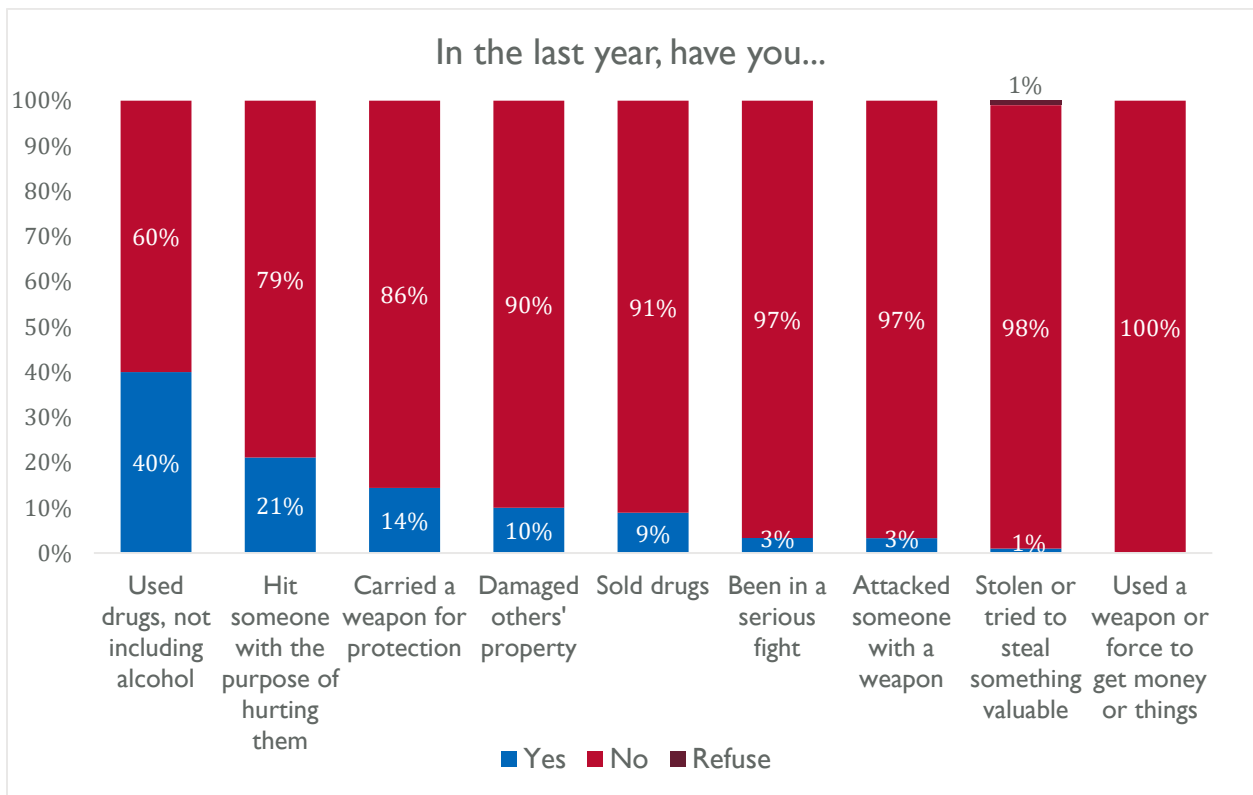
Figure 5: Distribution of answers for questions in the delinquency module

In the last year, how many of your close friends have...

	None	A few/Half	Most/All	Refuse
Used drugs, not including alcohol	38%	47%	16%	
Sold drugs	71%	24%	3%	1%
Been in a serious fight	72%	24%	2%	1%
Attacked someone with a weapon	83%	16%		1%
Damaged others' property	86%	13%		1%
Stolen or tried to steal something valuable	88%	10%		
Used a weapon or force to get money or things	94%	4%		1%
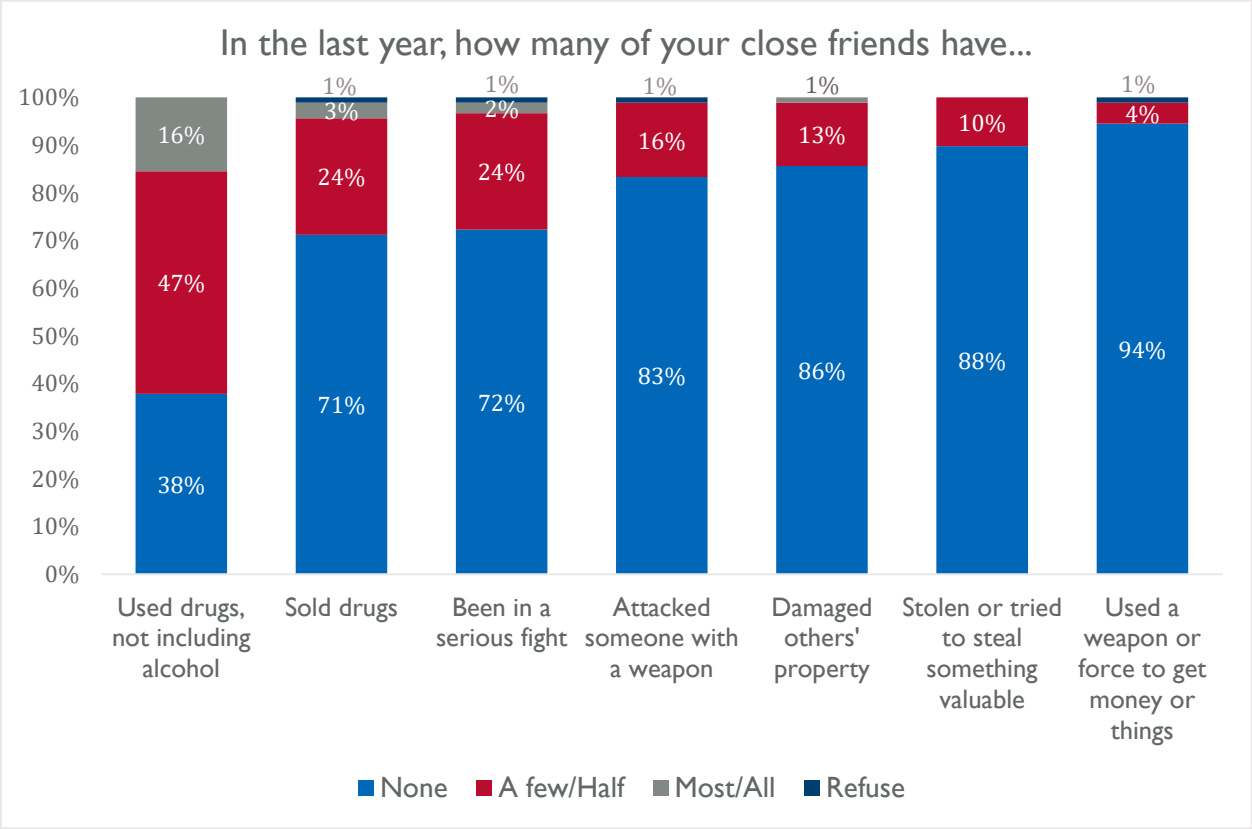
Figure 6: Distribution of answers for questions in the peer delinquency module

In our view, these findings indicate that respondents *are* willing to answer truthfully when survey best practices are adhered to. This conclusion accords with the findings of a recent meta-analysis comparing direct questions to list experiments designed to elicit honest answers to sensitive questions (Blair, Coppock, and Moor, 2020). Drawing on over 30 years' worth of data on sensitive questions across a wide range of domains, the meta-analysis found that "sensitivity bias is typically small to moderate, contra the evident expectation on either the authors' or their real or imagined reviewers' parts that misreporting was a large concern" (p. 1298). The study also provides a framework for understanding when sensitivity bias is likely to occur that asks researchers to consider three key questions in gauging the likelihood of sensitivity bias: Is there a clear "socially desirable" response? What is the perceived likelihood that respondents' perceptions will be revealed to an authority? And what are the perceived consequences of those revelations?

In the context of risk screens, although there is a clear "socially desirable" response to questions about involvement in crime, enumeration best practices can go a long way towards addressing the latter two questions and creating an environment in which the respondent recognizes the legitimacy of the research and trusts that their responses will not be revealed to authorities. Enumerators need to build a rapport with respondents, and to ensure respondents of the confidential nature of the risk screen. They need to obtain informed consent, and they need to explain in layperson's terms the steps that are taken to protect their data. With these assurances in place, the incentive to deceive is greatly reduced.

Risk screens should not be used for high-stakes decisions such as incarceration or eligibility for probation, where incentives to deceive are strong, and they should not be administered by personnel with obvious

conflicts of interest, such as probation officers, prison officials, or law enforcement officers. In addition, we recommend that the risk screen be used in conjunction with an interview conducted by trained social workers or program staff (we discuss these points in further detail in the "Guidelines" section).

Most of the respondents (96 percent) reported being comfortable or very comfortable during the survey, consistent with the enumerators' impressions. As the last question in this section, we asked respondents how honest they were when answering the survey, and if not completely honest, then why not. A small number, 2 percent, said they were only somewhat honest, and another 16 percent said mostly honest. When asked why they were not completely honest, some respondents said they did not know what their friends do or did not want to snitch on them, but most respondents just said "don't know" or did not give a coherent reason as to why they were not completely honest.
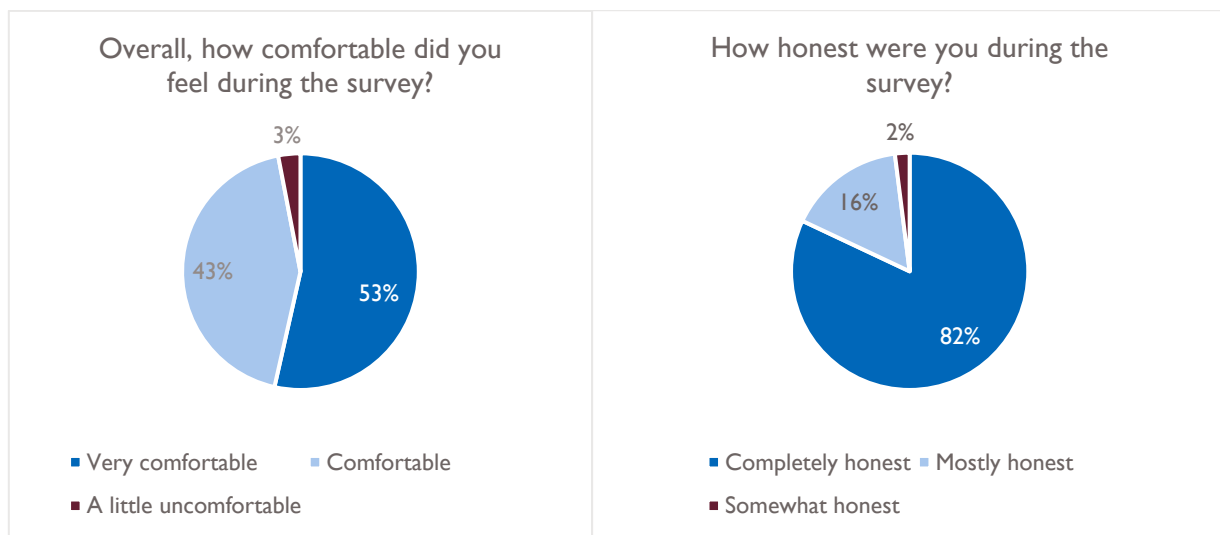
Figure 7: Respondents' comfort and overall (self-reported) honesty during the survey

The relatively high level of comfort, understanding, and honesty that respondents displayed during the survey is also evident in the 32 audio recordings. In reviewing these recordings, we found no instances in which a respondent became audibly angry, frustrated, or withdrawn. Rather, respondents sounded comfortable and engaged throughout the survey, and pauses between questions and answers (a potential sign of respondent hesitancy) were minimal. There were a handful of misunderstandings and miscommunications leading to confusion about the meaning of survey questions, but these were almost always due to a poor phone connection rather than question wording.

CHANGES MADE BASED ON FINDINGS FROM THE PRE-TEST

After looking at the answer distributions and listening to the audio recordings, we made a few decisions aimed at simplifying the tool. The historical factors and social support modules, which originally had "agree/disagree/neither agree nor disagree" answer options, became "yes/no" questions. The peer delinquency module went from five answer options as in the C-YSET ("none/a few/half/most/all") to only three ("none/a few/a lot").

One respondent asked whether alcohol was included in the question "In the last year, have you used illegal drugs that make you high?" To clarify that alcohol was not included, halfway through the pre-testing we modified this question to say "In the last year, have you used illegal drugs that make you high? Not including

alcohol." Alcohol is not included because an assessment of drinking habits would require at least 2-3 questions to establish how often the respondent drinks and how much each time. Because some of the Y-RISC respondents will be children and others will be young adults, we did not find questions, answer options, and scores for alcohol-related questions that would work well for respondents of any age.

One respondent told the enumerator that they did not think that marijuana was a drug. Anticipating that other respondents might have a similar opinion and not include marijuana among "illegal drugs that make you high" in their answers, we modified this question in the final tool to say "In the last year, have you used marijuana or other drugs that make you high? Not including alcohol."

Two questions on delinquency, one for self and one for peers, showed no variability and were dropped. The question asked whether they had used a weapon or force to get money or things from people. The exact wording of the questions was:

> *In the last year, have you used a weapon or force to get money or things from people?*

> *In the last year, how many of your close friends used a weapon or force to get money or things from people?*

Additionally, we dropped one of the decision-making questions that was redundant with another question in the module and had a very similar answer distribution. We also dropped one of the poverty questions that seemed to cause some confusion because it asked about childhood poverty rather than current poverty. The exact wording of these questions was:

> *I think about how the things I do will affect me in the long run (in the future or 'down the road').*

> *When you were in primary school, did any of your parents/guardians have periods of unemployment, when they were struggling to find a job?*

RISK SCORING

Figure 8 shows the distribution of risk scores. The mode for scores is a 7 and 8, and most respondents fell between 4 and 17 points. The average score was 9.55 and the maximum was 20, out of a possible maximum of 35 points. We discuss how to formulate "cut-points" for categorizing respondents as "at-risk" in the next section.
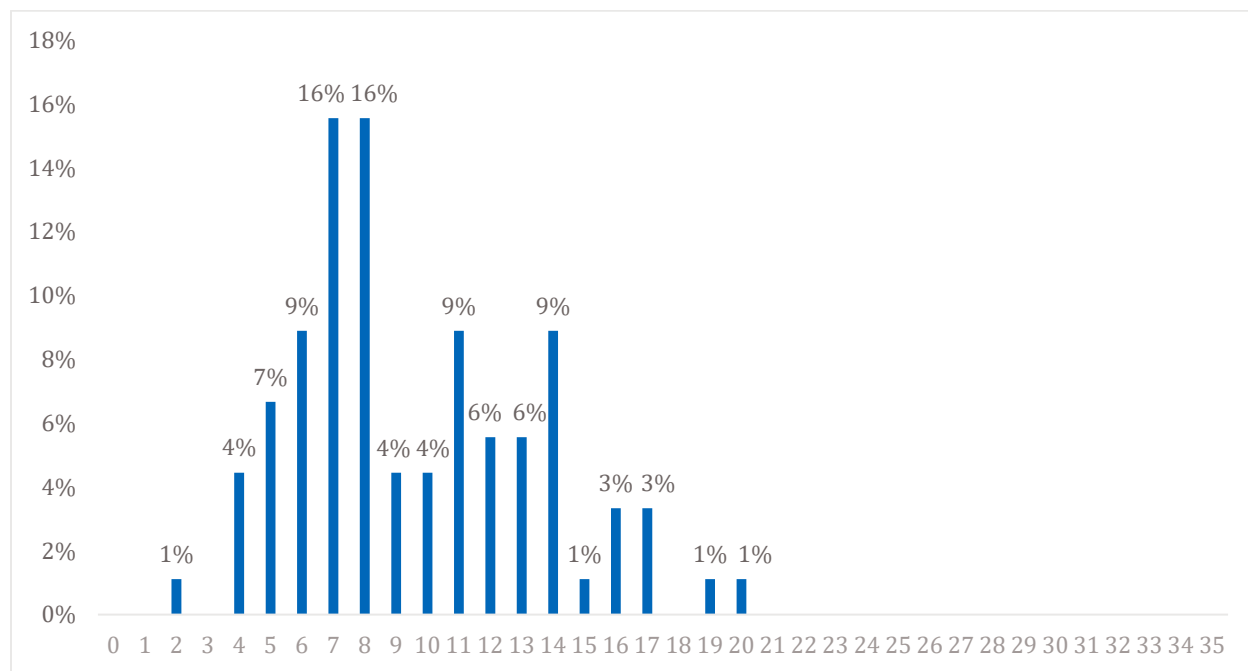
Figure 8: Share of respondents by risk score

GUIDELINES FOR USING THE Y-RISC

GUIDELINES FOR PROGRAM TARGETING

This section discusses guidelines for utilizing the Y-RISC for program targeting, including the intended sample for the assessment, considerations for assessment administration, risk scoring mechanics, determination of program inclusion thresholds, and the role of qualitative assessments by social workers/program staff and referrals in informing program targeting. An important point to reiterate is that the Y-RISC tool is a risk screen that can help identify youth who are most at risk and priority for program inclusion. It is not a needs assessment – the results will not provide insight on whether a particular young person will actually benefit from the program and whether the program will address the root causes of the youth's delinquency risk. Once a list of eligible youth has been identified, we recommend that program staff and/or social workers conduct quick interviews with each youth prior to finalizing the list of participants. This is discussed later in this section.

INTENDED SAMPLE OF THE Y-RISC

The Y-RISC should be utilized as a secondary prevention assessment tool administered with youth who may be at risk of engaging in violent or delinquent behavior due to the presence of environmental risk factors (e.g., youth who live in high-violence communities). The Y-RISC may not be suitable for youth who have already been in conflict with the law, where incentives to lie about illegal behavior are strongest. Furthermore, the tool is intended as a secondary prevention screen to identify youth most likely to engage in delinquent behavior *before* they enter the justice system or have had only minimal contact, rather than a tertiary prevention tool for assessing risk for youth already involved in the system. If it is to be used on this population, we recommend ensuring that a third-party researcher administers the survey, rather than a probation or parole officer, and that scores are shared with officers in a way that ensures they will not be used against the youth's interests.[20]

Y-RISC ADMINISTRATION

Recruiting and Training Interviewers: Once a target sample or setting is identified for the assessment, the first step for administering the Y-RISC would be to identify and train individuals who would be administering the assessment to youth. Depending on the implementing agency, these individuals could be social service workers, enumerators or other program implementing staff. These individuals should have some experience interacting with youth so they are equipped to build rapport with the youth being assessed. We do not recommend that probation or police officers conduct these assessments, as this may lead to dishonest responses, especially on items related to illicit behavior. If resources are available, ideally the Y-RISC would be implemented by third party enumerators who are not associated with the program or with the referring organization (e.g., the school). Given that implementers often face resource and time constraints, another option would be to identify qualified staff at the implementing organization who are not directly involved in program implementation to conduct these assessments. Both of these measures would help minimize bias introduced into the process by the people conducting the assessment and may help reduce the frequency of satisficing by respondents. The training should cover the purpose of the assessment, messaging to be used with stakeholders and assessment respondents when describing the assessment exercise, techniques for building rapport, best enumerator practices, item-specific protocols, and protocols for data management.

Securing Permissions and Consent: After this training is complete, the implementer would need to conduct outreach to obtain necessary approvals for conducting the assessment. The points for consideration would be similar to any other data collection exercise that would involve youth. For instance, if the target population for assessment are youth in school, then consent would need to be obtained from parents and school administration. If the target sample are youth in a particular community, then parental consent would need to be obtained. The implementer should develop a common outreach script that broadly explains the purpose and content of the assessment. During this process, it should also be communicated that the interviews will need to be conducted in private and that the detailed results of the assessment for a particular youth will not be shared. Youth may not feel comfortable answering honestly if their parents, teachers etc. are present in the room, or if they know that their responses will be shared later with these individuals.

---

[20] For instance, the third-party can share aggregate risk scores, rather than detailed scores for each item or each section of the assessment. This would mitigate concerns of consequences due to wrongdoing reported in the delinquency module.

Building and Maintaining Rapport: Prior to the start of each assessment interview, the enumerator should build rapport with the respondent by introducing themselves and properly explaining the purpose of the assessment. During this stage, it will also be crucial to assure youth that their responses will remain confidential and that the purpose of the assessment is not to record specific details about incidents from their past or about any illicit activities conducted by them or their friends. The Y-RISC tool contains helpful scripts for enumerators to read out prior to the start of each module, which help to reinforce these points (see Table 5).

Throughout the assessment interview, enumerators should be aware of verbal and non-verbal cues that the youth is feeling uncomfortable with the questioning. While the "Refuse to answer" option should not be read out, the respondent should be informed at the beginning of the interview that they always have the option to refuse to answer certain items, which may help mitigate discomfort. In addition, the enumerator can always pause the interview to allow the youth to calm down prior to proceeding. After completion of the assessment, enumerators should thank youth for their time and reiterate the confidentiality of their responses.

DATA MANAGEMENT

The Y-RISC has been designed to be easily set up as a spreadsheet that is filled in by the individuals conducting the assessment. Implementers can copy Table 5 into a spreadsheet program (like Microsoft Excel), add desired fields at the beginning of the form for Respondent Name/School, etc., and use this as the data collection form for the assessment. Youth responses can be recorded in the "Answer Options" column by highlighting the cell corresponding to their response, and the final score can be calculated using the "Scoring" column as a guide. If the implementer has the resources, the Y-RISC can be programmed as an online form using a data collection platform (like SurveyMonkey, SurveyCTO, or Qualtrics). The advantages of the online approach are that the interface is more user-friendly for the person entering the data, the data entry can be conducted on a smartphone, and all of the assessment data would be centrally and securely stored allowing for more systematic monitoring and evaluation. If the assessment data must be recorded on paper (perhaps because survey administrators do not have access to laptops or smartphones), the implementer should ensure that there are individuals designated to digitally enter these data soon after the assessments are conducted.

RISK SCORING

As discussed in the "Developing the Y-RISC Section", the scoring for the Y-RISC is a straightforward process. Answer options associated with delinquency risk factors are assigned 1 point in the "Scoring" column and all other answer options do not add points to the risk score (see Table 5). The overall risk score is a simple sum across all items. One thing to note about the scoring scheme for the Y-RISC is that when a respondent refuses to answer any of the historical factors, delinquency, or peer delinquency questions, they get a score of 1 for each refusal, as if they were answering yes. This was a conscious decision made by the research team based on the assumption that it is highly unlikely that respondents would refuse to answer if they, or their friends, had not exhibited delinquent behavior. For items in the historical factors section, we also assume that it is unlikely that respondents would refuse to answer if their home and community were relatively non-violent and stable. If the "Refuse to answer" option is not considered in the scoring scheme, it is possible that youth who are actually at the highest risk may end up

scoring low on the assessment if they refuse to answer multiple items in the historical factors and delinquency modules.

DETERMINING THRESHOLDS FOR PROGRAM INCLUSION

Although risk for future delinquency is a continuum, with higher scores indicating a greater risk of delinquency, determining program eligibility is easiest if a preset cut score can be determined. In this fashion, the assessments can be administered on an individual basis to determine whether a particular youth is eligible, or to a larger group to select a subset for program participation. The question then becomes how best to set the cut score, and how to allow some flexibility to override the score.

We assume that programs for at-risk youth will target communities and contexts with high violence and delinquency rates. Because youth from these neighborhoods, on average, are more likely than youth from less violent communities to engage in delinquent behavior, we can set a lower bar for the cut score. There is no set "rule" for determining the highest-risk group. However, studies from the Caribbean region have found between 20 and 25 percent of male respondents can be considered high-risk, that is, they are in a gang and/or engage in serious violence and delinquency. For example, a recent study in Jamaica found that one in four males from high-crime communities could be considered high-risk, that is, they were in a gang and/or had perpetrated a serious, violent crime. (H. Gayle, Violence Interruption Evaluation Report, 2018). It is important to note that these are aggregate figures that do not distinguish between the percentage of violent offenders and the percentage who engaged in gangs or other non-violent delinquent behaviors, and precise definitions for delinquency vary across different studies. Implementers should be aware that the percentage of youth likely to engage in serious violent behaviors would be much smaller than the percentage of youth likely to engage in gangs or other forms of delinquency.

This guidance on setting thresholds is predicated on using this assessment as a secondary prevention assessment tool, and conducting it with youth who are in environments that put them at greater risk of exhibiting violent and delinquent tendencies. If the tool were being used with a general population of youth, the recommended thresholds for designating youth as 'at-risk' and priority for program inclusion would be quite different, and likely much smaller than the top 20 – 25 percent.

Applying these principles to the pre-test risk score distributions, we see that the 75th percentile threshold falls between a risk score of 13 and 14 (see Figure 8). Therefore, we could consider setting the cut score at 14, and designating individuals who scored 14 points and above as 'at-risk' and priority for inclusion in programming. Although a cut score of 14 could be a reasonable threshold for other implementers to use, we caution that the pre-test was conducted with a small sample of only 90 individuals and that these individuals were young adults rather than juveniles. Further validation with a larger sample that includes juveniles is necessary to obtain more evidence to inform recommended cut points.

If implementers have the resources, they should conduct the Y-RISC themselves with at least 100 respondents to identify a cut score at the 75th percentile threshold that is relevant for their specific context. If these assessments will be conducted on a rolling basis with juveniles and young adults being referred or walking in, the Y-RISC can still be administered to obtain a risk score, but there may not be a large existing set of assessment data to contextualize the score and assess if it falls above the 75th percentile threshold in this particular setting. In this case, it would be equally important to conduct a clinical follow-up assessment with each individual to get a more qualitative sense of risk to inform program suitability. Though we recommend that the Y-RISC be implemented as a secondary prevention assessment tool,

implementers may end up administering the assessment to a sample that is, on average, at lower-risk. One instance where this might be the case would be if there was insufficient existing data to clearly identify a secondary prevention sample, meaning that the assessment might be conducted with a general sample of youth. In this case, the cut score would need to be adjusted upwards. We would recommend considering individuals with scores above the 90th percentile as 'at-risk' and conducting follow-up interviews with youth identified to further refine the list to include only those youth who would benefit from the program. It is crucial for those implementing and scoring the Y-RISC to have a sense of what type of population the assessment is being conducted on so that the threshold for program inclusion can be set accordingly. For instance, if conducting the Y-RISC with youth in a broad sample of schools (e.g., not targeting schools in high crime neighborhoods), keeping the cut score at the 75th percentile threshold would likely recommend a large number of youth for program inclusion who are not actually at risk and are not the target beneficiaries of secondary prevention programming.

All risk assessment tools acknowledge that risk prediction is not a perfect science. There will be false positives, assuming risk when it does not occur, and false negatives, where cases are missed. The cut scores, although based on prevalence data, also are not exact. This point links to the importance of supplementing assessment data with stakeholder consultation, which is further discussed in the next section.

ROLE OF REFERRALS AND QUALITATIVE ASSESSMENTS BY SOCIAL WORKERS AND PROGRAM STAFF

No risk assessment and cutoff threshold will be 100 percent accurate at identifying youth who are at the highest risk and should be considered for programming. Therefore, alongside the Y-RISC, implementers also need to ensure there is a referral mechanism in place to inform program inclusion. For instance, if the assessment is being conducted with youth at a school, the implementer should also be speaking with teachers, school counselors, and parents to hear their views on which students should be included. Referrals can reveal "false negatives" – youth who did not score above the inclusion threshold on the assessment but are actually high-risk and should be included in the program, as individuals making the referrals would have pertinent information that the Y-RISC does not capture. At the same time, evidence shows that individuals providing referrals may also allow their own implicit and explicit biases to influence who they recommend for programming (Noltemeyer et al., 2021). Risk assessments help to combat these biases by providing a more objective measurement of delinquency risk, though there is still a risk that respondents will not respond truthfully or that the risk assessment does not collect all relevant indicators. Therefore, we suggest that referrals should be considered alongside the results from the risk assessment as another data point to aid decision-making.

To take a hypothetical example – assume that there are 20 spaces available in an after-school mentoring program. You are conducting the Y-RISC assessment and soliciting referrals to identify youth to fill these spaces. From the assessment results, you identify 20 youth who scored the highest on the Y-RISC and during the referrals process you identify 5 additional youth recommended for program inclusion by staff at the school. In this case, we recommend that you conduct follow-up interviews with the 20 youth who scored highest on the Y-RISC and the 5 youth recommended via referrals to ascertain the 20 individuals who may benefit most from the program.

As discussed in the Diaz-Cayeros et al. (2020) evaluation of the CFYR *Family Matters* program, it is also possible that the inclusion thresholds could result in "false positives" – cases where youth are not actually high-risk but have scored high enough for inclusion in the program. The frequency of these "false positives"

can be greatly minimized at the pre-assessment stage by purposely selecting youth populations to conduct the assessment with, and by using information from administrative data and stakeholder consultations to identify where these secondary prevention populations are most likely to be. In our view, as long as the program is aimed at fostering positive youth development (rather than being restrictive and imposing sanctions on youth) and does not involve group-based activities, the potential risks posed by including these false positives in programming are minimal. If the program does involve substantial interaction between participants via group-based activities, implementers may need to consider measures to mitigate potential risks of exposing lower-risk youth to relatively high-risk peer groups. Some strategies may include setting a more conservative program inclusion threshold (e.g., 90th percentile rather than 75th percentile), or creating different cohorts within the participant group based on assessed risk level (both from the Y-RISC and other information sources).

Once the assessment is complete and a tentative list of youth to target has been identified, implementers can present this list to informed stakeholders to hear their views and further iterate on the list. Opening up this dialogue may reveal cases that would require follow-up interviewing or more consultation to make a determination on inclusion, so this consultative process would require more time and resources, which implementers should bear in mind.

Follow-up interviews conducted by program staff and/or social workers mentioned in the hypothetical example above, are an important part of the process as they can narrow down the list of potential participants from the risk assessment and referrals, by qualitatively verifying the information collected during the assessment and by providing an opportunity to conduct a brief needs assessment with each youth identified to ascertain if they are a good fit for the program. Program staff would need to be provided guidance on how to conduct these qualitative assessments, and synthesize information from the Y-RISC, referrals, and their own interviews and/or needs assessments with youth. Such guidance is outside the scope of this study, but one key guideline is that staff need to use multiple sources of information when making program inclusion decisions. One benefit of the Y-RISC's streamlined design is that it is less onerous to implement, therefore reducing the burden of the assessment process and freeing up more time and resources to conduct this triangulation process.

GUIDELINES FOR PROGRAM EVALUATION

As currently designed, the Y-RISC is a risk screening tool designed to establish a secondary risk sample, and not designed for program evaluation. However, in practice, we realize that programs may be able to adapt this tool for evaluation purposes if constructs measured align with their theory of change (for example, if a resilience-building program aims to improve decision-making skills). In this section, we suggest modifications to this tool to make it better suited for program evaluation purposes. The modified Y-RISC tool is provided in Appendix III. We also discuss the crucial precondition for using this tool to support program evaluation – that the dimensions measured by the Y-RISC align with intermediate outcomes expected in the program's theory of change.

APPLYING A THEORY OF CHANGE FRAMEWORK TO ASSESS SUITABILITY OF THE MODIFIED Y-RISC

Violence and delinquency are multiply determined and therefore complex outcomes to measure, but each program seeks to impact these outcomes in a specific way by addressing different aspects of the problem. If the program evaluation tool being used does not collect data related to the specific aspects that the program is designed to address, then it is not possible to assess whether the program is effecting change

as intended. Because violence and delinquency are multiply determined, a program may, in fact, be successful in modifying targeted risk or protective factors but simply not be sufficient in scope or intensity to impact violence and delinquency.

The main advantages of using the modified Y-RISC to support evaluation are that it is a relatively quick and simple tool to implement, and that it includes a validated set of indicators related to general social and emotional skills, contextual supports, and behaviors that have been shown to be strongly linked with violence and delinquency. If a program is a general positive youth development or resilience program looking to build decision-making skills, social support, and foster overall resilience, the modified Y-RISC tool would be reasonable to use as a starting point for program evaluation. In this case, we suggest that implementers could build on this modified tool and include additional indicators that could capture more detailed and nuanced information on outcomes that are a focus of the program.

If a program is more specialized or targets factors not measured in the Y-RISC, then more substantial adaptations will be required prior to using it to support evaluation. For instance, if the program is intended to improve family functioning, it would be critical to add at least one module with indicators specific to family functioning and implement these in conjunction with this tool. The modified Y-RISC only includes a couple of items related to physical characteristics of the household, and general social support available to youth – there are no items specifically referencing the respondent's family. A family functioning module would need to collect information on intra-household dynamics, parental monitoring, etc. Similarly, when evaluating an anger management program, it would be critical to include modules that specifically look at intermediate outcomes related to anger management and implement these alongside the modified Y-RISC.

For any implementer wanting to use the modified Y-RISC tool, we strongly suggest using the tool as just a *part* of the program evaluation assessment and strategy. The program's theory of change should play the biggest role in informing the program evaluation design.

DESIGN OF THE MODIFIED Y-RISC

We suggest a couple of modifications to the Y-RISC to make it better suited for program evaluation, and we list out the modifications made to create the version in Appendix III here. For one, the demographic module has been removed. Assuming that the Y-RISC (or similar assessment tool) has already been administered before the youth enter the program, demographic information for each youth would already be available so collecting this again would be unnecessary. The "historical factors" module has also been removed as this section pertains to items that are static (do not change over time). Again, assuming that these data have already been collected for risk assessment purposes, there would be no need to collect the information again because it would not be used for program evaluation.

The timeframe referenced in the "delinquency" and "peer delinquency" items has been reduced from "in the last year" to "in the last month". The key consideration here is that the timeframe for self-report should not conflict with the duration of the intervention. For example, if a program lasts only 3 months, it does not make sense to ask about behaviors from the past year at endline as the post-test data would then include reporting of behavior before the program began. A timeframe of one month is suggested as this is an easy timeframe for respondents to conceptualize and is a short enough period to mitigate concerns that post-test data would reference pre-intervention behaviors. Implementers can amend this time period keeping in mind that other time frames (like 3 months, or 6 months) can be harder for respondents to grasp and accurately report against. The use of major holidays as reference points is one

way to help anchor reference periods for respondents. Whichever timeframe is selected, it should remain constant across all assessment rounds to ensure that the data collected is comparable across rounds.

Lastly, a question has been added to the beginning of the survey to capture the data collection round (e.g., baseline, endline).

ADMINISTERING THE EVALUATION TOOL

Aspects related to training of staff in tool administration, building rapport with youth respondents, and maintaining awareness of respondent reaction to questioning discussed in the previous section remain applicable.

Two unique elements for program evaluation are the number of assessment rounds and the timing of administration of these rounds. To conduct a pre-post assessment, both a baseline and endline assessment are required, therefore the evaluation tool must be implemented at least twice. The decision on whether to conduct interim assessments would depend on program duration and available resources for evaluation. For a 3-month program, an interim assessment is unlikely to yield valuable information as these intermediate outcomes would not have had time to manifest, and repeated testing within a short time frame would be burdensome and potentially bias the responses. Conversely, for a 2-year program, an interim assessment after one year may be worthwhile.

With respect to the timing of the assessments, the baseline assessment should be administered in the 2 weeks before the start of the program. For logistical ease, the baseline could be conducted on Day 1 of the program immediately before the first session that is conducted. If the assessment is conducted too far in advance of the start of the program, this data would not represent a "true" baseline since the youth's situation may have changed substantially in the time between the assessment and the program's start. If the baseline is conducted after the program has started, this could change or bias the responses. Even if the Y-RISC and evaluation tool designs were identical, data from screening assessments are often not suitable to be used as a baseline for program evaluation for this reason. Screening assessments are typically conducted many weeks or months before a program starts. The objective of a baseline is to capture the status of respondents immediately before an intervention begins, therefore the timing aspect is crucial. To capture medium and longer-term movement in outcomes, the endline assessment should be conducted at least a few months after project close. The program's theory of change may help inform when an endline should be conducted if it has assumptions for when key outcomes should start to manifest. If this is not the case, then we suggest conducting the endline assessments six months after the program end. From a practical standpoint, after a program ends it would also become increasingly difficult to contact program participants. Therefore, implementers should plan ahead for the logistics of conducting these endline assessments (e.g., collecting preferred and alternate contact information from respondents, allocating staff and resources to conduct the assessments). If conducting any interim assessments, these assessments should be conducted at the same time for all program participants rather than being staggered so that data from different participants remain comparable.

To ensure that assessment data remains comparable across evaluation rounds, the assessment tool should remain the same from baseline through endline. This does come with a risk of respondents becoming accustomed to the assessment form and answering what they feel the "right" answers are, rather than responding truthfully. This concern of respondents becoming trained on which answers to give is common for many multi-round evaluations. In this case, this issue can be partially mitigated by 1) limiting the number

of interim assessments conducted to only what is considered necessary, 2) maximizing the elapsed time between assessment rounds, and 3) limiting the direct overlap between the evaluation tool and the risk assessment tool.

LIMITATIONS

This study comes with several limitations that implementers should consider before using the Y-RISC. The first is that, due to limitations to the scope of this study, the accuracy of the newly-developed Y-RISC at predicting future delinquency has not been verified. Doing so will require collecting longitudinal data on a sample of potentially at-risk youth, and using the Y-RISC score at time $t = 0$ to predict delinquency at time $t = 1$. The resulting data can be used to validate the Y-RISC, and to develop optimal weights for each question, following the approach outlined in Kshirsagar et al. (2017). To ensure the results of this validation exercise accurately capture real-world accuracy, the sample for this validation exercise should be drawn from a population that is as similar as possible to the population on which it will be used. For example, an organization implementing services for at-risk youth might use the Y-RISC as part of its standard intake procedures, tracking both those it does and does not admit into its programming for roughly one year. Researchers can then use these data to verify the accuracy of the Y-RISC. Once verified, the implementing organization can use the Y-RISC to inform program targeting, without the need to collect longitudinal follow-up data.

The importance of validating the Y-RISC on the population on which it will be used highlights another limitation to the use of risk screens for program targeting – even if accuracy is validated in setting X, there is no guarantee that accuracy will generalize to setting Y. Moreover, there is no guarantee that the tool's accuracy will generalize to the future even within setting X, as conditions could evolve in ways that attenuate the predictive power of the risk factors. The challenge of "external validity" or "generalizability" is not unique to risk screens – indeed it applies to all social science research – but it is important for practitioners to keep in mind when deciding whether to use the Y-RISC or any other risk assessment tool.

While there is no definitive solution to the challenge of generalizability, we have developed the Y-RISC with an eye towards capturing deep and universal risk factors, such as past delinquency, peer influence, and historical family factors, all of which have been shown to associate delinquency across a range of contexts. The hope is that by tapping into these deep and potentially universal predictors, we have developed a tool that is as generalizable across populations as possible.

A second limitation of this study is that it relies on self-reported data to measure delinquency and evaluate the accuracy of the YSET risk assessment tool. Self-reported data have both advantages and limitations, as compared with administrative data on crime. The main advantages are that it is the most cost-effective way to measure delinquency, and that when respondents answer truthfully, it captures all acts of delinquency. By comparison, administrative data from police agencies are much more difficult to access in the countries covered in this study, and only contains information on the small subset of crimes that are reported to the police and subsequently investigated. The main limitation to self-reported data is that respondents may not answer truthfully. Although we cannot rule this possibility out, a growing body of evidence suggests that, contrary to many people's priors, direct questioning is a surprisingly accurate measurement strategy (Blair, Coppock, and Moor (2020); Blattman et al. (2016)). When respondents are

assured of the confidentiality of their responses and trust the interviewer and researchers, the incentive to lie is greatly reduced.

A third limitation to our study relates to the sample on which we tested the Y-RISC. We elected to test the Y-RISC on a sample that partially overlaps with the sample from the Diaz-Cayeros et al. (2020) evaluation, which was previously administered the Y-SET in 2017, 2018, and 2019. The advantage of this sample is that because it was drawn from a population of USAID beneficiaries, it is similar to the types of populations on which the Y-RISC may eventually be used. It was also expedient, insofar as we were able to use the contact information from these earlier surveys to conduct surveys virtually amid the COVID-19 pandemic. The disadvantage of using this sample is that respondents will have seen some of the questions (or similar questions) previously, and this could influence how they responded during our pre-test. We believe this bias to be minimal, however, because at least two years had passed between their last exposure to the YSET and the time of our pilot in May 2021.

CONCLUSION

This study set out to develop a risk screen for youth and young adults with three criteria in mind: short and easy to administer, easy to score and aggregate into a risk score, and accurate at predicting future delinquency. To do this, we used longitudinal data from 2,393 potentially at-risk youth from Honduras and the Caribbean, applying machine learning and forecasting methods to identify which risk factors are the strongest predictors of future delinquency. Combining this subset of strongly predictive risk factors with a small subset of additional risk factors identified in emerging research on delinquency, we developed a streamlined risk assessment tool, the Y-RISC, that is easy to administer, easy to score, and potentially highly predictive of future delinquency.

We pre-tested this tool among a sample of 90 young adults across five Caribbean countries. Overall, the pre-test was successful at confirming high levels of respondent attention and comprehension throughout the survey, and subsequent analysis of pre-test data provided evidence that potentially sensitive questions about self and peer delinquency elicited truthful responses. Although it is impossible to decisively verify the veracity of responses to these survey questions, our assessment is that direct questioning remains the "least bad" approach to collecting sensitive information,[21] and the only viable option for risk screens designed to be inexpensive and easy to administer at scale.

In the final section of this study, we provide a set of guidelines for practitioners interested in using the tool to target their programs towards youth most at risk of delinquency, as well as guidelines for those interested in adapting the Y-RISC for use as a program evaluation tool.

An important limitation of our study is that we cannot decisively demonstrate the predictive accuracy of the Y-RISC. To fully validate the Y-RISC's predictive accuracy, we would need to deploy the tool on a substantially larger sample size than was used in our pre-test survey, and we would need to do so at multiple points in time, allowing us to evaluate whether past Y-RISC scores predict future delinquency. These activities were beyond the scope of this study and remain an important area for future research.

---

[21] See also Blair, Coppock, and Moor (2020), who come to similar conclusions in their meta-analysis comparing direct questioning to randomized response methods.

No risk screen will be 100 percent accurate at predicting which youth are most at risk and therefore should be targeted for programming. Any assessment exercise should also have a referral and clinical judgement process in place to complement and override assessment results. Nonetheless, the Y-RISC has the potential to be a highly useful tool to help inform decisions on program targeting, with its streamlined set of indicators and straightforward scoring process, which help to reduce the assessment burden while still providing useful information.

An important contribution of our study to the policy-research community is that we demonstrate how methods of variable selection from machine learning can be used to develop short, streamlined, and potentially highly accurate risk screens for program targeting. In so doing, we join a growing number of scholars and research-oriented practitioners who have begun to harness the power of machine learning to inform program targeting within the field of development. Over the past several years, machine learning methods have been used to develop "poverty scorecards" to accurately predict whether a household is poor with just ten questions (Kshirsagar et al., 2017); to measure female empowerment with just five survey questions (Jayachandran et al., 2021); and to target cash transfers to those most in need using mobile phone metadata (Blumenstock, 2020), to cite just a few examples. We view this as an encouraging trend with the potential to greatly improve strategic planning and program targeting.

REFERENCES

Beckert, Troy E. 2007. "Cognitive autonomy and self-evaluation in adolescence: a conceptual investigation and instrument development." *North American Journal of Psychology* 9(3): 579-594.

Blair, G., Coppock, A., & Moor, M. (2020). When to worry about sensitivity bias: A social reference theory and evidence from 30 years of list experiments. *American Political Science Review*, 114(4), 1297-1315.

Blattman, C., Jamison, J., Koroknay-Palicz, T., Rodrigues, K., & Sheridan, M. (2016). Measuring the measurement error: A method to qualitatively validate survey data. *Journal of Development Economics*, 120, 99-112.

Blumenstock, J. (2020). Machine learning can help get COVID-19 aid to those who need it most. *Nature.*

Canty-Mitchell, J. and G. D. Zimet. 2000. "Psychometric properties of the multidimensional Scale of Perceived Social Support in urban adolescents." *American Journal of Community Psychology* 28(3): 391-400.

Diaz-Cayeros, Alberto, Stephanie Gimenez Stahlberg, Rachel Pizatella-Haswell, Daniel Sabet and Julia Kresky. 2020. "Evaluation of Secondary Prevention in the Community, Family and Youth Resilience (CFYR) Program in St. Lucia, St. Kitts and Nevis, and Guyana." Prepared for the United States Agency for International Development (USAID). Arlington, VA: Social Impact.

Fawcett, T. (2006). An introduction to ROC analysis. Pattern recognition letters, 27(8), 861-874.

Gayle, H., Ministry of National Security Citizen Security and Justice Programme III. 2018 evaluation of the Violence Interruption Programme. Government of Jamaica.

Katz, Charles, Hyunjung Cheon, Scott H. Decker. 2019. "An evaluation of Proponte Más: A Honduran Secondary Prevention Program Technical Report." Prepared for the United States Agency for International Development (USAID). Arlington, VA: Social Impact.

Kshirsagar, V., Wieczorek, J., Ramanathan, S., & Wells, R. 2017. "Household poverty classification in data-scarce environments: a machine learning approach." Available at: https://deepai.org/publication/household-poverty-classification-in-data-scarce-environments-a-machine-learning-approach

James, G., Witten, D., Hastie, T., & Tibshirani, R. 2013. An introduction to statistical learning. New York: Springer.

Jayachandran, S., Biradavolu, M., & Cooper, J. 2021. "Using machine learning and qualitative interviews to design a five-question women's agency index". (No. w28626). *National Bureau of Economic Research.*

Hennigan, Karen M. Cheryl L. Maxson, David C. Sloane, Kathy A. Kolnick and Flor Vindel. 2014. "Identifying high-risk youth for secondary gang prevention." *Journal of Crime and Justice*, 37:1.

Hennigan, Karen M., Kathy A. Kolnick, Flor Vindel and Cheryl L. Maxson. "Targeting youth at risk for gang involvement: Validation of a gang risk assessment to support individualized secondary prevention." Children and Youth Services Review, 56: 86-96.

Katz, Charles, Cheon, Hyunjung, and Zheng, Yi. 2020. "A longitudinal validation of Caribbean Youth Services Eligibility Tool (YSET)." Prepared for the United States Agency for International Development (USAID). Phoenix, Arizona: Arizona State University.

Kraus, Molly, Kristine Chan, Alfonso Martin, Loraine Park, Jorja Leap, Laura Rivas, Kim Manos, Karen Hennigan and Kathy Kolnick. 2017. "GRYD Gang Prevention 2017 Report." Retrieved May 2020 from https://www.lagryd.org/sites/default/files/reports/GRYD%20Prevention%20Report_Final.pdf

Meeks-Gardner, J., Nancy G. Guerra, Kirk R. Williams, and Ian Walker. 2011. "The Jamaica Youth Survey: Assessing core competencies and risk for aggression among Jamaican youth." *Caribbean Child Development Quarterly,* 57(1), 35-53.

Monahan, John, and Jennifer L. Skeem. "Current Directions in Violence Risk Assessment - Jennifer L. Skeem, John Monahan, 2011." *SAGE Journals,* journals.sagepub.com/doi/10.1177/0963721410397271.

Noltemeyer, A, Kunesh, C, Harper, E, Davis, DR. "Using Conjoint Analysis to Examine Bias in Referral to a Behavior Intervention Program." *Psychol Schs.,* 2021; 58: 89–106. https://doi.org/10.1002/pits.22437.

Phillips, J, and F R Springer. "Presence of Caring - Individual Protective Factors Index 1992; Copyrighted: Phillips, J.; Springer, F." *Washington University Open Scholarship,* openscholarship.wustl.edu/bsltests/4328/.

APPENDIX I: LITERATURE REVIEW

RISK FACTORS

No single risk factor is deterministic in predicting youth involvement in risky and delinquent behavior, but the combination of several risk factors increases the chances that youth will engage in these kinds of behaviors. Multiple studies have found that the strongest predictor of future delinquency is past delinquency, as a literature review on juvenile delinquency has revealed (May et al., 2014). Delinquency, in turn, depends on several other risk factors, the most important of which are at the individual, family, and peers/community levels (Chioda, 2017). Individual risk factors involve school outcomes, and personality traits and behaviors such as empathy and impulsivity in decision-making.

Family risk factors include a safe and stable home environment, parenting skills, parental supervision, and living in poverty. Exposure to violence in the home during childhood, both as a victim or as a witness, increases the chances of emotional problems and of engaging in aggressive behavior (Sutton and Alvarez, 2017). The family's criminal history, particularly of the parents, is one of the strongest predictors of youth involvement with criminality (Chioda, 2017).

Peer relations are important as well, especially for the adolescent group. Adolescents are more likely to engage in delinquency if they have highly delinquent friends and they spend unstructured time socializing with them (Haynie and Osgood, 2005). The effect of peer pressure is stronger if the relationship between adolescents and their mothers is negative (Defoe et al., 2018). On the other hand, associating with prosocial peers reduces the likelihood of youth engaging in behaviors such as property offending and drug use (Walters, 2020).

Protective factors can help guard against negative behaviors by creating a layer of support. For instance, experiencing success in school and having engaged parents and other supportive adults help keep youth on a positive track. Decreasing the number and magnitude of risk factors and boosting protective factors can help prevent problem behaviors in youth (Sutton and Ruprah, 2017).

RISK SCREENS AND RISK ASSESSMENTS

In recent decades, the literature on crime prevention programs has focused on ensuring that programs are targeting the right youth for treatment and are offering services that can meet their specific needs (Savignac, 2010). Consequently, prior to administering a full risk assessment, experts advise program implementers to administer a screening tool in order to identify at-risk youth and recruit them for the program. Criminologists have long been aware that it is possible to predict risk with only a small number of variables (Jackson and Mendoza, 2020). As such, screening tools are typically relatively short instruments and use only a handful of items, such as past behavior, family and peer relationships, and environmental factors, to predict future risk levels.

Since the screening tool may produce false positives and false negatives, it is important to supplement this data with counselor assessments and administrative records, if available, to determine whether a youth may be a good candidate for a behavioral intervention. Several popular screening tools exist. Some are aimed toward more general behavioral and social problems, while others seek to identify particular behaviors that can be addressed by a specific intervention. Examples of screening tools include the

Behavioral and Emotional Screening System (BASC-2 BESS), the School Social Behavior Modules (SSBS), and the Problem-Oriented Screening Instrument for Teenagers (POSIT).

Once a youth is deemed eligible for a program, a risk assessment is typically administered, the results of which will allow for the development of a tailored and comprehensive approach to address the youth's situation. Common risk assessment tools include the Level of Service Inventory-Revised (LSI-R), the Structured Assessment of Violence Risk in Youth (SAVRY), and the COMPAS Core Risk and Needs Assessment (RNA), all of which are designed to capture in-depth information on the combination of factors that may have led to this particular youth's high risk level and which can be used by a program facilitator to identify which aspects of the intervention are best-suited to meet the youth's needs. As part of a risk assessment tool, it is recommended that both static and dynamic risk factors are evaluated to quantify a young person's risk or need for an intervention. Static factors do not change, such as criminal history, while dynamic factors include those that can change over time, such as family relationships and alcohol/drug problems. Assessment tools may also seek to identify protective factors that can promote resilience if they are reinforced as part of a targeted intervention.

Some instruments, like the Jamaican Risk Assessment and the YSET, serve as both screening tools and risk assessment tools. The Jamaican Risk Assessment tool is comprised of a screening tool of 13 factors followed by a 13-question assessment tool. It is only after the youth scores above a certain threshold on the screening questionnaire that they are administered the full risk assessment. The YSET, on the other hand, requires that all youth complete a full risk assessment to determine program eligibility. Therefore, even those youth who will ultimately be deemed ineligible for the program must undergo a lengthy risk assessment of approximately thirty to forty-five minutes, which may collect more information than is necessary to determine if the youth would be a good candidate for program participation.

Risk assessment tools vary in length, targeted behaviors, and question type, but most tools consist of a series of questions divided into different modules based on specific subject areas. For example, the Jamaican Youth Survey asks about five core competencies to assess the likelihood of engaging in aggression or violence: positive sense of self, self-control, decision-making skills, moral system of belief, and pro-social connectedness (Meeks-Gardner et al., 2011). The survey also collects self-reported measures of aggressive behavior. Similarly, the SAVRY tool assesses 24 items clustered under three risk domains: historical risk factors, social and contextual risk factors, and individual and clinical risk factors, and also protective factors such as social support, commitment to school and others.[22]

COMPAS Core RNA contains two main modules, the General Recidivism Risk Module and the Violent Recidivism Risk Module, as well as a set of need modules that can be adapted based on program requirements (Jackson and Mendoza, 2020). Other validated and well-known risk assessment tools we reviewed include the YLS/CMI 2.0, the Washington State Juvenile Court Assessment, and the Risk and Resilience Checklist (RRC).

Behavioral changes are typically assessed by analyzing youth responses to multiple-choice and Likert module questions, but there are also more unique approaches to measure risk. For example, the 5Essentials program in Chicago, formerly known as "My School, My Voice," focuses primarily on the school environment and argues that youth's likelihood to engage in risky behavior can be gauged by their level of automaticity, or the time it takes for them to understand and respond to any given situation (Vedantam,

---

2017). The authors argue that lower automaticity is correlated with positive behavioral outcomes and that this method can be applied to measure how likely a youth is to engage in risky behavior over time.

The predictive accuracy of risk assessments varies, but the evidence does not support using risk screens or assessments as the sole determinant for detention, criminal sentences, and release (Fazel et al., 2012). Some risk assessments are criticized on the basis of lacking transparency, even when they are used in high-stakes decisions by the criminal justice system (Rudin et al., 2020). A large systematic review and meta-regression analysis examined nine different risk screens and concluded that the SAVRY had the highest rates of predictive validity for antisocial, violent, and sexual risk (Singh et al, 2011). The instruments examined by this study produced higher rates of predictive validity among older and white populations, a finding which provides an important reason for creating and validating risk screens in contexts with a mostly young and non-white population.

Finally, and perhaps most importantly, in selecting a tool, it is essential to ensure compatibility between the tool, program implementation criteria, and program objectives (Savignac, 2010). The risk factors and issues central to the program should be similar to those targeted by the tool. If this condition is met and the tool is designed around the theory of change, then risk assessment tools can also be used as evaluation tools. The instrument developed for the impact evaluation of the *A Ganar* program in Honduras, for example, was designed with the theory of change in mind (Duthie et al., 2018). On the other hand, in the case of the evaluation of the CFYR secondary risk prevention activity in the Caribbean, questions from the endline survey included youth opinions of the police and other factors that are very unlikely to change with the intervention (Diaz-Cayeros et al., 2020). In general, an evaluation tool should only ask questions about behaviors and attitudes which the program seeks to change.

SURVEY DESIGN

SURVEY DESIGN EFFECTS

The literature on survey design provides insight into the types of challenges researchers face when interpreting results from surveys applied to youth. The problems include the cognitive capacity of younger respondents who may simply not quite understand what is being asked; the possibility of differences in the propensity to satisfice (answer questions without giving due thought, simply to shorten the survey); issues of questionnaire fatigue; and the propensity to lie or provide the socially desirable answers.

Younger children may face greater cognitive challenges in a face-to-face interview like the YSET, which may affect the quality of collected data (Fuchs, 2009). This is likely to influence non-response rates or the way in which they deal with numeric modules.

Being in school, children are familiar with a format of testing and may interpret a survey as a test in which they are expected to provide "correct" answers to questions. This phenomenon seems to be more prevalent when a survey is collected in schools (Scott, 2000). The literature suggests that older youth may manipulate the instruments by purposefully answering "as a joke," but this behavior may be countervailed by a greater nuance in their sense of morality and their desire to conform with peer pressure (De Leeuw, 2011). Younger children may be more likely to use satisficing when answering questions on subjects they find uninteresting (Borges et al., 2000).

Survey length can also influence the way youth respond to questions. Evidence of survey fatigue includes respondents answering questions faster at the end of questionnaires and providing more identical answers

to multiple questions in a row (Galesic and Bosnjak, 2009; Herzog and Bachman, 1981). Youth specifically have been found to fabricate answers when responding to longer instruments (Betrand et al., 2009). YSET survey length has been a concern for researchers since the instrument was implemented in Los Angeles (Kraus et al., 2017).

Social desirability bias is a common problem in social science research. Surveys referencing sensitive topics may elicit results that are distorted by social desirability bias (Krumpal, 2013). This could be especially true for youth, as social desirability has been found to account for variance in self-reported deviant behavior among youth (Camerini and Shulz, 2018).

RISK SCALING

Measuring risk utilizing multiple factors is commonly done by aggregating cumulative risk (Ettekal et al., 2019). Risk exposure tends to occur across multiple domains and systems rather than in isolation, so a cumulative risk module accounts for the co-occurrence of risk factors. Cumulative risk gathers several risk factors into a single index of risk, as is done in the YSET. Cumulative risk is a dichotomous variable where one is at risk and zero is not at risk. In some cases, an aggregate score is likely to have more predictive power than an individual risk factor alone, though aggregate scores are not always the best models to predict future behavior or a youth's psychological circumstances.

Alternative models to aggregate risk include a proportion-score cumulative risk index, a standardized z-score cumulative risk index, reflective indicator, formative indicator, and person-centered indicators. A proportion-score cumulative risk index is generated by dividing each individual score by the maximum score, and the value of the index is a continuous variable from zero to one. Individual risk variables that are continuous, as opposed to dichotomous, lend themselves to this method. Proportion-score cumulative risk rank-orders risk, so youth who would fall right below and above the cutoff on the cumulative risk index are considered close to the same risk level.

A standardized z-score cumulative risk index requires standardizing individual scores and summing the standardized individual scores. As with the proportion-score cumulative risk index, it maintains rank ordering and assumes risk is continuous. It does not weigh all risks the same and may be useful when certain risk factors contribute more to overall risk than others.

Reflective indicators require "specifying a measurement model using multiple risk indicators as effect indicators of an unobservable latent construct" (Ettekal et al. 2019). Individual risk factors can be either dichotomous or continuous. Thus, the model can be used to assess risk across latent, unobservable factors. Formative indicators "consist of specifying a multiple risk composite variable in which the indicators are predictors of the composite, as opposed to being reflective of it (as in the RI [reflective indicator] approach)" (Ettekal et al., 2019; Hall et al., 2010).

Person-centered methods identify groups or latent classes of youth who exhibit similar risk profiles based on a set of variables. They create risk profiles for latent classes such as individuals who exhibit high risk across multiple indicators or those who exhibit a mix of risk across indicators.

LITERATURE REVIEW BIBLIOGRAPHY

Borges, Natacha, Edith de Leeuw, and Joop Hox. 2000. "Children as Respondents in Survey Research: Cognitive Development and Response Quality." *Bulletin de Methodologie Sociologique.* Retrieved from: http://joophox.net/publist/bms66.pdf

Betrand, Annie, Melanie Beauvy-Sany, Selma Cilimkovic, Sita Conklin, and Selma Jahic. 2009. "Monitoring and Evaluation for Youth Workforce Development Projects." *The SEEP Network.* Retrieved from: https://seepnetwork.org/files/galleries/630_TN_Monitor_Eval_508.pdf

Camerini, A., and P. J. Schulz. 2018. Social Desirability Bias in Child-Report Social Well-Being: Evaluation of the Children's Social Desirability Short Module Using Item Response Theory and Examination of Its Impact on Self-Report Family and Peer Relationships. *Child Indicators Research* 11(4): 1159–1174.

Chioda, Laura. 2017. *Stop the Violence in Latin America: A Look at Prevention from Cradle to Adulthood.* Washington, DC: World Bank Group. https://doi.org/10.1596/978-1-4648-0664-3

De Leeuw, Edith. 2011. "Improving Data Quality when Surveying Children and Adolescents: Cognitive and Social Development and its Role in Questionnaire Construction and Pretesting." Retrieved from: https://www.aka.fi/globalassets/awanhat/documents/tiedostot/lapset/presentations-of-the-annual-seminar-10-12-may-2011/surveying-children-and-adolescents_de-leeuw.pdf

Defoe, I. N., J. S. Dubas, and M. A. G. van Aken. 2018. The relative roles of peer and parent predictors in minor adolescent delinquency: exploring gender and adolescent phase differences. *Frontiers in Public Health.* Retrieved from: https://www.frontiersin.org/articles/10.3389/fpubh.2018.00242/full#B2

Diaz-Cayeros, Alberto, Stephanie Gimenez Stahlberg, Rachel Pizatella-Haswell, Daniel Sabet, and Julia Kresky. 2020. "Evaluation of Secondary Prevention in the Community, Family and Youth Resilience (CFYR) Program in St. Lucia, St. Kitts and Nevis and Guyana." Arlington, VA: Social Impact, Inc. Retrieved from: https://pdf.usaid.gov/pdf_docs/PA00X2GZ.pdf

Duthie, Mike, Mateusz Pucilowski, Lisette Anzoategui, and Braden Agpoon. 2018. "A Ganar Alliance: Impact Evaluation Endline Report Honduras." Prepared for the United States Agency for International Development (USAID). Social Impact, Inc: Arlington, VA. Retrieved from: https://pdf.usaid.gov/pdf_docs/PA00T78R.pdf

Ettekal I, Eiden RD, Nickerson AB, Schuetze P. 2019. Comparing alternative methods of measuring cumulative risk based on multiple risk indicators: Are there differential effects on children's externalizing problems? *PLoS ONE* 14 (7): e0219134. Retrieved from: https://doi.org/10.1371/journal.pone.0219134

Fazel, Seena, Jay P. Singh, Helen Doll, and Martin Grann. 2012. "Use of risk assessment instruments to predict violence and antisocial behavior in 73 samples involving 24 827 people: systematic review and meta-analysis. *British Medical Journal (BMJ),* 345.

Fuchs, Marek. 2009. "The Reliability of Children's Survey Responses: The Impact of Cognitive Functioning on Respondent Behavior." Retrieved from: https://www150.statcan.gc.ca/n1/en/pub/11-522-x/2008000/article/10961-eng.pdf?st=CfumMDNm

Galesic, Mirta & Bosnjak, Michael. 2009. Effects of Questionnaire Length on Participation and Indicators of Response Quality in a Web Survey. *Public Opinion Quarterly* 73.

Hall JE, Sammons P, Sylva K, Melhuish E, Taggart B, Siraj-Blatchford I. 2010. Measuring the combined risk to young children's cognitive development: An alternative to cumulative indices. *British Journal of Developmental Psychology* 28: 219–238.

Haynie, D. and D. W. Osgood. 2005. Reconsidering peers and delinquency: how do peers matter? *Social Forces* 84(2): 1109-1130.

Herzog, A. Regula, Jerald G. Bachman. 1981. Effects of Questionnaire Length on Response Quality. *Public Opinion Quarterly*, 45 (4): 549–559.

Jackson, E., & Mendoza, C. 2020. Setting the Record Straight: What the COMPAS Core Risk and Need Assessment Is and Is Not. *Harvard Data Science Review*, 2(1). Retrieved from: https://doi.org/10.1162/99608f92.1b3dadaa

Kraus, Molly, Kristine Chan, Alfonso Martin, Loraine Park, Jorja Leap, Laura Rivas, Kim Manos, Karen Hennigan and Kathy Kolnick. 2017. *GRYD Gang Prevention 2017 Report.* Retrieved from: https://www.lagryd.org/sites/default/files/reports/GRYD%20Prevention%20Report_Final.pdf

Krumpal, I. 2013. Determinants of social desirability bias in sensitive surveys: a literature review. *Quality & Quantity* 47: 2025–2047.

May, J., K. Osmond and S. Billick. 2014. Juvenile delinquency treatment and prevention: a literature review. *Psychiatric Quarterly* 85: 295-301.

Meeks-Gardner, J., N. G. Guerra, K. R. Williams, and I. Walker. 2011. "The Jamaica Youth Survey: Assessing core competencies and risk for aggression among Jamaican youth." *Caribbean Child Development Quarterly,* 57(1), 35-53.

Rudin, Cynthia, Caroline Wang and Beau Coker. 2020. "The Age of Secrecy and Unfairness in Recidivism Prediction." *Harvard Data Science Review*, 2(1).

Singh, Jay P., Martin Grann and Seena Fazel. 2011. "A comparative study of violence risk assessment tools: a systematic review and metaregression analysis of 68 studies involving 25,980 participants." *Clinical Psychology Review* 31(3): 499-513.

Savignac, Julie. 2010. *Tools to Identify and Assess the Risk of Offending Among Youth.* National Crime Prevention Centre, Public Safety Canada: Ottawa

Scott, J. 2000. "Children as respondents: The challenge for quantitative methods." *Research with children: Perspectives and practices, 98-119.* New York: Psychology Press.

Sutton, H., & Alvarez, L. 2017. Violence against women and children: How big is the problem and is it connected to other violence? In H. Sutton & I. Ruprah (Eds.), *Restoring paradise in the Caribbean: Combatting violence with numbers* (pp. 59-80). Retrieved May 11, 2021, from https://publications.iadb.org/publications/english/document/Restoring-Paradise-in-the-Caribbean-Combatting-Violence-with-Numbers.pdf

Sutton, H., & Ruprah, I. 2017. Youth violence and delinquency: Reducing risk and enhancing protection. In H. Sutton & I. Ruprah (Eds.), *Restoring paradise in the Caribbean: Combatting violence with numbers* (pp. 81-98). Retrieved from: https://publications.iadb.org/publications/english/document/Restoring-Paradise-in-the-Caribbean-Combatting-Violence-with-Numbers.pdf

Vedantam, Shankar. 2017. Chicago Leaders Use Cognitive Behavioral Therapy to Combat Violent Crime [Audio podcast episode]. In *Hidden Brain*. February 28, 2017. National Public Radio. Retrieved from: https://www.npr.org/2017/02/28/517779745/chicago-leaders-use-cognitive-behavorial-therapy-to-combat-violent-crime

Walters, G. 2020. Prosocial peers as risk, protective and promotive factors for the prevention of delinquency and drug use. *Journal of Youth and Adolescence* 49: 618-630.

APPENDIX II: SURVEY FEEDBACK QUESTIONS

Table 1: Questions on survey feedback from respondents and enumerators

SURVEY FEEDBACK FROM RESPONDENTS	*That concludes the survey. Thank you for your time and cooperation. I would like to ask a few questions about what you think about the survey.*	
end_respondent1	Were there questions that were difficult to understand or answer?	Yes
		No
		Refuse to answer
end_respondent2	[if yes]: Which ones?	Open response
end_respondent3	Were there questions that you think other respondents would probably lie about?	Yes
		No
		Refuse to answer
end_respondent4	[if yes]: Which ones?	Open response
end_respondent5	Overall, how comfortable did you feel during the survey?	Very comfortable
		Comfortable
		A little uncomfortable
		Very uncomfortable
end_respondent6	[If very or a little uncomfortable]: Why were you uncomfortable?	[open response]
end_respondent7	Were you able to stay focused during the entire interview?	Yes
		No
		Refuse to answer
end_respondent8	[If not]: Why not? (*If needed, probe about survey length, tiredness, difficulty hearing or understanding the questions*)	[open response]
end_respondent9	How did you feel answering questions about breaking rules or laws in the last year?	[open response]
end_respondent10	How did you feel answering questions about what your friends have done in the last year?	[open response]
end_respondent11	Would you say you were completely honest throughout the survey, mostly honest, or only somewhat honest?	Completely honest
		Mostly honest
		Somewhat honest
		Refuse to answer
end_respondent12	[If mostly or somewhat honest]: Why were you not completely honest?	[open response]
end_respondent13	That was the last question. Thank you for speaking with us today. Do you have any other comments that you would like to add?	[open response]

SURVEY FEEDBACK FROM ENUMERATORS	[ENUMERATOR] These questions are for the enumerators to answer after ending the interview and hanging up with the respondent.	
end_enumerator1	How honest did the respondent seem to you overall?	Completely honest
		Mostly honest
		Somewhat honest
end_enumerator2	Were there any questions that were confusing to the respondent?	Yes
		No
end_enumerator3	[If yes]: Please make sure to note all questions (with question numbers) where the respondent was confused, with a brief explanation.	[open response]
end_enumerator4	Were there any questions that seemed to make the respondent uncomfortable?	Yes
		No
end_enumerator5	[If yes]: Please make sure to note all questions (with question numbers) where the respondent seemed uncomfortable, with a brief explanation.	[open response]
end_enumerator6	Do you feel that the respondent grew tired of the interview towards the end?	Yes
		No
end_enumerator7	Do you have any other comments or feedback?	[open response]

APPENDIX III: VERSION OF THE Y-RISC ADAPTED FOR PROGRAM EVALUATION

Implementers who choose to use this tool, should supplement it with indicators and additional modules that are applicable to their program's theory of change. Implementers may also want to add a demographic module or historical factors module at baseline in case that information has not already been collected, and consider modifying the timeframe against which respondents report behaviors (e.g., change from "last month" to "last three months").

Table 2: Y-RISC adapted for program evaluation

Respondent Name/ID No.			
Assessment Round	Select one: *Baseline* *Endline*		
Question or module name	**Question text**	**Answer options**	**Scoring**
PRODUCTIVE ENGAGEMENT	*The following is a list of activities of how young people spend most of their time. Please tell me which you are doing right now:*		
school	In school? Either full time or part time.	Yes	
		No	I
training	Getting job training, vocational training, or completing an internship/apprenticeship?	Yes	
		No	I
work	Working full or part time and/or self-employed?	Yes	
		No	I
family_business	Helping in a family business without pay?	Yes	
		No	I
volunteering	Volunteering or working for the community?	Yes	
		No	I
POVERTY	*Next, I will ask a few questions about your home. If you live in more than one place, please answer about your primary home.*		
poverty1	Does the place where you live have a refrigerator?	Yes	
		No	I
		Don't know	
		Not applicable	
poverty2	Does the place where you live have fewer than three rooms? Counting any kind of room, such as kitchen, bedroom, bathroom, etc.	Yes	I
		No	
		Don't know	

		Not applicable	
poverty3	Does anyone in your household have a car or truck?	Yes	
		No	I
		Don't know	
		Not applicable	
		Refuse to answer	
DECISION MAKING	*These next questions ask about making decisions, that is, the things that you make up your mind about. Please tell me how often you do these things:*		
decision_making1	When I am making up my mind about something important, I think about all the things that could happen (the consequences).	Always	
		Often	
		Sometimes	I
		Never	
decision_making2	I consider different choices before making up my mind about something.	Always	
		Often	
		Sometimes	I
		Never	
decision_making3	I think about how the things I do will affect others.	Always	
		Often	
		Sometimes	I
		Never	
DELINQUENCY	*People sometimes break rules or laws. Some of the questions in this section ask about the rules or laws you may have broken in the last **month**. Remember, your answers will stay private and will not be shared with anyone outside the **program** team.*		
delinquency1	In the last **month**, have you used marijuana or other illegal drugs? Not including alcohol.	Yes	I
		No	
		Refuse to answer	I
delinquency2	In the last **month**, have you sold illegal drugs?	Yes	I
		No	
		Refuse to answer	I
delinquency3	In the last **month**, have you purposely damaged or destroyed property that did not that belong to you?	Yes	I
		No	
		Refuse to answer	I
delinquency4	In the last **month**, have you carried a knife, gun, or other weapon for protection?	Yes	I
		No	

		Refuse to answer	I
delinquency5	In the last **month**, have you stolen or tried to steal something valuable?	Yes	I
		No	
		Refuse to answer	I
delinquency6	In the last **month**, have you hit someone with the purpose of hurting them?	Yes	I
		No	
		Refuse to answer	I
delinquency7	In the last **month**, have you attacked someone with a weapon?	Yes	I
		No	
		Refuse to answer	I
delinquency8	In the last **month**, have you been involved in fights where someone got seriously injured?	Yes	I
		No	
		Refuse to answer	I
SOCIAL SUPPORT	*This next set of questions asks about your relationship with others. Please tell me yes or no for the following statements:*		
social_support1	There are people I can depend on to help me if I really need it.	Yes	
		No	I
social_support2	I have an adult I can turn to for help when I am worried about something or have a problem.	Yes	
		No	I
social_support3	There is a special person in my life who cares about my feelings.	Yes	
		No	I
social_support4	I know where to go if I need advice about something.	Yes	
		No	I
social_support5	I know about opportunities available to me (such as jobs, trainings, religious activities, sports, clubs, etc.)	Yes	
		No	I
PEER DELINQUENCY	*This next set of questions asks about whether your friends are getting into trouble. But we don't want to know who specifically is getting into trouble, and we don't want to know anyone's name. Remember, this information is only for research and will not be shared with anyone outside the **program** team.*		
peer_delinquency1	In the last **month**, how many of your close friends used drugs that make you "high"? Not including alcohol.	None	
		A few	
		A lot	I
		Refuse to answer	
peer_delinquency2		None	

	In the last **month**, how many of your close friends sold drugs?	A few	
		A lot	1
		Refuse to answer	
peer_delinquency3	In the last **month**, how many of your close friends purposely damaged or destroyed property that did not belong to them?	None	
		A few	
		A lot	1
		Refuse to answer	
peer_delinquency4	In the last **month**, how many of your close friends stole or tried to steal something valuable?	None	
		A few	
		A lot	1
		Refuse to answer	
peer_delinquency5	In the last **month**, how many of your close friends attacked someone with a weapon?	None	
		A few	
		A lot	1
		Refuse to answer	
peer_delinquency6	In the last **month**, how many of your close friends been involved in fights where someone got seriously injured?	None	
		A few	
		A lot	1
		Refuse to answer	
TOTAL SCORE (Valid range is from **0-30**)			

U.S. Agency for International Development
1300 Pennsylvania Avenue NW
Washington, DC 20523