Empirical Insights for Improving Sexual Assault Prevention:

Early Evidence from a Cluster-Randomized Trial of IMPower and Sources of Strength

DRAFT VERSION, SEPTEMBER 2018

Evan Rosenman

Stanford University Department of Statistics

Clea Sarnquist

Stanford Prevention Research Center

Rina Friedberg

Stanford University Department of Statistics

Mary Amuyunzu-Nyamongo

Africa Institute for Health and Development

Gabriel Oguda

Africa Institute for Health and Development

Dorothy Otiento

Africa Institute for Health and Development

Michael Baiocchi

Stanford University Department of Statistics and Stanford Prevention Research Center

Abstract

The empirical science of measuring and preventing sexual assault is in its infancy, especially when considering adolescents in developing nations. We analyze pre-intervention data collected in a two-arm cluster-randomized controlled trial of a classroom-based sexual assault prevention program deployed to class 6 students around Nairobi, Kenya. We estimate that 7.2% of girls were raped in the prior 12 months. We identify school- and individual-level risk factors for rape. We isolate, as much as possible, variation in probability of rape attributable to a subset of these risk factors. We discuss statistical challenges and solutions in each of these domains.

Empirical Insights for Improving Sexual Assault Prevention:

Early Evidence from a Cluster-Randomized Trial of IMPower and Sources of Strength

Background

Sexual assault is a serious public health issue among adolescents. Worldwide, more than 10% of girls experience forced intercourse at some point in their lifetimes (UNICEF et al., 2014). Prior work has shown victimization rates as high as 31% among African girls, and perpetration rates as high as 40% among African boys (Shamu et al., 2015).

Research on sexual assault has typically focused on intimate partner violence among adults or older adolescents (Livingston et al., 2007) and typically in high-income countries (Koss et al., 2007; Ellsberg et al., 2015). The literature offers a limited understanding of sexual assault prevalence and prevention among children at an earlier stage of development. Vulnerable subgroups, such as economically and socially disadvantaged children living in informal settlements, are further understudied.

Previous work has established the effectiveness of sexual assault prevention programs in early adolescents in these settings (Jewkes et al., 2014). In particular, a randomized trial indicates that such programs can reduce the rate of rape in the prior twelve months by 3.7% (95% confidence interval: [0.4%, 8.0%]) from a baseline of 7.3% (Baiocchi et al., 2017). Such interventions – referred to as "empowerment" programs – place emphasis on (i) awareness of autonomy and right to refuse, (ii) understanding of common patterns of perpetration, and (iii) skill development to appropriately intervene and mitigate perceived threats.

This paper concerns data from a stratified sample of student participants in a cluster-randomized trial underway in five informal settlements around Nairobi, Kenya. The clusters in this intervention are the schools attended by the boys and girls. The tested intervention is a combination of a previously tested girls' intervention, IMPower, and a newly revised boys' intervention, Sources of Strength. The IMPower program traces many aspects of its model from a framework proposed in Koss & Rozee (2001). The Sources of Strength draws from frameworks developed around norm setting and bystander intervention (Gydicz et al., 2011).

All measurements used in this paper come from surveys before the start of any trainings associated with the randomized trial. Analysis of baseline data includes establishing point estimates and confidence intervals for outcome measures and covariates at baseline. Additional work involves an exploratory analysis of modifiable causal pathways (a.k.a. "mediating variables") for rape, and assessing predictors of each school's need for the intervention. Methodological challenges arise around estimating uncertainty, data adjudication, and modeling risk factors. This paper details the methodologies used to address each of these challenges. Code, data simulated to be similar to the actual data, and a walkthrough tutorial are made available to make the methodological considerations understandable, accessible, and replicable; these are available at: https://github.com/rinafriedberg/gbv-analytics/tree/master/Tutorial.

Related Literature

The design and analysis of surveys to gauge experiences of sexual assault is an active topic of research (Cook et al., 2011; Davis et al., 2014). The most widely used tool is the Sexual Experiences Survey, which measures unwanted sexual experiences from age 14 onward (Koss et al., 2007). In this and other surveys (Brener et al., 1999), the potential for inconsistent responses is well known. For example, Fisher and Cullen (2000) found that when surveying college women about their experience with sexual assault, nearly 20% of incidents initially classified as rape were later classified as "undetermined" because respondents provided inconsistent responses to follow-up questions. They hypothesize that "some 'real' rape victims were not counted because they did not understand questions in the incident report or wearied at having to answer a second round of questions about a potentially painful event in their lives” (Fisher and Cullen, 2000).

Due to these inconsistencies, researchers have suggested balancing the need for information collection against the possibility that follow-up questions may suppress disclosure (Koss et al., 2007). Current adjudication procedures frequently involve categorizing respondents according to their most severe reported assault experience (Davis et al., 2014). Researchers also make use of Cronbach’s Alpha (Cronbach and Meehl, 1955), a measure of the internal consistency of survey data, to validate that the responses are reliable (Alpha scores greater than 0.7 are often considered acceptable). However, explicit best practices for adjudicating conflicting responses were not found in a review of the relevant literature.

The cluster-randomized setting of this trial presents statistical challenges that are absent when treatments are randomized to individuals. For a full discussion, see Gail et al. (1996). Chief among these challenges is homophily, the tendency of units within a cluster to be more similar to one another (in terms of pre-intervention covariates) than to units in other clusters. Because of homophily, randomization can be challenging, as “cluster randomization does not guarantee balance of potential person-level confounders between conditions” (Vuchinich et al., 2012).

A second set of issues in cluster-randomized trials concerns sources of variability. If clusters are different sizes or if individuals are more variable within some clusters as compared to others, then treating each cluster as an independent draw from a population can yield incorrect inference. A potential solution is found via bootstrapping, originally proposed by Efron (1992). In a bootstrap procedure, “pseudo-datasets” are constructed by resampling with replacement from the original datasets. Estimates of the target parameter are then computed on each of the pseudo-datasets, and the spread of these estimates allows researchers to compute confidence intervals for the original sample estimate. In the cluster setting, researchers make use of a variant – the clustered bootstrap – which permits heteroskedasticity and within-cluster error correlation as long as there is a reasonably large number of clusters (Cameron et al., 2008).

Research on sexual assault prevention also frequently makes use of structural equation modeling (SEM) to identify pathways that lead to sexual assault (Shamu et al., 2015; Russell et al., 2014). SEM, also known as causal modeling, is a framework for thinking about data that researchers invoke to estimate causal relationships between a set of independent variables and dependent variables. In the SEM framework, a researcher posits a detailed causal relationship between variables (often represented by a directed graph), estimates coefficients associated with such a model of the data, and obtains goodness-of-fit measures to evaluate the plausibility of the relationships (Ullman and Bentler, 2012). SEM is widely used in the social sciences, and various guidelines exist for its practical utility (Anderson and Gerbing, 1988). In the context of sexual violence, a typical analysis might use SEM to identify the ways that gender attitudes mediate the relationship between socioeconomic status and rape frequency.

While SEM is quite flexible, it also imposes parametric assumptions that can be indefensible and biasing in practice (Ullman and Bentler, 2012). Alternatively, one can use a matching-based approach to the principal stratification technique proposed by Frangakis and Rubin (2002) in order to estimate the effect of intermediate variables. In this approach, a nonbipartite matching algorithm (Lu et al., 2001) is used to create pairs of individuals such that they are as similar as possible on all variables except a candidate modifiable variable of interest. Computing differences on an outcome measure across the matched pairs, the researcher attempts to isolate the relationship between the outcome and a candidate variable. Such an approach can emphatically not establish causality in the absence of a randomized controlled trial (nor could

SEM), but by isolating variation from baseline covariates and other candidate causal pathways, it may more reliably generate evidence in support of causal pathways of sexual assault.

Study Design

Partners and materials

These data come from a two-arm, parallel, cluster-randomized trial (CRT), where clusters are the schools that serve students from the informal settlements. The treatment is a combination of a girls' intervention, IMPower, and a boys' intervention, Sources of Strength. Participants are adolescent girls and boys in class 6, generally between the ages of 10-14 at baseline.

Data collection for the study began in January 2016 and will continue through December 2018. Baseline data were collected from January through October 2016. The study is being conducted in six informal settlements around Nairobi: Dandora, Huruma, Kariobangi, Kibera, Korogocho, and Mukuru. Due to proximity, and a small number of schools in Kariobangi, all subsequent analysis will classify schools in Kariobangi as being part of Dandora. Two of the five participating organizations are Kenya-based, non-governmental organizations: Ujamaa-Africa (Ujamaa) and the African Institute for Health and Development (AIHD). Ujamaa implemented the intervention and AIHD is the Nairobi-based research partner for the evaluation component of the study. The San Francisco based NGO No Means No Worldwide (NMNW) developed the intervention. Stanford University is the external evaluator of the intervention.

The primary outcome for the CRT is the change in incidence of self-reported rape over the prior 12 months among girls from baseline, compared to a life skills standard of care intervention. Secondary outcomes include determining the impact of the interventions on self-efficacy, self-esteem, gender attitudes and beliefs, and experiences of physical and emotional violence. The data reported in this manuscript were collected before any trainings.

Ethical approval for the study was given by the KEMRI and Stanford IRBs. School administrators determined the suitability of the program for their students. For the evaluation component (i.e., participation in surveys) learners were randomly sampled within school using a predetermined number of colored marbles to indicate selection into the study. The girls were then informed of their rights and potential risks and invited to participate; those that assented into the study then had their parents and caregivers given the opportunity to decline student participation. The trial is registered at ClinicalTrials.gov #NCT02771132. This research was supported by the South African Medical Research Council through the What Works to Prevent Violence Innovation Grant (#52069) as a result of the support obtained from the Secretary of State for International Development at the Department for International Development (DfID). This research was also conducted with Government support under and awarded by DoD, Air Force Office of Scientific Research, National Defense Science and Engineering Graduate (NDSEG) Fellowship, 32 CFR 168a. We are grateful for support from the Marjorie Lozoff Fund, Michelle R. Clayman Institute for Gender Research, Stanford University. Contents are solely the responsibility of the authors and do not necessarily represent the official view of the Secretary of State for International Development at the DfID or the South African Medical Research Council, or any other funding agencies.

Measures

There are few validated scales available for this group (10-14 year old girls living in informal settlements) regarding sexual abuse, and variables on the pathway to sexual assault. Therefore, the study draws from a variety of sources in creating the surveys. In order to measure the primary outcome, change in self-reported incidence of rape over the prior twelve months among girls from baseline, the study creates an index based on several variables. Key surveys that informed these questions included the Kenya VACS (Kenya, 2012) and surveys from the Stepping Stones project in South Africa (Jewkes et al., 2014).

Survey items include experiences of intimate partner violence, or IPV (for example, "In the past 12 months, how many times has your current or a previous boyfriend physically forced you to have sex when you did not want to?"). They also include sexual violence by non-partners ("In the past 12 months, how many times has a man who is NOT your boyfriend forced or persuaded you to have sex against your will?"). Validated, widely used scales are more readily available for secondary outcomes. We used the "Self-Efficacy Questionnaire for Children" (Muris, 2001) for self-efficacy. For experiences of social and emotional violence, we relied on the VACS and Stepping Stones surveys. For gender norms and relations, some questions were removed from existing surveys and modified for this population during piloting, and others were created based on the field team's knowledge of this population, then modified during piloting.

Methods

Adjudication

The baseline survey contains several questions pertaining to rape, covering a range of possible assailants and situations (for example: perpetrator, methods of coercion, threat types, physical assault during encounter, use of alcohol or drugs during assault). The survey also asks the girls how many times they have been assaulted at the end of that line of questioning. Logical inconsistencies are present in some of the student-provided answers, warranting consideration.

Because our primary outcome is rape in the prior 12 months, we focus on surveys in which a girl reports a rape in the past 12 months, but then later reports 0 for how many times she had been assaulted in her life. Such a pattern was present in 70 (23%) of the 298 surveys in which rape was reported in the prior 12 months. It appears that some respondents interpreted the questions as “asked and answered,” meaning that if they answered affirmatively on an earlier question then they removed that experience from consideration when answering future questions that might be overlapping. Given this line of reasoning, we resolved to adjudicate any inconsistent answers into an indication of a rape.

Formalizing the above: we classified a respondent as having been raped in the prior 12 months if a girl gave a nonzero answer to any of seven questions indicating prior-year forced sex. We compared this result against the girl’s response to the question “How many times in your life have you been forced to have sex against your will?” If the results were inconsistent, we counted the girl as having been raped, but flagged the survey as having inconsistent responses.

To empirically validate our adjudication procedure, we analyzed the covariate profiles of girls with conflicting survey responses. A logistic regression model would have been a reasonable option in this binary classification setting. However, because there are many potential interactions among our covariates, we opted instead for a more flexible, less parametric method: the random forest model (Breiman, 2001).We built a random forest predicting the likelihood of rape in the past 12 months, using as predictors all covariates other than reports of forced sexual activity. We refer to this as the “adjudication model.” We fit our model using only fully consistent reports from girls who were (i) raped (n = 228) or (ii) not raped (n = 3,827), excluding the 70 reports with internal inconsistency. Using this random forest, predictions were then generated for those 70 reports as “out of sample” observations.

The random forest model was fit to the data using the randomForest package in R (Liaw and Wiener, 2002). In all, 71 covariates were provided as candidates for the model. When the model was fit to the self-consistent data, the most important variables (as measured by the effect of the variable on out-of-bag prediction error) were whether a girl has been bribed for sex, whether she has had sex, and her average self-efficacy score. To validate the accuracy of the adjudication model, we used a ten-fold cross-validation. The data were split into 10 mutually exclusive equal-sized sub-samples (or "folds") via random permutation. We then looped over each fold, removing it from the dataset and fitting the model to the remaining 90% of the data. The model was used to predict the probability of being raped for each girl in the held-out fold, and the resultant predictions were recorded. This process was coded using built-in R functions. Predictions for each girl were compared against the true reports to assess the model's accuracy.

Bootstrap standard errors

Point estimates of baseline rape prevalence can be estimated using the survey data. As these measurements occurred prior to the intervention, such estimates combine both treatment levels from the study. Within each school, girls were sampled using a simple random sampling procedure for all girls in Class 6 who were in attendance on the day of study initiation. The familiar Wald interval fails to account for the structure of this study's sampling procedure, and thus may produce inaccurate measurements of sampling error. This is because we expect significant homophily within schools; that is, we expect two girls in the same school to be more similar to each other, as compared to two girls from two different schools.

There is a straightforward and intuitive statistical fix in the clustered bootstrap, which is appropriate in our setting because our number of schools exceeds the threshold of thirty described by Cameron et al. (2008). The bootstrap standard error estimate, as opposed to formulaic estimates, is a Monte Carlo estimate: we repeatedly sample individuals within schools with replacement. The number of students from a given school stays the same, but within a school it may be that a student is replicated a few times or not at all. This sample-resample procedure can be used to generate many pseudo-datasets that should have covariate patterns similar to the one we observed in our actual data set, because we imposed the clustered structure. Observing the proportion of girls who have been raped in each of these bootstrapped, pseudo-datasets yields a sampling distribution for the observed proportion.

Assessing candidate causal pathways

The dataset being analyzed was collected at one point in time (“cross-sectionally”) rather than at several time points (“longitudinally”). It is challenging to use such data to obtain causal insights without researcher-directed randomization. Researchers may be tempted to use SEM to tease out causal relationships, but there are philosophical as well as practical reasons to caution against this approach. SEM requires researchers to posit a model for all of the observed and latent variables, yet the paucity of causal research in this area indicates that such a model would be guesswork at best. Furthermore, there are substantive technical concerns over the possibility of omitted variable bias and the challenges posed by high multicollinearity among many of our measured variables. Thus we do not attempt an analysis that establishes causal links between proposed causal pathways and the outcome of rape in the prior 12 months. Instead, we attempt to isolate variation in the outcome as much as possible so as to inform the relative contribution of candidate causal pathways. At the end of the CRT, there will be three time points measured for participants, and similar causal pathway analyses will be performed.

Based on theoretical models of prevention, as well as findings in prior literature, we identified four candidate modifiable variables that may prevent rape: former experience of IPV, generalized self-efficacy, prior alcohol use, and belief in gender norms. These are girl-specific experiences and beliefs, not measures of the entire group. Self-efficacy was estimated using the girl’s average response across all 24 self-efficacy questions measured in the survey, spanning the domains of social, academic, and emotional self-efficacy. Higher scores represented greater self-efficacy. Gender-normative thinking was estimated by the girl's average response across 12 questions about the relationships between women and men, with higher scores representing greater gender-normative thinking. Alcohol use was estimated using a girl's self-reported frequency of consuming alcohol. IPV was estimated as the average of a girl's response across three questions about whether she had been slapped, insulted, or intimidated by a boyfriend.

There are three types of variables in this analysis: (i) covariates – i.e., socioeconomic indicators, parental violence, whether parents are alive or dead, and intimate partner status; (ii) candidate causal pathways – i.e., prior IPV, alcohol and drug use, generalized self-efficacy, and belief in gender norms; and (iii) outcome – i.e., whether a girl has been raped in the prior 12 months. Because there are four candidate modifiable variables under consideration, we perform four separate analyses. For each candidate modifiable variable of interest, girls were matched one-to-one based on a list of covariates. These included the three other pathway variables; a set of variables reflecting a girl's experience of parental violence, whether her parents were alive or dead, her socioeconomic status, and whether or not she had a boyfriend.

For the matching variables, missing values were imputed with the mean value across the girls. A missingness indicator was generated for each of the variables, with a large value (e.g. 100) denoting a missing value, and 0 denoting otherwise. These missingness indicators were also provided as matching variables to encourage matches between girls missing the same variables. For the self-efficacy and gender norms measures, which were composites of many questions, missingness was defined as lacking a response to more than one of the relevant questions. Girls were dropped from each analysis if they were missing a value for the associated pathway variable; this resulted in a loss of approximately 7% of the girls for each pathway explored.

For example, to isolate the effect of prior IPV on probability of rape in the past 12 months, we attempted to match two girls who were nearly identical on all covariates as well as identical on alcohol and drug use, generalized self-efficacy, and belief in gender norms, and the associated missingness indicators. To measure similarity between each pair of girls across these variables, we used the Mahalanobis distance (Mahalanobis, 1927). Simultaneously, we tried to make the two matched girls as different as possible in their prior experiences of IPV. To do so, we penalized candidate pairs that had similar values of this pathway variable.

This procedure produced a matrix describing the suitability of all potential matched pairs. This matrix was then used to generate the set of matched girls. Within a given matched pair, the two girls were nearly identical except for their experiences of prior IPV – and the girl with the higher value was considered "relatively exposed" versus the other girl within the pair. Once the matches were computed, balance on the match variables and the pathway variable were assessed. We then computed the difference between the two groups' probabilities of experiencing rape in the prior twelve months. These differences were normalized by the amount of separation we achieved between the groups on the pathway variable (IPV), as measured in standard deviations. We emphasize that this is purely for purposes of comparability across variables, as it is highly unlikely that a causal effect of one standard deviation would be achieved in practice.

We repeated the analysis described in the previous paragraphs three more times – for alcohol and drug use, generalized self-efficacy, and belief in gender norms. The matching algorithm was implemented using the nbpMatching package in R (Beck et al., 2016). Balance was assessed using standardized mean differences, with values less than 0.10 deemed acceptable.

Assessing and proposing predictors of school-level prevalence of rape

Two analyses are performed to provide insight on which kinds of schools interventionalists may want to approach first with the prevention program. That is, we try to find the best predictors of school-level rape prevalence. The first analysis uses easily accessible school-level covariates (e.g., number of students, test scores, building materials) to predict the proportion of girls at the school who were raped in the prior 12 months. The second analysis starts with individual girls' responses, aggregates these responses to school-level variables, and then predicts. The second analysis acknowledges that the questions asked of individual-level girls are much more private and difficult to obtain from school administrator before building trust within the school. But, the analysis of these aggregated responses – and their predictive power – may be useful for developing new questions interventionalists can ask of the administrators to assess the school's prevalence of rape before entering the school.

To investigate the associations between school-level covariates and school-level rape prevalence, univariate linear regressions were computed, in which the sexual assault frequency was the dependent variable and each covariate the independent variable in its own model. The outcome is continuous, but it is constrained to the interval [0, 1], which violates the assumption of Gaussian error. Nonetheless, as our goal is simply to find univariate associations for further study, we opt for the simplicity of linear regressions over more complex generalized linear models. Sixteen regressions were fit in total, using the standard lm() function in R. For numeric variables, the reported p-value is that of the t-test for the coefficient. For categorical variables, the reported p-value is that of the F-test for all levels of the variable. Coefficient signs were also reported for all numeric variables. All regressions were weighted by the number of girls interviewed at the school, down-weighting the influence of extreme values realized due to small sample sizes for small schools. The results were not corrected for multiple comparisons, though the majority of covariates were nonetheless not significant at the uncorrected $p = 0.10$ level.

An identical approach was used to understand the relationship between aggregated girl-level covariates and rape frequency. Covariates observed at the individual level (such as whether each girl has a boyfriend) were aggregated to school-level averages (such as “percentage of girls who have a boyfriend”). Univariate regressions were then fit using school-level prior-year rape frequency as the dependent variable and each aggregated covariate as the independent variable.

Lastly, to understand the relative predictive ability of these variables conditional on the presence of the other variables, we fit a LASSO model (Tibshirani, 1996) to the union of the school-level and aggregated individual-level covariates. The model was fit using the cv.glmnet() function in the glmnet package in R (Friedman et al., 2010). Lambda, the regularization penalty, was selected to minimize cross-validated error. Such a procedure generates parsimonious models with few nonzero coefficients, and can be viewed as an alternative to a stepwise variable selection procedure. Hence, the variables selected for model inclusion tend to be those most predictive when accounting for the effect of the other variables. Using such a model allows us to get a sense of which variables remain predictive even in the presence of other predictors.

Results

Adjudication

The accuracy of the adjudication model is summarized in the ROC curve given in Figure 1, generated with the roc() function from the pROC library (Robin et al., 2011). The curve is a common performance measure in binary classification problems. For any classifier, the ROC plots the false positive rate (i.e. the percent of the girls classified as raped who were not actually raped) that must be tolerated in order to achieve a desired true positive rate (i.e. the percent of all girls who were raped who are identified as such by the classifier). If the ROC curve is arced high above the 45-degree line, this indicates good performance, summarized by an AUC (“area under curve”) near 100%. In our case, the AUC is 84.11%, indicating strong predictive performance.

Now we apply our adjudication model (the random forest fit to all the consistent data) to the 70 girls with conflicting reports. Across the girls, we get an average prediction of 17.0% probability of rape in the prior 12 months. We ask: does this prediction make them more like the group of girls with non-conflicting reports of rape in the prior 12 months – or more like the group of non-victims? To answer this question, we compare against the mean predicted probabilities for each group, as generated in our cross-validation procedure. Note that we must compare against the cross-validated probabilities (rather against the predictions on each group using the model fit to all the consistent data) in order to make a fair comparison. Otherwise, we would be comparing within-sample predictions against out-of-sample predictions.

This approach gives an average out-of-sample predicted probability of rape of 19.9% for the girls who had non-conflicting reports of rape in the prior 12 months. For girls who had non-conflicting reports and did not report a rape in the prior 12 months, the corresponding average is 5.5%. By this measure, it is clear that the girls with conflicting reports are much more like the prior-year rape victims than the non-victims. Restating the above: the average prediction for the girls who gave contradictory prior-year rape reports is far closer to the average prediction for the girls with non-contradictory prior-year rape reports than to girls who were not raped in the prior 12 months. This lends credence to our adjudication choice.

Our adjudication procedure is consequential insofar as it alters the baseline estimate of rape prevalence over the prior 12 months. All girls who reported rape in the past 12 months were designated as victims, even if their survey responses were self-contradictory. This procedure yields a baseline estimate that 7.2% of girls were raped in the prior 12 months. Had only non-conflicting reports been counted, the estimate would have been notably lower: 5.5%.

Bootstrap standard errors

The effect of using bootstrap standard errors can be seen in our uncertainty estimates. We use 500 bootstrap replicates. Ninety-five percent confidence intervals for a selected set of variables can be found in Table 1. In most cases, the bootstrap confidence intervals are similar in width to the Wald Intervals, though they are somewhat wider in the case of measuring the proportion of girls who have had a boyfriend. The bootstrap intervals are also not necessarily symmetric about our point estimate, while the Wald intervals are symmetric by construction.

Assessing candidate causal pathways

We interrogated candidate modifiable causal variables, making use of the nonbipartite algorithm discussed previously. These analyses reduce differences in all variables except the candidate modifiable variable of interest, attempting to isolate the relationship between the outcome and this variable. We explored four potential causal pathways: the effect of a girl's self-efficacy, gender-normative thinking, alcohol use, and experience of intimate partner violence.

Covariate balance was quite good in all four analyses; all covariates had absolute standardized mean differences of less than 0.10. Balance tables are reported in the supplement.

Results are summarized in Figure 2. As we can see, intimate partner violence and alcohol use were associated with a higher proportion of girls being victimized in the prior 12 months, with "effect" sizes of about 3.5% and 2.5% per standard deviation of separation. A smaller effect size was associated with self-efficacy; a one standard deviation increase in self-efficacy was associated with about a 1% reduction in the proportion of girls raped in the prior 12 months. The association between gender-normative thinking and rape frequency was positive but small in magnitude. It may be the case that shifting gender norms within an individual, without shifting gender-norms for the surrounding community, does not have a profound effect on outcomes.

We were able to force separation of standardized mean differences for prior IPV, generalized self-efficacy, alcohol use, and gender norms of 0.48, 2.66, 0.67, and 2.34

respectively. These values can be read off of the balance tables (found in the supplement). Future interventionalists would need to assess how realistic it is for a given intervention to shift the candidate modifiable variable. It should be repeated that these are exploratory analyses and no claim of establishing causality is being made here. Though it is possible that the direction of causality is in the reverse of how it is discussed above, this analyses provides some empirical evidence to guide prioritization of which pathway variables to explore modifying.

Assessing and proposing predictors of school-level prevalence of rape

One trend immediately evident in the data is the high degree of variation in rape rates across the schools. We analyze 90 of the 95 schools – only the schools for which we have full school-level covariate data. At eleven of these schools, none of the girls reported being raped in the past 12 months, while at six of the schools, more than 20% of the interviewees reported being raped in the past 12 months. The overall distribution is shown in the left panel in Figure 3.

Certainly, some of this variation is attributable to the low number of girls interviewed at certain schools. With few interviewees, it is more likely to observe an extremely high or extremely low proportion of girls who report having been raped in the past year. If we exclude all schools with fewer than 20 interviewees, the most extreme values disappear from the distribution (as can be seen in the right panel in Figure 3). But substantial heterogeneity remains.

This heterogeneity is significant from a public health perspective. If the intervention is found to be effective at reducing the incidence of rape, then schools with higher rates of rape should be prioritized first for deployment of the intervention. But, as was mentioned earlier, there remains a technical challenge in obtaining estimated rates of rape or even reliable predictors at a given school, as this information is frequently sensitive to obtain and requires investments in trust and relationship-building at a given institution. In the interest of more effective triaging of resources, we explore the relationship between rape frequency and several easier-to-obtain administrative variables. A full list of the variables and how they were computed can be found in the appendix. To understand the variation in rate of rape from school-to-school, we analyzed the associations between each of our covariates and the prior-year sexual assault frequency at each school, using techniques described in the Methods section. These associations were computed via univariate linear regressions, where the prior rape frequency is the dependent variable and each covariate the independent variable in its own. Results are given in Table 2.

The relative dropout rate was significant at the $p = 0.03$ level, indicating a possible association. Relative dropout rate tracks the change in gender ratios within class, across classes 5-8. If a school has more boys relative to girls in class 8 than in class 5, then we say that the "relative dropout rate" is higher for girls. This variable is intuitively appealing, as lower retention of girls relative to boys may reflect underlying issues in a community – which may, in turn, affect the prevalence of sexual assault. But alternative explanations are also possible, including a "reversed" causal pathway whereby high rates of rape discourage girls from attending school, leading to higher dropout rates. This is not an extraordinarily strong predictor, and will require replication in future settings before it is considered for use in triaging intervention into schools. No other variables were strongly associated with prior-year rape frequency.

We hypothesized that some information that was harder to collect – only obtained from surveys completed by the girls, and then aggregated to the school level – would be predictive of the school-level rate of rape in the prior 12 months. This was borne out by the results. Aggregating individual-level covariates, we found that many of these were more predictive than the school-level covariates, with 37 of 127 aggregated covariates significant at the 0.05 level (though many of these variables are redundant). Variables concerning rates of intimate partner violence comprised five of the 10 variables most associated with rates of rape. The proportion of girls having a boyfriend was also positively associated with rape frequency at a school.

Somewhat counter intuitively, the proportion of girls not taught a No Means No Worldwide skill was associated with lower rape frequency at a school. This may indicate that trainers at our partner NGO Ujamaa have historically selected schools for training where they perceive the girls to be at greater risk of assault. If so, this indicates that a degree of "intervention triaging" is already occurring. We intend to follow up with trainers to learn whether there are other obtainable variables that they use in order to target at-risk schools.

To explore the effect of each variable conditional on the others, a LASSO was fit via the procedure described in the Methods section. The model is shown in Table 3. Note that all variables were standardized, so the coefficients are directly comparable. The resulting model is parsimonious, including just eight variables – all of which are aggregations of individual-level covariates, rather than school-level covariates. This further supports the idea that these variables are much stronger predictors of rape frequency. The selected variables include several measures of whether girls have been taught No Means No Worldwide skills, and several indicators of an abusive relationship with a boyfriend. The latter group further substantiates the view that there is a strong relationship between frequency of intimate partner violence and frequency of rape.

It is possible that school administrators could collect proxies for some of these variables, such as the estimated proportion of girls who have a boyfriend. This analysis may allow future interventionalists to target schools beyond the administrative variables discussed earlier.

Discussion

As the study of sexual assault prevention develops, the literature must advance statistical practices commensurate with the complexity and challenges of the data. It is crucially important to provide unbiased estimates of baseline rape prevalence in a given area, and valid confidence intervals around these estimates. In turn, academics can assist interventionalists by helping (i) to direct energies to the most in need ("triaging"), and (ii) developing evidence about what aspects of interventions are most beneficial for reducing rates of sexual assault ("causal pathways").

This paper aims to document analytical practices for the goal of transparency and improved statistical rigor. We have proposed both a particular approach and a novel validation procedure for data adjudication. Documenting our methods at this time will be useful, as similar issues with data inconsistency will likely manifest at the midline and endline of our ongoing study. Moreover, as conflicting survey responses are a common issue in the study of sexual assault, our validation procedure could be adapted by other researchers in this area.

While we can make no claims of causality from this baseline data, our pathways analysis yields insights that also can be compared with future results from the RCT. We find that, among otherwise-similar girls, those facing intimate partner violence or using alcohol have substantially higher rates of rape in the prior twelve months. To a lesser degree, those with high self-efficacy report lower rates of victimization. Our approach of pairing similar girls who differ only on the candidate pathway is also potentially appealing to researchers. It is both intuitively sensible and reduces many of the assailable parametric assumptions imposed by structural equation modeling.

Lastly, we take an empirical approach to developing school-level measurements to better triage interventions. We think it is of particular importance to separate variables based on ease of accessibility to researchers, and to assess the relative predictive value of easily obtainable variables versus those requiring live interviews. If aggregated summary statistics of questions currently asked in live interviews are highly predictive of the outcome of interest then it may be worthwhile to develop new questions that can be asked of administrators.

References

Anderson, J. C. and Gerbing, D. W. (1988). Structural equation modeling in practice: A review and recommended two-step approach. *Psychological Bulletin*, 103(3):411.

Baiocchi, M., Nyamongo, M., Oguda, G., Akinyi, D., Friedberg, R., Rosenman, E., Omondo, B., and Sarnquist, C. (2018). Prevalence and risk factors for sexual assault among class 6 students in unplanned settlements of Nairobi, Kenya: Methods from the Impower & Sources of Strength cluster randomized controlled trial.

Baiocchi, M., Omondi, B., Langat, N., Boothroyd, D. B., Sinclair, J., Pavia, L., Mulinge, M., Githua, O., Golden, N. H., and Sarnquist, C. (2017). A behavior-based intervention that prevents sexual assault: The results of a matched-pairs, cluster-randomized study in Nairobi, Kenya. *Prevention Science*, 18(7):818–827.

Beck, C., Lu, B., and Greevy, R. (2016). *nbpMatching: Functions for optimal non-bipartite matching*. R package version 1.5.1.

Breiman, L. (2001). Random forests. *Machine Learning*, 45(1):5–32.

Brener, N. D., McMahon, P. M., Warren, C. W., and Douglas, K. A. (1999). Forced sexual intercourse and associated health-risk behaviors among female college students in the united states. *Journal of Consulting and Clinical Psychology*, 67(2):252.

Cameron, A. C., Gelbach, J. B., and Miller, D. L. (2008). Bootstrap-based improvements for inference with clustered errors. *The Review of Economics and Statistics*, 90(3):414–427.

Cook, S. L., Gidycz, C. A., Koss, M. P., and Murphy, M. (2011). Emerging issues in the measurement of rape victimization. *Violence Against Women*, 17(2):201–218.

Cronbach, L. J. and Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52(4):281.

Davis, K. C., Gilmore, A. K., Stappenbeck, C. A., Balsan, M. J., George, W. H., and Norris, J. (2014). How to score the sexual experiences survey? A comparison of nine methods. *Psychology of Violence*, 4(4):445.

Efron, B. (1992). Bootstrap methods: another look at the jackknife. In *Breakthroughs in Statistics*, pages 569–593. Springer.

Ellsberg, M., Arango, D. J., Morton, M., Gennari, F., Kiplesund, S., Contreras, M., and Watts, C. (2015). Prevention of violence against women and girls: what does the evidence say? *The Lancet*, 385(9977):1555– 1566.

Fisher, B. S. and Cullen, F. T. (2000). Measuring the sexual victimization of women: Evolution, current controversies, and future research. *Criminal Justice*, 4:317–390.

Frangakis, C. E. and Rubin, D. B. (2002). Principal stratification in causal inference. *Biometrics*, 58(1):21– 29.

Friedman, J., Hastie, T., and Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, 33(1):1–22.

Gail, M. H., Mark, S. D., Caroll, R. J., Green, S. B., and Pee, D. (1996). On design considerations and randomization-based inference for community intervention trials. *Statistics in Medicine*, 15(11):1069–1092.

Gidycz, C. A., Orchowski, L. M., and Berkowitz, A. D. (2011). Preventing sexual aggression among college men: An evaluation of a social norms and bystander intervention program. *Violence Against Women*, 17(6), 720-742.

Jewkes, R., Gibbs, A., Jama-Shai, N., Willan, S., Misselhorn, A., Mushinga, M., Washington, L., Mbatha, N., and Skiweyiya, Y. (2014). Stepping stones and creating futures intervention: shortened interrupted time series evaluation of a behavioural and structural health promotion and violence prevention intervention for young people in informal settlements in Durban, South Africa. *BMC Public Health*, 14(1):1325.

Kenya, V. (2012). Violence against children in Kenya: Findings from a 2010 national survey. *Nairobi: UNICEF, Centers for Disease Control and Prevention & Kenya National Bureau of Statistics.*

Koss, M. P., Abbey, A., Campbell, R., Cook, S., Norris, J., Testa, M., Ullman, S., West, C., and White, J. (2007). Revising the SES: A collaborative process to improve assessment of sexual aggression and victimization. *Psychology of Women Quarterly*, 31(4):357–370.

Koss, Mary P.; Rozee, Patricia D. (December 2001). Rape: a century of resistance. *Psychology of Women Quarterly*. Sage. 25 (4): 295–311. doi:10.1111/1471-6402.00030

Liaw, A. and Wiener, M. (2002). Classification and regression by randomforest. *R News*, 2(3):18–22.

Livingston, J. A., Hequembourg, A., Testa, M., and VanZile-Tamsen, C. (2007). Unique aspects of adolescent sexual victimization experiences. *Psychology of Women Quarterly*, 31(4):331–343.

Lu, B., Zanutto, E., Hornik, R., and Rosenbaum, P. R. (2001). Matching with doses in an observational study of a media campaign against drug abuse. *Journal of the American Statistical Association*, 96(456):1245– 1253.

Mahalanobis, P. C. (1927). Analysis of race-mixture in Bengal. *Journal of the Asiatic Society of Bengal*, 23, 301-333.

Muris, P. (2001). A brief questionnaire for measuring self-efficacy in youths. *Journal of Psychopathology and Behavioral Assessment*, 23(3):145–149.

Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.-C., and Mller, M. (2011).

proc: an open-source package for r and s+ to analyze and compare roc curves. *BMC Bioinformatics*, 12:77.

Rosenberg, M. (2015). *Society and the adolescent self-image*. Princeton university press.

Russell, M., Cupp, P. K., Jewkes, R. K., Gevers, A., Mathews, C., LeFleur-Bellerose, C., and Small, J. (2014). Intimate partner violence among adolescents in Cape Town, South Africa. *Prevention Science*, 15(3):283–295.

Shamu, S., Gevers, A., Mahlangu, B. P., Jama Shai, P. N., Chirwa, E. D., and Jewkes, R. K. (2015). Prevalence and risk factors for intimate partner violence among grade 8 learners in urban South Africa: baseline analysis from the Skhokho Supporting Success cluster randomised controlled trial. *International health*, 8(1):18–26.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 267–288.

Ullman, J. B. and Bentler, P. M. (2012). Structural equation modeling. *Handbook of Psychology, Second Edition*, 2.

UNICEF, O. et al. (2014). Hidden in plain sight: a statistical analysis of violence against children. 2014. *Nova York*.

Vuchinich, S., Flay, B. R., Aber, L., and Bickman, L. (2012). Person mobility in the design and analysis of cluster-randomized cohort prevention trials. *Prevention Science*, 13(3):300–313.

Appendix

Variable definitions

The cluster-level covariates are summarized into two types of covariates. The first group of covariates was collected pre-randomization. There are 10 of these variables, six of which are numeric and four of which are categorical. The numeric variables are number of total classrooms, number of girls' toilets, number of total students, number of boys' toilets, the school's prior 12 months mean standardized test score, and number of total teachers. The categorical variables are summarized below.

- **Area**: the neighborhood in which the school is located, taking on five distinct values (Dandora, Huruma, Kibera, Korogocho, Mukuru)
- **Roof**: the material from which the roof is constructed, taking on seven distinct values (Ironsheets, Bricks, Tiles, Tiles & Asbestos, Asbestos, Concrete, Tiles and Concrete)
- **Floor**: the material from which the floor is constructed, taking on six distinct values (Cemented, Tiles, Concrete, Mud, Non-Cement, Earthen)
- **Sponsor**: the sponsor of the school, taking on four distinct values (Private, Government, Church/mission, Others)

Additionally we have six covariates that are composites of continuous covariates. We considered these variables because the give more precise statements of the relative resources available in the schools.

- **Relative dropout rate**: gender ratio (number of boys divided by number of girls) in grade 5 at a school, divided by gender ratio in grade 8. This value will be lower if the ratio increases through the grades, indicating that girls drop out at a higher frequency than boys.

- **Students per teacher**: total number of students at a school, divided by total number of teachers at that school
- **Boys' toilets per boy**: total number of boys' toilets at a school, divided by total number of boys at that school
- **Girls' toilets per girl**: total number of girls' toilets at a school, divided by total number of girls at that school
- **Gender ratio**: total number of boys at a school divided by total number of girls at the school
- **Students per classroom**: total number of students at a school, divided by total number of classrooms at that school Summary statistics for each of the numeric covariates are provided below.

Summary statistics for each of the numeric covariates are provided in the supplement.

Tables

Table 1

Point estimates and confidence intervals, using both Wald and bootstrap procedures

Quantity	Estimate	Wald 95% CI	Bootstrap 95% CI
Proportion of girls who have been forced to have sex	9.99%	(9.07%, 10.90%)	(9.10%, 10.80%)
Proportion of girls who have been forced to have sex in the past 12 months	7.22%	(6.43%, 8.01%)	(6.44%, 8.02%)
Proportion of girls who have had a boyfriend in last 6 months	21.31%	(20.00%, 22.57%)	(19.83%, 23.84%)
Proportion of girls who have consumed alcohol	11.60%	(10.62%, 12.57%)	(10.61%, 12.56%)
Proportion of girls who have tried drugs	0.66%	(0.41%, 0.90%)	(0.41%, 0.90%)
Proportion of girls who have had sex	7.37%	(6.57%, 8.16%)	(6.67%, 8.14%)

Table 2

Assessment of school-level prior-year rape predictors

Covariate	p-Value	Sign
Relative dropout rate	0.03	+
Students per classroom	0.13	+
Mean Test Score	0.26	–
Roof	0.30	NA
Girls’ toilets	0.44	–
Floor	0.51	NA
Area	0.54	–
Girls’ toilets per girl	0.55	–
Total students	0.58	+
Boys’ toilets	0.59	–
Total teachers	0.65	–
Total classrooms	0.68	–
Students per teacher	0.69	+
Boys’ toilets per boy	0.81	–
Sponsor	0.83	NA
Gender ratio	0.93	–

Table 3

LASSO model coefficients for union of school-level and aggregated individual-level predictors

Covariate	Coefficient
(Intercept)	–0.115
Proportion of girls who don't know if they have taken No-Means-No Worldwide	0.026
Proportion of girls who have not been taught a No-Means-No skill	–0.030
Proportion of girls who don't know if they have been taught a No-Means-No skill	0.004
Proportion of girls who do not have a boyfriend	–0.451
Proportion of girls who have been slapped by their boyfriends	0.063
Proportion of girls who have been threatened with a weapon by their boyfriends	0.183
Average self-reported ability to chat with an unfamiliar person	0.011
Average self-reported distance between girls' homes and sources of water	0.009

Figure Captions

Figure 1. ROC curve for the adjudication model, built using a random forest model.

Figure 2. Difference in percent of girls raped among girls paired via the bipartite matching algorithm. Distance is induced only on the candidate variable. Girls are similar in other respects.

Figure 3. Percentage of girls who have been raped across schools in the study. The right panel filters to schools with over 20 girls interviewed.

Figures

Figure 1

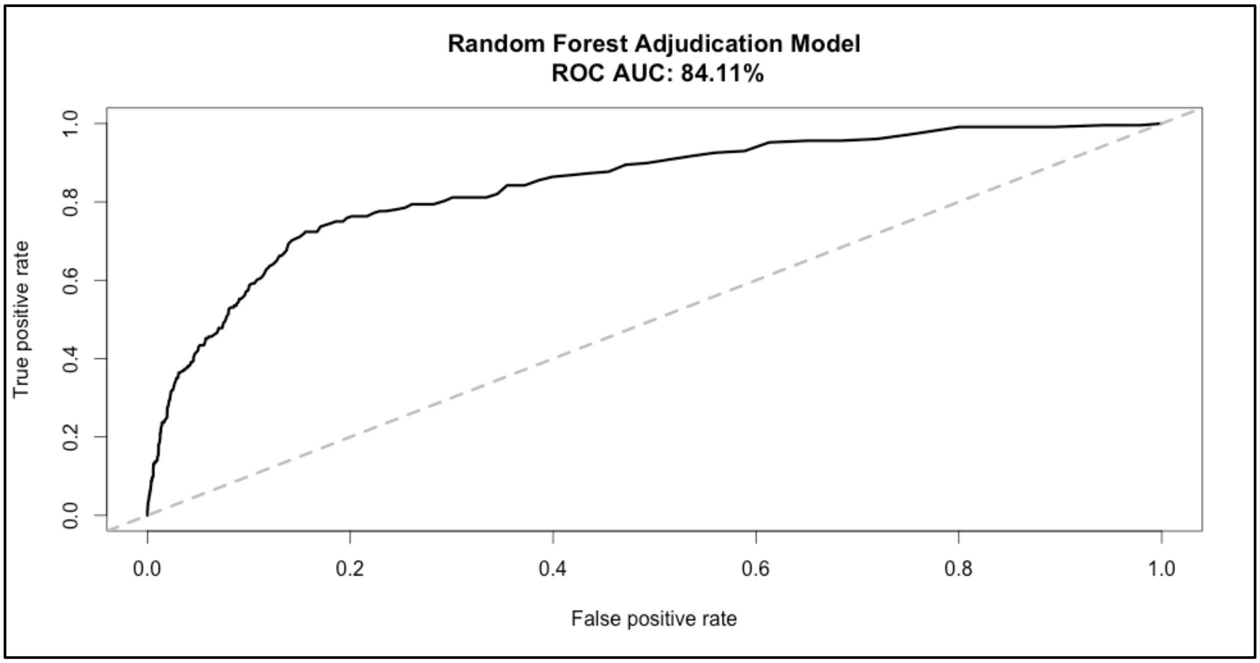


Figure 2

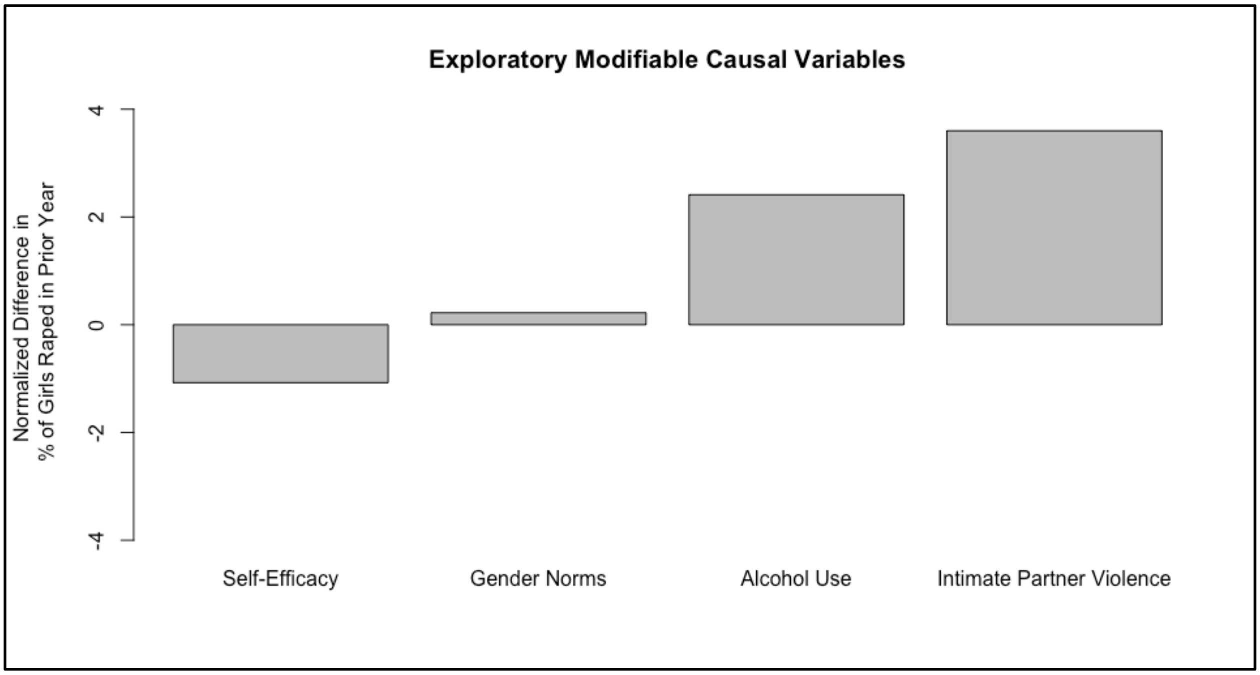
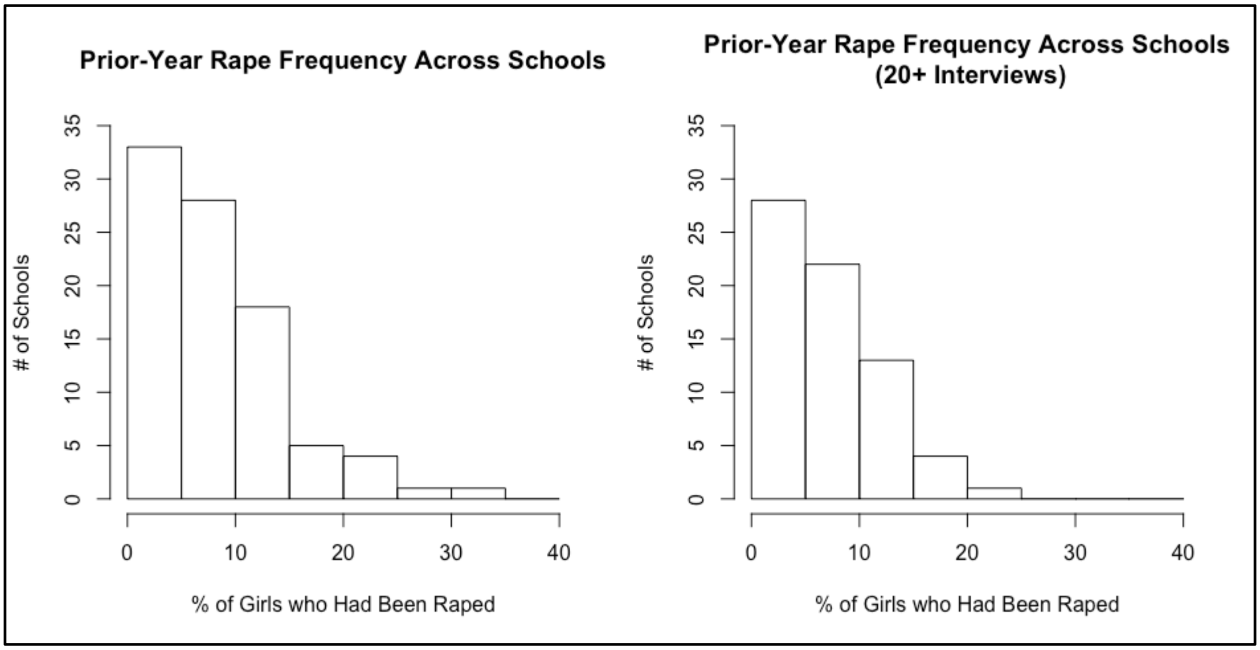

Exploratory Modifiable Causal Variables
Normalized Difference in
% of Girls Raped in Prior Year
4
2
0
-2
-4
Self-Efficacy
Gender Norms
Alcohol Use
Intimate Partner Violence

Figure 3


Prior-Year Rape Frequency Across Schools
# of Schools
0
5
10
15
20
25
30
35
0
10
20
30
40
% of Girls who Had Been Raped
Prior-Year Rape Frequency Across Schools
(20+ Interviews)
# of Schools
0
5
10
15
20
25
30
35
0
10
20
30
40
% of Girls who Had Been Raped